Lecture Notes in Electrical Engineering 173

Wei Deng
Editor

Future Control and Automation

Proceedings of the 2nd International Conference on Future Control and Automation (ICFCA 2012) - Volume 2

Springer

For further volumes:
http://www.springer.com/series/7818

Wei Deng (Ed.)

Future Control and Automation

Proceedings of the 2nd International Conference on Future Control and Automation (ICFCA 2012) - Volume 2

Springer

Editor
Wei Deng
Shenzhen University
Guangdong
P.R. China

ISSN 1876-1100 e-ISSN 1876-1119
ISBN 978-3-642-31002-7 e-ISBN 978-3-642-31003-4
DOI 10.1007/978-3-642-31003-4
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012939656

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

2012 2nd International Conference on Future Control and Automation (ICFCA 2012) will be the most comprehensive conference focused on the various aspects of advances in Future Control and Automation. The goal of this conference is to bring together the researchers from academia and industry as well as practitioners to share ideas, problems and solutions relating to the multifaceted aspects of Future Control and Automation. It aims to create a forum for scientists and practicing engineers throughout the world to present the latest research findings and ideas in the areas of control and automation.

ICFCA 2012 will be held on July 1–2, 2012, Changsha, China. Changsha is the capital city of Hunan Province, situated in the river valley along the lower reaches of the Xiang River, a branch of the Yangtze River. The recorded history of Changsha can be traced back 3,000 years. Tomb relics from the primitive periods witnessing the earliest human activities have been discovered in this region. The village of Shaoshan, about 130 kilometers south-west of Changsha is the hometown of Chairman Mao Zedong. Today, the village has become a memorial place for Chinese people to remember this extraordinary man. People erected a statue of the Chairman and have preserved the houses he lived as a tourist site. The city is now a major port and a commercial and industrial center, is one of China's top 20 "economically advanced" cities.

ICFCA 2012 provides an interdisciplinary forum in which researchers and all professionals can share their knowledge, experience and report new advances on Control and Automation. It is a medium for promoting general and special purpose tools, which are very essential for the evolution of conversional and new subjects of Control and Automation.

Future control and automation is the use of control systems and information technologies to reduce the need for human work in the production of goods and services. In the scope of industrialization, automation is a step beyond mechanization. Whereas mechanization provided human operators with machinery to assist them with the muscular requirements of work, automation greatly decreases the need for human sensory and mental requirements as well. Control and automation plays an increasingly important role in the world economy and in daily experience.

The editors were responsible for selection of reviewers for the papers, whom we thank for their promptitude and accurate work. The communication between authors and referees was managed by the editors. The organizing committee expresses its entire gratitude to all the authors who presented their works at ICFCA 2012 and contributed in this way to the success of this event. Special thanks are due to the authors from abroad for attending the conference and to the reviewers for their support in improving the quality of the papers and finally for the assurance the quality of this volume.

Wei Deng

Committee

Keynote Speech and Honorary Chairman

Jeng-Shyang Pan	Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan IEEE Fellow, IEEE Tainan Section SP Chapter Chair

General Chairs

Jin Wen	Chongqing University, China
Wei Lee	Melbourne ACM Chapter, Australia

Program Chairs

Wang Yu	Dalian Maritime University, China
Wensong	Shi, Dalian Maritime University, China

Publication Chair

Wei Deng	Shenzhen University, China

Local Chair

Zhang Bin	Hunan Normal University, China

Committee Members

Junwu Zhu	Western Australia University, Australia
Wei Deng	American Applied Sciences Research Institute, USA
Riza Esa	Kuala Lumpur ACM Chapter, Malaysia
Wei Lee	Melbourne ACM Chapter, Australia

David Chan	Macao ACM Chapter, Macao
Yuren Du	Yangzhou University, China
Zhibo Guo	Yangzhou University, China
Fen Wang	Jiangsu University of Science and Technology, China
Yongan Xu	Yangzhou University, China
Yongzhong Cao	Yangzhou University, China
Guansheng Zhen	Jiangsu University of Information and Engineering, China
Jiang Yi	Jiangsu University of Science and Technology, China
Hongyan Qi	Heilongjiang University of Science and Technology, China

Contents

Mathematical Modeling, Analysis and Computation

Control Engineering

Reliable Networks Design

Vehicular Communications and Networking

Automation and Mechatronics

A Construction Algorithm of Target Model for Moving Object

Junxiang Gao[1,2], Hao Zhang[2], and Yong Liu[2]

[1] College of Science, Huazhong Agricultural University, Wuhan 430070, China
gao200@gmail.com
[2] School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract. The target model is generally constructed manually in object tracking methods; moreover, current automatic construction technologies have some inherent drawbacks. Aiming at these problems, an automatic and accurate construction algorithm of target model is presented in this paper. At first, the moving object number and position are computed using connected components labeling algorithm. Subsequently, minimal circumscribed rectangles of the objects are obtained as the target models according to radius of gyration tensor method. Experiment results show that the presented method can compute the target models accurately, and the performance of the method outperforms projection method obviously.

Keywords: computer vision, target tracking, intelligent video surveillance, target model.

1 Introduction

Moving object tracking is an important component in many applications, such as video conferencing, distance learning, smart rooms, and video surveillance. Quite a few tracking algorithms have been proposed including the gradient descent method, the mean shift method, Kalman filter framework and particle filter framework. In these algorithms, a target model should be firstly defined in a reference image and then searched for in subsequent frames using a function that evaluates the similarity between the model and a candidate [1-3]. In current research, the construction of target model is generally performed manually, i.e., drawing a rectangle or an ellipse enclosing the target in a reference frame as the model [4-6]. Although automatic initialization methods are used in a few literatures, prior target knowledge is usually needed and thus the universality is reduced. For example, the approach proposed in [7] is only applicable for face tracking due to the use of color information. The projection method in [8, 9] is suitable for simple scene only; however the number, position or size of the targets may not be correctly computed if the projections of moving objects are overlapped.

To resolve these problems, we propose an automatic and accurate construction method of target model. Connected components labeling algorithm and radius of gyration tensor method are respectively used to compute the minimal circumscribed rectangles of the targets. The detailed steps are described in the following sections.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 1–7.
springerlink.com

2 The Analysis of Projection Method

In practice, the procedure of moving objects detection has been finished before target tracking, thus we suppose that the mask of moving objects is binary image $M(x, y)$.

$$M(x, y) = \begin{cases} 1 & (x, y) \in \{\text{moving object pixels}\} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

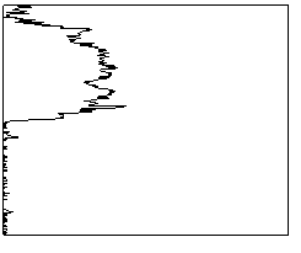
where (x, y) represent the position of moving object pixel. Image $M(x, y)$ can be obtained by any algorithm of moving object detection, for example, optical flow method, frame difference method, background subtraction method, and so on.

For a binary image $M(x, y)$, the vertical projection of every column is defined by the pixel amount where $M(x, y)=1$, and the horizontal projection of every rol is defined by the pixel amount where $M(x, y)=1$. In projection method, the vertical projection of $M(x, y)$ is firstly computed to get the amount, widths and horizontal positions of the targets. Consequently, the heights of the targets are determined by horizontal projection of binary image $M(x, y)$. The results of these two steps are shown in Fig. 1.

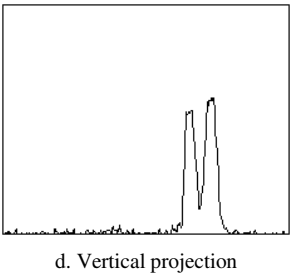
The drawback of projection method can be briefed as follows. For one thing, if there is more than one target in the scene and their projection is overlapped, the amount, positions and sizes of these targets can not be accurately computed according to the projection information. For another, the target models of projection method are upright circumscribed rectangles; therefore many background pixels will be included in the target models if the targets are not upright in the image.

a. Origin image

b. Binary image



c. Horizontal projection



d. Vertical projection

Fig. 1. The illustration of projection method

To conquer the problems above mentioned, two measures are taken in this paper. In the beginning, the uncertainties caused by overlapped projection are inherent defect of the project method. Thereby another solution, called connected components labeling algorithm, is adopted to get the amount, positions and sizes of the targets in this paper. Consequently, both upright circumscribed rectangle and minimal circumscribed rectangle can enclose all the target pixels. The projection method using the former, however, the latter can describe the target more accurately because of less background pixels in it. So we choose the latter to be the target model.

3 Automatic Initialization Method of Target Model

3.1 Connected Components Labeling Algorithm

Suppose a pixel set $M(x_i, y_i)=v$ is a subset of $M(x, y)$, where i=0, 1, 2,…, n, v=0 or v=1. If any two pixels can be connected by value v, the pixel set $M(x_i, y_i)$ is called a connected component. A labeled image LB can be obtained according to the algorithm detailed in literature [10]. The pixel values in labeled image LB are the labels of connected components at the same position, and the labels are integers to index connected components. A binary image with five connected components is shown in Fig. 2-a, and the labeled image is shown in Fig. 2-b.

1	1	0	1	1	1	0	1
1	1	0	1	0	1	0	1
1	1	1	1	0	0	0	1
0	0	0	0	0	0	0	1
1	1	1	1	0	1	0	1
0	0	0	1	0	1	0	1
1	1	0	1	0	0	0	1
1	1	0	1	0	1	1	1

a. Binary image

1	1	0	1	1	1	0	2
1	1	0	1	0	1	0	2
1	1	1	1	0	0	0	2
0	0	0	0	0	0	0	2
3	3	3	3	0	4	0	2
0	0	0	3	0	4	0	2
5	5	0	3	0	0	0	2
5	5	0	3	0	2	2	2

b. Labeled image

Fig. 2. Binary image and labeled image with five connected components

3.2 Minimal Circumscribed Rectangles of the Targets

If the target is not upright in the image, it is difficult to calculate the parameters of their minimal circumscribed rectangles. Therefore, it is necessary to move and rotate the target pixels to situate their center at the origin and let the axes of the targets point to the same direction with coordinate axes, while keeping the shape of the targets fixed. As a result, the upright circumscribed rectangles and minimal circumscribed rectangles are superposed, and it is easy to compute the position and size of the target. A detailed description of the procedures is given as follows.

Suppose LB_k (k=1, 2, …, K) is a connected component image including one target, the coordinates of n target pixels are (p_i, q_i), where i=0, 1, 2,…, n, then the means of

coordinates (p_i, q_i) of n target pixels, represented by (m, c) are easy to be computed. Consequently, the relative coordinates (u, v) are computed according to (2).

$$(u \quad v) = (p \quad q) - (m \quad c) \tag{2}$$

An n-by-2 matrix $\mathbf{X}$ can be constructed using (u, v) as follows

$$\mathbf{X} = \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \\ \vdots & \vdots \\ u_n & v_n \end{bmatrix} \tag{3}$$

Suppose 2-by-2 matrix $\mathbf{V}$ is the tensor of gyration radius of the target, i.e., $\mathbf{V}=(\mathbf{X}^{\mathrm{T}}\mathbf{X})/n$, then a matrix $\mathbf{U}$ can be constructed using the two eigenvectors of matrix $\mathbf{V}$. Rotation matrix $\mathbf{R}$ is determined according to (4).

$$\mathbf{R} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \bullet \mathbf{U} \tag{4}$$

The rotated coordinates (x_i, y_i) are computed as the following equation.

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \mathbf{R} \begin{pmatrix} u_i \\ v_i \end{pmatrix} \qquad i = 1, 2, \ldots, n \tag{5}$$

After that, the target is upright, and there is only one target in each connected component image LB_k. Therefore, the width W and height H of the target can be easily computed. In the 2-D space, rotation is defined by an angle α. The rotation matrix of a column vector about the origin is shown as follows

$$\mathbf{R} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \tag{6}$$

So α is determined by (7).

$$\alpha = \arcsin(\mathbf{R}[2,1]) \tag{7}$$

As a result, we can get the minimal circumscribed rectangle of the target in connected component image LB_k represented by a 5-dimension vector (m, c, W, H, $-\alpha$)$^{\mathrm{T}}$, where (m, c) represent the center of rectangle, (W, H) represent the width and height, $-\alpha$ is the rotation angle. The target model we need is the minimal circumscribed rectangle above mentioned.

4 Experimental Results

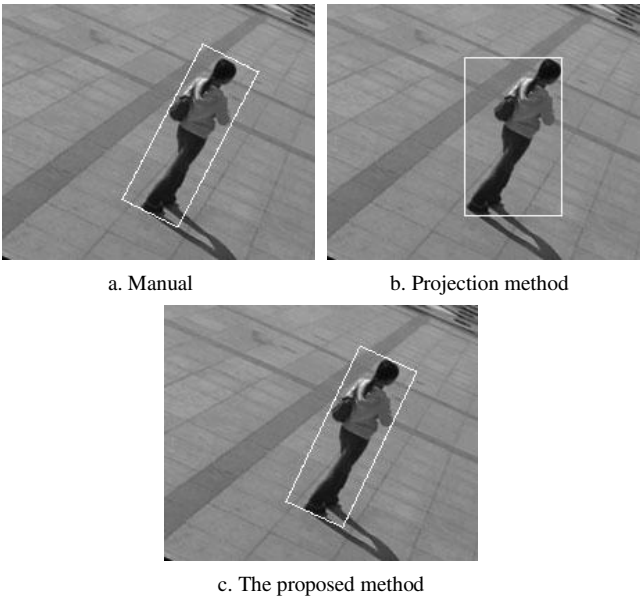
We test the proposed method on different sequences from public data set. Three of them are extracted to demonstrate the performance of the algorithm. The detailed information of these sequences is listed in Table 1.

Table 1. Detailed information of the sequences

title	targets	Data set
MOV045	***One lean person***	***http://210.44.184.112/shadowInv***
MOV031	***Three balls***	***http://210.44.184.112/shadowInv***
Test	***Two persons***	***http://peipa.essex.ac.uk/ipa/pix/pets/***

The results of the three sequences processed by three algorithms are shown in Fig.3-Fig.5. In every figure, the left image represent manual target model, while the mid one and the right one are results of projection method and the presented method, respectively. The target models are indicated by a white rectangle.

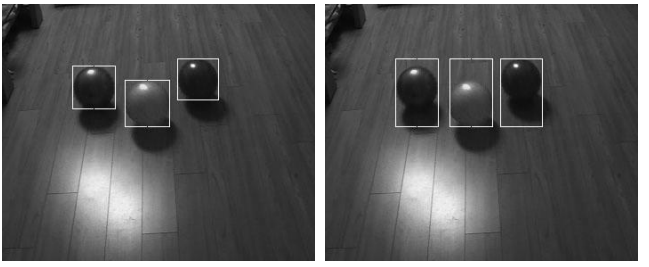
In Fig. 3, the target is lean in the image, which causes a large number of background pixels are included in target model in Fig. 3-b. However, the amount of misclassified pixels is sharply decreased in Fig. 3-c. The reason for this decrease is that the rotation of the rectangle can adapt lean status of the target. In Fig. 4, three balls can be recognized by vertical projection, yet horizontal projections of the balls are partly overlapped. The overlapped projection cause the height of the targets markedly increased in Fig. 4-b. Although the projections of the targets are overlapped, there is no occlusion among the balls in the image. Accordingly, it is not difficult to segment the targets. As a result, the proposed algorithm is not affected by overlapped projection, and it can accurately compute the target model in Fig. 4-c. In Fig. 5, both the vertical projection and the horizontal projection are overlapped, which results in the projection method mistakes two persons for one target. Similar to Fig. 4-c, the overlapped projections have no effect on the presented method, and the target model can be successfully computed in Fig. 5-c.



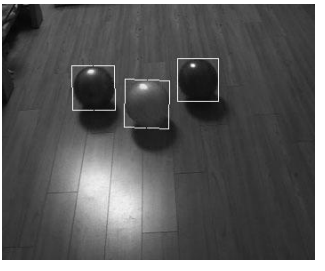
a. Manual b. Projection method

c. The proposed method

Fig. 3. Target model of sequence Mov045 computed by three methods



a. Manual b. Projection method



c. The proposed method

Fig. 4. Target model of sequence Mov031 computed by three methods

a. Manual b. Projection method

c. The proposed method

Fig. 5. Target model of sequence Test computed by three methods

5 Conclusion

In this paper, we propose an automatic and accurate method for initialization of target model in object tracking. Connected components labeling algorithm and radius of

gyration tensor method are respectively used to compute the minimal circumscribed rectangles of the targets. Experimental results show that the presented method can construct target model effectively, and it outperforms projection method obviously.

References

1. Comaniciu, D., Meer, P.: Mean Shift: A Robust Approach toward Feature Space Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 603–619 (2002)
2. Kabaoglu, N.: Target Tracking Using Particle Filters with Support Vector Regression. IEEE Transactions on Vehicular Technology 58, 2569–2573 (2009)
3. Leven, W.F., Lanterman, A.D.: Unscented Kalman Filters for Multiple Target Tracking with Symmetric Measurement Equations. IEEE Transactions on Automatic Control 54, 370–375 (2009)
4. Birchfield, S.T., Sriram, R.: Spatiograms Versus Histograms for Region-Based Tracking. In: Proceedings of International Conference on Computer Vision and Pattern Recognition, pp. 1158–1163 (2005)
5. Kyriakides, I., Morrell, D., Papandreou-Suppappola, A.: Sequential Monte Carlo Methods for Tracking Multiple Targets with Deterministic and Stochastic Constraints. Signal Processing 56, 937–948 (2008)
6. Maggio, E., Smerladi, F., Cavallaro, A.: Adaptive Multifeature Tracking in a Particle Filtering Framework. Circuits and Systems for Video Technology 17, 1348–1359 (2007)
7. Pernkopf, F.: Tracking of Multiple Targets Using Online Learning for Reference Model Adaptation. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 38, 1465–1475 (2008)
8. Wang, J.T.: Research on Object Detection, Tracking and Behavior Recognition in Video Sequences. Nanjing University of Science and Technology (2008)
9. Wu, C.D., Guo, L.F., Zhang, Y.Z., et al.: Method for Touching Objects in Multi-Vehicle Tracking. Journal of Northeastern University (Natural Science) 29, 1065–1068 (2008)
10. Shapiro, L.G., Stockman, G.C.: Computer Vision. Prentice-Hall (2001)

Adjacency List Based Multicast QoS Routing Problem Optimization with Advanced Immune Algorithm

Shanshan Wan, Ying Hao, and Dongwei Qiu

School of Computer Science
Beijing University of Civil Engineering and Architecture, Beijing, China
{wss,haoying,qiudw}@bucea.edu.cn

Abstract. QoS problem of multicast with uncertain parameters constrained is a challenging problem for the prevalent multimedia communications in Internet. In this paper the advanced bionic immune algorithm is used to solve the problem. Every node's priority based adjacency list is designed which guarantees the higher performance path has greater probability to be chosen. Immune algorithm's personal characteristic such as antibody and antigen considered it has ability to seek some information features or knowledge to optimize the process and restrain the compute evolutionary degradation. The solutions' fitness is obtained by the concentration mechanism and affinity function of immune algorithm. The immune operator and immune adjust strategy is helpful to the solution evolutionary process. With test on simulation multiple constraints QoS routing problem the advanced algorithm is proved to have high performance and efficiency.

Keywords: multicasts, QoS, immune algorithm, priority, adjacency list.

1 Introduction

With the computer network technology and the continuous development of multimedia technology, multimedia network applications shows explosive growth trends, such as remote education, collaboration, video conferencing, VoD, IP telephony, interactive games. These applications can be achieved using IP multicast technology. Some constraints such as delay, bandwidth constraints, packet loss rate and hop multicast network will be important influence indicator to the quality of service. QoS routing problem is a nonlinear combinatorial optimization problem in theory and it has been proved as one NP-complete problem and multiple constraints QoS routing problem has become an important research field in computer networks and distributed systems.

In recent years, lots of heuristic methods are extensively applied to this problem such as simulated annealing, ant algorithm search, genetic algorithm (GA) [1-3]. Genetic algorithm has the advantages of robust, suitable for parallel processing and

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 9–17.
springerlink.com

other notable features, but its obvious disadvantage is the slow rate of convergence, as well as easy to fall into premature convergence [4]. In this paper advanced immune algorithm is introduced to solve the QoS problem and the algorithm is advanced and some new strategies are put forward to improve the algorithm's performance. Advanced intelligent immune algorithm is tested on a network topology and three different scale routing demands are considered. The good adaptability, validity and stability performance are fully shown by the results.

2 Intelligent Immune Algorithm

Biological immune system consists of organs, tissues, cells and molecules which guarantee the immune function implementation [5]. The immune function is achieved through self-regulation of lymphocyte. When the antigenic comes into the body it activates the lymphocytes to produce antibodies. The antibodies are combined with the corresponding antigenic to remove the foreign antigen so that living organisms are protected from pathogens disoperation. Lymphocytes have mainly two types of B cells and T cells, which are the main roles in immune response cells. When the antibodies invasion occurs B cells are selected and stimulated under the T cell's reorganization and control. Then B cells proliferate and generate the activation antigen-specific antibody. At the same time, the stimulus and inhibition relations between antigens and antibodies, antibody and antibody make the formation of network structure to maintain the immune balance adjustment. Fig. 1 shows the schematic diagram of the reaction process in immune system. The basic procedure of immune algorithm is described in Fig. 2.

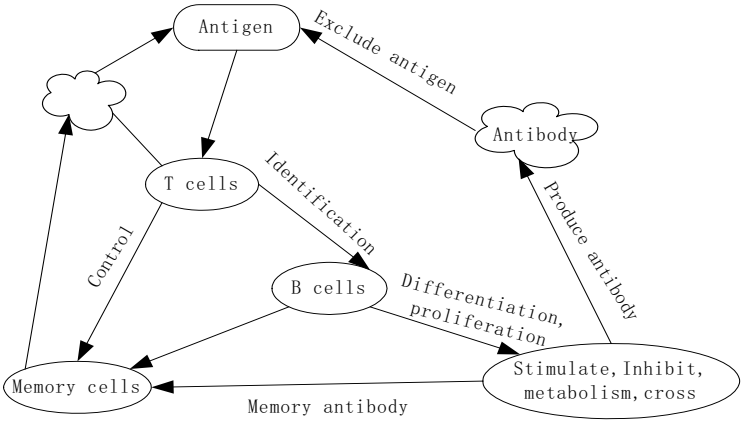


Fig. 1. The model of immune system

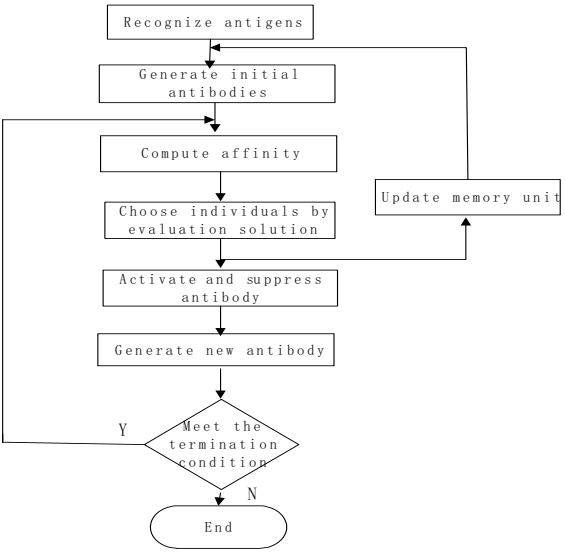
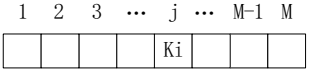


Fig. 2. Flowchart of immune algorithm

To demonstrate the diversity of all the antibodies and measure the individual differences Toyoo Fukuda introduced the concept of information entropy and through evaluation of the affinity among antibodies, antibody and antigen to assess the evaluation of antibodies [6]. The evaluation method based on the information entropy of individual affinity is proved to be a better method. An antibody can be described as follows.



K_i is the allele, i=1,2,…,S. We supposed that one immune system is composed of N antibody and each antibody has M gene positions. Each gene position has S allele values to choose. So the information entropy of the N antibodies is,

$$H(N) = \frac{1}{M}\sum_{j=1}^{M} H_j(N). \tag{1}$$

Where, (1) is the information entropy for the Nth antibody's jth position. P_{ij} is the probability for the jth to choose K_i. Based on the individual genes' information entropy we can define the affinity equation of two antibodies v, w.

$$ay_{vw} = 1/(1+H(2)). \tag{2}$$

Affinity for antigen and antibodies shows the identification level for antigen. Affinity of antibody v and the antigen v is,

$$ax_v = \text{fitness}(v). \tag{3}$$

Fitness(v) is the fitness function for antigen (the problem) and antibodies (solutions).

The concentration of solution v is evaluated as,

$$c_v = \frac{1}{N}\sum_{w=1}^{N} ac_{vw}. \tag{4}$$

Where, $ac_{vw} = \begin{cases} 1, ay_{vw} \geq Tac1 \\ 0.\text{otherwise} \end{cases}$, Tac1 is a pre-determined threshold value.

The expected reproductive rate of the antibody v that's selected rate is defined as $e_v = ax_v / c_v$.

Immune algorithm combines the functions of classifier, neural network and the advantages of machine reasoning. It has the ability of data analysis and classification capabilities. Therefore, it is used for data mining, information processing and other fields. Here we use it to solve multiple constraints based multicast QoS problem.

3 Multiple Constrains QoS Routing Problem

Multiple constrains QoS routing problem aims to find the minimum cost path to connect the source node and target node in the network and the path should meet the requirements of a number of independent path constraints.

The problem is defined as follows:

For one connected undirected weighted graph G = (P, E), it has P nodes, E edges. $E=\{(i,j|i, j \in P)\}, E \subseteq P \times P$. Quaternion group ($td_i$, dj_i, pl_i, co_i) is used to describe the node i in P, in which td_i, dj_i, pl_i and co_i express the time delay, delay jitter, packet loss rate and costs of the node i respectively. And quaternion group ($td_{i,j}$, $du_{i,j}$, $bw_{i,j}$, $co_{i,j}$) is used to describe the link i→j in E, in which $td_{i,j}$, $du_{i,j}$, $bw_{i,j}$, and $co_{i,j}$ express the time delay, delay jitter, packet loss rate and costs of the link i→j respectively. Here we suppose one routing demand R=(P_s, P_d, TD, DJ, PL, BW, CO). P_s is the source node and P_d is the destination node. TD, DJ, PL, BW and CO represent the demanded R's constraints of the maximum delay, the maximum jitter, maximum packet loss rate, minimum bandwidth and maximum cost.

So every path from P_s to P_d should satisfy the followings constraints.

Constraints considered the objective function of QoS routing problem can be described as (5).

$$F(R, P_{optimal}) = \max_{u \in \Omega}\left(\sum_{i=1}^{5} w_i \cdot f_i(u, R)\right). \tag{5}$$

$$\text{s.t.}\begin{cases}\sum_{i\in u} td_i + \sum_{(i,j)\in u} td_{(i,j)} \le TD \\ \sum_{i\in u} dj_i + \sum_{(i,j)\in u} dj_{(i,j)} \le DJ \\ 1-\prod_{i\in u}(1-pl_i) \le PL \\ \min_{(i,j)\in u} bw_{(i,j)} \le BW \\ \sum_{i\in u} co_i + \sum_{(i,j)\in u} co_{(i,j)} \le CO \end{cases}$$

Where, u is a path from P_s to P_d. w_i(i=1,2,3,4,5) represents the weight of the delay, delay jitter, packet loss rate, bandwidth and cost for path u respectively. $f_i(u,R)$ is the path u's satisfaction function to the i constraint of demand R. Ω is the collection of all the paths which meet the constraint equation.

4 Advanced Immune Algorithm

Compared with some optimization problem such as QAP and TSP, multiple constrains QoS routing problem has its own Characteristics. It has more complexities and its objective function is determined by more constraints such as length, bandwidth, delay, delay_jitter. Special characteristics considered the following strategies are designed to solve multiple constrains QoS routing problem using immune algorithm.

4.1 Advanced Immune Algorithm Strategy

- Prune the topology network. Among the given topology network some paths can not meet the bandwidth demand so the preprocessing is necessary to predigest the problem. After preprocessing the reserved bandwidth for each linkage should meet the path bandwidth-constrained. If not, remove the link from the network topology.
- Make the adjacency list. Use adjacency list to store the pruned topology network. Each node's adjacent nodes are sorted by the priority which represents the link's performance. The priority is calculated as:

$$L_{ij}=1/(w_1 td_{ij}+w_2 dj_{ij}+ w_5 co_{ij}).\ j=1,2,3,\ldots,R. \tag{6}$$

R is the number of the adjacent vertexes for node i. The lower L_{ij} obtained the higher priority.

$$P_{ij} = L_{ij} / \sum_{j=1}^{R} L_{ij}. \tag{7}$$

- Double directions antibody encoding mechanism. For all the individuals, the first node of the path is the source node. And the last one is one of the destination nodes. The encoding length of problem is not fixed. So for each destination node we designed double directions encoding mechanism. For each feasible solution the first node and last node are fixed. Each individual try to complete the solution from the source node and destination node. Form each direction the new node is produced by the link's probability P_{ij}. If the new nodes are the same one then the search ends. If the node produced from one direction has existed in the gene position then encode is connected by this node and the redundant gene positions are deleted.
- Target immune operator to keep the antibody's diversification. Target immune operator is a commonly used immune operations operator. The individuals' high concentration in a short time may cause the premature and the diversification is decreased so the target immune operator refers to conduct the immune mutation operation only in certain gene position of the high concentration individuals. The individual's concentration is obtained according to the affinity between each individual and all other individuals. The antibody selection probability is updated by the concentration. To avoid the local optimum value of algorithm the given link's priority is decelerated correspondingly by its concentration.

  The probability P_{ij} is modified as follows.

$$P_{ij} = P_{ij} + C_t \cdot \alpha. \tag{8}$$

  Where C_t is concentration proportion of the individual i among all the feasible solutions α is the correction factor $\alpha \in [0.2,0.5]$. The strategy guarantees that the individual whose fitness is superior to the average will obtain larger choice probability increment for its linkage nodes. And choice probability of the individual whose fitness is lower than the average will be decreased.

5 Simulation Experiment

The algorithm is implemented in VC++ on personal computer, Core2Duo, 2.4Ghz, 2G inner Memory, Windows Xp operating system.

The Experimental network topology is given by Salama network topology algorithm. The algorithm guarantees an average node degree four, which is close to the actual network environment. The network topology in Fig. 3 is used to test the algorithm.

The link property is described with quaternion group (delay, delay jitter, bandwidth, cost). The QoS indexes of each edge are generated randomly, the index range are given by certain range respectively. The range of delay is that td $\in$ (0.0, 30). The range of delay jitter is that dj$\in$ (0.0, 10). The range of bandwidth is that bw$\in$ (0.0, 90.0). The cost range is that co$\in$ (0.0, 120).And here we suppose all nodes' packet loss rate can meet the constraint demand.

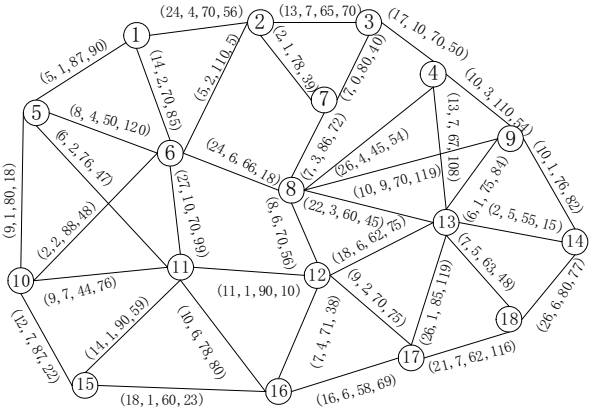


Fig. 3. Experimental network topology

Suppose the source node S is node 1, the bandwidth constraint is 60, delay constraints is 30ms, Delay jitter Constraints is 10ms. In the experiment three sets of values are chosen as destination nodes: D1 = (7,12,15), D2 = (9,10,13,16), D3 = (4,11,13,16,18). Though preprocess some links are pruned because they can not meet the bandwidth constraint demand. They are 5->6, 4->8, 10->11, 13->14, 16->17.

For each node's possible linkage the priority and choice probability is calculated by (6) and (7). The path which has greater priority obtains larger probability to be chosen. It is consistent to Dijstra K shortest path algorithm. The priority based adjacency list is shown in table 1.

Some important parameters are given as α =0.25. The pop-size of chromosome individuals is initialized according the scale of problem. In table 2 the pop-size and the optimal results are shown where the optimal paths, the delay, total cost and the average compute time are given.

The advanced immune algorithm is compared to GA immune algorithm. The cost, average evolutionary time and the proportion of optimal individuals in solution colony are recorded in table 3. The results illustrate the advanced algorithm's performance is better than GA obviously.

Table 1. Priority Based Adjacency list

Node	Priority Sort	Node	Priority Sort
1	5-6-2	**10**	5-6-15
2	6-7-3-1	**11**	12-5-15-16-6
3	7-2-4	**12**	11-16-8-17-13
4	9-3-13	**13**	18-9-8-12-17
5	10-11-1	**14**	9-18
6	2-10-8-1-11	**15**	10-16-11
7	2-3-8	**16**	12-15-11
8	12-7-6-13-9	**17**	12-18-13
9	4-13-8-14	**18**	13-14-17

Table 2. Results of the algorithm

Destination Node	Pop Size	Results	Time Delay	Total Cost
7,12,15	25	1->2->7 1->2->6->11->12 1->5->10->15	119	195
9,10, 13,16	40	1->6->8->9 1->6->10 1->2->3->4->13 1->5->11->12->16	160	246
4,11,13, 16,18	50	1->2->6->8->4 1->5->11 1->6->8->13 1->5->11->15->16 1->2->7->3->4->9->13 ->18	259	378

Table 3. Comparison of algorithm performance

Three Destinations Set	GA	Advanced Immune Algorithm
Total Cost	214/277/456	195/246/378
Average time(s)	4.5/12.1/15.7	4.2/8.0/11.2
Proportion	1/0.9/0.92	1/1/0.98

6 Conclusions

Experiments show that priority based adjacency list is reasonable and helpful to guide the immune evolutionary process. The double directions search and target immune operator based on concentration mechanism and affinity function optimize the process and restrain the compute evolutionary degradation. The advanced immune algorithm is more adaptive to solve such problem as virus detection, network hacking, fault monitoring and diagnosis, network security, logistics, system maintenance and so on.

Acknowledgment. The authors wish to acknowledge the support of Beijing University of Civil Engineering and Architecture Science Research Foundation of China (No.100903808) and Beijing Technology Plan General Projects (No. KM201010016008).

References

1. Li, Y., Li, C.L.: QoS Multicast Routing Algorithm Based on GA. J. Journal of Systems Engineering and Electronics 15(1), 90–97 (2004)
2. Yuan, X.: Heuristic Algorithms For Multi-constrained Quality of Service Routing. J. IEEE/ACM Transactions on Networking (10), 244–256 (2002)
3. Chen, J., Zhang, H.W.: QoS Multicast Routing Algorithm Based on Adaptive Ant Colony Algorithm. J. Computer Engineering (34), 200–203 (2008)

4. Cui, Y., Wu, J.P., Xu, K., Xu, M.W.: Research on Internetwork QoS Routing Algorithms: a Survey. Journal of Software (11), 340–344 (2002)
5. Cai, Z.X.: The Progress of the Immune Algorithm Research. J. Control and Decision 19(8), 841–846 (2004)
6. Dasgupta, D.: Artificial Immune Systems and Their Applications. Springer, Heidelberg (1999)

Simple and Efficient Direct Solution to Absolute Orientation

Huaien Zeng* and Qinglin Yi

Key Laboratory of Geological Hazards on Three Gorges Reservoir Area, Ministry of Education
China Three Gorges University
Yichang, People's Republic of China
zenghuaien_2003@yahoo.com.cn

Abstract. Absolute orientation is a fundamental and important task in photogrammetry and robotics. It aims to find the best transformation parameters which relate the stereo model coordinates and geodetic coordinates. The paper presents a simple and efficient direct solution based on optimization theory and rigid body transformation rule. It doesn't require the good approximation of transformation parameters or linearization process, neither iterative computation. A numerical case is illustrated, and the result shows the presented solution is efficient and adapted to any rotation angles case.

Keywords: absolute orientation, direct solution, rigid body transformation, transformation of super large rotation angles.

1 Introduction

Absolute orientation is a fundamental and important task in photogrammetry. It involves recovering the transformation from image coordinate system of the stereo model developed from pairs of photographs to a geodetic system. Usually the 7-parameter (3 translation parameters, 3 rotation angles and 1 scale factor) similarity transformation is used to represent the transformation, thus absolute orientation aims to find the optimal 7 parameters, given the two sets of coordinates of control points in the two coordinate system. It is also important in robotics community, which relating the camera coordinate to the coordinate in a coordinate system attached to the mechanical manipulator. Traditional solution to absolute orientation is numerical iterative solution [1], [2], which needs linearization process, namely requires a good approximation of transformation parameter, especially the rotation angles. In some cases, e.g., unmanned aircraft vehicle photograph, it is difficult even impossible to gain the approximation and the iterative calculation will fail. So, it is full of meaning to develop direct (non-iterative) solution to absolute orientation. Reference [3] and [4] presented direct unit quaternion solution. Reference [5] presented a direct matrix square-root solution. Reference [6] and [7] presented direct solution based singular

* Corresponding author.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 19–25.
springerlink.com

value decomposition (SVD). Reference [8] presented a closed-form solution based on linear subspace method.

This paper will present a simple and efficient direct solution to absolute orientation, based on optimization method and physical meaning of rigid body transformation. The seven transformation parameters are determined individually and stepwise. Firstly, the three translation parameters are obtained by means of optimization method in the least squares sense, whose formula is the function of the scale factor and rotation matrix. Secondly, by backward substituting formula of translation parameters into transformation model, and optimization method, the scale factor is determined, whose expression has a variable i.e. rotation matrix. Lastly but most crucially, the rotation matrix is found by use of the physical meaning of rigid body transformation.

The remainder of the paper is organized as follows. Sect. II presents a new direct solution to absolute orientation, and derives the basic mathematical model and explicit expression of transformation parameter. In sect. III, numerical case study is carried out to valid the presented solution. Lastly, Conclusions are made in sect. IV.

2 Direct Solution to Absolute Orientation

2.1 Solution of Translation Parameters

The 7-parameter similarity transformation can be written as

$$a_i = t + sRb_i, \tag{1}$$

where, $a_i = \begin{bmatrix} X_i & Y_i & Z_i \end{bmatrix}^T$, $b_i = \begin{bmatrix} x_i & y_i & z_i \end{bmatrix}^T$, $(i = 1, 2, \cdots, n)$ are the coordinates in target coordinate system and source coordinate system (transformed), n denotes the number of control points, $t = \begin{bmatrix} \Delta X & \Delta Y & \Delta Z \end{bmatrix}^T$ is the translation parameters vector, s is the scale factor, and R is the rotation matrix including the 3 rotation angles. Introducing the following matrix form of the coordinates as

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^T, B = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}^T, \tag{2}$$

then the 7-parameter similarity transformation is rewritten as

$$A = \mathbf{1}t^T + sBR^T, \tag{3}$$

where, $\mathbf{1} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T}_{n}$, superscript T stands for transpose.

Considering the coordinates include errors, the model (3) is transformed as

$$A = \mathbf{1}t^T + sBR^T + E, \tag{4}$$

where E is the transformation error matrix. The criterion of least squares can be expressed by the Frobenius matrix norm as

$$\|E\|^2 = tr(E^T E) = \min, \tag{5}$$

where tr denotes trace operation of matrix, substituting the expression of E easily obtained from (4) into (5), one can obtain

$$\begin{aligned} \|E\|^2 &= \|A - \mathbf{1}t^T - sBR^T\|^2 \\ &= tr((A - \mathbf{1}t^T - sBR^T)^T (A - \mathbf{1}t^T - sBR^T)) \\ &= tr((A - sBR^T)^T (A - sBR^T) \\ &\quad -2(A - sBR^T)^T \mathbf{1}t^T + t\mathbf{1}^T\mathbf{1}t^T) \end{aligned}, \tag{6}$$

$$\frac{\partial \|E\|^2}{\partial t^T} = 2(\mathbf{1}^T\mathbf{1})t - 2(A - sBR^T)^T\mathbf{1}. \tag{7}$$

Let (7) be 0, one obtain the least squares solution of translation parameters vector

$$t = \frac{1}{n}(A - sBR^T)^T\mathbf{1}. \tag{8}$$

Obviously, t is the function form of s and R.

2.2 Solution of Scale Factor

As soon as one backward substitutes t into (3) and makes a arrangement, one will get a centralized model as

$$(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)A = (I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)sBR^T, \tag{9}$$

where $(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)$ is the centering matrix, and $\Delta A = (I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)A$, $\Delta B = (I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)B$ are the centralized coordinates matrix. Equation (9) is then written as

$$\Delta A = s\Delta BR^T, \tag{10}$$

and transformation error matrix $E = \Delta A - s\Delta BR^T$, substituting it into (5), one gets

$$\begin{aligned} \|E\|^2 &= tr((\Delta A - s\Delta BR^T)^T(\Delta A - s\Delta BR^T)) \\ &= tr(\Delta A^T \Delta A - 2s\Delta A^T \Delta BR^T + s^2 R\Delta B^T \Delta BR^T), \end{aligned} \tag{11}$$

$$\frac{\partial \|E\|^2}{\partial s} = -2tr(\Delta A^T \Delta BR^T) + 2s \times tr(R\Delta B^T \Delta BR^T). \tag{12}$$

Let (12) be 0, one gets the least squares solution to scale factor

$$s = \frac{tr(\Delta A^T \Delta BR^T)}{tr(\Delta B^T \Delta B)}, \tag{13}$$

the derivation of (13) makes use of the properties of trace operation, i.e.,

$$tr(R\Delta B^T \Delta BR^T) = tr(\Delta B^T \Delta BR^T R) = tr(\Delta B^T \Delta B). \tag{14}$$

2.3 Solution of Rotation Matrix

If one expands centralized coordinate matrix ΔA and ΔB as

$$\Delta A = \begin{bmatrix} \Delta a_1 & \Delta a_2 & \cdots & \Delta a_n \end{bmatrix}^T, \tag{15}$$

$$\Delta B = \begin{bmatrix} \Delta b_1 & \Delta b_2 & \cdots & \Delta b_n \end{bmatrix}^T, \tag{16}$$

thus $\Delta a_i, \Delta b_i, (i = 1, \cdots, n)$ is the corresponding two sets of centralized coordinate vector. According to the physical meaning of rigid body transformation, i.e,

$$s = \frac{\|\Delta a_i\|}{\|\Delta b_i\|}, (i = 1, \cdots, n), \tag{17}$$

and (10), one gets

$$\delta a_i = R\delta b_i, (i = 1, \cdots, n), \tag{18}$$

where

$$\delta a_i = \frac{\Delta a_i}{\|\Delta a_i\|}, \delta b_i = \frac{\Delta b_i}{\|\Delta b_i\|}. \tag{19}$$

Defining

$$\delta A = \begin{bmatrix} \delta a_1 & \delta a_2 & \cdots & \delta a_n \end{bmatrix}, \tag{20}$$

$$\delta B = \begin{bmatrix} \delta b_1 & \delta b_2 & \cdots & \delta b_n \end{bmatrix}, \tag{21}$$

thus,

$$\delta A = R\delta B\,, \tag{22}$$

One can get the solution of R as

$$R = \delta A\delta B^T(\delta B\delta B^T)^{-1}. \tag{23}$$

Once the estimation of R is obtained, one can compute the rotation angles by using the following formulae directly [1],

$$\varphi = \tan^{-1}(-\frac{r_{13}}{r_{33}}),\ \omega = \sin^{-1}(-r_{23}),\ \kappa = \tan^{-1}(\frac{r_{21}}{r_{22}})\,. \tag{24}$$

where the r_{ij} is the elements of the estimate of R in the *i*th row and *j*th column.

3 Numerical Case Study and Discussion

In order to validate the presented direction solution to absolute orientation, the paper simulates and processes the data through the following several steps. Firstly, the coordinates of ground control points are given as in Table 1, and 3 sets of transformation parameters of absolute orientation are given as in Table 2, which take account of the cases of small rotation angles, large rotation angles and super large rotation angles, respectively. Then, the stereo model coordinates are rigidly computed by similarity transformation model, as listed in Table 3. Thirdly, the transformation parameters of absolute orientation are recovered with the presented direction solution in the paper in terms of the above simulative coordinates, and the result is showed in Table 4. Besides the calculated errors of transformation parameters, the residuals of transformation coordinates are also important criterion to evaluate the direction solution. So the residuals are computed meanwhile and listed in Table IV, where M_x, M_y, M_z, and M_p stands for the mean squares root error of *X* direction, *Y* direction, *Z* direction, and point location.

Seen from Table 4, the errors of transformation parameters are very small, and doesn't become bigger and bigger evidently with the increase of rotation angles. Also seen from Table 4, the statistics of residuals are also very small and close to each other considering the 3 sets transformation. Thus, the resented direct solution to absolute orientation is efficient and suitable for any rotation angles case.

Table 1. Coordinates of Ground Control Points

Point no.	*X* (m)	*Y* (m)	*Z* (m)
1	3395720.844	531102.361	550.321
2	3395750.936	531978.647	613.145
3	3394890.464	532026.834	561.822
4	3394911.953	531118.463	490.291

Table 2. Parameters of Absolute Orientation

Set no.	ΔX (m)	ΔY (m)	ΔZ (m)	φ	ω	κ	s
Set 1	3392566.687	533870.514	270.736	50′	80′	20′	100
Set 2	3392566.687	533870.514	270.736	30°	45°	20°	100
Set 3	3392566.687	533870.514	270.736	60°	70°	85°	100

Table 3. Simulative Stereo Model Coordinates

Point	Set 1			Set 2			Set 3		
no.	x (m)	y (m)	z (m)	x (m)	y (m)	z (m)	x (m)	y (m)	z (m)
1	31.417678	-27.802910	2.980305	17.058935	-37.084280	10.134270	-32.108230	-21.070659	17.147694
2	31.778749	-19.029859	3.400030	19.813444	-31.196711	4.216318	-29.011931	-21.496953	8.931602
3	23.170370	-18.507059	3.000893	13.619351	-25.676631	6.603524	-22.525829	-16.164530	10.939727
4	23.321901	-27.606060	2.494096	11.085493	-32.136758	12.512666	-26.174464	-15.969757	19.289647

Table 4. Errors between Recovered Values and True Values of Absolute Orientation Parameters and Statistics of Residuals

Set no.	Solution error of absolute orientation parameters							Statistics of residuals			
	ΔX (m)	ΔY (m)	ΔZ (m)	φ (″)	ω (″)	κ (″)	s	M_x (m)	M_y (m)	M_z (m)	M_p (m)
Set 1	-2.4×10^{-4}	5.5×10^{-5}	8.0×10^{-6}	8.0×10^{-3}	2.4×10^{-3}	8.2×10^{-6}	2.8×10^{-8}	2.5×10^{-4}	1.9×10^{-7}	2.5×10^{-4}	3.5×10^{-4}
Set 2	-2.3×10^{-4}	4.6×10^{-5}	1.5×10^{-6}	1.1×10^{-2}	-3.3×10^{-3}	2.0×10^{-3}	9.4×10^{-9}	2.5×10^{-4}	1.3×10^{-7}	2.5×10^{-4}	3.5×10^{-4}
Set 3	-2.4×10^{-4}	6.5×10^{-5}	8.4×10^{-6}	4.7×10^{-3}	9.0×10^{-3}	-2.9×10^{-2}	-1.2×10^{-8}	2.5×10^{-4}	1.8×10^{-7}	2.5×10^{-4}	3.5×10^{-4}

4 Concluding Remarks

Absolute orientation is a fundamental and important task in photogrammetry and robotics. It aims to solve the seven transformation parameters of absolute orientation. The paper presented a simple and efficient direct solution based on optimization theory and rigid body transformation rule. It doesn't require the good approximation of transformation parameters or linearization process, neither iterative computation. The numerical case is illustrated, and the result shows the presented solution is efficient and adapted to any rotation angles case.

Acknowledgment. The work of this paper is jointly supported by Youth Science Foundation of China Three Gorges University (grant No.KJ2009A004) and Talent Research Start-up Fund of China Three Gorges University (grant No.KJ2009B008). The first author is grateful to the support and good working atmosphere provided by his research team in China Three Gorges University.

References

1. Wang, Z.: Principles of Photogrammetry. Wuhan: Press of Wuhan Technical University of Surveying and Mapping (1990) (in Chinese)
2. Zhang, Z., Zhang, J.: Digital Photogrammetry. Wuhan University Press, Wuhan (2002) (in Chinese)
3. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. J. Optical Soc. Am. 5(7), 1127–1135 (1987)
4. Jiang, G., Wang, J., Zhang, R.: A close-form solution of absolute orientation using unit quaternions. Journal of Zhengzhou Institute of Surveying and Mapping 24(3), 193–195 (2007)
5. Horn, B.K.P., Hilden, H.M., Negahdaripour, S.: Closed-form solution of absolute orientation using orthonormal matrices. J. Optical Soc. Am. 5, 1127–1135 (1988)
6. Arun, K.S., Huang, T.S., Blostein, S.D.: Least-squares fitting of two 3D point sets. IEEE Trans. Pattern Analysis and Machine Intelligence 9(5) (September 1987)
7. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. IEEE Trans. Pattern Analysis and Machine Intelligence 13(4) (April 1991)
8. Wang, Z., Jepson, A.: A new closed-form solution for absolute orientation. In: IEEE Conf. Computer Vision and Pattern Recognition, pp. 129–134 (1994)
9. Fiore, P.D.: Efficient linear solution of exterior orientation. IEEE Transactions on Pattern Anlysis and Machine Intelligence 23(2), 140–148 (2001)
10. Triggs, B.: Camera pose and calibration from 4 or 5 known 3D points. In: Proc. Int'l Conf. Computer Vision, ICCV, pp. 278–284 (1999)
11. Zhao, S., Guo, Q., Luo, Y., Wu, W.: Quaternion-based 3D Similarity Transformation Algorithm. Geomatics and Information Science of Wuhan University 34(10), 1214–1217 (2009) (in Chinese)
12. Grafarend, E.W., Awange, J.L.: Nonlinear analysis of the three-dimensional datum transformation [conformal group C7(3)]. J. Geod. 77, 66–76 (2003)
13. Shen, Y.-Z., Chen, Y., Zheng, D.-H.: A quaternion-based geodetic datum transformation algorithm. J. Geod. 80, 233–239 (2006)
14. Yi, Q., Zeng, H., Wu, Y., Huang, S.: A Quaternion-Based Solution of Non-linear 3D Coordinate Transformation Parameters. In: Proc. of IEEE International Conf. on Information Engineering and Computer Science, Wuhan, China, pp. 183–186 (2009)
15. Zeng, H.: A 3D coordinate transformation algorithm. In: Proc. IEEE the 2nd Conf. Environmental Science and Information Application Technology, Wuhan, China, July 17-18, pp. 195–198 (2010)
16. Zeng, H., Yi, Q.: A new analytical solution of nonlinear geodetic datum transformation. In: Proc. the 18th International Conference on Geoinformatics, Beijing, China, June 18-20 (2010)

Information System Attribute Reduction Parallel Algorithm Based on Information Entropy

Chunlin Yang[1], Zhonglin Zhang[2], and Jun Zhang[2]

[1] School of information engineering
Lanzhou University of Finance and Economics
Lanzhou, China
[2] School of electronic and information engineering
Lanzhou Jiaotong University
Lanzhou, China
yangcl@ lzcc.edu.cn

Abstract. Attribute reduction is an important treatment in information system. Classification is the base of attribute reduction, but it is inefficient to reduce attribute directly on a large data set. This paper put forward an attribute reduction parallel algorithm based on information entropy. The algorithm reduces attribute at the same time breakdowns original information system layer by layer, as a result, achieve attribute reduction parallel calculate and shrink the search space.

Keywords: information system, attribute reduction, entropy of information, parallel calculate.

1 Introduction

The knowledge in knowledge base (attribute) is not equally important, some of which is redundant, is not conducive to make the right decisions simply and effectively. The so-called attribute reduction is to maintain the knowledge base in the same conditions of classification ability, delete irrelevant or unimportant attributes. People always expect to find the smallest reduction of attributes, but this has proven to be an NP-hard problem[1]. Existing information systems reduction algorithm, mainly based on attribute importance[2], discernibility matrix[3] and information entropy[4]. However, as computer and data acquisition technology has been progressed, the accumulation of data is not only in the data object, but also the number of dimensions is rapidly growth, due to the non-parallelism of these algorithms, they will increase more and more complex as the data size becomes lager.

This paper presents a parallel attribute reduction algorithm based on information entropy for information system. The original large-scale data sets were broken down into a sub-table, reduce the size of information systems, and also in various sub-table on calculation, reducing the computing time.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 27–34.
springerlink.com

2 Related Concepts

Definition 1[5,6]: We have formally called quaternion group S=(U, C, V, f) is an information system, in which $U = \{x_1, x_2, \ldots, x_n\}$ is the object of non-empty finite set called the universe; $C = \{a \mid a \in C\}$ is the property of non-empty finite set; $V = \bigcup V_j (1 \le j \le m)$ is the property range of the function f; f is S's information function.

Definition 2: In the information table S, for each attribute subset BA, you can define an indiscernibility relation IND (B) = ((x, y) | b (x) = b (y), (x, y) ∈ U2, b ∈ B). Obviously IND (B) is an equivalence relation. That is the equivalence class of the IND (B) which contains the element x.

Definition 3: Let U is a universe, P and Q are two equivalence relations family on the domain U. The P and Q are defined two random variables in the subset of U consisting of algebra σ^-, its probability distribution can be determined as follows:

P and Q located at the delineation of domains were derived X and Y, which

$$\begin{aligned} X &= U / IND(P) = \{X_1, X_2, \ldots, X_n\} \\ \mathrm{Y} &= U / IND(Q) = \{Y_1, Y_2, \ldots, Y_n\} \end{aligned} \tag{1}$$

The P and Q in the domain of a subset of U composed of the probability distribution defined on algebra σ^- were

$$\begin{aligned} [X; p] &= \begin{bmatrix} X_1 & X_2 & \ldots & X_n \\ p(X_1) & p(X_2) & \ldots & p(X_n) \end{bmatrix}, \\ [\mathrm{Y}; p] &= \begin{bmatrix} Y_1 & Y_2 & \ldots & Y_n \\ p(Y_1) & p(Y_2) & \ldots & p(Y_n) \end{bmatrix} \end{aligned} \tag{2}$$

Which

$$\begin{aligned} p(X_i) &= \frac{|X_i|}{|U|}, i = 1, 2, \ldots, n \\ p(\mathrm{Y}_i) &= \frac{|Y_j|}{|U|}, j = 1, 2, \ldots, n \end{aligned} \tag{3}$$

Symbols |E| represents the base of set E.

The information entropy of knowledge P is defined as H (P).

$$H(P) = -\sum_{i=1}^{n} p(X_i) \log p(X_i) \tag{4}$$

Information entropy H (P) can also be expressed as general functions of the form

$$H(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i \tag{5}$$

We called $H(p_1, p_2, \ldots, p_n)$ entropy function.

The conditional entropy of knowledge Q relative knowledge P is:

$$H(Q \mid P) = -\sum_{i=1}^{n} p(X_i) \sum_{j=1}^{m} p(Y_j \mid X_i) \log p(Y_j \mid X_i) \tag{6}$$

where, $p(Y_j \mid X_i) = \dfrac{Y_j \cap X_i}{X_i}$.

Definition 4: Let U be a universe, P is the equivalence relation family on domain U, and Q $\subseteq$ P, if P and Q satisfy the following conditions:

(1) H(Q)=H(P);

(2) For any $q \in Q$, $H(q \mid Q - \{q\}) > 0$.

P called a reduction of Q .

3 Information Entropy Analysis

From the previous definition of information entropy, we can get the following properties:

Property 2.1: If H(C)=0, then C has only one category on the U.

Proof: Since $H(C) = -\sum_{i=1}^{|U/C|} P(X_i) \log P(X_i)$, $(i = 1, 2, \ldots, r)$, $P(X_i)$ is always satisfied $0 \le P(X_i) \le 1$ $(i = 1, 2, \ldots, r)$,and in the information entropy the logarithm is always taken more than 1,so $\log P(X_i) \le 0$,then $-\log P(X_i) \ge 0$, $-P(X_i) \log P(X_i) \ge 0$. If $H(C)=0$, then $-P(X_i) \log P(X_i)$ must be 0 or $P(X_i)=1$. According to the certainty of entropy function in the $P(X_i)$ only one probability weight equal to 1, others probability weight equal to 0,so C has only one category on the U.

Property 2.2: Set U / a = (X1, X2, ..., Xn), X1, X2, ..., Xn more evenly, n is larger, the H (a) is greater.

Proof: By the entropy function of convex [6] known, there is a maximum entropy. Due to $p_1, p_2, \ldots, p_n$ be met $\sum_{i=1}^{r} p_i = 1$ .Therefore, the maximum value of $H(p_1, p_2, \ldots, p_n)$, should be bound by the constraints under maximum conditions.

According to condition maximum method, make an auxiliary function:

$$F(p_1, p_2, \ldots, p_n) = H(p_1, p_2, \ldots, p_n) + \lambda[\sum_{i=1}^{n} p_i - 1]$$

$$= -\sum_{i=1}^{n} p_i \log p_i + \lambda[\sum_{i=1}^{n} p_i - 1]$$

Under this method, the probability distribution corresponding the maximum value of $H(p_1, p_2, \ldots, p_n)$ is:

$$p_i = \frac{1}{n}(i = 1, 2, \ldots, n)$$

As a result, the maximum value of $H(p_1, p_2, \ldots, p_n)$ is

$$\begin{aligned} H(p_1, p_2, \ldots, p_n) &= H(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}) \\ &= -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} \\ &= \log n \end{aligned}$$

Therefore, under normal circumstances $H(p_1, p_2, \ldots, p_n) \le \log n$ .This shows that the number of objects in $X_1, X_2, \ldots, X_n$ were equal, that is when the domain was the most evenly divided by attribute a, the H(a) is maximum. When n greater, the number of sub-class is greater, the maximum value of H (a) is greater too. Because of information entropy function is convex, so when the more uniform of $X_1, X_2, \ldots, X_n$, n bigger, more close to the maximum value, the greater its entropy.

Infer 2.1: Set U/a={X1, X2,… ,Xn} , if H (a) the greater the X1, X2, ..., Xn more tend to be uniform, n greater.

Proof: Since the convexity and extreme of the entropy function[6], the greater the value of information entropy, the closer the maximum value, X1, X2, ..., Xn become more uniform, n greater.

Property 2.2: the information system set U is divided into n sub-table U1, U2, ..., Un, while doing the same operate on the sub-table. The more uniform of the object number contained in the sub-table, the less time-consuming complete operate on all the objects.

Proof: Let the total of objects in information system U is m, the objects were randomly divided into sub-table Ui, i = 1,2, ..., n,. If the time to operate one object is t, then the time to operate on all objects in Ui are t*|Ui|. So when all the objects of the domain to complete operation, the time is t * max (| Ui |). For t * max (| Ui |) tends to the minimum, you have to max (| Ui |) tends to the smallest, When | Ui | = $\left\lfloor \frac{m}{n} \right\rfloor$ can make the max (| Ui |) became smallest. That is, the more uniform of the object number contained in the sub-table, the less time-consuming complete operate on all the objects.

Table 1. Information table

	a	b	c	d	e
x_1	0	0	1	1	1
x_2	1	1	1	1	1
x_3	1	1	0	0	1
x_4	0	1	0	1	1

4 Information Systems Reduction Analysis Based on Classification

According to the definition of information system reduction, attribute classification is the basis of reduction.

Table 1, where (abcde) for the attribute set C, (x1, x2, x3, x4) for the domain of U, its reduction is (ac), (bcd), (abd).

Using the method of gradually add the property to calculate the reduction of this form. According to previous analysis, we choose the attribute a that can form the table most uniform and can decompose the information table down into the maximum class. This the time to calculate sub-table at least. Then U / a = ((x1, x4), (x2, x3)), because the U / a ≠ U / C, so need to add attributes, and add the attribute is essentially equal to the U / a was further divided, then U can be divided into two domains of U1 and U2 [7].The attribute which can distinguish {x1 , x4} and {x2 , x3} can be added. If no such attribute, the attribute which can became classification smallest and most uniform can be added. That is add the attribute which make U1 and U2 change biggest. And in the rest attribute set {bcde}, the attribute can make U1 and U2 change biggest only c. U/ac = U/C, and ac is independent, so ac is a reduction and is a minimum reduction.

5 Information Attribute Reduction Parallel Algorithm Based on Information Entropy

algorithm statement

By the above analysis a new kind of enlightenment parallel attribute reduction algorithm as follows has been present:

Input: information table S=<U,C,V,f>, C is attribute set

Output: reduction of the information table

1) Initialization: reduction $R = \varnothing$, $Leave = C$, $k = |Leave|$, $Long = |U|$

// k means attribute number of $Leave$, $Long$ means object number of domain.

2) if Long=1 then return “There is only one attribute in the information table, it is insignificant to reduce”

For information table U , calculate information entropy of attribute a_i , $i = 1, 2, \ldots, k$ in $Leave$.

If $\max[H(a_i)] = 0$

Return "The objects in this information table are same, it is insignificant to reduce."

Else

Random select attribute b within $\{a_i \mid \max[H(a_i)]\}, U_j = U/b, j = 1,2,\ldots,n.$ where $n = |U/b|$;

$R = R \cup \{b\}$;

$Leave = Leave - \{b\};$

3) For every sub-table do judge and dismantling operations at the same time. Let r as sub-table number.

If $|U_j| = 1$

Stop the operate to U_j, $flag = 1$;

Else

For attribute $a_i, i = 1,2,\ldots,k$ in $Leave$, calculate their information entropy $H_j(a_i), j = 1,2,\ldots,r$;

If $\max[H_j(a_i)] = 0, i = 1,2,\ldots,k$

Stop the operate to U_j, $flag = 1$;

Else

Random select attribute b which in $\{a_i \mid \max[\sum_{j=1}^{r}(|U_j|/Long)H_j(a_i)]\}$

to dismantling the sub-table again;

$R = R \cup \{b\}$;

$Leave = Leave - \{b\};$

Repeat step (3) for every sub-table.

If the $flag = 1$ for every sub-table , return R .

Property 4.1: When all the sub-table can not be classified, $H(R) = H(C)$.

Proof: To proof $H(R) = H(C)$, only need to proof H(C)-H(R)=0。

$$\begin{aligned} & H(C) - H(R) \\ &= H(C - Leave + Leave) - H(R) \\ &= H(C - Leave) + H(Leave \mid C - Leave) - H(R) \\ &= H(R) + H(Leave \mid R) - H(R) \\ &= H(Leave \mid R) \end{aligned}$$

Let $H(a_i)$ as the information entropy of attribute a_i in dismantling table of m layer, where $a_i \in Leave$, according to conditional entropy definite

$\sum_{j=1}^{n}(|U_j|/Long)H_m(a_i) = H_m(a_i \mid R)$, where $a_i \in Leave$.

Analysis of algorithmic time complexity

For the information table $S =< U, C, V, f >, |U|$ is the object number of the table, $|C|$ is the attribute number of the table. In the traditional heuristic attribute reduction algorithms, the information systems attribute reduction's time complexity is $O(|C|^3|U|^2)$ [8].

Using the algorithm was proposed in this paper, select attribute a to dismantling information table U, get sub-table U_i ,the object number is $|U_i|$. The time complexity to dismantling first layer is $O(|C||U|^2)$. If using the cardinal number sort in [9], the time complexity to dismantling first layer will fall to $O(|C||U|)$. For not loss of generality, U_n be an example to dismantle second layer. If there are j sub-table U_{nj}, the time complexity to dismantling U_n is $O((|C|-1)|U_n|)$. Because the dismantling to U_n is simultaneously, the time complexity to dismantling whole second layer is $O((|C|-1)(\max|U_i|))$ .And for the same reason, the time complexity to dismantling kth layer is $O((|C|-k+1)(\max|the\ k-1\ level\ of\ sub-table|))$. For the information table S, maximum layer number to dismantle is |C|, the time complexity of it is $O(\max|\text{the } |C|-1 \text{ level of sub-table}|)$.

To sum up, the time complexity of the algorithm is

$$\begin{aligned} &O(|C||U|)+(|C|-1)(\max|U_i|) \\ &+\ldots+(|C-R|+1)(\max|\text{the}|C-R|-1 \text{ level of sub-table}|)) \end{aligned} \tag{7}$$

The above formula shows that more uniform of the table in every layer be dismantled or the more same object in the table, the less time complexity. When the first layer sub-tables are |U| or 1,time complexity is minimum which equal to $O(|C||U|)$;when exist C layer sub-table and there are two sub-tables in every layer, there is 1 object in the one of sub-table, time complexity is worst which equal to $O(|C|^2|U|)$. In most instances, this algorithm has $\max|\text{the } |C-R|-1 \text{ level of sub-table}|<|U|$ especially to a larger scale information system. In practical dismantling process, since exist sub-tables do not need be classified intermediate stages in dismantling process, that will save time. The more information table scale, the less time expended of this algorithm compare to traditional algorithm. When in practical use, the time complexity of this algorithm is much less than traditional rough set algorithm.

6 Experiment

Selected the Mushroom Database in UCI machine learning Database and selected 15 condition attributes to compose a new Database, the experiment was made on PIII800. The algorithms of attribute reduction proposed in [8] and in this paper were used, the time of each algorithm expended shows in table 2.

Table 2 shows that time spending has a big difference between the two algorithm using on a same data set. This paper's algorithm superior the algorithm of [8] in time spending. The attribute reduction algorithm in [9] added a step to calculate attribute core and do not dismantle table. So this paper's algorithm superior the algorithm of [9] in efficiency.

Table 2. Reduction algorithm comparer on data set

attributes	Instance scale	Time spending of algorithm in [8](s)	Time spending of this paper algorithm (s)
15	3000	584.130	0.190

7 Conclusion

The information system attribute reduction is an important step of data processing, but found a table's minimum reduction is a NP-hard problem. The present algorithms calculate for whole set of information table. When system scale becomes larger, time spending is inestimable. The paper present a parallel reduction algorithm based on information entropy. The data table is dismantled, so the information system is scale down very much. The time complexity of the new algorithm was analyzed. Experiment shows that the new algorithm is more efficiency than traditional algorithm.

References

1. Wong, S.K.M., Ziarko, W.: On Optimal Decision Rules in Decision Tables, pp. 693–696. Bulletin of Polish Academy of Science, Poland (1985)
2. Xiaoju, W., Yun, J., Yonghua, L.: Significance of attribute evaluation based on dependable difference. Computer Technology and Development 19(1), 67–70 (2009)
3. Xu, F., Li, T., Li, H.: An improved discernibility matrix and the computation of the core. Computer Engineering & Science 31(2), 53–55 (2009)
4. Miao, D., Hu, G.: A heuristic algorithm for reduction of knowledge. Journal of Computer Research & Development 36(6), 681–684 (1999)
5. Mao, D., Li, D.: Rough Sets Theory Algorithms and Applications, pp. 139–170. Tsinghua University press, Beijing (2008)
6. Jiang, D.: Information Theory & Coding, 2nd edn., pp. 1–80. University of Technology Science and Technology of China Press, Hefei (2004)
7. Tian, H., Yan, L., Liu, X.: Data-table decomposition algorithm based on rough set. Computer Engineering and Design 30(5), 1198–1200 (2009)
8. Hu, X.H., Nick, C.: Learning in relational databases: A rough set approach. International Journal of Computational Intelligence 11(2), 323–338 (1995)
9. Ge, H., Li, L., Yang, C.: Improvement to quick attribution reduction algorithm. Journal of Chinese Computer Systems 30(2), 308–312 (2009)

Research of the Optimal Wavelet Selection on Entropy Function

Hong-feng Ma[1,2], Jian-wu Dang[1], and Xin Liu[2]

[1] College of Electronics and Information Engineering
Lanzhou Jiaotong University
Lanzhou, 730070, China
[2] Department of Electronic and Information Engineering
Lanzhou Polytechnic College Lanzhou, 730050, China
mhf418@yahoo.com, dangjw@mail.lzjtu.cn, liuxin1185@163.com

Abstract. The selection of the appropriate wavelet is the key issue in the wavelet packet de-noising. In this paper, it takes the entropy function as the evaluation criteria for the best packet basis function. The entropy is calculated by the coefficients of wavelet packet decomposition of speech signal to determine the appropriate decomposition. At the same time, four kinds of packet basis will be used to denoise in computer simulation experiments with wavelet packet threshold algorithm. The simulation results show that the optimal wavelet bases, which should be selected by two entropy function, not only can eliminate background noise to a large extent, but also raise the SNR of voice signal.

Keywords: Wavelet-packet analysis, Entropy function, Speech de-noising.

1 Introduction

It should inevitably suffer from a variety of interference noise in the speech communication process, such as environment noise, the noise inside the device and so on. The noise will make the speech processing system deteriorated, so speech enhancement is a kind of effective method to solve the noise pollution. The purpose of speech enhancement is to extract the pure signal from the contaminated signal and improve sound quality, so that the listeners will not feel tired and it can also improve speech intelligibility [1].

Wavelet transform is a new tool of signal processing, which is developed rapidly in last decade. It has the characteristics of multi-resolution analysis and the ability of denoting local signal characteristics in time domain and frequency domain, so it is a time-frequency analysis method which is a fixed window size but variable shape [2]. Because of the diversity of wavelet function, different wavelets will produce different results on the same issue. So how to select the optimal wavelet is a very important problem. In this paper, noisy signal will be dealt with wavelet packet threshold denoising algorithm, which is combined with entropy function to determine the methods of signal decomposition. It should compare to the effect of four wavelet base for selecting the appropriate wavelet base. The simulation results indicate the optimal wavelet base can suppress noise and reduce the loss of the signal to a certain extent.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 35–42.
springerlink.com

2 The Theory of Wavelet Packet Analysis

Wavelet packet analysis is an extension of wavelet transform. Each subband in various scales is divided once more and the high-frequency parts are further decomposed with the aid of wavelet packet decomposition. So it is more sophisticated than the wavelet decomposition.

If the signal can be expressed as $f(t)$ and the space can be expressed as V_j, that is $f(t) \in V_j$. If the scale space can be expressed as V_{j+1} and the wavelet space can be expressed as W_{j+1}, $f_{2n}^{j+1}(k)$ stands for the coefficients in scale space and $f_{2n+1}^{j+1}(k)$ stands for the coefficients in wavelet space. So wavelet packet decomposition algorithm is as follows[3]:

$$\begin{cases} f_{2n}^{j+1}(k)=\sum_m h_0(m-2k) f_n^j(m) \\ f_{2n+1}^{j+1}(k)=\sum_m h_1(m-2k) f_n^j(m) \end{cases} \tag{1.1}$$

Wavelet packet reconstruction algorithm is as follow[3]:

$$f_n^j(k)=\sum_m h_0(k-2m) f_{2n}^{j+1}(m)+\sum_m h_1(k-2m) f_{2n+1}^{j+1}(m) \tag{1.2}$$

Wavelet packet decomposition is that signal is projected onto the wavelet basis and it gets a series of coefficients in scale space and wavelet space. According to the actual need, these coefficients are processed and the retained coefficients are reduced to keep the original signal. Because the wavelet packet decomposition is orthogonal decomposition, there is no redundancy and the signal will be complete and reliable [4].

3 Entropy Function

In the wavelet packet decomposition process, the projection of the signal decomposed by best wavelet basis will be as large as possible in each subspace. Therefore, it needs to identify a criterion for selecting the best wavelet basis with minimal cost in possible wavelet basis. Generally, this criterion is known as “Information Cost Function”, which is a mathematical criterion to measure the effectiveness of the wavelet transformation. The criterion can also be referred to as “Entropy Function”. The smaller the entropy value, the better the corresponding wavelet.

Entropy function is an additive cost function with concentration. Entropy function can be expressed as E; signal can be expressed as S; the projection coefficients of the signal on the orthogonal wavelet can be expressed as S_i. So it is as follow:

$$\begin{cases} E(0)=0 \\ E(s)=\sum_i E(s_i) \end{cases} \tag{2.1}$$

Here are two common entropy function [5]:

(1) Shannon entropy

If $E_1(s_i) = -\sum_i s_i^2 \log(s_i^2)$, then

$$\begin{cases} E_1(s) = -\sum_i s_i^2 \log(s_i^2) \\ 0\log(0) = 0 \end{cases} \tag{2.2}$$

(2) SURE entropy

$$E(s) = -n + A(2 + p^2) + B \tag{2.3}$$

n is the length of a sequence for the entropy; p is representative of the threshold; A is expressed as the number that is greater than p^2 in the sequence S_i^2; B is expressed as the sum that is not greater than p^2 in the sequence S_i^2.

Shannon entropy is a monotonically increasing function from the above definition. Therefore, any small changes caused by the random variables will not lead to the mutation of the entropy, and it has a very good applicability. SURE entropy need to consider the threshold which will directly impact on the entropy, so it has certain restrictions. In the applications, it needs to consider the characteristics of the signal and actual needs to select the appropriate entropy function for improving the results of denoising, which can determine the optimal wavelet packet decomposition.

4 The Process of Wavelet Packet Threshold Denoising Based on Entropy Function

The mathematical mode of noisy signal can be expressed as follows [6]:

$$f(t) = s(t) + n(t) \tag{3.1}$$

The noisy speech signal, the pure signal without adding noise and the Gaussian White Noise denote by $f(t), s(t), n(t)$ respectively.

The noise must be inhibited in order to recover $s(t)$ from the original signal $f(t)$.

In the Matlab environment, the entropy function can combine with wavelet packet threshold algorithm for denoising and the process is as follow [7].

Firstly, choose a wavelet basis and determine the decomposition level. Noisy signal will be decomposed with N layer by "wpdec" function and get a complete binary tree. In this tree, each node has a set of wavelet packet coefficients called $\omega_{j,k}$, which includes the coefficients of pure signal called $c_{j,k}$ and the coefficients of the Gaussian White Noise called $d_{j,k}$.

Secondly, select an entropy function and calculate the entropy of each node in the complete binary tree with wentropy function. Then compare the value of each node from bottom to top for finding the best wavelet packet tree with "besttree" function.

Thirdly, select the appropriate threshold which can be expressed as λ. The right threshold for the coefficients of the best wavelet packet tree can be done with "wpthcoef" function to get the estimated wavelet coefficients called $\hat{\omega}_{j,k}$ for making the difference between $\hat{\omega}_{j,k}$ and $c_{j,k}$ as small as possible, which is formulated that $\left\| \hat{\omega}_{j,k} - c_{j,k} \right\|$.

Finally, the tree of the right threshold can be done for noise reduction and reconstruct speech signal with "wpdencmp" function.

The wavelet packet threshold algorithm based on entropy function can decompose both the high frequency part and low frequency part from the above process. At the same time, this algorithm can choose the right decomposition to generate a full binary tree and compare the entropy of each node in the tree. If the sum of the value of child node is greater than the value of parent node, the child node should be removed and find the best wavelet packet tree. According to the characteristics of the correlation both the signal and noise in the wavelet domain, the right threshold of the coefficients in best tree will be done for denoising and removing the useless information, then the left coefficients can be reconstructed speech signal.

5 Simulation Results and Analysis

The simulation experiment for speech denoising should be done with wavelet packet threshold algorithm in the Matlab7.0 environment, in which the different wavelet basis can be analyzed with the Shannon entropy or SURE entropy. The experimental requirements are as follow:

(1) Collect a voice as the original signal such as "Da Jia Zao Shang Hao" and the sample frequency is 8kHz.

(2) The added noise is Gaussian White Noise which is randomly generated from Matlab7.0 software.

(3) Use four wavelet basis such as db4, sym4, coif3 and bior2.2; the decomposition lever is four; the uniform threshold $\lambda = \delta\sqrt{2\ln N}$; Use the soft threshold function [8].

Therefore, the simulation graphs are as follow:

(1) Four wavelet basis in the condition of Shannon entropy for speech denoising.

Fig.1-4 are the simulation graphs of four wavelet basis in the condition of Shannon entropy for speech denoising, where the abscissa is time and vertical axis is amplitude.

From Fig.1-4, it can be seen that the characteristics of original speech signal should be remained in the experiment of four wavelet basis and the interference can be effectively reduced. However, there are some distortions in the simulation graphs. For example, the distortion on the first vowel is more obvious from Fig.1, Fig.2 and Fig.4, but the distortion is smaller and it has better effect in denoising on Fig.3.

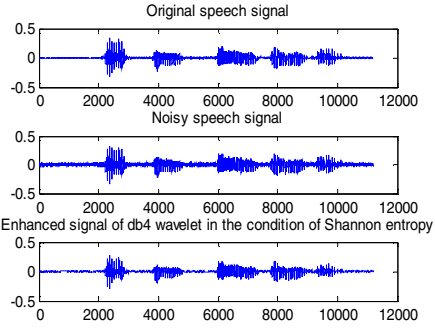


Fig. 1. The simulation graph of db4 wavelet

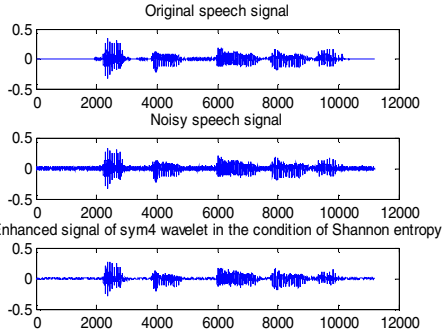


Fig. 2. The simulation graph of sym4 wavelet

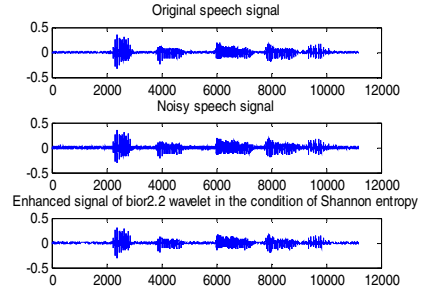


Fig. 3. The simulation graph of bior2.2 wavelet

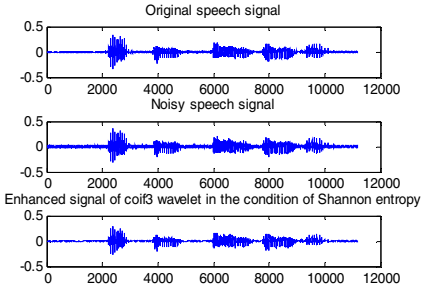


Fig. 4. The simulation graph of coif3 wavelet

(2) Four wavelet basis in the condition of SURE entropy for speech denoising.

Fig.5-8 are the simulation graphs of four wavelet basis in the condition of SURE entropy for speech denoising, where the abscissa is time and vertical axis is amplitude.

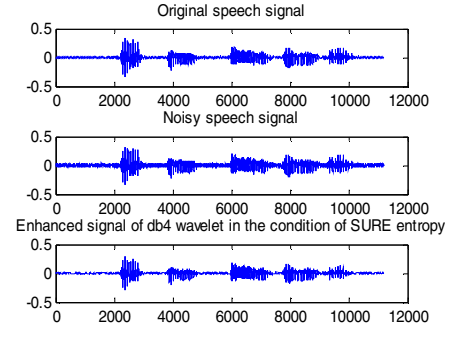


Fig. 5. The simulation graph of db4 wavelet

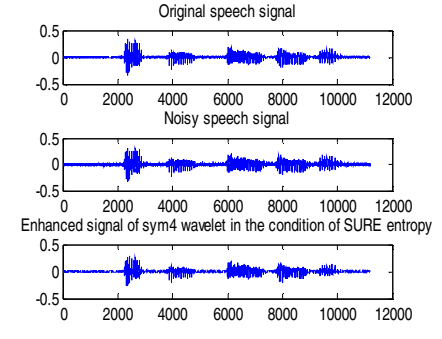


Fig. 6. The simulation graph of sym4 wavelet

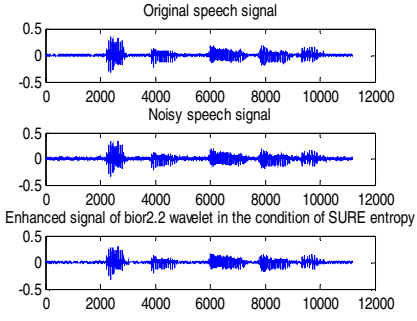


Fig. 7. The simulation graph of bior2.2 wavelet

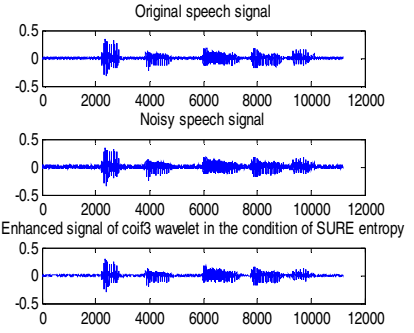


Fig. 8. The simulation graph of coif3 wavelet

From Fig.1-8, it can be seen that the characteristics of original speech signal should be remained in the experiment of four wavelet basis and there are some distortions in the simulation. The effect of bior2.2 wavelet is still the best. Compared Fig.3 and Fig.7, it can be seen that there are more discontinuous shock point which will cause the distortion for part of the enhanced speech in Fig.3, but in Fig.7 the simulation graph is smoother and the effect is better.

The standard for denoising evaluation is signal to noise ratio (SNR). If the difference of SNR is greater, the effect of denoising will be better and the reconstructed signal will be closer to pure speech. So the simulation results of SNR are as follow:

(1) Table 1 describes the change of four wavelet basis about SNR in the condition of Shannon entropy.

Table 1. The change of four wavelet basis about SNR in the condition of Shannon entropy

Entropy function	***Wavelet basis***	***Input signal to noise ratio*** SNR_{in}	***Output signal to noise ratio*** SNR_{out}	***The difference of SNR***
Shannon	db4	10.1767	14.8147	4.6380
	sym4	10.3917	14.9243	4.5326
	bior2.2	10.2738	18.3174	8.0436
	coif3	10.3151	14.8459	4.5308

It can be seen from tableI that the algorithm with four wavelet basis has improved the noisy signal to some extent, but enhanced signals have some residual noise. In particular, SNR for bior2.2 wavelet can increase significantly and the effect for denoising is more obvious.

(2) Table 2 describes the change of four wavelet basis about SNR in the condition of SURE entropy.

Table 2. The change of four wavelet basis about SNR in the condition of SURE entropy

Entropy function	*Wavelet basis*	*Input signal to noise ratio* SNR_{in}	*Output signal to noise ratio* SNR_{out}	*The difference of SNR*
SURE	db4	11.2636	14.7650	3.5014
	sym4	9.8913	14.8824	4.9911
	bior2.2	8.7304	33.3983	24.6679
	coif3	10.4619	14.8066	4.3447

Table 2 shows that the difference of SNR of bior2.2 wavelet is the greatest in the condition of Shannon entropy and SURE entropy. Not only a substantial increase for SNR_{out} of bior2.2 wavelet in the condition of SURE entropy, but enhanced speech signal is also clear and barely hear the residual noise, so it is close to the original voice.

The comparison shows that the effect of bior2.2 wavelet in graph and SNR is quite satisfactory, so bior2.2 wavelet is more appropriate wavelet.

6 Conclusion

The experiment in this paper mainly studies the three parts about the graph, SNR and hearing results by combining the wavelet packet threshold algorithm with entropy function to select the optimal wavelet bases. And the experimental results show that bior2.2 wavelet selected by two entropy function is the optimal wavelet basis, which can effectively suppress noise, improve SNR and have the better hearing. Due to the variety of entropy function, how to select the appropriate entropy function for selecting the best wavelet packet basis is studied the main problem in future work.

Acknowledgements. The work is supported by the National Education Commission Doctor-Training Units Foundation (No. 20060732002), Natural Science Foundation of Gansu Province (No. 096RJZA084) and the Tutor Research Projects on Education Department of Gansu Province (No. 0814-4).

References

1. Zhang, X., Chen, L., Yang, J.: Modern Speech Processing Technology and Applications. Machine Industry Press, Beijing (2003)
2. Guo, J., Wang, F.: The study of signal denoising based on wavelet packet Transform. Modern Electronic Technology 19(238), 55–2 (2007)
3. Ge, Z., Sha, W.: Wavelet Analysis Theory and MATLAB R2007 Implementation. Electronic Industry Press, Beijing (2007)
4. Cohen, I., Raz, S., Malah, D.: Orthonormal shift-invariant wavelet packet decomposition and representation. Signal Processing 57(3), 251–270 (1997)

5. Fei Sike Technology R&D Center. Wavelet Analysis Theory and MATLAB7 Implementation. Electronic Industry Press, Beijing (March 2005)
6. Deng, Y.: The study of speech denoising based on wavelet packet threshold algorithm. Speech Technology 09(33), 65–5 (2009)
7. Zhang, L., Qin, H., Yu, C.: The study based on wavelet threshold algorithm for denoising. Computer Engineering and Applications 44(9), 172–174 (2008)
8. Donoho, D.L.: Denoising by soft thresholding. IEEE Trans. on Information Theory 41(3), 613–627 (1995)

Designing an Efficient Scheduling Algorithm for P2PTV System

Kai Zhang and Kan Li*

Beijing Key Lab of Intelligent Information Technology, School of Computer Science and Technology
Beijing Institute of Technology
Beijing, China
{zhangk,likan}@bit.edu.cn

Abstract. Scheduling the transmission of stream chunks is one of the main challenges in P2PTV system. In this paper, in order to improve the performance of push-based scheduling algorithm by exploiting chunk and peer's characteristics in the heterogeneous scenario, we first present a 2-stage buffer window model and define redistribution capacity factor to describe chunk and peer's characteristics. Then, we propose a new scheduling algorithm, called Urgent then Latest chunk / Redistribution Capacity peer (UL/RC). The UL/RC includes two components: UL chunk scheduler and RC peer scheduler. UL chunk scheduler uses the urgent chunk priority strategy to achieve low loss rate. RC peer scheduler employs high redistribution capacity peer priority strategy to achieve low delivery delay. Experimental results indicate that our scheduling algorithm outperforms state-of-the-art approaches.

Keywords: scheduling algorithm, P2PTV, heterogeneity, delivery delay, loss rate.

1 Introduction

With the development of broadband network and multimedia technology, P2PTV systems are becoming important applications in the Internet. A number of P2PTV systems (e.g., PPLive, UUsee and Zattoo) have emerged. These system relay on distributed, swarm-like dissemination mechanisms: the stream is divided into small parts, so-called chunks, which follow random, independent paths in the peer population.

The scheduling algorithm is one of main drivers of the performance of P2PTV systems, and refers to the decision of which chunk will be sent to (or retrieved from) which peer. The scheduling algorithm can be categorized as push-based or pull-based algorithms according to depending on whether chunk transmission is initiated by the sender or receiver respectively. Compared with pull-based algorithms, push-based algorithms can achieve lower delivery delay which is desirable for real-time system.

* Corresponding author.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 43–50.
springerlink.com

Push-based scheduling algorithms have attracted significant interest from research communities in recent years 1267.

Push-based scheduling algorithms can achieve good performance in homogeneous scenario. However, their performance worsen in heterogeneous scenario where peers have different upload bandwidth 12. We argue that the reason is that push-based algorithms do not fully exploit chunk and peer's characteristics. These characteristics include chunk factors such as *urgent* chunk (close to playback deadline), *latest* chunk (recently generated at the source) and peer factors such as *deprived* peer (having more missing chunks), high *capacity* peer (having high upload bandwidth). Although some algorithms exploit some factors mentioned above, there has not been a scheduling algorithm taking all factors into consideration so far.

In this paper, we focus on push-based algorithms. Our aim is reducing loss rate and delivery delay by carefully exploiting chunk and peer characteristics in heterogeneous scenario. Our contributions are the following. Firstly, we present a 2-stage buffer window model to tag urgent and latest chunk and map the urgent chunk between peers. Secondly, we define a redistribution capacity factor which can easily identify the high capacity and deprived peers. Finally, we propose a scheduling algorithm, called Urgent then Latest chunk / Redistribution Capacity peer (UL/RC). In this algorithm, peers prefer to select the urgent chunk, and select the latest chunk under not available the urgent chunk. Peers choose a destination peer with a probability proportional to its redistribution capacity factor. Simulation shows that the UL/RC algorithm can achieve better performance and outperform other two typical algorithms.

The remainder of this paper is organized as follows. Section 2 describes some related work. In section 3, we first present 2-stage model and define redistribution capacity factor, then propose our scheduling algorithm (UL/RC). Section 4 conducts simulations to evaluate the performance of our algorithm. Finally, we draw our conclusions and describe future work in Section 5.

2 Related Work

The performance of P2PTV system critically depends on the scheduling algorithm, much research focuses on developing efficient scheduling algorithms. Several pull-based algorithms have been proposed. Coolstreaming 3 proposed rare-first scheduling algorithm which assigns higher priority to the chunk with the smallest number in its neighbors. The authors in 4modeled the pull-based scheduling problem as a classical min-cost network flow problem, and proposed a distributed heuristic algorithm to maximize network throughout. The author in 5 proposes a simple queue-based chunk scheduling method to achieve a higher streaming rate in the P2P streaming system.

Some push-based scheduling algorithms have been proposed and analyzed. In 6, the authors prove the rate optimality of the algorithm, called Deprived peer / Random Useful chunk (DP/RU), which a peer distributes a chunk to a neighbor with the largest number of missing chunks. In 7, delay optimality of the Random Peer / Latest Blind chunk (RP/LB) algorithm is proven assuming all peers are characterized by the same upload bandwidth. More recently, in 1, the authors proposed several scheduling

algorithm, such as Latest Useful chunk / Useful Peer (LU/UP), Latest Useful chunk / Deprived Peer (LU/DP), and show these algorithms achieve near-optimal rate and delay performance. But these algorithms apply only to homogeneous scenario.

Several preliminary investigations have been carried out in heterogeneous scenario. In 8, the authors obtained lower delay bounds and show that heterogeneity may be exploited to reduce the delivery delay. But they do not devise an efficient scheduling algorithm. In 2, the authors proposed a scheduling algorithm, called Latest Useful chunk / Bandwidth aware Peer (LU/BP), which peers prefer to schedule the latest chunk to the peer with high upload bandwidth. The LU/BP can optimize delay by carefully exploit peers' upload bandwidth. But the LU/BP algorithm only focuses on achieving lower delay. Compared with LU/BP, our UL/RC algorithm not only devotes to reduce loss rate performance, but also tries to reduce delivery delay as possible in heterogeneous scenario.

3 Prepare Your Paper Before Styling

In this section, we propose Urgent then Latest chunk, Redistribution Capacity peer (UL/RC) scheduling algorithm. Before describing our algorithm, we present a 2-stage buffer window model and define redistribution capacity factor to describe chunk and peer's characteristics.

3.1 2-Stage Buffer Window Model

Chunk characteristics (urgent, latest) are important to design an efficient chunk scheduling algorithm. But these characteristics are ambiguous or not defined clearly. We need a method to precisely describe these characteristics. In this section, we present a 2-stage buffer window model to tag urgent chunk and latest chunk, and map urgent chunk between peers.

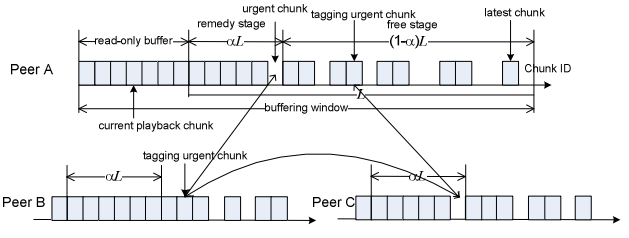


Fig. 1. Buffer window model and chunk map

Figure 1 shows a 2-stage buffer window model. According to the playback time of chunks, the window on the right side of the read-only buffer is divided into two stages: the remedy stage and the free stage. The remedy stage contains the chunks with the most urgent playback time (or smallest Chunk IDs) and the free stage contains the chunks with the latest playback time (or largest Chunk IDs). With the sliding forward of buffer window, a chunk first enters the free stage then arrives at

remedy stage. When a missing chunk enters the remedy stage, the missing chunk becomes the *urgent chunk* and the peer becomes an *urgent peer*. To acquire the urgent chunk, the urgent peer generates a signaling message and sends to a neighboring peer who has the urgent chunk. Once receiving the message, the neighboring peer tags the urgent chunk as *tagging urgent chunk*. The map of tagging urgent chunk and urgent peer between peers is also shown in Figure 1.

The number of tagging urgent chunks is determined by the length of the remedy stage. On the one hand, we hope that the number of tagging urgent chunks is as small as possible because the latest chunk has more chance to be scheduled so that peers can achieve low delivery delay. On the other hand, more tagging urgent chunks scheduled can help peers to reduce chunk loss rate. Obviously, there is a delivery delay/loss rate performance trade-off which can be achieved by selecting the proper length of remedy stage. Theoretically, the length of remedy stage l_i should be

$$l_i = max\left\{RRT_{ij} \mid j \in N(i)\right\}, \tag{1}$$

where RRT_{ij} is the round-trip-time between peer i and peer j, and $N(i)$ is the set of neighbors of peer i. Practically, it is difficult to determine l_i because of traffic congestion. In this paper, we give a factor $\alpha \in [0, 1]$ to tune l_i, and $l_i = \alpha L$, where L is the total length of remedy and free stages. We will figure out a proper α by simulation in Sect. 4.

3.2 Redistribution Capacity Factor

Since peer scheduling favoring high capacity peers and deprived peers can improve the performance. We need a method to identify these peers. We define the ratio of the incoming data that can be redistributed to other peers as a redistribution capacity factor, and label it with k. $k(i)$ can be estimated as an upload capacity rate $c_u(i)$ divided to download capacity rate $c_d(i)$ of the peer i, that is $k(i)=c_u(i)/c_d(i)$. k varies from zero to infinity. Assume that a chunk is delivered to a peer during a time slot. If k =1, it means that the peer can redistributes the full chunk. If k=2, it means that the peer redistributes two copies of the chunk. If k=0.5, it means that the chunk redistribution needs two time slots.

In an ideal world, each peer can sustain its streaming rate and utilize its uplink bandwidth fully. Then each peer has a constant redistribution capacity factor because k only depends on the peer's upload bandwidth. In reality, the k values of the peers would change over time. We usually have a much diverse set of redistribution capacity factor.

- k is lower than 1
  - Free-riders (k =0)
  - Peers with low uplink bandwidth (k<1)
- k is higher than 1
  - High capacity peers, i.e., peers with high upload bandwidth
  - Deprived peers, i.e., peers that do not receive enough chunks, but redistribute more than one copies of chunk to other peers.

Obviously, we can easily identify deprived or high capacity peers by checking k value of peers. It can direct us to design a new peer scheduling strategy to achieve better performance. Note that the peer is a free-rider when k=0. This indicates that we can design an incentive compatible scheduling algorithm. We do not analyze this case because it is beyond the scope of this paper.

3.3 UL/RC Scheduling Algorithm

The proposed UL/RC scheduling algorithm includes two schedulers: UL chunk scheduler and RC peer scheduler. The UL chunk scheduler is responsible for selecting the proper chunk. The idea of UL chunk scheduler is that the peers prefer to schedule the tagging urgent chunk and schedule the latest chunk under not available the tagging urgent chunk. The RC peer scheduler is responsible for choosing a destination peer. The idea of RC scheduler is choosing a destination peer with a probability proportional to its redistribution capacity factor. The rationale behind RC scheduler is simple: the destination peer has more capacity in redistributing chunks.

We first give some symbols used here. $C_u(i)$ is the set of tagging urgent chunks of peer i has received, $N_u(i,c)$ is the set of urgent peers that miss a urgent chunk c, $C(i)$ is the set of chunks that peer i has received, and $N(i,c)$ is the set of peers that do not receive chunk c. The UL/RC algorithm can be described as following three steps:

1) Chunk selection: Peer i inspects its buffer window. If $C_u(i)\neq\varnothing$, peer i random selects a tagging urgent chunk c from $C_u(i)$; otherwise, peer i selects the latest chunk c.

2) Peer selection: After selecting a chunk c, peer i starts to select a destination peer j. If c is a tagging urgent chunk, peer j is selected from $N_u(i,c)$; otherwise, peer j is chosen from the set N(i,c). The probability $\pi_i(j)$ of selecting peer j can be expressed as

$$\pi_i(j)=\frac{k(j)}{\sum_{l\in S}k(l)};\quad with\ S=\begin{cases}N_u(i,c) & if\ c\in C_u(i)\\ N(i,c) & otherwise\end{cases} \tag{2}$$

3) Chunk scheduling: Once j is selected, peer i starts transmitting chunk c. After the transmission is completed, a new push phase starts. Peer i then delivers c to another peer p in $N_u(i,c)$ or N(i,c) according to $\pi_i(p)$. In case all neighbors already have c, peer i selects another chunk and schedules it using the same process mentioned above.

4 Experiment

In this section, we validate and evaluate our scheduling algorithm by conducting extensive simulations, and compare our algorithm with two typical approaches: LU/UP and LU/BP which have been introduced in related work.

4.1 Network, Bandwidth Models and Simulation Setting

We use the P2PTVSim simulation tool developed within the Napa-Wine EU project 9. The simulator allows describing bandwidths and delays between peers and building many overlay topologies. The overlay model we use here can be described by a random graph that belongs to the class of *Quasi Regular Random Graphs* (QRRGs) in which the arcs are randomly placed 10. Each peer selects d_0 random neighbors. Links are bidirectional, thus the number of neighbor is $2d_0$ on average.

We model bandwidths distribution as three-class model: low-bandwidth peers (having a bandwidth equal to $0.5B$), mid-bandwidth peers (having a bandwidth equal B), and high-bandwidth peers (having a bandwidth equal $2B$). The fraction of high-bandwidth peers in the system is $h/3$, the fraction of low-bandwidth peers is $2h/3$, and the fraction of mid-bandwidth peers is $1-h$. As a result, the average bandwidth is B. We call h as *heterogeneity factor*, and $h \in [0,1]$. $h=0$ corresponds to the homogenous case. $h>0$ corresponds to the heterogeneous case.

The overlay network considered consists 1000 peers and an average node degree is 16 ($d_0=8$). The source is a special additional node with a bandwidth of 5.0Mb/s. Chunk size is equal to 0.1Mb. The source emits chunks at the rate of 1Mb/s. The average bandwidth B is 1Mb. The video content consists of 500 chunks. Longer video has been simulated, leading to a similar conclusion. The size of buffering window is 5s, i.e., 50chunks. The read-only buffer is set 10 chunks, and α is set 0.2.

4.2 Simulation Results and Analysis

We first evaluate the performance of our algorithm by comparing with two typical algorithms: LU/UP, LU/BP. Figure 2 reports the mean delivery delay as a function of the heterogeneity factor for the three different algorithms respectively. It is observed that three algorithms exhibits similar delivery delay performance in homogenous scenario ($h=0$). But the case is different in heterogeneous scenario. LU/UP increases the delivery delay with increasing h. On the contrary, LU/BP and UL/RC reduce the delivery delay with increasing h. This indicates that the delivery delay performance can benefit from heterogeneity by preferring to schedule high bandwidth or redistribution capacity peers. The reason is that these peers can contribute more to chunk diffusion. We also notice that the delivery delay of UL/RC is slightly higher than that of LU/BP. This result indicates that bandwidth might be a more influence factor to chunk diffusion.

Let us now consider chunk loss rate as a function of the heterogeneity factor for the three different algorithms. The results are reported in Figure 3. It can be observed that three algorithms achieve similar chunk loss rate performance in homogenous scenario ($h=0$). With increasing h, however, the chunk loss rate of UL/RC is significantly lower than that of two other algorithms. This indicates that UL chunk scheduling strategy plays important roles in reducing chunk loss. By observing Figure 2 and Figure 3, we can draw such conclusion: UL/RC outperforms other two algorithms on the whole although UL/RC achieves a little higher delivery delay than LU/BP.

We also check the impact of α value on the performance in our UL/RC algorithm. Figure 4 shows delivery delay under different value of α∈[0, 1] and different value of *h*. We find that delivery delay is small and changes little when α is set small. But delivery delay increases rapidly when α is greater than 0.2. An intuitive explanation is that the latest chunks have little chance to be scheduled when α is set large. Figure 5 shows the loss rate under different value of α and *h*. We observe that peers experience high loss for small α and the loss rate rapid decreases with increasing α. We also notice that the loss rate tends to zero when α reaches 0.6. This is because that the more urgent chunks are scheduled for large α so that peers can avoid chunk loss. Analyzing the results of Figure 4 and Figure 5, α=0.3 might be a good choice. In this case, UL/RC can achieve relative small delivery delay and loss rate.

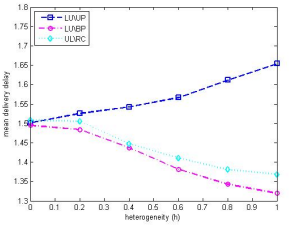


Fig. 2. Delivery delay versus heterogeneity factor *h*, with α=0.2

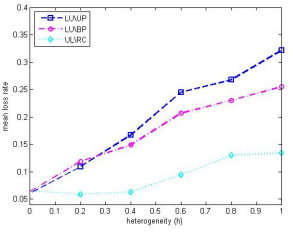


Fig. 3. Loss rate versus heterogeneity factor *h*, with α=0.2

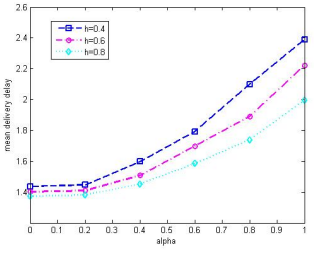


Fig. 4. Delivery delay versus α

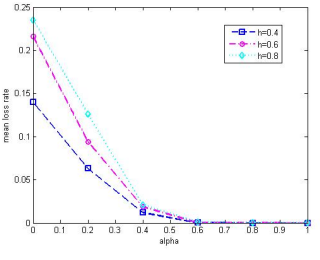


Fig. 5. Loss rate versus α

5 Conclusions and Future Work

In this paper, we investigate the characteristic of chunk and peer in P2PTV system, and show that the scheduling algorithm can achieve low delivery delay and loss rate by carefully exploiting these characteristics in heterogeneous scenario. We first present a 2-stage buffer window model and define redistribution factor to describe chunk and peer's characteristics, and then propose our UL/RC scheduling algorithm which fully exploits chunk and peer's characteristics. Simulations show that our UL/RC scheduling algorithm outperforms state-of-the-art approaches.

As a future work, we will evaluate our UL/RC algorithm in dynamic network.

Acknowledgment. The research was supported in part by Natural Science Foundation of China (No.60903071), the Ministerial Level Advanced Research Foundation, and Beijing Key Discipline Program.

References

1. Bonald, T., Massoulie, L., Mathieu, F., Perino, D., Twigg, A.: Epidemic Live Streaming: Optimal Performance Trade-Offs. In: ACM SIGMETRICS 2008, Annapolis, MD (June 2008)
2. Couto da Silva, A.P., Leonardi, E., Mellia, M., Meo, M.: Exploiting heterogeneity in p2p video streaming. IEEE Trans. on Computers (2010) (in press)
3. Zhang, X., Liu, J., Li, B., Yum, T.-S.P.: CoolStreaming/DONet: A data-driven overlay network for live media streaming. In: Proceedings of IEEE INFOCOM 2005, Miami, FL, USA, pp. 2102–2111 (March 2005)
4. Zhang, M., Xiong, Y., Zhang, Q., Yang, S.: Optimizing the Throughput of Data-Driven Peer-to-Peer Streaming. IEEE Trans. on Parallel and Distributed System 20(1), 97–110 (2009)
5. Guo, Y., Liang, C., Liu, Y.: AdaptiveQueue-based Chunk Scheduling for P2P Live Streaming. In: IFIP Networking (May 2008)
6. Massoulie, L., Twigg, A., Gkantsidis, C., Rodriguez, P.: Randomized decentralized broadcasting algorithms. In: IEEE INFOCOM 2007, Anchorage, AL (May 2007)
7. Sanghavi, S., Hajek, B., Massoulie, L.: Gossiping with multiple messages. In: IEEE INFOCOM 2007, Anchorage, AL (May 2007)
8. Picconi, F., Massoulie, L.: Is there a future for mesh-based live video streaming? In: IEEE P2P 2008, Aachen, Germany (September 2008)
9. Network-Aware P2P-TV Application over Wise Networks, `http://www.napa.wine`
10. Bollobas, B.: Random Graphs. CambridgeUniversity Press (2001)

A Timely Occlusion Detection Based on Mean Shift Algorithm

Ai-hua Chen, Ben-quan Yang, and Zhi-gang Chen

Taizhou University
Taizhou 318000, China
chen_1216@163.com

Abstract. Mean shift algorithm has attracted much attention in computer vision and has recently shown promising performance in the challenging problem of visual tracking, but it is difficult to deal with occlusion. In this paper, a timely occlusion object detection based on mean shift is proposed. By analyzing occlusion process, it is evident to find that occluded size is increasing and occlusion patch lies to edge of objects at the beginning. So object model is divided into several parts. In order to reduce computation, only edge patches is considered. If Bhattacharyya coefficient of one patch decreases greatly and other patches change faintly, it means that object is occluded in this area. In this method, object model is divided into two parts, one is occlusion part, the other is no occlusion part, the no occlusion part of the object can be obtained to track object continuatively until object is occluded totally. Experiments show that compared with whole object model judges, it is timely to detect occlusion and deals with occlusion successfully.

Keywords: object tracking, occlusion detection, mean shift.

1 Introduction

Object tracking is an important and challenging task within the field of computer vision. The growth in this area is being driven by the increased availability of inexpensive computing power and image sensors, as well as the inefficiency of manual surveillance and monitoring systems. Detection and tracking of moving objects can be applied in many different applications such as surveillance, perceptual user interfaces, augmented reality, object-based video compression and driver assistance [1].

Recently, mean shift algorithm has attracted much attention in computer vision and has recently shown promising performance in the challenging problem of visual tracking [2, 3]. Hager, Dewan, and Stewart propose a modification using multiple kernels [4]. The mean shift algorithm can also be applied to tracking objects with scale changes by Collins [5]. However, all these approaches require the entire object patch in each frame be visible. So object occlusion will cause tracking failure because of the invisibility of object occlusion parts [6]. Traditional mean shift tracking algorithms do not consider occlusion effects caused by inevitable object overlapping

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 51–56.
springerlink.com

for occlusion objects. Consequently, during the process of tracking occluded objects, the pixels that belong to the occluding objects are regarded as those of the occluded objects. This causes disturbances that lead to tracking failures [7].

By analysis of occlusion course, it is evident to find that occluded size of object is increasing. Firstly occluded patch is very small, then the size of occluded patch becomes bigger and bigger as time goes away. According to analyzes above, a timely occlusion object detection method based on mean shift is proposed in this paper. The remainder of this paper is organized as follows. Section 2 presents relation work about occlusion object tracking. Section 3 reviews the relevant aspects of mean shift algorithm briefly. Section 4presents proposed method to detect and track occlusion object. The experimental analysis of the proposed approach and the conclusions are given in Section 5 and 6.

2 Relation Work

Plenty of work has been done on occlusion object tracking. Various methods have been proposed to detect and handle occlusion object tracking. Senior et al. proposes an approach to track objects through occlusion using appearance models [8]. The appearance model is applied to determine the depth ordering of object during occlusion. But heavily and totally occluded object cannot be solved by this approach. Kaucic et al. and Amitha Perera et al. propose an approach to solve total occlusion during long gaps [9, 10]. The two segment of the moving object trajectory should be linked when the object has been occluded by the background for a long time. When the moving object has been totally occluded by the foreground objects in a long gap, the efficient approach of handling spilt-merge would be used to maintain the identity through long sequences. Han et al. suggests a multiple object tracking algorithm [11]. During tracking, the moving object's current state is estimated based on a state sequence instead of one previous state. So the optimal state sequence that maximizes the joint state-observation probability can be found. Temporal distractions such as occlusion, background clutter, and multi object confusion can be conquered. Zheng Li introduces a method to tracking occlusion object which based on velocity and appearance in [7]. Occlusion object is divided into two parts. One is occlusion patch, other is no occlusion patch. When object is occluded, according to no occlusion to locate object, but it does not deal with how to judge when object is occluded. Most of the tracking techniques use subjective evaluation method. E. Loutas proposes mutual information based metrics as measures of tracking reliability. Moreover, the use of the metric is extended to the analysis of partial and total occlusion in object tracking [12]. But the computation of method is huge, it is difficult to realize in object tracking.

3 Review Mean Shift Algorithm

The heart of mean shift algorithm is computation of an offset from location vector x to a new location $x^{'} = x + \Delta x$, according to the mean-shift vector.

$$\Delta x = \frac{\sum_a K(a-x)w(a)a}{\sum_a K(a-x)w(a)} - x \tag{1}$$

w is the weight. K is a suitable kernel function and the summations are performed over a local window of pixels a around the current location x. A "suitable" kernel K is one that can be written in terms of a profile function k such that $K(y) = k(\|y\|^2)$ and profile k is nonnegative, nonincreasing, piecewise continuous [4].

An important property of the mean-shift algorithm is that the local mean shift vector computed at position x using kernel k points is opposite to the gradient direction of the convolution surface

$$J(x) = \sum_a g(a-x)w(a) \tag{2}$$

Where g satisfies $g(x) = -k'(x)$. $J(x)$ is designed by using Bhattacharyya coefficient to measure the similarity of two kernel histograms representing the object model and the candidate model respectively. The algorithm finds the maximum of the Bhattacharyya coefficient given an object model and a starting region. Based on the mean shift vector, which is an estimation of the gradient of the Bhattacharyya function, the new object location is calculated. This step is repeated until the location no longer changes significantly.

4 Proposed Tracking Method

Analysis of occlusion process, it is easy to find that size of the occluded object is increasing and occlusion patch lies to the edge of object at the beginning. In order to timely detect when object is occluded and the occluded area, object model is divided into several parts. The number of patches is concerned with the size and shape of tracking objects, which is decided before tracking. In order to reduce computation, we only consider edge patches. As is shown in Fig. 1, object model is divided into three sections with two modes and we get four edge patches.

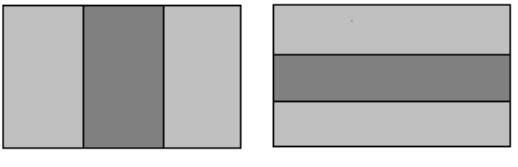


Fig. 1. Schematic diagram of model partition

During normal tracking process, Bhattacharyya coefficient of edge patches changes little among frames. When object is occluded, Bhattacharyya coefficient of occlusion part decreases greatly and other no occlusion parts change faintly. So occlusion is detected and the occlusion area is found. According to whether is occluded, object is

divided into two parts. At the beginning, only partial object is occluded, object can be located based on no occlusion patch. According to information of velocity of object moving, the ratio of these two parts is adjusted automatically. When the proportion of occlusion part exceeds a threshold which is given before tracking, the remainder can not describe characteristic of tracking object. It is regarded that object is occluded totally.

When object is occlusion totally, all information of object is lost, traditional tracking algorithm can not locate object. In object tracking, Kalman Filter is often used to forecast object position [13]. In normal tracking, the state of object moving is recorded which is used to initialize Kalman Filter parameters. In order to avoid that forecast position of object appears large error, forecast process appends restriction which is based on object moving fitting curve getting according to object motion before occluded.

After a few frames, object maybe appears in the scene again, it is necessary to find object again. Whether finds object again is decided by performance of forecast algorithm and time consumption of occluding. We plan to search object around forecast position. The size of search area is determined by the length of occlusion time.

5 Experimental Results

The proposed algorithm is tested using VC.NET programs on an AMD 3.0GHz machine. In order to show performances of suggested method, the performances of proposed method is compared with traditional method.

Fig. 2 shows the tracking results of car sequence in which there are 500 frames of 720×576 pixels. In the sequence, car is covered by tree. The object model is marked with 120×60 pixels rectangle in sequence. It is divided into three sections in horizontal direction and four sections in vertical direction, so we get four edge patches. As is shown in Fig. 2, occlusion is detected in frame 188. Object is occluded totally in frame 196. Frame 205 shows forecast position and object is found again in frame 216.

Fig. 2. The result of proposed method is shown (the frames 188,196, 205 and 216)

In general, occlusion is detected based on the whole model. That is to say, when Bhattacharyya coefficient of whole model decreases greatly, occlusion is detected. Fig. 3 shows traditional method to detect occlusion. As occlusion is not detected timely, tracking location deviates its actual position. Occlusion is detected in frame 194. Compared with proposed method, the performances of tracking become worse.

Fig. 3. The result of traditional method is shown (the frames 192,194, 205 and 216)

6 Conclusions

Occlusion detection is a challenging task in robust object tracking. A timely occlusion object detection based on mean shift is proposed in this paper. It divides object model into several parts and selects edge sections to compute Bhattacharyya coefficient. If Bhattacharyya coefficient of one of these sections decreases greatly and the other changes little, it means object is occluded. Then object is divided into two parts, according to no occlusion part, tracking process continues until object is occluded totally. During totally occlusion, Kalman Filter is used to forecast object position. Moreover, proposed method appends object moving fitting curve to restrict that forecast position appears large error. Experiments show that it is timely to detect occlusion and deals with occlusion successfully.

References

1. Ying, M., Jingjue, J.: Detection and tracking of moving objects under occlusion and shadow. In: Proc. of SPIE, vol. 6044 (2005)
2. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-Based Object Tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence 25(5), 564–577 (2003)
3. Peng, N., Yang, J., Liu, E.: Model update mechanism for mean-shift tracking. Journal of Systems Engineering and Electronic 16(1), 52–57 (2005)

4. Fan, Z., Wu, Y., Yang, M.: Multiple collaborative kernel tracking. In: Proc. of CVPR, pp. 502–509 (2005)
5. Collins, R.T.: Mean-Shift Blob Tracking through Scale Space. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2003, vol. 2, pp. 234–240. IEEE, Vancouver (2003)
6. Li, Z., Tang, Q.L., Sang, N.: Improved mean shift algorithm for occlusion pedestrian tracking. Electronics Letters 44(10) (2008)
7. Li, Z.: Improved mean shift algorithm for multiple occlusion target tracking. Optical Engineering 47(8) (2008)
8. Senior, A., Hampapur, A., Tian, Y.L., Brown, L., Pankanti, S., Bolle, R.: Appearance models for occlusion handling. In: 2nd Intl. Workshop Performance Eval. Track. Surveill. Syst. (2001)
9. Kaucic, R., Perera, A.G., Brooksby, G., Kaufhold, J., Hoogs, A.: A unified framework for tracking through occlusions and across sensor gaps. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 990–997. IEEE (2005)
10. Amitha Perera, A.G., Srinicas, C., Hoogs, A., Brooksby, G., Hu, W.: Multiobject tracking through simultaneous long occlusions and split-merge. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 666–673. IEEE (2006)
11. Han, M., Xu, W., Tao, H., Gong, Y.: An algorithm for multiple object trajectory tracking. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1864–1871. IEEE (2004)
12. Loutas, E., Nikou, C., Pitas, I.: Information theory-based analysis of partial and total occlusion in object tracking. In: Proceedings 2002 International Conference on, vol. 2 (2002)
13. Zou, X., Li, D., Liu, J.: Real-time vehicles tracking based on Kalman filter in an ITS. In: Proc. of SPIE, vol. 6623 (2008)

Niche Particle Swarm Algorithm and Application Study*

Chao-li Tang, You-rui Huang, and Ji-yun Li

Institute of Electrical and Information Engineering
Anhui University of Science and Technology
Huainan, Anhui232001, China
{chaolitang,hyr628}@163.com

Abstract. To the multi-modal function optimization problem, after analyzing characteristics and deficiencies of traditional niche genetic and niche clonal selection algorithms, it proposes a niche particle swarm algorithm (NPSA), which based on principle of particle swarm algorithm. By the convergence analysis and simulation experiments of four typical multi-modal functions, conclusions show that the NPSA has eximious simplicity and high efficiency.

Keywords: multi-modal function optimization, niche technology, particle swarm algorithm.

1 Introduction

After make mathematical modeling for practical problems, it can be abstracted as an optimization problem of valued functions. But in massive solution calculations of practical optimization problems, it needs not only to find global optimal solution in the feasible domains, but also to search for more global optimal solutions and significant optimal solutions, providing multiple choices or many-sided information. This problem is called multi-modal or multi-peak functions optimization problem. And functions optimization problem can transform into seeking maximum problem. It should be studied that how to construct an optimization algorithm, finding out all global maximum points and multiple partial maximum points as possible.

To the multi-modal functions optimization problem, traditional heuristic information search algorithms, such as simulated annealing method and artificial neural network, exist the problem, how to avoid losing local extreme points. The characteristics of Genetic Algorithms(GA) and Artificial Immune Algorithm(AIA) are using population search, operating factor, and distributed geospatial information inheritance. Remarkably, these effects are better than traditional methods, single point of search and heuristic guidance, which characteristic is neighborhood local information inheritance and approximation[1]. But the population in GA and AIA, has convergence property of single pattern, and can't ensure many patterns coexistence for a long term easily. For using GA and AIA to find more optimal solutions, various niche algorithms are proposed, such as niche genetic algorithm[2], niche genetic

* This work is supported by the National Natural Science Foundation of China (61073101).

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 57–65.
springerlink.com

algorithm based on sharing mechanism of fitness[3] and niche clonal selection algorithm[4]. These improved algorithms can relieve pattern convergence degree, and find most extreme points. But distance value L of the niche in literature[2] and sharing radius in literature[3] are difficult to determine, which are more sensitive for the concrete optimization objective. Although N niches are formed in literature[4], it has strengthened diversity of antibodies formally and disorganized the balance of initial population in feasible region, which cause convergence of single pattern. That's because d antibodies with low fitness are instead by new produced antibodies in the formation process of next antibodies. Compared to GA and AIA, particle swarm algorithm has higher efficient information sharing mechanism and better instructiveness[5-7]. So, theoretically, particle swarm algorithm has faster convergence. But it makes ants excessively centralized when optimizing, and maybe all ants move to global optimal point in the end, and the algorithm can't be used to multi-modal functions optimization. The niche particle swarm algorithm in this paper realizes the multi-modal functions optimization better and be easy to programming realization. The results of simulation show that this algorithm is better others in literature[2-4].

2 Basic Principle of PSO Algorithm

PSO simulates the predatory behavior of birds. Imagine a scene: a group of birds are searching for food randomly. There is only one piece of food, and all birds don't know the position. But they know its distance from themselves. What is the optimal strategy? The most simple and effective one is to search surrounding area, where the distance from birds is shortest. From this model, PSO gets enlightenment and uses it to optimization problem. Every solution of PSO is one bird in search space, called "particle". Each particle has a fitness value, determined by the function optimized, and a velocity value, determining the birds' direction and distance. And then particles are searching in the solution space, following the optimal particle.

PSO is initialized to a group of random particles, P (random solution). Find the optimal solution by iteration. In iteration, particle, P, updates itself by tracking two "extreme values". The first optimal solution is found by particle itself and called personal extreme, pbest. The other one is found by the whole populations and called global extreme, gbest. After that, particle update its' velocity and new position by the follow formulas:

$$v_{ij}^{k+1} = v_{ij}^{k} + c_1 * r_1 * (pbest_{ij}^{k} - x_{ij}^{k}) + c_2 * r_2 * (gbest_{j}^{k} - x_{ij}^{k}) \quad (1)$$

$$x_{ij}^{k+1} = x_{ij}^{k} + v_{ij}^{k+1} \quad (2)$$

Which, subscript, i, represents the i particle; subscript, j, represents the j dimension of velocity (or position); superscript, k, represents the iteration number. For example, v_{ij}^{k} and x_{ij}^{k} are respectively the velocity and position of the i particle (Pi) on the j dimension in the k iteration. And both of them are limited to certain range. c1 and c2

are learning factors, in range of [0,4]. r1 and r2 are random number between 0 and 1. $pbest_{ij}^{k}$ is coordinate of personal extreme on the j dimension for particle Pi. $gbest_{j}^{k}$ is coordinate of global extreme on the j dimension for populations.

3 Real Coding of Niche Particle Swarm Algorithm

It can simulate a scene to multi-modal functions optimization problem: a group of birds are searching for food randomly. There are some pieces of food, having different sizes. All birds don't know the position, but they know they can find the position of recent food from themselves. What is the optimal strategy for each bird to find the recent food quickly and have a tendency for the biggest block? Analyzing formula (1), the updating of particle's velocity is composed by three parts: the first one is the anterior velocity; second "cognitive part", expressing thinking of particle; third "social sector", expressing the information sharing of particles. According to the above and mechanism of particle swarm algorithm, it takes real coding, which is suitable for NPSA to solve multi-modal functions optimization. Each particle updates their velocity by the follow formula:

$$v_{ij}^{k+1} = w_{\mathrm{k}} v_{ij}^{k} + c * r * (pbest_{ij}^{k} - x_{ij}^{k}) \tag{3}$$

The formula (3) comes from formula (1). Based on c_2 equaling to zero in formula (1), it makes particle's anterior velocity multiply variable w_{k} (w_{k} is a weight value changing from big to small); particle's new position is calculated by formula (2). This algorithm has no social sector, weakening information sharing and enhancing local search ability. Then, each particle has a trend toward the recent peak (extreme), forming niche.

NPSA enhances the local searching ability of particles and weakens information sharing between them, so particles don't concentrate on an extreme point (unless there is a peak). Conversely, if weakening information sharing, whether particles just find local extreme points? By simulation experiment analysis, this situation, no finding the global optimal points, won't occur in NPSA.

The basic thinking of NPSA is: produce N particles' initial velocities and positions randomly to compose of an initial population and calculate fitness value (function value) of every particle according to objective function. And then calculate particle's personal optimal solution, pbest and global optimal solution, gbest. Update the particles' velocities and positions according to formula (3) and (2). Because enhancing local search ability, each particle has a trend toward the recent extreme, forming N niches.

There is no calculation about the three part when updating velocity in NPSA, so it is simplified. The relation between time complexity and population size is linear in NPSA, but square relation in literature[4] and n-th power in literature[4]. So it won't affect computational complexity and velocity, when increasing particles, and it will in literature[2-4].

The NPSA is described as follows:

For each particle

Coding for particles' initial positions (limited to feasible region) and velocities (limited to [Vmin,Vmax])

END

Do % cyclic iteration

For each particle

Calculate the particle P_i's personal fitness value according to objective function. If it is better than the current one, $pbest_i$, and then make the $pbest_i$ be equal to it.

End

Find the maximum from all personal fitness values, and make $gbest_k$ be equal to it.

For each particle

Update P_i's velocity according to formula (3), limiting to range of [Vmin,Vmax].

Update P_i's position, limiting to feasible region

End

While (the maximum iteration number designed)

4 Astringency Analysis

From formula (3) and (2), v_{ij}^{k} and x_{ij}^{k} are multi-dimensional variables, and each dimension is independent. So the algorithm can be simplified to one-dimensional. Assumed that, there is no change to the optimal positions found by particle itself and the population, they are denoted by p and g respectively. The range of w_k is between 0 and 1, and it don't change when calculating. c is a constant. Formula (3) and (2) are simplified as follows:

$$v(k+1) = wv(k) + c(p - x(k)) \tag{4}$$

$$x(k+1) = x(k) + v(k+1) \tag{5}$$

By formula (4) and (5), we get:

$$v(k+2) = wv(k+1) + c(p - x(k+1)) \tag{6}$$

$$x(k+2) = x(k+1) + v(k+2) \tag{7}$$

Take formula (5) and (6) to formula (7), we get:

$$x(k+2) + (c - w - 1)x(k+1) + wx(k) = cp \tag{8}$$

It is second-order constant coefficient inhomogeneous differential equation. There are many methods to solve it, and the typical one is characteristic equation[5]. First, take

the formula (8): $\lambda^2 + (c - w - 1)\lambda + w = 0$. It is divided into three situations according to solutions of equation of two degree.

(1) when $\Delta = (c-w-1)^2 - 4w = 0$, $\lambda = \lambda_1 = \lambda_2 = -(c-w-1)/2$, and then $x(k) = (A + Bk)\lambda^k$, which, A and B are undetermined coefficients. By v(1) and x(1), we get:

$$A = x(1), \quad B = \frac{(1-c)x(1) + wv(1) + cp}{\lambda} - x(1) \cdot$$

(2) when $\Delta = (c-w-1)^2 - 4w > 0$, $\lambda_{1,2} = \frac{w - c + 1 \pm \sqrt{\Delta}}{2}$, and then $x(k) = A + B\lambda_1^k + C\lambda_2^k$, which, A, B and C are undetermined coefficients. Make a=x(1)-A, b=(1-c)x(1)+wv(1)+cp-A, and then we get:

$$A = p, \quad B = \frac{\lambda_2 a - b}{\lambda_2 - \lambda_1}, \quad C = \frac{b - \lambda_1 a}{\lambda_2 - \lambda_1}$$

(3) when $\Delta = (c-w-1)^2 - 4w < 0$, $\lambda_{1,2} = \frac{w - c + 1 \pm i\sqrt{-\Delta}}{2}$, and then $x(k) = A + B\lambda_1^k + C\lambda_2^k$, which, A, B and C are undetermined coefficients. We get:

$$A = p, \quad B = \frac{\lambda_2 a - b}{\lambda_2 - \lambda_1}, \quad C = \frac{b - \lambda_1 a}{\lambda_2 - \lambda_1}$$

If $k \to \infty$, x(k) has limit, and is trend to finite value, representing iterative convergence. From this, its conditionsare $\|\lambda_1\| < 1$ and $\|\lambda_2\| < 1$, if satisfying the above three situations to keep x(k) convergent.

Take the follow conclusions:

When $\Delta = 0$, convergence region is: parabola $w^2 + c^2 - 2wc - 2w - 2c + 1 = 0$ 且 $w \in [0,1)$;

When $\Delta > 0$, convergence region is: the area composed by $w^2 + c^2 - 2wc - 2w + 1 > 0, c > 0$ and $2w - c + 2 > 0$.

When $\Delta < 0$, convergence region is: the area composed by $w^2 + c^2 - 2wc - 2w - 2c + 1 < 0$ and $w \in [0,1)$.

Combining the above three situations, convergence region is: the area composed by w<1, c>0 and $2w - c + 2 > 0$. It shows as shaded parts in Fig1.

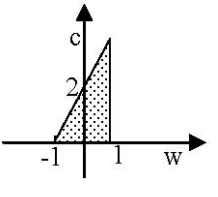


Fig. 1. Convergence region of algorithm

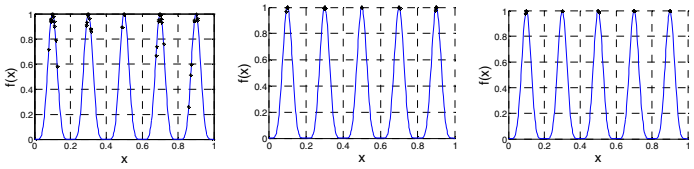
5 Simulation Research

In simulation, take four typical multi-modal functions showed in table1. Which, F1 is a function with equal peaks, single variable, and multi maximum, which is used to the test whether optimization algorithm can find all optimal solutions. F2 is a function with changing peaks, single variable and multi maximum (eight maximum points, and one of them is global), which is used to test whether it can find global optimal solution and all local optimal solutions. F3 has thirteen-six peaks, including four global peaks, and the global maximum is (0, 0). F4 also has many peaks with larger difference, and conventional multi-modal function optimization algorithm can't get all local optimal solutions. If finding all global and local optimal solutions in F4 quickly and accurately, algorithm is stable, reliable and good performance of multi-modal functions optimization.

Table 1. Testing functions

function name	test functions	value range of variable	function feature
F1	$\mathrm{Sin}^6(5\pi x)$	$x \in [0,1]$	equal peaks, single variable and multi maximum
F2	$x + 10 * \sin(5x) + 7 * \cos(4x)$	$x \in [0,9]$	Changing peaks, single variable and multi maximum
F3	$\cos(2\pi x) * \cos(2\pi y) * e(-((x^2 + y^2)/10))$	x、y ∈[-1,1]	Changing peaks, multi variable and multi maximum
F4	$x*\sin(4\pi x) - y*\sin(4*\pi y + \pi + 1)$	x、y ∈[-1,2]	Changing peaks, multi variable and multi maximum

First, seek the optimal solutions of function F1 by literature[3], literature[4] (NCSA) and NPSA respectively. literature[3] uses binary coding, and the parameters are: the number of chromosomes 50, the length of chromosomes 22, evolutionary generations 30, α 1, σ_{share} 4, crossover probability of single point 0.6, mutation probability 0.01. NGA uses binary coding, and the parameters are: the number of antibodies is 50, the length of antibodies 22, the clone number of each antibody 10, evolutionary generations 30, high-frequency of mutation probability 0.1. The parameters in NPSA are: the number of particles is 50, iteration times 30, c 2 and inertia weight $w_i = w\max - \frac{w\max - w\min}{i_{\max}} * i$ (i is the current generation, and imax is the whole iteration times, and wmax is equal to 1, and wmin 0). Realize the algorithm by MATLAB language on a same computer, and results show in Fig2.

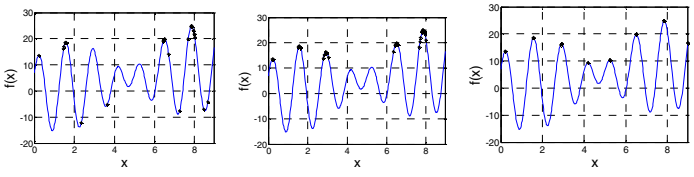


(a) chromosomes (b) antibodies (c) particles

Fig. 2. The results of function F1 with three different algorithms

The optimization results of function F1, using literature[3], literature[4] (NCSA) and NPSA, show in Fig2(a), Fig2(b) and Fig2(c) respectively. By adjusting parameters in literature[3] for many times, it took 2.1320 seconds to get result, showing in Fig2(a); it took 2.9763 seconds for NCSA to get, showing in Fig2(b); it took 1.0160 seconds for NPSA to get, showing in Fig2(c). To explain the search ability of NPSA, made many simulations and found all optimal solutions for each time. But there always are some chromosomes away from extreme in literature[3]. Although all individuals move to extreme in NCSA and NPSA, it takes longer time to run for NCSA than NPSA. If expanding the population sizes, NCSA's running time increases exponentially, and NPSA's increases linearly. So NPSA's optimization result is better than literature[3] and [4].

Take the optimization simulation to function F2 by three algorithms respectively, with the same parameters in function F1. The results are shown in Fig3.



(a) chromosomes (b) antibodies (c) particles

Fig. 3. The results of function F2 with three different algorithms

The optimization results of function F2, using literature[3], literature[4] (NCSA) and NPSA, show in Fig3(a), Fig3(b) and Fig3(c) respectively. And running times are 2.1450 seconds, 3.0174 seconds and 1.0820 seconds respectively. There is no change for running time between function F2's optimization and function F1's. But obviously, it is different to search all extremes, and NPSA is better than literature[3] and [4]. To avoid unstable effects from initial population produced randomly and show better the search ability of NPSA, it is necessary to keep the parameters invariant. Statistical results, operated 10 times independently by three different algorithms, are showed in Table2. It shows that NPSA's parameters are insensitive to different optimization objectives, and NPSA find not only global optimal solution of multi-modal functions, but also all local optimal solutions.

Table 2. The statistical results of function F2 optimized by three different algorithms

times	niche genetic algorithm based on fitness sharing mechanism		niche clonal selection algorithm		Niche Particle Swarm Algorithm	
	maximum point	number of maximum points searched	maximum point	number of maximum points searched	maximum point	number of maximum points searched
1	find	4	find	5	find	8
2	find	3	find	4	find	8
3	find	4	find	5	find	8
4	find	3	find	3	find	8
5	find	3	find	6	find	8
6	find	4	find	4	find	8
7	find	4	find	7	find	8
8	find	3	find	5	find	8
9	find	5	find	5	find	8
10	find	4	find	4	find	8

Optimize two-dimensional function F3 and F4 by NPSA, changing the number of ants to 150 and other parameters don't change. The results show in Fig4 and Fig5.

From optimization results to two-dimensional function F3 and F4, it shows that NPSA can find only all global optimal solutions, but also all local ones. This verifies the algorithm's correctness adequately. To avoid the effects from initial population produced randomly, make more simulation experiments, and it finds all extreme points each time. Particles final distributions are all up to ideal effect of algorithm.

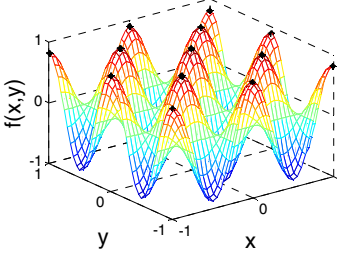


Fig. 4. The result of F3 by NPSA

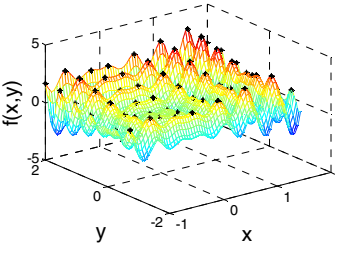


Fig. 5. The result of F4 by NPSA

6 Conclusions

Comparing the simulation results in literature[3-4], it gets conclusions: the relation between time complexity of NPSA and population size is linear; its parameters are

easy to select; it has good convergence; NPSA parameters are insensitive to different optimization objectives; NPSA is suitable for one-dimensional and multi-dimensional function optimization and multi-modal functions optimization problem with searching for some optimal solutions or local optimal solutions.

References

1. Li, M.-Q., Kou, J.-S.: Coordinate multi-population genetic algorithms for multi-modal function optimization. Acta Automatica Sinica 28(4), 497–504 (2002)
2. Yuan, L., Li, M., Xiao, Q., et al.: Multiple hump function optimization based on niche genetic algorithm. Journal of Nanchang Institute of Aeronautical Technology 20(4), 1–4 (2005)
3. Goldberg, D.E., Rechardson, J.: Genetic Algorithms with Sharing for Multimodal Optimization. In: Proceedings of the Second International Conference on Genetic Algorithms, pp. 69–76. Lawrence Erlbaum Associates (1987)
4. Wang, X.-L., Li, H.-J.: Niche Clonal Selection Algorithm for Multi-modal Function Optimization. Journal of Gansu Sciences 18(3), 64–68 (2006)
5. Gao, S., Tang, K., Jiang, X., et al.: Convergence Analysis of Particle Swarm Optimization Algorithm. Science Technology and Engineering 6(12), 1625–1628 (2006)
6. Jian, L.: Differential genetic particle swarm optimization for continuous function optimization. In: 3rd International Symposium on Intelligent Information Technology Application, vol. 3, pp. 524–527 (2009)
7. Wang, Y.-J.: Improving particle swarm optimization performance with local search for high-dimensional function optimization. Optimization Methods and Software 25(5), 781–795 (2010)

The Design and Analysis of Logistics Information System Based on Data Mining

Xiaoping Wu[1], Juming Liu[1], Hongbin Jin[2], and Lejiang Guo[2]

[1] Training Department
Air Force Radar Academy
Wuhan, China
[2] Department of Early Warning Surveillance Intelligence
Air Force Radar Academy
Wuhan, China
{radar_boss,lezzabay}@163.com

Abstract. With the development of modern society, it makes the growing demand for information, data mining is the process of extracting knowledge. Data mining is frequently obstructed by privacy concerns. Overview the past and present of data mining, focusing on analysis of the current direction of the branch of data mining, and data mining technology is prospected. Data preprocessing, Data classification, Association rule mining and Prediction are the foundation and linkage of the whole data mining process life cycle. the main application of mining theories and technologies in spatial data mining, the main methods and examples of visualization spatial data mining, this paper also presents a reference model of mining-based transport information system. Logistics information services will be more and more important, the quality and efficiency of logistics services will also continue to improve.

Keywords: Data mining, logistics data, the target data, data warehouse.

1 Introduction

Data mining, as the name suggests, is mined useful information from mass data, it is a non-trivial process that is from the data of abundant, incomplete, be noise, ambiguous, the random practical application, to find the information and knowledge which is implicit, regularity, people not known in advance, but potentially useful and ultimately comprehensible. Data mining as a relatively new, comprehensive subject, that includes database technology, machine learning, statistics, pattern recognition, information retrieval, neural network, knowledge-based system, artificial intelligence, high performance computing and data in visualization, primarily study based on computer science, statistics and other subjects. Many methods for data mining research are derived from two branches, one is machine learning, and the other is statistics, in particular, diverse computing statistics. Machine learning related with computer science and artificial intelligence, is used to detect relationship and rule among the data, these relationships and rules can be expressed in universal law.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 67–74.
springerlink.com

Modern logistics system is a large and complex system, especially the whole logistics, including transportation, warehousing, distribution, transport, packaging and logistics re-processing, and many other links, each link has a huge information flow. In particular, application of modern information logistics network system expands the scale of the original database constantly, results in a huge data flow, so that the enterprise is difficult to collect accurately and efficiently and handle timely on these data, thereby helping policy makers to make decision rapidly, accurately, to achieve control of logistics processes, reduce logistics costs throughout the process. Data mining technology can help companies, in management of the logistics information system, collect analyze customer, market, sales and all kinds of inner information of the whole enterprise timely and accurately, analyze the customer behavior and market trends effectively, understand the different customer's interest, thus targeted product and services can be provided to customers, which greatly enhance the variety of customers' satisfaction on business and product.

2 Data Mining

Data mining is the process that from the data of abundant, incomplete, noise, ambiguous, the random practical application, draw the information and knowledge which is implicit, not known in advance, but potentially useful. Data mining is a new information processing technology, its main feature is to extract, transform, analyze and other modeling handle a lot of business data in the enterprise database (storehouse), to extract the key data assisting business decision-making. Therefore, data mining can be described as: It is an advanced and effective method that in accordance with enterprise established business objective, explore and analyze a large number of enterprise data, reveal the hidden, unknown or verify the known law, and modeling further. The object of data mining can be database, file system or any other organized data collection.

2.1 The Concept of Data Mining (DM)

Data mining, in artificial intelligence, also customarily known as knowledge discovery in databases (KDD), someone take the data mining as an essential step in the process of knowledge discovery in the database. Knowledge discovery process composes the following three phases

1) The data preparation

Data preparation can be divided into 2 sub-steps: data selection, data pre-processing. The purpose of the data selection is to determine have found the operation object of the task that is the target data, which is based on the user's need to extract a set of data from the original database. Generally, data preprocessing includes elimination of noise, derived calculation missing values data, elimination duplicate records, accomplish data type conversion (such as transforming the continuous data into discrete data, so it's easy to sum the symbolic; or transforming the discrete data into continuous data, then it will be easy to the neural network calculation) and the reduction of data dimension (that is, to find out really useful data mining features from the initial feature to reduce the number of the considering variables).

2) Data mining as the core of data mining technology, mainly is, on the basis of determining the mining task, to select the appropriate data mining techniques and algorithm, and based on this, repeatedly iterative search, extract hidden, novel model from the data collection. Such as neural networks, decision trees, cluster analysis technology, association discovery and sequence discovery techniques etc and the ID3 algorithm, BP algorithm and so on.

3) The expression and interpretation of results. Data mining can interact with the user or knowledge base. Not all of the information discovery tasks are treated as data mining. For example, the use of database management system to find individual records, or through the Internet search engine to find a specific web page, it is information retrieval. Although these tasks are important, may involve the use of complex algorithms and data structure, but they are mainly rely on conventional computer science and technology and the distinct features of data to create index structure, thus to effectively organize and retrieve information. However, data mining has also been used to enhance the capacity of information retrieval system.

2.2 The method and mission of DM

1) Method of DM

Data mining methods mainly are statistical analysis, decision trees, genetic algorithms, inductive learning methods, imitation biotechnology、Bayesian belief network, neural networks, fuzzy sets, rough sets and so on. In practical application process, generally combined with the practical need, select a number of methods as public to get the best results.

2) Mission of DM

In general, data mining tasks include: correlation analysis, cluster analysis, classification, prediction, time series models, deviation analysis etc.

2.3 System Architecture of DM

In the commonly used data mining systems, in general, it is divided into three layers, as shown below. Among them, the first layer is the data source, including database, data warehouses. The second is sapper, using various data mining methods provide by data mining system analyze and extract data in the database, to meet the needs of users. The third layer is the user layer, using a variety of ways to reflect obtained information and the found knowledge to the user.

3 The Application of Data Mining in the Logistics Information System

With the establish of integrated logistics management information system and the application of network technology, EDI, artificial intelligence, bar code and POS and other advanced technology, the commercialization of logistics information, logistics information collection database and the code, electronic and computerized of logistics information processing, organically combination of the excavated the rules and all aspects of logistics management, can greatly improve the competitiveness of

enterprises. Logistics decision-making system is a new type operate decision-making system combined data mining and artificial intelligence, mainly through the artificial intelligence to collect large amounts of information from all aspects including the raw material purchase, process manufacture, distribution and dispatching to goods sold, and using data warehouses and data mining technology analyzes and handles, and accordingly this to determine the appropriate business strategy.

3.1 The Build of Data Warehouse

Data warehouse as the basis of data mining, is different from the traditional online transaction processing system, it has the characteristic of the facing the theme, integrated, cannot update and changes with time. Various online transaction processing system as the original data source of data warehouse, provide business large amounts of data and reports collected in the daily activities, including purchase order, inventory orders, accounts payable, trade terms, and customer information as documents, while there are a lot of external information and other data. Data warehouse through ETL process (extract, transform and load) processes the interface files, and organizes according to different subject domain, stores and manages the customer data. Through the data warehouse interface, analyze on-line and data mining the data of warehouse data. After the establishment and complete enterprise-level information data warehouse in Fig.1, can do data mining work based on the data warehouse platform.

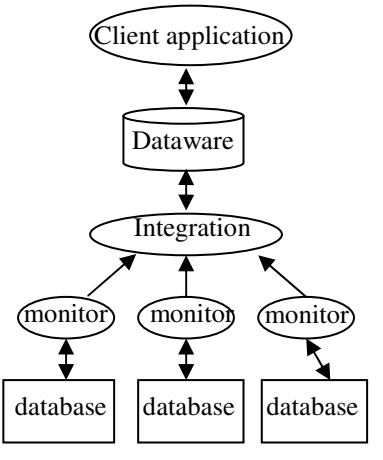


Fig. 1. The architecture of dataware

3.2 The Architecture of Llogistics Information System Based on DM

In Fig.2, the architecture of logistics information system based on data mining is main component from the following:

1) Logistics information collection and processing: Record he various types of information in logistics management activities, collect, process, transport a variety of

information in logistics activities, and store in the data warehouse in a unified format. Architecture of data Warehouse is the construction of infrastructure, only the solid data warehouse infrastructure can support flexible and various application of data warehouse. To the enterprise, the construction of data warehouse is a systematic project, is an ongoing establishment, development, improvement process, which requires enterprises based on the overall reference frame of data warehouse system, gradually build a complete, robust data warehouse system.

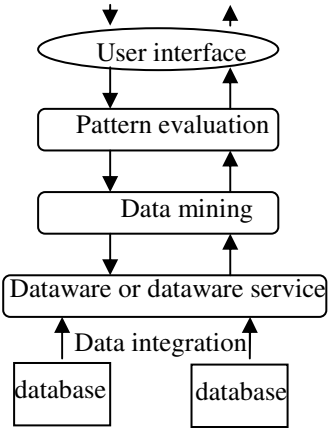


Fig. 2. Data mining mode

2) Logistics information management system of data mining: Gather the collected data into a data warehouse, then according to the data from the data mining, provides management decision-makers the latest and most valuable information or knowledge, to help them quickly and correctly make a decision.

3)DM: According to characteristics of the question suggested by policy makers, determine the excavated task or purpose, streamline and preprocess the relevant data in the warehouse data, then excavate the new and effective knowledge from streamlined data, and provide the data to data mining logistics information management system, which will provide effective knowledge to decision makers.

4) Knowledge base: Include the composition structure of data warehouse based on sector, membership function, and so on.

5) Developer and expert interface: Developers and experts define and maintain the knowledge in the knowledge base through this interface.

6) Data warehouse: The main storage of various data related to logistics management

This system joins up links like the raw materials, purchase, transportation, distribution, storage, packaging, order processing, inventory control , to completely realize a integrity and efficient circulation process that the goods from the production line to the customer, to provide high quality service.

3.3 Features of the Logistics Information System Based on Data Mining

Compare with traditional logistics management information system, the logistics information system based on data mining has the following characteristics.

(1) Traditional logistics management system, generally is divided into many functional modules according to the function, information sharing scope and logistics information management, have a far cry from particularly, information high transparency and rapid response of the requirements of supply chain management. Moreover the data mining technology based logistics management information system, uses data warehouse technology to organize and manage data, it can describe various enterprise data of each object related completely and consistently, so that it can bind each department of enterprise, upstream manufacturers and other downstream retail, to achieve maximally the information sharing.

(2) The general database in order to improve efficiency of the system is often as little as possible to retain historical information. While the data warehouse has an important feature, it is that the general has a long history of data storage. The purpose of long history of data storage is to make a long-term trend analysis, forecast changes in inventories in period time of future, achieve forward-looking allocate and transfer, enhance the ability to adapt to unforeseen factors, provide strong data support for policy-makers over the long term.

(3) In the traditional system, model base and knowledge base are often independently design and achieve lack of internal unity. Knowledge model from expert, and is difficulty to update. System based on data mining, separates the knowledge discoverer and knowledge users of these two types of the role, policy makers do not have a deep understanding of decision-making system and does not require in-depth study of data warehousing, data mining and other related knowledge. The professionals can access the relative server at any time by the TCP/IP protocol, to manage the system, maintain, expand knowledge base and etc.

(4) Uncertain DM, Traditional data mining processing position has been precisely given object, but in practical application, due to the limitation of measuring instrument will result in inaccurate measurements, data uncertainty is inevitable. The uncertainty of data can be divided into the existing uncertainty and the value uncertainty of two categories, existing uncertainty refers to the uncertainty object or whether tuple exists, such as a certain tuple of relational database associate with a probability groups associated to signify the credibility of this tuple, the value uncertainty refers that the existence of a tuple is certain, but its value is uncertain. Research of the uncertain data mining has become a hotspot, in the cluster analysis, association rules, space excavation and other aspects have break, classic K-means algorithm is extended to the UK-means algorithm, Apriori algorithm is extended to UApriori algorithm.

(5) Rules optimization not only has been a difficult problem in data mining but also a core issue. There are many extraction rules, which used frequently are the black box method, adaptive extraction criteria, BP network, input and output binary criteria, full guidelines for some guidelines tray extraction algorithm.

(6) The operation efficiency of algorithm. The main efficiency evaluation of the algorithm is time complexity and space complexity. ID3 and some other algorithms can not be achieved on the classification of continuous attributes, and is mainly used

for smaller dataset, which mainly restrictions limit lies in the memory; that can deal with continuous attribute, but due to all the attribute values should be sequenced in the partitioning, the time is wasted. It also save the train set in-memory.

(7) Scalability of algorithm and decision tree is mainly for capacity of the training set size. As the algorithm is mainly set the training collection in memory, processing small data sets has more efficient, but processing of large data sets has certain issues about effectiveness and scalability; the CART algorithm made improvements on scalability, the data set is stored mainly in external memory, but the corresponding problem, the internal memory and external memory data scheduling, is caused.

3.4 The Development Tendency of DM

Data mining tasks and the diversity of data mining made many challenging research problems, in the future there will be a larger climax which focus on the following research aspects: Study devotes to discovery a formalized and standardized language in data mining; research for data mining process visualized method to making the process of knowledge discovery can be understood by user, and easily interacted with human; study In Network and Distributed Data Mining environment, especially in the Internet to build data mining server with database server to establish data mining; strengthen the various unstructured data mining, such as text data, graphics and video data, multimedia data; explore scalable and interactive data mining method, and comprehensively improve the overall efficiency of mining process, especially the efficiency of the data set in large scale data mining; to expand the scope of data mining applications, such as financial analysis, biomedical research, criminal investigation etc.; develop method which adapt multiple data types and allow noise of the mining to solve the heterogeneous data set of data mining; dynamic data and knowledge of data mining.

4 Conclusions

Data mining involves a multi-disciplinary technology such as: Database technology, statistics, machine learning, high performance computing, pattern recognition, neural networks, data visualization, information retrieval and spatial data analysis and so on. Therefore, data mining has a very promising research area, with the continuous development of data mining technology; it will be widely and deeply applied in the human society's all fields. At present, although the data mining technology was applied to a certain extent, and have achieved significant results, but there are still many unsolved problems, such as data preprocessing, mining algorithms, pattern recognition and interpretation, visualization, and so on. For the business process, the most important data mining problem is how to combine spatial and temporal characteristics of business data to mine out the knowledge representation, that is, the problem of temporal and spatial of knowledge representation and interpretation mechanisms. As people deeply research data mining, data mining technology will be in wider areas of application, and achieve more significant results.

With the increase in the level of logistics information, the logistics strategy has changed from internal integration to external integration. Supply chain management

has become a very important competitive strategy component. Data mining methods has effectively promoted the company's business process restructuring, to improve and enhance customer service, strengthen enterprise's asset liability management, to promote market optimization, to accelerate cash flow, realize enterprise scale optimization, effectively improve the competitiveness of enterprises.

References

1. Liu, X., Du, J., Li, W., Zuo, M., Han, Z.: 'Data warehousing for data mining based on OLAP technology. In: China-Ireland International Conference on Information and Communications Technologies, CIICT 2008, pp. 176–179 (2008)
2. Zhang, Y., Li, W., Chen, Y.: The study of multidimensional-data flow of fishbone applied for data mining. In: 7th ACIS International Conference on Software Engineering Research, Management and Applications, pp. 86–91 (2009)
3. Varela, M.L.R., Aparício, J.N., Silva, S.C.: A web-based application for manufacturing scheduling. In: Proceedings of the IASTED International Conference on Intelligent Systems and Control, pp. 400–405 (2003)
4. Li, L., Yang, B., Zhou, F.: A framework for object-oriented data mining. In: 5th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2008, pp. 60–64 (2008)
5. Tanel, T., Hele-Mai, H., Vello, K., Kääramees, Marko: A rule-based approach to web-based application development. In: 7th International Baltic Conference on Databases and Information Systems, pp. 202–208 (2006)
6. Jin, H., Liu, H.: Research on visualization techniques in data mining. In: 2009 International Conference on Computational Intelligence and Software Engineering, CISE 2009, pp. 111–116 (2009)
7. Mejia, B.J.F., Paolo, F., Maurizio, M.: A web-based application to verify open mobile alliance device management specifications. In: 1st International Conference on Advances in System Testing and Validation Lifecycle, VALID, pp. 13–18 (2009)
8. Chen, C.-H., Tseng, V.S., Hong, T.-P.: Cluster-based evaluation in fuzzy-genetic data mining. IEEE Transactions on Fuzzy Systems 16(1), 249–262 (2008)
9. Patnaik, D.: Accelerator-oriented algorithm transformation for temporal data mining. In: NPC 2009 - 6th International Conference on Network and Parallel Computing, pp. 93–100 (2009)
10. Afshin, S., Mohammad Reza, K.: Designing an extensible query language for data mining. In: 2008 International Conference on Data Mining, DMIN 2008, pp. 348–354 (2008)

An Overview of Dynamic Balanced Scorecard

Tiezhu Zhang

Institute of Management Science and Engineering
Zhejiang Gongshang University
Hangzhou, China
ztz0647@sina.com

Abstract. The rapid changing competitive environment of modern enterprise sets a higher request to enterprise's performance management. The studies of integrating the Balanced Scorecard and system dynamics to carry though performance management get more and more attention. Dynamic Balanced Scorecard compensates for the weaknesses of the Balanced Scorecard's simplified causal relationship and dynamics. It is helpful to the implement of the organization's strategic performance management. On the basis of the retrospect of the studies on the Dynamic Balanced Scorecard, proposes a set of Specific implementation procedures of it; gives the expectation of the development trend and some suggestions on the subsequent research. It is of great practical significance to promote the implementation of the Balanced Scorecard and improve the organization's performance management level.

Keywords: performance management, system dynamics, balanced scorecard, overview.

1 Introduction

Competition between enterprises has been identifying increasingly with the economic globalization, which sets a higher request to enterprise's performance management. Performance management has become a strategic competitive advantage for enterprises to cultivate. To meet this need, more and more enterprises want to set up an overall performance management system to ensure the strategy's implementation and achievement. Balanced Scorecard is such a tool which provides an important strategic performance management system.

Balanced Scorecard has been prevalent in the world, while its theoretical limits are questioned. Firstly, it can not sort out the complexity of the dynamic relations within the enterprise. Secondly, it can not simulate the relations between strategic targets and performance indicators dynamically, and so the manager could concern about the short-term performance and ignore longer-term performance. Thirdly, Balanced Scorecard does not take the time delay within the causal relationship into account, so the problems encountered are ready to be overcorrected. And all these shortcomings of Balanced Scorecard are just the strengths of system dynamics.

In recent years, more and more researchers put their attention into the combination of system dynamics and the Balanced Scorecard. After combining Balanced Scorecard

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 75–82.
springerlink.com

and system dynamics, Dynamic Balanced Scorecard makes the whole performance management system into a loop process which conforms to the actual situation of enterprises. In this paper, based on the overview of literature about the combination of system dynamics and the Balanced Scorecard, a set of implement steps are proposed, which could promote the implementation of the Balanced Scorecard more effectively and improve the enterprise's strategic performance management.

2 Disadvantage of the Balanced Scorecard

As an important strategic performance management tool, balanced scorecard introduces three non-financial aspects of evaluation indicators, which compensate for the traditional performance evaluation's shortage which only emphasize on financial indicators. From a strategic viewpoint, it can help the managers concentrate on the most important elements which lead to success, and ensure the implement of enterprise's strategy is system thinking. The core idea of the Balanced Scorecard is balancing a series of indicators such as short-term and long-term indicators, financial and non-financial indicators, leading and lagging indicators, inner and outer indicators. With strategy map, balanced scorecard can describe the causality of the indicators in different aspects. After the application of strategy map and Balanced Scorecard, the focal point of performance management turns from short-term targets into strategic objects, and from reflection on the results to real-time analysis of the problem. However, there are a number of deficiencies in theory within the balanced scorecard.

2.1 Ignoring the Time Delay

The existence of time delay makes the results of the actions can only be produced with a gradual manner. If ignored or not fully understood, it is easy to lead the activities taken by managers to exceeding the proper limits in righting a wrong. Actually, there exist time delay between the activities and its feedback. When putting their attention on improving the leading indicators in Balanced Scorecard, the managers should take the time delay into account.

Balanced Scorecard takes emphasis on strategy formulation and performance evaluation, but it is unable to describe the strategy's impletion dynamically. Ignoring the time delay could also make the formulation of strategy lose the focus. As for the causal relationship's time delay between causes and consequences, Balanced Scorecard can not provide an appropriate solution.

2.2 Dynamic Insufficiency

Balanced scorecard takes emphasis on the static evaluation on the performance management which is lack of dynamic. It is unable to describe the time-based dynamic performance which engender when enterprises implementing strategies. Balanced scorecard introduces non-financial and leading indicators, but it still estimates the enterprise's current performance and can not reveal the enterprise's future performance. Therefore, this static evaluation can only give decision-makers a current intuitive status of the enterprise; can not forecast the future development of the enterprise.

2.3 One-Way Causality

Balanced Scorecard can not reflect the causal feedback within the system effectively. The one-way causality of Balanced Scorecard is unable to reflect the non-linear causality in the complex business activities of enterprises, which is only one aspect of dynamic complexity within the enterprise. One result may be caused by many reasons, and balanced scorecard is ready to lose sight of the main or real causation. for example, the causality" service quality's improvement can attract more customers" can not show such important information on reverse: more customers will also cut down the service time offered and furthermore will reduce the service quality.

When turn the strategy into activities using balanced scorecard's simple liner causality, it usually lead to interest conflict in different departments. Non-linear causality refers to that the functional relationship between two variables is not a linear relationship, but a curvilinear relation. In most cases, the relationship between factors within the system is non-linear, and to handle this, balanced scorecard is not efficiency enough.

2.4 Summery

When turning enterprise's strategy into departments' targets or even each employee's goals using Balanced Scorecard, the key is to clarify the causal relationship between these indicators. In actual operation, Balanced Scorecard only provides a framework and with which it is hard to establish a causal relationship between indicators. Secondly, balanced scorecard is unable to handle the non-linear causality and time delay problems. These shortcomings are just the strengths of system dynamics. Actually, the founder of Balanced Scorecard Kaplan and Norton are also cognizant of these shortcomings. They also suggest using of system dynamics to make up for these deficiencies, especially in dynamic and complex situations. Combining balanced scorecard and system dynamics can assist the establishment of the causality; promote the implementation of the Balanced Scorecard effectively.

3 Related Researches on the Integration of Balanced Scorecard and System Dynamic

According to the investigation of Lewy and Dumee from Holland organizations, 70 percent of which failed to achieve their strategy goals. In view of theoretic shortages of balanced scorecard, many problems are encountered in the practical implementation process. Scholars began to improve balanced scorecard's shortcomings, the majority of which are integration of balanced scorecard and system dynamics. The researches are classified into three phases based on the overview.

3.1 Phase I

Some scholars initially pointed out that Balanced Scorecard was lake of dynamics, but did not propose corresponding solutions. Representative studies are introduced as follows:

Atkinson, Waterhouse, Wells fingered out that strategy should be changed according to the changing environment, and balanced scorecard was lake of elasticity and need to be improved[1].

Otley pointed out that the financial indicators are related to the non-financial ones, though there did not exist a systematic tool which could indicate the causal relationships of the indicators in different aspects[2].

Mooraj pointed out that balanced scorecard was one-way thinking and was unable to measure the two-way operation of enterprises[3].

Norrtklit fingered out that the relationship between indicators in all levels of balanced scorecard were beyond one-way causal relationship, or even a two-way complex relationships, but this complex causal relationship can not be manifested in the Balanced Scorecard[4].

Todd and Palmer developed a full system dynamics model which could be used for simulation and what-if predictions; it involved over twenty meetings with management and affected staff [5].

Through the empirical study in Finland, Malmi found that most companies using the Balanced Scorecard were uncertainty about the relations between the indicators[6].

Kasurinen's research showed that there were some obstructive factors such as confuses, frustrates and delayers[7].

McCunn pointed out that if the enterprise were only a blind pursuit of new management tool, and could not take the problems with system thinking, it would result that the implementation of the Balanced Scorecard was not as effective as expected[8].

Ahn related and analyzed the framework of the balanced scorecard and found out that it could only provide a framework and was unable to provide a set of steps for setting up the relationships between the goals and indicators. From the standpoint of management control, the balanced scorecard was just a diagnostic control system but an interactive control system, and it is hard to lead to double-loop learning[9].

Olve suggested simulating the balanced scorecard with software in computer, by which the enterprise could find out the real structure of the balanced scorecard and furthermore provide a decision support system for the activities[10].

The literature above pointed out the balanced scorecard's shortcomings such as one-way causality, non-liner, time delay and so on. In the complex dynamic environment, the existing Balanced Scorecard's one-way causal relationship is unable to evaluate the dynamic and complex business activities of enterprises. The strategy map and balanced scorecard can only describe the enterprise's strategy statically, and can not explain the consequence of the dynamic changes of circumstances and strategy. Therefore, it is necessary to set up a dynamic management tool which is able to monitor the index value as well as simulation analysis.

3.2 Phase II

Recognizing the lack of the dynamic nature of the Balanced Scorecard, and system dynamics is given to solve the problem preliminary. Representative studies are as follows:

The founder of Balanced Scorecard Kaplan and Norton also are cognizant of these shortcomings. They also suggest using of system dynamics to make up for these deficiencies, especially in dynamic and complex situations[11].

With the viewpoint of feedback loop, Wolstenholme put the Balanced Scorecard into the process of the objectives and its implementation to form a complete circuit. This will be amendments to form a continuous feedback loop[12].

Sloper introduced systems thinking and system dynamics into the Balanced Scorecard to sort out the complex causal relationship between indicators, increased the understanding of the system, thus the performance evaluation is executed more effectively[13].

Roy found it was often difficult to distinguish the complex causal relations in the implementation process of the Balanced Scorecard. The time delay problem was even more difficult to cope with, leaving the original expectations were not achieved. The use of system dynamics could compensate for this deficiencies; and the combination can guarantee the smooth implementation of the strategy[14].

Schoenebron pointed out that the balanced scorecard was based on the partial factors' simple causal relations, and the implement of the balanced scorecard were often failed because of the dynamic complex which should be controlled by causal feedback using system dynamics[15].

Most of the researches above have a clear account of the ways which is to design the process of combining the Balanced Scorecard and the system dynamics to establish a causal relationship. However, how to use the system dynamics model to continue the next step for balanced scorecard after the model construction, very few researches referred to. The researches above focused mainly on theoretical arguments of both the feasibility and necessity of the combination of balanced scorecard and system dynamics which is lack of examples to verify the reasonableness of his theory.

3.3 Phase III

Some researchers did empirical analysis of the combination of system dynamics and balanced scorecard, demonstrated the superiority of the combination. On this basis, put forward some suggestion of the combination. Representative studies are as follows:

Rydzak.F pointed out the Balanced Scorecard was lack of dynamics, and took use of system dynamics to establish a dynamic balanced scorecard model for a business advisory service organization in United States. The simulation can clearly show the results of different strategies and provide support for managers to choose the most rational strategy[16].

Carmineianchi pointed out that with the emergence of dynamic complexity in public administration, managers need an innovative planning and control systems to improve the efficiency. The study built up a dynamic balanced scorecard management system for an urban water supply sector and verified its effectiveness[17].

Davis and O'Donnell through the following four-stage process to help companies build Balanced Scorecard performance management system: firstly, conform the company's target; secondly, using system thinking to establish causal connection between objectives and actions; thirdly, build up system dynamic flow to identify the key leading performance indicators; finally, find out the ways which can improve the key leading performance indicators[18].

In view of the side effects triggered by activities which are caused by the implementation of the Balanced Scorecard performance system and TQM improvement, Sterman took system dynamics method to diagnose the problem. The

results showed that it was caused by the complex relationship between cause and effect in the Balanced Scorecard system, when the individual performance indicators were improved, the overall situation could become worse. The figures showed that the system dynamics could indeed be used to improve the design of the Balanced Scorecard[19].

Linard pointed out that systemic concept have made minimal headway in the field of performance indicators and the balanced scorecard methodology in particular. They have suggested that failure to account for the systemic dimensions erodes the credibility of the performance management systems. They outlined the broad approach being taken in the applied research at the University, applying systems concepts in all aspects of the design and development of a balanced scorecard, at whole of government level and for elements of the bureaucracy[20].

HA Akkermans and KE van Oorschot described a case study in which system dynamics modeling and simulation was used to overcome both kinds of problems. In a two-stage modeling process, the balanced scorecard was developed for management of one organizational unit of a leading Dutch insurer. This research illustrates how, through their involvement in this development process, management came to understand that seemingly contradictory goals such as customer satisfaction, employee satisfaction and employee productivity were, in fact, better seen as mutually reinforcing[21].

The studies above have demonstrated the feasibility and necessity of the combination of the Balanced Scorecard and system dynamics, and provide some examples for related industries to implement the Balanced Scorecard.

4 The Implemental Steps of the Dynamic Balanced Scorecard

From the analysis above we can conclude that some of the balanced scorecard scholars questioned the simple cause-and-effect relationship, and then suggest using the system dynamics to make up for the balanced scorecard's lack in tacking the dynamic complex; but they did not give out specific in-depth steps of the implementation. The other scholars only proposed some problems when combing the balanced scorecard and systems dynamics. This article firstly analyses and discusses the relevance of the balanced scorecard and the service-profit chain and the study on the integration of the balanced scorecard and systems dynamics, then proposes specific steps when applying the dynamical balanced scorecard into the service industry. This article then describes a case study in a service enterprise in which system dynamics modeling and simulation was used to overcome both kinds of problems.

In this paper, we use system dynamics as an approach to overcome the limitations to current balanced scorecard theory. We suggest a two-stage and seven-steps process of system dynamics modeling for balanced scorecard development. As the distinction in system dynamics, we call these two stages the qualitative and the quantitative stage.

Stage one, qualitative stage: Develop the qualitative conceptual model from the reciprocity of the business management; draw the strategic map and then refine the indicators and determine the relationship between them.

Stage two, quantitative stage: Transform the cause-and-effect graph into the systems dynamics flow, establish a quantitative simulation model, and then carry on the

simulation analysis. The specific implemental steps of Dynamic Balanced scorecard are as follows:

4.1.1 Do a detailed understanding of the organizations which is willing to implement the dynamic balanced scorecard, determine the reasons why the organization exists, the vision and the goal of the organization through various means of discussion.

4.1.2 Translate the strategically goals of the organization into operational objectives, and then set up corresponding performance indicators in view of the specific objectives. Identify the four perspectives of the balanced scorecard and draw out the strategy map of the organization.

4.1.3 In accordance with the cause-and-effect relationship between the indicators of the strategy map, establish the cause-and-effect graph.

4.1.4 Establish quantitative system dynamic model on the base of the four perspectives of the balanced scorecard. Define each factor and the quantitative relations between them; use Vensim to set up the quantitative system dynamic flow.

4.1.5 Use systems dynamics method to exam the validity of the system dynamic model. Revise the model until it passes the exam.

4.1.6 Simulate and analyze the model and then compare theresults with the real system. Revise the model constantly to accord with the real situation.

4.1.7 Re-design the relations between the factors in the model. In view of the result provided by the model, apply this to the real system to change the current status of the system. Provide these results to the managers to support the organization's strategically performance management.

References

1. Atkinson, A.A., Waterhouse, J.H., Wells, R.B.: Approach to strategic performance measurement. Sloan Management Review 38(3), 25–27 (1997)
2. Otley, D.: Performance management: a framework for management control system research. management. Accounting Research 109(4), 363–382 (1999)
3. Mooraj, S., Oyon, D., Hostettler, D.: The Balanced Scorecard: a Necessary Good or an Unnecessary Evil. European Measurement Journal 17, 481–489 (1999)
4. Norrtklit, H.: The balance on the balanced scorecard-a critical analysis of some of its assumptions. Management Accounting Research 11, 65–88 (2000)
5. Todd, D., Palmer, E.: Development and design of a dynamic scorecard in local government. In: The European Operations Management Association 8th International Annual Conference, Bath, United Kingdom, pp. 65–70 (2002)
6. Malmi, T.: Balanced scorecard in finish companies: a research note. Management Accounting Research 12(2), 207–220 (2001)
7. Kasurinen, Tommi: Exploring management accounting change: the case of balanced scorecard implementation. Management Accounting Research 13(3) (September 2002)
8. Mccunn, P.: The balanced scorecard...the elevnth commandment. Magement Accounting, London 76(11), 34–36 (1998)
9. Ahn, H.: Applying the balanced scorecard concept: an experiential report. Long Range Planning 34, 441–461 (2001)

10. Olve, N., Roy, J., Wetter, M.: Performance drivers: a practical guild to using the balanced scorecard. John Wiley & Sons Ltd., England (1999)
11. Kaplan, R., Norton, D.: The strategy focused organization. Harvard Business School Press, Boston (2000)
12. Wolstenholme, E.: Balanced strategies for balanced scorecards: the role of system dynamics in supporting balanced scorecards and value based management. In: The 16th International System Dynamics Conference, Quebec, Canada, pp. 25–63 (1998)
13. Sloper, P., Linard, K.T., Paterson, D.: Towards a dynamic feedback framework for public sector performance management. In: The 17th International System Dynamics Conference, Wellington, New Zealand, pp. 19–25 (1999)
14. Roy, S., Roy, J.: Balanced scorecard in a dynamic environment. In: CD-ROM Proceeding of 2000 International System Dynamics Conference, Bergen, Norway (2000)
15. Schoeneborn, F.: Linking balanced scorecard to system dynamics. In: The 21st International System Dynamics Conference, New York, USA, pp. 165–182 (2003)
16. Rydzak, F., Magnuszewski, P., Pietruszewski, P., Sendzimir, J., Chlebus, E.: Teaching the Dynamic Balanced Scorecard. In: Proceedings of 22nd International Conference, System Dynamics Society (2004)
17. Bianchiy, C., Battista, G., Montemaggiore: Building dynamic balanced scorecards to enhance strategy design and planning in public utilities: key-findings from a project in a city water company. Revista de dinámica de sistemas 2(2), 3–35 (2006)
18. Davis, S., Albright, T.: An investigation of the effect of balanced scorecard implementation on financial performance. Management Accounting Research 15, 135–153 (2004)
19. Sterman, J.D., Repenning, N.P., Kofman, F.: Unanticipated side-effects of successful quality programs: exploring a paradox of organizational improvement. Management Science 43(4), 503–521 (1997)
20. Linard, K.L.: A dynamic balanced template for public sector agencies. In: CD-ROM Proceeding of 20th International Conference of the System Dynamics Society (2000)
21. Akkermans, H., Van Oorschot, K.: Developing a balanced scorecard with system dynamics. In: The 20th International System Dynamics Conference, Palermo, Italy (2002)

Joint Waveguide Invariant and Moving Target Parameters Estimation

Yun Yu, Anbang Zhao, Jingwei Yin, and Junying Hui

National Laboratory of Underwater Acoustic Technology
Harbin Engineering University
Harbin, 150001, China
yuyuntc@163.com

Abstract. It is significant to estimate the waveguide invariant and the moving target parameters accurately, because the former is a physical parameter demonstrating the propagation characteristics and environmental characteristics of the ocean, while the latter is the foundation of passive ranging and target recognition. Hough transform based waveguide invariant and the moving target parameters estimation algorithm in which neither the distance information nor the closest point of approach of the moving target is needed, is proposed combining LOFARgram and bearing-time records, more attentions are paid to the joint estimation of the waveguide invariant and the heading angle. And the moving target parameters also include the ratio of range at the closest point of approach (CPA) to source speed and the time at CPA. The feasibility of the parameter estimation algorithm has been validated by the simulation researches, which also indicate that the parameter estimation algorithm has high estimation precision.

Keywords: interference structure of low-frequency acoustic field, waveguide invariant, moving target parameters estimation, Hough transform.

1 Introduction

The waveguide invariant[1-2], usually designated as β, was proposed by S.D.Chuprov, a Russian scholar, at 1982, which is used to describe the continuous spectrum interference striation on the LOFARgram obtained by processing a moving broadband source acoustic signals. The invariant β relates the slope of the interference striations, $\mathrm{d}\omega/\mathrm{d}r$, to the range r from the source and the frequency ω, summarizes the dispersive propagation characteristics of the acoustic field, and provides a descriptor of constructive/destructive interference structure in a single scalar parameter. For a general waveguide, the waveguide invariant is not really unchanged, whose value is weakly dependent on mode order, source frequency, source/receiver depth, azimuth and so on. The β is canonical value 1 for the pekeris waveguide; Whereas it is -3 in surface waveguide with a linear variation in the square index of refraction, $n^2(z)=1-az$ [2]; In most cases, β is not a constant, but is treated as a

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 83–93.
springerlink.com

distribution[3], and the range is $-3 \le \beta \le 1$. In reality, β is approximate 1 in the shallow water environment.

Since the waveguide invariant theory[3-4] was proposed, it has been extensively researched and achieved some important applications[6-12] on many fileds, such as acoustic signal processing and environmental parameters invertion. Spain and Kuperman[4] extended an invariant expression which is suitable in the shallow water environments that vary in range and azimuth and is used to interpret and predict the stripe on the spectrograms obtained in the sea experiment. Rouseff et al. [3] modeled the waveguide invariant as a distribution. Scholars expected to improve the performance of Time-Reversal Mirror[5-6], Matched-Field Processing[7] and the Virtual Receiver[8] algorithms by combining them with waveguide invariant concepts. Turgut et al. [9] applied the waveguide invariant theory to horizontal line array beamformer output, estimated the invariant β and the moving parameters based on the generalized Radon transforms to localize moving broadband noise sources from measured acoustic intensity striation pattern. K.D.Heaney[10] et al. exploited a rapid geoacoustic characterization and environmental assessment algorithm applied to range-dependent environments by adopting the overall transmission loss, the time spread and the invariant β. In addition to the aspects mentioned above, the waveguide invariant theory is also developed to suppress the interference [11] and reverberation [12].

It is significant to accurately estimate the waveguide invariant, a physical parameter demonstrating the propagation characteristics and environmental characteristics of the waveguide, and the moving target parameters, such as the ratio of range at the closest point of approach (CPA) to source speed, the bearing and the heading angle, which are the premise of passive ranging and target recognition. Hough transform based waveguide invariant and the moving target parameters estimation algorithm is proposed combining LOFARgram and bearing-time records, the focus is on the joint estimation of the waveguide invariant and the heading angle. Neither the distance information which is demanded in [13] nor the closest point of approach of the moving target which is demanded in [14] is needed by the parameters estimation algorithm. Compared with [9], three-dimensional searching is omitted, but two-dimensional parameter space is calculated. So the algorithm in the present paper is more universal and less computational cost.

2 The Principle of Parameter Estimation

2.1 The Motion Equation

Provided that the target radiates continuously broadband signals and moves in a uniform rectilinear motion, the speed is v, and the range at the closest point of approach (CPA) is r_0.The sketch map of target movement is shown in Fig.1.

The origin of coordinates is located at the acoustic center of the single sensor or the array, so by analyzing the geometric relationship shown in Fig.1, the trajectory equations of the moving target are given as follows:

$$r(t) = \frac{r_0}{\sin(\theta - \varphi)} \tag{1}$$

$$r(t) = \sqrt{v^2\tau^2 + r_0^2}\ , \quad \tau = t - t_0 \tag{2}$$

Where θ is target bearing, φ is the heading angle which is defined as the angle between the positive axis of x and the target's moving direction, and t_0 is the time at CPA, the definition of the remaining variables are as mentioned above.

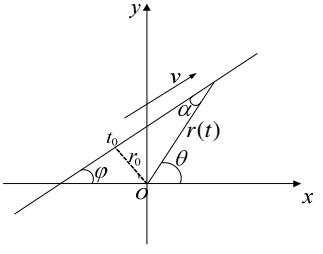


Fig. 1. The sketch map of target movement

Fig.1 shows that:

$$\tan(\alpha) = \frac{r_0}{v(t-t_0)} = \frac{r_0/v}{(t-t_0)} \tag{3}$$

$$\alpha = \theta - \varphi \tag{4}$$

Substituting the Eq. (4) into Eq. (3) yields

$$\tan(\varphi) = \frac{-\frac{r_0}{v}\cos\theta + (t-t_0)\sin\theta}{\frac{r_0}{v}\sin\theta + (t-t_0)\cos\theta} \tag{5}$$

2.2 The Waveguide Invariant

According to the definition, the waveguide invariant in the range-independent waveguide can be expressed as [2]:

$$\beta = \frac{d\omega}{dr}\frac{r}{\omega} = -\frac{\Delta S_{pmn}}{\Delta S_{gmn}} \tag{6}$$

Where ω is the frequency of acoustic signal, r is the range from the source, β is the waveguide invariant, and $\Delta S_{pmn} = 1/v_m - 1/v_n$ is the mode phase slowness difference, $\Delta S_{gmn} = 1/u_m - 1/u_n$ is the mode group slowness difference, v_m and u_m is the phase velocity and group velocity of the m th mode, respectively.

Seen from the Eq. (6), two methods can been obtained to estimate β. The first term in the Eq. (6) shows that the value of β can been estimated by extracting the slope of the interference striations on the $r-f$ interference pattern based on the image processing, which is the principle the [13] based on; while the second term in the Eq.

(6) indicates that the value of β can also be calculated numerically by modeling the acoustic field to get the mode phase velocity and group velocity, which requires accurate prior knowledge of the wave propagation environment.

2.3 The Description of the Interference Striation

Power spectrum analysis to the received signal at each discrete time is one section of the LOFARgram, which is parallel to the horizontal axis (f axis), and the temporal evolution of all sections integrate the LOFARgram. The $r-f$ interference pattern mentioned above may be observed in the propagation experiment, whereas sonar can only get the $t-f$ interference pattern (namely LOFARgram) in most scenarios. For the high signal to noise ratios, just single-element can obtain clear stripes on the LOFARgram. However, for the low signal to noise ratios cases when the target is far, the array processing is resorted to and spectrum analysis should be implemented to the tracking beam output to get a clear interference pattern, which has a farther detection distance and can tell multiple sources at different azimuths.

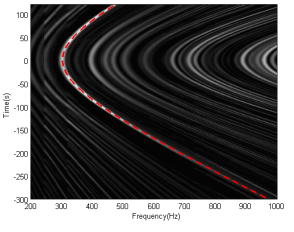


Fig. 2. The interference structure of target in intermediate or short range: range at the Closest Point of Approach $r_{CPA}=1200$ m, the moving velocity is 6m/s and the sea depth is 55m.

The Fig.2 obtained in the simulation shows the interference structure of target in intermediate or short range, which exhibits the obvious striated bands of intensity maxima and minima corresponding to constructive and destructive interference of normal modes, and the striations are a family of quasi hyperbolas. The slope of the striations can be written as:

$$\frac{df}{d\tau}=\frac{df}{dr}\cdot\frac{dr}{d\tau} \tag{7}$$

And the Eq. (6) can be expressed as:

$$\frac{df}{dr}=\frac{f}{r}\beta \tag{8}$$

Derivation of the Eq. (2) yields

$$\frac{dr}{d\tau}=\frac{v^2\tau}{\sqrt{v^2\tau^2+r_0^{\ 2}}} \tag{9}$$

Substituting the Eq. (8) and Eq. (9) into Eq. (7), then integrating both side of the equation and rearrangement yields

$$f = f_0 \left[1 + (\frac{v}{r_0})^2 \tau^2 \right]^{\beta/2} \tag{10}$$

Supposing t_s is the reference time, and the coordinates of some interference striation at this time is (f_s, t_s), then the expression of the interference striation can be obtained by variables separation and integration to the Eq. (8), shown in Eq. (11).

$$\left(\frac{f(t)}{f_s} \right)^{1/\beta} = \frac{r(t)}{r_s} \tag{11}$$

Where $r_s = r(t_s)$ and $f_s = f(t_s)$ are called the reference distance and the reference frequency.

The moving equation corresponding to the reference time t_s can be known from Eq. (1), rewriting it as follows:

$$r_s = \frac{r_0}{\sin(\theta_s - \varphi)} \tag{12}$$

Substituting the Eq. (1) and Eq. (12) into Eq. (11), one gets

$$f(t) = f_s \left[\frac{\sin(\theta_s - \varphi)}{\sin(\theta(t) - \varphi)} \right]^{\beta} \tag{13}$$

All the Eq. (10), Eq. (11) and Eq. (13) are the descriptions of the interference striation. The Eq. (10), selecting the time at CPA t_0 to be the reference time, describes the interference striations in the LOFARgram, and if the $\beta = 1$, the striations are a family of hyperbolas. While the Eq. (11) describes the interference striation in the $r - f$ interference pattern, which requires prior knowledge of target's distance information. And the Eq. (13), relating the frequency information of the interference striation in the LOFARgram to the target's bearing in the bearing-time records at the same time, is the function of the target's heading angle and the waveguide invariant, and it can select an arbitrary time t_s as the reference time, not be limited to adopt t_0. This is the principle of the parameter estimation algorithm proposed in this paper.

For the shallow water environment, $\beta \approx 1$ as mentioned above, so the Eq. (13) can be simplified as

$$\tan\varphi = \frac{f_2 \sin\theta_2 - f_1 \sin\theta_1}{f_2 \cos\theta_2 - f_1 \cos\theta_1} \tag{14}$$

Where θ_1 and θ_2 are bearings of the target at t_1 and t_2, respectively. f_1 and f_2 are frequencies of some light striation corresponding to t_1 and t_2, while t_1 and t_2 are two different arbitrary time. Known from this equation, only two frequencies on

the interference striation and corresponding two bearings at time interval $\Delta t = t_1 - t_2$ are required to realize the estimation for φ. It is called the double points measurement algorithm, which is simple and whose calculation cost is not expensive, but it assumes $\beta = 1$.

2.4 Parameter Estimation by Hough Transform

Hough transform [15], a special case of generalized Radon transform, has been extensively applied in computer vision and pattern recognition, object detection, feature recognition and other fields. Hough transform is an effective edge detection method by mapping the image domain to a parameter domain, and it is suitable to detect arbitrary curve.

The essence of the Hough transform is providing a mapping from points on the same curve in the image space to a family of curves intersected at one point in the parameter space, the coordinate of the intersection reflect the parameter of the curve in the image space. The intensity of each element (a,b) in the parameter space is the cumulative intensity of the points on the curve characterized by the parameters (a,b) in the image space, so the parameter of the curve we concerned can be achieved by searching the maximum element in the parameter space.

The joint estimation of the waveguide invariant and the heading angle can be achieved by combing the LOFARgram and the bearing-time records when Eq. (13) is used as a Hough transform template. The LOFARgram is just the image space, while the parameter space is the plane with the heading angle as the horizontal axis and the waveguide invariant as the vertical axis.

Of course, the Eq. (10) can also be the Hough transform template to process the LOFARgram, so the r/v_0 and β can be estimated synchronously; in the same way, r/v_0 and φ can be estimated synchronously when Eq. (5) is adopted to be the Hough transform template to process the bearing-time records. Those parameters estimation algorithms have been studied in [9] and [14], so not repeat them here. If the heading angle has been estimated by the double points measurement algorithm, Eq. (5) can be used as a Hough transform template to process the bearing-time records, so r/v_0 and t_0 can be estimated synchronously without three-dimensional searching.

3 Simulation Researches

Simulation researches are conducted in this section to verify the correctness and feasibility of the parameters estimation algorithm. For simplicity, single vector sensor who can estimate the object direction is adopted in the simulation.

The conditions used in the simulation are as follows: the Pekeris model is used. The sea depth is $H = 55$ m. The acoustic velocity and density of water are $c_1 = 1500$m/s and $\rho_1 = 1000$kg/cm3 respectively. While the acoustic velocity and density of bottom medium are $c_2 = 1610$m/s and $\rho_2 = 1900$kg/cm3, respectively. The depth of the vector sensor is $z_r = 30$ m. Supposing that the target cruises in the same

depth which is $z_s = 4$ m, the speed of navigation is $v = 12$ m/s, and the range at the CPA is $r_0 = 1200$ m, so the ratio of range at the CPA to source speed is 100s. The time at the CPA is set as the 0 time, and the time is defined negative when the target moves towards the receiver, and positive the opposite case. The heading angle is 30 degree. The working band is 200~1000Hz. The acoustic field is modeled by the KRAKENC program. Then the flow chart for the simulation is shown in Fig.3.

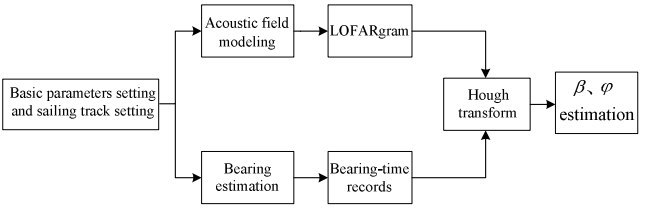


Fig. 3. The simulation flow chart of the parameter estimation

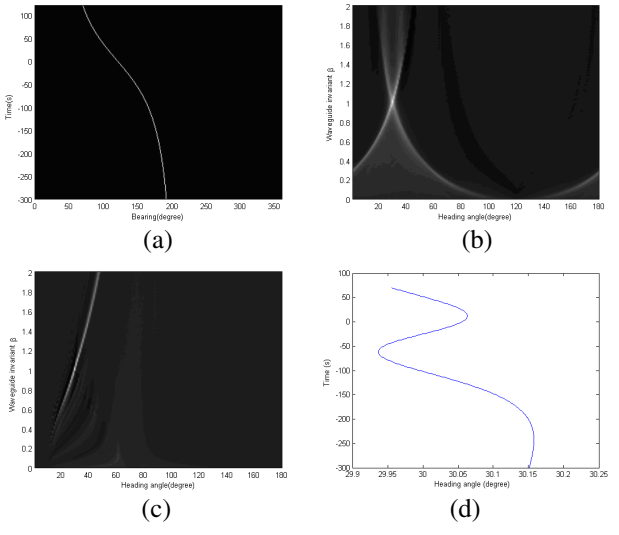


Fig. 4. The simulation results of the scenario with the CPA: (a) Bearing-time records, (b) Parameter space ($t_s = 0$ s, $f_s = 304$ Hz), (c) Parameter space ($t_s = -300$ s, $f_s = 970$ Hz), (d) The result of the double points measurement ($\Delta t = 50$ s).

The simulation results of the scenario with the CPA are shown in Fig.4 when $t = -300 \sim 120$ s, and the corresponding LOFARgram is shown in Fig.2, where Fig.4(a) is the simulated bearing-time records; Fig.4(b), selecting $t_s = 0$ s, $f_s = 304$ Hz and $\theta_s = 120^\circ$ as the reference, is the parameter space of the Hough transform using the Eq. (13) as a template and; and Fig.4(c) has the same meaning as Fig.4(b) excepting that $t_s = -300$ s, $f_s = 970$ Hz and $\theta_s = 191.6^\circ$ is selected as the

reference. The simulation results indicate that although different reference is selected and images of the Fig.4(b) and Fig.4(c) are also different, the same parameter values $\beta = 1.01$, $\varphi = 30^\circ$ are achieved corresponding to the maximum in the Fig.4(b) and Fig.4(c), which verifies the accuracy of the algorithm and is also the result from extracting the same interference striation, the brightest one shown as the dotted line in Fig.2. When the time interval is $\Delta t = 50$ s, the heading angle estimated by the double points measurement algorithm is shown in Fig. 4(d) which indicate the high precision of the algorithm.

The simulation results of the scenario without the CPA are shown in Fig.5 when $t = -700 \sim -280$ s, where Fig.5(a) is the LOFARgram accordingly, which shows the interference striations formed by the target in the distance are a family of straight lines; Fig.5(b) has the same meaning as Fig.4(b) selecting $t_s = -300$ s, $f_s = 970$ Hz and $\theta_s = 191.6^\circ$ as the reference, and the parameter values $\beta = 1.09$, $\varphi = 31^\circ$ are achieved corresponding to the maximum in parameter space. Compared with the results with the CPA, the peak in the parameter space is not sharp enough and some highlights (not only one) appear which affect the accuracy of the parameter estimation. Fig.5(c) is one section of the parameter space when $\beta = 1$, the estimated heading angle is $\varphi = 30^\circ$, which is entirely consistent with the true value. And the Fig. 5(d) is one section of the parameter space when $\varphi = 30^\circ$, the estimated waveguide invariant is $\beta = 1.01$. That is to say, if one parameter is known as prior knowledge, the other can be accurately estimated.

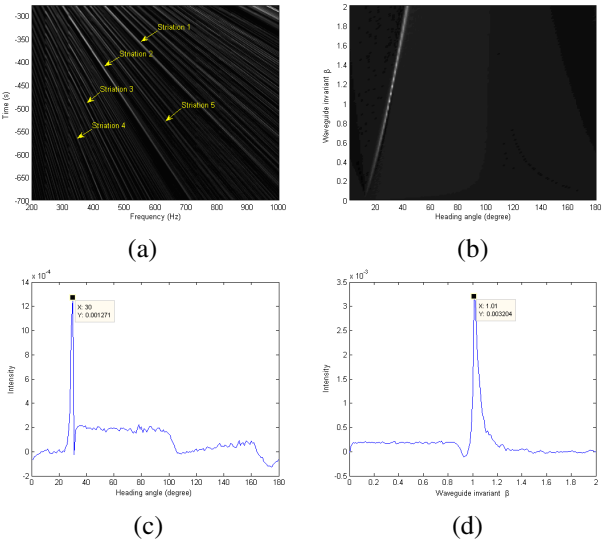
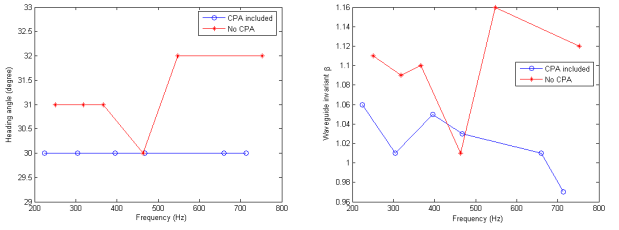


Fig. 5. The simulation results of the scenario without the CPA: (a) LOFARgram, (b) Parameter space ($t_s = -300$ s, $f_s = 318$ Hz), (c) The section of $\beta = 1$, (d) The section of $\varphi = 30^\circ$.

Obtained by double points measurement, the TABLE.1 is the heading angle estimation results of the five interference striations signed in the Fig.5 (a), which shows that each striation has high accuracy of heading angle estimation. This algorithm is simple, and has much less computation cost than the Hough transform, nevertheless the waveguide invariant should be known as prior knowledge.

Table 1. Heading Angle Estimation Results Obtained by Double Points Measurement

The number of the striations	1	2	3	4	5
Heading angle /degree	30.11	29.79	29.62	29.90	30.04

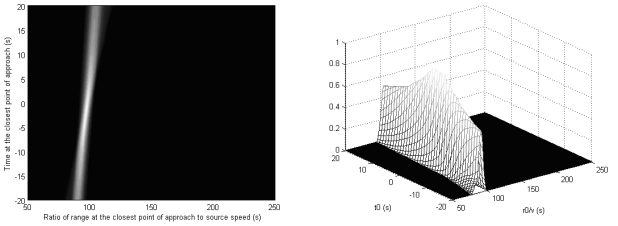


(a) (b)

Fig. 6. The parameter estimation results of different reference points: (a) The estimation of the heading angle, (b)The estimation of the waveguide invariant.

Fig.6 is the parameter estimation results of different reference points, where solid line with circle symbols is the results corresponding to the scenario with the CPA when the reference time $t_s = 0$ s, while solid line with star symbols is the results corresponding to the scenario without the CPA when the reference time $t_s = -300$ s. The results indicate that: the parameter estimation accuracy of the scenario with the CPA is better than the scenario without the CPA; if different reference frequencies are adopted, in other words, extracting the parameter of different interference striations, the corresponding results of parameter estimation accord with each other. The parameter estimation of the scenario without the CPA also has satisfactory results without considering the local noise, so the parameter estimation algorithm proposed here is more universal, because the CPA is not necessary to present in the sailing track of the moving target.

The focus is on the estimation of r_0/v after finishing the estimations of the heading angle and β. Substituting φ=29.62° estimated by double points measurement using the interference striations 3 into Eq. (5), then processing the bearing-time records by adopting the Eq. (5) as the Hough transform template, the estimation of the r_0/v and t_0 can be achieved which is shown in Fig. (7), whose horizontal axis is



(a) (b)

Fig. 7. The results of the Hough transform to the bearing-time records: (a) Quasi three-dimensional display, (b) Three-dimensional display

r_0/v and vertical axis is t_0. A bright band with large slope can be seen in Fig. (7), which shows r_0/v can be obtained with high precision, but estimation accuracy of t_0 is low. The maximum in the parameter space is the estimation values, which is $r_0/v = 97$ s and $t_0 = -2$ s, while the true values are 100s and 0s, respectively. So the estimation accuracy can be acceptable.

4 Conclusion

Waveguide invariant can be used to describe the interference striations in the low-frequency sound field interference structure, can be used for geoacoustic inversion, and can be used to explore new underwater acoustic detection technology. Waveguide invariant and moving target parameter estimation algorithm is proposed here based on the Hough transform to the LOFARgram and the bearing-time records. The focus is on the joint estimation of the waveguide invariant and the heading angle, neither the distance information nor the closest point of approach of the moving target is needed, so the algorithm is more universal. The simulation researches verify the correctness and feasibility of the algorithm, and the results show that: the parameter estimation accuracy of the scenario with the CPA is better than the scenario without the CPA; if different reference frequencies are adopted, in other words, extracting the parameter of different interference fringes, the corresponding results of parameter estimation accord with each other. The parameter estimation of the scenario without the CPA also has satisfactory results without considering the local noise.

Acknowledgment. This work was supported by National Natural Science Foundation of China(Grant No. 51009041), the Science and Technology Foundation of State Key Laboratory of Underwater Acoustic Technology (Grant No. 9140C2002100802) and Special Funding of Marine Scientific Research and Public Service Sectors (Grant No. 201005001-05).This paper is funded by the International Exchange Program of Harbin Engineering University for Innovation-oriented Talents Cultivation.

References

[1] Chuprov, S.D.: Interference structure of sound field in the layered ocean in Ocean Acoustics. In: Brekhovskikh, L.M., Andreeva, I.B. (eds.) Modern State, Nauka, Moscow, pp. 71–91 (1982)

[2] Brekhovskikh, L.M., Lysanov, Y.P.: Fundamentals of Ocean Acoustics. Springer, New York (2002)

[3] Rouseff, D., Spindel, R.C.: Modeling the waveguide invariant as a distribution. In: AIP Conference Proceedings, vol. 621(1), pp. 137–160 (2002)

[4] D'Spain, G.L., Kuperman, W.A.: Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth. J. Acoust. Soc. Am. 106(5), 2454–2468 (1999)

[5] Song, H.C., Kuperman, W.A., Hodgkiss, W.S.: A time-reversal mirror with variable range focusing. J. Acoust. Soc. Am. 103(6), 3234–3240 (1998)

[6] Hodgkiss, W.S., Song, H.C., Kuperman, W.A.: A long-range and variable focus phase-conjugation experiment in shallow water. J. Acoust. Soc. Am. 105(3), 1597–1604 (1999)

[7] Thode, A.M., Kuperman, W.A., D'Spain, G.L., Hodgkiss, W.S.: Localization using Bartlett matched-field processor sidelobes. J. Acoust. Soc. Am. 107(1), 278–286 (2000)

[8] Thode, A.M.: Source ranging with minimal environmental information using a virtual receiver and waveguide invariant theory. J. Acoust. Soc. Am. 108(4), 1582–1594 (2000)

[9] Turgut, A., Orr, M., Rouseff, D.: Broadband source localization using horizontal-beam acoustic intensity striations. J. Acoust. Soc. Am. 127(1), 73–83 (2010)

[10] Heaney, K.D.: Rapid Geoacoustic Characterization Applied to Range-Dependent Environments. IEEE Journal of Oceanic Engineering 29(1), 43–50 (2004)

[11] Yang, T.C.: Beam intensity striations and applications. J. Acoust. Soc. Am. 113(3), 1342–1352 (2003)

[12] Goldhahn, R., Hickman, G., Krolikc, J.: Waveguide invariant broadband target detection and reverberation estimation. J. Acoust. Soc. Am. 124(5), 2841–2851 (2008)

[13] Tian, L.A., Liu, F.C., Zhou, S.H.: Waveguide invariant estimation using Hough transform. Acoustics and Electronic Engineering (4), 22–24 (2009)

[14] Yu, Y., Hui, J.Y., Yin, J.W., Hui, J., Wang, Z.J.: Moving target parameter estimation and passive ranging based on waveguide invariant theory. Acta Acoust (2010) (accepted)

[15] Hough, P.V.: A method and means for recognizing complex patterns. U.S. Patent: 3069654

Multicore-Based Performance Optimization for Dense Matrix Computation

Guoyong Mao[1], Xiaobin Zhang[2], Yun Li[2], Yujie Li[2], and Laizhi Wei[2]

[1] Department of Electronic Information and Electric Engineering
Changzhou Institute of Technology
Changzhou Key Lab for Research and Application of Software Technology Changzhou 213002, China
[2] College of Information Engineering Yangzhou University Yangzhou 225009, China
zxb0412@163.com, maogy@czu.cn

Abstract. To make the traditional applications benefit from multicore processors, the traditional Gaussian Elimination algorithm is improved to enhance its parallel performance under multicore architecture by matrix partition. The stability of the original algorithm is guaranteed. The hit rate of cache is improved by adjusting the computation sequence, the experiment shows that the speedup can reach 1.8 under duo core CPU environment when evaluating the inverse of dense matrix.

Keywords: Gaussian elimination, matrix block, multicore, parallel computing.

1 Introduction

Nowadays, to improve the processing ability of CPU, the chip manufacturer focuses more on multicore processors[1],that is, integrate more processing units on single chips, which has a big influence on performance upgrade method designed for traditional applications. As a result, the TLP (Thread Level Parallelism) under multicore environment attracts more attention than the traditional ILP (instruction level parallelism) now.

Though multicore chip can improve the computation precision and performance, it still faces big challenge --``memory wall'', that is, delay of memory fetch[2][3]. From Sun's formula we can find that when such delay is fixed, the speedup can be improved effectively by increasing more data in L1 cache to decrease the memory fetch time, as the data fetching speed in L1 cache is generally regarded the same as the speed of processor. In this way, the efficiency of multicore processors can be improved[2].

The computation of dense matrix is widely applied in numerical computing, it usually occupies most of the time in such computations like solving the numerical problems using finite elements methods. The research on computation of dense matrix under multicore architecture gets more and more attention now, various high efficiency algorithms under different types of CPUs are brought forward[4]. However, the traditional algorithm like BLAS and LAPACK are based on block matrix, the efficiency of the corresponding parallel version like ScaLapack is not good, as it can

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 95–102.
springerlink.com

not make full advantage of the multicore processors[5].Hence, in this paper, based on existing BLAS library, we will focus on how to use the multicore processors more efficiently, such that the traditional applications can also benefit from multicore processors.

The evaluation of the inverse of matrix is widely applied in numerical computation. The complete Gaussian pivoting elimination method is very stable in evaluating the inverse of matrix, with the appearance of multicore processor, and together with the ILP technology provided by newly developed processors, this method can ensure good performance. We will use this method as the base in our experiment, try to overcome the "memory wall" problem for multicore processors, and optimize the traditional Gaussian method using block matrix.

2 Evaluate the Inverse of Dense Matrix

2.1 Serial Gaussian Elimination Method

The evaluation of the inverse of dense matrix using Gaussian elimination method can be found in many papers, $\sum_{i=1}^{n}(n-i+1)\times(2n-i)+\sum_{i=2}^{n}(i-1)\times n$ multiplications and $\sum_{i=1}^{n-1}(n-i)\times(3n-i)+\sum_{i=2}^{n}(i-1)\times n$ subtractions are needed for this computation, the time complexity reaches $O(n^3)$. This method is actually serial, and cannot benefit from multicore processors; on the other hand, the cache usage in this algorithm is not good, as data read and write from memory is needed by every step of computation; the cache hit rate for this algorithm is low, which will further deteriorate the overall performance of computation.

2.2 Parallelized Gaussian Elimination Method

For standard Gaussian elimination algorithm, lots of matrix row transformations are needed. For matrix with n rows, n(n-1) row transformations are needed to transform the original matrix into diagonal matrix. As the change of the order of row transformation will not affect the final result, we can partition the original matrix based on rows into many sub matrices and do the row transformation for these sub matrices in parallel, then merge the transformation result, partition the matrices again and do the transformation. The final result can be reached by doing the steps repeatedly. Hence, the parallel computation of Gaussian elimination can be implemented by partition the sub matrix of original matrix. It is obviously that the more the sub matrices, the more the parallelization. When the number of processors p is in direct proportion to the number of rows n, the time complexity is $O(n^2)$.

To increase the fetch speed of data, L1 cache is added into processors. Though the speed of L1 cache is very high, the capacity of it is very low because the limitation in manufacturing technology and cost. For example, the capacity of L1 cache is only 16KB for some types of Intel CPUs. Hence, it means little for large matrix as L1 cache can only accommodate very small matrix. On the other hand, the size of L2 cache is usually greater than 1MB, some even reach 8MB, so that it can accommodate relatively large matrix; though the speed of L2 cache is lower than that of L1 cache, it is much faster than that of memory. Therefore, if the data hit rate of L2 cache can be improved,

the performance of Gaussian elimination algorithm can be improved effectively. To increase the operation times of data in L2 cache, we need to add as much as possible the rows of matrix to increase the times of row transformation, which will lead to the decrease of columns, however, the increase of parallelization of block matrix demands less rows in block matrix, to reach a balance, the sub matrix is set to square matrix, and a n ×2n matrix is partitioned as illustrated in Figure 1. The traditional Gaussian elimination algorithm to evaluate the inverse of matrix has to be changed in accordance with such matrix partition and with the operation condition of multicore processors. The changed algorithm and the time complexity will discussed below.

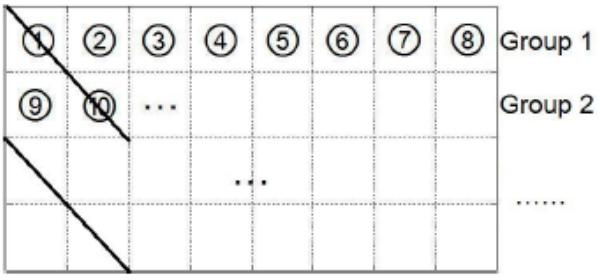


Fig. 1. Partition of matrix

Before we do row transformation for n×2n augmented matrix C, it is needed to partition augmented matrix C into multiple s×s square sub matrices logically, the size of sub matrix matches the size of L2 cache, and the partitioned sub matrices is illustrated in figure1. For the sake of simplicity, all sub matrices in one row is referred to as sub matrix group, for example, sum matrices 1, 2, ….8 are in sub matrix group 1. After partition, the number of sub matrix group is l = n/s. The improved algorithm is listed below.

1) Make elementary row transformation to sub matrix in the first row, and transform it into identity diagonal matrix, and store the transformation coefficient in the original sub matrix.

2) Set index i of sub matrix group to 1.

3) Make elementary row transformation to sub matrix group with index i based on data in sub matrix group i, and transform the sub matrix indexed i in the i+1 row into O matrix.

4) Make elementary row transformation to sub matrix group indexed i+1, till the i+1 sub matrix is transformed into identity diagonal matrix, and store the transform coefficient in this sub matrix.

5) Make elementary row transformation to l-i-1 sub matrix group indexed from i+2, i+3, …, l, based on data in sub matrix group i, till the i sub matrix in these sub matrix groups are transformed into O matrix.

6) Reset the index i of sub matrix to i+1, if i≥ n-1, go to step 3; otherwise, go next.

7) Reset the index i of sub matrix group to n-1.

8) Make elementary row transformation to sub matrix group indexed i based on data in the i+1 sub matrix group, till the i+1 sub matrix in this row is transformed into O matrix.

9) Make elementary row transformation to i-1 sub matrix group indexed from 1, 2, ….., l, based on data in sub matrix group i+1, till the i+1 sub matrix in these sub matrix groups are transformed into O matrix.

10) Make elementary row transformation to i-1 sub matrix group indexed from 1, 2, ….,l based on data in sub matrix group i, till the i sub matrix in these sub matrix groups are transformed into O matrix.

11) Reset the index i of sub matrix group to i-1 , if i>2, go to step 8, otherwise, the algorithm ends.

When the procedures listed above are finished, matrix E is transformed into inverse matrix A^{-1}. This algorithm needs a little more computation than the traditional algorithm, as the computation order is adjusted in this algorithm. However, the time complexity is still $O(n^3)$. Furthermore, all the steps other than step \ref{step01} and step 1 can be executed in parallel, i.e. these steps can be executed in parallel in multiple processing units.

This algorithm is highly stable and scalable. From the computation procedures listed above we can find out that during the transformation process for every sub matrix group, the transformation only related to m coefficients stored in one sub matrix located in the left column of this row, and only two sub matrix need elementary row transformation, the transformation has nothing to do with other sub matrix. Therefore, even the transformation process for two sub matrix in the same group can be distributed into different processing units, and finished in parallel, which guarantees the scalability of the algorithm. On the other hand, when the data is loaded into L2 cache, the data will be computed as much times as possible, such that the cache hit rate and the efficiency of data usage of this algorithm is improved. In fact, the average times of sub matrix be loaded into cache is (l-1)/2, that is, the average load times for every row is (l-1), which means a significant cache hit rate improvement over the traditional Gaussian algorithm with average load times n× (n-1).

3 Implementation of Improved Gaussian Elimination

3.1 Implementation of Algorithm

To implement the improved algorithm listed above, and to improve parallelism, the transformation matrix and its corresponding identity matrix will be divided, as shown in figure1, that is:

C = A|E

$$= \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1m} & E_{11} & E_{12} & \dots & E_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} & E_{21} & E_{22} & \dots & E_{2m} \\ & & \dots & & & \dots & & \\ A_{m1} & A_{m2} & \dots & A_{mm} & E_{m1} & E_{m2} & \dots & E_{mm} \end{pmatrix}$$

Where A_{11}, A_{12}, ……, A_{1m-1}, $A_{21,}$ ….., A_{mm-1} are all k-order square matrices, A_{1m}, A_{2m}, ….., A_{m-1m} are k× l matrices and $A_{m1,}$ $A_{m2,}$……, A_{mm-1} are l× k matrices, and A_{mm} is l× l matrices. Three sub programs are needed to implement this algorithm, the first sub program is used to diagonalize the sub matrix block, the second sub program is used to transform the sub matrices in the same row after diagonalization, and the third is used to transform the sub matrices that are not in the same row after diagonalization. The three sub programs are illustrated in algorithm 1, algorithm 2 and 3 respectively.

Algorithm 1 DIAGONALIZING(A_{ss}, k)

```
1: for i := 1 to k do
2:     a_i ← a_i / a_ii
3:     {where a_i is the ith row vector of A_ss, a_ii is the
       element in A_ss at i row and i column}
4:     for j := 1 to k do
5:         if j ≠ i then
6:             a_j ← a_j − a_ji · a_i
7:         end if
8:     end for
9: end for
```

We should notice that square matrix Ass will be transformed into the identity diagonal matrix after algorithm 1, so it is not necessary to save the identity matrix. Therefore, in algorithm 1, it is the coefficients of each row transformation that are saved in Ass, not the identity matrices after transformation. The coefficients saved in Ass will be used by subsequent algorithms to do transformation directly.

After A_{ss} is diagonalized, the corresponding row transformation for sub matrices A_{st} and E_{sp} in the same row is needed, such transformation procedure is shown in algorithm 2.

Algorithm 2 POSTDIAGONALIZED(A_{st}, A_{ss}, k)

```
1: for i := 1 to k do
2:     {a_i is the ith row vector of A_st or E_sp, a_ii is the
       element in A_ss at i row and i column}
3:     a_i ← a_i / a_ii
4:     for j := 1 to k do
5:         if j ≠ i then
6:             { a_j is the jth row vector of sub matrix
               A_st or E_sp, a_ji is the element in A_ss at
               j row and i column}
7:             a_j ← a_j − a_ji · a_i
8:         end if
9:     end for
10: end for
```

Algorithm 3 HANDLECOLBLOCK($A_{rt}, A_{st}, A_{rs}, k$)

```
1: for i := 1 to k do
2:     for j := 1 to k do
3:         {a_i is the ith row vector of A_rt or E_rp, a_ij
           is the element in A_rs at i row and j column,
           a_j is the jth row vector in A_st.}
4:         a_i ← a_i − a_ij · a_j
5:     end for
6: end for
```

Algorithm 4 INVERSEMATRIX(A, k, m)

```
1: DIAGONALIZING(A_11, k)
2: for parallel s := 2 to m do
3:     if s ≠ m then
4:         DIAGONALIZING(A_ss, k)
5:     end if
6:     for t := s + 1 to m do
7:         POSTDIAGONALIZED(A_st, A_ss, k)
8:         for r := 1 to m do
9:             if r ≠ s then
10:                HANDLECOLBLOCK(A_rt, A_st, A_rs, k)
11:            end if
12:        end for
13:    end for
14:    for p := 1 to m do
15:        POSTDIAGONALIZED(E_sp, A_ss, k)
16:        for q := 1 to m do
17:            if q ≠ s then
18:                HANDLECOLBLOCK(E_qp, E_sp, A_qs, k)
19:            end if
20:        end for
21:    end for
22: end for
```

The transformation procedure for sub matrices Art and E_{rp} that are not in the same row is shown in algorithm 3.

Algorithm 4 illustrates the procedure of evaluating the inverse of matrix A.

Notice that in algorithm 4, the loop started from line 2 can be executed in parallel, so that good speedup can be reached.

3.2 Experimental Environment and Results

In the improved algorithm to make elementary row transformation, data in two sub matrixes are needed, also, the transformation procedure needs many variables, the number of these variables are the same as the dimensions of sub matrix block. Taking these factors into account, the size of sub matrix block is slightly smaller than half of

the size of L2 cache. In duo core CPU of Intel, the L2 cache is shared by both cores, we can simply imagine that each core get half capacity of the L2 cache, so the size of sub matrix block will get 1/4 of the total capacity of L2 cache. BLAS is a very popular library, and has been optimized by various processor manufactures based on different type of processors, so we will use BLAS library specially optimized for some processors when dealing with matrix vector computations.

The experiment is done on Intel Core 2 CPU, which has 2 processing units, each of them has 3MB L2 cache, the frequency of the processor is 2.26GHz, the memory size is 4GB, the operating system is Redhat Linux AS 5, and the compiler is Intel C++ compiler 10. The source code of BLAS library is taken from Intel MKL10, and is recompiled based on this Intel Core 2 CPU. The experiment data is a randomly generated 4096× 4096 double precision matrix. To ensure that the computation can be successfully accomplished, the full pivoting procedure is finished first. The speedup reaches 1.8 when this algorithm is used.

The improved algorithm can significantly improve the usage of data in cache, data is processed as much times as possible, hence, the times that data needs to be repeatedly loaded from memory to caches is reduced, so the performance of program is improved. On the other hand, though the existing processors have good data processing ability, the data transfer ability is relatively poor. The analytical result shows that in chips with 4 processing units, the memory transfer speed can only satisfy 2 processing units; other 2 units have to be in the status of idle, as they are lack of data to be processed. In the improved algorithm, the data in cache is processed repeatedly, so the contest between every processing unit to the data bus is significantly reduced, such that the abilities of all processing units in multicore chips can be used.

4 Conclusion

The Gaussian elimination algorithm is modified to improve parallelism and scalability in this paper. By increasing the hit rate of data in caches in the computation process, and by increasing the usage of data in caches, the execution efficiency of program is effectively improved, related implementation issues are discussed in detail. The experiment studies a way to improve the performance of traditional applications under multicore environments.

Acknowledgment. This paper is supported by the Natural Science Foundation of Jiangsu Province, China, project number: No. BK2009697; the ``Six Talent Peaks Program'' of Jiangsu Province of China, and Natural Science Research Project of Changzhou Institute of Technology, Project Number: YN0816, The authors are grateful for the anonymous reviewers who made constructive comments.

References

1. Pollack, F.J.: New microarchitecture challenges in the coming generations of CMOS process technologies. In: MICRO 32: Proceedings of the 32nd Annual ACM/IEEE International Symposium on Microarchitecture, p. 2. IEEE Computer Society, Washington, DC (1999)

2. Sun, X., Chen, Y.: Reevaluating Amdahl's law in the multicore era. Journal of Parallel and Distributed Computing 70, 183–188 (2010)
3. Hill, M.D., Marty, M.R.: Amdahl's law in the multicore era. IEEE Computer 41(7), 33–38 (2008)
4. Kurzak, J., Alvaro, W., Dongarra, J.: Optimizing matrix multiplication for a short-vector SIMD architecture – CELL processor. Parallel Computing 35, 138–150 (2009)
5. Chan, E., Quintana-Orti, E.S., Quintana-Orti, G., van de Geijin, R.: Supermatrix out-of-order scheduling of matrix operations for SMP and multicore architectures. In: 19th Annual ACM Symposium on Parallel Algorithms and Architectures, SPAA 2007, pp. 116–125 (2007)

Research on the Quantifying and Calculating Model of the Software Component Reusability

Qi Wang

Department of Computer Science Technology, YunCheng University, YunCheng, China
wqjsj@126.com

Abstract. Reusability quality directly affects the software quality and application development. The metric of software component reusability has been a hot spot for the research of the CBSD (component-based on software development). The fault of using AHP to measure reusability is that the specific attribute value of reusability and its sub-property cannot be figured out and the purpose of measurement cannot be achieved. To solve this problem, in this paper, with the decomposition model of reusability attribute, a method for quantifying and calculating software component reusability is proposed. In this way, to a certain extent, the calculation of component reusability values would be more intuitive and understandability while given a important basics to improve software quality. And compared with the AHP(The Analytic Hierarchy Process), the experiments show the validity and rationality of this method.

Keywords: Software reuse, Software component, Reusability, Quality metrics, Evaluation model.

1 Introduction

Along with the continuous increase of the reused component number, in order to develop software, how to choose the component with the better reusability from the component library is an important problem for the managers of the component library and the persons of reusing components. The metrics and the evaluation of the component reusability are all involved[1]. And the metric of the reusability has been a hotspot in the research field of the CBCD.

The metrics of the reusability is used to evaluate the reusability and quality of the component[2]. Through the metric, the useful products or components can be distinguished from the Hereditary System, and the components with the high reusability and high quality can be controlled and ensured. Meanwhile, the meritorious information of the component quality metrics can be provided for the person to reuse the components. In this paper, the metrics of the component reusability is defined as the quantification calculation of the reusability.

The traditional AHP (Analytic Hierarchy Process) can not figure out the difference of reusability between components and the quantized value of every sub-attribute. Meanwhile, the AHP can not give the component reusability an effective metrics and evaluation. Aiming at the problem of the AHP, a method for quantifying and calculating software component reusability based on the Colony Decision is proposed

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 103–109.
springerlink.com

in this paper. The thought of this method is that firstly a decomposition model of the component reusability attributes is given; and then based on this model, the value of the reusability can be calculated by layer from the bottom to the top through this method which combines an improved AHP and the Colony Decision.

2 The Decomposition Model of the Reusability

The reusability is the essential attribute of the component. In fact, a satisfactory metrics and evaluation for the component reusability is relatively difficult, because the component reusability is affected by more factors from the external world. The quality elements of the reusability are a combination of extracting the quality character and sub-attribute of the component quality[3].

2.1 The Index System of the Reusability

According to the literatures[3,4,5], the software reusability is evaluated from the point of the component managing and the component assembling. And the quality feature of the component is described by five indexes (function, easily-used, reliability, maintaining, transplanting). Meanwhile, the data compatibility, the model standardization and the custom-made are elected as the sub-properties of the quality. So in this paper, these attributes and sub-attribute are all used to evaluate the reusability. And the sub-properties are described as Table 1.

Table 1. The Index System of the Quality's Sub-attribute among the Component's Reusability

Sub-properties	Description
Data compatibility	The data pattern is or isn't compatible with the international data pattern in common use
Model standardization	The component accords doesn't accord with one certain component model.
Custom-made	The embedding ability of the custom-made and allocation to support the component internal function
Help Document quality	Help system/training/demonstration The document integrality
Complexity of interface	The complexity degree of the component interface. For the component users and managers, component should has better capsulation. So the evaluation of the complexity only needs to consider the complexity degree of the component interface, and needn't to take the component internal complexity into account.
Reused number Code reuse Function reuse	The number of the components which has been reused The ratio of the code reuse The ratio of the function reuse
Easily-changed Easily-tested Self-described	The ability of being changed when the component is transplanted to a system. Whether Provide the self-testes or the text tool or not. The ration of the annotation lines
Software dependence Hardware dependence	The dependence on the special software or component when the component is transplanted to the software system. The Dependence on the special hardware
Component Scale	Bytes number, Code lines

2.2 The Decomposition Model of the reusability

A tree-type decomposition model of the component reusability is given as below in order to illustrate the method of calculating the component reusability. The tree-type structure of quantizing the component reusability is shown as figure 1.

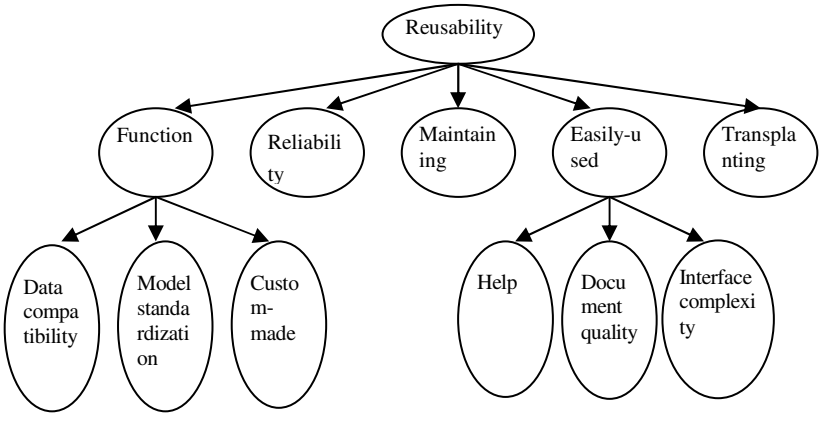


Fig. 1. Decomposition Model of Component's Reusability

Reusability = function, reliability, maintaining, easily-using, transplanting;
Function = data compatibility, mode standardization, custom-made;
Reliability = the reused number, the code reuse, the function reuse
Maintaining = easily-used, easily-tested, self-described;
Easily-used = help, document quality, interface complexity;
Transplanting = software dependence, hardware dependence, component scale.

3 The Method of Quantizing and Calculating the Reusability

The attribute value of the reusability is calculated from the leaf nodes. Firstly, the attribute values of the leaf nodes are obtained from the result set of the users' evaluation. Then, according to the judgment matrix provided by the experts, the composite weight coefficient of the node can be got. Finally, by the use of the RV arithmetic, the attribute value of the reusability can be calculated by layer from the bottom to the top.

3.1 The Attribute Value of the Leaf Nodes

In this paper, the result sets of the users' evaluation are divided into five grades: A excellent (1.0-0.8), B favorable (0.8-0.6), C moderate (0.6-0.4), D general (0.4-0.2), E bad (0.2). Wherein, the number in the brackets is the score of each question, and the full score of each question is 1.0. According to the user's answer to the question, the value of the leaf node can be obtained.

3.2 The Calculation of the Node Weight

In the decomposition model, the importance of the child node to the father node is different, because the components can be used in different field. So here, the weight value is adopted to show the importance of the child node to the father node. The node weight is calculated mainly through three steps. Firstly, calculate the node weight by layer from single expert. Secondly, calculate the weight of each expert (the reliability of the expert). Finally, combine the two weights to get the final weight of the whole expert group.

3.2.1 The Calculation of the Node Weight from Single Expert

The AHP is based on the hierarchy thought to form a model of a multilayered analytical structure, which decomposes the question into many different factors. By the use of the AHP, the system analysis is finally turned into the calculation of the weight of the relative importance for the bottom layer to the high layer or the sort of the relative quality. Thereby, it can help select the strategic scheme.

The AHP normally needs to check out the consistency of the judgment matrix. Here, we use an improved AHP which needn't to checkout the consistency. Thereby, it can calculate the weight more simply and more feasibly.

The calculation steps of the improved AHP are described as below.

Step 1 Construct the judgment matrix A'_k, and get the optimal consistent judgment matrix A_k.

Step 2 Construct the judgment matrix.

$$A_k = \{a_{ij}^k\} \text{ 且 } a_{ij}^k > 0, \quad a_{ji}^k = \frac{1}{a_{ij}^k}, \quad i, j=1,2,3,\cdots,n; k=1,2,\cdots,m$$

Step3 Calculate the characteristic vector corresponding to the max characteristic value $\lambda_{\max}^k$ of the matrix A_k, and then turn it to the scope from 0 to 1. Then the node weight w_i^k can be got(i represents the node, i=1,2,3, …; k represents the expert, k=1,2,3,…).

3.2.2 The Calculation of the Expert's Relative Weight by the Colony Decision[5]

Due to the difference of the each expert's knowledge background, the difference of the professional level and the different degree of knowing things, the true degree of the judgment matrix is different by the use of the AHP. Through the method from 3.2.1, the optimal judgment matrix A_k can be obtained, and then the corresponding ratio C_{RI}^K can be calculated.

The expert weight p_k can be calculated by using the following formula (1).

$$p_k = \frac{1}{1+gC_{RI}^K}, a>0, k=1,2.....m \quad \text{(k represents the number of the experts, k=1,2,3,…)} \tag{1}$$

Manage p_k to the scope from 0 to 1 and then get the expert weight coefficient p_k^*.

$$p_k^* = p_k \Big/ \sum_{k=1}^{m} p_k \tag{2}$$

Therein, the parameter g acts as a regulator, and in the practical use, the value of g is normally set 10.

3.2.3 The Calculation of the Composite Weight of the Node by the Colony Decision[6]

On the basis of the node weight w_i^k, the corresponding expert weight p_k^*, the composite weight can be calculated by the formula(3).

$$w_i = \sum_{k=1}^{m} w_i^k \bullet P_k^* \tag{3}$$

归一化处理 and the composite weight w_i^* can be obtained.

$$w_i^* = w_i \Big/ \sum_{i=1}^{n} w_i \tag{4}$$

3.3 The Attribute Value of the No-Leaf Nodes

In the decomposition model of the component reusability, the no-leaf nodes all have one child node or more, and the root node is one of the no-leaf nodes. The calculation of the reusability value must take the attribute value and the weight of the child nodes into count, because different child node affects the parent node differently. Through 3.1, the values of the leaf nodes can be got. Through 3.2, the weights of the all nodes can be obtained. So, to begin with the leaf nodes, the attribute values of the leaf nodes and the final value of the reusability can be calculated by layer from the bottom to the top.

Suppose that the child nodes of the component C's reusability are $i_1, i_2, \ldots, i_n$, and the corresponding node values and weights are respectively $V_c(i_1), V_c(i_2), \ldots, V_c(i_n)$ and w_1^*, w_2^*, … w_n^*. Then the attribute values of the no-leaf nodes can be calculated by the use of the following algorithm.

Input: The child nodes of the component C's reusability i_1, i_2, …, i_n, the corresponding node values $V_c(i_1), V_c(i_2), \ldots, V_c(i_n)$, and the corresponding weights w_1^*, w_2^*, … w_n^*.

Output: The value of the component C's reusability RV_c.

(1) Get the max weight w_{best} among the child nodes of the reusability.

w_{best} =Max（ w_1^* , w_2^* , … w_n^* ）

(2) Calculate the ratios of the nodes' weights i_1, i_2, …, i_n towards w_{best}.

$M_{w_1} = w_1 / w_{best}$, $M_{w_2} = w_2 / w_{best}$,... $M_{w_n} = w_n / w_{best}$

(3) X=Min（$V_c(i_1) * M_{w_1}$, $V_c(i_2) * M_{w_2}$, ..., $V_c(i_n) * M_{w_n}$ ）

(4) Y= $w_1^* * V_c(i_1) + w_2^* * V_c(i_2) + \cdots + w_n^* * V_c(i_n)$

(5) $RV_c = XY^{1/2}$.

4 The Calculation Result and Analysis

In this paper, three components of report and print are elected from one certain component library to test and verify by experiments. The three components all have the functions of report and print, but the user chooses a component with better reusability very difficultly. Here, the component reusability is evaluated separately by the use of the traditional AHP and the method proposed in this paper.

Firstly estimate the component reusability by using AHP, and then elect the perfect component according to the composite weight. Secondly quantize and calculate the component reusability through the method proposed in this paper, and then select the perfect component according to the value of the component reusability. The comparison of the estimate results of AHP and the calculation results of the reusable values is shown as figure 2.

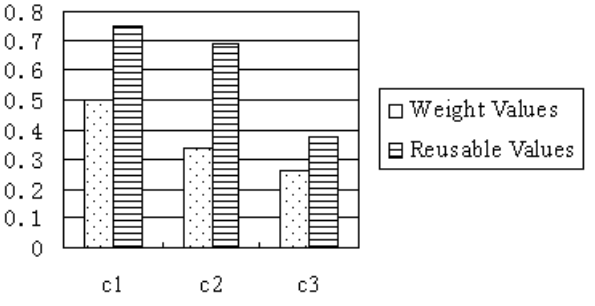


Fig. 2. The Composition of the Weight Sequencing's Results and the Reusable Values

That illustrates this method is reasonable and effective. Compared with the metrics result of the AHP, this method not only can select the perfect reusable component, but also can calculate the values of all attributes. This method makes users know exactly in the heart, and makes the metrics result more specific and clear. Meanwhile, the metrics result of the sub-attributes can provide an important reference to improve the component quality.

5 Conclusions

Aiming at the shortcomings of the traditional AHP, a metrics and calculation model of the component reusability is proposed based on the Colony Decision. The experiments

show that this method is effective and to the max limit, reduces the subjectivity of the calculation result by the Colony Decision. Meanwhile, this method can make the results more reliable, truer and more specific. But in the decomposition model, the rationality of the norm system still needs to be more perfect and the assignment of the leaf nodes needs the support of the automate tools. These are all the new research contents in future.

Acknowledgement. This work has been supported by the R@D Project NO. 20091150 of Science and Technology in the institution of higher learning of Shanxi Province, the Natural Science Funds Project No. 2009011022-1 of Shanxi Province and the Project of Yuncheng University No. JC-2009015.

References

1. Yang, C., Yang, H., Jin, M., Gao, Z.: Software component quality metrics. Computer Engineering and Design 27(3) (2006)
2. Mei, H., Xie, T., Yuan, W., Yang, F.: Component Metrics in Jade Bird Component Library System. Journal of Software 11(5), 634 (2000)
3. Wang, S.: Research on evaluation model of measurement for software component reusability. Computer Engineering and Design 29(10) (2008)
4. Bansiya, J., Davis, C.: A hierarchial model for object-oriented design quality assessment. IEEE Trans. on Software Engineering 28(1), 4–17 (2002)
5. Wu, D., Li, D.: Shortcomings of Analytical Hierarchy Process and the Path to Improve the Method. Journal of Beijing Normal University (Natural Science) (2), 264–267 (2004)
6. Zhou, G., Huang, D., Zeng, Q.: Quantifying and Calculating Security of Software Component Based on Decomposition of Security Attributes. Microelectronics & Computer 25(10) (2008)
7. Li, X., Ding, J., Zhang, Y.: Research on Modeling Methods for Software Metrics Data. Application Research of Computers (2005)

The Application of Sequential Indicator Simulation and Sequential Gaussian Simulation in Modeling a Case in Jilin Oilfield

Xiaoyu Yu[1] and Xue Li[2]

[1] Exploration and Development Institute
Jilin oilfield Company
Jilin, China
Yuxy_jl@petrochina.com.cn
[2] College of Marine Geosciences
Ocean University of China, OUC
Qingdao, China
lix010@126.com

Abstract. Conditional simulation approach is the major trend and reseach focus of geostatistic development. In this paper, simulation result, statistical parameters and variogram is analysized by applying sequential indicator simulation to reservoir modeling, besides, the comparison between kriging and sequential simulation method has been made. The result proves the characteristic of conditional simulation and enriches the theory. This paper discusses indicator variation function and indicator model, the smoothing effect of kriging estimation can be seen from simulation results. Finally, the sequential conditional method has been applied to reservoir simulation modeling and theoretical basis and example provides a wide range of choices to describe the real geological conditions.

Keywords: sequential indicator simulation, reservior modeling, variogram, kriging estimation.

1 Introduction

Conditional simulation technology is a new geologistic tool that develops rapidly after the kriging geostatistic estimation technique.The kriging estimation is an optimal unbiased estimation method, an estimate on the average as well, and it only gives a single numerical model. Relatively speaking, conditional simulation can achieve more than conditional simulation. The differences between these results reflect the uncertainty of geological variables' spatial distribution. In recent years, conditional simulation is widely applied in stochastic reservoir modeling. It often results in reservoir's uncertainty and multiple solutions because of inadequate survey data. Conditional simulation not only can achieve the structure related to reservoir properties's spatial distribution, but also conditioning the known well data and get more results to satisfy the description and analysis of reservoir properties. Therefore, conditional simulation is widely applied in geological reservoir modeling and prediction in current.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 111–118.
springerlink.com

2 Differences between Conditional Simulation and Kriging

Compared with kriging, conditional simulation methods places much attention on reflecting volatility of the spatial data, while the Kriging method focuses on minimizing the estimation error. The differences are detailed in three aspects:

- The Kriging method is a kind of interpolation method, it only takes the accuracy of local estimation into consideration, including optimal assessment (minimum estimated variance) and unbiased estimates (the mean estimate value and the observation point), rather than specifically consider the ratio of the estimated spatial correlation value. Although the conditional simulation uses kriging algorithm to estimate point, it just estimates the points distribution that is to be estimated, it firstly consider the overall nature simulation results and correlation of statistical spatial value, followed by the accuracy of local estimation;
- Kriging method has a certain smoothing function, if the observation data is discrete data, the discreteness will be reduced by the smoothing estimation results with interpolation, ignoring subtle changes between wells: the conditional simulation makes the estimation results more true than interpolation by means of the "random noise" which is used in interpolation model system. "Random noise" is subtle changes between wells, although the value of each simulation for the local point is not entirely true, the estimation variance may be even greater than the interpolation method, the simulation curves can better manifest the real curves' fluctuations;
- Kriging method can only get a definitive result by interpolation, while the conditional simulation may achieve more results that is to say it can produces a number of optional model, the difference between each model is the reflection of the spatial uncertainty.

3 Advantages of Sequential Indicator Simulation

Sequential indicator simulation not only it can be used to simulate the continuous variables, but also to simulate the discrete characteristics with the fundamental principles of Kriging indicator, the indicator variables is determined by average frequency and variation function which characterizes its spatial continuity. This method speeds slower than the Sequential Gaussian Simulation, but its advantages is that it can simulate the complex heterogeneity, such as different relevance, anisotropy and different lithofacies, etc.. Sequential Gaussian Simulation is mainly used to simulate property parameters such as porosity and permeability.

4 Data Analysis and Variation Function

Data analysis is a very important step in property modeling:plays a key role in explaining the data, identifing geological features, the results can be directly used in property modeling module. Data analysis is divided into discrete and continuous data analysis. It is generally applied to discrete data. Discrete data is usually integer, the

object of analysis can be well log or discrete well points the whole model. As to facies modeling, for example, it can analyze discrete attributes and relativity of continuous variables' attributes.

Variation function which describes the spatial variation of the reservoir, takes a dominant action in the data analysis, the variation function produced from input data can be used for property modeling. Variogram should first set the main direction of the parameters such as bandwidth, search radius, step, etc., and then set the second direction and vertical parameters. Range, among all the various parameters, is particularly important. In the range area, the smaller the distance is between two points, the better the similarity will be. As to other points, the similarity is irrelevant to distance.

5 Specific Case

5.1 Data Preparation

The hierarchical data used in this modeling is the comparison data of sand group and sand body of each layer. Hierarchical data can be divided into different sand group and small layer, using sedimentary cycles contrast and log analysis method, each group includs a number of layers which contains sand body and mud. The following picture(Fig.1) shows the well section diagram of three wells' previous seven layers of qn3-I.

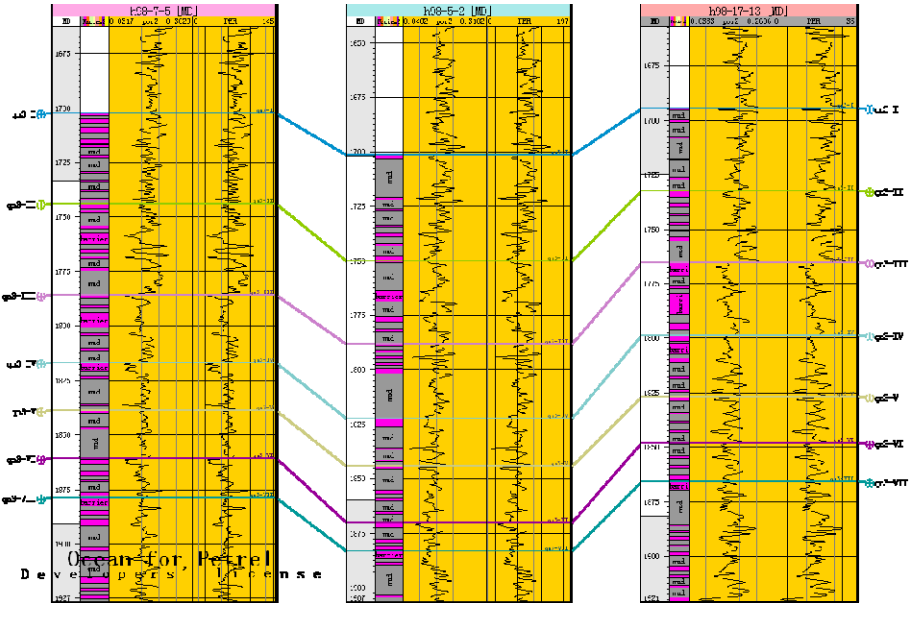


Fig. 1. well section diagram of part stratum(h98-7-5,h98-5-2 and h98-17-13)

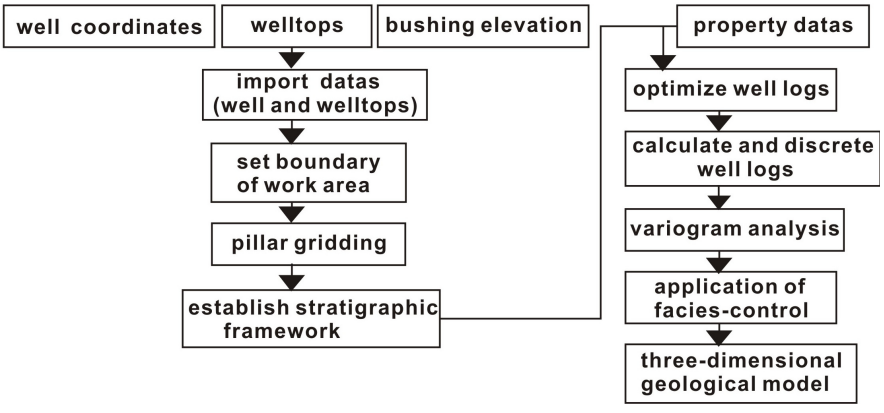


Fig. 2. Working flow chart used in simulation

Sequential indicator simulation applied in Daqingzi oilfield is based on the preliminary work as follows. Fig.2 is the working flow chart previous the facies and property modeling.

With the discrete property data, we calculate average well point data. The following contour map (including porosity and permeability contour map) is an expression of the average datas, take qn3-I sand group for example.

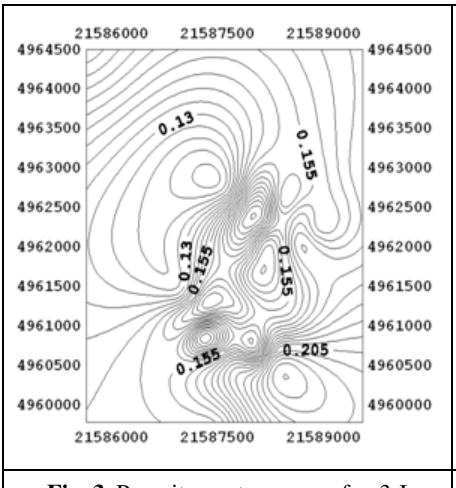


Fig. 3. Porosity contour map of qn3-I

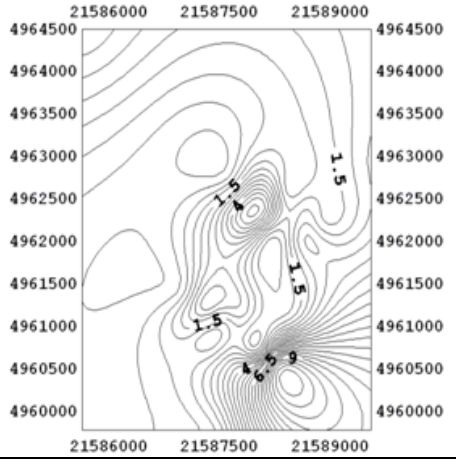


Fig. 4. Permeability contour map of qn3-I

6 Facies and Property Model with the Use of Sequential Simulation

In order to study the sequential simulation, one of the parameters for the simulation is facies, and the other is porosity and permeability. Sedimentary facies are discrete

parameters, using sequential indicator simulation method for simulation. Porosity and permeability parameters are continuous, using sequential gaussian simulation. The following graphs (Fig.5-Fig.7) are the facies and property model simulated by sequential simulation.

Fig.8-Fig.10 is the 3D fence map of the entire study area.

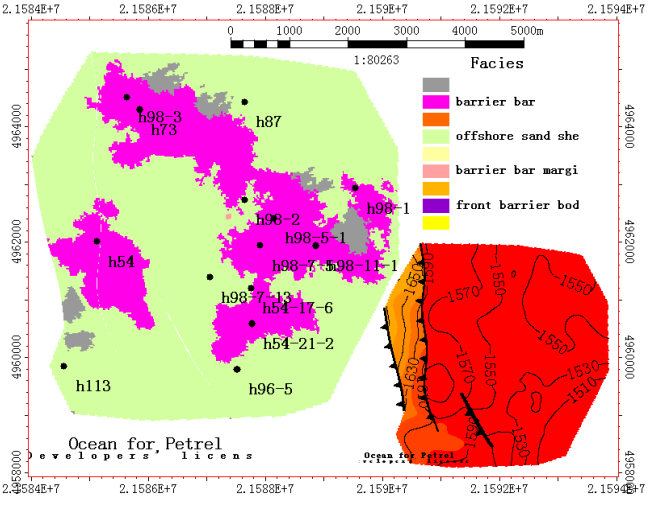


Fig. 5. Facies model of qn3-I

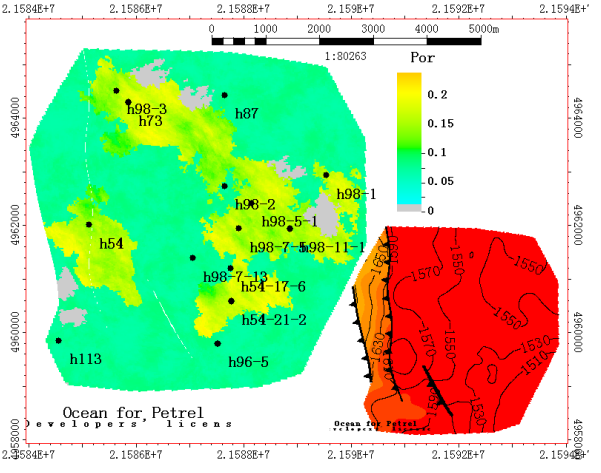


Fig. 6. Porosity model of qn3-I

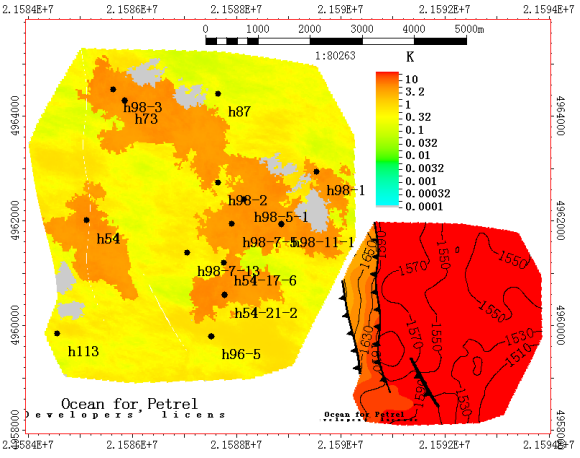


Fig. 7. Permeability model of qn3-I

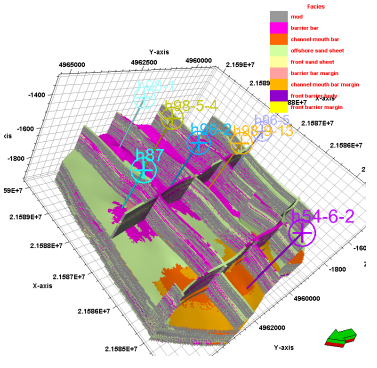


Fig. 8. 3D Fence map of facies model

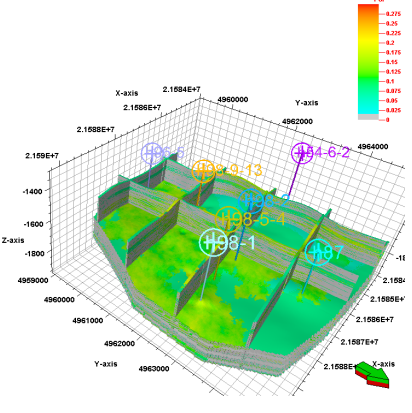


Fig. 9. 3D fence map of porosity model

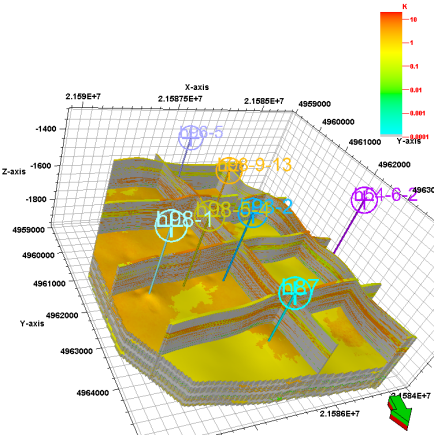


Fig. 10. 3D fence map of permeability model

7 Conclusion

a) It makes the smoothing function of Kriging methods more superior by the comparison of the results achieved by applying Sequential Indicator Simulation and Kriging methods for specific reservoir modeling.

b) Variance function is the groundwork of Kriging estimation and simulation, the variance function obtained by choosing different cut-off value is different, so we have to integrate all the data and select reasonable cut-off value.

c) More results can be achieved with the application of Sequential Indicator Simulation in reservoir modeling and its applicability for reservoir modeling can be best elaborated as well.

Acknowledgment. The authors thank the workers of Jilin oilfiled for supplying researching data and assistance.

References

1. Zhang, J., Jinkai, W.: Application of the 3D Computer Modeling and Numerical Simulation in the Lower-Middle Series of Strata of Hetaoyuan Formation in Xiaermen Oilfield. In: 2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing, pp. 201–205 (2008)
2. Zhang, J., Ma, X., Wang, J.: 3D Geological Modeling of Complex Reservoir Geometries in a Lacustrine-delta System, Sha19 Block of Shanian Oilfield. Sencond ISECS International Colloquiim on Computing, Communication, Control, and Management 4, 455–458 (2009)
3. Journel, A.G.: Stochastic modeling of a fluvial reservoir:a comparative review of algorithms. Journal of Petroleum Seience and Engineering 21, 95–121 (1998)

4. Zhang, J., Ma, X., Zhang, Z., Zhang, X.: The Importance of User Interaction In 3D Geological Modeling and Reservoir Exploitation. In: Proceeding of First International Conference on Modelling and Simulation, pp. 290–293 (2008)
5. Journel, A.G., Alabert, F.: New Method for reservoir mapping. Petroleum Technology 42(2), 212–218
6. Haldorsen, H., Damsleth, R.: Stochastic modeling. JPT 42(4), 404–412 (1990)

Research on the Multi-classifier Fusion Model Based on Choquet Integral

Tie-song Li[1], Cheng Jin[2], Yong-hua Cai[1], and Yong Mi[3]

[1] Department of Mathematics and Computer Science
Hebei Normal University for Nationalities
Chengde, China
litiesong@126.com
[2] Department of Physics
Baoding College
Baoding, China
[3] Qualification & Measurement Management Center
Chengde Iron & Steel Co.
Chengde, China
jincheng1973@126.com, cyhcyh_sjz@126.com, cdmiyong@163.com

Abstract. In order to improve the performance of the classifiers, multi-classifier fusion is a good strategy. The multi-classifier fusion models based on Choquet integral have shown successful applications in practice since the adopted fuzzy measure in the integral can effectively describe not only the importance degree of the classifiers respectively but also the interaction among the classifiers. Considering many real problems, allowing fuzzy measure to assume interval or fuzzy numbers seems more reasonable. In this paper, we define a generalized Choquet integral in which an interval-valued fuzzy measure is adopted and its value is a crisp real number. Furthermore, the defined Choquet integral is applied in the multi-classifier fusion. PSO is used to determine the values of the interval-valued fuzzy measure. Experimental results demonstrate the rationality of the definition and the feasibility and effectiveness of the application in the multi-classifier fusion.

Keywords: Fuzzy measure, Choquet integral, Interval-valued Multi-classifier fusion.

1 Introduction

Multiple classifiers in different application areas have different terminologies, such as multiple classifier combination in classifier systems, classifier fusion, mixture of experts, neural networks committee, classifier combination, etc. [3]. Classifiers (classifier selection) and classifier fusion (classifier fusion) is a multi-classifier combination of the two main strategies [4]. Device selection in the classification, each classifier only containing focal of features Kongjian information, only the right feature space of Duiyingbufen bear Zhuanjiazeren, Duidaishibie sample Shouxian select a one or a few classifier Duiqi classified then the output of the classifier for

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 119–126.
springerlink.com

processing, the final result of classification, the classifiers are complementary relationship between, such as dynamic selection classifier method [5-6], classifier construction and the grouping and so on. The integration of the classifier, each classifier are seen as the feature space on the experts, contains information about the characteristics of the information space, separation between the segments is competition. Applications, are treated each classifier to classify the sample identification, and then by the fusion operator to the output of each classifier to integrate the results obtained the final classification. The output of the classifier fusion methods are many, often with majority voting [7], Borda count method [3,4,8], Dempster-Shaffer theory of evidence [9], Bayesian rules and all fusion operator, such as the weighted average operator, the value operator, to take big / take small operator, multiplication operator, OWA operator, fuzzy integral [10-15] and so on. If the output of each classifier output for the hardware, which samples the output of the class is marked to be identified, the majority voting method, used in the fusion; if the classifier output for the sort class standard, commonly used for the Borda count fusion; if classifier output for the soft output, that output is to type the number for the dimension of non-negative real vectors, fusion operator can choose more, such as the weighted average operator, to take big / small operators to take, in value operator, multiplication operator, OWA operator, fuzzy integral, Bayesian methods and so on.

Choquet integral, as the development of Lebesgue integral, has been widely applied in practice. In its application to Multi-classifier fusion, the importance level of every classifier and the interactive effects among different classifiers can be described by Fuzzy measure. Three interactions are considered: (1) redundancy and passive coordination, that is, the synthetic evaluation capacity of two classifiers is no bigger than the sum of their respective evaluation capacity. (2) Complementation and active coordination, that is, the synthetic evaluation capacity of two classifiers is bigger than the sum of their respective evaluation capacity. (3) Independency, the intermediate between the previous two situations. When the synthetic evaluation capacity of two classifiers is equal to the sum of their respective evaluation capacity, they are independent. Fuzzy measure is a non-negative and non-additive set function, and the nonadditivity can right describe the interactions between two classifiers. Superadditivity means the synthetic evaluation capacity of two classifiers is bigger than the sum of their respective evaluation capacity, namely, two classifiers are actively coordinated, while subadditivity refers to the passive coordination when the synthetic evaluation capacity of two classifiers is smaller than the sum of their evaluation capacities. Additivity means no interactions exist between two classifiers.

It has become a research focus that fuzzy integral is used to combine multiple classifiers in order to obtain higher classification accuracy. Fuzzy measure plays an important role in the fusion system of Multi-classifier, and impacts its function to a large extent. The definition of fuzzy measure directly influences the functions of classification accuracy, robustness, etc. Therefore, fuzzy measure should be studied before the calculation of fuzzy integral. Two methods are involved. (1) Preliminary fuzzy measure is set by realm experts based on their prior knowledge. This method is relatively subject and has low classification accuracy to some databases. (2) Fuzzy measure is set through the study of dataset. This method is relatively objective, which makes higher classification accuracy. There are many ways to set Fuzzy measure

through data study, such as genetic algorithm, neural network, optimization problem and heuristic search.

2 Extension of Choquet Integral

Previous studies in Multi-classifier fusion process used classical Choquet integral, in which fuzzy measure was real-valued, but interval values or fuzzy values are more rational in solving practical problems. So Choquet integral should be extended. The integration result should be real-valued since it is the result of classification. Therefore, in our definition of Choquet integral, integrand is a real value, the classification result of classifiers, fuzzy measure is an interval value, and integration result is real-valued.

$N_{[0,1]}$ represents a set of all closed intervals in $[0,1]$.

Definition 1: Let $(X, P(X))$ be a measurable interval. The interval valued function $\overline{\mu}: P(X) \to N_R$ is called an interval-valued fuzzy measure. Iff the following conditions hold:

(1) $\overline{\mu}(\varnothing) = [0, 0]$;

(2) if $A, B \subset P(X)$, $A \subset B$, then $\overline{\mu}(A) \leq \overline{\mu}(B)$;

It should be explained that the monotone property should not be excluded; otherwise the condition of fusion operator can not be satisfied[1]. Therefore, the monotone property of interval-valued fuzzy measure needs defining in the extension of Choquet integral, and the condition of fusion operator satisfied in the extended Choquet integral. The definition of monotone property of interval-valued fuzzy measure is closed related to the extension of Choquet integral. The following part will define the extended Choquet integral, and then the monotone property of interval-valued fuzzy measure.

The definition only covers finite set, because the domains of relevant practical problems are finite set.

Let $\overline{\mu}(A) = [\mu_l(A), \mu_r(A)]$, $\mu_l(A)$, $\mu_r(A) \in [0,1]$.

Definition 2: let $\overline{\mu}$ be an interval-valued fuzzy measure in the defined interval of $(X, P(X))$, the function $f: X \to [0,1]$ is measurable, and then the Choquet Integral of f with respect to $\overline{\mu}$ is defined as

$$(c)\int f d\overline{\mu} = \sum_{i=1}^{n} (f(x_i) - f(x_{i-1})) \frac{\mu_l(A_i) + \mu_r(A_i)}{2}$$

Without loss of generality, suppose $f(x_1) \leq f(x_2) \leq \cdots \leq f(x_n) \leq 1$ (Otherwise, the elements in X should be rearranged as $\{x_1^*, x_2^*, \cdots, x_n^*\}$ to make the corresponding function value $f(x_i^*)$ satisfy this relation), $A_i = \{x_i, x_{i+1}, \cdots, x_n\}$, $f(x_0) = 0$.

Definition 3: $\bar{\mu}(A) \le \bar{\mu}(B)$, iff $\frac{\mu_l(A)+\mu_r(A)}{2} \le \frac{\mu_l(B)+\mu_r(B)}{2}$.

Example 1: Let $X = \{x_1, x_2, x_3\}$, and f is a real-valued function on the domain of X:

$$f(x) = \begin{cases} 0.1 & if \quad x = x_1 \\ 0.3 & if \quad x = x_2 \\ 0.5 & if \quad x = x_3 \end{cases}$$

$\bar{\mu}$ is a fuzzy measure on interval values of domain $P(X)$, and the value selection is shown in table 1:

Then the Choquet Integral of f with respect to $\bar{\mu}$ is

$$\begin{aligned} \int f d\bar{\mu} &= 0.1 \times \bar{\mu}(\{x_1, x_2, x_3\}) \\ &+ (0.3-0.1)) \times \bar{\mu}(\{x_2, x_3\}) + (0.5-0.3) \times \bar{\mu}(\{x_3\}) \\ &= 0.1 \times \frac{1+1}{2} + 0.2 \times \frac{0.6+0.8}{2} + 0.2 \times \frac{0.4+0.6}{2} \\ &= 0.34 \end{aligned}$$

3 Determination of the Interval-Valued Fuzzy Measure in Fusion Process

Particle swarm optimization is used to optimally determine the value of interval-valued fuzzy measure of extended Choquet Integral. The corresponding algorithm is as follows:

Step 1: Randomly initialize particle swarm, particle location and speed. Every particle location represents an interval-valued fuzzy measure denoted as $\bar{\mu}_i = \{\bar{\mu}_{ij}\}$; particle speed indicates the variable quantity in present iteration, which is denoted as $\bar{v}_i = \{\bar{v}_{ij}\}$. Thereinto, $\bar{\mu}_{ij} \in N_{[0,1]}, \bar{v}_{ij} \in N_{[-1,1]}, i = 1, 2, \cdots, m, j = 1, 2, \cdots, 2^n - 2$, with m being particle number and n property number. And then initialize $pbest_i$ and $gbest$ with 0, $pbest_i$ and $gbest$ indicating respectively the best places that till the present iteration every particle and the whole particle swamp have ever been to.

Step 2: Use monotonicity adjustment strategy to ensure that every $\bar{\mu}_i$ meets constraint monotonicity conditions, calculate the fitness value of every particle, and define fitness function as classification accuracy. And then update $pbest_i$ and $gbest$. If fitness value is bigger than the given standard or the biggest iteration time is bigger than the biggest given iteration time, then follow Step 5; otherwise, Step 3.

Step 3 : The following formula is to update particle speed:

$$\overline{v}_{ij}(t+1) = w_i \cdot \overline{v}_{ij}(t) + c_1 \cdot rand() \cdot (pbest_{ij}(t) - \overline{\mu}_{ij}(t)) + c_2 \cdot rand() \cdot (gbest_j(t) - \overline{\mu}_{ij}(t))$$

w_i shall be calculated as follows:

$$w_i(t) = \left\{ \frac{(t_{\max} - t)}{(t_{\max})} \right\} \cdot (w_{initial} - w_{final}) + w_{final}$$

There into, t is the t iteration, $c_1 = c_2 = 2$, $rand()$ stands for random numbers evenly disposed in [0,1], $t_{\max}$ is the biggest given iteration time, $w_{initial}$ the initial weight, and w_{final} the suspension weight. If the updated speed $v_{ij} > v_{max}$, then $v_{ij} = v_{max}$; if $v_{ij} < -v_{max}$, then $v_{ij} = -v_{max}$ with $v_{max} = 1$.

Step 4: Particle location shall be calculated as follows:

$$\overline{\mu}_{ij}(t+1) = \overline{\mu}_{ij}(t) + \overline{v}_{ij}(t+1)$$

There into, t has the same meaning as it is in Step 3. If particle place $\overline{\mu}_{ij} > 1$ or $\overline{\mu}_{ij} < 0$, then $\overline{\mu}_{ij} = 1$, $\overline{\mu}_{ij} = 0$. And then follow Step 2.

Step 5 : Output $gbest$ and the corresponding classification accuracy.

4 Experiment

Experiments are carried out in Iris Plants Database、Glass Identification Database and Pima India diabetes Database in UCI[2] databases to verify the feasibility and validity of Choquet integral in multiple classifier fusion system. Iris Plants, a classification problem database of 3 classes, has 150 samples, with each one containing 4 characteristic attributes and 1 categorical attribute. Glass Identification Plants, a classification problem database of 7 classes, has 214 samples, with each one containing 9 characteristic attributes and 1 categorical attribute. 30% of the experiment data is used as testing data, and the other 70% as training data. Pima India diabetes, a classification problem database of 2 classes, has 532 samples, with each one containing 8 characteristic attributes and 1 categorical attribute.

The steps are as follows:

Divide the data into training data set and testing data set.

Train three neural networks classifiers n1, n2 and n3 on the training data set. Different characteristic attributes and training samples are used to reduce the similarity of neural networks.

Test the trained neural networks on the training data set and count the result.

Optimally determine by PSO the value of interval-valued fuzzy measure of extended Choquet Integral.

Test the value on the testing data set.

Analyze the fusion result.

We mainly take classification accuracy as the measurement of classification performance index and evaluate the advantages and disadvantages of the methods of Choquet integral by comparing correction rates and error rates of the models on data sets.

The results of the experiments are as follows:

1. Select randomly 70% of the samples in Iris database as training sample set, and the remaining 30% as testing sample set. There are three neural networks classifiers n1, n2 and n3 after training. The results of the three experiments on testing set can be seen in Table 2, in which the last line is the results that the extended Choquet integral we define fuses separate classifiers.

2. The experiment results on Glass database can be seen in Table 3, in which the last line is the results that the extended Choquet integral we define fuses separate classifiers.

3. Table 4 is the experiment result on Pima database, with the last line showing the results that the extended Choquet integral we define fuses separate classifiers.

Tables 2, 3 and 4 indicate that the results that the extended Choquet integral we define fuses separate classifiers are better than any single classifier, and the performance of classified system is improved. Besides, the constringency speed is faster and the running time is reduced when PSO is adopted to determine the interval-valued fuzzy measure.

Table 1. Interval-valued Fuzzy measures

$\bar{\mu}(\varnothing)$	$[0,0]$	$\bar{\mu}(\{x_2\})$	$[0.1,0.5]$	$\bar{\mu}(\{x_1,x_2\})$	$[0.3,0.5]$	$\bar{\mu}(\{x_2,x_3\})$	$[0.6,0.8]$
$\bar{\mu}(\{x_1\})$	$[0.2,0.4]$	$\bar{\mu}(\{x_3\})$	$[0.4,0.6]$	$\bar{\mu}(\{x_1,x_3\})$	$[0.5,0.7]$	$\bar{\mu}(\{x_1,x_2,x_3\})$	$[1,1]$

Table 2. Experiment results on Iris database

	N1	N2	N3	C
Classification Accuracy 1	0.8333	0.9014	0.9333	0.9437
Classification Accuracy 2	0.9667	0.8	0.8204	0.9688
Classification Accuracy 3	0.9211	0.9400	0.9333	0.9503
Average Accuracy	0.9070	0.8805	0.8957	0.9608

Table 3. Experiment results on Glass database

	N1	N2	N3	C
Classification Accuracy 1	0.6313	0.6977	0.6537	0.7203
Classification Accuracy 2	0.5897	0.6279	0.6423	0.6987
Classification Accuracy 3	0.68	0.5608	0.5960	0.7062
Average Accuracy	0.6337	0.6288	0.6307	0.7084

Table 4. Experiment results on Pima database

	N1	N2	N3	C
Classification Accuracy 1	0.7069	0.6783	0.7451	0.7509
Classification Accuracy 2	0.6500	0.7360	0.6876	0.7430
Classification Accuracy 3	0.7230	0.6900	0.7362	0.7680
Average Accuracy	0.6933	0.7014	0.7230	0.7539

The following preliminary conclusions can be drawn from the experiments:

(1) It is feasible and effective to adopt the definition of extended Choquet integralin in multi-classifier fusion, which can ensure higher classification accuracy than that of single classifier.

(2) There are some advantages such as parameters set, faster constringency speed and feasibility, when PSO is adopted to determine the interval-valued fuzzy measure.

The search process of swarm evolution optimization is random, so the search result has some uncertainty. For example, there may be multiple solutions and we have to select those which can ensure the least fitness, which may result in over-fitting. Aiming at this problem a certain error rate can be granted in training process.

In addition, the fuzzy measure is defined on the power set of classifier assemblage, and the parameter of interval-valued fuzzy measure doubles that of ordinary fuzzy measure. As a result, the more classifiers are involved, the more parameters it will produce, which leads to difficulties in searching process. Particle swarm can be improved to solve this problem.

References

1. Feng, H.-M.: A Study on the Fuzzy Measure for Multi-classifier Fusion. Hebei: Hebei University, Master degree thesis (2005)
2. UCI Repository of Machine Learning Databases and Domain Theories, ftp://ftp.ics.uci.edu/pub/machine-learning-databases
3. Kuncheva, L.I., Bezdek, J.C., Duin, R.P.W.: Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recognition 34, 299–314 (2001)
4. Ruta, D., Gabrys, B.: An Overview of Classifier Fusion Methods. Computing and Information Systems 7, 1–10 (2000)
5. Kuncheva, L.I., Bezdek, J.C., Duin, R.P.W.: Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recognition 34, 299–314 (2001)
6. Ruta, D., Gabrys, B.: An Overview of Classifier Fusion Methods. Computing and Information Systems 7, 1–10 (2000)
7. Battiti, R., Colla, M.: Democracy in neural nets: Voting schemes for classification. Neural Networks 7(4), 691–707 (1994)
8. Ho, T.K., Hull, J.J., Srihari, S.N.: Decision combination in multiple classifier systems. IEEE Trans. Pattern Analysis and Machine Intelligence 16(1), 66–75 (1994)
9. Rogova, G.: Combining the results of several neural network classifiers. Neural Networks 7(5), 777–781 (1994)
10. Cho, S.-B., Kim, J.H.: Combining Multiple Neural Networks by Fuzzy Integral for Robust Classification. IEEE Trans. Syst. Man Cybern. 25(2), 380–384 (1995)
11. Cho, S.-B., Kim, J.H.: Multiple Network Fusion using Fuzzy Logic. IEEE Trans. Neural Networks 6(2), 497–501 (1995)
12. Marichal, J.-L.: On Sugeno Integral as An Aggregation Function. Fuzzy Sets and Systems 114, 347–365 (2000)
13. Grabisch, M.: Fuzzy Integral for Classification and Feature Extraction, Fuzzy measures and integrals: Theory and Applications, pp. 415–434. Physical-Verlag (2000)
14. Auephanwiriyakul, S., Keller, J.M., Gader, P.D.: Generalized Choquet Fuzzy Integral Fusion. Information Fusion 3, 69–85 (2002)
15. Labreuchea, C., Grabisch, M.: The Choquet integral for the aggregation of interval scales in multicriteria decision making. Fuzzy Sets and Systems 137, 11–26 (2003)

Application of Cashmere and Wool Drawing Machine by Using PLC

Yunhui Yang[1] and Yiping Ji[2]

[1] School of Computer Science and Software Engineering, Tianjin Polytechnic University, Tianjin, China
[2] School of Textiles, Tianjin Polytechnic University Tianjin, China
yang.yh@163.com

Abstract. The application of control system with PLC on cashmere and wool drawing machine is introduced. With the adoption of PLC and frequency converter, the motor realizes continuously variable. Through practical operation of this system, it shows that the control system can achieve the desired requirements.

Keywords: Cashmere and wool drawing machine, PLC, frequency converter, control system.

1 Introduction

At present, high level garment fabrics are lightweight and soften by using fine yarn and animal hair. Because spinning yarn number of cashmere wool was limited by the number of fibers from yarn cross section, so the cashmere wool fiber fineness become the critical factors of light and thin fabrics. But the number of high count fine cashmere wool is far less sufficient to meet the needs of market. Hence, people start to find new ways to make the cashmere and wool fiber fine. Cashmere and wool drawing technique is one of main methods to realize it.

Cashmere and wool drawing machine can accomplish the drawing of cashmere and wool. It is consisted of wool top pretreatment, cashmere and wool drawing device, setting device and dryer as shown in Fig.1. The four units are drived by one 1.5kW AC asynchronous motor (M4) and three 2.2kW AC asynchronous motor (M1-M3). M1 is used to control feeding of wool tops pretreatment and cashmere and wool drawing. Setting and dryer device is controlled by M2 and M3.M4 gives the wool tops false twisting. The system requires the four units can achieve synchronous operation.

This drive system is multi-point drive. Because the inevitable variability on the structural and characteristics of motor, practical speed of each motor can not achieve full accord even using the same brands and models. So in practical work, the load of electrical motor is different in multi-point drive system. Generally speaking, the load of fast motor is larger and slow is less. The unbalance in the output power of this kind of motor can affect normal playing of production capacity. Severe causes motor

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 127–131.
springerlink.com

overload, burn down etc. Especially for belt transmission mechanism, it can cause the belt thermal agitation, even can lead to other mechanical system failure.

So, the system must ensure the output power of motor maintain the same in the multi-point drive. In other words, the practical speed of each motor can be adjusted in real time. This can resolve a series of problem by mismatching of motor characteristic and speed-reducer drive.

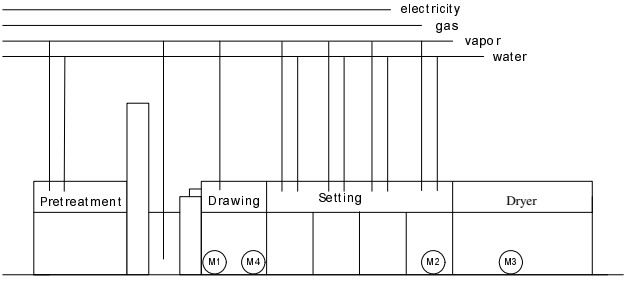


Fig. 1. Schematic diagram of cashmere and wool drawing machine

2 Application of Frequency Converter

In the past, the textile machinery always used gear reducer or cam mechanism to achieve different purpose of driving. Different axis has different rotational speed, and was related each other. So the mechanical structure was very complex. Furthermore the noise was big, the power was high and the abrasion affected controlled resolution in the driving process [1].

At present, frequency converter has superior properties of speed regulation and braking. Because it is high-efficiency, power frugal and widely used, it is deemed to the most promising speed control method.

As a high-tech product, frequency converter has two features of energy saving and environmental protection. So it gradually replaces traditional speed control method and is of special importance in the driving system[2].

In recent years, with the level of mechantronics in the textile machinery improved continuously, AC frequency converter becomes more and more popular. The basic structure of AC frequency converter is as shown in Fig.2.

AC frequency converter changes the constant voltage and frequency AC power to adjustable voltage and frequency. It can drive the motor smooth adjustable speed and achieves soft start and automation.

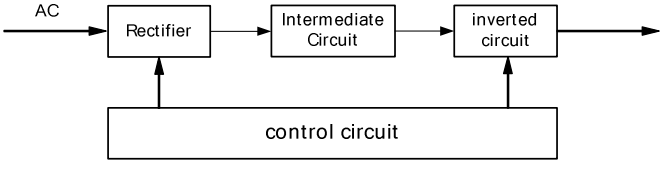


Fig. 2. Basic structure of AC frequency converter

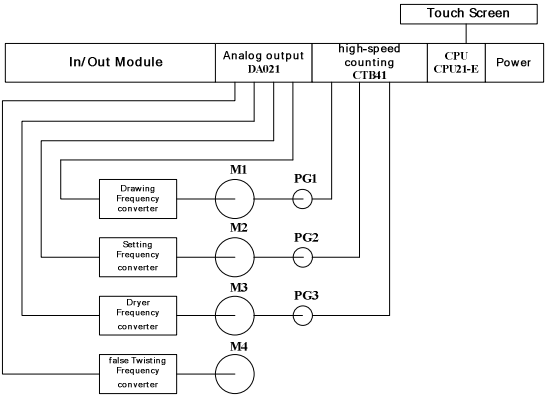


Fig. 3. Function charts of synchronous control system

Fig. 3 shows that the cashmere and wool drawing machine use four Emerson CT EV3000 series frequency converters to control the 4 motors.

The system need auto adjust load distribution to effectively use and protect the device. So it adopts master / slave structure. The motor of drawing device is master machine, others are slave.

Multi-motor synchronous transfer control is the key problem of cashmere and wool drawing machine[3]. Among the four motors, it has coupling effect each other and its working state can be affected each other. So the system adopts control methods based on compensation principle: the speed output of previous motor is the base of next motor speed. The difference value will be added in the input of master motor through compensator. This method can greatly improve the synchro control accuracy and anti-disturbance ability[4].

3 Design of Control System

When automatic control system using frequence converter, in many cases the use of programmable logic controller (PLC) and inverter is compatible use.

PLC is a computer with embedded CPU, together with the I/O modules, specifically designed for industrial environments. It developed a new control devices integrated with traditional relay technology, computer technology and the communication technology[5].

The cashmere and wool drawing system adopt ORMON's CQM1 Programmable Controller. CQM1 is a modular structure; it is mainly made up of the CPU module, power module, the bus connection cable and expansion modules. CPU module mainly consists of CPU and I/O two portions, but the function is limited and has no analog I/O. So it must be extended by special function module sometimes. CQM1 has special analog output module DA021, through extend, it can achieve fine control to motor.

The system attains full automatic control to all technological parameter and program. The function charts of synchronous control system as Fig.3.

PLC is the key of system. The DA021 analog output module output 4 control signals to relevant frequency converter. So the linear rate of drawing, setting and dryer units can be in accordance with predetermined synchronizing coefficient. At same time it also need guarantee the drawing unit and false twisting drive keep synchronous.

Linear speed of drawing unit (roller) is stored in DM0050 of PLC's data storage area (DM). The separate adjustment coefficients of M2 are stored in DM0051 and DM0054. The separate adjustment coefficients of M3 are stored in DM0071 and DM0074. The separate adjustment coefficients of M4 are stored in DM0091 and DM0094.

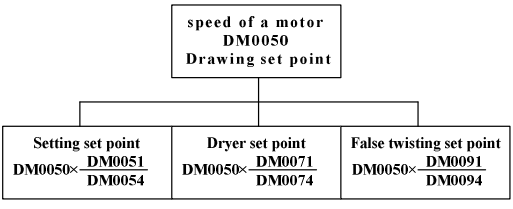


Fig. 4. Setting of motor speed

Speed setting is shown in figure 4.This setting is overall coordination, and belongs to open-loop setting. It equivalents to coarse adjustment to meet the basic needs of technics with the adjustment of each DM number. And the false twisting device

To realize pinpoint accuracy synchronization relationship, open-loop is not enough. The system must adopt the feedback of motor by sensor. The system uses three POG10DN1024I encoders (PG1-PG3) of HUBNER which placed at shaft of 3 motor.

As shown in Fig.4, angular-movement of motor transforms the linear displacement by optical-electrical conversion. Because of the difference of svelocity ratio and roller diametermm each unit, we must transform the pulse data to reflect linear speed as shown in Fig.5.

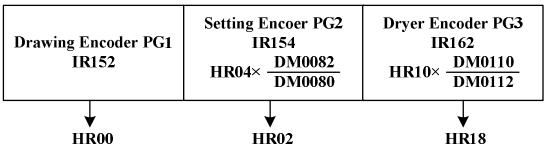


Fig. 5. Transformation coefficient setting of encoder

The difference of pulse data output by PG1 and PG2 can be symbol discriminated and Computing process. The second output of analog output module DA021 controls setting unit motor. This can make the setting module keep strict synchronization with drawing unit.

4 Design of Man-Machine Interface

PLC plays an important role in Industrial automation. Continuous development of technology accelerates improvement of control function and level With PLC control system as the core. Meanwhile, there are more and more requirements on its control method and operational level. So interactive interface, alarm record and print are also main parts of the control system. It shows special weightiness in complicated, with more control parameters process like hair slenderizing. The emergence of touch-screen provides a simple, feasible solution.

The control system of hair slenderizing adopts domestic manufactured eView series MT510T industrial touch-screen. Through abundant graphical interfaces, one can easily select, set, adjust and modify various process parameters and procedures. That is, it's possible to realize real-time display and control of the equipment on production site. At the same time, touch switch can be set on the screen which is safe, intuitive and simple operated. Functions like monitoring, fault diagnosis and alarm greatly improve the reliability of production[5].

5 Conclusion

The speed control of each unit is the key issue to fulfill the requirement of processing technic. Through adopting PLC and frequency converter, the machine realize infinite variable speed to simplify transmission organ. This makes transformation process simple, easy and certain. And monitor and control function become easier by using touch screen and has better stability and lower fault rates.

With the feedback of the data of motor gathered by encoder, the system can revise and compensate frequency converter in real time. Through practical operation of this system, it shows that the control system can achieve the desired requirements.

References

1. Zhang, Y.: Application of Frequency Control. China Machine Press (2001)
2. Yang, G.: Common Inverter Applications. Electronic Industry Press (2006)
3. Yang, S.: A Study of Processing Equipment and the Drawing Technology in the Slenderizing Wool Fibers. Journal of Textile (6), 32–33 (2003)
4. Sha, L.: Multi-motor synchronization system based on variable gain intelligent control. Journal of Tianjin Polytechnic University (10), 58–60 (2007)
5. Weikexin: Comprehensive Application Control. China Machine Press (2007)

Reliability Assessment of Carbon Sequestration Market Simulation*

Bo Tang and Jianzhong Gao

College of Economics and Management, Northwest A&F University Yangling, Shaanxi, China
{tangbo0817,gaojianzhong2003}@yahoo.com.cn

Abstract. This paper focuses on China Carbon Sequestration Market (CSM) and establishes the reliability assessment index frame of Carbon Sequestration Market simulation. It draws the conclusion that the CSM simulation is "More reliable" totally and the multi-agent dynamic gaming simulation of forest sequestration market is reliable and necessary. But these results also mean that this research exist some flaws to some extent. So more testify is needed in the aspect of simulation necessities. In terms of simulation condition, we also should enhance the construction of research basis and improve the quality of researchers continuously. It lays down the foundation for future study on China Sequestration Market.

Keywords: Reliability Assessment, Carbon Sequestration Market, Simulation.

1 Introduction

The research and practice of China Carbon Sequestration Market (CSM) is relatively slower than others. Plenty of literatures focus on the regime and measurement of carbon sequestration, which just simply reference the experience of overseas' carbon market, but less on simulation before establishing the market and overlook China national situation. If we have made a simulation of the CSM using the complex adaptive system theory, whether it is feasible and reliable? It is really necessary to assess the reliability of CSM simulation before taking into practice. So this research will construct system simulation index and make an assessment of the CSM simulation under such circumstance, then give some comments about the system simulation and propose some ideas for future studies on China Carbon Sequestration Market.

2 Reliability Assessment Index Frame of Carbon Sequestration Market Simulation

Carbon sequestration market is a complex system which needs to be observed from different aspects. So it is necessary to study the reliability assessment of carbon

* This paper is part of TANG Bo's master degree thesis, "Simulation Research on Multi-Agent Dynamic Game in China Carbon Sequestration Market Based on Complex Adaptive System Theory." which has been approved.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 133–140.
springerlink.com

sequestration market simulation in the perspective of complex system. In this paper, it explores 3 first-class indexes, including simulation necessities B_1, Simulation conditions possibilities B_2, simulation results rationality B_3, and 10 second-class indexes, which reflect the true feature of China Carbon Sequestration Market simulation under the circumstance of special situation. The detail reliability assessment indexes of carbon sequestration market simulation are constructed as follows.

Table 1. Reliable Assessment Index Frame of Csm Simulation

Simulation necessities B1	Establishing necessities C1
	Feature of CSM C2
	Complex system perspective C3
	National conditions C4
Simulation conditions possibilities B2	Collaborators' quality C5
	Research foundations C6
	Hard &soft ware supports C7
simulation results rationality B3	Simulation results verisimilitude C8
	Technical potential and innovation C9
	Outcome promotion prospects C10

3 Selection of the Evaluation Model

Carbon sequestration market simulation is a complex system and its agents and environment contains fuzziness. So it is advisable to evaluate the simulation reliability using the Fuzzy Comprehensive Evaluation model (FCE).

In the system engineering fields, system evaluation appears an uncertain complicated feature since the evaluation object containing both the subjective and objective information. However, it is insurmountable that the subjective cognition, preference and intuition from the valuator have a deeply impact on the index weight in the process of weighting the evaluation indexes (Jin,2008). This paper adopt the algorithm which tests and amends the judgment matrix on the basis of traditional Analytic Hierarchy Process(AHP) in the process of weighting,. So this paper will apply the improved Fuzzy Analytic Hierarchy Process, namely the Fuzzy Analytic Hierarchy Process based on Accelerated Genetic Algorithm (AGA-FAHP)(Wang, et al,2011; Yan,2011).

Accelerated Genetic Algorithm(AGA)is one kind of global general method simulating the survival of the fittest and internal information exchanging mechanism of organisms, which amends the consistency check of judgment matrix. Fixed judgment matrix and treatment of each element of the single sorting weights is the biggest difference of AGA compared to the traditional AHP. Suppose the judgment matrix $X = \{x_{ij}\}_{n\times n}$, each element of the single sorting weights of X is $\{w_k | k = 1 \sim n\}$, so the optimum consistency matrix:

$$\min\ CIC\ (n) = \sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_{ij} - b_{ij}\right| / n^2 + \sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_{ij} - w_i\right| / n^2$$

The CIC(n) implies the consistency index coefficient (CIC), and d is nonnegative parameters range from [0,0.5] according to the experience. The judgment matrix posses completely consistency only in the condition that the global minimum (CIC (n) =0) and the judgment matrix and accuracy of its single sorting matrix has consistency. If not, we should adjust the parameter d or judgment matrix repeatedly until we get the satisfactory consistency.

4 Reliability Assessment of Carbon Sequestration Market: A Fuzzy Comprehensive AHP Method Based on Genetic Algorithm

A. Judgment matrix and its consistency test

Usually, the weighing methods have Delphi, AHP and Information Entropy method, etc. We weight the indexes using the AHP in this paper and amend the weight adopting the Genetic Algorithm. Commonly, the element of judgment matrix use the scale of 1 to 9 and their reciprocal, then the judgment matrix of floor B can be expressed as $A=(a_{ij})_{n_b\times n_b}$, which means relative importance of Bi toward Bj in the perspective of A. Floor C's judgment matrix of element Bk: $B_k=\{b_{ij}^k \mid i,j=1\sim n_c\},k=1\sim n_b$. After three times' consulting through the Delphi approach, we get the judgment (Xu, et, al, 2011; Yang ,2009)

$$A=\begin{bmatrix} 1 & 1/3 & 2 \\ 3 & 1 & 3 \\ 1/2 & 1/3 & 1 \end{bmatrix} \quad B_1=\begin{bmatrix} 1 & 1/3 & 1/4 & 1/5 \\ 3 & 1 & 1/2 & 1/3 \\ 4 & 2 & 1 & 1/3 \\ 5 & 3 & 3 & 1 \end{bmatrix} \quad B_2=\begin{bmatrix} 1 & 1/5 & 1/3 \\ 5 & 1 & 3 \\ 3 & 1/3 & 1 \end{bmatrix} \quad B_3=\begin{bmatrix} 1 & 1/3 & 1/2 \\ 3 & 1 & 3 \\ 2 & 1/3 & 1 \end{bmatrix}$$

Presently, there are many ways of consistency test of judgment, such as product method, power method, and minimum deviation method and so on. This paper mainly adopts the improved AGA-AHP(Huang,2009). Usually, the single sorting weight should compare to the random consistency index function value. If the consistency index function value is lower than 0.1, accept it. If not, we should adjust the judgment matrix repeatedly, until we get the satisfactory results. But the real complicated system and Subjective unilateralism resulted from people's cognition diversity may generate the lower consistency since there is no unified judgment criteria while using the Delphi method to obtain the judgment matrix. So this paper introduces the Accelerated Genetic Algorithm to solve the problem of consistency test. It will improve the satisfaction of judgment consistency when it is used as a common global optimization scheme (Chen,2011; Jin,2005).

Average random consistency index value table of Accelerated Genetic Algorithm AHP (AGA-AHP) judgment matrix is listed as follows.

Table 2. Judgment matrix average random consistency INDEXES function RCIF (n) values

Order(n)	3	4	5	6	7	8	9
RCIF	1.62	1.64	1.98	1.86	1.69	1.81	1.95

Calculate the weight sorting using the single sorting scheme of judgment matrix respectively, and figure up the consistency test function value, get sorting results of each judgment matrix as table 3.

Table 3. Single Sorting and Consistency Function Value Figured Up from Different Method

Weigh sorting method	Judgment matrix	Weigh sorting				Consistency index function value
		w1	W2	W3	W4	
Product method	A1	0.25185	0.58889	0.15926		0.046469
Power method	A1	0.24931	0.59363	0.15706		0.046225
Minimum deviation method	A1	0.2485	0.5932	0.1583		0.043117
AGA-AHP	A1	0.2476	0.5955	0.1569		0.003958
Product method	B1	0.07816	0.16696	0.25198	0.5029	0.073799
Power method	B1	0.0709	0.16592	0.25214	0.51104	0.041064
Minimum deviation method	B1	0.09742	0.16313 7	0.24108 7	0.49836	0.040728
AGA-AHP	B1	0.10813	0.16723	0.22736	0.49728	0.016072
Product method	B2	0.10616	0.63335	0.2605		0.033375
Power method	B2	0.10473	0.63699	0.25828		0.033199
Minimum deviation method	B2	0.10991	0.63274	0.25735		0.030287
AGA-AHP	B2	0.10823	0.63724	0.25453		0.012731
Product method	B3	0.15926	0.58889	0.25185		0.046469
Power method	B3	0.15706	0.59363	0.24931		0.046225
Minimum deviation method	B3	0.1582	0.6037	0.2381		0.037691
AGA-AHP	B3	0.15116	0.59482	0.25402		0.00685

B. Fuzzy comprehensive evaluation analysis

1) Reliability rating division

To evaluate the simulation reliability rating of Carbon Sequestration Market clearly, this paper divide the simulation reliability of CSM into 4 parts, labels it Vi and gives the semantic scale, assigns values 9,7,5,3,1, that is $V=\{V_1,V_2,V_3,V_4\}$. V_1: Absolutely reliable,V_2: More reliable, V_3: Reliable, V_4: Less reliable, V_5: Absolutely not reliable. Refer to rating table 4.

Table 4. Reliability rating interval division

Reliability Ranking	V1	V2	V3	V4	V5
Scoring Interval	[8,10]	[6,8]	[4,6]	[2,4]	[1,2]
Reliability Rating	Absolutely reliable	More Reliable	Reliable	Less reliable	Absolutely not reliable

2) Fuzzy evaluation matrix and membership

This paper took the knowledge of subject on CSM and economics system simulation into consideration while selecting the subjects, since the subjects' interest and research area will have a great influence on the quality of questionnaire. We selected 5 masters and doctors from the field of forest economics, 3 postgraduates from economic information system area and 2 experts from system simulation field. The first round investigation may have less validity due to the less known about the China Carbon Sequestration Market simulation, although they are experts from their own fields. Gathering the results of the second survey by designing questionnaires adopting Delphi method, Pick up and get the data as table 5:

Table 5. Reliable Assessment Scores Of System Simulation

First-class indexes and weights Wi	Second-class indexes and weights Wij	Reliability rating and scores				
		V1 [9,10]	V2 [7,8]	V3 [5,6]	V4 [3,4]	V5 [1,2]
Simulation necessities B1 (0.247)	Establishing necessities (0.108)	3	6	5	2	0
	Features of CSM (0.167)	2	7	4	2	1
	Complex system perspective (0.227)	4	5	6	1	0
	National conditions (0.498)	6	6	3	1	0
Simulation conditions possibilities B2 (0.596)	Collaborators' quality (0.108)	4	5	6	1	0
	Research foundations (0.637)	2	6	5	2	1
	Hard &soft ware supports (0.255)	5	7	2	2	0
simulation results rationality B3 (0.157)	Simulation results verisimilitude(0.151)	3	7	5	1	0
	Technical potential and innovation (0.595)	4	6	2	4	0
	Outcome promotion prospects (0.254)	5	7	3	1	0

As the data expressed in above table, 3 experts give the V_1 rank for element C_1 in the index B_1, 6experts give the V_2 rank, 5 experts give V_3 rank, V_4 from 2, and no experts think the system simulation is absolutely not reliable. Membership of each element evaluation is figured up as follows.

$r^1_{11} = d^1_{11} / d = 3/16 = 0.1875$, $r^1_{12} = 6/16 = 0.375$, $r^1_{13} = 5/16 = 0.3125$, $r^1_{14} = 2/16 = 0.125$, $r^1_{15} = 0/16 = 0$

Similarly, we get the membership matrix of evaluation element C_2, $r^1_{21} = d^1_{21} / d = 2/16 = 0.125$, $r^1_{22} = 0.4375$, $r^1_{23} = 0.25$, $r^1_{24} = 0.1875$, $r^1_{25} = 0.0625$;

Membership matrix of C_3: $r^1_{31} = 0.25$, $r^1_{32} = 0.3125$, $r^1_{32} = 0.375$, $r^1_{33} = 0.1875$, $r^1_{34} = 0.0625$, $r^1_{35} = 0$;

Membership matrix of C_4: $r^2_{41} = 0.25$, $r^2_{42} = 0.3125$, $r^2_{43} = 0.375$, $r^2_{44} = 0.0625$, $r^2_{45} = 0$;

So the 4*5 order fuzzy judgment matrix of B_1:

$$R_{\sim 1}=\begin{bmatrix} 0.1875 & 0.375 & 0.3125 & 0.125 & 0 \\ 0.125 & 0.4375 & 0.25 & 0.125 & 0.0625 \\ 0.25 & 0.3125 & 0.375 & 0.0625 & 0 \\ 0.375 & 0.375 & 0.1875 & 0.0625 & 0 \end{bmatrix}$$

Similarly, 3*5 order judgment matrix of B_2:

$$R_{\sim 2}=\begin{bmatrix} 0.25 & 0.3125 & 0.375 & 0.0625 & 0 \\ 0.125 & 0.375 & 0.3125 & 0.125 & 0.0625 \\ 0.3125 & 0.4375 & 0.125 & 0.125 & 0 \end{bmatrix}$$

3*5 order judgment matrix of B_3:

$$R_{\sim 3}=\begin{bmatrix} 0.1875 & 0.4375 & 0.3125 & 0.0625 & 0 \\ 0.25 & 0.375 & 0.125 & 0.25 & 0 \\ 0.3125 & 0.4375 & 0.1875 & 0.0625 & 0 \end{bmatrix}$$

3) Comprehensive evaluation analysis

(1) First grade evaluation

Comprehensive evaluation vector of simulation necessities B_1:

$$B_{\sim 1}=W_{1j}\circ R_{\sim 1}=(0.108\ \ 0.167\ \ 0.227\ \ 0.498)\begin{bmatrix} 0.1875 & 0.375 & 0.3125 & 0.125 & 0 \\ 0.125 & 0.4375 & 0.25 & 0.125 & 0.0625 \\ 0.25 & 0.3125 & 0.375 & 0.0625 & 0 \\ 0.375 & 0.375 & 0.1875 & 0.0625 & 0 \end{bmatrix}$$

$$=(\ 0.2845\quad 0.3712\quad 0.2541\quad 0.0797\quad 0.0105\ )$$

Comprehensive evaluation vector of Simulation conditions possibilities B_2:

$$B_{\sim 2}=W_{2j}\circ R_{\sim 2}=(0.108\ \ 0.637\ \ 0.255)\begin{bmatrix} 0.25 & 0.3125 & 0.375 & 0.0625 & 0 \\ 0.125 & 0.375 & 0.3125 & 0.125 & 0.0625 \\ 0.3125 & 0.4375 & 0.125 & 0.125 & 0 \end{bmatrix}$$

$$=(\ 0.1863\quad 0.3841\quad 0.2715\quad 0.1182\quad 0.0398\ )$$

Comprehensive evaluation vector of simulation results rationality B_3 :

$$B_{\sim 3}=W_{3j}\circ R_{\sim 3}=(0.151\ \ 0.595\ \ 0.254)\begin{bmatrix} 0.1875 & 0.4375 & 0.3125 & 0.0625 & 0 \\ 0.25 & 0.375 & 0.125 & 0.25 & 0 \\ 0.3125 & 0.4375 & 0.1875 & 0.0625 & 0 \end{bmatrix}$$

$$=(\ 0.2564\quad 0.4003\quad 0.1692\quad 0.174\quad 0\ )$$

(2) Second grad evaluation

We can get the second grade evaluation vector B~ according to the results of first grade evaluation.

$$B_{\sim}=\begin{bmatrix} 0.2845 & 0.3712 & 0.2541 & 0.0797 & 0.0105 \\ 0.1863 & 0.3841 & 0.2715 & 0.1182 & 0.0398 \\ 0.2564 & 0.4003 & 0.1692 & 0.174 & 0 \end{bmatrix}$$

So the result of second grade evaluation:

$$U = W_i \circ R_- = (0.2476 \quad 0.5955 \quad 0.1569) \begin{bmatrix} 0.2845 & 0.3712 & 0.2541 & 0.0797 & 0.0105 \\ 0.1863 & 0.3841 & 0.2715 & 0.1182 & 0.0398 \\ 0.2564 & 0.4003 & 0.1692 & 0.174 & 0 \end{bmatrix}$$

$$= (0.2216 \quad 0.3835 \quad 0.2512 \quad 0.1175 \quad 0.0263)$$

(3)Reliability ranking of comprehensive evaluation vector

Reliability ranking scores of simulation necessities B_1:

$V_{B_1} = (9 \quad 7 \quad 5 \quad 3 \quad 1)[0.2845 \quad 0.3712 \quad 0.2541 \quad 0.0797 \quad 0.0105]^T = 6.6792$. Reliability ranking scores of Simulation conditions possibilities B_2: $V_{B_2} = 6.1175$.

Reliability ranking scores of simulation results rationality B_3: $V_{B_3} = 6.4783$.

Reliability ranking scores of CSM system simulation:.V=6.3132.

5 Conclusions

From the evaluation results of AGA-FAHP, we can see that the reliability of the whole carbon sequestration market system simulation ranks V2, ranges the interval of [6,8], that is "More Reliable" . And the values of other three first class indexes also belongs the scale of [6,8], totally reliable. But these results also mean that this research exist some flaws to some extent. Such as in terms of simulation necessities, more testify is needed from multi perspective; in terms of simulation condition, we should enhance the construction of research basis and improve the quality of researchers continuously, and so on.

This paper makes an assessment of China Carbon Sequestration Market and gets the conclusion that multi-agent dynamic gaming simulation of forest sequestration market is reliable and necessary. However, construction of China Carbon Sequestration Market is long term system engineering and need the joint effort of different administrations according to the reliability assessment consequence. Only on the basis of fully investigation and simulation practice, can we hope to establish the real China Carbon Sequestration Market (CSM).

References

1. Tang, B., Gao, J.Z.: Review of Multi-Agent Dynamic Gambling Simulating Research on China Carbon Sink Market Based on Complex Adaptive System Theory. In: 2010 The 3rd International Conference on Computational Intelligence and Industrial Application, vol. 6, pp. 306–309 (2010)
2. Yang, B.W., Ren, J.: Forming Process of Game of Interests Groups in Complex Adaptive System. Chinese Journal of Systems Science 18(1), 39–41 (2010) (in Chinese)
3. Xu, H., Qiu, T., Zhao, J.S.: Synchronized method for construction a fuzzy judgment matrix and checking its matrix consistency. Journal of Tsinghua University (Science and Technology) 50(6), 913–916 (2010) (in Chinese)
4. Jin, J.L., Yang, X.H., Wei, Y.M.: System Evaluation Method Based on Fuzzy Preferential Relation Matrix. Systems Engineering-Theory Methodology Application 14(4), 364–368 (2005) (in Chinese)

5. Yang, J., Qiu, W.H.: Research on consistency test and modification approach of fuzzy judgment matrix. Control and Decision 24(6), 903–906 (2009) (in Chinese)
6. Wang, L.P., Jia, Z.Y.: Solving complete job shop scheduling problem using genetic algorithm in feasible domain. Journal of Dalian University of Technology 51(2), 205–209 (2011) (in Chinese)
7. Yan, L.W., Chen, S.H.: Solving Nonlinear Equations Based on Improved Genetic Algorithm. Acta Scientiarum Naturalium Universitatis Sunyatseni 50(1), 9–13 (2011) (in Chinese)
8. Kuo, R.J., Yang, C.Y.: Simulation optimization using particle swarm optimization algorithm with application to assembly line design. Applied Soft Computing Journal 11(1), 605–613 (2011)
9. Huang, S.Z.: Research and application of wavelet neural networks of particle swarm optimization algorithm in the performance prediction of centrifugal compressor. Advanced Materials Research 187, 271–276 (2011)
10. Chen, T.: Selective SVM ensemble based on accelerating genetic algorithm 28(1), 139–141 (2011) (in Chinese)
11. Huang, Y.H., Zhu, J.F.: Study and Application of PPC Model Based on RAGA in the Clustering Evaluation. Systems Engineering 27(11), 107–110 (2009) (in Chinese)

A Customer Requirements Rating Method Based on Fuzzy Kano Model

Lei Xie[1] and Zhongkai Li[2]

[1] Vocational College of Zibo, Zibo, China
xielei1999@tom.com
[2] School of Mechatronics Engineering, China University of Mining and Technology, Xuzhou, China
lizk@cumt.edu.cn

Abstract. Aiming at the problems of lacking classification ability in customer requirements rating computation, a customer requirements classification and importance rating method based on fuzzy Kano model was proposed. In the requirements classification stage, the customer survey data were counted and analyzed with the fuzzy Kano model, to divide the requirements into three types such as attractive, must-be and one-dimensional, which improving the classification ability for uncertain demand information. In the following importance rating stage, the adjustment functions for importance degree were introduced, during which different adjustment functions were used to calculate the weights of needs on the basis of products' basic information. The requirements category could be combined with the related importance calculation to guide enterprises getting the maximum customer satisfaction with the minimum cost input. The effectiveness of the proposed method was illustrated by the customer satisfaction evaluation of an enterprise's combine harvesters.

Keywords: customer satisfaction rate, fuzzy Kano model, requirements importance, importance adjustment, combine harvester.

1 Introduction

With the increase of market competition, the market transformed to be customer-oriented from enterprise-oriented. The companies should mining the market demand and concern the quality, price and life cycle of product at the same time. The customer requirements information is first importance for product design. How to extract the customer requirements in the maximum degree and analyze the relative importance ratings are the basis for product design and development.

In the requirements classification and rating literatures, Yang [1] recommended eight satisfaction types of the Kano model, which dividing the traditional attractive, must-be, one-dimensional and indifference quality to improve the classification accuracy. Tontini [2] introduced an adjustment factor to assign higher weights to factors with "cause customer satisfaction" and "eliminate customer dissatisfaction". Chan [3] used the fuzzy analytic hierarchy process (AHP) to analyze requirements

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 141–146.
springerlink.com

importance, but the accuracy of AHP relied heavily on expert experience. Schneider [4] proposed the American customer satisfaction index (ACSI) model to improve customer satisfaction indices for service industry. This model could aggregate customer satisfaction with past consuming experiences and forecast the future performance, but it lacked the quantitative analysis capability.

With the drawbacks of traditional customer requirements rating methods, a customer requirements rating method based on fuzzy Kano model is proposed. The fuzzy Kano questionnaire and fuzzy processing are proposed to classify the requirements with fuzzy information. Then, the adjustment functions for importance rating are introduced to calculate the weights of requirements in different categories, in order to get the maximum customer satisfaction with the minimum resource cost.

2 Computational Model for Requirements Rating

The aim of the customer requirements rating method based on fuzzy Kano model is to improve the customer satisfaction degree, during which the requirements are classified and the weights are adjusted with different functions to develop products with high customer satisfaction.

The computational model for customer requirements rating is illustrated in Fig. 1, and described as follows:

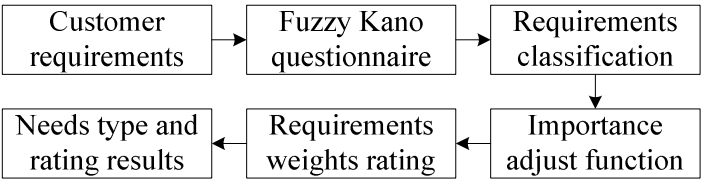


Fig. 1. Computational framework for requirements rating with fuzzy Kano model

The requirements items of customers for the product should be defined firstly with the previous market survey. Next, the fuzzy Kano question tables should be developed, hand out and analyzed to acquire the items initial importance, outside competitive strength and the original data for fuzzy Kano model. Then, the fuzzy Kano model will be used to classify the requirements to different categories, and the different requirements adjusting functions can be introduced to adjust the weights rating with different requirements category. Finally, the validity of the rating results should be verified, and the final ratings of requirements can be used to instruct the product design.

3 Requirements Classification with Fuzzy Kano Model

Customers can only select a most satisfied option from the given options in traditional Kano survey [5]. But it ignores the uncertain feature of human thought. Considering the uncertainty of customer satisfaction, the fuzzy Kano model [6] for requirements classification is proposed.

The first difference between traditional and fuzzy Kano models is the format of questionnaire. Traditional Kano model only permits customers to choose a most satisfied answer to the positive or negative question. While fuzzy Kano model permits customers to give fuzzy satisfactory values to a number of survey items. The satisfactory values are expressed with the form of percentage degree, and the values are between 0 and 1 with the sum of the row elements is equal to 1. The fuzzy Kano is similar to traditional Kano method in the data treatment stage. The basic steps are as follows:

1) Supposing a fuzzy Kano questionnaire example as: MP = [0.5 0.5 0 0 0], MN = [0 0 0 0.2 0.8]. The interaction matrix MS is generated:

$$MS_{5\times5} = MP_{1\times5}^{\mathrm{T}} \times MN_{1\times5} = \begin{bmatrix} 0 & 0 & 0 & 0.10 & 0.40 \\ 0 & 0 & 0 & 0.10 & 0.40 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{1}$$

2) By comparing the matrix MS with the classification assignment table in Kano model [5], a requirement's category association vector T can be acquired as:

$$T = \left\{ \frac{0.4}{\mathrm{M}}, \frac{0.4}{\mathrm{O}}, \frac{0.1}{\mathrm{A}}, \frac{0.1}{\mathrm{I}}, \frac{0}{\mathrm{R}}, \frac{0}{\mathrm{Q}} \right\}. \tag{2}$$

3) To acquire more accurate and reliable data, a threshold "α" is introduced to filtering the data in vector T. Set $\alpha = 0.4$, when $T_{i,j} \geq 0.4$, $T_{i,j} = 1$, which represents the requirement following the relative category viewed by the customer. Otherwise, $T_{i,j} = 0$. The requirement association vector is acquired as $T_0 = \{1, 1, 0, 0, 0, 0\}$. It can be seen that the requirement not only is must-be (M), but also tend to be one-dimensional (O) at the same time.

4) Repeat steps 1 to 3 for all the customer survey for this requirement item, and count the item category number. The one with highest frequency is the category of the requirement item. If two types have the same frequency, the priority list for requirements category is: M > O > A > I.

5) Repeat steps 1 to 4 for all the requirement items to classify the design requirements.

4 Importance Rating with Adjustment Functions

4.1 Adjusting Functions

The relationship between customer satisfaction and product performance in Kano model can be represented with a function with parameters: $S = f(k, p)$, where S is satisfaction rate, p is product performance, and k is adjusting coefficient in Kano model. Supposing S_0 and p_0 are current satisfaction and performance levels, and S_1, p_1 are expecting level. So it can be deduced that

$$\frac{S_1}{S_0} = \frac{cp_1^k}{cp_0^k} = \left(\frac{p_1}{p_0}\right)^k . \tag{3}$$

$$IR_{\text{adj}} = (IR_0)^{1/k} . \tag{4}$$

where IR_{adj} is the performance improving coefficient, IR_0 is the initial satisfaction improving factor. k is a factor with different value in different requirements categories. In the fuzzy Kano model, the k value for attractive, one-dimensional and must-be requirements can be set as 0.5, 1 and 2 according to computational experience. k can also be acquired with function $k = \log_{p_1/p_0} S_1/S_0$.

4.2 Importance Rating Process

Based on the product requirements and customer satisfaction factors analysis, the customer satisfaction rating computational process is shown in Fig. 2. Some key steps are described as follows:

1) According to the advices from design and sales departments, n customer requirements should be identified which most impact customer satisfaction. The questionnaires are designed with these items and hand out to make market survey.

2) In the initial importance, internal analysis and competitive analysis process, numbers 1 to 5 are used to illustrate customer satisfaction, during which 5 represents very satisfied, and 1 is to be very dissatisfied.

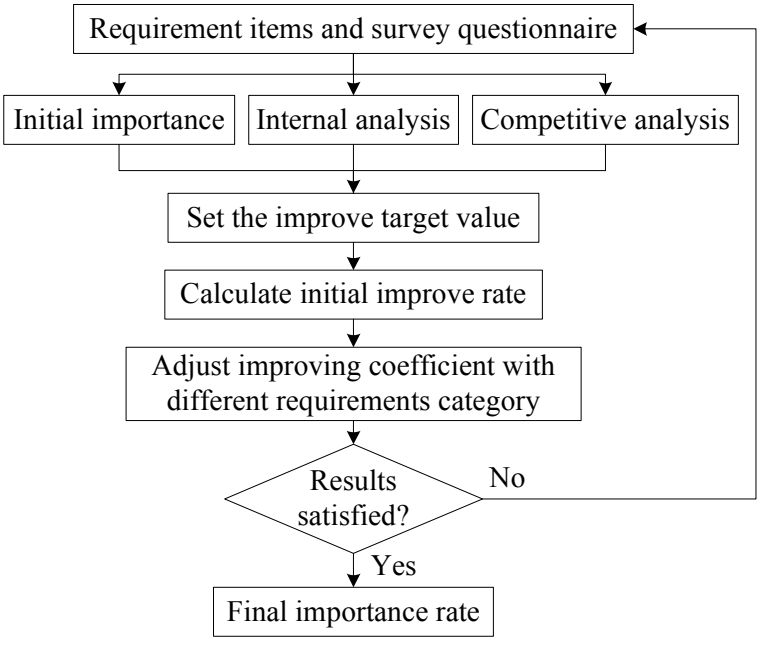


Fig. 2. Computational framework for requirements importance adjustment and rating

3) The improve target value should be set with the customer survey. It is the target value for satisfaction rate, and can be represented with a value in 1 to 5. There are two comments that firstly if the satisfaction rate is lower than competitive, then it should be set as the same level; If the level is higher than competitive, then it can be not changed.

4) The initial improve rate can be calculated as: improve rate = target value / initial importance. It represents the current improving degree for the requirement item.

5) The improving coefficients are adjusted with different requirements categories, which select different k value for requirement items in different Kano categories.

6) The final importance rate is defined as: final importance = improved coefficient × initial importance. If the final rating results are reasonable, we output the rates to guide product design, else, return to step 1 to make another round market survey.

5 Case Study

Combine harvester is a multi-function machine to harvest crops. With the after-sales experience, seven items such as flexible operation (R_1), GPS (R_2), speed adjustment (R_3), operating room user- friendly design (R_4), high efficiency threshing (R_5), reasonable price (R_6) and beautiful shape (R_7) are take as requirements to make Kano survey.

80 fuzzy Kano questionnaires were hand out, during which the survey objects are the harvester customers, repairing customers and product designers. 73 questionnaires were feedback, and 9 questionable surveys were deleted and 64 effective copies were inputted to the fuzzy Kano model. The requirements surveys and their affiliations are shown in Table 1, during which the type with the maximum number is the classification for the respective requirement. So R_1 and R_5 are must-be requirements, R_3, R_6 and R_7 belong to one-dimensional type, and R_2 and R_4 are attractive.

The requirements rating with adjustment functions are listed in Table 2. After the adjusting process, the must-be requirements R_1 and R_5 improved their weights. The weights of one-dimensional requirements R_3, R_6 and R_7 are not changed, and the attractive R_2 and R_4 lowered their weights. These illustrated that more cost should be put in to guarantee the must-be requirements.

Table 1. Requirement items frequency for different Kano classification

Requirements No.	M	O	A	I	R	Q	Calssification
R_1	*36*	17	12	6	0	0	M
R_2	8	15	*39*	7	0	0	A
R_3	19	*32*	17	5	0	0	O
R_4	8	21	*35*	6	0	0	A
R_5	*41*	14	11	5	0	0	M
R_6	17	*38*	16	4	0	0	O
R_7	14	*36*	15	8	0	0	O

Table 2. Customer requirements rating results with fuzzy Kano model for combine harvester

No.	Initial value	Kano	Internal	Competitive	Target	Coefficient	Improved factor	Final result
R_1	5	M	4	5	5	1.25	1.56	7.80
R_2	4	A	2	3	3	1.50	1.22	4.88
R_3	5	O	4	3	4	1.00	1.00	5.00
R_4	3	A	2	3	3	1.50	1.22	3.66
R_5	4	M	3	4	4	1.33	1.77	7.08
R_6	4	O	4	3	4	1.00	1.00	4.00
R_7	3	O	3	4	4	1.33	1.33	4.00

6 Summary

The method is important for enterprises to acquire requirements and innovative product design. From the previous market survey, the requirements are classified with fuzzy Kano model to eliminate uncertain information. The adjusting functions are developed to change the weights for requirements in different categories. Requirements analysis for combine harvester proved its effectiveness.

Acknowledgements. This research is finically supported by the National Natural Science Foundation of China 51005237.

References

1. Yang, C.C.: Total Quality Management 16, 1127 (2005)
2. Tontini, G.: Total Quality Management 18, 599 (2007)
3. Chan, F.T.S., Kumar, N., Tiwari, M.K.: Int. J. of Prod. Res. 46, 3825 (2008)
4. Schneider, B., Macey, W.H., Lee, W.C.: J. of Ser. Res. 12, 3 (2009)
5. Sharif, A.M.M., Tamaki, J.: Sys. Eng. 14, 154 (2011)
6. Lee, Y.C., Huang, S.Y.: Exp. Sys. With App. 36, 4479 (2009)

The State of Solution for a Nonlinear Age-Dependent Population System Based on Weight

Dai Lili

Department of Mathematic
TongHua Normal University
TongHua, China
drx820115@126.com

Abstract. Population control is the process of forcing a population in order to obtain a certain behaviour of it. This article discussed the solution of the non-linear Age-Dependent population system, proved the existence of solution and the solution continually depended on the control variable.

Keywords: nonlinear population system, age-dependence, the existence of solution, depended on the control variable.

1 Introduction

In current, more and more people attention the ecological balance of population and the sustainable development of society Mathematical biology. With the continuous development of mathematical biology, people use all sorts of mathematical tools to establish from simple to complex mathematical modeling process of life. Using mathematical methods research population problem and people take effective measures to control population problem. It is expected to make the ideal state of development. It is important significance to research the system of population and optimal control problems. Therefore, This article discussed the solution of the non-linear Age-Dependent population system (P), proved the existence of solution and the solution continually depended on the control variable .

Consider the following non-linear Age-Dependent population system based on weight (P):

$$\begin{cases} \dfrac{\partial p}{\partial a}+\dfrac{\partial p}{\partial t}+\mu(a,t;S)p(a,t)=0, \quad (a,t)\in Q=(0,T)\times\Omega \\ p(0,t)=\int_0^A \beta(a,t;S)p(a,t)da+u(t)\ ,\ t\in(0,T)\ , \\ p(a,0)=p_0(a)\ ,\ a\in(0,A)\ , \\ S(t)=\int_0^A \omega(a,t)p(a,t)da\ ,\ t\in(0,T). \end{cases} \tag{1.1}$$

The system (P) is described by the initial and boundary value problems of the non-linear integral-partial differential equations in this paper. In this mathematical model, the state

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 147–153.
springerlink.com

function $p(a,t)$ for the system (P) denotes the distribution of age density at time t and age a; $p_0(a)$ is the initial distribution of age density when $t=0$; $S(t)$ denotes the weighted total quantity of the population at time t; ω is weighted function; $u(t)$ is the control function; β, μ partly denote the birth and death-rate functions of the population ;in this paper β, μ both have some connection with ω ,which incarnates the actual influence of the weighted size on the dynamic process for the population, so it has the actual significance even more.

2 Hypotheses

Throughout this paper we work under the following assumptions, excepting the case when we will explicitly mention other hypotheses:

(H_1) $\mu(a,t;y)\geq 0$ is nonnegative on $Q\times R_+$, measurable in a, t, twice continuously differentiable in y, and $|\mu(a,t;y)|+|\mu_y(a,t;y)|+|\mu_{yy}(a,t;y)|\leq C_1$; , (H_2) $\beta(a,t;y)$ is nonnegative on $Q\times R_+$, measurable in a, t, twice continuously differentiable in y, and $|\beta(a,t;y)|+|\beta_y(a,t;y)|+|\beta_{yy}(a,t;y)|\leq C_2$; (H_3) $p_0(a)\in L^\infty(0,A)$, $0\leq p_0(a)\leq C_3$ a.e. $(0,A)$, $\omega(a,t)\in L^\infty(Q)$, $0\leq\omega(a,t)\leq C_4$; C_1,C_2,C_3,C_4 are constant.

3 The Existence of Solution to (P)

Definition 3.1. By a solution to (P) we mean a function $p(a,t)\in L^\infty(Q)$,absolutely continuous along almost every characteristic line (of equation $t-a=\text{const}$, such that

$$\begin{cases} Dp(a,t)=-\mu(a,t;S(t))p(a,t), & a.e.\ (a,t)\in Q,\\ \lim\limits_{\varepsilon\to 0^+} p(\varepsilon,t+\varepsilon)=\int_0^A \beta(a,t;S(t))p(a,t)da+u(t), & a.e.\ t\in(0,T),\\ \lim\limits_{\varepsilon\to 0^+} p(a+\varepsilon,\varepsilon)=p_0(a), & a.e.\ \ a\in(0,A),\\ S(t)=\int_o^A \omega(a,t)p(a,t)da, & a.e.\ \ t\in(0,T). \end{cases}$$

$$Dp(a,t)=\lim_{\varepsilon\to 0^+}\frac{p(a+\varepsilon,t+\varepsilon)-p(a,t)}{\varepsilon}=\frac{\partial p}{\partial a}+\frac{\partial p}{\partial t},$$

(Dp is a directional derivative).

LEMMA 3.1 Fixed $S\in L^\infty(0,T)$, $S(t)\geq 0$, for any $t\in(0,T)$, Using the definition of a solution to (P) we can obtain(by integration along the characteristic lines)

$$p(a,t;S)=\begin{cases} p_0(a-t)\Pi(a,t,t;S), & a\geq t\\ b(t-a;S)\Pi(a,t,a;S), & a<t \end{cases} \tag{3.1}$$

which

$$\Pi(a,t,s;S)=\exp\left\{-\int_0^s \mu(a-\tau,t-\tau;S(t-\tau))d\tau\right\}, s\in(0,\min\{a,t\}) \tag{3.2}$$

Thus $b(\cdot\,;S)$ satisfies the following Volterra equation

$$b(t;S)=F(t;S)+\int_0^t K(t,s;S)b(t-s;S)ds, \qquad t\in(0,T) \tag{3.3}$$

Which

$$F(t;S)=\int_0^\infty \beta(a+t,t;S)\,p_0(a)\Pi(a+t,t,t;S))da+u(t) \tag{3.4}$$

$$K(t,a;S)=\beta(a,t;S)\Pi(a,t,a;S) \tag{3.5}$$

The fuction p_0,β,Π continuation zero outside of these domain.

We shall give the proof only for the case $T>A$, we omit the proof for the case $T\le A$.

LEMMA 3.2 Existing the constant $C_6>0$, $C_7>0$,

$C_8>0$,for almost any $t\in(0,T)$,we have that

$$\left|F(t;S_1)-F(t;S_2)\right|\le C_6\left(\left|S_1(t)-S_2(t)\right|+\int_0^t\left|S_1(s)-S_2(s)\right|ds+\left|u_1(t)-u_2(t)\right|\right) \tag{3.6}$$

$$0\le b(t;S_1)\le C_7 \tag{3.7}$$

$$\left|b(t;S_1)-b(t;S_2)\right|\le C_8\left(\left|S_1(t)-S_2(t)\right|+\int_0^T\left|S_1(s)-S_2(s)\right|ds+\left\|u_1(t)-u_2(t)\right\|_{L^\infty(0,T)}\right) \tag{3.8}$$

which $S_i(t)=\int_o^A \omega(a,t)\,p_i(a,t)da$, $p_i(a,t)=p(a,t;u_i)$, $i=1,2$ Proof. Firstly estimated $\left|F(t;S_1)-F(t;S_2)\right|\cdot$

By(3.2), (3.4) and the Hypotheses (H_1)-(H_4),for almost any $t\in(0,T)$,We have that

$$\begin{aligned}
&\left|F(t;S_1)-F(t;S_2)\right|\\
&\le\int_0^\infty\left|\beta(a+t,t;S_1)-\beta(a+t,t;S_2)\right|\left|p_0(a)\Pi(a+t,t,t;S_1)\right|da\\
&+\int_0^\infty\left|\beta(a+t,t;S_2)\,p_0(a)\right|\left|\Pi(a+t,t,t;S_1)-\Pi(a+t,t,t;S_2)\right|da+\left|u_1(t)-u_2(t)\right|\\
&\le\int_0^\infty C_2\left|S_1(t)-S_2(t)\right|\left|p_0(a)\Pi(a+t,t,t;S_1)\right|da\\
&+\int_0^\infty C_1\left|\beta(a+t,t;S_2)\,p_0(a)\right|\int_0^t\left|S_1(s)-S_2(s)\right|ds\,da+\left|u_1(t)-u_2(t)\right|\\
&\le AC_2C_3\left|S_1(t)-S_2(t)\right|+AC_1C_2C_3\int_0^t\left|S_1(s)-S_2(s)\right|ds\\
&+\left|u_1(t)-u_2(t)\right|\\
&\le C_6\left(\left|S_1(t)-S_2(t)\right|+\int_0^t\left|S_1(s)-S_2(s)\right|ds+\left|u_1(t)-u_2(t)\right|\right)\\
&C_6=\max\left\{AC_2C_3,AC_1C_2C_3,1\right\}
\end{aligned} \tag{3.9}$$

Secondly estimated $b(t;S_1)$

Relations(3.3), for almost any $t\in(0,T)$,we obtain that

$$b(t;S_1)=F(t;S_1)+\int_0^t K(t,s;S_1)b(t-s;S_1)ds$$

$$= \int_0^{\infty} \beta(a+t,t;S_1)\, p_0(a)\Pi(a+t,t,t;S_1))da + u_1(t) \qquad + \int_0^t K(t,s;S_1)b(t-s;S_1)ds$$

$$\le AC_2C_3 + \overline{u} + C_2\int_0^t b(t-s;S_1)ds\,,$$

By Bellman inequality we get

$$b(t;S_1) \le (AC_2C_3 + \overline{u})\exp(C_2T) = C_7\,.$$

Finally estimated $\left|b(t;S_1) - b(t;S_2)\right|$.

By(3.3) and (3.6) we get

$$\left|b(t;S_1) - b(t;S_2)\right| \qquad \le \left|F(t;S_1) - F(t;S_2)\right| + \int_0^t \left|K(t,s;S_1)b(t-s;S_1) - K(t,s;S_2)b(t-s;S_2)\right|ds$$

$$\le C_6\left(\left|S_1(t) - S_2(t)\right| + \int_0^t \left|S_1(s) - S_2(s)\right|ds + \left|u_1(t) - u_2(t)\right|\right)$$

$$+ \int_0^t \left|K(t,s;S_1)b(t-s;S_1) - K(t,s;S_2)b(t-s;S_2)\right|ds\,.$$

$$\le C_6\left(\left|S_1(t) - S_2(t)\right| + \int_0^t \left|S_1(s) - S_2(s)\right|ds + \left|u_1(t) - u_2(t)\right|\right)$$

$$+ \int_0^t b(t-s;S_1)\left|K(t,s;S_1) - K(t,s;S_2)\right|ds \quad + \int_0^t K(t,s;S_2)\left|b(t-s;S_1) - b(t-s;S_2)\right|ds$$

$$\le C_9\left(\left|S_1(t) - S_2(t)\right| \quad + \quad \int_0^t \left|S_1(s) - S_2(s)\right|ds \quad + \quad \left\|u_1(t) - u_2(t)\right\|_{L^{\infty}(0,T)}\right)$$

$$+ C_2\int_0^t \left|b(t-s;S_1) - b(t-s;S_2)\right|ds$$

$$= \eta(t) + C_2\int_0^t \left|b(t-s;S_1) - b(t-s;S_2)\right|ds \tag{3.10}$$

which $C_9 = \max\left\{C_6, C_6 + C_2C_7 + C_1C_2C_7\right\}$,

$$\eta(t) = C_9\left(\left|S_1(t) - S_2(t)\right| + \int_0^t \left|S_1(s) - S_2(s)\right|ds + \left\|u_1(t) - u_2(t)\right\|_{L^{\infty}(0,T)}\right).$$

By Gronwall inequality we get

$$\left|b(t;S_1) - b(t;S_2)\right| \le \eta(t) + \int_0^t C_2\eta(s)\exp\left(\int_s^t C_2 d\tau\right)ds$$

$$\le \eta(t) + C_2\exp(TC_2)\int_0^t \eta(s)ds \tag{3.11}$$

which

$$\int_0^t \eta(s)ds \le (C_9 + C_9T)\int_0^t \left|S_1(s) - S_2(s)\right|ds$$

$$+ C_9T\left\|u_1(t) - u_2(t)\right\|_{L^{\infty}(0,T)} + (C_9 + C_9T)\left|S_1(t) - S_2(t)\right|$$

$$\le (1+T)\eta(t) \tag{3.12}$$

Relations (3.12) and (3.11), we conclude that

$$\left|b(t;S_1) - b(t;S_2)\right| \le \eta(t) + C_2\exp(TC_2)(1+T)\eta(t)$$

$$\le \eta(t)\left[1 + C_2\exp(TC_2)(1+T)\right]$$

$$\le C_8\left(|S_1(t)-S_2(t)|+\int_0^t|S_1(s)-S_2(s)|ds+\|u_1(t)-u_2(t)\|_{L^\infty(0,T)}\right)$$

$$C_8 = C_9\left[1+C_2\exp(TC_2)(1+T)\right].$$

Definition 3.2. $H=\{v\in L^\infty(0,T;L^1(0,A));a.e.v(a,t)\ge 0\ \}$

define

$$\Lambda: H\to L^\infty(0,T;L^1(0,A))$$
$$(\Lambda v)(a,t)=p(a,t;V),$$
$$V(t)=\int_0^A\omega(a,t)v(a,t)da$$

$p(a,t;V)$ such as (3.1), where V replace S,

obviously $\Lambda v\in H$.

Lemma 3.3. Existing the constant $C_{10}>0$, for any $v_1,v_2\in H$, we have that

$$\|(\Lambda v_1)(\cdot,t)-(\Lambda v_2)(\cdot,t)\|_1\le C_{10}\int_0^t\|v_1(\cdot,s)-v_2(\cdot,s)\|_1 ds \tag{3.13}$$

which $\|\cdot\|_1=\|\cdot\|_{L^1(0,A)}$

Proof. For any $v_1,v_2\in H$ we define

$$V_i(t)=\int_0^A\omega(a,t)v_i(a,t)da, i=1,2.$$

Using (3.1), (3.2) and Lemma 3.2, we get that

$$\|(\Lambda v_1)(\cdot,t)-(\Lambda v_2)(\cdot,t)\|_1=\int_0^A|p(a,t;V_1)-p(a,t;V_2)|da$$
$$\le\int_0^t b(t-a;V_1)|\Pi(a,t,a;V_1)-\Pi(a,t,a;V_2)|da$$
$$+\int_0^t\Pi(a,t,a;V_2)|b(t-a;V_1)-b(t-a;V_2)|da$$
$$+C_3\int_t^A|\Pi(a,t,t;V_1)-\Pi(a,t,t;V_2)|da \tag{3.14}$$

Relations (3.7) and (3.8), (3.11), (3.14), we get that

$$\|(\Lambda v_1)(\cdot,t)-(\Lambda v_2)(\cdot,t)\|_1\le C_7\int_0^t\Big|\exp\{-\int_0^a\mu(a-\tau,t-\tau;V_1(t-\tau))d\tau\}$$
$$-\exp\{-\int_0^a\mu(a-\tau,t-\tau;V_2(t-\tau))\}d\tau\Big|da$$
$$+\int_0^t|b(t-a;V_1)-b(t-a;V_2)|\exp\{-\int_0^a\mu(a-\tau,t-\tau;V_2(t-\tau)d\tau\}da$$
$$+C_3\int_t^A\Big|\exp\{-\int_0^t\mu(a-\tau,t-\tau;V_1(t-\tau))d\tau\}-\exp\{-\int_0^t\mu(a-\tau,t-\tau;V_2(t-\tau))\}d\tau\Big|da$$
$$\le(TC_1C_7+C_8+TC_8+AC_1C_3)C_4\int_0^t\|v_1(\cdot,s)-v_2(\cdot,s)\|_1 ds\qquad\le C_{10}\int_0^t\|v_1(\cdot,s)-v_2(\cdot,s)\|_1 ds$$

Which $C_{10}=C_4(TC_1C_7+C_8+TC_8+AC_1C_3)$.

We are now able to state the main result in this paper, the uniqueness and existence of solution to the system (P).

Theorem 3.1. Let the hypothesis (H_1)-(H_4) be satisfied,then for any $u \in U_{ad}$,the system (P) has one and only one nonnegative solution $p^u \in L^\infty((0,T);L^1(0,A))$, bounded consistent in u and $\forall t \in (0,T), u_1, u_2 \in U_{ad}$,then

$$\left\| p^{u_1}(\cdot,t) - p^{u_2}(\cdot,t) \right\|_{L^1(0,A)} \le C_{11} \left\| u_1(t) - u_2(t) \right\|_{L^\infty(0,T)} \tag{3.15}$$

$$\left\| p^{u_1}(\cdot,t) - p^{u_2}(\cdot,t) \right\|_{L^\infty(Q)} \le C_{12} \left\| u_1(t) - u_2(t) \right\|_{L^\infty(0,T)} \tag{3.16}$$

Proof. We have defined equivalence norm in H

$$\left\| h \right\|_* = Ess \sup_{t\in(0,t)} \left\{ e^{-\lambda t} \left\| h(\cdot,t) \right\|_1 \right\}, \lambda > B_4 \tag{3.17}$$

By Lemma 3.3, we get that

$$\begin{aligned} \left\| \Lambda h_1 - \Lambda h_2 \right\|_* &\le Ess \sup_{t\in(0,t)} \left\{ e^{-\lambda t} \left\| (\Lambda h_1)(\cdot,t) - (\Lambda h_2)(\cdot,t) \right\|_1 \right\} \\ &\le C_{10} Ess \sup_{t\in(0,t)} \left\{ e^{-\lambda t} \int_0^t e^{\lambda s} e^{-\lambda s} \left\| h_1(\cdot,s) - h_2(\cdot,s) \right\|_1 ds \right\} \\ &\le C_{10} \lambda^{-1} \left\| h_1 - h_2 \right\|_* \end{aligned}$$

Then by the compressed map theorem, Λ had unique fixed point, $\Lambda\hat{h} = \hat{h}$, we take $\hat{h} = p^u$, $p^u \in H$,is the solution of the system $(P)$and obviously bounded consistent in u,the proof of （3.15） is similar the proof of (3.3).

$$\begin{aligned} &\int_0^A \left| p^{u_1}(a,t;S^{u_1}) - p^{u_2}(a,t;S^{u_2}) \right| da \\ &\le \int_0^t b(t-a;S^{u_2}) \left| \Pi(a,t,a;S^{u_1}) - \Pi(a,t,a;S^{u_2}) \right| da \\ &+ \int_0^t \Pi(a,t,a;S^{u_1}) \left| b(t-a;S^{u_1}) - b(t-a;S^{u_2}) \right| da \\ &+ C_3 \int_t^A \left| \Pi(a,t,a;S^{u_1}) - \Pi(a,t,a;S^{u_2}) \right| da \\ &\le C_7 \int_0^t \left| \Pi(a,t,a;S^{u_1}) - \Pi(a,t,a;S^{u_2}) \right| da \\ &+ \int_0^t C_8 \left(\int_0^{t-a} \left| S^{u_1}(s) - S^{u_2}(s) \right| ds + \left| S^{u_1}(t-a) - S^{u_2}(t-a) \right| + \left\| u_1(t) - u_2(t) \right\|_{L^\infty(0,T)} \right) da \\ &+ C_1 C_3 \int_t^A \int_0^t \left| S^{u_1}(s) - S^{u_2}(s) \right| ds da \\ &\le (C_1 C_4 C_7 + C_4 C_8 T + C_1 C_3 C_4 A + C_4 C_8) \int_0^t \left\| p^{u_1}(a,s) - p^{u_2}(a,s) \right\|_{L^1(0,A)} ds \\ &+ C_8 T \left\| u_1(t) - u_2(t) \right\|_{L^\infty(0,T)} \end{aligned}$$

By Bellman inequality, we get that

$$\left\| p^{u_1}(\cdot,t;S^{u_1}) - p^{u_2}(\cdot,t;S^{u_2}) \right\|_{L^1(0,A)}$$
$$\le C_8 T \left\| u_1(t) - u_2(t) \right\|_{L^1(0,T)} \exp \int_0^t (C_1C_4C_7 + C_4C_8T + AC_1C_3C_4 + C_4C_8)ds \quad = C_{11} \left\| u_1(t) - u_2(t) \right\|_{L^\infty(0,T)}$$

which $C_{11} = C_8 T e^{(C_1C_4C_7 + C_4C_8T + AC_1C_3C_4 + C_4C_8)T}$

The following we prove that

$$\left\| p^{u_1}(\cdot,t) - p^{u_2}(\cdot,t) \right\|_{L^\infty(Q)} \le C_{12} \left\| u_1(t) - u_2(t) \right\|_{L^\infty(0,T)}$$

When $a \ge t$, by (3.15), we get that

$$\left| p^{u_1}(\cdot,t;S^{u_1}) - p^{u_2}(\cdot,t;S^{u_2}) \right| \le p_0(a-t) \left| \Pi(a,t,t;S^{u_1}) - \Pi(a,t,t;S^{u_2}) \right|$$
$$\le C_1C_3C_4 \int_0^t \left\| p^{u_1}(a,s) - p^{u_2}(a,s) \right\|_{L^1(0,A)} ds$$
$$\le C_1C_3C_4C_{11}T \left\| u_1(t) - u_2(t) \right\|_{L^\infty(0,T)}$$
$$= C_{12} \left\| u_1(t) - u_2(t) \right\|_{L^\infty(0,T)}$$

which $C_{12} = C_1C_3C_4C_{11}T$.

In the same way, when $a < t$ we get that

$$\left| p^{u_1}(\cdot,t;S^{u_1}) - p^{u_2}(\cdot,t;S^{u_2}) \right| \le C_{12} \left\| u_1(t) - u_2(t) \right\|_{L^\infty(0,T)}$$

Consequently

$$\left\| p^{u_1}(\cdot,t) - p^{u_2}(\cdot,t) \right\|_{L^\infty(Q)} \le C_{12} \left\| u_1(t) - u_2(t) \right\|_{L^\infty(0,T)}.$$

References

1. Gurtin, M.E., MaCamy, R.C., Hoppensteadt, F.: Nonlinear Age-dependent Population Dynamics. Arch. Rat. Mech. Anal. 54, 281–300 (1974)
2. Webb, G.F.: Theory of Nonlinear Age-dependent Population Dynamics. Pure and Applied Mathematics. Dekker, New York (1985)
3. Chan, W.L., Guo, B.Z.: Global Behavior of Age-dependent Logistic Population Models. J. Math. Biol. (28) (1990)
4. Anita, S.: Optimal Harvesting for a Nonlinear Age-dependent Population Dynamics. J. Math. Anal. Appl. (1998)
5. Anita, S., Iannelli, M., Kim, M.Y., Park, E.L.: Optimal Harvesting for Periodic Age-dependent Population Dynamics. J. Appl. Math. 58(5), 1648–1666 (1998)
6. Gurtin, M.E., Murphy, L.F.: On the Optimal Harvesting of Persistent Age-structured Populations. J. Math. Biol. 13 (1981)
7. Anita, S.: Analysis and control of age-dependent population dynamics. Kluwer Academic Publishers, Dordrecht (2000)

Mining Max Frequent Patterns over Data Streams Based on Equal Weight Clique

Zhufang Kuang[1], Guogui Yang[2], and JunShan Tan[1]

[1] School of Computer and Information Engineering
Central South University of Forestry & Technology
Changsha, China
[2] School of Computer
National University of Defense Technology
Changsha, China
{Zfkuangcn,ggyang}@nudt.edu.cn
tan_junshan@yahoo.com.cn

Abstract. Graph theory is introduced to model the problem of frequent pattern mining over data stream. The equal weight clique is proposed in this paper. The problem of mining max frequent pattern is transformed into the problem of solving max equal weight clique. A max frequent pattern mining algorithm EWCFPM which based on equal weight clique is proposed in this paper. In order to decease the processing time, we design pruning strategy. The IBM synthesizes data generation which output customers shopping a data are adopted as experiment data. The EWCFPM algorithm not only has high precision for mining frequent patterns, but also has high performance for data processing.

Keywords: data streams, data mining, max frequent pattern, equal weight clique.

1 Introduction

Recently, along with the development of compute network and communications technology, a new data model—data stream appear. Such as, network stream surveillance, web log analysis, wireless sensor network, traffic real-time surveillance and so on. The frequent pattern mining over data stream which is an important data mining issue, become an important research issue recently [1].

The issue of mining frequent pattern has attract many attention in recent years, and there are some research production [2-8].

C.Giannella[3] proposed the FP-stream algorithm, which integrate FP-Growth and incline time window model. Chang and Lee[4] propose a frequent pattern mining algorithm estDec which adopts the time decay measure to decease the support of old pattern and distinguish the pattern of recent transaction and the pattern of past transaction. Leung and Khan[5] propose a frequent pattern tree DStree to maintain pattern in the time window over data stream. Each node in DStree has one more counter for counting the support of the item in recent data stream. LI Guo-Hui[6]

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 155–161.
springerlink.com

proposed the MSW algorithm, which integrates time decay model and sliding window model mining frequent pattern for data stream. The MSW algorithm not only could get frequent pattern in any size sliding window, but also delete outdate and non frequent pattern in order to decease the memory requirements. WU Feng[7] proposes the DFPMiner algorithm based on landmark window model and time decay model. Yang Bei[8] propose a Top-K frequent item dynamic increment approximate algorithm TOPSIL-Miner which based on landmark window model over data streams.

The work of this paper is different from the recent work. The graph theory is introduced in this paper. We model the problem of mining frequent pattern through graph theory. The adjacency matrix is used as the synopses structure. The concept of equal weight clique is proposed in this paper. The problem of mining max frequent pattern is transformed into the problem of solving the equal weight clique. A max frequent pattern mining algorithm EWCFPM which based on equal weight clique is proposed in this paper. In order to decease the processing time, we design pruning strategy. The IBM synthesizes data generation which output customers shopping a data are adopted as experiment data. The EWCFPM algorithm not only has high precision for mining frequent patterns, but also has high performance for data processing.

2 EWCFPM Synopses Structure

It suppose that $I=\{i_1,i_2,...,i_n\}$ is the set of item, $D=\{T_i\}$ is the set of transaction. $T_i=\{td_i,x_1,x_2,...,x_m\}$, the td_i is the unique identifier of transaction T_i, which increase by 1, the $\{x_1,x_2,...,x_m\}$ is the sub-set of I .The pattern $A\subseteq I$ is the set of item. The transaction T_i includes the pattern A, if and only if $A\subseteq\{x_1,x_2,...,x_m\}$ .If the number of item is r in pattern A, we call pattern A as r-pattern.

The graph theory is introduced to model the problem. $G=\{V,E,W\}$ .The V is the vertex set. The E is the edge set. The W is the weight of edge. In the modeling, the vertex denotes the item of item-set, and the edge of graph denotes that items appear in the same transaction, and the weight of edge denotes the number of that items appear in the same transaction. The adjacency matrix is used as store structure of graph as following:

$$\begin{bmatrix} w_{11} & w_{12} & w_{13} & \cdots & w_{1m} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2m} \\ w_{31} & w_{32} & w_{33} & \cdots & w_{3m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ w_{m1} & w_{m2} & w_{m3} & \cdots & w_{mm} \end{bmatrix}$$

Fig. 1. Adjacency matrix W

The w_{ij} denotes the number of the x_i and x_j appearing in the same transaction, $w_{ij}=W(x_i,x_j)$.

Definition 1 (complete sub-graph, clique): undirected graph $G=\{V,E\}$, if $U=\{V',E'\},V'\subseteq V,E'\subseteq E$ and any $u,v\in V'$, $(u,v)\in E$,we call U as the complete sub-graph of G. The complete sub-graph of G is the clique of G, if and only if there are not more large complete sub-graph of G which includes U.

Definition 2 (equal weight clique): undirected graph $G=\{V,E,W\}$. The W denotes the weight of edge, $U=\{V',E',W'\},V'\subseteq V,E'\subseteq E,W'\subseteq W$ .The U is equal weight clique of G, if and only if U is clique of G ,and any $e,e'\in E',W'(e)=W'(e')$. The $W'(U)$ denotes the weight of equal weight clique. We call the number of vertex of equal weight clique U as the level.

We integrate the above modeling method and the definition of equal weight clique. When users provide the support s, we transformed the problem of mining max frequent pattern into the problem of solving the equal weight clique of graph G.

3 EWCFPM Algorithm

The adjacency matrix is used as the synopses structure for storing the information of data stream. EWCFPM algorithm contains EWCPreserve and EWCQuery two parts. When new data stream is coming, EWCPreserve algorithm takes charge of maintaining the adjacency matrix. When users query the frequent pattern, EWCQuery algorithm takes charge of print the frequent pattern.

3.1 Maintaining Synopses Structure

EWCPreserve algorithm takes charge of maintaining the adjacency matrix. The fundamental of EWCPreserve: at the beginning, the element of adjacency matrix is equal to 0, $w_{ij}=0$, which means not any two item appear in the same transaction. When a new transaction $T_k=\{tid,x_1,x_2,...,x_n\}$ is coming, $w_{ij}=w_{ij}+1$. The algorithm is as following:

Algorithm 1: EWCPreserve Algorithm
Input: adjacency matrix W,
New transaction $T_k=\{tid,x_1,x_2,...,x_n\}$.
Output: adjacency matrix W.

```
1. N++;
2. for(i=1;i<n;i++)
3.    for(j=1;j<n;j++){
4.       w_ij = w_ij +1;
5.       }
```

3.2 Frequent Pattern Generation

EWCQuery algorithm takes charge of queries from users, and print frequent pattern. The fundamental of EWCQuery is: at any time, when users provide the support s.

Firstly, in order to decease the processing time, the no frequent pattern is pruned according to the support s. The $w_{ij} < N * s$ means that the number of item x_i and item x_j appears meantime is less than the support s. The pruning operation is $w_{ij} = 0$ concretely. Secondly, the function CliqueBacktrack is called. We use depth first search to solve the pruned adjacency matrix. Everyone of equal weight clique denote a max frequent pattern. The set of equal weight clique is the set of max frequent pattern. The Algorithm is as following:

Algorithm 2: CliqueBacktrack Algorithm
Input: adjacency matrix W.
Output: Frequent pattern set, The Count of frequent pattern.

```
1. void CliqueBacktrack(int **w,int i)
2. {// computing equal weight clique
3.    if (i > n) {//reach to leaf, get a Clique
4.     Count++;
5.     for (int j = 1; j <= n; j++)
6.        if (x[j]) printf(x[j]);
7.     return;
8.    }
9.   // Checking the vertex i whether link with current Clique
10.  int OK = 1;
11.  for (int j = 1; j < i; j++)
12.   if (x[j] && w[i][j] == 0) {
13.      //vertex i and j no adjacency
14.      OK = 0;   break;}
15.  if (OK) {//enter left sub tree
16.     x[i] = 1;
17.     CliqueBacktrack(a,i+1);
18.     x[i]=0;
19.  }
20.  x[i] = 0;
21.  CliqueBacktrack(a,i+1);
22.}
```

Algorithm 3: EWCQuery Algorithm
Input: adjacency matrix W, Support s.
Output: Frequent pattern set, The Count of frequent pattern.

```
1. Pruning the item which could not generate frequent pattern;
2. for(k=⌈N*s⌉;k<Max_w;k++)
3.   for(i=0;i<n;i++)
4.      for(j=0;j<n;j++){
5.           if (w[i][j]==k) ans[i][j]=w[i][j]
6.           else ans[i][j]=0;
          //solving max frequent pattern with the support more than s
7.            CliqueBacktrack(ans,0)
8.      }
```

4 Experimental Results

All experiment of this paper carry on a computer with Intel Core2 (TM) Duo CPU 2.2GHz, memory 3GB, and the operate system is Windows XP profession. The SWFPM algorithm implemented adopting Visual C++. We compare the precision, processing time, memory requirement of the DFPMiner algorithm with EWCFPM algorithm.

Experimental data sets are generated by IBM synthetic data generator [9]. The experiment adopts the T20I10D100K data est. T represents the average length of transaction of data set, I represents the average length of the potential frequent item. D represents the number of the transaction. The data set has 1K distinct items; the other parameter of data generator is default.

4.1 Comparison Precision

We compare the precision of EWCFPM and DFPMiner algorithm with different support s. Figure 2 show the result. We can see from experimental result, the precision of DFPMiner vary along with the change of support s. But, the precision of EWCFPM is still 100%. This is because DFPMiner is approximate algorithm, which adopts time decay model, and counts the support of pattern through time exponential decay function. The counting method is not precision compare with our method. Our algorithm distinguishes with DFPMiner algorithm. The adjacency matrix is used as the synopses structure for storing the information of data stream. Hence, our algorithm could mine the entire max frequent pattern.

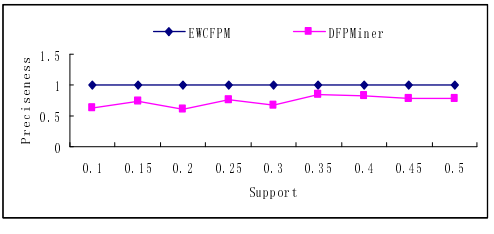


Fig. 2. Precision with different support s

4.2 Processing Time

We compare the processing time of EWCFPM and DFPMiner algorithm with different support s. Figure 3 show the result. We can see from experimental results. The processing time of EWCFPM and DFPMiner algorithm vary with the change of support s. The more the support s is, the less the processing time is. This is because the more the support is, the less the frequent pattern is, and the shorter the processing time also. The processing time of our algorithm is higher than the DFPMiner algorithm. But, when the support is more than 0.5, the processing time of our algorithm is less than the DFPMiner algorithm. This is because the effect of pruning strategy is excelled when the support s is more than 0.5.

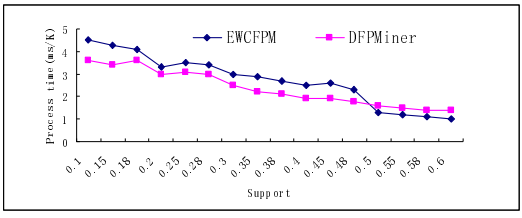


Fig. 3. Processing time with different support s

4.3 Memory Requirement

We compare the memory requirement of EWCFPM and DFPMiner algorithm with different support s. Figure 4 show the result. We can see from experimental results.

The memory requirement of DFPMiner algorithm changes with the support s. But the memory requirement of our algorithm keeps still no change. The memory requirement of synopses structure is $O(m^2)$, the m is the number of item of item set. In this experiment, the m=100. Hence, the memory requirement of our algorithm is 10000 store unit.

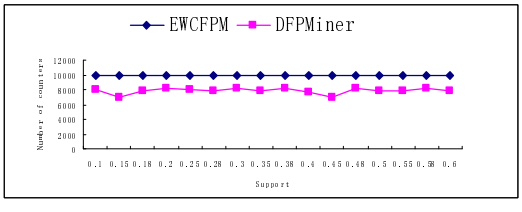


Fig. 4. Memory requirement with different support s

5 Conclusions

The graph theory is introduced in this paper. The problem of mining max frequent pattern is modeled as graph. The adjacency matrix is used to synopses structure for data stream. The equal weight clique is proposed in this paper. The problem of mining max frequent pattern is transformed into the problem of solving max equal weight clique, which is an innovation in this paper. A max frequent pattern mining algorithm EWCFPM which based on equal weight clique is proposed in this paper. The experiment results show that the EWCFPM algorithm not only has high precision for mining frequent patterns, but also has high performance for data processing. But decease the processing time of the BFFPM must be research farther.

Acknowledgment. This research was supported in part by the National High Technology Research and Development Program of China (863 Program) under grant No. 2008AA01A201, the National Natural Science Foundation of China under grant No 60873082, 60902044, the Planned Science and Technology Project of Hunan

Province of China under grant No 2009FJ3204.the Scientific Research Fund of Hunan Provincial Education Department under grant No 08B091and 08C943.

References

1. Babcock, A.K., Babu, S., Datar, M.: Model and issues in data stream systems. In: Popa, L. (ed.) Proc. of the 21st ACM SIGACT-SIGMOD-SIGART Symp. on Principles of Database Systems, pp. 1–16. ACM, Madison (2002)
2. Charikar, M., Chen, K., Farach-Colton, M.: Finding Frequent Items in Data Streams. In: Widmayer, P., Triguero, F., Morales, R., Hennessy, M., Eidenbenz, S., Conejo, R. (eds.) ICALP 2002. LNCS, vol. 2380, pp. 693–703. Springer, Heidelberg (2002)
3. Giannella, C., Han, J., Pei, J., Yan, X., Yu, P.S.: Mining frequent patterns in data streams at multiple time granularities. In: Data Mining: Next Generation Challenges and Future Directions, pp. 191–212 (2004)
4. Chang, J.H., Lee, W.S.: Finding recent frequent itemsets adaptively over online data streams. In: Lise, G., Ted, E.S., Pedro, D., Christos, F. (eds.) Proc. of the 9th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, pp. 487–492. ACM Press, Washington (2003)
5. Leung, C.K.S., Khan, Q.I.: DStree: A tree structure for the mining of frequent sets from data streams. In: Clifton, C.W., Zhong, N., Liu, J.M., Wah, B.W., Wu, X.D. (eds.) Proc. of the 6th Int'l Conf. on Data Mining, pp. 928–932. IEEE Press, Hong Kong (2006)
6. Li, G.-H., Chen, H.: Mining the Frequent Patterns in an Arbitrary Sliding Window over Online Data Streams. Journal of Software 19(10), 2585–2596 (2008)
7. Wu, F., Zhong, Y., Wu, Q.-Y.: Mining Frequent Patterns over Data Stream under the Time Decaying Model. Acta Automatica Sinica 36(5), 674–684 (2010)
8. Yang, B., Huang, K.: Mining Top-K Significant Itemsets in Landmark Windows over Data Streams. Journal of Computer Research and Development 47(3), 463–473 (2010)
9. `http://www.almaden.ibm.com/cs/projects/iis/hdb/Projects/data_mining/datasets/syndata.html#instructions`[CP/OL]

UAV Path Re-planning of Multi-step Optimization Based on LRTA* Algorithm*

Li Fu and Kun Zhu

School of Automation, Shenyang Aerospace University
37 Daoyi South Street, *Liaoning Province, China*
ffulli@yahoo.com.cn

Abstract. For the UAV with the flight environment of unknown threats in advance, the paper combine self-learning A* real-time algorithm with node expansion method based on Five-fork Tree, then application of multi-step optimizing search algorithm re-plan path. When the UAV's detectors detect unknown threats in flight environment, the algorithm modify the affected tracks according to new environmental information in order to get a new track. Finally through a numerical simulation demonstrates the effectiveness of the algorithm.

1 Introduction

Now, with the widespread use of unmanned aircraft, unmanned aircraft and unmanned probe, in the field of intelligent search, UAV path planning has get considerable attention, when we have the complete and precise environmental information, we use global planning to get a optimal trajectory from the beginning to the end. But in reality, it is difficult to obtain the complete and precise global information. In real battlefield environment, the enemy arrange many air defenses, and some of them are known in advance, and the others are detected by UAV's radar during the flight, the threat is generally unknown random threats[1-3] (pop-up). At this time the previous path planning can not meet the requirements, and the UAV need to deal with new threats during the flight, UAV need several re-planning to get optimal trajectory relying on the information detected by radar [4-5]. The paper combine self-learning A* real-time algorithm with node expansion method based on Five-fork Tree according to the characteristics of UAV, then application of multi-step optimizing search algorithm do partial correction for appearing pop-up treatments during the flight.

2 Description of Relevant Issues

A. Self-learning A* Real-Time Algorithm

Self-learning A* real-time algorithm[6-7] is proposed a heuristic search algorithm by E. Korf[8], which plan optimal trajectory through the establishing and updating of the

* This work is partially supported by the Open Research Fund of State Key Laboratory under Grant 2009SY01 and the National Natural Science Funds under Grant 61074090.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 163–169.
springerlink.com

cost of assessment form each state node to the target node. Planning phase and implementation phase alternating in practical application process, and it meets the needs of real-time planning by adjusting the relative proportions of planning time and execution time and very effective to static target path search[9-10].

Before the algorithm search, at first, we gave the value of assessment for each state node in the state list, if the assessed value is not feasible, then it is set to zero. After the exploring of the search space in order to close the true value through studying the cost of each state point. We assume that x is the current position of the search process, and h(x) is the cost of assessment which is path cost from the current position x to the target point.

LRTA* finish optimal path search by repeating the following steps:

(1) To each adjacent nodes of the current node, calculating $f(x') = h(x') + k(x, x')$, where $k(x, x')$ is the value of track costs from current node x to neighbor y. $h(x')$ is estimate of track costs from neighbor node to target node- heuristic information, which is a key of LRTA*.

The paper X is:

$$h(x') = w_1 f + w_2 d \ . \tag{1}$$

Where, f is costs of threat, d is costs of distance, w_1 is weigh of f, w_2 is weigh of d, $w_1 + w_2 = 1$. Here, $h(x')$ took into account two indicators, which is the threats and the distance, and the simulation results are more close to reality.

(2) Move to the child node of the smallest value, and the child node as the current node.

(3) The track cost $h(x)$, which is from current position x to the target point, using the smallest value of $f(x')$ instead.

Where $f(x')$ is the cost value of evaluation function from the current node x pass its neighbor child nodes to the target nodes. Because each path which reaches the target point must pass the adjacent nodes, and evaluation value $h(x')$ is less than the true cost value, the value of real costs is certainly not less than the smallest value of $f(x')$, updated x will be closer to the true cost value.

Through the above analysis, all of the nodes are positive, so it must have the path form each node to the target nodes. To the point, self-learning A* real-time algorithm meets the requirements of completeness. If the initial assessed value of the cost is less than the value of real costs, eventually converge to the costs of path value in accordance with the optimal path to reach the target point by repeated search and updated the costs value.

B. Five-Fork Tree Node Expansion

In the traditional A* algorithm process of node expansion[11-12], the way of node expansion is also using the 8-neighborhood or the 24、48-neighborhood node, which is high time complexity and greatly increasing the number of nodes expanded. The

paper use node expansion method based on Five-fork Tree under meeting the kinematics equation of UAV [13].

According to planning assumptions, and assuming UAV level cruise flight in a certain height, simplify the equations of UAV's motion and establish UAV's kinematics equations on the ground coordinate system[14]:

$$\begin{aligned} \theta_{k+1} &= \theta_k + \theta_0 \Delta \\ x_{k+1} &= x_k + S_0 \cos(\theta_{k+1}) \\ y_{k+1} &= y_k + S_0 \sin(\theta_{k+1}) \end{aligned} \quad (2)$$

Where θ_k is the flight path azimuth of current aircraft in level flight; θ_{k+1} is the flight path azimuth of aircraft in the next moment; θ_0 is the smallest unit of angle changes when the course of aircraft changes; Δ is the size of course angle when the course of aircraft changes; S_0 is the step of route planning. $A(x_k, y_k)$ and $B(x_{k+1}, y_{k+1})$ are the coordinates of the current and the next route waypoint of aircraft coordinates. Moreover, the minimum turning radius of the UAV is:

$$R_{\min} = v_{\min}^2 / g * \sqrt{n_{y\max}^2 - 1} . \quad (3)$$

Where, $v_{\min}$ is the minimum flight speed of the UAV, $n_{y\max}$ is the maximum normal overload of the UAV. According to the minimum turning radius and the step of route planning, the maximum heading angle is:

$$\theta_{\max} = \arcsin(S_0 / (2R_{\min})) . \quad (4)$$

Next, we use node expansion method which based on Five-fork Tree under not exceeding the maximum heading angle. For example, assuming the maximum steering angle of UAV is 90 and is not to backward maneuvering, and any node up to flight the nodes of the follow five the direction and format the five-tree structure.

Assuming the entrance angle of a node is α, then the course angle of flowing nodes node may only be the one of α、$\alpha \pm 45$、$\alpha \pm 90$.

We will put the intersection of the grid as the node to be expanded in the planning area, so a problem resulting which are nodes by using the aircraft kinematics equations to obtain are not necessarily strict in the intersection of the grid, so we need to put the nod to count the intersection of the grid as expanded nodes.

Steps to expand the child nodes:

(1) The expanded nodes are whether or not in the intersection of the grid, if it is yes, directly as a node to be expanded.

(2) If the expanded nodes are not the intersection of the grid, determine which grid is located in and determine four vertices of the grid and determine if each vertice is in the danger zone, if it is yes, do not as the expanded nodes.

(3) To compute the distance form remaining vertice of the grid to the target node and select the nearest vertex from the target node as to be expanded.

When it has n nodes, the traditional number of the node expansion is $8^1+8^2+\ldots+8^n$, while the number of node expansion based on Five-fork Tree is $5^1+5^2+\ldots5^n$, so decreasing node expansion and improving algorithm speed are very important for real-time route planning of the UAV.

3 Multi-step Optimization Algorithm Based on LRTA* Algorithm

The paper combine self-learning A* real-time algorithm with node expansion method based on Five-fork Tree, then application of multi-step optimizing search algorithm re-plan path.

When UAV flights along the initial track, if the radar of UAV detects the section of track falling into the danger zone of enemy, then fix track form the next track of the track. When the track is fixing, assuming number of prediction steps is N, it needs to remove the N-1 steps of the before and after in the section of track except removing the dangerous section of track, then application of multi-step optimizing search algorithm re-plan path in order to advanced to avoid threats and ensure re-planning track close to the optimal track [15].

The essence of multi-step optimizing search algorithm express that, the initial track is a straight line segment from A to G, when the UAV flights from A to B, the new threat is appearing in the initial track, and if prediction steps N is three, we fix the track from the point B, which pull in the straight line of N times the step length from the point B in order the distance can be nearest from the end of the line segment to the target point and get N-step planning track section which is named BI. The UAV flights from B to H according to the planned track, re-planning to generate a new track segment is HK, which is planning N-steps and implementing one step, planning and implementation are synchronization and meet the real-time requirements of re-planning.

4 Simulation Results and Analysis

Here, we use this approach to simulation, under the condition of the real-time of re-planning and UAV constraints, it supposed that we set the planning area of a size of 200*200(km), the step of UAV's route planning is S_0=2.5km, and the radar detection range of UAV is 20km, flight speed is 0.15km/s.

Figure 1 shows a size of 20*20(km) planning area after treatment, we regard all of the treated threats as threat circle, the starting point coordinates of UAV is (0,0) which is expressed by'○', the end point coordinates of UAV is (180,170) which is expressed by'✳'.

The reference trace shows in the blasé-spot solid line of figure 2, when UAV flight following the reference trace in the planning area, the radar of UAV detects a new threat which is expressed by threat circle, its position shows in figure 2.

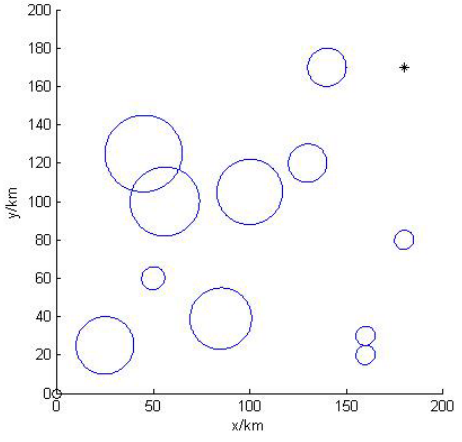


Fig. 1. Threat distribution

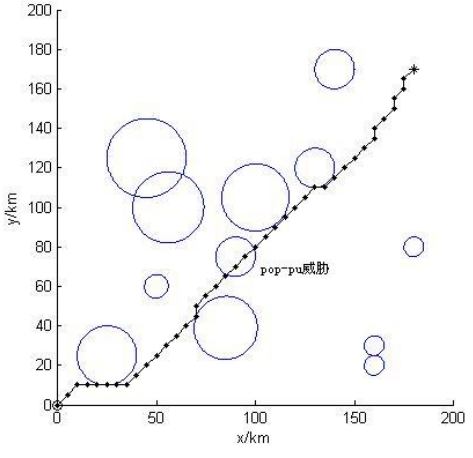


Fig. 2. Initial trace

The threat just located on the reference trace in fig. 2. If UAV flight following the reference trace, it will clearly fall into the danger zone, so we need to re-plan the reference trace during the flight in order to avoid new threats. The UAV's re-planning trace which is application of this algorithm show in the blasé-spot solid line of figure 3.

The simulation results show that there is angular trace which is not conducive to flight; it needs smooth re-planning trace, after a smooth flight trace show in figure 4.

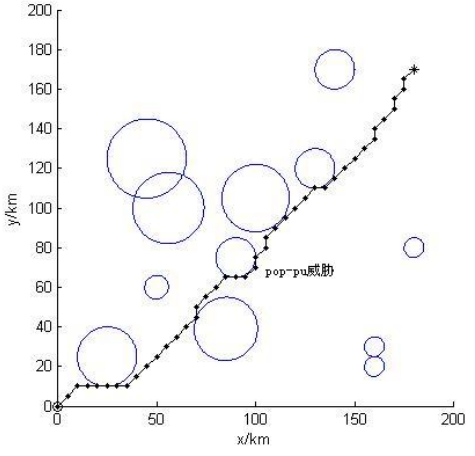


Fig. 3. Re-planning trace

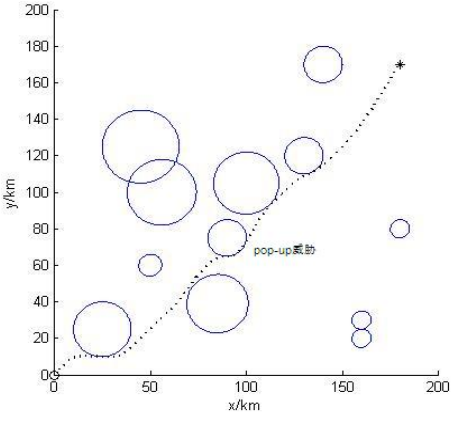


Fig. 5. Smooth flight trace

5 Conclusion

For the security requirements of the trace planning of UAV in this paper, for the suddenly the threat of emerging during the flight, the paper combine self-learning A* real-time algorithm with node expansion method based on Five-fork Tree, then application of multi-step optimizing search algorithm re-plan path. A special feature in this paper is combining node expansion method based on Five-fork Tree with UAV's kinematics equations, which are decreasing node expansion and improving algorithm speed and meeting the constraint of UAV's maneuvering performance. Simulation results show that the algorithm can effectively avoid the pop-up threat and solve the problem of UAV flight path re-planning.

References

1. Zheng, C., Ding, M., Zhou, C., Li, C.: A re-planning algorithm for aircraft track online. J. Huazhong Univ. of Sci. & Tech. 31(2), 90–93 (2003)
2. Li, C., Zhou, C., Ding, M.: Vehicle real-time three-dimensional route planning in the dynamic environment. Journal of a Stronautics 24(1), 38–42 (2003)
3. Ba, H.: The study of UAV's path planning. Northwestern Polytechnical University, Xian (2004)
4. Brown, T., Doshi, S., Jadhav, S., et al.: Test bed for a wireless network on small UAVs. In: Proc. of AIAA 3rd Unmanned Unlimited Technical Conference, pp. 20–23 (2004)
5. Ren, M., Shen, L.: A Fast Trajectory Planning Algorithm for Air Vehicle Based on Hopfield Artificial Neural Networks. Tactical Missile Technology 1(5), 45–48 (2007)
6. Li, J., Sun, X.: Route replanning's method for unmanned aerial vehicles based on multi-step optimizing search algorithm. Systems Engineering and Electronics 31(10), 2510–2513 (2009)
7. Korf, R.E.: Real-time heuristic search. Artificial Intelligence 42(2-3), 189–211 (1990)
8. Zhang, Z., Wang, Y.: Resoure Allocation Using Timed Petri-nets and Heuristic Search. Journal of Beijing Institute of Technoloy 9(2), 148–154 (2000)
9. Rathbun, D., Kragelund, S., Pongpunwattana, A., et al.: An evolution based path planning algorithm for autonomous motion of a UAV through uncertain environments. IEEE Trans. on Aerospace and Electronic Systems 2(8D2), 1–12 (2002)
10. Yang, G., Kapila, V.: Optimal path planning for unmanned air vehicles with kinematic and tactical constraints. In: Proc. of the 41st Conference on Decision and Control, vol. 2(1), pp. 1301–1306 (2002)
11. Zhou, M., Zhou, J.: Algorithm of 4D Real-time Path Planning based on A* Algorithm. Fire Control and Command Control 33(8), 98–101 (2008)
12. Song, J., Li, K.: 3D Route Planning Algorithm for Long Range Missiles Based on A* Algorithm. Transactions of Beijing Institute of Technology 27(7), 613–617 (2007)
13. Zeng, J., Zhou, D., Ma, Y.: UCAV Path Planning Based On search Algorithm with Five-fork Tree. Missile and Guidance 25(3), 266–268 (2005)
14. Li, J., Sun, X.: A Route Planning's Method for Unmanned Aerial Vehicles Based on Improved A-Star Algorithm. Acta Armamentar 29(7), 788–792 (2008)
15. Xiao, Q., Gao, X.: The study of UAV's local path replanning algorithm. Flight Mechanics 24(1), 85–88 (2006)

Research and Implementation on the Three-Dimensional Modeling of Interior Building Component

Hongjuan Wang[1,2] and Xinming Lu[1,2]

[1] College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao, 266510, China
[2] Shandong Lionking Software Company Ltd., Taian, 271000, China

Abstract. To solve the existing problems of the interactive modeling of buildings, a fast, automatic and detailed modeling algorithm is introduced. Through customizing property library, drawing a plane figure of the building and building components, and configuring properties for them, the algorithm could automatically complete the whole modeling process. The three-dimensional model can be automatically modified with the property information of GIS. The algorithm is widely applied into LionKingSoft digital city system and meets the requirements of rapid modeling of the large group of buildings. The experiment results indicate the algorithm is rapid, accurate and practical.

Keywords: building, GIS, three-dimensional modeling, digital city.

1 Introduction

With the applications of virtual reality**,** scientific visualization and computer simulations, the three-dimensional (3D) modeling of digital city has developed rapidly and become a hot topic in the current research. How to implement the 3D model of the building rapidly to meet the extensive application in these areas is still a question. In the system of 3D GIS, the ground objects model is another expression of ground object which is independent from the terrain, and it usually has large quantities and much difference on the complexity (Chunyu et al., 2001). For the ground objects which require the rich detailed modeling, we can adopt the individual method, while the individual method is not practical for the large groups of buildings (Gang et al., 2002). Hongyuan et al., 2005 analyzed the computer modeling methods of interior architecture combined with 3ds max and gave an in-depth exploration, but every method needs manual interactive operation and requests the users of the highly specialized skills**.** Gang et al., 2008 applied the technique of texture features analysis into the modeling of the polygon, and introduced a surface subdivision algorithm based on the terrain elevation value to construct the multi-resolution and virtual model of buildings. The algorithm was simple and it was easy to realize, but the side of the building model was lack of the texture mapping, there were some defaults in the sense of reality of the building. With the emergence of three-dimensional

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 171–178.
springerlink.com

modeling software, it was popular to use AutoCAD, 3D Max and Maya to aid three-dimensional building modeling by a means of human-computer interaction. However, the modeling process was complicated, especially for the large groups of buildings of the large areas, the method of interactive modeling was laborious, time-consuming and not practical, and it was very particular about the skills of the users (Xidao et al., 2008).

As can be seen from the above analysis, the current software of the building modeling has some shortcomings such as inefficiency, manual interventions and complicated operations. In order to meet the requirements of the construction of digital city, a new modeling algorithm based on the parameters and attributes of GIS is introduced. Because different objects have different geometrical characteristics, in the process of modeling, we use different methods for the different objects to ensure the vivid model, and integrate them into the LionKingSoft digital city system. At the same time, we sufficiently consider seamless match of the building model and the terrain, and truly realize the visualization of virtual buildings. LionKingSoft digital city system integrates three-dimensional GIS, CAD, database, visualization of three-dimensional model, has the functions of the collecting, managing, retrieving and displaying the three-dimensional model data.

In this paper, the modeling is based on the attribute library. The users only draw the plane figure of the buildings and configure their attributes. The algorithm automatically completes the process of 3D modeling without manual operation. This method reduces the data storage and achieves the goal of the detailed modeling accurately. The process of building modeling is similar to a real project. Firstly all the main parts of building are created, such as walls, doors and windows, balconies and stairs. Other models of building and the furniture could directly be imported from the model library, and the detailed geometry model is created by using LionKingSoft 3D solid modeling and Boolean operation system.

2 The GIS Attribute LIBRARY

LionKingSoft digital city system has the advanced functions of general GIS. In the process of modeling, we mainly use the customization of attribute library, the attribute management and attribute association to define the structure of library and store the parameters of GIS attribute. The customization of attribute library, shown in the Fig 1, defines the structures of attribute library. We can use it to define the object types, the attribute library structures and the dictionary formats. All defined objects can be conveniently incorporated into the contents of GIS management. The customization of attribute library is the key part of modeling system based on GIS, and all the information of modeling solid is stored into the attribute library. Along with the technical advance in GIS, building modeling function of establishing modeling automation software based on GIS is more and more powerful, which can rely on the GIS spatial database, shorten the modeling period, increase modeling quantity and economize modeling cost.

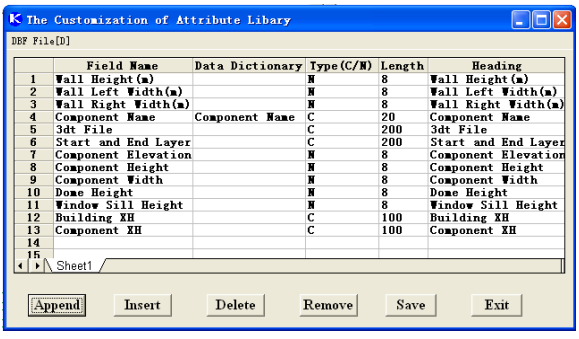
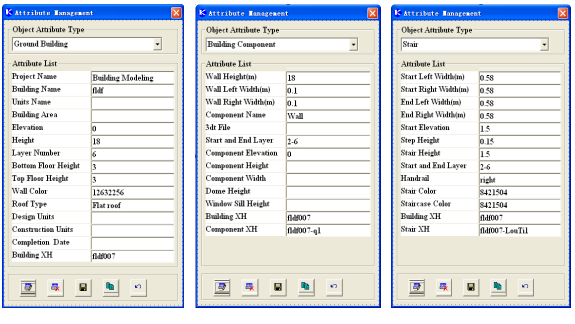


Fig. 1. The customization of attribute library

The attribute management is shown in the Fig 2. In detailed modeling system of the buildings, attribute libraries mainly include the ground building, building component and stair. The attribute library of ground building, shown in the Fig 2(a), mainly includes building common attributes. The field Building XH is used to uniquely identify the ground building. The attribute library of building component, shown in the Fig 2(b), mainly includes individual information of building component. The field Component XH is used to uniquely identify the building component. The attribute library of stair, shown in the Fig 2(c), mainly includes the information of stairs. The field Stair XH is used to uniquely identify the stair. The field Building XH of the building component and the stair is equal to the field Building XH of the ground building.



(a) The ground building (b) The building component (c) The stair

Fig. 2. The attribute management

For building components with the same attribute, the system provides the function of attribute association shown in the Fig 3, which can ensure the uniqueness of attribute in the attribute library. At the same time, for the components with the same attributes and sizes, the system provides the tools such as copy, mirror and array, which can help to reduce the input operation.

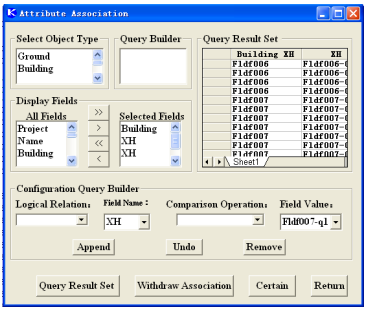


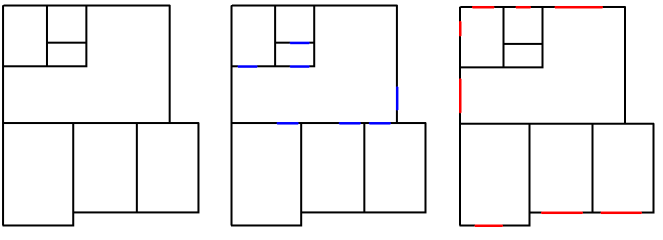
Fig. 3. The attribute association

3 The Description of the Main Algorithm

After completing the customization of attribute library, we can draw the plane figure of building components and configure the attributes for them. Through reading the configured attributes, the programs automatically complete the operation which converts the two-dimensional (2D) plane figure into the 3D building model. Firstly, we generates the main body of 3D wall using the algorithm, and then we edits the body of 3D wall according to the number of the doors and windows, finally we imports the doors and windows to complete the whole modeling. In addition, we could use LionKingSoft solid modeling and Boolean operation system to create the models of other subsidiary building components. The concrete step is shown below:

Step 1: Creating the layer. Before drawing a plane figure, we firstly create the six layers: the ground building, wall axis line, door axis line, window axis line, floor and stair.

Step 2: Drawing a plane figure of the building and building components. In the ground building layer, we draw the outline of a building. In the other corresponding layer, we draw other building components, such as wall, door and window. The wall axis line is shown in the Fig 4(a). The related position of the door in the wall is shown in the Fig 4(b). The related position of the window in the wall is shown in the Fig 4(c).



(a) The wall axis line (b) The door axis line (c) The window axis line

Fig. 4. The plane figure of the wall, door and window

Step 3: Configuring the attribute. For the components with the different attributes, we configure the different attributes for them. For the building components with the same attributes, we use the function of attribute association to complete the configuration of attribute. The plane figure with attributes is saved as the file format of "Net" of LionKingSoft.

Step 4: Reading the .net form attribute file of LionKingSoft. According to the attribute table name and the serial number, the corresponding attribute library is connected and the attributes are read. The different components are stored into different tables according to the layer number of components.

Step 5: Generating the 2D wall line according to the left width and right width of axis line. The left point and the right point of each line can be calculated through this step. The wall axis line has length as well as direction, and its data structure mainly includes the start point, end point, left width, right width as well as left start point, left end point, right start point, right end point. According to the width of the line, the left and right edges of the line can be calculated. Here are the examples which display three conjunction relations shown in the Fig 5(b). P_0 is the start point and is adjacent to line e_1, line e_2 and line e_3. P_1, P_2 and P_3 are the end point. We can use the vector method to calculate the left point and right point of P_0. Let V_1 be a unit vector of the vector $\overrightarrow{P_0P_1}$, V_2 be a unit vector of the vector $\overrightarrow{P_0P_2}$ and V_3 be a unit vector of the vector $\overrightarrow{P_0P_3}$. According to the formula (1), (2), (3), P_{01}, P_{02} and P_{03} are calculated.

$$P_{01} = P_0 + V_1 \times W_6 + V_3 * W_1, \tag{1}$$

$$P_{02} = P_0 - V_3 \times W_2, \tag{2}$$

$$P_{03} = P_0 + V_2 \times W_5 + V_3 * W_4, \tag{3}$$

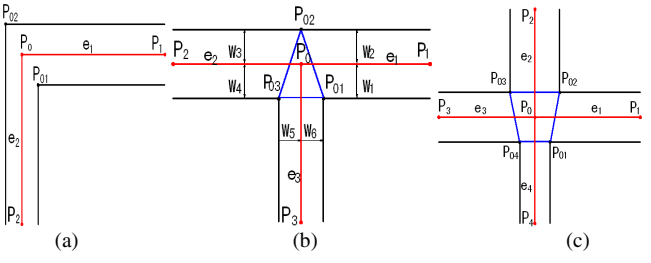


Fig. 5. The conjunction relationship of the point and line

The two left points and two right points are connected into one line respectively. In this step, using vector method, the algorithm does not need intersecting operation of line and line, which avoids the influence of operation error, and improves the robustness of the algorithm.

Step 6: Generating the 3D wall body. According to the wall height, the left lines and right lines are stretched, and the corresponding lines are connected into the faces. If there are no doors or windows in the wall, the two left lines and the two right lines are directly connected into one face respectively. On the contrary, for the doors or

windows in the wall, the steps are taken as follow: according to the attributes of the wall, door and window, the points of the doors and windows are calculated and the points are added into the point table and edge table, at the same time, the number of door or window is recorded. For the wall with the door or window, using the Delaunay triangulation algorithm, all the points are triangulated. The wire frame of 3D wall body is shown in the Fig 6 (a). The 3D wall body is shown in the Fig 6 (b). The wall body with the hole of door and window is shown in the Fig 6 (c).

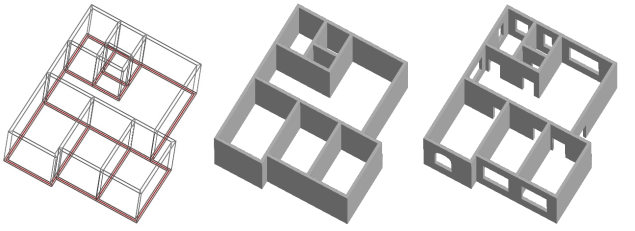


Fig. 6. The 3D wall body

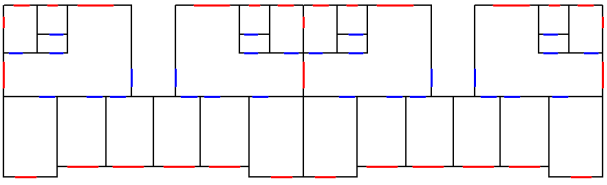
4 The Instanc

In the end, a sample of a six-storey building model is shown in the Fig 7. We draw a plane figure shown in the Fig 7(a), it takes about five minutes to draw it, and then it takes ten minutes to configure their attributes shown in the Fig 7(c). Finally, it takes one minute to generate the 2D wall lines shown in the Fig 7(b) and the 3D model shown in the Fig 7(d) which is varied with attributes and has real size with thickness. Comparing with the traditional software modeling such as 3d max, the building modeling based on GIS has the following main advantages:

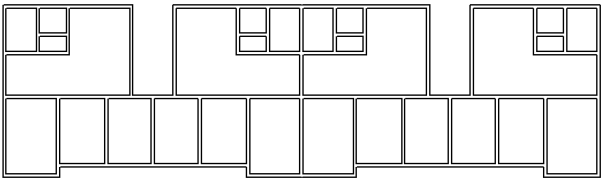
(1) It is rapid. We only use simple command to draw a plane figure of the building components and configure their attributes. For the components with the same attributes and sizes, we may use the function of attribute association and the tools such as copy, mirror and array to reduce the input operation and save more time. According to GIS attributes, the whole process of modeling is completed and the 3D model with the attribute information is generated through this algorithm without manual interventions.

(2) It is convenient for modification. Using the 3d max, the modification of 3D model is time-consuming, and it needs a lot of manual operations. Using this algorithm, only the 2D axis lines and their attributes are modified to complete the whole change of 3D model. For example, through modifying the windows shown in the Fig 7(d) and changing the 2D attribute of windows from the Fig 7(c) to the Fig 7(e), the 3D model is automatically modified. The modified model is shown in the Fig 7(f).

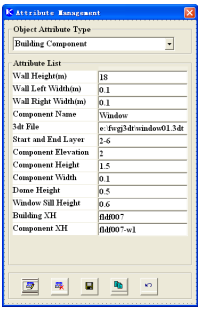
(3) Based on the GIS spatial database, the algorithm shortens the modeling period, increases modeling quality and economizes modeling cost and storage space. It can be applied for the large groups of buildings modeling. The interior model of six-storey building is shown in the Fig 8.

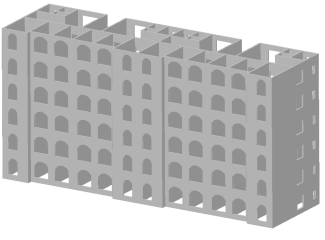


(a) The plane figure of wall, door and window

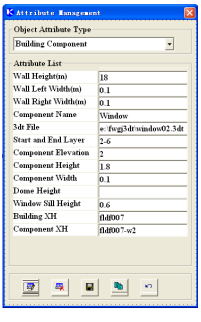


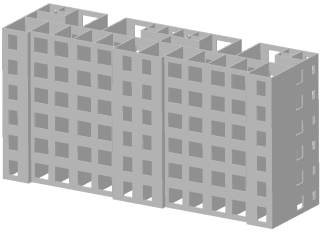
(b) The 2D wall line with width





(c) The attribute of window (d) The wall body with door and window





(e) The attribute of window (f) The wall body with door and window

Fig. 7. The 3D modeling of six-storey building

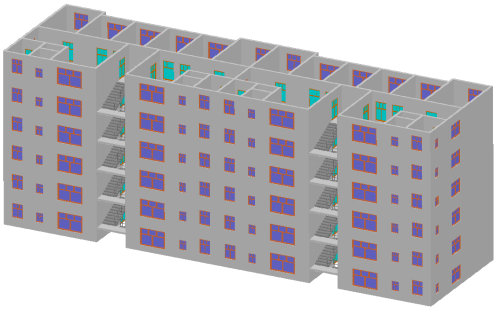


Fig. 8. The interior model of six-storey building

5 Conclusion

This paper mainly analyzes the building modeling method based on GIS, studies the characters of the GIS modeling technology. On this basis, a set of intelligent digital city system that integrates three-dimensional GIS, CAD, database, visualization of three-dimensional model are developed. The system has the functions of the collection, management, retrieval of the three-dimensional GIS and the function of displaying three-dimensional model data. It can meet the needs of three-dimensional modeling based on GIS. The algorithm of this paper is widely used in LionKingSoft digital city system and can be applied for the large groups of buildings modeling. The instance results indicate the algorithm is rapid, accurate and practical.

Acknowledgment. The study described in this research was supported by National High-Tech Research and Development Program of China (863 Program) grants 2009AA062700.

References

1. Chunyu, L., Yuqi, B., Ge, C.: The Construction Method of Geometric Objects In 3D Terrain. Computer Applications 5, 29–30 (2001)
2. Gang, W., Gang, C., Xiong, Y.: Research on Geometry Modeling of Terrain Features in Virtual City. Acta Geodaetica et Cartographica Sinica 31, 60–65 (2002)
3. Hongyuan, L., Chikun, C., Gao, W.: Study of Computer Modeling Methods in Interior Architecture. Journal of Engineering Graphics, 23–28 (2005)
4. Gang, X., Rubo, G., Tingjin, L.: Modeling and Real-time Rendering Technology of Multi-resolution 3D Buildings. Computer Engineering 34, 260–262 (2008)
5. Xidao, L., Long, Y., Yuxiang, X.: Advances in Study of 3D Modeling. Computer Science 35, 208–210 (2008)

Preplanned Recovery Schemes Using Multiple Redundant Trees in Complete Graphs

Wei Ding[1] and Yi Shi[2]

[1] Department of Basic Education Zhejiang Water Conservancy and Hydropower College Hangzhou, Zhejiang Province 310000, China
[2] Department of Computing Science University of Alberta Edmonton, Alberta T6G 2E8, Canada
dingweicumt@163.com, ys3@cs.ualberta.ca

Abstract. Zehavi and Itai have suggested a conjecture that implies we can use k disjoint trees to achieve up to k-1 simultaneous edge failures restoration for k-edge connected graphs or up to k-1 simultaneous node failures restoration for k-node connected graphs. In this paper, we firstly point out that a complete graph with n nodes is both (n-1)-edge connected and (n-1)-node connected. This implies that, provided that Zehavi's conjecture is right, we can use n-1 disjoint trees for restoration. Although we have not demonstrated the correctness of Zehavi's conjecture, we indeed construct two types of recovery schemes using multiple redundant trees for complete graphs, Hamilton-based recovery scheme and star-based recovery scheme, based on two types of decomposition of complete graphs. In complete graphs with n nodes, the latter can recover from any up to n-2 simultaneous link or node failures, and the former can recover from any up to n-2 simultaneous link or node failures if n is odd and any up to n-3 failures if n is even.

Keywords: Restoration, Redundant trees, Complete graph, Decomposition.

1 Introduction

Protection and restoration in high-speed networks are important issues that have been studied extensively [3], [12], [14], [17], [21], [22], [23]. They have important implications in both synchronous optical network (SONET) and wavelength-division multiplexing (WDM) networks [16], [18], [25]. In past decade, many efficient protection schemes have been proposed, e.g. path protection and link protection schemes in [3], [21], [22], [23], p-cycle protection scheme in [8], protection cycles and self-healing rings in [6], [9], [10], [25]. In this paper, we concentrate on the preplanned recovery scheme using redundant trees. Itai and Rodeh in [11] proposed to use two trees to achieve single link or single node failure restoration for two-connected graphs. Later, Médard *et al* in [13], [14], [15] proposed a preplanned recovery scheme for two-connected graphs, which allows the use of tree routings and redundancy for recovery from failures. Further, Xue *et al* in [26], [27] and Zhang *et al* in [29] studied *Quality of Protection* (QoP) and *Quality of Service* (QoS) of recovery scheme. These schemes are applicable to IP, WDM, SONET and ATM networks to

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 179–187.
springerlink.com

provide multicast protection and restoration in case of single link or node failure [9], [16], [17], [22], [23], [24].

In [28], Zehavi and Itai proposed to use three disjoint trees to achieve any two edge failures restoration for three-edge connected graphs or two node failures restoration for three-node connected graphs, and suggested a conjecture that in general we can use k disjoint trees to achieve any $k-1$ edge failures restoration for k-edge connected graphs or $k-1$ node failures restoration for k-node connected graphs. Note that failures may occur simultaneously.

Provided that Zehavi's conjecture is correct, we propose to use the following preplanned recovery scheme to achieve a fast recovery. We can use the first tree $T^{(1)}$ in k trees as the *working* tree and the other $k-1$ trees $T^{(2)},T^{(3)},\ldots,T^{(k)}$ as *backup* trees for recovery. When one node (other than the root node) or edge fails, the root node cannot reach some nodes via $T^{(1)}$, then we use $T^{(2)}$. If the root node also can not reach some nodes via $T^{(2)}$ since other failures occur, we use $T^{(3)}$, and so on. In a word, the root node can reach each of other nodes via one of k disjoint trees.

We indeed construct preplanned recovery schemes using multiple redundant trees as above for complete graphs while we have not proved the correctness of the conjecture. In this paper, we firstly prove the complete graph with n nodes to be both $(n-1)$-edge connected and $(n-1)$-node connected, and then introduce the *Hamilton decomposition* and the *star decomposition* of complete graphs. Finally, we present two types of preplanned recovery schemes based on two types of decomposition. Both of them can guarantee a fast recovery from any simultaneous multiple node or edge failures.

The rest of this paper is organized as follows. In Sect. 2, we complete basic preliminaries. In Sect. 3, we introduce two types of decomposition of complete graphs. In Sect. 4, we present two types of preplanned recovery schemes using multiple redundant trees for complete graphs. In Sect. 5, we illustrate two schemes. In Sect. 6, we conclude this paper.

2 Preliminaries

As in [27], a computer and communication network can be represented as an undirected graph $G=(V,E)$, where V denotes the set of *vertices* and E denotes the set of *edges*. We use edge and *link* interchangeably, as well as vertex and *node*. All vertices in G are labeled by numbers $1,2,\ldots,|V|$ in sequence. Thus, $V=\{1,2,\ldots,|V|\}$. We use $[i,j]$ to denote an edge in G between a pair of nodes i and j, $[i_1,i_2,\ldots,i_k]$ to denote a *path* comprising $k-1$ edges $[i_1,i_2],\ldots,[i_{k-1},i_k]$, and $[i_1,i_2,\ldots,i_k,i_1]$ to denote a *cycle* consisting of k edges $[i_1,i_2],\ldots,[i_{k-1},i_k],[i_k,i_1]$. Each path $[i_1,i_2,\ldots,i_k]$ is associated with two *directed paths*, i.e. $(i_1,i_2,\ldots,i_k)$ from i_1 to i_k and $(i_k,i_{k-1},\ldots,i_1)$ from i_k to i_1. Two paths are called *edge-disjoint* (abbreviated to *disjoint*) if they have no edges in common, and *node-disjoint* if they have no nodes in common except two end nodes of path. We call G a *k-edge connected* (or *k-node connected*) graph if there are at least k edge-disjoint (or

node-disjoint) paths between any pair of nodes in G. We refer readers to [7] for other graph theoretic notations not defined here.

We use $K_n = (V_n, E_n)$ to denote an *undirected complete graph*, and DK_n to denote a *bi-directed complete graph* induced by K_n. (DK_n is constructed by replacing each edge $[i, j]$ in K_n with a pair of arcs (i, j) and (j, i).)

Theorem 1: K_n is both (n-1)-edge connected and (n-1)-node connected.

Proof. We mainly prove K_n to be $(n-1)$-edge connected, and similarly prove it to be $(n-1)$-node connected.

Given any two nodes i and j of K_n, $1 \le i < j \le n$, there exist at least $n-1$ different $i-j$ paths between i and j, including $[i, j]$ and $[i, z, j], z \in \{1, \ldots, n\} \setminus \{i, j\}$. It is easy to verify that any two paths have no edge in common. So there are at least $n-1$ edge-disjoint paths between i and j. Thus, K_n is at least $(n-1)$-edge connected. On the other hand, we see that there are at most $n-1$ edge-disjoint paths between i and j since the degree of each vertex in K_n is $n-1$. Thus, K_n is at most $(n-1)$-edge connected.

3 Decomposition of Complete Graphs

In this section, we introduce the Hamilton decomposition and the star decomposition of complete graphs.

3.1 Hamilton Decomposition

A cycle covering all vertices of G is called a *Hamilton cycle* of G. Given a K_n with a designated node r, we see that $[r, i_1, \ldots, i_{n-1}, r]$ is a Hamilton cycle of K_n, denoted by $H_{n,r}$. Two Hamilton cycles are called *disjoint* if they have no edges in common.

Based on some related results in [1], [2], [7] on Hamilton decomposition, we decompose K_n as follows:

(i) If n is odd, we use (1) to decompose K_n into $\frac{n-1}{2}$ disjoint Hamilton cycles, (addition in (1) is modulo $n-1$)

$$\begin{aligned} H_{n,r}^{(k)} &= [r, i_1 + k, i_2 + k, \ldots, i_{n-1} + k, r], \quad k = 0, 1, \ldots, \tfrac{n-3}{2}, \\ H_{n,r}^{(0)} &= [r, 1, 2, n-1, 3, n-2, \ldots, \tfrac{n+3}{2}, \tfrac{n+1}{2}, r], \end{aligned} \tag{1}$$

(ii) If n is even, we use (2) to decompose K_n into $\frac{n}{2}-1$ disjoint Hamilton cycles, (addition in (2) is modulo $n-1$)

$$H_{n,r}^{(k)} = [r, j_1 + k, j_2 + k, \ldots, j_{n-1} + k, r], \quad k = 0,1,\ldots,\tfrac{n}{2} - 2,$$
$$H_{n,r}^{(0)} = [r,1,2,n-1,3,n-2,\ldots,\tfrac{n}{2},\tfrac{n}{2}+1,r], \tag{2}$$

3.2 Star Decomposition

A spanning subgraph of K_n is called a *star* of K_n if it has single node with degree $n-1$ and $n-1$ nodes with degree one. The node with degree $n-1$ is called the *center* of star and each node with degree one is called a *leaf* of star. A star with center i is denoted by $S_n(i)$. Next we introduce two kinds of directed stars. When each edge of $S_n(i)$ is attached by a direction toward i, we obtain an *in-star* of DK_n with center i, denoted by $S_n^{\mathrm{I}}(i)$. When one edge $[i, j]$ of $S_n(i)$ is attached by a direction toward i and each of the other edges is attached by a direction toward its leaf, we obtain a *fork-star* of DK_n with center i and *root* j, denoted by $S_{n,j}^{\mathrm{F}}(i)$.

Two directed stars are called *disjoint* if they have no arcs in common. Given a DK_n with a designated node r, we observe that all $S_{n,r}^{\mathrm{F}}(i), i \in \{1,\ldots,n-1\}$ are disjoint. Therefore, we propose to use (3) to decompose DK_n into n disjoint directed stars including one in-star $S_n^{\mathrm{I}}(r)$ and $n-1$ fork-stars $S_{n,r}^{\mathrm{F}}(i), i \in \{1,\ldots,n-1\}$,

$$DK_n = S_n^{\mathrm{I}}(r) + \sum_{i \in \{1,2,\ldots,n-1\}} S_{n,r}^{\mathrm{F}}(i) \quad . \tag{3}$$

4 Preplanned Recovery Schemes Using Multiple Redundant Trees

Provided that Zehavi's conjecture in [28] is correct, we conclude from Theorem 1 that we can use $n-1$ disjoint trees in sequence to achieve a fast recovery from any up to $n-2$ simultaneous edge or node failures for a complete graph with n nodes. This essentially forms preplanned recovery schemes using multiple redundant trees. In this section, we construct two types of such recovery schemes, Hamilton-based recovery scheme and star-based recovery scheme. For each scheme, we design one convenient algorithm to construct it.

4.1 Hamilton-Based Recovery Scheme

Guo *et al* in [9], [10] proposed that one Hamilton cycle can recover from any single edge or node failure. For instance, a Hamilton cycle $H_{n,r} = [r, i_1, \ldots, i_{n-1}, r]$ of K_n is associated with a pair of *directed Hamilton paths*, $h_{n,r}^0 = (r, i_1, \ldots, i_{n-1})$ and $h_{n,r}^1 = (r, i_{n-1}, \ldots, i_1)$. When any single edge or node fails, r can reach each node via $h_{n,r}^0$ or $h_{n,r}^1$. Thus, we can recover from any single edge or node failure by firstly using $h_{n,r}^0$ and then $h_{n,r}^1$. This inspires us of an idea that we can recover from more

edge or node failures by using more pairs of directed Hamilton paths in sequence as long as all paths are disjoint. The scheme of realizing the idea is called the *Hamilton-based recovery scheme* (HR). Algorithm 1 can construct an HR for K_n in $O(n^2)$ time, see Theorem 2.

A directed Hamilton path $(r, i_1, \ldots, i_{n-1})$ of K_n can be stored in a $1 \times n$ row vector $\langle r, i_1, \ldots, i_{n-1} \rangle_{1\times n}$, and then $\lfloor \frac{n-1}{2} \rfloor$ pairs of directed Hamilton paths can be stored in a $2\lfloor \frac{n-1}{2} \rfloor \times n$ array in which one pair of paths are stored in adjacent two rows. As a consequence, a Hamilton-based recovery scheme for K_n can be stored as H_n in (4),

$$\mathrm{H}_n = \left(\left(h_{n,r}^{(0,0)}\right)^{\mathrm{T}}, \left(h_{n,r}^{(0,1)}\right)^{\mathrm{T}}, \ldots, \left(h_{n,r}^{(\lfloor \frac{n-1}{2} \rfloor - 1,0)}\right)^{\mathrm{T}}, \left(h_{n,r}^{(\lfloor \frac{n-1}{2} \rfloor - 1,1)}\right)^{\mathrm{T}} \right)^{\mathrm{T}}. \quad (4)$$

where $h_{n,r}^{(i,0)}$ and $h_{n,r}^{(i,1)}$ ($i = 0, 1, \ldots, \lfloor \frac{n-1}{2} \rfloor$) are the i -th pair of directed Hamilton paths of K_n.

Algorithm 1: H_n **=HRA** (n)

$\mathrm{H}_n := \varnothing_{2\lfloor \frac{n-1}{2} \rfloor \times n}$;

for $k \in \{0, 1, \ldots, \frac{n-3}{2}\}$ (if n is odd) or $k \in \{0, 1, \ldots, \frac{n}{2} - 2\}$ (if n is even) **do**

Construct a pair of directed Hamilton paths $h_{n,r}^{(k,0)}$ and $h_{n,r}^{(k,1)}$ according to (1) if n is odd or according to (2) if n is even. Store $h_{n,r}^{(k,0)}$ in the $(2k+1)$ -th row and $h_{n,r}^{(k,1)}$ in the $(2k+2)$ -th row of H_n ;

endfor

Theorem 2: Algorithm 1 can correctly construct a Hamilton-based recovery scheme H_n for K_n in $O(n^2)$ time, which can recover from any up to $n-2$ simultaneous edge or node failures if n is odd and any up to $n-3$ simultaneous edge or node failures if n is even.

Proof. Firstly, we demonstrate the correctness of algorithm 1. Since $\lfloor \frac{n-1}{2} \rfloor$ Hamilton cycles obtained by (1) or (2) are disjoint and one pair of directed Hamilton paths associated with each Hamilton cycle can survive any single edge or node failures, the union of $2\times\lfloor \frac{n-1}{2} \rfloor$ directed Hamilton paths can survive any up to $2\times\lfloor \frac{n-1}{2} \rfloor - 1$ simultaneous edge or node failures. If n is odd, then $2\times\lfloor \frac{n-1}{2} \rfloor - 1 = 2\times\frac{n-1}{2} - 1 = n-2$. Else if n is even, then $2\times\lfloor \frac{n-1}{2} \rfloor - 1 = 2\times(\frac{n}{2} - 1) - 1 = n-3$.

Next, we discuss the time complexity of algorithm 1. For each $k \in \{1, \ldots, \lfloor \frac{n-1}{2} \rfloor\}$, algorithm 1 spends $O(n)$ time to put a pair of directed Hamilton paths into H_n. Therefore, the time complexity of algorithm 1 is $O(\lfloor \frac{n-1}{2} \rfloor) \times O(n) = O(n^2)$.

4.2 Star-Based Recovery Scheme

We discover that any two fork-stars can recover from any single edge or node failure. In fact, for any fork-star $S_{n,r}^{\mathrm{F}}(i)$, it has a directed path (r,i,z) for each $z\in\{1,\ldots,n-1\}\setminus\{i\}$. Given any two fork-stars $S_{n,r}^{\mathrm{F}}(i_1)$ and $S_{n,r}^{\mathrm{F}}(i_2)$, r can reach z via (r,i_1,z) or (r,i_2,z) when any single edge or node fails. This inspires us of an idea that we can recover from more edge or node failures by using more fork-stars in sequence. The scheme of realizing this idea is called the *star-based recovery scheme* (SR). Algorithm 2 can construct an SR for K_n in $O(n^2)$ time, see Theorem 3.

A fork-star $S_{n,r}^{\mathrm{F}}(i)$ of K_n can be stored in a $1\times n$ row vector $\langle r,i,1,\ldots,i-1,i+1,n-1\rangle_{1\times n}$, and then $n-1$ fork-stars in (3) can be stored in a $(n-1)\times n$ array. As a result, a star-based recovery scheme for K_n can be stored as Ψ_n in (5),

$$\Psi_n=\left(\left(S_{n,r}^{\mathrm{F}}(1)\right)^{\mathrm{T}},\left(S_{n,r}^{\mathrm{F}}(2)\right)^{\mathrm{T}},\ldots,\left(S_{n,r}^{\mathrm{F}}(n-1)\right)^{\mathrm{T}}\right)^{\mathrm{T}}. \tag{5}$$

Algorithm 2: Ψ_n **= SRA** (n)

$\Psi_n:=\varnothing_{(n-1)\times n}$;
for $k\in\{1,2,\ldots,n-1\}$ **do**
Store $S_{n,r}^{\mathrm{F}}(k)=(r,k,1,\ldots,k-1,k+1,\ldots,n-1)$ in the k-th row of Ψ_n;
Endfor

Theorem 4: Algorithm 2 can correctly construct a star-based recovery scheme Ψ_n for K_n in $O(n^2)$ time, which can recover from any up to $n-2$ simultaneous edge or node failures.

Proof. Firstly, we demonstrate the correctness of algorithm 2. For each node $z\in\{1,\ldots,n-1\}$, each fork-star $S_{n,r}^{\mathrm{F}}(i)$ where $i\in\{1,\ldots,n-1\}\setminus\{z\}$ has a directed path (r,i,z) and $S_{n,r}^{\mathrm{F}}(z)$ has a directed path (r,z). The union of $n-1$ fork-stars can survive any up to $n-2$ simultaneous edge or node failures.

Next, we discuss the time complexity of algorithm 2. For each $k\in\{1,\ldots,n-1\}$, algorithm 2 spends $O(n)$ time to put a star-based recovery scheme into Ψ_n. Therefore, the time complexity of algorithm 2 is $O(n-1)\times O(n)=O(n^2)$. □

Both algorithm 1 and algorithm 2 output a fixed array in a quadratic time as long as the input is a fixed number. Based on above discussions, we draw a conclusion that, in terms of capability of recovery, Hamilton-based recovery scheme is as good as star-based recovery scheme when the number of nodes is odd and a little worse when even.

5 Illustrative Examples

In this section, we illustrate Hamilton-based recovery scheme and star-based recovery scheme with two sample complete graphs where the number of nodes takes on 4 and 5.

For a complete graph with 4 nodes, we use algorithm 1 together with (2) and (4) to construct an HR comprising two directed Hamilton paths, and use algorithm 2 together with (5) to construct an SR comprising three fork-stars, see Fig. 1. Suppose that edge $[1,2]$ fails. In HR, we firstly use $(r,1,2,3)$, then $(r,3,2,1)$. Root node r can reach 1 but can not reach 2 and 3 via $(r,1,2,3)$, then reach 3 and 2 via $(r,3,2,1)$. In SR, we firstly use $S^{\mathrm{F}}_{4,r}(1)$, then $S^{\mathrm{F}}_{4,r}(2)$, finally $S^{\mathrm{F}}_{4,r}(3)$. Root node r can reach 1 and 3 but can not reach 2 via $S^{\mathrm{F}}_{4,r}(1)$, then reach 2 via $S^{\mathrm{F}}_{4,r}(2)$. Suppose that node 2 fails. In HR, root node r can reach 1 but can not reach 3 via $(r,1,2,3)$, and then reach 3 via $(r,3,2,1)$. In SR, root node r can reach 1 and 3 via $S^{\mathrm{F}}_{4,r}(1)$. Furthermore, we observe that HR can not recover from another failure on edge $[2,3]$ since r still can not reach 3 and 2 via $(r,3,2,1)$ while SR also can achieve a recovery since r can reach 2 via $S^{\mathrm{F}}_{4,r}(2)$.

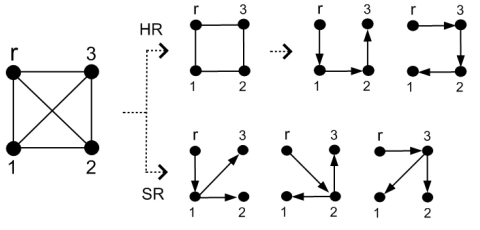


Fig. 1. HR and SR on a complete graph with 4 nodes

In a similar way, we analyze and compare HR and SR on a complete graph with 5 nodes, see Fig. 2.

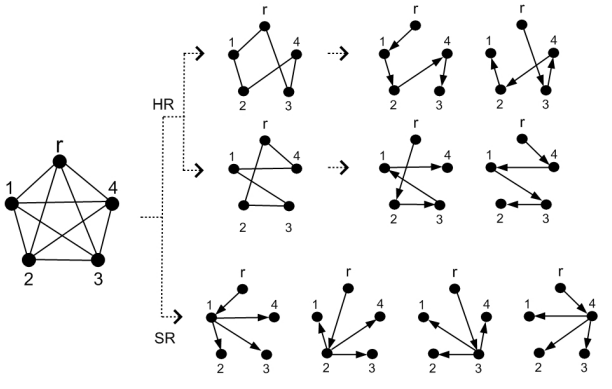


Fig. 2. HR and SR on a complete graph with 5 nodes

6 Conclusions

In this paper, we have proposed two types of preplanned recovery schemes using multiple redundant trees for complete graphs. Both of them can achieve a fast recovery from any multiple simultaneous edge or node failures. As in [26], [27], [29], it is also of interests to study their QoP and QoS issues. Moreover, how to construct a recovery scheme to guarantee a fast recovery for general k-connected graphs is an important research direction.

References

1. Alspach, B., Bermond, J.C., Sotteau: Decompositions into cycles I: Hamilton decompositions, cycles and rays, pp. 9–18. Kluwer Academic Press (1990)
2. Alspach, B., Gavlas, H.: Cycle decompositioins of Kn and Kn-I. J. Combin. Theory, Ser. B 81, 77–99 (2001)
3. Anand, V., Qiao, C.: Dynamic establishment of protection paths in WDM networks, part I. In: IEEE ICCCN 2000, pp. 198–204 (2000)
4. Choi, H., Subramaniam, S., Choi, H.-A.: On double-link failure recovery in WDM optical networks. In: IEEE Infocom 2002, pp. 808–816 (2002)
5. Choi, H., Subramaniam, S., Choi, H.-A.: Loopback recovery from double-link failures in optical mesh networks. IEEE/ACM Transactions on Networking 12, 1119–1130 (2004)
6. Ellinas, G., Hailemariam, A.G., Stern, T.E.: Protection cycles in mesh WDM networks. IEEE Journal on Selected Areas in Communications 18, 1924–1937 (2000)
7. Gross, J.L., Yellen, J.: Handbook of Graph theory. CRC Press (2004)
8. Grover, W.D., Stamatelakis, D.: Cycle-oriented distributed preconfigur-ation: ring-like speed with mesh-like capacity for self-planning network restoration. In: IEEE ICC 1988, pp. 537–543 (1998)
9. Guo, L., Wang, X.W., Du, J., Wu, T.F.: A new heuristic routing algorithm with Hamiltonian Cycle Protection in survivable networks. Computer Communications 31, 1672–1678 (2008)
10. Guo, L., Wang, X.W., Wei, X.T., Yang, T., Hou, W.G., Wu, T.F.: A New Link-Based Hamiltonian Cycle Protection in Survivable WDM Optical Networks. In: AICT 2008, pp. 227–231 (2008)
11. Itai, A., Rodeh, M.: The multi-tree approach to reliablity in distributed networks. Information and Computation 79, 43–59 (1988)
12. Jukan, A., Monitzer, A., de Marchis, G., Sabella, R. (eds.): Restoration methods for multi-service optical networks. Optical Networks: Design and Modelling, pp. 3–12. Kluwer Academic Publishers (1999)
13. Médard, M., Barry, R.A., Finn, S.G., He, W., Lumetta, S.S.: Generalized loop-back recovery in optical mesh networks. IEEE/ACM Trans. Net. 10, 153–164 (2002)
14. Médard, M., Finn, S.G., Barry, R.A.: A novel approach to automatic protection switching using trees. In: IEEE ICC 1997, pp. 272–276 (1997)
15. Médard, M., Finn, S.G., Barry, R.A., Gallager, R.G.: Redundant trees for preplanned recovery in arbitrary vertex-redundant or edge-redundant graphs. IEEE/ACM Trans. Net. 7, 641–652 (1999)
16. Mohan, G., Ram Murthy, C.S., Somani, A.K.: Efficient algorithms for routing dependable connections in WDM optical networks. IEEE/ACM Trans. Net. 9, 553–566 (2001)

17. Mukherjee: WDM optical networks: progress and challenges. IEEE Journal on Selected Areas in Communications 18, 1810–1824 (2000)
18. Mukherjee: Optical Communication Networks. McGraw Hill (1997)
19. Ouveysi, I., Wirth, A.: On design of a survivable network architecture for dynamic routing: optimal solution strategy and an efficient heuristic. European Journal of Operational Reshearch 117, 30–44 (1999)
20. Ouveysi, I., Wirth, A.: Wirth, Minimal complexity heuristics for robust network architecture for dynamic routing. Journal of the Operational Research Society 50, 262–267 (1999)
21. Qiao, D.X.: Distributed partial information management (DPIM) schemes for survivable networks, part I. In: IEEE Infocom 2002, pp. 302–311 (2002)
22. Ramamurthy, S., Mukherjee, B.: Survivable WDM mesh networks, part I-protection. In: IEEE Infocom 1999, pp. 744–751 (1999)
23. Ramamurthy, S., Mukherjee, B.: Survivable WDM mesh networks, part II-restoration. In: IEEE ICC 1999, pp. 2023–2030 (1999)
24. Sahasrabuddhe, L., Ramamurthy, S., Mukherjee, B.: Fault management in IP-over-WDM networks: WDM protection versus IP restoration. IEEE Journal on Selected Areas in Communications 20, 21–33 (2002)
25. Wu, T.H., Way, W.I.: A novel passive protected SONET bidirectional self-healing ring architecture. IEEE Journal of Lightwave Technology 10, 1314–1322 (1992)
26. Xue, G., Chen, L., Thulasiraman, K.: QoS issues in redundant trees for protection in vertex-redundant or edge-redundant graphs. In: IEEE ICC 2002, pp. 2766–2770 (2002)
27. Xue, G., Chen, L., Thulasiraman, K.: Quality of service and quality protection issues in preplanned recovery schemes using redundant trees. IEEE Journal on Selected Areas in Communications; Optical Communications and Networking series 21, 1332–1345 (2003)
28. Zehavi, A.I.: Three tree-paths. J. Graph Theory 13, 175–188 (1989)
29. Zhang, W.Y., Xue, G., Tang, J., Thulasiraman, K.: Faster algorithms for constructing recovery trees enhancing QoP and QoS. IEEE/ACM Trans. Net. 16, 642–655 (2008)

Three-Dimensional Numerical Simulation of Duct Conveying Using Smoothed Particles Hydrodynamics

Peigang Jiao

Department of Construction Machinery, Shandong Jiaotong University
Jinan, Shandong Province, 250023, China
jiaopeigang@163.com

Abstract. An efficient smoothed particle hydrodynamics (SPH) method for duct conveying is proposed. We have implemented a 3D SPH code for duct conveying, which extends industrial application field of SPH method. The simulation results of 3D SPH accord with the experimental results. Numerical simulation results show that major physics of the duct conveying can be well captured in the simulation. The SPH method proposed can be easily extended to deal with many other problems of duct flow if simulation control parameters such as input condition, boundary condition and corresponding equation of state etc. are defined properly.

Keywords: smoothed particle hydrodynamics (SPH), numerical simulation, duct conveying, meshfree method.

1 Introduction

Duct conveying is commonly seen in many material handling systems such as mining, pharmacy and chemical industry. During the duct conveying process, the movement of material in the duct is very intricate, which is affected by many factors. Theoretical and experimental investigations of duct conveying have been carried out in the past few decades because of its importance [1]. However, theoretical solutions to the duct conveying are only limited to some simple cases; experimental studies need to resort to expensive testing facilities. Recently more and more analyses of duct conveying are based on numerical simulations with the development of the computer hardware and computational techniques. However, numerical simulations of duct conveying problem are generally very difficult for traditional grid-based numerical methods due to some features such as large deformations, moving material interfaces, deformable boundaries and so on.

Recent developments in so-called meshless methods provide alternates for traditional numerical methods in modeling the duct conveying. Among the meshless methods, smoothed particle hydrodynamics (SPH) methodology is unique in the computational fluid dynamics [2]. The nice combinations of meshless, Lagrangian and particle properties of SPH make it fairly attractive in simulating the duct conveying and related physics [3].

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 189–197.
springerlink.com

This paper presents our implementation of the SPH method as well as its application to simulate three-dimensional duct conveying problems. In Section II, the basic SPH methodology, governing equation and some important numerical aspects are described. The numerical example of duct conveying is presented in Section 3. Major physics of the duct conveying can be captured in the simulations.

2 SPH Methodology and Numerical Aspects

The combination of meshless and Lagrangian nature inherent in the SPH methodology makes it very attractive in treating large deformations and moving boundaries in the extremely transient duct conveying process.

A. SPH Formulation

In the SPH method, for a function *f*, the approximation of its function value at a certain location can be expressed as summation interpolations over the neighbor particles using a smoothing kernel function W with the smoothing length *h* [4]:

$$< f(x_i) >= \sum_{j=1}^{N} \frac{m_j}{\rho_j} f(x_j) W_{ij} \tag{1}$$

where W_{ij} is the smoothing function of particle *i* evaluated at particle *j*, and is closely related to the smoothing length *h* (see Figure 1).

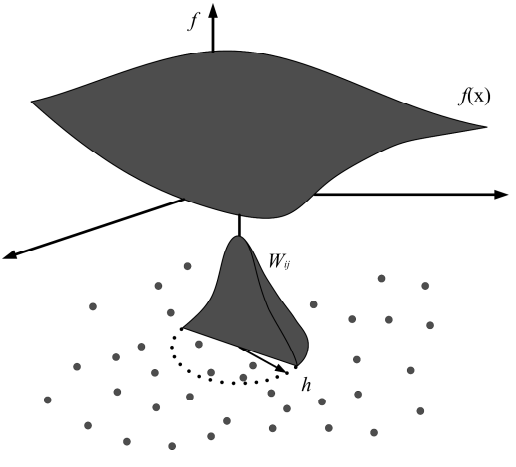


Fig. 1. Reproducing kernel approximation of a function

The Euler equation can be used to model the duct conveying process with suitable equation of state,

$$\begin{cases} \dfrac{D\rho}{Dt} = -\rho \nabla \cdot v \\ \dfrac{Dv}{Dt} = -\dfrac{1}{\rho}\nabla p \\ \dfrac{Du}{Dt} = -\dfrac{p}{\rho}\nabla \cdot v \\ p = p(\rho, u) \end{cases} \tag{2}$$

where v—— velocity vector
u—— internal energy
ρ—— density
p—— pressure
t —— time instant

The first three equations in Eq.(2) state the conservation of mass, momentum, and energy, while the fourth one is the equation of state.

Based on the concept of kernel and particle approximation in SPH, discretized SPH equations of motion are used to model duct conveying problem.

$$\begin{cases} \dfrac{D\rho_i}{Dt} = \sum_{j=1}^{N} m_j v_{ij}^{\beta} \cdot \dfrac{\partial w_{ij}}{\partial x_i^{\beta}} \\ \dfrac{Dv_i^{\alpha}}{Dt} = -\sum_{j=1}^{N} m_j \left(\dfrac{p_i}{\rho_i^2} + \dfrac{p_j}{\rho_j^2}\right) \dfrac{\partial w_{ij}}{\partial x_i^{\alpha}} \\ \dfrac{De_i}{Dt} = \dfrac{1}{2}\sum_{j=1}^{N} m_j \left(\dfrac{p_i}{\rho_i^2} + \dfrac{p_j}{\rho_j^2}\right) v_{ij}^{\beta} \dfrac{\partial w_{ij}}{\partial x_i^{\beta}} \\ \dfrac{Dx_i}{Dt} = v_i \end{cases} \tag{3}$$

Where i —— Reference particle
j —— Neighboring particle of i

B. Equation of State

In the standard SPH method for solving compressible flows, the particle motion is driven by the pressure gradient, while the particle pressure is calculated by the local particle density and internal energy through the equation of state. Every theoretically incompressible fluid becomes actually compressible by using artificial compressibility. Therefore, it is feasible to use a quasi-incompressible equation of state to model the incompressible flow.

In the implementation, we employ the following state equation which has been applied to model free surface flows by Monaghan [5].

$$p(\rho) = B\left[\left(\frac{\rho}{\rho_0}\right)^{\gamma} - 1\right] \tag{4}$$

Where γ is constant, different material has different value of γ. A zero subscript denotes reference quantities. B is a problem dependent parameter, which sets a limit for the maximum change of the density. Generally, B is taken as the initial pressure. It can be seen from Eq. (4) that a small oscillation in density may result in a large variation in pressure.

C. Boundary Treatment

In the application of SPH, boundary treatment is both its advantage and its weakness at present.

In our work, virtual particles are used to treat the solid duct boundary condition. There are two types of virtual particles as illustrated in Figure 2. The virtual particles of the first type (type 1) are located right on the solid boundary and are similar to what Monaghan used. The virtual particles of the second type (type 2) fill in the boundary region and are similar to what Libersky and Petscheck used. The second types of virtual particles are constructed in the way below, for a certain real particle *i* inside the boundary, a virtual particle is placed symmetrically on the outside of the boundary. These virtual particles possess the same pressure and density as the corresponding real particles but opposite velocity. The second types of virtual particles are usually used in the instance characterized by continuous change of boundary condition. The first types of virtual particles are used to simulate the wall of duct in this paper.

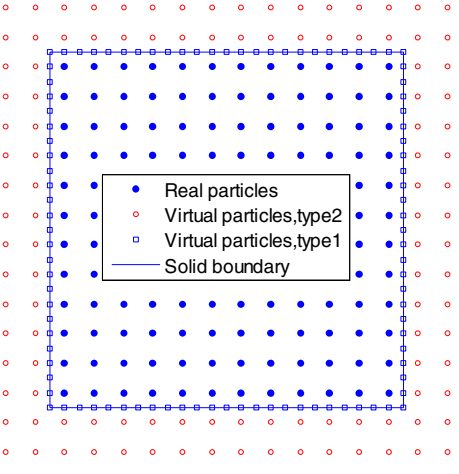


Fig. 2. Illustration of real particles and the two types of virtual particles as well as the solid boundary

D. Structure of SPH Code

The basic SPH methodology and the accompanied algorithms for various numerical aspects of SPH result in some special features in the SPH coding. These special features are generally involved under the main loop of time integration process, including particle interaction calculation, boundary treatment, etc.

The structure of a typical SPH code is schematically shown in Figure 3.

E. Time Integration

Some standard methods such as Leap-Frog (LF), predictor-corrector and Runge-Kutta (RK) schemes can be used to perform the numerical integration of ordinary differential equation for physical variables at every particle.

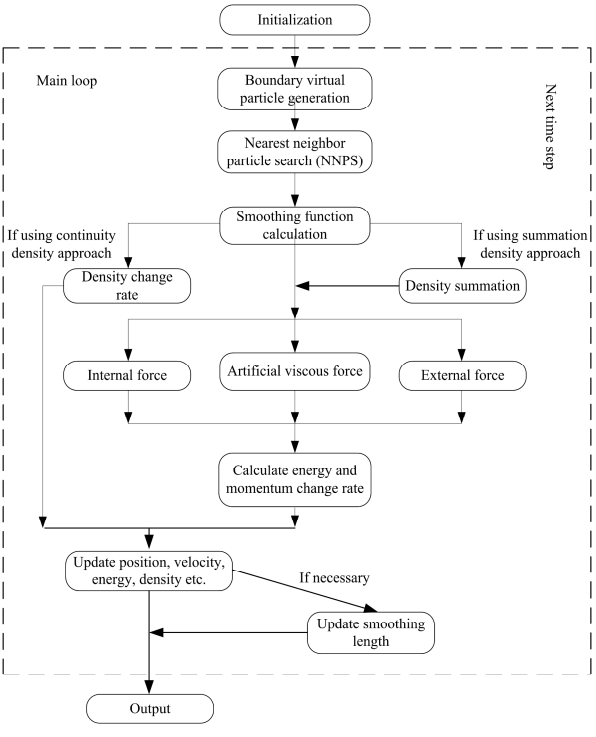


Fig. 3. Structure of the SPH code

In our work, the Leapfrog method is used in the time integration for its low memory storage in the computation and efficiency for one force evaluation per step. The particle density, velocity, internal energy and position can be upgraded in the following formulations:

$$
\begin{aligned}
t &= t + \Delta t \\
\rho_i(t + \Delta t/2) &= \rho_i(t - \Delta t/2) + \Delta t \cdot D\rho_i(t) \\
v_i(t + \Delta t/2) &= v_i(t - \Delta t/2) + \Delta t \cdot Dv_i(t)
\end{aligned} \tag{5}
$$

$$u_i(t+\Delta t/2) = u_i(t-\Delta t/2) + \Delta t \cdot Du_i(t)$$
$$x_i(t+\Delta t) = x_i(t) + \Delta t \cdot v_i(t+\Delta t/2)$$

Note that the LF scheme is conditionally stable. The stability condition is the so-called CFL (Courant-Friedrichs-Levy) condition, which typically results in a time step proportional to the smoothing lengths. In our work, the time step is taken as,

$$\Delta t = \min(\xi h_i / [h_i \nabla \cdot v_i + c_i + 1.2(\alpha_\Pi c_i + \beta_\Pi |\nabla \cdot v_i|)]) \quad (6)$$

where ξ is the Courant number, taken around 0.3.

3 Numerical Example

The following three-dimensional numerical example is given to validate the proposed duct conveying SPH method, nondimensional parameters are employed in the simulation in order to research real effect of every parameters.

The model consists of a bottomed conical silo connected with a horizontal-vertical-horizontal duct through a concentric outlet. In the operation, the material for conveying is initially quiescent as shown in figure 4. Then the pellets discharge downwards by their own gravity forces. After entering the horizontal duct, they are driven to the right along the duct by blower in the left.

A total of 1134 real particles are used to discretize the pellets domain and an additional 10821 virtual particles to represent the solid boundaries of the silo and the duct. These virtual particles have the same properties as the real particles except that they are permanently stationary. A time step of 0.5ms is used and the total 20000 time steps are employed in the simulation.

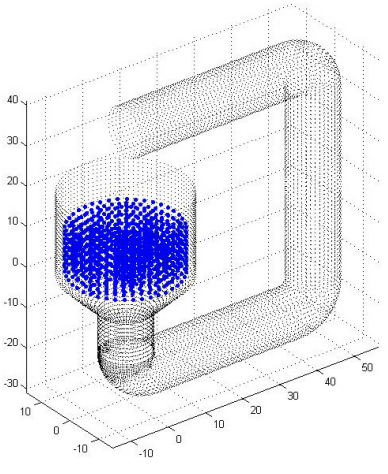


Fig. 4. Initial geometry and particle distribution

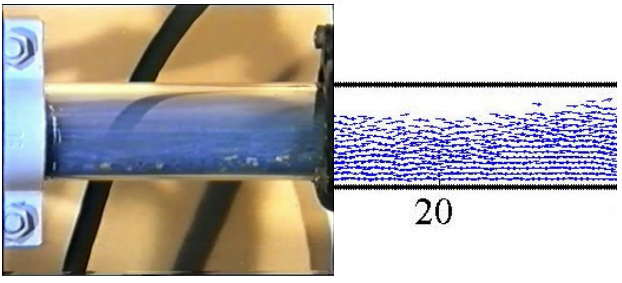


Fig. 5. Particles conveying results of horizontal duct Left: experimental; Right: SPH

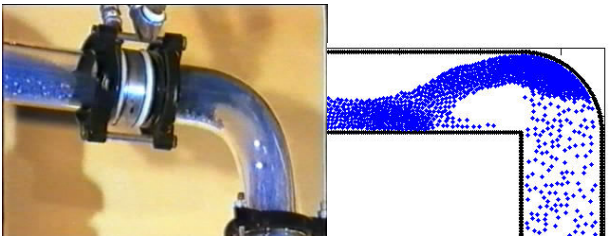


Fig. 6. Particles conveying results of Vertical duct Left: experimental; Right: SPH

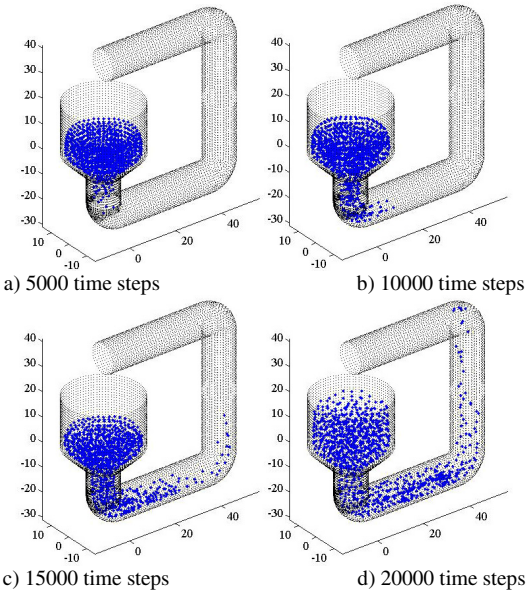
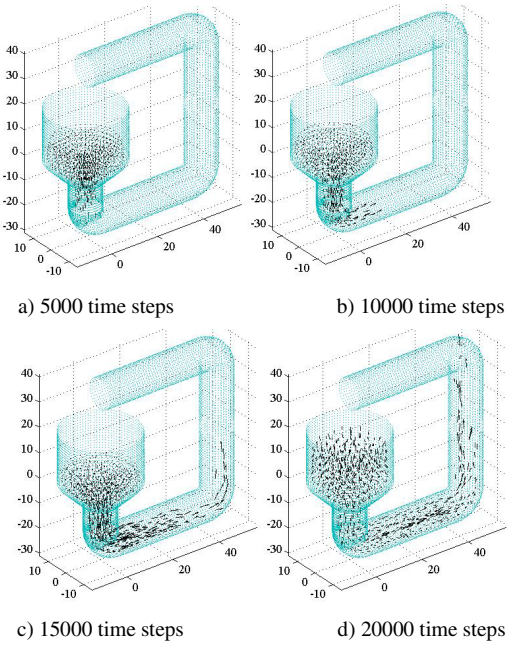


Fig. 7. Particle distribution in the duct conveying process at different time steps

Horizontal duct: Because gravity is perpendicular to drag force, Particles of horizontal duct have a trend towards the bottom, as shown in fig. 5, the left is experimental result, the right is SPH result, and we can see that two results obtained by using different methods are the same.

Vertical duct: Particle in the vertical duct suffers not only thrust force but also its own gravity, in the case thrust force is greater than gravity, particles still move upwards, and these particles will move towards the left along the duct when they reach the corner as illustrated in Figure 6.

Figure 7 shows particle distribution during the duct conveying process at four different instants. It can be seen that the discharged pellets were prevented from penetrating through the boundary of silo or duct under the control of virtual particles of the first type located right on the solid boundary. At the beginning the discharged pellets move downwards by their own gravities, after they met the bend, these pellets are accelerated to the right gradually along the horizontal duct under the repulsive force of virtual particles and blower pressure. An area of no pellets in the upper left of horizontal duct is produced. When the particles reach the right bend, they move upwards along the vertical duct.



a) 5000 time steps b) 10000 time steps

c) 15000 time steps d) 20000 time steps

Fig. 8. Velocity distribution in the duct conveying process at different times

Figure 8 shows particle velocity distribution during the duct conveying process at four different instants. It can be seen that the speed of pellets increase gradually driven by the left blower in the horizontal duct till pellets reach a certain steady speed.

Due to conical bottom of the silo, when pellets move downward, they will gather towards central axis in the conveying process. This aggregation process increases the

pressure around central axis. As a result, some pellets on the center of the top will produce a transient upward motion as shown in Figure 8d.

4 Conclusions

Three-dimensional duct conveying process is simulated using the pure Lagrangian meshless SPH methodology. The distributions and velocities of particles in the conveying process are analyzed. Experimental results accord with simulation results, which demonstrate the ability of SPH to produce good quality and reasonably accurate solutions.

This work is mainly focused on the modeling of duct conveying with SPH method. SPH method is firstly employed to simulate 3D duct conveying process, which extends industrial application field of SPH method.

The SPH method presented in this paper is of generality and simplicity, which can be easily extended to deal with many other problems of duct conveying in engineering system if simulation control parameters such as input condition, boundary condition and corresponding equation of state are modified properly.

Acknowledgment. This work is supported by research fund of Shandong Jiaotong University (Z201027) and sponsored by scientific research foundation for Ph.D of Shandong Jiaotong University.

References

1. Jiao, P., Zhou, Y., Fang, J., Chen, L., Cui, X.: Smoothed particle hydrodynamics for numerical simulation of duct conveying. In: 2008 Pacific-Asia Workshop on Computational Intelligence and Industrial Application, Wuhan, China, vol. 2, pp. 634–639 (2008)
2. Monaghan, J.J., Gingold, R.A.: Shock simulation by the particle method SPH. J. Comput. Phys. 52, 374–389 (1983)
3. Liu, M.B., Liu, G.R., Lam, K.Y., Zong, Z.: Smoothed particle hydrodynamics for numerical simulation of underwater explosion. Comput. Mech. 30(2), 106–118 (2003)
4. Monaghan, J.J.: Simulating free surface flow with SPH. Journal of Computational Physics 110, 399–406 (1994)
5. Monaghan, J.J.: An introduction to SPH. Comput. Phys. Commun. 48, 89–96 (1988)

A New Data Mining Approach Combing with Extension Transformation of Extenics

Honghai Guan

Mudanjiang Normal University

Abstract. Extenics is a new discipline is to solve the problem, research data mining is mining useful knowledge from data sets the rapid development of the information society, information, knowledge to overload age, how to excavate the useful and change of knowledge, puts forward a new task of data mining, establish the extenics provides new ideas and methods for data mining and opens up new research direction for data mining, using the extenics method of combining the data mining technology not only can access to static knowledge, but can dig to realization of knowledge, and illustrate the effectiveness of the proposed method.

Keywords: data mining, knowledge, extenics.

1 Introduction

The history of mankind is a history of resolving contradictions. Whether in the past, present, and future, people are the time to deal with problem.

The more accurate say are producing extenics to resolve conflicts. Extenics research formal model of the possibility and thinking expansion open the rules and methods to solve the problem and the innovation, the absurd science. Extenics research ridiculous objective world problems, at present, gained advertisement formal description model of advantage, the worst extension and extension transformation of thing, it builds extension methods and extension engineering methods of contradiction to solve problems. On the other hand, extenics provides new methods and tools for data mining is studied. Data mining application extenics are known extension data mining method. The extension data mining method is based on the extension theory and extension method, our objective is to mining "cannot change yes" rule. And the differences of traditional a-one data mining is data mining traditional knowledge discovery, however extension data mining are not only mining knowledge and skills, but also learn to relevant rules and expand trans -, in order to form the knowledge mining.

The extension data mining advertising extensions of combining the data mining, it studies the theory and method in extenics mining knowledge related transformation solve this contradiction database, including promotion class of problems knowledge, transfer knowledge and other related transformation knowledge, they say extension knowledge. At present, we study the basic method and computer to realize the extension data mining problems.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 199–205.
springerlink.com

2 The Basic Theory Knowledge

2.1 Theory Overview of Extenics

Extenics is a new discipline, Chinese scholars independently created, it mainly studies "concept, the human brain's way of thinking" managing incompatible problems and computer simulation optimization problem [1]. Extenics study the possibility of formal model rules of ideas and expand the opening of the innovation advertising method of absurdity of advertising to solve the problem of science. Research object is extenics, including the contradiction between not real world problems and opposition problem, its purpose is to make the disharmonious problems of compatibility becomes an issue, against problem into the coexistence of the problem. Research direction is to explore the legal advertising method problem solving contradictions [2].

2.2 The Concept of Data Mining

From a lot of data mining, incomplete, noisy, fuzzy and random data extraction implied useful information and knowledge of people don't know in advance but potential [3].

At present, the task of data mining are as follows: correlation analysis, timing model, clustering, classification, correcting error detection and prediction.

3 Concept and Theory of Extension Data Minig

3.1 Concept of Extension Data Mining

Extenics is to use the extension transformation, that is to say, from changing ideas that fake proposition come true ideas and unknown problems existed problems, to be done by the problem of conversion of the actual problem. The extension data mining, and on the basis of the data mining in get static knowledge, according to change the extension transformation obtain knowledge, that is, containing the extension transformation rule knowledge transformation between the extension or with the corresponding transfer rules knowledge conversion process [4].

3.2 Theory of Data Mining Extension

Extension data mining theory is based on the following two theorems [5]:

Theorem l. For the two types of rules:

$$\wedge a_i \rightarrow P \tag{1}$$

$$\wedge b_i \rightarrow N \tag{2}$$

If existence the conditions of the extension transformation T_{if} :

$$T_{if}(b_j) = a \tag{3}$$

Description: For $j \geq i$ there is $T(b_k) = a_k$, k = 1, 2 ... i.

And existence of the conclusions of the extension transformation $T_{conclude}$:

$$T_{conclude}(N) = P \tag{4}$$

The rules knowledge of Extension transformation is established (change knowledge):

$$[(b_j) = a_j \rightarrow T(N) = P] \tag{5}$$

Namely

$$\text{If} \quad (b_j) = a_j \quad \text{Then} \quad T(N) = P \tag{6}$$

Theorem 2: For the same two rules:

$$A \rightarrow P \tag{7}$$

$$C \wedge B \rightarrow P \tag{8}$$

If there is extension transformation

$$T(B) = A \tag{9}$$

So set up: the rules knowledge transformation of extension

$$T(B) = A \rightarrow P \tag{10}$$

That is, if T (B) = A then P.

3.3 The Process of Extension Data Mining

From the extension data mining theorem, extension data mining process can be sued the following [6]:

The first pace: for classification problem using data mining methods to obtain classification rule that obtain knowledge (1) and (2).

The second step: make sure to existing rules extension transformation in the premise and the presence of extension transformation in short, that is, to meet the FND extended transformation (3) advertising (4).

Step 3: use theorem 1 and theorem 2 widely knowledge (5) or (10).

4 Mining Based on Extension to Learn Extension Transformation of Knowledge Instance

Knowledge of primitive extenics, the relationship between the basic elements of extended from analysis, can obtain the expanded knowledge; extension transformation can be expanded knowledge. So, how to find useful new knowledge outside, still can from existing knowledge? And how to discover the method based on existing knowledge? I am through the relationship between tourism and climate that use this machine extension method from unearthed knowledge base in the extension transformation of knowledge.

4.1 Acquire Rule Knowledge through Data Mining

There are four conditions database: characteristics of weather make people feel equivocal, temperature C2, humidity, C3, wind C4. The weather has 3 value of sunny, cloudy, rain, Temperature hot, medium and cold three values; Humidity have high, normal value; two - The wind has "wind", "no wind" two cases. Conclusion of phosphorus and nitrogen classes, P says, they can go to travel, N said not to travel. Data in table 1 and table shows the extension unearthed from a table and hope that can be useful knowledge.

Table 1. CUM TE YUQIG SET

No	Features				categories d
	weather C1	temperature C2	humidity C3	wind C4	
1	cloudy	hot	high	windy	N
2	cloudy	hot	high	windy	N
3	rain	hot	high	windless	P
4	rain	moderate	high	windless	P
5	rain	cold	normal	windless	P
6	cloudy	cold	normal	windy	N
7	sunny	cold	normal	windy	P
8	sunny	moderate	high	windless	N
9	rain	cold	normal	windless	P
10	sunny	moderate	normal	windless	P
11	cloudy	moderate	normal	windy	P
12	cloudy	moderate	high	windy	P
13	rain	hot	normal	windless	P
14	sunny	moderate	high	windy	N

According to the data sheet can be obtained the tourism of O's 14 four-dimensional state information element.

$$I_i = \begin{bmatrix} Travel\,\mathrm{O}, & C1, & \mathrm{v}_{1\mathrm{i}} \\ & C2, & \mathrm{v}_{2\mathrm{i}} \\ & C3, & \mathrm{v}_{3\mathrm{i}} \\ & C4, & v_{4i} \end{bmatrix}, \quad \mathrm{i}=1,2,\ldots,14$$

Conclusion information element has two pairs of ay one class state information element

$$D_1 = (O, d, P) \qquad D_2 = (O, d, N)$$

For ay one class state information elements:

$$I = \begin{bmatrix} O, & C1, & V_1 \\ & C2, & V_2 \\ & C3, & V_3 \\ & C4, & V_4 \end{bmatrix}$$

Which

V_I = {sunny, cloudy, rain} V2 = {hot, moderate, cold}

V_3 = {high, normal}, V4 = {the wind, no wind}

Listed in the preceding paragraph class data, using the decision tee data mining method, using the ID3 prior decision tree decision tree data mining method, in order to obtain the following decision tree knowledge, figure 1 shows. By the following sub - rule knowledge.

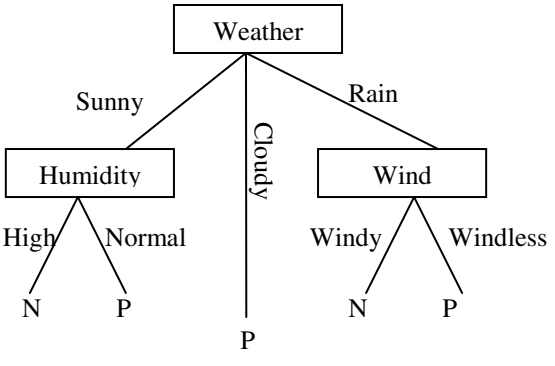


Fig. 1. Decision tree

$$[(O, C1, sunny) \wedge (O, C3, normal)] \Rightarrow [O, C1, normal) \wedge (O, d, P) = DJ]$$

$$[(O, C1, cloudy)] \Rightarrow [(O, C1, normal) \wedge (O, d, P) = D_1]$$

$$[(O, C1, rain) \wedge (O, C4, windless)] \Rightarrow [(O, d, N) = D_2]$$

$$[(\quad normal) \wedge (O, C3, high)] \Rightarrow [(O, d, N) = D_2]$$

$$[(O, C1, rain) \wedge (O, C4, windless)] \Rightarrow [(O, d, N) = D2]$$

4.2 The Conditions Allow the Extension Transformation

1) Condition transformation

$$T_1(O, C1, sunny) = (O, C1, cloudy)$$

$$T_2(O, C1, sunny) = (O, C1, rain)$$

$$T_3(O, C1, rain) = (O, C1, cloudy)$$

$$T_4(O, C1, cloudy) = (O, C1, sunny)$$

$$T_5(O, C1, rain) = (O, C1, sunny)$$

$$T_6(O, C1, cloudy) = (O, C1, rain)$$

$$T_7(O, C3, high) = (O, C3, normal)$$

$$T_8(O, C3, normal) = (O, C3, high)$$

$$T_9(O, C4, windless) = (O, C4, windy)$$

$$T_{10}(O, C4, windy) = (O, C4, windless)$$

2) Allows the conclusions of transformation

$$TD_2 = D_1, \quad TD_1 = D_2$$

4.3 The Use of Data Mining Extension Theorem 1 and Theorem 2 Can Be Obtained on the Transformation of Knowledge

3) The categories of change in knowledge:

$$\{[(O, C1, sunny) \wedge [T_7(O, C3, high) = (O, C3, normal)]\}$$
$$\Rightarrow [TD_2 = D_1]$$
$$\{[(O, C3, high) \wedge [T_1(O, C1, sunny) = (O, C1, cloudy)]\}$$
$$\Rightarrow [TD_2 = D_1]$$
$$\{[(O, C1, rain) \wedge [T_{10}(O, C4, windy) = T_9(O, C4, windless)]\}$$
$$\Rightarrow [TD_2 = D_1]$$
$$\{[(O, C4, windy) \wedge [T_3(O, C1, rain) = (O, C1, cloudy)]\}$$
$$\Rightarrow [TD_2 = D_1]$$
$$\{[(O, C1, sunny) \wedge [T_8(O, C3, normal) = (O, C3, high)]\}$$
$$\Rightarrow [TD_1 = D_2]$$
$$\{[(O, C1, rain) \wedge [T_9(O, C4, windless) = T_9(O, C4, windy)]\}$$
$$\Rightarrow [TD_1 = D_2]$$

4) The categories of knowledge does not change

$$\{[(O, C3, normal) \wedge [T_1(O, C1, sunny) = (O, C1, cloudy)]\} \Rightarrow D_1$$
$$\{[(O, C4, windless) \wedge [T_3(O, C1, rain) = (O, C1, cloudy)]\} \Rightarrow D_1$$

$$\{[(O, C4, windless) \wedge [T_8(O, C1, cloudy) = (O, C1, rain)]\} \Rightarrow D_1$$
$$\{[(O, C3, normal) \wedge [T_4(O, C1, cloudy) = (O, C1, sunny)]\} \Rightarrow D_1$$

Tourism and climate relationship in the extension knowledge, there is an extension of knowledge for

$$\{[(O, C1, rain) \wedge [T_3(O, C1, rain) = (O, C1, cloudy)]\}$$
$$\Rightarrow [TD_2 = D_1]$$

Now "weather" appeared in "rain" into "Yin" is still "wind" as extension knowledge reasoning can draw the conclusion.

Although climate is still "windy," when the weather is "rain" into "dark" trip, so you cannot become travel.

5 Conclusions

These paper preliminary studies of extenics extension transformation use and data mining technology, the combination of both methods in mining changing knowledge from knowledge base, this combination is a new method, the development, the traditional data mining a series of economic and social fields has important practical value. It not only make the enterprise in product development early detection of new products, advertising can help doctors found symptoms fundamental change, find out the best option to improve effectiveness of treatment effect, in the computer industry, as early as possible to develop new software products to solve the network problems. The present study also is an abecedarian only, worthy of further discussion and study.

References

1. Cai, W., Yang, C.: Extenics engineering approach. Science Press, Beijing (2001)
2. Cai, W., Yang, C., He, B.: Extenics logic initial. Science Press, Beijing (2003)
3. Han, J., Kambr, M.: The concept and technology of data mining and. Xiao-f eng Meng translation. Mechaical Industry Press, Beijing (2001)
4. Cai, W., Yang, C., Chen, W., Li, X.: Extension Set and Extension Data Mining. Science Press, Beijing (2008)
5. Huang, J., Chen, W.: the conception and theory of extension data mining. Computer Engineering and Applications I4, 7–9 (2006)
6. Chen, W.: Research of mining the mutative knowledge with extension data mining. Engineering Sciences 18(11), 70–73 (2006)

Iterative Learning Control with Initial State Learning for Non-affine Nonlinear System

Jingcai Bai*, Xiao Yang, and Junxiao Wu

Henan Mechanical and Electrical Engineering College, Xinxiang 453002, China
okbjc@163.com

Abstract. In this paper, for non-affine nonlinear system, a closed-loop PD-type iterative learning control algorithm with initial state learning is proposed and the sufficient condition for convergence is put forward. Using the contraction mapping method, it is proved that the output of the system with an arbitrary initial state can track the desired trajectory completely after iteration. The problem of convergence with initial state unknown is solved. The simulation results testify that the proposed algorithm is effective.

Keywords: Iterative learning control, Initial state learning, Non-affine nonlinear system.

1 Introduction

Iterative learning control (ILC) is an attractive technique which doesn't depend on the accurate mathematical model of plant. It can make the output of the system approach ideal value as far as possible by multiple iterations to generate optimal input signal and by repeatedly performing the same task to reduce the error. It comes from the repeat tracking control of nonlinear system, with strong engineering background. The initial conditions of ILC are the limited conditions of system initial point at the start of each iterative process in order to guarantee the convergence of the control system. That is the relationship of system initial state and convergence in the control process. It has received a great deal of attention from many researchers, and many articles about it have been published in recent years [1].

The research of initial value problem is mainly divided into three types, namely: fixed initial value problem, any initial value problem and initial value unknown problem. Early research focused on the fixed initial value problem. At present, more research is any initial value problem, that is, the initial value deviates from the desired state, not a fixed location, or changes in the neighborhood of expected initial state, or changes over a large area, which is still strict positioning in some sense. Although many documents have relaxed the initial conditions, the widest initial condition does not require the initial positioning operation, that is, the initial value is unknown, for the initial values of many systems can not be measured or are difficult to measure. There are a lot of literature on this issue now, such as a class of affine systems using the

* Corresponding author.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 207–214.
springerlink.com

open-loop D-type, PD-type iterative learning algorithm were studied respectively in [2,3,4], a class of affine nonlinear systems using the closed-loop PD-type algorithm were studied in [5,6]. But in many cases, the system is like a darkroom, most of the information is unknown, that is, the system is non-affine and nonlinear. The most basic ILC algorithms are the open-loop and closed-loop. But the open-loop algorithm is less robust without feedback, more iterative times and longer cycle, early in the iteration divergence and excessive overshoot are prone to occur, limiting the application of ILC. And the closed-loop algorithm uses the last control and the current tracking error, the learning speed is greatly increased.

In this paper, the closed-loop PD-type ILC algorithm with initial state learning is used to control the most general non-affine nonlinear system and proved in convergence, not only the system control input learning but also the iterative initial state learning. The learning factor of the proposed initial state learning law doesn't depend on the input learning law and any information of system. The initial state can be set arbitrarily in convergence conditions, so as to ensure the robustness of the system on the initial positioning. Finally, the simulation results illustrate the effectiveness of the propose algorithm.

2 Problem Formulation

Consider the following non-affine nonlinear system

$$\begin{cases} \dot{x}(t) = f(x,u,t) \\ y(t) = g(x,u,t) \end{cases} \tag{1}$$

where $t \in [0,T]$, $x \in R^n$, $u \in R^r$, $y \in R^m$ are state, input, and output variables respectively. Assume that all the parameters above-mentioned satisfy the properties and bounds stated as follows.

Assumption 1. The function $f(x,u,t)$ is uniformly semi-globally Lipschitz in x and u, that is, there exists a constant $l_f > 0$, such that for $x_1, x_2, u_1, u_2 \in R^n$

$$\left\| f(x_1,u_1,t) - f(x_2,u_2,t) \right\| \le l_f \left(\left\| x_1 - x_2 \right\| + \left\| u_1 - u_2 \right\| \right).$$

Assumption 2. The function $g(x,u,t)$ is uniformly semi-globally Lipschitz in x and u, that is, there exists a constant $l_g > 0$, such that for $x_1, x_2, u_1, u_2 \in R^n$

$$\left\| g(x_1,u_1,t) - g(x_2,u_2,t) \right\| \le l_g \left(\left\| x_1 - x_2 \right\| + \left\| u_1 - u_2 \right\| \right).$$

Assumption 3. For all x and u, $h_u(x,u) = \dfrac{\partial h(x,u)}{\partial u} \neq 0$, $h_x(x,u) = \dfrac{\partial h(x,u)}{\partial x} \neq 0$, and $h_u > 0$, $h_x > 0$.

Assumption 4. There exists the only expected input $u_d(t)$ such that the corresponding state variables are just $x_d(t)$ and the output is $y_d(t)$.

The initial value of each iteration is different, the k-th iteration initial value is $x_k(0)$.

The tracking error is

$$e_k(t) = y_d(t) - y_k(t) \tag{2}$$

The learning law of initial state is

$$x_{k+1}(0) = x_k(0) + Le_k(0) \tag{3}$$

The input learning law using the closed-loop PD-type ILC law is as follows:

$$u_{k+1}(t) = u_k(t) + P_c e_{k+1}(t) + D_c \dot{e}_{k+1}(t) \tag{4}$$

When $k \to \infty$, the tracking error $y_d(t) - y_k(t)$ can converge to zero for all $t \in [0,T]$.

3 Convergence Analysis

Solving state equation, we denote $\delta x_k = x_k - x_{k+1}$, then

$$\delta x_k = x_k - x_{k+1} = x_k(0) - x_{k+1}(0) + \int_0^t [f(x_k, u_k, t) - f(x_{k+1}, u_{k+1}, t)] d\tau \tag{5}$$

Taking the norm of the above equation, and substituting (3) and (4) into (5), then we have

$$\begin{aligned} \|x_k - x_{k+1}\| &\le \|x_k(0) - x_{k+1}(0)\| + l_f \int_0^t (\|x_k - x_{k+1}\| + \|u_k - u_{k+1}\|) d\tau \\ &\le L\|e_k(0)\| + \int_0^t \left[l_f \|x_k - x_{k+1}\| + l_f (P_c e_{k+1} + D_c \dot{e}_{k+1}) \right] d\tau \end{aligned} \tag{6}$$

According to the Lemma of Bellman-Gronwall, we know that

$$l_f \int_0^t e^{l_f(t-\tau)} \|D_c \dot{e}_{k+1}\| d\tau = l_f D_c \bullet \left[\|e_{k+1}\| - \|e_{k+1}(0)\| e^{l_f t} + l_f \int_0^t e^{l_f(t-\tau)} \|e_{k+1}\| d\tau \right] \tag{7}$$

Substituting (7) into (6), and taking the norm, then

$$\|x_k - x_{k+1}\|_\lambda \le L\|e_k(0)\|_\lambda + l_f D_c \|e_{k+1}(0)\|_\lambda + \left[l_f D_C + \left(l_f P_C + {l_f}^2 D_C \right) \frac{1 - e^{(l_f - \lambda)T}}{\lambda - l_f} \right] \|e_{k+1}\|_\lambda \tag{8}$$

Let

$$a_1 = l_f D_c,\ a_2 = l_f D_C + \left(l_f P_C + {l_f}^2 D_C\right)\frac{1-e^{(l_f-\lambda)T}}{\lambda - l_f}.$$

Thus, (8) can be written as

$$\left\|x_k - x_{k+1}\right\|_\lambda \le L\left\|e_k(0)\right\|_\lambda + a_1\left\|e_{k+1}(0)\right\|_\lambda + a_2\left\|e_{k+1}\right\|_\lambda \tag{9}$$

Theorem: For system (1), if the iterative learning laws are given as (3) and (4), and all the parameters satisfies Assumption 1 to 4, and there exists ρ such that

$$\frac{m_1 c}{1 - m_2 c - m_3 c a_2} \le \rho < 1 \tag{10}$$

Then, for all $t \in [0,T]$

$\lim_{k\to\infty} y_k(t) = y_d(t)$.

Proof: Let $\delta u_k = u_{k+1} - u_k$, according to Taylor theorem, we know

$$\begin{aligned} y_{k+1} - y_k &= h(x_{k+1}, u_{k+1}, t) - h(x_k, u_k, t) \\ &= h(x_k + \delta x_k, u_k + \delta u_k, t) - h(x_k, u_k, t) \\ &= \delta x_k h_x(\xi_k) + \delta u_k h_u(\xi_k) = \delta x_k h_x(\xi_k) + (P_c e_{k+1} + D_c \dot{e}_{k+1}) h_u(\xi_k) \end{aligned} \tag{11}$$

where $\xi_k = (x_k + \gamma\delta x_k, u_k + \gamma\delta u_k, t), 0 \le \gamma \le 1$, then the k-th tracking error is

$$e_{k+1} = e_k + y_k - y_{k+1} = e_k - \delta x_k h_x(\xi_k) - (P_c e_{k+1} + D_c \dot{e}_{k+1}) h_u(\xi_k) \tag{12}$$

Let $h_x = h_x(\xi_k),\ \ h_u = h_u(\xi_k)$.

Then, the above equation can be written as

$$h_u D_c \dot{e}_{k+1} = e_k - \left[1 + h_u P_c\right] e_{k+1} - \delta x_k h_x \tag{13}$$

Taking the integration of (13), we have

$$h_u D_c \left(e_{k+1} - e_{k+1}(0)\right) = \int_0^t e_k d\tau - \left[1 + h_u P_c\right]\int_0^t e_{k+1} d\tau - h_x \int_0^t \delta x_k d\tau \tag{14}$$

So

$$h_u D_c e_{k+1} = h_u D_c e_{k+1}(0) + \int_0^t e_k d\tau - \left[1 + h_u P_c\right]\int_0^t e_{k+1} d\tau - h_x \int_0^t \delta x_k d\tau \tag{15}$$

Using $\left(h_u D_c\right)^{-1}$ to left-multiply both sides of (15) and taking the norm, then it becomes

$$\begin{aligned}\|e_{k+1}\| &\le \|e_{k+1}(0)\| + \left\|\left(h_u D_c\right)^{-1}\right\| \int_0^t \|e_k\| d\tau + \\ &\left\|\left(h_u D_c\right)^{-1}\left(1+h_u P_c\right)\right\| \int_0^t \|e_{k+1}\| d\tau + \left\|\left(h_u D_c\right)^{-1} h_x\right\| \int_0^t \|\delta x_k\| d\tau \\ &= \|e_{k+1}(0)\| + m_1 \int_0^t \|e_k\| d\tau + m_2 \int_0^t \|e_{k+1}\| d\tau + m_3 \int_0^t \|\delta x_k\| d\tau \end{aligned} \quad (16)$$

where $m_1 = \left\|\left(h_u D_c\right)^{-1}\right\|$, $m_2 = \left\|\left(h_u D_c\right)^{-1}\left(1+h_u P_c\right)\right\|$, $m_3 = \left\|\left(h_u D_c\right)^{-1} h_x\right\|$.

Using $e^{-\lambda t}\left(t \in [0,T]\right)$ to multiply both sides of (16) and taking the λ-norm, then it becomes

$$\|e_{k+1}\|_\lambda \le \|e_{k+1}(0)\|_\lambda + m_1 \frac{1-e^{-\lambda T}}{\lambda}\|e_k\|_\lambda + m_2 \frac{1-e^{-\lambda T}}{\lambda}\|e_{k+1}\|_\lambda + m_3 \frac{1-e^{-\lambda T}}{\lambda}\|\delta x_k\|_\lambda \quad (17)$$

Let $c = \dfrac{1-e^{-\lambda T}}{\lambda}$, and substitute (9) into (17), we have

$$\|e_{k+1}\|_\lambda \le \frac{m_1 c}{1-m_2 c - m_3 c a_2}\|e_k\|_\lambda + \frac{m_3 c L}{1-m_2 c - m_3 c a_2}\|e_k(0)\|_\lambda + \frac{1+m_3 c a_1}{1-m_2 c - m_3 c a_2}\|e_{k+1}(0)\|_\lambda \quad (18)$$

If λ large enough, we have $1-m_2 c - m_3 c a_2 > 0$, there must be

$$\|e_{k+1}\|_\lambda \le \rho\|e_k\|_\lambda + M_1\|e_k(0)\|_\lambda + M_2\|e_{k+1}(0)\|_\lambda$$

where $\rho = \dfrac{m_1 c}{1-m_2 c - m_3 c a_2}$, $M_1 = \dfrac{m_3 c L}{1-m_2 c - m_3 c a_2}$, $M_2 = \dfrac{1+m_3 c a_1}{1-m_2 c - m_3 c a_2}$.

The next, we analyze $\|e_k(0)\|_\lambda$.

(14) also can be written as

$$h_u D_c e_{k+1} - h_u D_c e_{k+1}(0) + \left[1+h_u P_c\right]\int_0^t e_{k+1} d\tau = \int_0^t e_k d\tau - h_x \int_0^t \delta x_k d\tau \quad (19)$$

Using $\left(h_u D_c\right)^{-1}$ to left-multiply both sides of (19) and taking the λ-norm, then it becomes

$$\left(1+m_2 c\right)\|e_{k+1}\|_\lambda - \left\|e_{k+1}(0)\right\|_\lambda \le m_1 c\|e_k\|_\lambda + m_3 c\|\delta x_k\|_\lambda \quad (20)$$

When $t=0$, $\|e_{k+1}(0)\|_\lambda \le \dfrac{m_1 c + m_3 c L}{m_2 c}\|e_k(0)\|_\lambda$.

When $\frac{m_1 c + m_3 cL}{m_2 c} < 1$, $\lim_{k\to\infty} \|e_k(0)\| = 0$, simultaneously choosing sufficiently large positive λ, so that $\rho < 1$, then the proposed theorem is proved.

4 Simulations

We consider the following non-affine nonlinear system

$$\begin{cases} \dot{x}(t) = 5x + \sin(3u\cos x) \\ y(t) = 2(x+u) \end{cases}$$

The desired trajectory is set as following:

$$y_d(t) = t^2(1-3t), t \in [0,1]$$

Suppose the initial state of each iterative is $x(0) = 0.45$, and the initial control is $u(0) = 0$.

In order to verify the control performance of the closed-loop ILC algorithm, under the same conditions the system is simulated with the open-loop and closed-loop algorithm respectively.

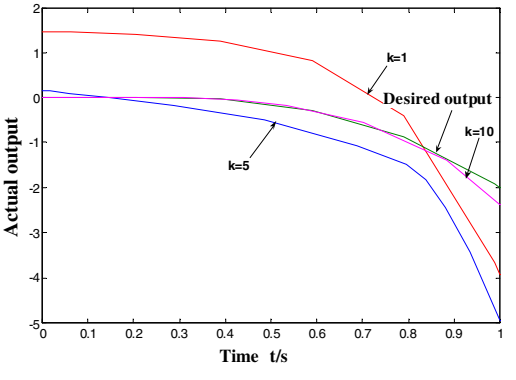


Fig. 1. Tracking curves with open-loop PD-type ILC

The above studied system is controlled using ILC with initial state learning. Figs. 1-3 show the simulation results. Through the contrast of Figs. 1 and 2, it can be seen that under arbitrary initial conditions, the actual output of the open-loop system don't track the desired trajectory completely after 10 iterative learning. But only after 5 times iteration with the closed-loop algorithm, the actual output can fully track the desired output. It also can be seen from Figs. 3 that the maximum output error decreases gradually with the increase in the number of iteration. The learning speed of the

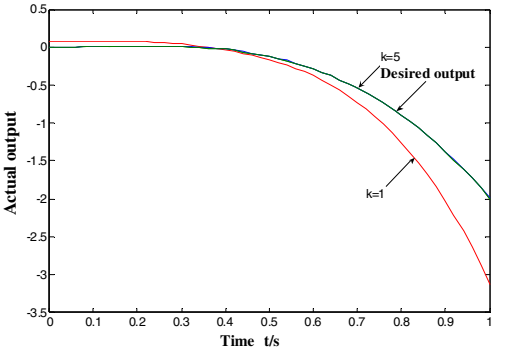


Fig. 2. Tracking curves with closed-loop PD-type ILC

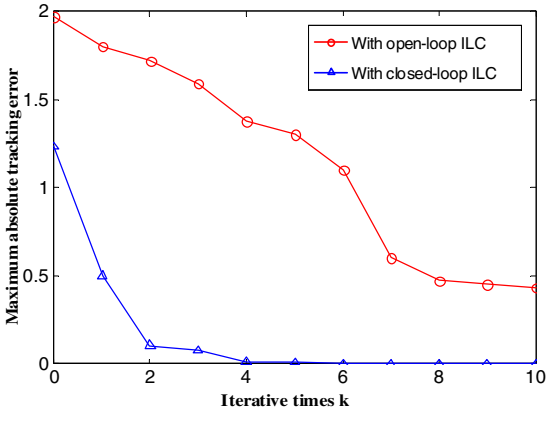


Fig. 3. Output error curves

open-loop algorithm slows down after several iterations and the tracking error can't approach zero. While, the maximum output error of the closed-loop algorithm can tend to zero with the increase in the number of iterations.

5 Conclusions

The closed-loop PD-type ILC algorithm with initial state learning for non-affine nonlinear systems is discussed in this paper. With this algorithm the actual output of the system tracking the desired output doesn't depend on the desired state and the desired input. In the case of the initial value is unknown, the initial state learning allows a certain positioning error between the initial state and the desired state at the beginning of each iterative process, thereby relaxing the requirements and ensuring the robustness on the initial state positioning. This is different from the conventional ILC which requires the initial conditions must be positioned in a specific location or a neighborhood of expected initial state. The algorithm still maintains the unique

advantage of not knowing the exact mathematical model of the system. The proposed closed-loop PD-type ILC algorithm adopting appropriate feedback, not only can ensure the stability of the system and speed up the learning convergence, but also can use the internal advantage of feedback restraining the effects of interference to enhance the robustness of ILC. Finally, the simulation results demonstrate the effectiveness of the proposed algorithm.

References

1. Sun, M., Huang, B.: Iterative Learning Control. National Defense Industry Press, Beijing (1999)
2. Sun, Y., Li, Z.: Open-loop D-Type Iterative Learning Control for a Class of Nonlinear Systems with Arbitrary Initial Value. Journal of System Simulation 20, 6767–6770 (2008)
3. Sun, M.: Iterative learning control with initial state learning. Control and Decision 22, 848–852 (2007)
4. Sun, Y., Li, Z.: Open-Loop PD-Type Iterative Learning Control for a C lass of Nonlinear Systems With Control Delay and Arbitrary Initial Value. Measurement & Control Technology 28, 51–57 (2009)
5. Cao, W., Dai, X.: Closed-loop PD-type Iterative Learning Control with Initial State Learning. Journal of Wuhan University of Technology 32, 98–102 (2010)
6. Li, S., Sun, M.: Iterative learning control with initial state learning for non-affine nonlinear system. Journal of Zhejiang University of Technology 8, 268–272 (2010)

Non-Additivity Effects Analysis on Risks of Construction Schedule

Junyan Liu[1] and Huifeng Chen[2]

[1] Institute of Engineering Management, Hohai University, Nanjing, China
junyan.liu@unsw.edu.au, l_junyan@live.cn
[2] Qingdao Maritime Safety Administration, Qingdao, China
hfchen025@foxmail.com

Abstract. Many project activities can be affected by uncontrolable factors. Those risk factors may occur individually or simultaneously, existing research is limited in considering that when several risks take palce at the same time, the total impacts on project activity duration are non-additive impacts or additive impacts, and appropriate methodologies for analyzing non-additivty are not existent. In this paper, Bayesain network are proposed to settle this issue, and the results from the calculating example shows that there is non-additive effects when many risks impact on a activitiy simultaneously.

Keywords: Construction project, Schedule; Risk analysis, Bayesian network, Non-additivity.

1 Introduction

During the process of construction, project activities can be impacted by one risk factor or several risk factors simultaneously, in academic circles and the project field, there is no research to test whether the total impacts is the summation of their individual impacts or not. Without knowing additivity or non-additivity effects will reduce the accuracy of measures adopted by decision-maker when risk events occur.

So far, non-additivity was analyzed in the field of biological engineering, medical science research, ecology system. [1] analyzed the gene expression non-additivity in immature ears of a heterotic F1 maize hybrid; [2] showed an non-additivity analysis on biological chain; [3] discussed the non-additivity in protein–DNA binding; [4] showed that the nonadditivity relation of the Tsallis entropies in nonextensive statistical mechanics has a simple physical interpretation for systems with fluctuating temperature or fluctuating energy dissipation rate; [5] proposed that many major questions in quantum information theory can be formulated as additivity problems; [6] investigated the effects of correlations between the outcomes in different options, then analyzed the effects of investments that are fungible across project options.

In this paper, non-additivity is taken into account in analyzing the schedule risks, and Bayesian network(BBN) is adopted to be a method for analyzing non-additive impacts when serveral risks arise at the same time.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 215–222.
springerlink.com

2 Bayesian Network

Bayesian networks (BBNs) were be proposed first by Pearl[7, 8], which also called Bayesian belief networks. BBNs are graphical tools used to represent a high-dimensional probability distribution. They are convenient for making inferences about uncertain states when limited information is available [9]. So far, BBN have been used for making diagnosis in medica[10] and engineering applications[11], fault prediction [12], artificial intelligence [13], and are common in system reliability assessment[14]. However, more research need to be done in the area of civil engineering, espcially in construction risk management.

A. Concept of BBN

A BBN is a directed acyclic graph(DAG)(illustrated in Fig.1), it provids two kinds of information, one is qualititative information which is defined by a DAG to show the direct independent and dependent relationship between variables, another is quantitative information which is described by conditional probibilites to show the correlation of variables.

A BBN is composed by notes and links. Notes represent stochastic variables and links express the dependence relationships between variables (X) [15]. A note without any input arrow is called root note; the one with input arrows is called subnote or child and the one with a link pointing to other notes is called parent described as $\pi(X)$. Each note contains two kinds of data, possible states and the probability under each state. Root-notes and subnotes have their own marginal probabilities $P(X)$ and conditional probabilities given parents notes $P(X \mid \pi(X))$ respectively. For example, in Fig.1, $P(X_3 \mid X_1, X_2)$ means the probability of X_3 given parents notes X_1 and X_2.

B. Calculation of BN

For variables $X_i (i = 1, \cdots, n)$ given $\pi(X_i)$, X_i is conditionally independent on all non-parents nodes, a joint distribution probability of *n* variables can be decompsed acording a chain rule such as:

$$P(X_1, \cdots, X_i) = \prod_{i=1}^{n} P(X_i \mid X_1, \cdots, X_{i-1}) = \prod_{i=1}^{n} P(X_i \mid \pi(X_i)) \tag{1}$$

Where when $\pi(X_i) = \varnothing, P(X_i \mid \pi(X_i))$ is marginal probability of X_i, $P(X_i)$.

In order to do an inference in BBN, prior probabilities and posterior probabilities need to be obtained.

Set X, Y are two stochastic variables, and suppose that $X = x$ and $Y = y$ that is an evidence. Before considering the evidence $Y = y$, $P(X = x)$ the probability of the event $X = x$ should be estimated first, that is a prior probability. After taking into account of the evidence $Y = y$, $P(X = x \mid Y = y)$ the probability estimation of $X = x$

is a posterior probability. Acording to Bayes theorem, the calculation formula of a posterior probability is:

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{P(X = x)P(Y = y \mid X = x)}{P(Y = y)} \tag{2}$$

Where $P(X = x, Y = y)$ is the probability of the joint event $X = x \wedge Y = y$ [15]. If X, Y are independent, then $P(X = x \mid Y = y) = P(X = x)$.

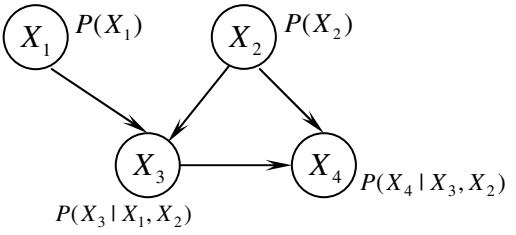


Fig. 1. A simple BBN

C. Building BBN

To built a BBN, the probabilities of variables and a construction should be detemined using expert elicitation or historical data if available. This information permits the determination of the representative variables, their possible states and probability estimates for the construction of the network. The general process to construct a BBN is as follows[16]:

1). Define the relevant variables $\{X_1, X_2, \cdots, X_n\}$ and their order $\alpha = \langle X_1, X_2, \cdots, X_n \rangle$, where X_1 is the first in the ordering, X_2 is the second, etc. Note that different order correspond to different complicated structure. Therefore, it is recommended that dependent variables are considered first in the order, then utilize the cause and effect relationship to confirm next one.

2). Define the relationship among variables. For each variable, set $\pi(X_i)$ parents to be a subset of $\{X_1, X_2, \cdots, X_{i-1}\}$ such that X_i is conditional independence to all other members of $\{X_1, X_2, \cdots, X_{i-1}\}$ given $\pi(X_i)$.

3). Define the states of the variables.

4). Estimate conditional probabilities of the relationships in a probability table. For example, a table for node X_3 in Fig.1 shown as Table 1. Assuming each variable with two states, where State_{11} means the first state of X_1, State_{12} means the second state of X_1,etc..

Table 1. Probability Table for Variable X_3

X_1		$State_{11}$		$State_{12}$	
X_2		$State_{21}$	$State_{22}$	$State_{21}$	$State_{22}$
X_3	$State_{31}$	0.0	0.0	0.0	0.0
	$State_{32}$	0.0	0.0	0.0	0.0

3 Non-additivity Analysis of Schedule RISKS Based on BBN

A. Calculation of construntion schedule risks

In this paper, a BBN structure of schedule risk is built using the cause and effect relationship; the probabilities and possible states of weather, labor efficiency and activity are obtained from a expert workshop(shown in Fig.2 and Table 2).

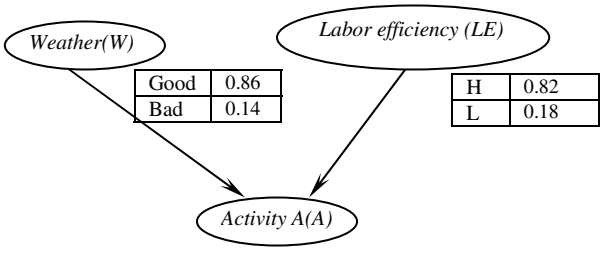


Fig. 2. A simple BBN of schedule-cost risk

Table 2. Probability Table for Activity A

Weather(W)		Good		Bad	
Labor efficiency (LE)		High(H)	Low(L)	High(H)	Low(L)
Activity A	5days	0.78	0.27	0.39	0
	10days	0.22	0.52	0.41	0.32
	13days	0	0.21	0.3	0.68

Then using (1) and (2), the posterior probabilities of schedule risk will obtain such as:

$$\begin{aligned} P(A=5days) &= P(A=5|W=good,LE=H)P(W=good)P(LE=H) \\ &\quad + P(A=5days|W=good,LE=L)P(W=good)P(LE=L) \\ &\quad + P(A=5days|W=bad,LE=H)P(W=bad)P(LE=H) \\ &\quad + P(A=5days|W=bad,LE=L)P(W=bad)P(LE=L) \\ &= 0.78\times0.82\times0.86+0.27\times0.86\times0.18+0.39\times0.14\times0.82+0 \\ &= 0.64 \end{aligned}$$

Similarly,

$$P(A=10days)=0.29,$$
$$P(A=13days)=0.07.$$

B. Non-additivity effects analysis

Case 1:
Suppose that weather condition changes, the probabbility of the state "bad" is 1, and the situation of labor efficiency changes nothing, then probabilities of the duration of Activity A were changed accordingly, the calculation as follows:

$$\begin{aligned}P(A=5days)&=P(A=5days\,|\,W=good,LE=H)P(W=good)P(LE=H)\\&\quad+P(A=5days\,|\,W=good,LE=L)P(W=good)P(LE=L)\\&\quad+P(A=5days\,|\,W=bad,LE=H)P(W=bad)P(LE=H)\\&\quad+P(A=5days\,|\,W=bad,LE=L)P(W=bad)P(LE=L)\\&=0+0+0.39\times1\times0.82+0=0.3198\end{aligned}$$

Similarly,

$$P(A=10days)=0.3938,$$
$$P(A=13days)=0.2864.$$

Case2:
Suppose that labor efficiency changes, the probabbility of the state "Low" is 1, and the situation of weather condition changes nothing, then probabilities of the duration of Activity A as follows:

$$\begin{aligned}P(A=5days)&=P(A=5days\,|\,W=good,LE=H)P(W=good)P(LE=H)\\&\quad+P(A=5days\,|\,W=good,LE=L)P(W=good)P(LE=L)\\&\quad+P(A=5days\,|\,W=bad,LE=H)P(W=bad)P(LE=H)\\&\quad+P(A=5days\,|\,W=bad,LE=L)P(W=bad)P(LE=L)\\&=0+0.27\times0.86\times1+0+0=0.2322\end{aligned}$$

Similarly,

$$P(A=10days)=0.492,$$
$$P(A=13days)=0.2758.$$

Case3:
Suppose that labor efficiency in "Low" and weather condition is "bad" simultaneously, then probabilities of the duration of Activity A are:

$$\begin{aligned}P(A=5days)&=P(A=5days\,|\,W=good,LE=H)P(W=good)P(LE=H)\\&\quad+P(A=5days\,|\,W=good,LE=L)P(W=good)P(LE=L)\\&\quad+P(A=5days\,|\,W=bad,LE=H)P(W=bad)P(LE=H)\\&\quad+P(A=5days\,|\,W=bad,LE=L)P(W=bad)P(LE=L)\\&=0+0+0+0=0\end{aligned}$$

Similarly,

$$P(A=10days)=0.32,$$
$$P(A=13days)=0.68.$$

Showing as Table 3, when "weather", "labour efficiency" in worse state respectively or simultaneously, the increment summation of case 1 and case 2 (0.7212) is bigger than the increment of case3(0.6366).

Table 3. Non-additivity impacts analysis (1)-(3) for Activity A

A	initial	Δcase1	Δcase2	Δcase3	Δcase1+case2
5days	0.64	0.3168	0.4044	0.6366	0.7212
10days	0.29	-0.1030	-0.2012	-0.0292	-0.3043
13days	0.07	-0.2138	-0.2032	-0.6074	-0.4169

Case 4:
Assume that "weather" in "good" state and "labour efficiency" with the original probabilities of its states, then probabilities of the duration of Activity A shown as:

$$P(A = 5days) = 0.6882,$$
$$P(A = 10days) = 0.274,$$
$$P(A = 13days) = 0.0378.$$

Case 5:
Assume that "labour efficiency" in "High" state and "weather" with the original probabilities of its states, then probabilities of the duration of Activity A such as:

$$P(A = 5days) = 0.7254,$$
$$P(A = 10days) = 0.2466,$$
$$P(A = 13days) = 0.028.$$

Case 6:
Assume that "labour efficiency" in "High" state and at the same time "weather" in "good" state, then probabilities of the duration of Activity A as:

$$P(A = 5days) = 0.78$$
$$P(A = 10days) = 0.22,$$
$$P(A = 13days) = 0.$$

Showing as Table 4, when "weather", "labour efficiency" in good state respectively or simultaneously, the increment summation of case 4 and case 5 (0.1404) was less than the increment of case6 (0.1434).

Table 4. Non-additivity impacts analysis (4)-(6) for Activity A

A	initial	Δcase4	Δcase5	Δcase6	Δcase4+case5
5days	0.64	-0.0516	-0.0888	-0.1434	-0.1404
10days	0.29	0.0168	0.0442	0.0708	0.0609
13days	0.07	0.0348	0.0446	0.0726	0.0794

Therefore compareing Table 3 with Table 4, the conclusion is that when risk factors occurs simultaneously with not good state, the total impacts are less than the summation of their individual impacts; however when risk factors in good states at the same time, the total impacts are bigger than the summation of their individual impacts.

4 Conclusion

In light of the analysis above, the result is that when serveral risks occured simultaneously, the total impacts are not equal the summation of their individual impacts. In same cases the total impacts are bigger than the summation of their individual impacts, however sometime the total impacts are less than the summation of their individual impacts, so decision-makers could make more acceptable descions corresponding to risk events.

Acknowledgments. Financial assistance was provided by Excellent Doctoral Dissertation Project of Hohai University and Graduate Education Innovation Project of Jiangsu Province (CX09B_058R).

References

1. Pea, G., Ferron, S., Gianfranceschi, L., et al.: Gene expression non-additivity in immature ears of a heterotic F1 maize hybrid. Plant Science 174, 17–24 (2008)
2. Vonesh, J.R., Osenberg, C.W.: Multi-predator effects across life-history stages:non-additivity of egg- and larval-stage predation in an African treefrog. Ecology Letters 6, 503–508 (2003)
3. Flanagan, R.A.O., Paillard, G., Lavery, R., et al.: Non-additivity in protein-DNA binding. Bionformatics 21, 2254–2263 (2005)
4. Beck, C.: Non-additivity of Tsallis entropies and fluctuations of temperature. Europhysics Letters 57, 329–333 (2002)
5. Shor, P.W.: Equivalence of additivity questions in quantum information theory. Comm. Math. Phys. 246, 453–472 (2004)
6. Vassolo, R.S., Anand, J., Folta, T.B.: Non-Additivity in Portfolios of Explotation Activities:A Real Options-Based Analysis of Equity Alliances In Biotechnology. Strategic Management Journal 25, 1045–1061 (2004)
7. Pearl, J.: Fusion, propagation and structuring in belief network. Artificial Intelligence 29, 228–241 (1986)
8. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. In: Proc. National Conference on AI, pp. 133–136 (1988)
9. Bedford, T., Cooke, R.M.: Probabilistic risk analysis: foundations and methods. Cambridge University Press, Cambridge (2001)
10. Lin, C.Y., Ma, L.H., Yin, J.X., et al.: A Medical Image Semantic Model ing Based on Hierarchical Bayesian Networks. Journal of Biomedical Engineering 26, 400–404 (2009)
11. Fan, C.F., Yu, Y.C.: BBN-based software project risk management. The Journal of Systems and Software 73, 193–203 (2004)

12. Xu, L.J., W.H.J., Long, B.: Fault prediction of complex systems based on Bayesian network. Systems Engineering and Electronics 30(4), 780–784 (2008)
13. Russell, S.J., Norvig, P.: Artificial intelligence: a modern approach. Prentice Hall/Pearson Education, Upper Saddle River, N.J. (2003)
14. Yin, X.W., Qian, W.X., Xie, L.Y.: A Method for System Rel iabiltiy Assessment Based on Bayesian Networks. Acta Aeronau Tica Et Astronau Tica Sinica 29, 1482–1489 (2008)
15. Zhang, L.W., Guo, H.P.: Introduction to Bayesin Network. Science Pulisher, Beijing (2006)
16. Liu, J.Y.: Bayesian Network Inference on Risks of Construction Schedule-Cost. In: 2010 International Conference of Information Science and Management Engineering. IEEE CS, Xi'an (2010) (in press)

Water-Saving Irrigation Intelligent Control System Based on STC89C52 MCU

Jiang Xiao and Danjuan Liu

School of Technology of Beijing Forestry University
International Conference on Remote Sensing
Beijing, China
xiaojiang56@gmail.com, liudj1987@126.com

Abstract. The level of automation of irrigation systems represent the state of development of agricultural modernization, while China's current level of agricultural modernization is low. This paper created the control model of irrigation system based on STC89C52 MCU (Micro Control Unit). The system uses the communication between PC and the microcontrollers with microcontroller control systems controlling valve switches, to achieve the purpose of saving water.

Keywords: MCU; Water-saving Irrigation; Control.

1 Introduction

The study about automation of China's water-saving irrigation is in the initial stage, and the degree of automation is low. The current development of automatic irrigation control system is still in the research and trial stage. In China, although there are a variety of irrigation controllers, most of them are small scale, limited to experimental and theoretical study, and the developed product is expensive. Farmers know these products can save energy, water and increase production, but a one-time investment is too much. Most farmers cannot afford it, so it cannot be widely applied.

This paper selects management water saving technology, according to the water requirement of crop, soil conditions and other factors, to do timely and appropriate amount of scientific irrigation, so achieve scientific and rational using of water for irrigation scheduling [1]. This system controls the irrigation volume and time to study a set of cheap and easy use water-saving irrigation control system.

2 The Working Principle and the Overall Structure

2.1 The Working Principle

The system does real-time detection of soil’s humidity through the microcomputer system, and then sends the test value to PC machine. PC machine calculates the amount of soil required for irrigation, convertes the irrigation quantity into irrigation time, then sents the time value to the MCU. The MCU system controls electromagnetic valves’

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 223–230.
springerlink.com

open and close, achieving the timing and the quantificational irrigation of soil. A PC machine can control the number of SCM (single chip microcomputer) system, achieving the distinction irrigation between different objects.

The calculation of the irrigation quantity required irrigation area, soil's bulk density, required soil's soaked depth, soil's water holding capacity of field before irrigation and the soil's moisture before irrigation. Among them, the soil bulk density, also known as "soil fake density", is the weight of unit volume without the destruction of the soil's natural structure, usually expressed in g/cm^3 [2]. Soil holding capacity, also known as water capacity, including the absorption of water and capillary water (the part absorbed and used by the root system is all called the capillary water). Soil's soaked depth is an one-time irrigation depth of wet soil, and its value varies according to crops.

Certain area of irrigation quantity is calculated as follows:

$$Q=S\times P\times D\times (FWC-SM) \tag{1}$$

Among them,
Q -- Irrigation volume, unit: m^3;
S -- Irrigated area, unit: m^2;
P -- Soil bulk density, unit: g/cm^3;
D -- Required depth of soil wetting, the unit: m;
FWC -- Pre-irrigation soil water holding capacity field;
SM -- Current value of soil's moisture.

After calculating the irrigation quantity, according to water velocity, pipe cross-sectional area and the number of nozzles, we can calculate the required time of irrigation. The formula is as follows:

$$t=Q/(v\times s\times n) \tag{2}$$

Among them,
v -- Water velocity, unit: m/s;
s -- Pipe cross-sectional area, unit: m^2;
n -- Number of nozzles.
Different targets can also use different control strategies.

2.2 The Overall Structure

Using PC's serial port (or USB-232 serial line) to achieve the communication with MCU to control a number of SCM systems, control system's overall structure is shown in Figure 1.

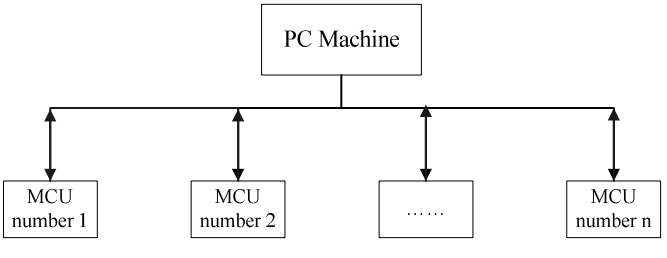


Fig. 1. The overall framework of the irrigation system

Because the slave machines are all the same microcomputer system using STC89C52 chip as the core, so this paper only describes one of the MCU system. This paper designs a set of microcontroller to communicate with PC machine. Microcontroller receives the control strategy sent by the PC machine to control valves. The block diagram of the control system is shown in Figure 2.

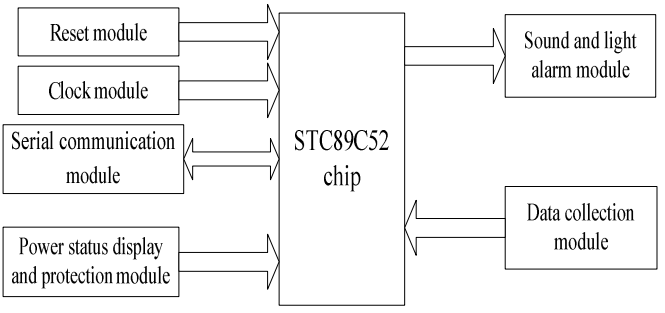


Fig. 2. Block diagram of microcomputer control system

3 The Hardware Design of Control System

Microcomputer control system consists of microcomputer minimum system, power protection circuit, moisture collecting circuit and so on [3].

3.1 Microcomputer Minimum System

SCM chooses STC89C52 as its core chip, which has 8K of program memory, 512 bytes of data memory, 8 interrupt sources, three timers/counters and a good capability/price ratio. For faster processing speed, crystal oscillator's frequency selects 14.7456MHz, so a machine cycle is only about 0.814us.

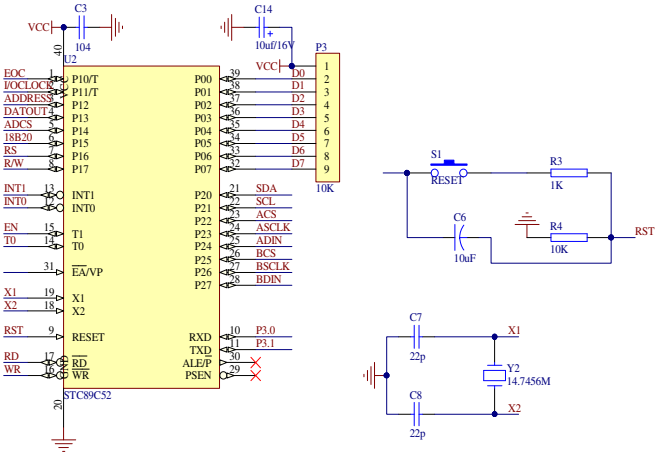


Fig. 3. Microcomputer minimum system

Microcomputer minimum system includes the main controller STC89C52, clock circuit and reset circuit, specifically shown in Figure 3.

3.2 Power Status Display and Protection Circuit

The system designs a simple over-current protection and filter circuit. The total power lines series with a fuse to prevent the over-current burning out the circuit, and filter circuit can improve anti-interference ability of SCM; besides, in lines parallel connect a light-emitting diode to display power supply work status. The circuit is shown in Figure 4.

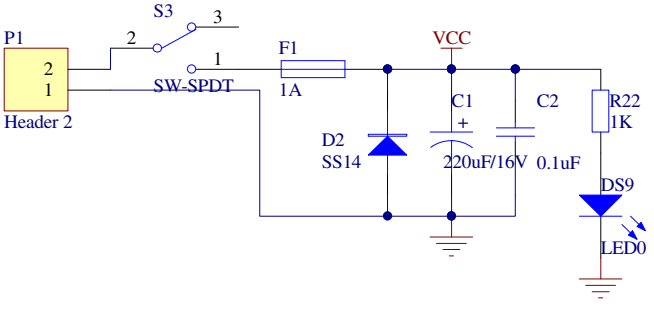


Fig. 4. Power status display and protection circuit

3.3 Data Collection Circuit

In order to convert humidity value into digital data which can be processed by the SCM, this system selects the TLC1543 as its A/D conversion chip. TLC1543 is a multi-channel and low price ADC produced by a U.S. company TI. It uses serial communication interface, has multi-input channels and is more cost-effective. It's easy to interface with microcontroller, so it can be widely applied to various data collection system. TLC1543 is able to capture 9 analog signals and it's a 10-bit switched capacitor successive approximation ADC. Signal collection circuit is shown in Figure 5.

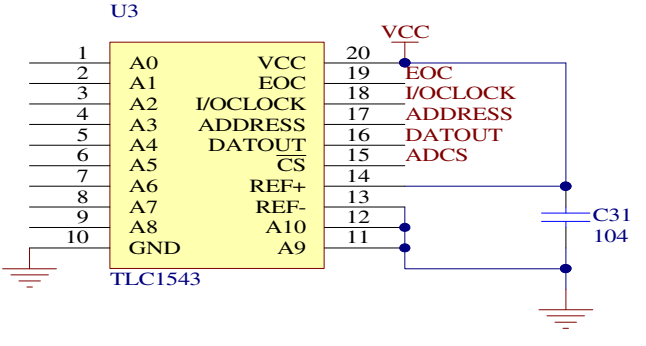


Fig. 5. Data collection circuit

3.4 Alarm Circuit

We should alarm, when the soil moisture value is too high or too low. This paper designs a simple sound and light alarm circuit, using the buzzer's sound and flashing LED lights to alarm. The circuit is shown in Figure 6.

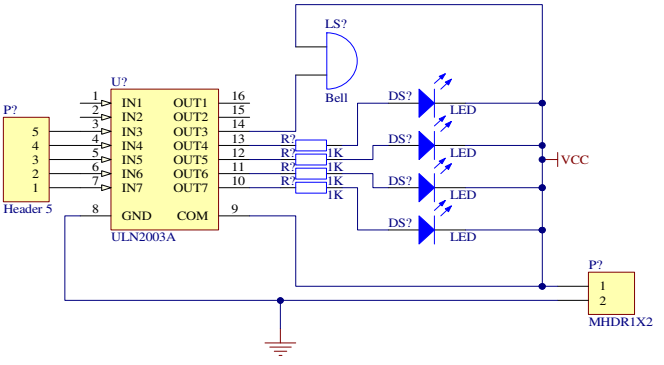


Fig. 6. Sound and light alarm module circuit

3.5 STC89C52 MCU and PC Serial Communication Interface Circuit

The communication between PC and MCU uses asynchronous serial communication standard. The interface standard is the serial bus standard RS-232 officially which is announced by the American Electronics Industry Association [4].

PC's COM port's input and output are RS-232C level, while STC89C52 Serial port's input and output are TTL level. Because TTL level and RS-232C level are not compatible, we need level conversion. Level converter chip selects the MAX232 chip, which is produced by company MAXIM, including two-way receiver and driver IC chip. Its internal has a power supply voltage converter, needing +5 V power supply, and enabling convert TTL level to RS-232 level, and also enabling convert RS-232 level to TTL level, and using easily. The circuit is shown in Figure 7.

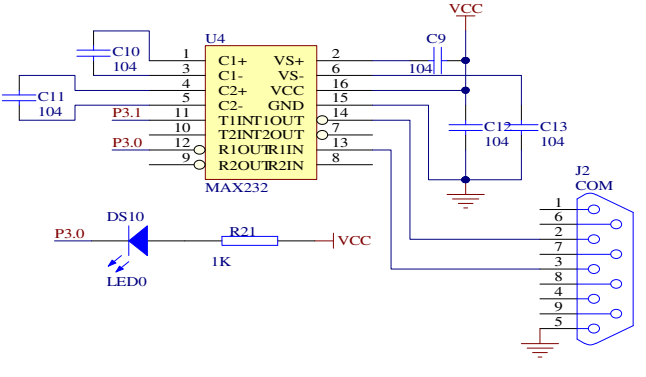


Fig. 7. Serial communication circuit

4 The Software Design

4.1 PC-Programming Design

PC machine is the operation terminal and information center of the control system. This system uses VB6.0 to program in the WindowsXP environment. Three interfaces are designed: Welcome and working condition monitoring interface, the control objects setting interface and irrigation system parameters setting interface. We can also be able to design an information repository to record the information of control object. PC's main program flow chart is shown in Figure 8.

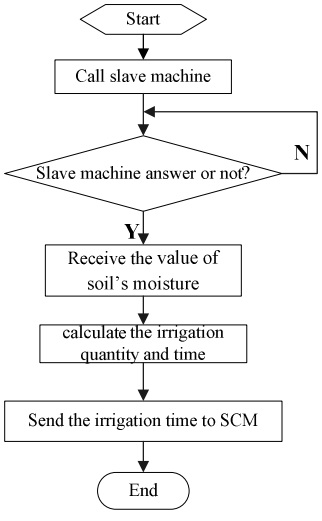


Fig. 8. The main program flow chart of the host computer

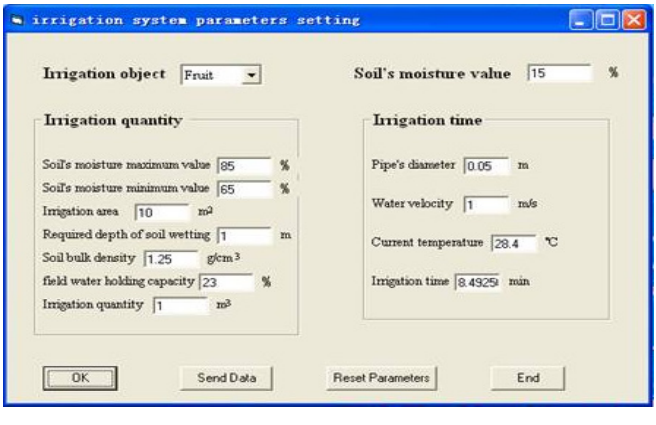


Fig. 9. Using fruit as an example of irrigation amount and irrigation time computing interface

The most important interfaces are the serial port parameters settings interface and the irrigation parameters setting interface, through setting the parameters of MSCOMM control to achieve serial communication between PC computer and mircrocomputer. For example, a sprinkler irrigates 10m2 of land, required soil moisture depth is 1m, field capacity is 0.23 and the value of soil moisture is transferred to 0.15. According to the equation (1) and equation (2) can figure out its irrigation amount and irrigation time. The interface is shown in Figure 9.

4.2 Microcontroller Programming Design

SCM software programming design includes: the main program design, sampling subroutine design, data processing subroutine design, alarm subroutine design, serial communication subroutine design and so on.

MCU's main program flow chart is shown in Figure 10.

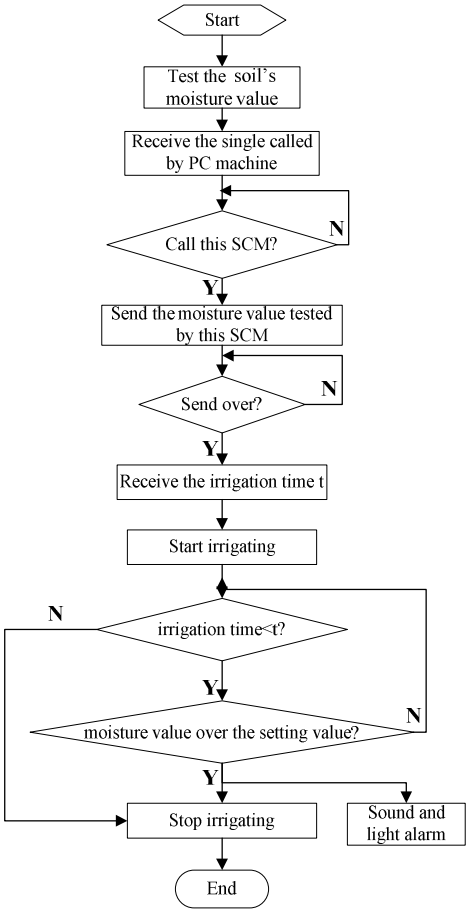


Fig. 10. MCU main program flow chart

5 Conclusions

The system achieves accurate quantitative and timing irrigation, up to the purpose of water saving. SCM system can use modular design. It's easy to use, cost low, flexible, easy to operate and very reliability, so it is easy to propagate.

References

1. http://www.nx12396.cn/itemdate
2. Baidu Encyclopedia
3. Zhang, J.: Single chip intermediate course - principles and applications, 2nd edn. Beijing University of Aeronautics and Astronautics Press
4. Zhang, H., Wang, S., Lu, L., Liu, C.: 51 microcontroller and PC serial communication system design. Chemical Engineering 32(4), 39–41 (2005)
5. Pan, X., Wang, Y.: Microcomputer control technology. Higher Education Press
6. Yang, S.: Principle and application of micro-computer system (version 2). Tsinghua University Press
7. Kang, H., Zou, S., Qin, Z.: Basic digital electronics section, 5th edn. Higher Education Press

Strength Properties of Structural FJ Chinese Fir Lumber

Haibin Zhou

Research Institute of Wood Industry, Chinese Academy of Forestry, Beijing 100091, P.R. China
zhouhb@caf.ac.cn

Abstract. As a good building material, structural Finger-jointed (FJ) lumber are used mainly in structural applications. The paper evaluated strength properties of structural FJ Chinese fir lumber. Lumber was sawn from the logs following a pattern typically used in China to maximize the volume of recovered sawn timbers. After kiln-drying, the rough-sawed lumber was planned to 4.5 cm thick, 9.0 cm wide and 100cm long. The lumber pieces were assigned to two groups according to their dynamical MOE. FJ lumber was produced by cutting a series of sloping fingers on the end of the wood pieces to be joined and interlocking the two pieces by MDI glue. The FJ lumber was nondestructively tested using edge-wise bending, longitudinal and transversal vibration methods respectively. After three methods, tensile and bending strengths of FJ lumber were tested. The results showed that machine grading can help the finger joint played more active role, and dynamical MOE was a good indicator of the static MOE.

Keywords: finger-jointed lumber, tensile strength, bending strength, machine grading, nondestructive method.

1 Introduction

In the world the fast shift of natural forest to cultivation forest calls for efficient utilization of timber resources. For example, timber-processing industry generates large volumes of wood residue, most of which are reportedly suitable for the production of high value-added products such as finger-jointed lumber. However, in some companies solid sawmill lumber off-cuts (residue) is presently not well utilized. Structural FJ lumber is used mainly in structural applications including glue-laminated beams, wooden I-joists, open web joists and more recently in parallel chord wood trusses. Structural finger-joints were developed to reduce the waste of high-quality lumber that resulted from machining of scarf joints, and they are reported to be one of the most economical ways of wood utilization. Strength requirements vary through a wide spectrum from studs on the low end to machine-stress-rated (MSR) lumber and glulam beams on the high end. Chinese fir has the biggest grown area among all plantation species. Undeniably, it will be a potential building material for timber construction in future. The precondition of structural use of FJ Chinese fir lumber is to first determine its load-bearing strength properties according to the full-size test methods.

The bending test is considered the most convenient and practical test for an extensive preliminary study of finger-joints, and can also be used for quality control

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 231–238.
springerlink.com

after qualification. The Canadian National Lumber Grades Authority (NLGA) recommends a two-point loading test to evaluate finger-jointed lumber [1]. Although the classic static test is recognized as a more desirable method of determining wood properties, static testing may be difficult to carry out and may be time consuming. A fast, reliable, and easy-to-use method for predicting bending properties may have value in machine stress grading before or after finger jointing for FJ Chinese fir lumber produced from sawmill lumber residues or off-cuts. Nondestructive wood testing permits strength and modulus of elasticity (MOE) values of individual lumber pieces determined destructively to be correlated with MOE measured nondestructively in order to assign property values without damage due to overloading, thereby improving the efficiency of timber utilization [2]. The main objective of this study was to obtain structural performances in the strength distribution of structural FJ Chinese fir lumber and evaluate the effectiveness of using vibration techniques before and after finger jointing as means of nondestructively predicting its strength properties.

2 Materials and Methods

2.1 Materials

2.1.1 Preparation of Specimens

A total of 75 standing trees of the plantation Chinese fir (*Cunninghamia lanceolata*) with diameter at breast height ranging from 200 to 350 mm were collected from Huangshan in Anhui province. The DBH range can represent Chinese fir tree commonly sawn for commercial production in China. Lumber was sawn from the logs following a pattern typically used in China to maximize the volume of recovered sawn timbers. The lumber was then kiln-dried to an average moisture content of 12% and planed to the final dimension of 45×90×4000 mm. The 4000mm long lumber was finally sawn into four 1000 mm long specimens in length direction. The dynamical MOE of specimens were measured by the longitudinal vibration method. All specimens were separated into two sampled groups according to the target average values of Young's modulus (8.49 and 10.56 GPa respectively). The former group is called the L specimen (low-grade lumber) and the latter group H specimen (high-grade lumber) in the following section.

2.1.2 Preparation of Finer-Jointed Samples

The finger-jointing was done under factory conditions and in accordance with the NLGA rules. For each group, the 50 mm long area from an end of each specimen has no visional defects. The end was cut to make male-female type joints with different finger length *(d)* and vertical finger-joint orientation. The finger length was 20, 25 and 35 mm respectively in same quantity. The tips and roots width difference was 0.1-0.5 mm, and slope was 1 in 10. For each finger length, one third of the L-grade group were jointed with another one third of the same group; the left one third of the L-grade group were jointed with one third of the H-grade group; the left two third of the H-grade group were jointed each other. Therefore, each finger length lumber had three grade groups (L-L, L-H and H-H). A finger-jointer equipped with woodworking, gluing, and pressing components was used. The MDI glue was mixed in accordance with the supplier's

instructions and according to the normal practice in the producing factory. Glue was applied by hand on one side of the joint at a temperature of about 15℃. The joint was held open for about 60 seconds before mating and applying end pressure. Three different end pressures used for pressing different finger length was shown in Table 1. Each finger jointed lumber was placed for 24 hours in a conditioned room.

Table 1. End pressure for each finger length

Finger length [mm]	End pressure [MPa]
20	8.6
25	7.4
35	6.7

2.2 Test Methods

2.2.1 Test of Dynamic MOE

Dynamic MOE was determined by the longitudinal vibration and transversal vibration techniques for both finger-jointed samples. Dynamic MOEs of the finger-jointed lumber were tested after jointing. The technique involved introducing vibration into the specimen by mechanical impact using a hammer. For the longitudinal vibration, a microphone at the other end received the sound and transmitted it into a FFT Analyzer. For the transversal vibration, a sensor attached on the other end received the sound and transmitted it into a FFT Analyzer. The FFT Analyzer can measure the fundamental resonance frequency of each sample.

2.2.2 Static Test

Four pieces of finger jointed lumber were randomly selected from each grade group for the tensile test and the remaining pieces were bending tested. The edge-wise bending test were conducted under four-point loading using a bending test machine with a static loading capacity of 500kN. The machine was set at a crosshead speed of 20 mm/min until ultimate failure. The finger-jointed specimen was positioned on the supports such that the finger-joint was at center of the 1620mm span. Each bending specimen replicate was tested under a four-point loading arrangement in accordance with ASTM D198 [3]. Deflection was measured within the test span, using two transducers positioned at each side of the finger-joint. Each specimen was tested to destruction to determine the bending strength and MOE. Tensile tests were conducted with the tensile test machine (Metriguard Model 401). The tension machine was equipped with serrated plates to grip the specimens. Test spans are 120cm and the finger-jointed specimen was positioned at center of the span. Test time to failure was about 2-3 min.

3 Results and Discussion

3.1 Strength Properties of FJ Chinese Fir Lumber

Table 1 shows the differences of strength properties between FJ lumber and NFJ lumber. The statistical values of strength properties for FJ lumber was based on all

grade groups. The 5th percentiles of strength distributions were obtained by the nonparametric method according to ASTM D2915 [4].The results showed that they wouldn't at least reduce the strength of products under the free grouping condition that the finger jointing process met the standard requirement. Therefore, the FJ lumber with final machine grades can be assigned to the design value the same with NFJ lumber.

The pre-grading of lumber units can change the blind jointing existing in factories. When lumber units were selected by pre-grading before finger jointing, the more scientific jointing process instead of visual checking would increase the whole strength properties as shown in table 2. Compared with the L-L group, the H-H group increased 10-30% in different strength properties. In fact, if the lumber unit grades were known in advance in the practical production, the L-H products would be not recommended. Theoretically speaking, the strength properties of L-L products should be near to that of L-H ones. In table 2 the small difference of the strength properties between L-H and L-L should be caused by the testing methods as described above.

According to the design code, the L lumber units with the average value of 8.49 belong to M10 grade in GB50005 [5]. Normally the FJ lumber from its grade groups should be assigned to the allowable properties of M10. This point could also be proved that the test results were more than the characteristic values of M10 lumber in the standard. Therefore, the pre-grading method not only enhances the properties of FJ products and increases the FJ lumber products effectively; more importantly but also become a possible to cancel the grading process after jointing. The machine pre-grading process should become an important step in production of the finger jointed lumber. It will be necessary to develop the pre-grading machines especially for nonstandard and shorter lumber with given dimension range.

Table 2. The 5th percentile values of strength properties between FJ and NFJ lumber

Strength properties	FJ lumber	NFJ Lumber (No finger joints)
MOE [GPa]	7.91	7.78
Bending strength [MPa]	21.33	21.20
Tensile strength [MPa]	17.46	16.67

Table 3. Strength properties for different grade group

Strength properties		Grade group		
		L-L	L-H	H-H
MOE [GPa]	5th percentile	7.27	8.25	9.38
	50th percentile	8.88	9.70	10.77
Bending strength [MPa]	5th percentile	20.43	22.18	22.77
	50th percentile	28.57	30.03	33.66
Tensile strength [MPa]	Average	22.37	23.08	27.80

Tensile failure modes were generalized and shown in figure 1. Six failure modes [6] behaved in tensile testing, while other five failure modes happened in bending test except the fourth mode. Failure modes 1 and 2 make a distinction between less than 70 % wood failure and more than 70 % wood failure. Mode 1 is the most undesired failure for FJ products. Failure modes 5 and 6 have a difference between failure beginning from joint and away from joint.

Fig. 1. Tension failure modes

The finger profile has been a research hotspot for scientists of FJ products in past years [7,8]. The finger length is not an exception, because it is one of important parameters of the finger profile. Table 3 shows the strength properties and failure modes for different finger length. According to the general viewpoint, the longer the finger length the higher the strength properties. The 5th percentile value of MOE might be a good example. However, the similar changing behavior for other strength properties wasn't reflected in table 3.

Table 3. Strength properties for different finger length

Strength properties		Finger length [mm]		
		20 (0.1)	25 (0.3)	35 (0.5)
MOE [GPa]	5th percentile	7.08	7.83	8.54
	50th percentile	10.05	9.99	9.90
Bending strength [MPa]	5th percentile	27.57	25.39	19.77
	50th percentile	35.23	34.31	27.94
	Failure mode with maximum percentage	Mode 5	Mode 5 and 2	Mode 2 and 3
Tensile strength [MPa]	Average	25.69	25.36	23.20
	Failure modes with maximum percentage	Mode 6	Mode 5	Mode 3

Note: values in parentheses are the width difference between tips and roots (mm).

As we know, there are the spaces between fingers for side locking press by manufacturing the width difference which finger tip is more than finger root in width. But if the space is bigger, the failure would happen in the weaker areas near finger tips or roots under the ultimate load while the MOE, which determined from the elastic deformation, might not be greatly influenced. Failure mode 3 was usually caused by the bigger space. This might be the reason why the strength of FJ products with the finger length of 35mm was lower than that with finger length below 35mm.

3.2 Relationship between Strength Properties

Of course, pre-grading was just a tentative process in study. Traditionally, many studies were focused on using vibration techniques after finger jointing to nondestructively predict its strength properties. The MOEs of finger jointed lumber were statistically determined on the base of all different finger lengths and grade groups, and the results were presented in Table 4. MOE for the longitudinal vibration was a few higher than both those for transversal vibration and static bending.

Table 4. MOEs for different methods

Method	MOE [GPa]		
	5th percentile	50th percentile	Average
Longitudinal vibration	8.19	10.06	10.02
Transversal vibration	6.59	9.89	9.15
Static bending	7.96	9.82	9.79

Regression of dynamic MOE on static bending MOE for different finger lengths was performed, and the results were presented in Table 5. The ranges of correlation coefficients obtained for the finger-jointed samples were 0.42 to 0.91 for longitudinal vibration and 0.35 to 0.78 for transversal vibration respectively. These values were a little lower than those obtained for finger jointed lumber produced using the clear samples. But correlation coefficients for the regression of dynamic MOE on static bending MOE for the combined data could indicate a good correlation between dynamic MOE and static bending MOE (0.85 for longitudinal vibration and 0.72 for transversal vibration), thus confirming the linearity of the relationship between the two properties. The regression results seemed to indicate that both the correlations between statically and dynamically established MOE were strong for longitudinal vibration and transversal vibration, but the longitudinal vibration might be a better strength predictor of the static MOE of finger jointed lumber than transversal vibration.

Table 5. Summary of regression parameters for relationships between dynamic MOE and static bending MOE

Finger length	Grade	Sample size	Regression model	correlation coefficient (r)
20 [mm]	Total	83	E_s=0.8605E_l+1.0883	0.81
			E_s=0.3883E_t +6.1812	0.65
25 [mm]	Total	83	E_s=0.8611E_l+1.2146	0.85
			E_s=0.4286E_t +5.9226	0.74
35 [mm]	Total	83	E_s=0.8723E_l+1.069	0.91
			E_s=0.4785E_t +5.4052	0.77
Total		249	E_s=0.8646E_l+1.1239	0.85
			E_s=0.4286E_t +5.8688	0.72

Dynamic MOE and static bending MOE were each separately correlated to bending strength and the results were presented in Table 6. The results showed that the correlation between static bending MOE and bending strength was only slightly higher than that between dynamic MOE and bending strength for finger-jointed specimens. The regression results showed that, generally, there was low correlation between MOE and bending strength for finger-jointed lumber. The failure in the bending test wasn't often the place in the middle of the test span, but defect position such as knots, check, skip or slope of grain. H-H and L-L grade groups showed relatively higher correlation coefficient than the L-H grade group. Thus the big variation on strength along the length direction within a piece of lumber might be a main reason for the low correlation coefficient. The statistically low correlation coefficients seemingly indicated that both static bending MOE and dynamic MOE might be not good indicators of the bending strength if the finger-jointed lumber had many defects from nature, drying and process.

Table 6. Summary of regression parameters for relationships between MOE and bending strength (F_b)

Method	Grade group	Sample size	Regression model	Correlation coefficient (r)
Longitudinal vibration	Total	212	$F_b = 3.0454E_l + 16.22$	0.36
Transversal vibration	Total	212	$F_b = 1.7018E_t + 31.23$	0.35
Bending method	Total	212	$F_b = 3.6237E_s + 11.275$	0.44

Dynamic MOE and static bending MOE were each separately correlated to tensile strength and the results were presented in Table 7. The regression results showed that, generally, there was low correlation between MOE and tensile strength for finger-jointed lumber. The results showed that the correlation between dynamic MOE from longitudinal vibration and bending strength was much higher than that between static bending MOE and bending strength for finger-jointed lumber. The statistically low correlation coefficients seemingly indicated that both static bending MOE and dynamic MOE might also be not good indicators of the tensile strength of finger-jointed Chinese fir lumber.

Table 7. Summary of regression parameters for relationships between MOE and tensile strength (F_t)

Method	Sample size	Regression model	Correlation coefficient (r)
Longitudinal vibration	36	$F_t = 2.6626E_l - 1.8491$	0.60
Transversal vibration	36	$F_t = 1.1128E_t + 14.504$	0.37
Bending method	36	$F_t = 1.9476E_s + 5.8473$	0.40

4 Summary

The FJ Chinese fir lumber could increase the strength properties when it improves the wood utilization. The machine grading process should be selected as an important step in production of the finger jointed lumber in despite of before or after jointing. It is necessary to develop the grading machines especially for shorter lumber. The control of the space between finger tip and root is absolutely taken by increasing the finger length in production of FJ lumber to obtain the higher performance.

Dynamic MOE was well correlated to static bending MOE for finger-jointed Chinese fir lumber. The correlation between dynamic MOE and static bending MOE was only slightly lower for transversal vibration compared with longitudinal vibration. Both static bending MOE and dynamic MOE cannot better predict the tensile strength of finger-jointed Chinese fir lumber.

References

1. National Lumber Grades Authority. Special products standard for finger-joined structural lumber. SPS1, NLGA, Ganges, B.C., Canada (2003)
2. Bodig, J., Jayne, B.A.: Mechanics of Wood and Wood Composites. Van Nostrand Reinhold company, US (1982)
3. American Society for Testing and Materials. Standard methods of static tests of timbers in structural sizes. ASTM D 198 (2004)
4. American Society for Testing and Materials. Standard practice for evaluating allowable properties for grades of structural lumber. ASTM D 2915 (2003)
5. State Ministry of Construction. Code for design of timber structures. GB50005 (2005)
6. American Society for Testing and Materials. Standard test method for evaluating structural adhesives for finger jointing lumber. ASTM D 4688 (2005)
7. Bustos, C., Beauregard, R., Mohammad, M., et al.: Structural performance of finger-jointed black spruce lumber with different joint configuration. Forest Products Journal 53(9), 72–76 (2003)
8. Konnerth, J., Valla, A., Gindl, W., et al.: Measurement of strain distribution in timber finger joints. Wood Science and Technology 40(8), 631–636 (2006)

Experiment Study on Grinding Force of 65Mn Steel in Grinding-Hardening Machining

Judong Liu, Jinkui Xiong, and Wei Yuan

School of Mechanical Engineering, Jimei University, Xiamen 361021, China
liujd@jmu.edu.cn, xiongjinkuixjk@126.com, yuawie@qq.com

Abstract. 65Mn steel was grind-hardened on a surface grinder with corundum wheel and the changing characteristic of the grinding force during the process of grinding-hardening was studied deeply. The result shows that, the stage of steady grinding existed in precision grinding was not appeared in grinding-hardening machining. The essence of grinding-hardening is grinding burn of steels with harden ability. It is different from the precision grinding, the grinding force ratio raised with the increasing of the depth of cut during the process of grinding-hardening machining. Owing to the extrusion effect that the grinding wheel squeezed the materials of workpiece surface which was soften by high temperature, the grinding force ratio was larger and the value changed between 2.9 and 8.9. Especially in reciprocating grinding-hardening, the grinding force was far greater than precision grinding.

Keyword: Grinding, Surface hardening, Depth of cut, Grinding force, Force ratio.

1 Introduction

Grinding-hardening is a new technology that utilizes the grinding heat induced in grinding process to quench the surface of steel directly. It realizes the integrated manufacturing of grinding and surface hardening; and decreases the heat treatment process and the investment of special parts heat treatment equipment and personnel. It also reduces the pollution of heat treatment equipment emission to the environment and has remarkable economic and social benefits [1]. In recent years, scholars both at home and abroad studied on grinding-hardening machining and achieved some innovation results [1-6].

Grinding force is an important physical quantity in grinding-hardening machining. It not only affects the deformation of the grinding technology system, but also influences the energy consumption and grinding heat generation during the process of grinding-hardening and further affects the work condition of the grinding wheel's surface and the surface integrity of the grounded workpiece. It also influences the quality of the grinding-hardening, such as the microstructure of harden layer, micro-hardness, residual stress and burr size, etc. However, at present the related researches about the grinding force in the grinding-hardening were confined to the one-pass grinding-hardening experiment of 40Cr steel and 48MnV steel[7-11], and

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 239–246.
springerlink.com

researches about the changing characteristic of the grinding force and grinding force ratio were not shown. This paper based on the surface grinding-hardening experiment and analysis the change rules of the grinding force during the process of grinding-hardening in 65Mn steel. It also studies the effects of the depth of cut, table speed and grinding method on the grinding force ratio. It reveals the mechanism of the grinding-hardening processing from the view of grinding machine, and provides references for the engineering application in the selection of grinder and grinding wheel.

2 Experimental Materials and Method

The test material was steel 65Mn, initially annealed. The size of test piece was (length × width × height):100mm× 5mm×15mm. Grinding-hardening experiment was performed on a surface grinder M7130, and the grinding conditions are listed in Table 1.The parameters with underline are basic parameter. In order to ensure the reliability of date and the reproducibility of experimental results, each test condition repeated three times. The grinding force was measured online with a three -dimensional dynamometer (Kistler 9265B).

Table 1. Grinding condition

Grinding wheel	A60L6V
Wheel speed v_s[m/s]	26
Table speed v_w[m/min]	0.2, 0.4, 0.6, 0.8
Depth of cut a_p[mm]	0.2, 0.4, 0.6, 0.8
Grinding method	down grinding, up grinding, down-up grinding, up-down grinding, down-up-down grinding, up-down-up grinding
Cooling	Dry grinding

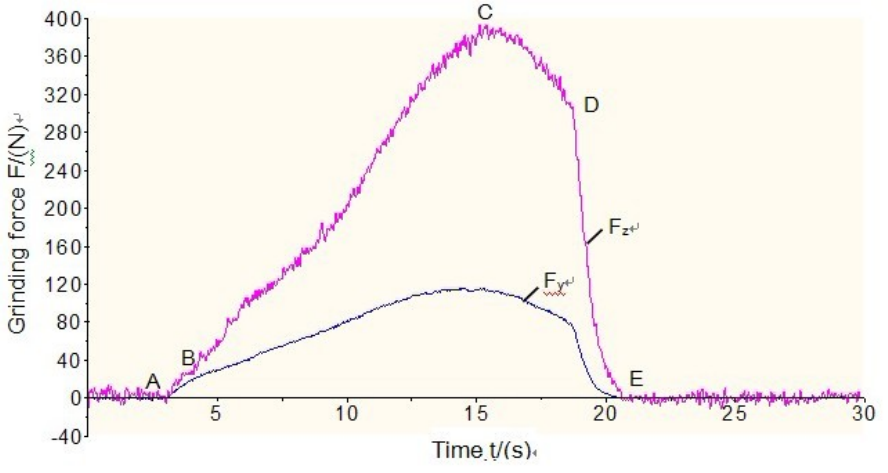
3 Results and Analysis

3.1 The Changing Characteristic of the Grinding Force

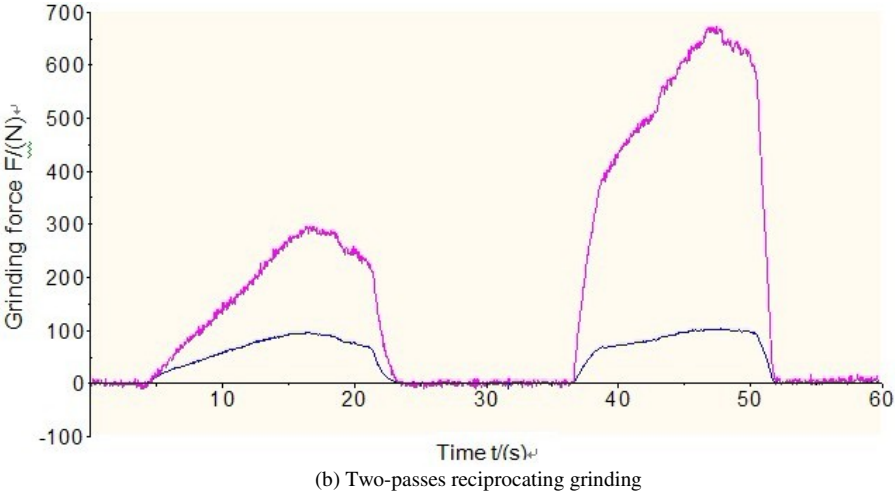
The typical signals of grinding force that measured on the grinding-hardening experiment was smoothing as shown in Fig. 1, the calculation uses the average of each signal's absolute value.

As shown in the Fig 1, in the grinding-hardening machine, the changing characteristic of vertical grinding force F_V was the same as level grinding force F_H. According to the changing characteristic, it can be divided into four stages: grinding force raised sharply stage (AB stage) , grinding force raised stage (BC stage), grinding force decreased stage (CD stage), grinding force decreased sharply stage (DE stage).Compared with precision grinding, the changing characteristic of grinding force in grinding hardening machine was caused by the change of grinding method. The concrete analysis was as follows:

(1) Grinding force raised and decreased sharply stages. These stages were located in the cut in and cut out region. From the beginning (point A) of grinding wheel contacted with the workpiece, as the increasing of the actual depth of cut, vertical grinding force F_V and level grinding force F_H raised sharply on linear. When arrived at point D, as the decreasing of the actual depth of cut, vertical grinding force F_V and level grinding force F_H decreased sharply on linear. Compared with precision grinding, owing to the lager depth of cut and littler table speed that grinding-hardening used leaded to the contact time of the workpiece and grinding wheel prolonged accordingly. Thus the lift rate was slightly small.



(a) One-pass grinding



(b) Two-passes reciprocating grinding

Fig. 1. The typical signals of grinding force in grinding-hardening machining

(2) Grinding force raised and decreased stage. These stages were located in the full cutting region. Owing to lake of cooling、lubrication and flushing of coolant, when grinding wheel arrived at the full cutting depth region (point B), a great of grinding heat passed into the surface of workpiece make the thermal expansion of the workpiece intensified and the actual depth of cut increased; at the same time, the grinding wheel was blocked more serious; on the other hand, the grinding heat make the workpiece's material in the surface layer softening intensified, and the material's yield strength decreased. Owing to the influence of the former on the growth of grinding force was more than the later, so that the grinding force continued to increase. As the grinding-hardening was going on, the temperature of the workpiece's surface layer continued to rise, and blockage of the grinding wheel tended to be saturated. Because of the influence of the soften material in the workpiece's surface layer on the grinding force and the blockage of the grinding wheel leaded to the grinding force presented downtrend. Fig 2 presents the macrostructure of grind-hardened surface layer. As shown in it, the material in the surface layer which was soften by high temperature and squeezed by the grinding wheel generated plastic flow and overlapped in the grinding surface constantly, so as to form the morphology.

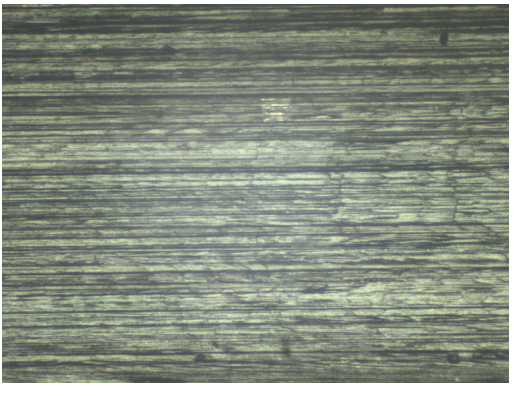


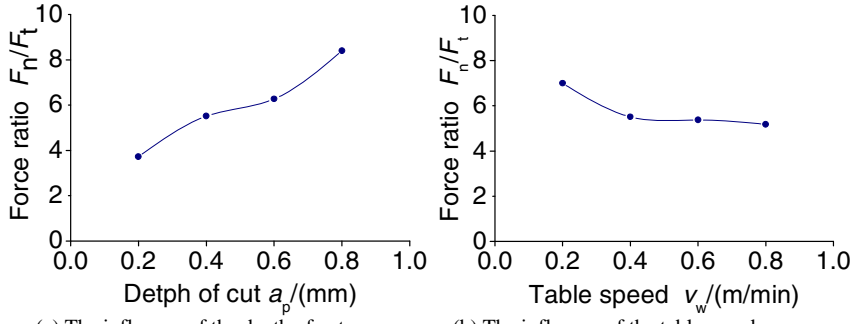
Fig. 2. The surface of grind-hardened layer ×75

The changing characteristic of the grinding force and the macrostructure of grind-hardened surface layer show that: the essence of grinding-hardening was grinding burn of steels with harden ability.

3.2 The Influence of the Depth of Cut on the grinding Force Ratio

Owing to the max value of the depth of cut was 0.8mm in this experiment, the influence on the normal and tangential grinding force could be ignored. So the vertical grinding force F_V that measured on dynamometer could take place of the normal grinding force F_n, and the tangential grinding force F_t was replaced by the level grinding force F_H. Fig 3 presents the change rule of grinding force ratio under different depth of cut and table speed. As shown in Fig 3(a), as the increasing of the depth of cut a_p, the grinding force ratio increased accordingly. When a_p raised from 0.2mm to

0.8mm, the grinding force ratio F_n/F_t increased from 3.7 to 8.4, this is different from the change rule that was mentioned in general literature. The reason is that: as the increasing of a_p, the grinding force increased, the temperature also raised, the attrition and blockage of the grinding wheel intensified, the softening and expansion of the material in the workpiece's surface increased accordingly. But compared with the tangential grinding force, the normal grinding force increased more and leaded to the grinding force ratio F_n/F_t increased with the raising of the depth of cut a_p.



(a) The influence of the depth of cut a_p (b) The influence of the table speed v_w

Fig. 3. The influence of grinding parameter on the grinding force ratio

As shown in Fig 3(b), as the raising of v_w, the grinding force F_n/F_t ratio decreased. While the v_w increased from 0.2m/min to 0.8m/min, the grinding force ratio decreased from 7.0 to 5.2. This is the same with the change rules that mentioned in general literature.

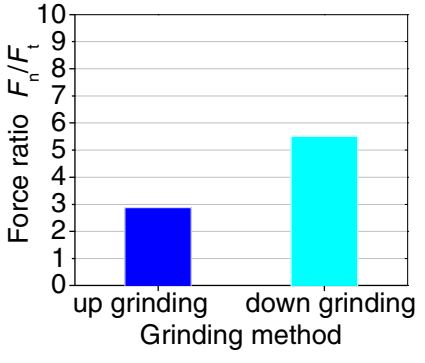
From the experiment results, it can be concluded that, compared with the table speed v_w, the influence of the depth of cut on the grinding force ratio was more obvious in grinding-hardening machine of 65Mn steel.

3.3 The Influence of Grinding Method on the Grinding Force Ratio

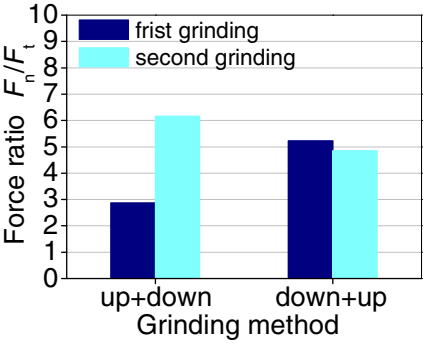
Fig. 4 presents the change rule of the grinding force ratio with different grinding method. As show in Fig. 4(a), when the other grinding conditions are the same, the grinding force ratio is 5.5 in one pass down grinding and it is 2.9 in up grinding. The grinding force ratio in down grinding is greater than that in up grinding. The reason is that: compared with up grinding, the curvature of the abrasive's motion trail、the undeformed chip's thickness and the practical negative rake angle of the abrasive was bigger in down-grinding and leaded to the increasing of the chip's deformation. Moreover, during the process of down grinding, the highest temperature of the arc area's surface was close to the middle part of the arc area, and when the abrasive began to cut the workpiece, the strength of material that located in the area with the maxim cutting thickness was relatively higher, thus leaded to the increasing of grinding force and the raising of the grinding temperature. On the other hand, the grinding temperature of the arc area was higher in down grinding [12]. These two aspects caused the thermal

expansion of the material in the workpiece's surface layer became more serious and leaded to the normal force F_n raised sharply so that the grinding force ratio increased accordingly.

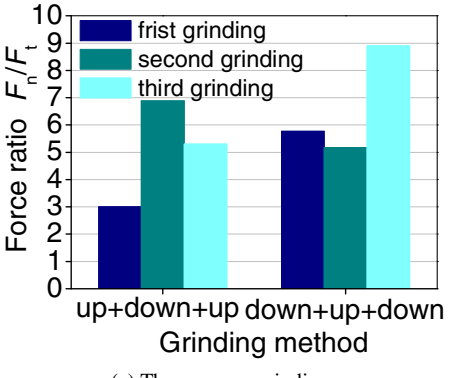
It can be seen from Fig. 4(b) and Fig. 4(c) that: when the other grinding conditions are the same, as the increasing of the grinding trip, the grinding force ratio raised accordingly no matter the grinding method was down grinding or up grinding. In two passes grinding, the grinding force ratio was 6.2 when the second grinding method used down grinding; and the value was 4.9 when it used up grinding. In three passes reciprocating grinding, when the third grinding method was down grinding, the grinding force ratio was 8.9, but the value was only 5.2 in up grinding. The reason is that: In two-passes or three passes reciprocating grinding, owing to most of grinding process was going on the previous grind-hardened layer, and the working position of the grinding wheel was not changed. Thus, as the increasing of the grinding trip, the attrition and blockage of the grinding wheel became more serious, so that leaded to the grinding force ratio increased accordingly.



(a) One-pass grinding



(b) Two-passes grinding



(c) Three-passes grinding

Fig. 4. The influence of grinding method on the grinding force ratio

4 Discussion

As known to all, In grinding ordinary steel parts, the grinding force ratio is 1.6-1.8; in grinding quenched steel parts, the grinding force ratio is 1.9-2.6 and it is 2.7-3.2 in grinding cast iron. But the grinding force ratio is as high as 2.9-8.9 when grinding-hardening 65Mn steel in this experimental condition. This is because that: in order to get satisfactory depth of hardened layer, the depth of cut and table speed that used in grinding-hardening machine usually chose the value between ordinary grinding and creep feed grinding, and used dry grinding. Due to the chips that generated in grinding-hardening process was thin and long leaded to the grinding wheel was blocked easily. Add to the grinding force of single abrasive was smaller, so that the self-sharpening of the grinding wheel was weak and the abrasive blunted quickly. These two aspects caused the grinding temperature to rise continuously. The high grinding temperature not only made the material of the workpiece soften, leaded to the "adhesion" and blockage of the grinding wheel intensified so that the cutting ability of the grinding wheel dropped sharply; but also leaded to the actual depth of cut increased owing to the thermal expansion of the material. Finally, all of this aspect made the normal grinding force rose dramatically, and the grinding force ratio increased to. Thus, it can be considered that: the grinding-hardening is the process that the grinding wheel deprived of cutting ability squeezed the material that softened by high temperature and pushed it down form the workpiece. The macrostructure of grind-hardened surface layer presented in Fig 2 and the author's research results related to the microstructure of grind-hardened layer and its residual stress [6] can prove the rationality of the above analysis.

5 Conclusion

(1) Owing to the effect that caused by the thermal expansion and softening of the material and the blockage and attrition of the grinding wheel, the grinding force in grinding-hardening machine was always in the changing process and the stage of steady grinding that existed in precision grinding was not appeared in grinding-hardening machining;

(2) In grinding-hardening machine, the changing process of grinding force contains raised and decreased stages; from the view of grinding machine, the essence of grinding-hardening was grinding burn of steels with harden ability;

(3) In this experimental condition, owing to the extrusion effect that the grinding wheel squeezed the materials of workpiece surface, the grinding force ratio is bigger and the value is between 2.9 and 8.9; especially in reciprocating grinding-hardening, the grinding force was far greater than precision grinding;

(4) It is different from the ordinary grinding, the grinding force ratio raised with the increasing of the depth of cut during the process of grinding-hardening machining.

Acknowledgments. The authors are grateful to the NSF of Fujian Province (2009J01260), Foundation for Innovative Research Team of Jimei University, China (2009A001) for support of this project.

References

1. Brockhoff, T.: Annals of the CIRP 48(1), 255–260 (1999)
2. Zarudi, I., Zhang, L.C.: Journal of Materials Science 37(18), 3935–3943 (2002)
3. Fricker, D.C., Pearce, T., Harrison, A.: Pearce and A Harrison: Journal of Engineering Manufacture 218(10), 1339–1356 (2004)
4. Chryssolouris, G., Tsirbas, K., Salonitis, K.: Journal of Manufacturing Process 7(1), 1–9 (2005)
5. Nguyen, T., Zarudi, I., Zhang, L.C.: International Journal of Machine Tools and Manufacture 47(1), 97–106 (2007)
6. Liu, J.D.: Study on the Forming Mechanism of the Grinding-Hardening Machining and Its Application Fundament (Ph.D. Dissertation, Jiangsu University, China (2005) (in Chinese)
7. Zhang, L.J.: Modern Manufacturing Engineering (7), 82–84 (2006) (in Chinese)
8. Zhang, L.: Theoretical analysis and experimental research on one-pass surface grinding-hardening. Ph.D. Dissertation, Shandong University, China (2006) (in Chinese)
9. Xiao, B., Su, H.H., Ding, W.F., Fu, Y.C., Xu, J.H.: Key Engineering Materials 316, 15–19 (2006)
10. Han, Z.T., Zhang, L.J., Gao, D.: Journal of Materials Engineering (12), 35–38 (2007) (in Chinese)
11. Yang, G., Du, C.L., Han, Z.T.: Manufacturing Technology and Machine Tool (5), 102–104 (2008) (in Chinese)
12. Wager, J.G., Gu, D.Y.: Annals of the CIRP 40(1), 323–326 (1991)

Thought Intervention through Biofield Changing Metal Powder Characteristics Experiments on Powder Characterisation at a PM Plant

Mahendra Kumar Trivedi[1], Shrikant Patil[1], and Rama Mohan. R. Tallapragada[2]

[1] Trivedi Foundation, 15111 Hayden Road Suite 160 #305, Scottsdale, AZ 85260 USA
[2] A3-103, BuddhadevVihar, Chitalsar, Manpada, Thane (West) 400607 India
shrikant@trivedifoundation.org

Abstract. In earlier papers the effect of Mr. Trivedi's thought intervention through biofield in his physical presence on the atomic, crystalline and particle characteristics of first series of transition metal powders, group four metals and carbon allotropes are discussed. In the present paper we demonstrate this unusual effect on sieve size distribution, apparent density and flow of several metal powders under PM plant conditions.

Keywords: Biofield, Sieve analysis, Apparent density, Metal powders, Particle size.

1 Introduction

Biofield is a cumulative effect exerted by human body on the surroundings. It is known that electrical currents along with their associated magnetic fields arepresent in the bodies that are complex and dynamic. These are associated with dynamical processes such as heart and brain function, blood and lymph flow, ion transport across cell membranes, and many other biologic processes on many different scales [1]. One possible influence of biofield phenomena is that they may act directly on molecular structures, changing the conformation of molecules in functionally significant ways. Another influence is that they may transfer bio-information carried by very small energy signals interacting directly with the energy fields of life, which is more recently known as the biofield [1].

Biofield transmitted by Mr. Trivedi through his thought intervention has transformed the characteristics of various living and non-living materials. The details of several scientific investigations and the results in the form of original data are reported elsewhere [2- 5].

The present paper reports the changes in the characteristics of several metal powders after exposure to the thought intervention of Mr. Trivedi thorough biofield both in his physical presence as well as from a long distance.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 247–252.
springerlink.com

2 Experimental

The experiments conducted were of two types. The first sets of experiments are performed by thought intervention of Mr. Trivedi in his physical presence on Hoganas PASI60 iron and PP. Patelcopper powders after which these are characterised. This experiment is termed as 'Thought intervention in physical presence'. In the second set of experiments the powders are at first characterised for sieve analysis, flow and apparent density (control samples). These are then kept on a table and Mr. Trivedi who was at about 100 Km distance are treated by **'**Long distance thought intervention'.

3 Results

The weights of various sieve fractions in both the control and treated (subjected to the thought intervention of Mr. Trivedi) powders are given in table 1.

It can be noticed that after treatment some sieve fractions decreased in weight while some others showed increase. The coarse sieve fraction above 152µm decreased by 10.71% in P iron powder and 100% in P copper powder. The corresponding values for L powders are respectively 54.26 and 100%. Other sieve fractions showed moderate increase and decrease.

Table 1. Comparison of sieve fractions in metallic powders treated by thought intervention in physical presence [P] as well as by long distance [L]

Powder	Mesh	Range in micrometers	Weight% control Wc	Weight% treated Wt	Percent Change 100 (wt-wc)/wc
P Iron powder	100+	152+	2.40	2.14	-10.71
	150	152-104	14.35	21.79	51.82
	200	104-76	24.00	29.69	23.72
	300	76-53	25.05	24.64	-1.63
	350	53-44	6.70	4.64	-30.70
	350-	44-	27.50	17.09	-37.85
P Copper powder	100+	152+	0.14	0.30	117.41
	150	152-104	5.99	9.98	66.56
	200	104-76	25.38	26.36	3.86
	300	76-53	27.63	29.95	8.43
	350	53-44	7.79	8.41	7.96
	350-	44-	33.07	24.99	-24.44

Table 1. (*continued*)

L Iron powder	100+	152+	6.88	3.15	-54.26
	150	152-104	18.74	17.42	-7.06
	200	104-76	25.63	24.38	-4.88
	300	76-53	23.12	22.40	-3.11
	350	53-44	9.30	7.72	-16.96
	350-	44-	16.33	24.94	52.69
L Copper powder	100+	152+	0.15	0.00	-100.00
	150	152-104	8.37	8.47	1.23
	200	104-76	27.84	29.67	6.58
	300	76-53	23.99	24.28	1.23
	350	53-44	8.62	7.03	-18.42
	350-	44-	31.03	30.54	-1.58

Table 2. Flow and apparent density of metallic powders treated by thought intervention in physical presence [P] as well as by long distance [L]

Powder	Flow control g/s	Flow treated g/s	Percent change	Apparent density control g/cc	Apparent density treated g/cc	Percent change
P Iron powder	8.33	6.25	-24.97	3.19	3.31	3.76
P Copper powder	7.14	6.25	-12.47	3.27	3.26	-0.31
L Iron powder	11.02	10.69	-2.99	3.15	3.2	1.59
L Copper powder	9.43	9.48	0.53	3.22	3.28	1.9

The bulk properties such as apparent density and flow of the three types of powders are given in table 2. The flows of iron and copper P powders have decreased by 24.97 and 12.47% respectively which correlate with decrease in percent of coarse

sieve fraction. The corresponding values for L powders are respectively 2.99% and an increase by 0.53%. Apparent density has increased by 3.76% for P iron powder, while P copper powders showed a negligible decrease of 0.31%. The corresponding values for L powders are respectively an increase by 1.59 and 1.9%.

Table 3 shows a comparison of characteristics for both powders. The data for iron powder shows that the long distance thought intervention seems to be more effective in reducing almost all the coarse sieve fractions thus contributing to an increase in finest fraction by 52.69%. Physical presence seems to be more effective in reducing the fine fractions as well as reducing the flow.

Table 3. Comparison of powder characteristics between treated powders P (thought intervention by physical presence) and L (powders treated by long distance thought intervention)

Powder	Range in micrometers	difference in weight % 100*(wt-wc)/wc		Change in flow %		Change in apparent density %	
		P	L	P	L	P	L
Iron powder Hoganas PASI60	152+	-10.71	-54.26	24.97	2.99	3.76	1.59
	152-104	51.82	-7.06				
	104-76	23.72	-4.88				
	76-53	-1.63	-3.11				
	53-44	-30.70	-16.96				
	44-	-37.85	52.69				
Copper powder PP Patel	152+	117.41	-100.00	12.47	0.53	0.31	1.9
	152-104	66.56	1.23				
	104-76	3.86	6.58				
	76-53	8.43	1.23				
	53-44	7.96	-18.42				
	44-	-24.44	-1.58				

The data for copper powder shows that the long distance thought intervention seems to be more effective in reducing almost all the sieve fractions except in the middle ranges, as opposed to thought intervention in physical presence which has increased the sieve fractions. Physical presence seems to be more effective in decreasing the flow.

4 Discussion

The results of the present experiments confirm our hypothesis presented in an earlier paper on transition metal powders where it was reported that d_{50} and d_{99} particle sizes as well as crystallite sizes showed significant increases and decrease [3] indicating that the thought intervention had caused deformation and fracture as if the powders have been subjected to high energy milling. The reason for this is likely to be the change in lattice parameters of the unit cell which in turn changed the crystallite size and density [3]. Thus, the computed weight and effective nuclear charge of the atom varied leading to the speculation that the thought intervention acts on the nucleus through some reversible weak interaction of larger cross section causing changes in the proton to neutron ratios. Hence it is reasonable to suppose that the effect would have been felt by all the atoms, and hence the unit cell and single crystal grain. Therefore the stresses generated in turn can cause deformation or fracture of the weak interfaces in crystallites as well as in particles. How the processes that are known to occur under external pressure and heat readily take place with a mere thought is a challenge for the materials scientists. Perhaps we are not able to perceive some unknown subtle energy which can be mobilised with least expense. Mr. Trivedi is currently in USA and is willing to help in keen researchers by conducting experiments through intervention in their presence.

5 Conclusions

1. Mr. Trivedi's thought intervention through biofield both in physical presence as well as by long distance has substantially changed the sieve fractions of iron and copper powders. The bulk properties are substantially changed by thought intervention in physical presence as compared to long distance thought intervention, whereas the weight of individual sieve fractions changed more by long distance thought intervention.

2. The results showed increase as well as decrease in weights of various sieve fractions indicating increase in size due to possible plastic deformation, welding, agglomeration, sintering etc. The decrease in size is due to fracture probably at crystallite, grain and agglomerate boundaries.

3. The results obtained are in agreement with the changes observed in particle size and size distribution by laser particle size analyser and BET surface area. At the atomistic level the mass and charge of the atoms are changing which are reflected in the changes in size of unit cell, crystallite and particle. How this can happen with just by thought intervention cannot be understood with our current scientific knowledge.

References

1. Rubik, B., Becker, R.O., Flower, R.G., Hazlewood, C.F., Liboff, A.R., Walleczek, J.: Bioelectromagnetics applications in medicine. In: Alternative Medicine: Expanding Medical Horizons, pp. 45–65. U.S. Government Printing Office, Washington, DC, NIH Publication No. 94-066 (1994b)

2. Information on,
http://www.trivedifoundation.us/default2.aspx?cid=41
3. Trivedi, M.K., Tallapragada, R.M.: Metal Powder Report 63(9), 22–28, 31 (2008)
4. Dabhade, V.V., Trivedi, M.K., Tallapragada, R.M.: Bulletin of Materials Science 32(5), 471–479 (2009)
5. Trivedi, M.K., Tallapragada, R.M.: Materials Research Innovations 13(4), 473–480 (2009)

Analysis of Influencing Factor on Fracture Energy of Concrete Containers for Nuclear Waste

Li Yi, Zhao Wen, and Qujie

Northeastern University
College of Res. and Civ.Eng.
ShenYang, China 110004
liyi@mail.neu.edu.cn, qujie_hu@126.com

Abstract. The anti-fracture property of concrete container for nuclear waste was investigated to ensure its long-time durability. According to RILEM Standard the three-point-bending test was carried out by using the hybrid fiber reinforced concrete(HFRC) and the plain high-strength concrete (PHSC) (SP-13) for comparison at normal temperature and 150℃respectively. Results show that the fracture energy of HFRC is 6.3～12.0 times higher than PHSC before heating and is 12.0～16.7 times higher than PHSC after the heating .The big-size steel fibers(Types A and C) are superior to both the small-size fibers (Types B) and polypropylene fibers in absorbing fracture energy, and the Types C steel fibers(shear-indented type) is superior to Type A(Shear-threaded type)and Type B(ultra-short/fine type)in improving the fracture energy. Based on the comprehensive flexure toughness index and the comparative analyses of range and variance, a conclusion is drawn that the Types A and C are applicable to the materials used to make concrete containers for nuclear waste.

Keywords: fiber reinforced concrete,containers for nuclear waste, fracture energy, temperature.

1 Introduction

Concrete containers of nuclear waste have been widely used around the world because of their easy removal, radiation protection and low-cost. The material used to contain nuclear waste in Chinese is specified to ordinary concrete by the specification “Concrete vessel using to store the medium-low radioactive solid waste "(EJ914-2002). The disintegration and heat releasing of nuclear waste can cause the deterioration of material performance and inhomogeneous deformation of the stored vessel thus leading to metamorphism of concrete and waste material leaking. and with the growth of time concrete is easy to be metamorphic, vessels damage and radioactive materials leak. Mixing hybrid fiber into concrete can prevent initiation of internal cracks in concrete at different levels, and effectively improve the ductility and toughness of concrete.

Fracture energy is defined as the energy consumption by forming unit-area of fractures, which stands for the capacity of materials to prevent the expansion of the cracks. At present, the study about fracture energy of the ordinary concrete at normal

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 253–259.
springerlink.com

temperature is relatively mature [1, 2], but is rare for the hybrid fiber reinforced concrete subjected to medium or high-temperature. In this paper, the anti-fracture property of concrete container for nuclear waste was investigated to ensure its long-time durability. According to RILEM Standard, the three-point-bending test was carried out comparatively for the hybrid fiber reinforced concrete and the plain high-strength concrete (SP-13) at normal temperature and at 150℃ (The temperature of the nuclear waste store is usually controlled below 150℃). Based on the comprehensive flexure toughness index and the comparative analyses of range and variance, the effect of ingredients on fracture energy of hybrid fiber reinforced concrete is given.

2 Test Design and Raw Material

The steel fiber in the test is produced by *changhong* steel fiber company in Anshan. Steel fiber A is sheared thread with a length of 32mm. The length/diameter ratio is 31and tensile strength is 600Mpa ; steel fiber B is super short and fine with high toughness and is 13mm long with length/diameter ratio of 64. The tensile strength is 2000Mpa ; steel fiber C is shear indentation with a length of 34mm and length/diameter ratio of 32. The tensile strength is 1200MPa ; polypropylene fiber provided is by *Lierdetong* new type material scientific limited company in Changzhou with a length of 19mm and density of 0.91 g/cm^3. the melting point is 165～175℃; grade-Ⅰ fly ash is provided by thermoelectricity field in Shenyang with a density of 2059kg／m^3 ; slag powder is provided by steel group slag development company in Anshan with a density of 2910kg／m^3 and specific surface area of 430m^2／Kg. The density of ordinary Portland cement 42.5 with a brand of *Huari* is 3000kg/m^3 ; the density of medium sand is 2630kg/m^3 and the fineness modulus is 2.6 ; the density of gravel is 2930kg/m^3 and maximum grain size is 31.5mm ; powdery defoamer P803 and *Huawang* naphthalene series superplasticier are provided by *Shantelonghu* scientific industry company. In the 13 Groups of samples,SP-1～SP-8 is the high performance concrete incorporated steel fiber A ,steel fiber B and polypropylene hybrid fiber and fly ash. and the mixed proportion is according to $L_8(2^7)$ orthogonal design. SP-9～SP-11 is the high performance concrete incorporated steel fiber A ,steel fiber B and fly ash; SP-12 is fiber reinforced concrete incorporated steel fiber C; SP-13 is the high strength concrete E3 in literature [3]used as storage of nuclear waste (28-day compressive strength is 80 MPa), which has the best impermeability. The reference strength of SP-1～SP-12 is C50 and the mixed proportion is shown in Table 1.

The process for mixing concrete adopts dry mixing method. Each group has 3 specimens with a size of 150×150×550mm for the bending test. A steel plate of 2mm thickness is inserted at the middle of the span in the pouring surface of the beam before vibrating, and the insert depth is 25mm. The plate is pulled out after initial setting. The test at normal temperature is carried out after specimen has been fostered according to GBJ81—85. The specimen was then put in electric blasting drying oven and heated to 150℃ for 24 hours and then was cooled to normal temperature.

Table 1. The mixed proportion designed by orthogonal for experiment

number of group	water ／ Kg	cement ／Kg	fly ash/Kg	slag powder ／Kg	sand ／Kg	gravel ／Kg	steel fiber A /Kg	steel fiber B /Kg	steel fiber C /Kg	PP fiber /Kg	water reducer %
SP-1	180	450	50	0	593	1202	40	8	0	0.3	0.6
SP-2	180	350	150	0	578	1174	80	8	0	0.3	0.6
SP-3	180	350	150	0	578	1174	40	24	0	0.3	0.6
SP-4	180	450	50	0	593	1202	80	24	0	0.3	0.6
SP-5	180	350	150	0	578	1174	40	8	0	0.9	0.6
SP-6	180	450	50	0	593	1202	80	8	0	0.9	0.6
SP-7	180	450	50	0	593	1202	40	24	0	0.9	0.6
SP-8	180	350	150	0	578	1174	80	24	0	0.9	0.6
SP-9	180	400	100	0	585	1188	100	26.8	0	0	0.6
SP-10	180	400	100	0	585	1188	120	29.6	0	0	0.6
SP-11	180	400	100	0	585	1188	140	32.4	0	0	0.6
SP-12	180	500	0	0	599	1216	0	0	120	0	0.6
SP-13	158	428.7	0	183.7	618.5	1145.1	0	0	0	0	2

3 Testprocess and Result

According to RILEM standard, the span of specimen is 500mm. the specimen is loaded by using the WDW-T100 computer-controlled electronic universal testing system at the middle of the span. The displacement is controlled at the rate of 0.2mm/min. Figure 1 shows the loading test and measurement chart; the graph 2 shows load-deflection （$P-\delta$） curve before the deflection of the beam reaching to 3.5mm. The following formula can be used for nuclear waste containers to evaluate the fracture energy of the hybrid fiber reinforced concrete comprehensively:

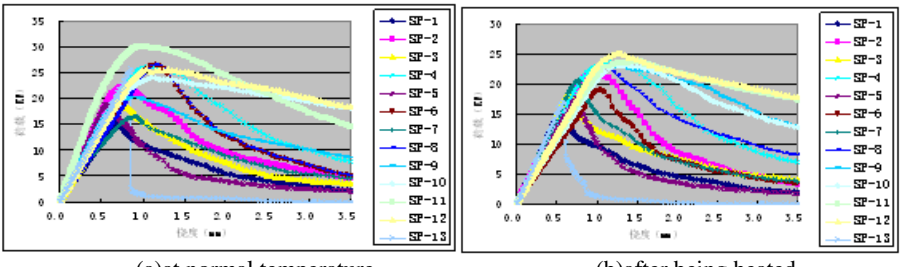
$$f_{eq,2}=\frac{3}{2}\left[\frac{D_{BZ,2}^{f}}{0.50}\right]\frac{L}{bh_{sp}^{2}},$$

$$f_{eq,3}=\frac{3}{2}\left[\frac{D_{BZ,3}^{f}}{2.50}\right]\frac{L}{bh_{sp}^{2}},$$

$$G_F=\left(W_0+mg\delta_0\right)/A_{lig} \tag{1}$$

Where $f_{eq,2}$, $f_{eq,3}$ is the equivalent bending strength when deflection of beam is δ_2, δ_3 respectively; G_F is the fracture energy of concrete; L、b、h_{sp} is the span, cross-section width and effective height of specimen respectively; $\delta_2=\delta_L+0.65$,

$\delta 3 = \delta_L + 2.65$; $D^f_{BZ,2}$、$D^f_{BZ,3}$ is the energy absorbed by the fiber when the deflection of beam is δ_2、δ_3 respectively; $D^f_{BZ,2} = D_{BZ,2} - D^b_{BZ}$, $D^f_{BZ,3} = D_{BZ,3} - D^b_{BZ}$, $D_{BZ,2}$、$D_{BZ,3}$ $D^f_{BZ,3}$ is the common energy absorbed by the fiber and concrete when the deflection of beam is δ_2、δ_3 respectively; $\delta_2 = \delta_L + 0.65$, $\delta 3 = \delta_L + 2.65$; δ_L is the deflection corresponding to the maximum load before the deflection of the beam reach 0.05mm; D^b_{BZ} is the single energy absorbed by the concrete, which is equal to the triangle area surrounding by F_L, $\delta_L + 0.3$ and coordinate origin; W_0 is the area surrounding by load-defection curve and Abscissa; $m = m_1 + 2m_2$, m_1 is the quality of the beam between the two blocks; m_2 is the quality of loading equipment independent to the testing system; g is the acceleration of gravity and is taken as 9.8 m/s^2; δ_0 is the deflection after damage; A_{lig} is rupture area and is taken as the projected area of the sample ruptures area.



(a)at normal temperature (b)after being heated

Fig. 2. The curve of $P - \delta$

4 The Analysis of the Test Result

4.1 The Analysis of Impacting Factors before and after Heated

By gradation analysis:

- Before heated, the effect of the various factors on the concrete fracture energy is in the order of: steel fiber A> steel fiber B > pp fiber > fly ash. Except pp fiber has negative effect, all have positive effect.
- After heated, the effect of the various factors on the concrete fracture energy is in the order of: steel fiber A> steel fiber B > fly ash. > pp fiber. All have positive effect.

By analyzing from F. value and Sig., it can be concluded that:

- Before heated, the effect of the various factors on the concrete fracture energy is: steel fiber A and steel fiber B have highly significant effect; pp fiber and fly ash have no significant influence.
- After heated, the effect of the various factors on the concrete fracture energy is: steel fiber A and steel fiber B have significant effect; pp fiber and fly ash have no significant influence.

Table 2. The calculated test results of fracture energy for concrete under bending

State	Category of specimen	F_L /KN	$D^f_{BZ,2}$ /(N•mm)	$D^f_{BZ,3}$ /(N•mm)	$f_{eq,2}$ / MPa	$f_{eq,3}$ / MPa	fracture energy/N/m
before being heated	SP-1	13.808	5810	16264	3.718	2.082	1162.72
	SP-2	18.993	10461	29703	6.695	3.802	2077.96
	SP-3	16.167	8229	22459	5.267	2.875	1771.16
	SP-4	19.641	13081	43391	8.372	5.554	3309.94
	SP-5	17.100	6660	14893	4.262	1.906	1157.08
	SP-6	23.445	11931	31136	7.636	3.985	2555.84
	SP-7	9.044	7644	25994	4.892	3.327	1892.83
	SP-8	17.833	11654	44051	7.459	5.639	3226.97
	SP-9	14.618	10115	36936	6.474	4.728	3701.31
	SP-10	16.916	11795	54253	7.549	6.944	7042.31
	SP-11	17.335	14230	62379	9.107	7.985	6440.25
	SP-12	22.400	12880	56677	8.243	7.255	7035.03
	SP-13	23.400	-	--	-	-	588.72
after being heated	SP-1	12.400	5438	15181	3.480	1.943	1068.94
	SP-2	16.412	10714	27700	6.857	3.546	1942.48
	SP-3	16.221	6905	21077	4.419	2.698	1689.85
	SP-4	14.081	11252	41399	7.201	5.299	2905.19
	SP-5	15.507	5347	11921	3.422	1.526	1124.94
	SP-6	16.916	7867	19511	5.035	2.497	1626.51
	SP-7	18.543	7616	20955	4.874	2.682	1735.06
	SP-8	20.256	10783	34878	6.901	4.464	3166.51
	SP-9	17.775	11422	47765	7.323	6.114	4558.97
	SP-10	21.412	11514	44292	7.369	5.669	4423.42
	SP-11	16.489	11362	52523	7.272	6.723	5936.22
	SP-12	14.171	11520	54211	7.373	6.939	6171.92
	SP-13	12.455	2574	3128	1.647	0.400	370.44

4.2 The Comparative Analysis for the Fracture Energy before and after Heated

1) According to Fig 2, compared with the high strength concrete of SP-13, the plumpness of the curve $P-\delta$ is significantly improved. area surrounded by curve and abscissa is obvious increased. Fiber significantly increased the bearing capacity and energy absorption capacity of concrete beam. Besides, in Table 3 the gradation of the fracture energy of the reinforced fiber concrete is positive, which shows that fracture energy of the fiber reinforced concrete have significant advantage compared with the high strength concrete without fiber. The fracture energy of fiber reinforced

concrete is 6.3～12.0 times than that of plain high strength concrete (SP-13) before heated and is 12.0～16.7 times after heated.

2) From Table 2, it can be seen that $f_{eq,2}$ is larger than $f_{eq,3}$ for the same sample, i.e., the bearing capacity of beams is decreasing with the increasing of deflection after the ultimate load. Compared with SP-13，indices of SP-9～SP-11, except f_L is slightly low at normal temperature（C50）, are all better than those of SP-13, no matter hybrid steel fiber or sole steel fiber is mixed. The advantage of flexural toughness is more obvious after being heated.

3) The fracture energy of fiber reinforced concrete degrade after heated and the retention ratio is about 90% except for SP-6(the retention ratio of SP-6 is 64%); the fracture energy retention ratio of plain high strength concrete(SP-13) after heated is 63%.This shows that for the storage concrete under bolometric radiation, the capacity of preventing propagation of cracks is reinforced through the bridge effect of fiber mixed at different levels.

4) The effect of fiber type on fracture energy

a) Large size steel fiber (A and C) is better than small size steel fiber (B) and polypropylene fiber in aspect of energy absorption capacity. Among the influencing factors of fracture energy of the steel-polypropylene hybrid fiber concrete, the polypropylene fiber is not as obvious as steel fiber for the effect of transferring load.

b) If the size of steel fiber is near each other, the character of fiber (compressive strength) and the differences of the surface shape are important factors to fracture energy: the C type (shear crimped shape) steel fiber is rarely broken during test because of its high tensile strength. The bite force and friction are larger than A (shear thread shape) and B (ultra-short and ultra-fine shape) type steel fiber because of its concave character. The effect of C type steel fiber is more obvious on improving fracture energy of high performance concrete.

5 Conclusion

The fracture energy of fiber reinforced concrete is 6.3～12.0 times than that of plain high strength concrete(SP-13) before heated and is 12.0～16.7 times after heated .The big-size steel fibers(Type A and C) are superior to both the small-size fibers (Type B) and polypropylene fibers in absorbing fracture energy, and the C Type steel fiber(shear-indented type) is superior to Type A(Shear-threaded type)and Type B(ultra-short/fine type)in improving the fracture energy. Based on the comprehensive comparison and analyze of flexure toughness index, range and variance, a conclusion is drawn that the A type and C type are applicable to be added in concrete used for nuclear waste container.

Acknowledgment. This work presented is jointly supported by the Chinese NNSF (No. 50508008); Fundamental Research Funds for the Central Universities (N090401009); Natural science Foundation of Liao Ning Province (No. 20082024).

References

1. Sivakumar, A., Santhanam, M.: Mechanical properties of high strength concrete reinforced with metallic and non-metallic fibres. Cement and Concrete Composites, 587–646 (2007)
2. Wang, T.: Experiment on the effects of steel fibers on the flexural toughness and fracture energy of high-performance concrete. Industrial Constructi, 65–68 (2007)
3. Bao, G.-Q., Ye, Q., Wang, J.-D.: Research on high performances concrete applied to nuclear waste container. Concrete, 49–52 (2003)

A Line Segment Based Inshore Ship Detection Method

Jiale Lin[1], Xubo Yang[1], Shuangjiu Xiao[1], Yindong Yu[1], and Chengli Jia[2]

[1] Department of Software Engineering
Shanghai Jiao Tong University, Shanghai, China
[2] Institute of Remote Sensing, Beijing, China
{carew4134,yangxubo,xiaosj,Kyohirooo}@sjtu.edu.cn,
jclnudt@163.com

Abstract. On the problem of ship detection in optical remote sensing images, it is difficult to separate the ships from the harbor background to detect the inshore ships. This paper shows a new method based on the line segment feature, which can detect the shape of the ship by the line segment feature and remove the false alarm by the bounding box of the ship. The experimental results present that the method is effective to detect the inshore ships in linear time.

Keywords: inshore ship detection, optical remote sensing image, LSD, line segment detection, ship detection model.

1 Introduction

With the development of remote sensing technology, the remote sensing images have high resolution under meter scale. Nowadays, the optical remote sensing images are widely used. Obviously, there is a need for automatic detection of targets of interest. However, in the literature of the ships detection, the researches almost focused on the detection of ships in sea while few methods about inshore ships detection can be found. So it would be valuable to present an efficient and effective method to detect the inshore ships.

Among the researches on inshore ships detection, most of them [1, 2] followed the matching approach, which, however, needs to establish a large template database and for each candidate object in the image, make a match with every templates in database. It would be computational expensive so that it can't be applied in the surveillance system. Though the AIAC matching approach in [2] is quite fast compared with other matching approaches, it would be difficult to extract the continuous edge information in the remote sensing image. Furthermore, in [3], it is effective to deal with the problem of rotation, scale invariant and the difference of the ships by detecting the heads of the ships. But the method in [3] is only able to detect the ships with the "V" shape heads. It would fail even lack of a little part of the "V" shape head.

In this paper, we present a state-of-the-art method to detect the inshore ships. At first, we extract the line segment features in the boundary between the sea and the harbor. Then, we can use the line segments to detect the heads of the ships as the

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 261–269.
springerlink.com

feature and to build the line table which can provide with the spatial information of the line segments. Furthermore, we find the main axis near the ship head as the orientation of the ship so as to deal with the rotation invariant problem. Finally, we validate the ship by checking line segments in the bounding box of the ship.

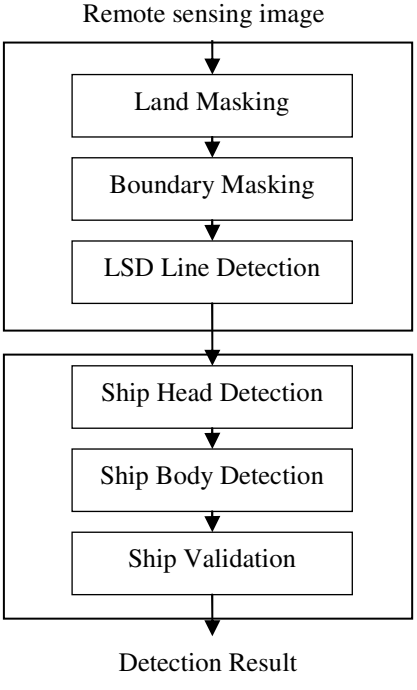


Fig. 1. The framework of the ship detection algorithm

The rest of this paper is structured as follows: in Section 2, we present the whole process of the proposed algorithm. The LSD line segment detector and the line table are reported in Section 3. In Section 4, the ship detection model is presented. We discuss the complexity of the algorithm and experimental results in Section 5.

2 The Algorithm Procedure

Fig.1 shows the whole framework of the ship detection algorithm, which can be divided into two parts. In the first part, we mask the pixels as the land whose gray-levels are higher than the threshold which is the lowest value between the two peaks of the histogram and define the soft boundary between the land and the sea. The detail can be reviewed in [3]. Furthermore, we extract the line segment features in the soft boundary. In the other part, we propose a ship detection model, in which we detect the ship head as the main feature, search the lines near the ship head for the ship body, and finally, validate the ship and output the result.

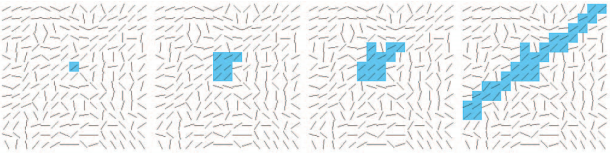


Fig. 2. The process of the region growth algorithm for the aligned points[6]. The support region is in blue and the segments present the gradient orientation of the pixels. The left is the feed point which can form the final support region on the right.

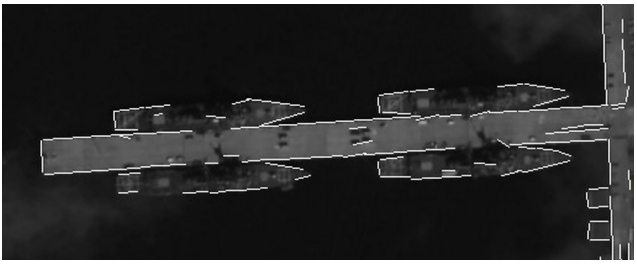


Fig. 3. The result of line detection

3 Line Segment Detection

Line segments provide with rich geometric information of the image. Therefore, the line segment detection is one of the most popular researches in computer vision. Classic approaches apply the Canny edge detector [4] followed by a Hough Transform [5], but the result isn't quite well both in the computational time and accuracy. However, the recent researches made huge advance. In [6], a line segment detector is reported, in which detects the line segment by the region growth method using the gradient direction with false alarm control. The LSD detector can detect the line segment accurately in linear time. So we try to apply the LSD line segment detector in ship detection in optical remote sensing image.

The whole process of the LSD line segment detection method is listed as follows:

- Sort the pixels by the descending order of gradient magnitudes.
- Following the order, choose the pixel never used and start the region growth method to acquire the support region (See Fig.2). During the process of region growth, the orientation difference of the candidate pixel and the support region must be in a small certain range.
- Associate the line-support region (a set of pixels) with a line segment (actually, a rectangle). A line segment is determined by its endpoints and its width.
- Validate the line segment with the false alarm control. The false alarm is the pixel whose gradient orientation is quite different from the orientation of the line-support region. If the NFA is larger than the threshold, reject the line segment.

When the line segments are detected in the soft boundary, we propose to utilize a data structure called line table (See Fig.3), which divide the image into several grids in w

width and h height to provide the spatial information of the line segment. With the line table, each line segment can search its neighbor line segments efficiently. We record the ID of line segment to the grid through which the line segment passes.

The line segment detection result in optical remote sensing image is showed in Fig.4. From the result, we can see that LSD line segment detector is so accurate. However, it may miss some line segments which are some important for ship detection. So we need to have an effective ship detection model in this condition.

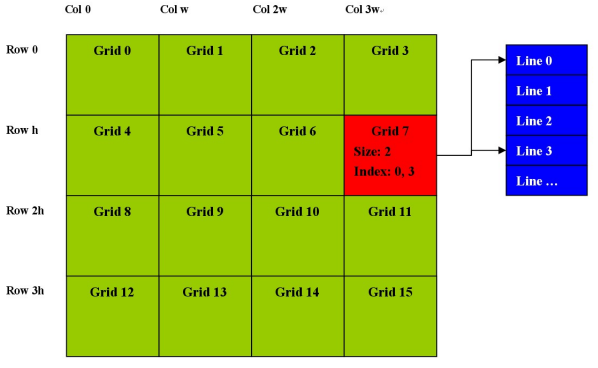


Fig. 4. The line table. The red part marks that Line 0 and Line 3 go through Grid 7.

4 Ship Detection Model

After extracted the line segments in the soft boundary with LSD line segment detector and built the line table, we propose a ship detection model. Because the background of the harbor is complicated so that the ship and the background can't be easily separated, the line segment is always interceptive (See Fig.4). Besides, the ship detection is quite different from the pedestrian and car detection that the orientations of the ships in the image aren't the same. So we should deal with the rotation invariant problem in the ship detection other than the scale invariant problem. Like the method in [3], we firstly detect the head as the main feature. However, besides the ships with the "V" shape heads, we can detect more kinds of ships than that in [3] with the line segment information. Furthermore, the ships may park side by side. So the priori knowledge of park position in [3] would fail in this situation. Therefore, quite different from [3], we detect the ship body so as to detect the whole ship after detected the head with the line segments information.

4.1 The Ship Head Detection

The ship head detection can be described as the following steps:

- For each line segment, find the nearest line segments of the two endpoints and record their IDs when the distance is less than a small threshold.
- For each line segment, separately check the two nearest line segments of the two endpoints. When the angle between the nearest line segment and the

current line segment is in a range from 12 to 75 degree, we consider the two line segments as a "V" shape head (See Fig.5). In Fig.5, the line segment between the two line segments presented the direction of the ship head.

- For each line segment, if the current line segment can't make up the head and the two endpoints both have their nearest line segments, check the two nearest line segments. When the angle between the two nearest line segments is in a range from 12 to 75 degree, we consider the three line segments as another kind of head (See Fig.6). In Fig.6, the middle line segment is the current line segment while the line segments in two sides are two nearest line segments of two endpoints.

After the above three steps, we can detect the candidate ship heads with the IDs of the line segments. Furthermore, we can calculate the average direction of the two line segments made up the heads as the orientation of the head.

Fig. 5. One result of ship head detection

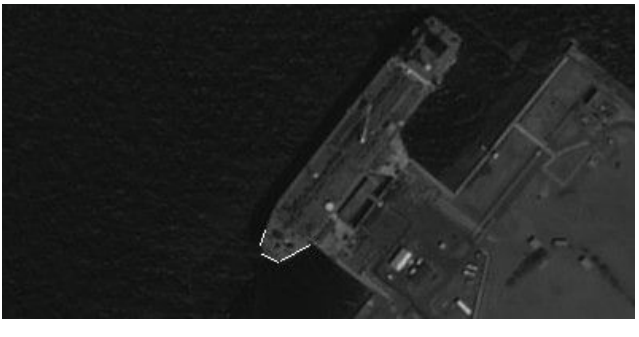


Fig. 6. One result of ship head detection

4.2 The Ship Body Detection

For each candidate ship head, there are two endpoints near the ship body. Therefore, for each endpoint, we search the line segments in the grid of the line table belonging to the endpoint. For each endpoint P, the candidate line segments satisfying the following conditions are considered as the ship body:

- The candidate line segment is parallel with the ship head.
- The candidate line segment is the lateral part of the ship head. The intersection point of the candidate line segment and the line segment of the endpoint P should be outside of the line segment of the endpoint P or be the endpoint P.

If there is at least one side of the ship body detected, we consider that the ship body is successfully detected.

4.3 The Ship Validation

In the optical remote sensing image, due to the complicated harbor background and the drawback of the line detection method, the line segments information may be missed so that parts of the ship may lack the geometric information. Therefore, we validate the ship in the bounding box so as to remove the false alarm.

After detected the ship head and the ship body, we can easily calculate the width of the ship and according to the statistics, the length of the ship is supposed to be at most 12 times of the width. So started from the ship head, along to the ship body, we can get the bounding box of the ship.

For each line segment in the bounding box, we consider the farthest line segment from the ship head as the ship tail, which can be found by following steps:

- Calculate the direction of the bounding box and mark the angle by α .
- With the following Hough Transform equation:

$$\gamma(\theta) = x_0 \cdot \cos\theta + y_0 \cdot \sin\theta \tag{1}$$

XY coordinate can be transformed to $\gamma(\alpha)\ \gamma(\beta)$ coordinate.

For each short edge of the bounding box, substitute one point on the edge into the following equation:

$$\gamma(\alpha) = x_0 \cdot \cos\alpha + y_0 \cdot \sin\alpha \tag{2}$$

For each long edge of the bounding box, substitute one point on the edge into the following equation:

$$\gamma(\beta) = x_0 \cdot \cos\beta + y_0 \cdot \sin\beta \tag{3}$$

(α and β are perpendicular).

Finally, we get the new bounding box in the new coordinate: $\gamma(\alpha)_0$ 、 $\gamma(\beta)_0$, $\gamma(\alpha)_1$ 、 $\gamma(\beta)_1$ 。

- For each line segment in the grids of the line table which the bounding box covers, substitute the two endpoints into (2) and (3). If both of the two endpoints are inside the new bounding box, record the distance from the current line segment to the ship head.
- If the farthest distance isn't long enough to make up a ship, we reject the candidate ship. According to statistics, the length of the ship is at least 4 times of the width.

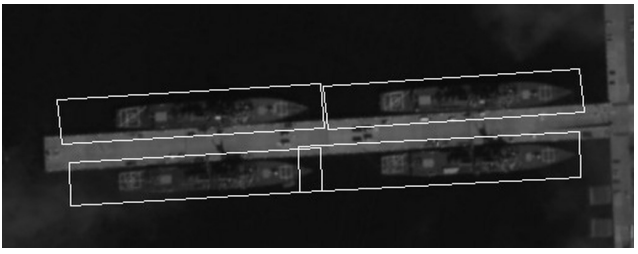


Fig. 7. The inshore ship detection result

One detection result is demonstrated in Fig. 7.

5 Experimental Evaluation

5.1 Complexity of the Algorithm

At the stage of land masking, the classic method is used that the pixels are marked as the land whose gray-levels are higher than the threshold which is the lowest value between the two peaks of the histogram. The complexity of land masking process is O(n).

To detect the line segment in the soft boundary with LSD line segment detector, which traverses the pixels in constable times to grow the line-support region, it only needs O(n) time complexity.

In the ship detection model, the main operation is to find the neighbor line segments for each line segment. With the line table, the line segment only needs to traverse the line segments in the grid of the line table in which it lies. For each grid with w width and h height, there are at most w x h line segments (each pixel is a line segment). However, there are only a small number of line segments in each grid. Therefore, it needs O(L) time complexity where L is far less than n.

In a conclusion, the whole algorithm of ship detection is O(n) which can detect the ships in harbor in linear time.

5.2 Experimental Results

The proposed algorithm for inshore ship detection is proved to be quite effective and efficient after tested a set of remote sensing image. The experiment is in an Intel Pentium D 2.8GHz、1G DDR2 PC with a set of images of size 612 x 612, which are quite different from the background and the shapes of the ships. The results are listed in Table 1.

From the table, the results are quite satisfactory both in the runtime and the accuracy. From the accuracy, the algorithm can detect all the ships except in PIC 5 and PIC 7, including the ship in sea in PIC 6. In PIC 5, both sides of the body of the missed ship lack line segments information when detecting the ship body. And the two missed ships in PIC 7 are covered by a piece of cloud while the head of the third missed ship is messed with the background so that the LSD line detector misses a line

segment of its head. The false alarm is detected because its shape is like the shape of ship. Finally, it is remarkable that the algorithm can detect the inshore ships in the test images in 300 ms on average.

Table 1. Experimental results

PIC ID	Corre	Miss	False	Time(ms)
1	1	0	0	297
2	7	0	0	328
3	7	0	1	516
4	3	0	0	296
5	5	1	0	297
6	1	0	0	188
7	3	3	0	343

6 Conclusion

In this paper, we utilize the LSD line segment detector to extract the line segments in the soft boundary between the sea and the land and propose a ship detection model to deal with the inshore ship detection problem which few researches focus on. The experimental results show the algorithm is satisfactory in inshore ship detection. Our algorithm can solve the tough problems in inshore ship detection, including the detection of different shapes of ship and the detection of ships which park side by side. However, experimental results also show that the story is not complete, just starting. The line segment detector may become the most important part to be improved in the future because the line segment detected by LSD detector would be cut off. For the ship detection model, it is easy to extend to combine with other methods like HOG [7] and SVM to remove the false alarm. Finally, the occlusion, like the cloud occlusion, should be considered in the future algorithm.

Acknowledgment. The authors are indebted to their collaborators for many remarks and corrections, and more particularly to Feng Chen, Liang Ding and Tianli Bi. The research was partially financed by a grant from the 863 Program of China.

References

1. Lei, L., Su, Y.: An inshore ship detection method based on contour matching. Remote Sensing Technology and Application 22, 622–627 (2007)
2. Jiang, L.-B., Wang, Z., Hu, W.-D.: An AIAC-based inshore ship target detection approach. Remote Sensing Technology and Application 22, 88–94 (2007)
3. Hu, J.-H., Xu, S.-S., Chen, H.-L., Zhang, Z.: Detection of ships in harbor in remote sensing image based on local self-similarity. Journal of Image and Graphics 14, 591–597 (2009)
4. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Analysis and Machine Intelligence 8, 679–698 (1986)

5. Ballard, D.H.: Generalizing the Hough Transform to detect arbitrary shapes. Pattern Recognition 13, 111–122 (1981)
6. von Gioi, R.G., Jakubowicz, J., Morel, J.M., Randall, G.: LSD: A fast line segment detector with a false detection control. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 722–732 (2010)
7. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 886–893 (2005)

A Novel Fault Diagnosis Method for the Plant with Min-Disturbance*

Yongqi Chen and Xiangsheng Yang

College of Science and Technology, Ningbo University, Ningbo, China
lingfen7781@163.com, yangxiangsheng@nbu.edu.cn

Abstract. For estimating sensor faults for the plant with min-disturbance, a novel fault diagnosis method based on support vector interval regression is proposed. This method can reduce the influence of min-disturbance. When faults occur in this plant, the interval of regression model can not contain outputs of the plant and the faults can be detected. Experiments are given to demonstrate the efficiency.

Keywords: Fault diagnosis, support vector interval regression model, min-disturbance.

1 Introduction

When unexpected faults occur in complex industrial systems, they will lead to serious damage, and even be hazardous to the systems and environment. During the last few decades, Lots of efforts have been invested in model based fault detection and isolation (FDI). Some important papers of FDI are given[1-8]. For this years, support vector machines (SVM), with high generalization performance, has been developed very quickly. Many fault diagnosis methods based SVM are proposed and experiments are used to demonstrate the effectiveness of SVM[9-10]. But these fault diagnosis methods are valid only the data samples of plant are crisp input-crisp output. But, Available information is often uncertain and imprecise when there is min-disturbance in the plant. If classic SVM is used to detect fault of the plant with min-disturbance, wrong alarm can often occur. For this reason, it is very necessarily to develop a new robust fault diagnosis method for a class plant with min-disturbance. In this note, a robust fault diagnosis method based support vector interval regression model is constructed because interval regression model is suitable to estimate imprecise data. Firstly, the regression model of the plant without fault is constructed by support vector interval regression. Then, this regression model is used to estimate the plant. when there are no faults, outputs of the plant with min-disturbance locate between upper side and lower side of the regression model. But when faults occur in

* This work was supported by natural science foundation of Ningbo(No. 2009A610074) A Project Supported by Scientific Research Fund of Zhejiang Provincial Education Department (Y200803444); Advance Research Fund of College of Science and Technology (003-21020901); Education Research Fund of Ningbo University (007-200903).

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 271–276.
springerlink.com

this plant, the interval of regression model can not contain outputs of the plant and the faults can be detected.

This paper is organized as follows. Section 2 introduces support vector interval regression model. In Section 3, the robust fault diagnosis method based support vector interval regression model is applied in the plant with min-disturbance Finally, Section 4 presents the concluding remark. s

2 Interval Regression Analysis

In this section, support vector interval regression models are briefly introduced. The interval regression by quadratic programming is firstly proposed by Tanaka and Lee [11]. Based on the interval regression method of Tanaka and Lee[11], Hwang[12] presented support vector interval regression model(SVIRM) for crisp input and output data in 2006. SVIRM uses the high generalization of standard support vector regression approach to yield the robust estimate output. The prime optimization problem of SVIRM is as follows:

$$\min_{a,c} \frac{1}{2}(\|a\|^2 + \|c\|^2) + C\left(\sum_{i=1}^{n}\xi_{1i} + \sum_{i=1}^{n}(\xi_{2i} + \xi_{2i}^*)\right)$$

$$\text{subject to}\begin{cases} \langle c\cdot\phi(|x_i|)\rangle \le \xi_{1i} \\ y_i - \langle a\cdot\phi(x_i)\rangle \le \xi_{2i} + \varepsilon \\ \langle a\cdot\phi(x_i)\rangle - y_i \le \xi_{2i}^* + \varepsilon \\ \langle a\cdot\phi(x_i)\rangle + \langle c\cdot\phi(|x_i|)\rangle \ge y_i \\ \langle a\cdot\phi(x_i)\rangle - \langle c\cdot\phi(|x_i|)\rangle \le y_i, i = 1\cdots n \end{cases} \tag{1}$$

The following quadratic programming (QP) problem is deduced by Lagrangian theorem and kernels for dot products: where $a_{1i}, a_{2i}, a_{2i}^*, a_{3i}, a_{3i}^*$ are the nonnegative Lagrange multipliers. By solving the above dual QP problem, the Lagrange multipliers a_{1i}, a_{2i}, a_{2i}^*, a_{3i}, a_{3i}^* can be obtained. Therefore, the weight vectors a and c can be showed as:

$$\begin{aligned} a &= \sum_{i=1}^{N}(a_{2i} - a_{2i}^*)\phi(x_i) + \sum_{i=1}^{n}(a_{3i} - a_{3i}^*)\phi(x_i) \\ c &= \sum_{i=1}^{N}(a_{3i} + a_{3i}^*)\phi(|x_i|) - \sum_{i=1}^{n}a_{1i}\phi(|x_i|) \end{aligned} \tag{2}$$

The upper and lower bounds of the interval regression model constructed by SVIRM are

$$\begin{aligned} F\_up(x) &= \langle a \cdot \phi(x) \rangle + \langle c \cdot \phi(|x|) \rangle \\ F\_down(x) &= \langle a \cdot \phi(x) \rangle - \langle c \cdot \phi(|x|) \rangle \end{aligned} \tag{3}$$

The functional form of $\phi(x_i)$ need not to be known since it is defined by the kernel function $k(x_i, x_j) = \phi(x_i)\phi(x_j),\ i = 1 \cdots n,\ j = 1 \cdots n$ [9-11]. Different kernel functions present different mappings from the input space to the high dimension feature space. The well used kernels for regression problem are given as follows:

linear kernel: $k(x_i, x_j) = (x_i, x_j)$

polynomial kernel of degree: $k(x_i, x_j) = [(x_i, x_j) + 1]^d$

RBF kernel: $k(x_i, x_j) = \exp(-\dfrac{\left\| x_i - x_j \right\|^2}{2\sigma^2})$

3 Experiments

In this section, the fault diagnosis method based SVIR is applied in sensor faults detection for the plant with min-disturbance. Sensor is important for the plant to achieve its optimal performance. All sensor faults must be detected accurately rapidly to prevent serious accidents. Consider the plant:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + N_1 d(t) + N_2 f(t) \end{aligned} \tag{4}$$

where $x \in R^n$ are the states of the plant, $u \in R^m$ are the control inputs, $y \in R^p$ are the measurable outputs of the plant, $d \in R^l$ are unknown min-disturbance, $f \in R^l$ are sensor faults. f and d are the uncoupled forms. Let:

$$\begin{aligned} A &= \begin{bmatrix} 0 & 1 \\ -2 & -2 \end{bmatrix} \qquad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ C &= \begin{bmatrix} 1 & 0 \\ 0.5 & 1 \end{bmatrix} \\ N_1 &= \begin{bmatrix} 1 \\ 1 \end{bmatrix}, N_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{aligned} \tag{5}$$

The unknown min-disturbance is assumed as following:

$$d(t) = 0.1 - 0.2 * 1 * \mathrm{rand}( ) \tag{7}$$

It denotes noises generated in the interval [-0.1, 0.1] at random. The sensor fault is given as:

$$f(t) = \begin{cases} 0.3\sin(t) + 0.1 & 4s \le t \le 8s \\ 0 & t < 4s, t > 8s \end{cases} \tag{7}$$

To detect the sensor faults, one must obtain the regression model of the plant without faults. By comparing the estimated outputs of the regression model and the observed outputs of the plant, the sensor faults can be detected. Support vector regression (SVR), which owns high generalization performance, is an effective method to construct regression model. But the regression model based on traditional SVR only presents crisp outputs and can not describe the effect of min-disturbance in plant. For this reason, traditional SVR is not fit to detect the sensor faults of the plant with min-disturbance. It is likely to regard the disturbance as sensor faults and give a wrong alarm. In this section, SVIR is used to construct the interval regression model for the plant with min-disturbance. When there are no sensor faults, outputs of the plant with min-disturbance will locate in the interval of regression model. But if sensor faults occur, outputs of the plant will be beyond the interval of regression model.

In order to construct SVIR fault diagnosis model, one must obtain the training samples of the plant without sensor faults. the training samples for the plant can be described as follows:

$$X = \begin{bmatrix} y_1 & \cdots & y_m & u_1 & \cdots & u_n \\ y_2 & \cdots & y_{m+1} & u_2 & \cdots & u_{n+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{p-m} & \cdots & y_{p-1} & u_{p-n} & \cdots & u_{p-1} \end{bmatrix}, S = \begin{bmatrix} y_{m+1} \\ y_{m+2} \\ \vdots \\ y_p \end{bmatrix} \qquad \mathrm{y_i} = \mathrm{y(t_i)},$$

$$\mathrm{u_i} = \mathrm{u(t_i)}, \mathrm{t_i} = \mathrm{t_{i-1}} + \Delta t, i = 1,2 \cdots m \tag{8}$$

Δt is the sample time. m is output parameter's delay rank, n is input parameter's delay rank. In this experiment, $\Delta t = 0.2s$, m=4, n=4. Simulation time is chosen as 10 second. Fig. 1 shows interval regression estimates and output of plant without sensor faults. As shown in Fig. 1-2, although there is disturbance in the plant, the outputs of plant always locate in the interval of regression model. SVIR performs smoothly and gives no wrong alarm.

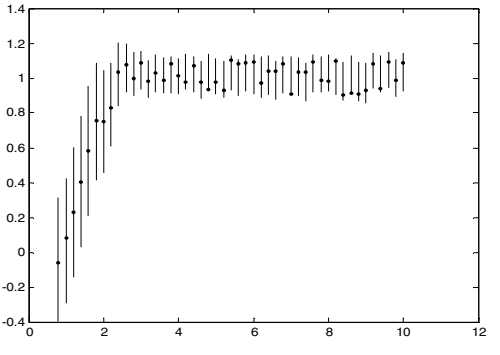


Fig. 1. Fault diagnosis method based SVIR estimates of interval bounds and the first output parameter of the plant with disturbance

When sensor faults occur, output of the plant will beyond the interval. Fig. 2 shows interval regression estimates and output of plant with sensor faults. As shown in Fig. 3-4, because there are sensor faults in the plant between 4s and 8s, the outputs of plant also is beyond the interval between 4s and 8s. SVIR is successful in detecting the sensor faults as early as possible.

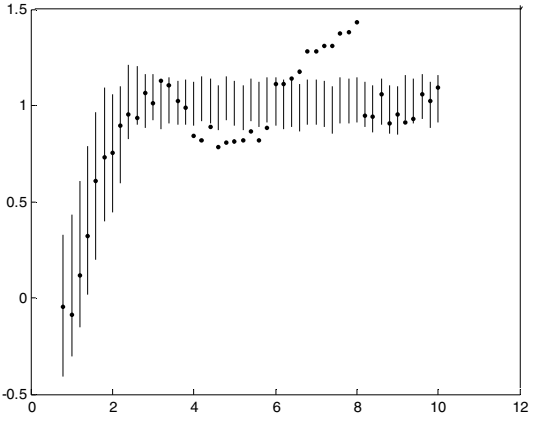


Fig. 2. Fault diagnosis method based SVIR estimates of interval bounds and the first output parameter of the plant with sensor faults and disturbance

4 Conclusions

In this note, a robust fault diagnosis method based support vector interval regression model is proposed to estimate the fault signals of the system with min-disturbance. Simulation results are given to demonstrate the efficiency of the proposed observer.

References

1. Isermann, R.: Process fault detection based on modeling and estimation methods: a survey. Automatic 20(2), 387–404 (1984)
2. Wang, H., Daley, S.: Actuator fault disgnosis: An adaptive observer based technique. IEEE Trans. Autom. Control 41(2), 1073–1078 (1996)
3. Tyler, M.L., Asano, K., Morari, M.: Application of Moving Horizon Estimation Based Fault Detection to Cold Tandem Steel Mill. International Journal of Control (2000)
4. Zhang, X., Polycarpou, M., Parisini, T.: A robust detection and isolation scheme for abrupt and incipient faults in nonliear systems. IEEE Trans. Autom. Control 47(4), 576–593 (2002)
5. Isermann, R.: Model-based fault-detection and diagnosis-status and application. Ann. Rev. Control 29(2), 81–85 (2005)
6. Tao, G., Chen, S.H., Joshi, S.M.: An adaptive control scheme for systems with unknown actuator failures. Automatic 38(6), 1027–1034 (2002)
7. Demetriou, M.: Robust adaptive technique for sensor fault detection and diagnosis. In: Proc. IEEE Conf. Dec. Control, pp. 1143–1148 (1998)

8. Tao, G., Tang, X.D., Chen, S.H., Fei, J.T., Joshi, S.M.: Adaptive failure compensation of two-state actuators for a morphing aircraft lateral model. IEEE Trans. Control Syst. Technol 14(1), 157–164 (2006)
9. Zhang, L.-J., He, Z.-S., Zheng-Jia: Fault diagnosis approach based on hidden markov model and support vector machine. Chinese Journal of Mechanical Engineering 8(3) (2003)
10. Tian, H., Wang, A.: A novel fault diagnosis system for blast furnace based on support vector machine ensemble. ISIJ International 5(20) (2010)
11. Tanaka, H., Lee, H.: Interval regression analysis by quadratic programming approach. IEEE Transactions on Fuzzy Systems 6(4), 473–481 (1998)
12. Hwang, C., Dug, H.H., Kyung, H.S.: Support vector interval regression machine for crisp input and output data. Fuzzy Sets and Systems (157), 1114–1125 (2006)

Adaptive Wireless Network Routing Strategies for City Illumination Control System

Yang Wang

Department of Electronic Engineering
Shenzhen Polytechnic
Shenzhen, Guangdong Province, China
wyang@oa.szpt.net

Abstract. In order to control every single street lamp in city illumination system, this paper designed one wireless control system based on ZigBee technology. In this system, there is one ZigBee control module in every street lamp, which can communicate with the city illumination control system server. This ZigBee module can receive the wireless command from the control centre, and control the street lamp base on the command, such as opening or closing operation. Therefore, ZigBee module also reports the status of street lamp to system automatically. So in the city illumination system, there should be more efficient routing strategies for the ZigBee modules. Base on the feathers of city illumination system, this paper proposed some routing strategies for ZigBee, which have been practiced in one road with 30 street lamps.

Keywords: city illumination system, ZigBee module, routing strategies.

1 Introduction

The electricity consumption of city illumination in China is about 30% in the whole lighting electricity consumption. How to use various advanced scientific way to reduce city illumination electricity consumption is the purpose of energy-saving target. Furthermore, every street lamp is isolated in information system. So, the city illumination system can't control the pointed street lamp directly or get the status of the pointed street lamp. That means it is unable to realize effective centralized management, and hardly to satisfy increasingly needs of street lamp management already. Now, the city illumination system has the following deficiencies:

1) Unable to modify lamps opening or closing time effectively: At present, the lights control methods usually adopt "clock control" or "light-sensor control", but both of them can't control street lamp rationally and scientifically. For example, in practical operation, that to open lamp or to close lamp not in time usually happens. These will not only cause a huge waste of electric, but also affect traffic safety.

2) Unable to carry on collecting and querying the street lamp working state: Due to the street lamps are now isolated information island, the working condition of each light lamps can not be transmitted to outside. Now collecting and querying the street lamp working state is only to rely on artificially management, which leads to low efficiency.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 277–282.
springerlink.com

3) Unable to realize fine energy-saving function effectively: Now, the control unit of street lamps is based on sections, which is usually several miles long. So the opening or closing operation of lamps will be same in one section, which can not effectively distinguish the needs of lighting and saving-energy requirements.

4) Unable to report alarming information: Street lamps' cable and accessories are frequently stolen in some lonely roads; especially the batteries of solar street lamps are stolen in a serious condition. In these lonely roads, without the real-time alarm report system, the theft can't be found and stopped in time, which cause huge economic losses.

The problems above can't be solved in the existing city illumination management. So some new technologies should be used to overcome and solve the current problems. How to put every street lamp into the internet to realize city illumination information management? What kinds of technologies adopted to realize information transmission between each street lamp are based on the reliability, the development potentials, and the extensive adaptability. There are two kinds of technologies, which can meet the requirements of city illumination information management. One is to power line carrier communication, and the other one is wireless transmission. Power line carrier communication often is restricted because it depends on power lines. By comparison, wireless transmission is suitable for most application environment [1] [2]. This paper adopts ZigBee communication technology to realize the requirements of city illumination control system [3] [4]. Firstly, this paper designed a city illumination based on ZigBee, and then proposed the routing protocols for ZigBee wireless system appropriate for the features of city illumination system [5] [6].

2 City Illumination Wireless Control System Design

City illumination digital information management system design is the based on ZigBee wireless network technology, including street lamp terminal controller design, background management control software design, network planning and design. The focus of the whole scheme is to choose a method of communications which can bring every street lamp into internet to realize information management. On this basis, the paper discusses the network planning, functional design, and other key technical problems. As shown in Fig. 1, the whole system consists of several parts as following basically:

1) Monitoring center: It is in charge of all the functions of the whole city illumination system, including monitoring and controlling every street lamp, and collecting operation information of each one. The software system of this monitoring center can be extended in the original urban management GIS system, and also can be designed independently as a set of self-owned system.

2) 3G networks information transmission: As shown in Fig. 1, each centralized controller is in charge of information exchanging between ZigBee network and 3G system, which can connect the street lamp nodes to internet in fact.

3) ZigBee network transmission: In every road section of the system (including lots of street lamps), the monitoring center transmit control instruction to internet, and then through 3G system to centre controller of city illumination. The centre controller

receives the 3G information and then forward to ZigBee network. On the contrary, the lamps can also transmit operating status information, alarm information of themselves through the ZigBee network to the monitoring center.

4) Control node: Every street lamp as a node of information chain, not only can receive every instruction from the monitoring center, but also send its own operation information to the monitoring center actively.

5) Centre controller: As the unit, one section is including tens of street lamps, and each section has one centre controller, which can exchange information between 3G and ZigBee network. As shown in Fig. 2, centre controller as one device in city illumination system has both the 3G interface and ZigBee interface on its board at the same time.

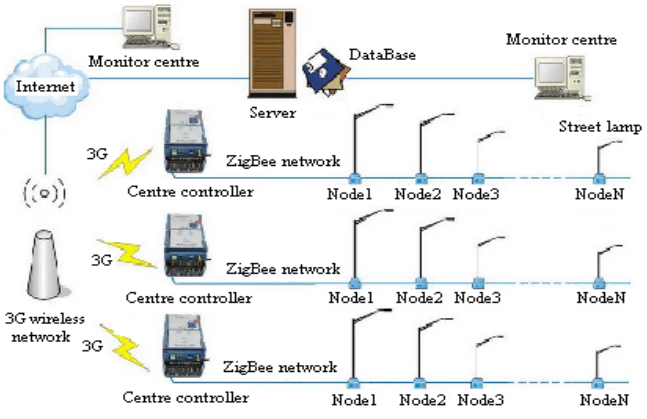


Fig. 1. City illumination system structure diagram

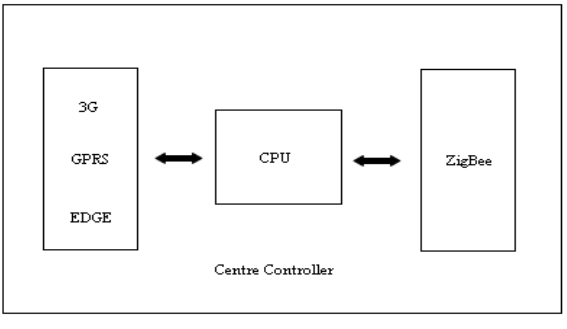


Fig. 2. Centre controller structure

6) Server: It runs server software of the monitor center, which supports C/S architecture, and allows the client through the internet to access monitor server, then realizes the remote login and manages the operation of city illumination system.

7) Database: It copies and backups the alarm message and the operation data of the street lamp in city illumination system, which can meet requirements of inquiring and searching with kinds of conditions and realize the function of data analysis.

Through the integrated background management system, the subsystems which can be realized include the following:

1) Automatic timing control mode: According to the geographical position (longitude and latitude) and the all year weather conditions of the city, designs open and close timing schedules, in this mode control node will turn on (including: daily, holidays, weekend time, etc.) or turn off the lights daily automatically.

2) Temporary operation and control: When the special conditions happen, such as important events, maintenance patrol, it can formulate temporary control strategy according to real request, and the system can execute temporary control according to the temporary strategies.

3) Immediate operation: In unexpected situations such as deteriorating weather, and on-site operation and maintenance, etc., it can control designated areas or the designated street lamp sections through the background.

4) Single point control functions: In the background management system, it can even control the single lamp in the designated area or designated sections separately, including opening lamp, closing, and adjusting brightness.

5) Remote measurement function (remote sensing): The system can get the voltage and current value of every street lamp. According to these data, it can determine whether the street lamp is normal working and calculate illumination of lamp. The system also can calculate the active power, power factor, the rate of lighting, environmental intensity of illumination, and collect the switch state of terminal control cabinet's door, and collect the operating status of the city illumination system.

6) Remote video monitoring (remote viewing): In the monitoring center, manager can monitor night scene directly through the installed camera in road.

7) Statistical functions: The system counts the measurement data, the breakdown maintenance conditions, and the alarm information, and then makes forms for reporting statistics.

8) Remote alarm functions: The system can handle the alarm information of overvoltage and over current. If the lamp is opened and closed abnormally, the system can give an alarm. And the system will report an alarm when one of the followings happens: the rate of lighting is lower than setting level; terminal control cabinet door is opened illegally; the pole's door is opened illegally; the communication link is broke down; the cable is broke down, or short circuit. And all the alarm information can be recorded and to collect statistics.

9) Large screen display functions: Screen rear projection TV wall will be adopted, which can simultaneously display computer real-time image and monitor image, and dynamically show the working condition of the all sections, each branch and each single lamp. The main contents of display includes: showing the geographical information; showing the state of the system operation; showing the parameters of the system operation; showing video; displaying alarming information.

10) Statements print: All kinds of statistical date can be print. The system can set daily, monthly and yearly report, and print any data, such as the rate of lighting, the number of broken light and its number, power, clock, date, current, voltage, the

parameters of the instantaneous high and low voltage, and etc. In addition, the system can automatically make statistics.

11) Inquiring historical operation records: According to the chosen condition, the alarm information, the fault information, and the operation records all can be inquired.

12) Remote management system: The system supports C/S framework, the client can telnet server to inquire and manage the lamp.

3 Routing Strategies Design

The city illumination is the fixed application, and street lamps location is very regular, as shown in Fig. 3.

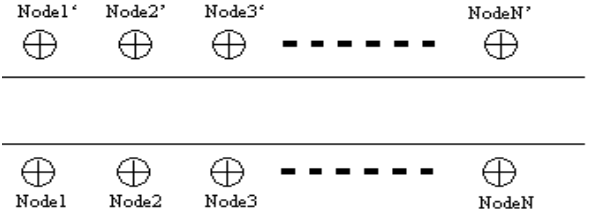


Fig. 3. Street lamp location schematic diagram

In ZigBee system, every ZigBee node should connect to the coordinator directly or indirectly. So there is star topology, mesh topology, tree topology and so on. But the location feature of street lamp in city illumination needs more efficient topology. Like in the fieldbus control system, the most appropriate topology for street lamp nodes is linear topology, which is similar to tree topology. This paper designed some routing strategies for street lamp nodes as follows:

1) All routers: The distance interval between two adjacent street lamps usually is 30 to 40 meters, but the distance of ZigBee transmitting is usually within 100 meters. So every street lamp ZigBee node should forward the information to coordinator, it means every node is router in ZigBee system.

2) Linear topology: Every ZigBee node of street lamp should obey the rule of linear topology, it means in the router table of street lamps, the nodes are one by one in order like a chain.

3) Node lost process: If one of ZigBee node is lost in system, it can't forward the information to other nodes. So, in router table, the lost node will be noted, and the two adjacent nodes will connect directly. If the system finds the lost node recovered, then the router table will restore the origin set. As shown in Fig. 4, the node1 firstly only communicates with node2, and the node 3 also only communicates with node2. But when the node2 is lost, node1 and node3 can't communicate to node2. So they will repair their router table, noting node2 lost. And then, node1 and node3 will communicate each other directly.

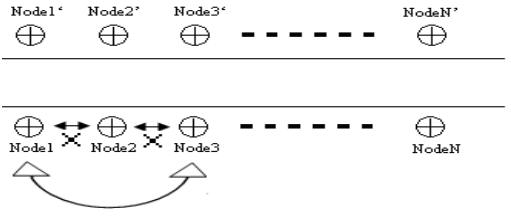


Fig. 4. Node lost working condition

4) Efficient jump: When the system initializes, every node should test their real router table, to find the longest node they can reach. And when node wants to send some unicast messages, it can jump to the longest node at first, and then reduce the transmitting period. As shown in Fig. 4, for example, node1 will find that the longest node to communicate is node3. If node1 wants to send one unicast message, it can directly communicate with node3.

4 Conclusions

There are many requirements in city illumination system now, such as information management. This paper designed one control system based on ZigBee wireless technology. Every street lamp in city illumination system can be equipped with one ZigBee module. Through this ZigBee module, every lamp can connect to internet directly, and exchange the information with outside. It means every single lamp can be controlled and monitored directly by system through internet. This can change the entire city illumination system management mode. In addition, this paper also proposed some routing strategies for the city illumination system, which can improve the efficiency of ZigBee communication. And the practice shows that ZigBee technology can change the entire city illumination management mode.

References

1. Kemal, A., Mohamed, Y.: A survey on routing protocols for wireless sensor networks. Ad Hoc Networks 3(3), 325–349 (2005)
2. Fukuhara, T., Izumikawa, H., Ishikawa, H.: Novel Multiple Path Routing Technology for Multi-Hop Wireless Networks. In: The Proceedings of the IEEE Wireless Communications and Networking Conference, pp. 362–366 (2001)
3. Pottle, G., Kaiser, W.: Wireless integrated network sensors. Communications of the ACM 43(5), 551–558 (2000)
4. Wameke, B., Last, M., Leibowitz, B., Pister, K.S.J.: Smart dust:communication with a cubic-millimeter computer. IEEE Computer Magazine 34(1), 44–51 (2001)
5. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An application-specific protocol architecture for wireless microsensor networks. IEEE Transactions on Wireless Communications 1(4), 660–670 (2002)
6. Mikko, K., Jukka, S., Mauri, K., et al.: Network signaling channel for improving ZigBee performance in dynamic cluster-tree networks. EURASIP Journal on Wireless Communications and Networking 8(3), 46–52 (2008)

Development and Application of Chinese Hamster Information Management System

Bing Kou, Tianfu Liu, Guohua Song, and Zhaoyang Chen

Laboratory Animal Center
Shan xi Medical University
030001 Taiyuan, Shan xi, China
koubing@sohu.com

Abstract. In order to improve the efficiency of traditional Chinese hamster records management, an information management system of hamsters was introduced in this paper, which combined the idea of software engineering and object-oriented techniques. This system was developed in Visual Basic6.0, based on Access2000. During the management of hamsters, this system improved the efficiency as well as reduced manual mistakes largely which was a satisfied result.

Keywords: Chinese hamster, Software Engineering, Information Management.

1 Introduction

Laboratory animal is an important support condition for the development of modern life science and the basic method and necessary way for the teaching and research of the life science and the evaluation of drugs' safety, and it has become the international "language" of the exchange of medical science and the appraisal of achievement [1], especially in biological medicine field, laboratory animal is known as "living reagents" and "the most precise instrument". During 1901 to 2008, 67.5% of the Nobel Prize in physiology or medicine research achievements was acquired by using laboratory animal or animals which involved 25 species and 119 times of animals [2]. Along with the process of China's reform and opening-up policy and world's economy and technology integration, laboratory animal has become an important factor which can affect the establishment of research subjects and the levels of research results in the fields of life science, medical science and pharmacy and so on.

The group inbred Chinese hamster in Shanxi Medical University has been domesticated from wild hamster since 1980 by our center and it was the high incidence of diabetes inbred Chinese hamster which was bred successfully in 1991 latter after 20 consecutive generations inbreeding. The traditional laboratory animal information management is filled out by hands which wastes time and effort and has low efficiency. It is more important that adopts modern information technology to establish Chinese hamster information management system.

This paper introduces a hamster-based information management system which uses object-oriented language Visual Basic6.0 and Access2000 as backstage database. The system's development has remarkable function in the aspects of data storage and

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 283–289.
springerlink.com

updating and so on. The establishment of the system achieves hamster management's informatization, provides scientific and reliable experimental data for animal experiments and provides reference for laboratory animal's informationalized management. This paper makes a detailed introduction to the system's development process according to the software's development flow [3].

2 The Environment of System's Implementation

A. The environment of hardware

The recommendation of disposition is Intel processor which is above 1.6GHz, the memory is above 512M, and the hard disk space is above 40M and printer.

B. The environment of software

This system is completed in Windows XP operating system, and the system adopts VB6.0 as the development language. VB6.0 has a strong capacity in the application and development of database, high design efficiency, powerful function and it is easy to study. It can compile the beautiful user interface and is one of the fastest and simplest programming tools under Windows platform [4].The database of backstage adopts Access2000 [5], Access is the relational data management system which is based on Windows desktop and introduced by Microsoft company. It provides table, inquiry, window, report form, page, macros and module 7 species of objects that can be used to establish the database system, provides many kinds of guides, generators and templates to make the standardization of the operations of data storage, data queries, interface designing and reports generation and the like, provides convenience for the establishment of database management system which has perfect functions, and makes common users complete most of data management tasks without compiling codes.

3 The Overall Design of the System

C. The overall structure of the system

According to the requirements and tasks of the works of hamster information management, this system mainly includes the following several modules: system setup, hamster information management, breeding information management, disease information management, hamster information statistics and system maintenance. The specific function modules of the system are as shown in figure 1.

D. The design of the database

The database's construction is the core and foundation of the construction of hamster information management system, the perfect database system is the foundation of the system's normal operation [6].

After the design of data module and the management of data's integrity and uniformity, this system mainly includes the following two tables.

- User tables, the main fields are: user ID, user name, user group, the name of user group, user level and the password.
- Hamster information tables, the main fields are: the hamster ID, family, gender, father number, mother number, parents' relations, algebra, Reproductive frequency, birth date and death date.

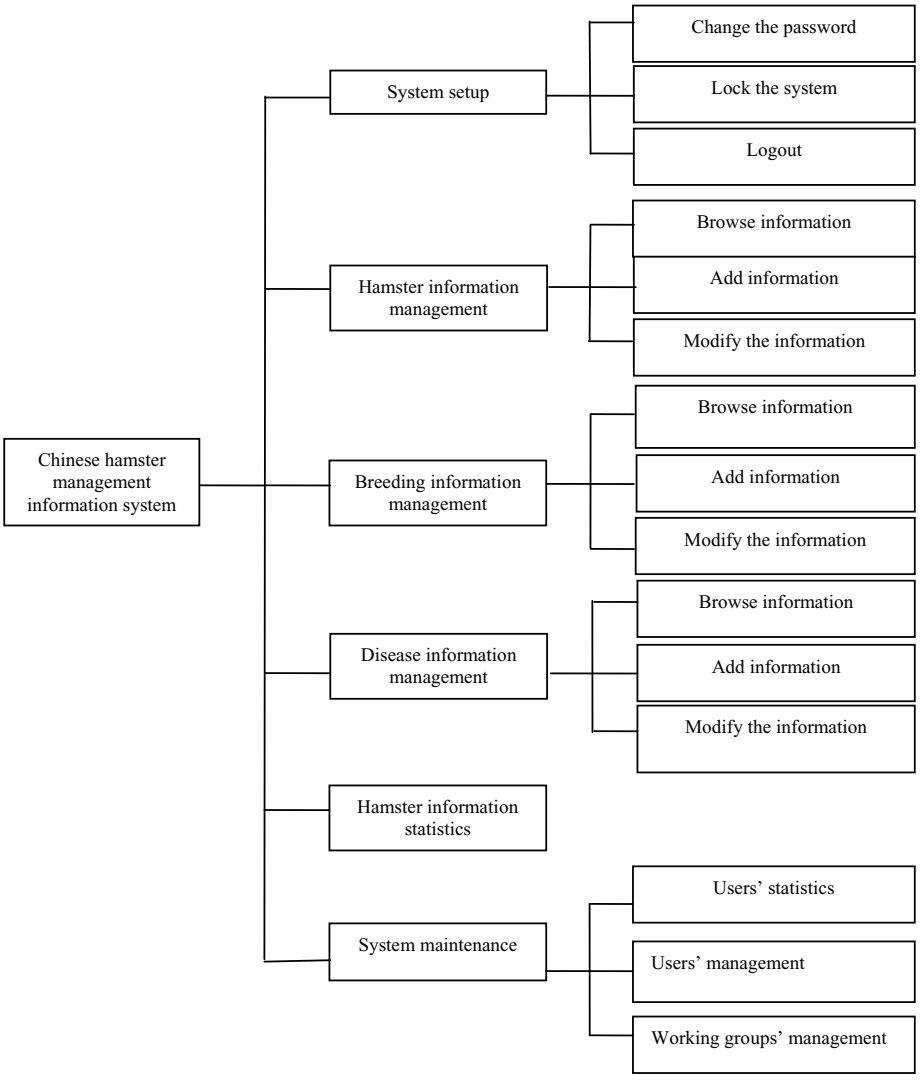


Fig. 1. The diagram of system's function module

4 The Detailed Design of the System

The system firstly established the friendly interface of system; the subject is the personalized welcome screen, it includes the button of "start", "change password", "system maintenance" and "exit system" of various types of information management window. In order to make the interface friendly and operating easily, it establishes the appropriate windows for various types of information, the controls in the windows

should be corresponded to the information which needs to be demonstrated. Windows can be achieved to add, modify, delete, query and other maintenance works for the laboratory animals' files in the windows after establishing, and it does not need to face the humdrum and bald database interface.

The data entry process of the system uses Chinese menu type driving completely. It is intuitive and easy and the operation is easy to master. The design of the system fully takes the data's security into consideration. When the system inputs the data, it carries on entering the corresponding textboxes and inputting contents after checking the integrity. It carries on the logical inspection to all the various inputted data which relate to the hamster and inquires, modifies or eliminates the too large or too small data which surpass the scope of logical inspection. It sets the internal memory function to different pedigrees and different annual material which has been inputted, and reminds whether or not to re-operate on time to reduce the unnecessary repetitive effort which is caused by people's memory errors.

The following makes the introduction for the main function modules.

E. The login interface

The users must input the right passwords when logging in, if the password is not right, the system will prompt the corresponding error message. The design of this interface well guarantees the data's security. The system interface's operating is shown in figure 2.

Fig. 2. The login interface

F. The data's adding

Because the adding processing of hamster information, breeding information and disease information of this system is similar, so this paper only directs one module to introduce. The main information is inputted through text control of VB and the inputted information will demonstrate in Grid control, it will use Grid control to input information to Access after confirming no mistakes, it firstly established a textbox

text in the window which includes Grid.vbx, the textbox's location, length and width can input data arbitrarily when established, because Grid's row corresponds to the field of database table, so it can achieve the textbox to be in accordance with the inputted rows' (namely the fields in database table) location, length, width and letterform and font through code programming, thus when inputs content to textbox, it completely equates inputting content to each row of Grid gridding externally, then transmits the textbox's content to the ranks which correspond to Grid gridding, and finally transcribes several rows of content which has joined in the gridding into the database relevant tables. And hamster information's adding is as shown in Figure 3.

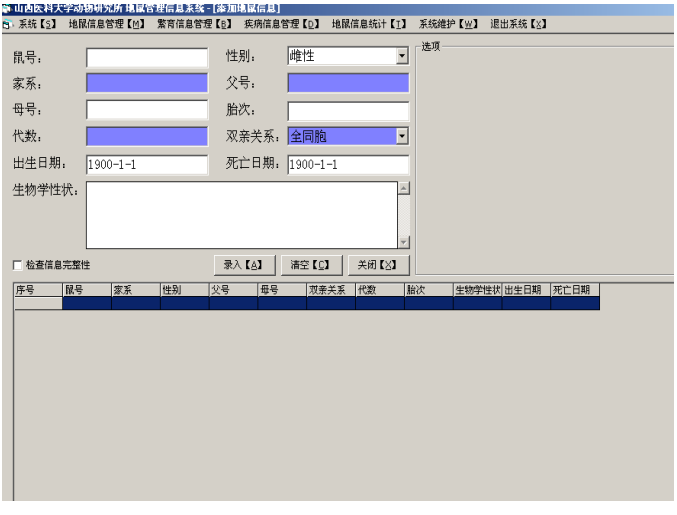


Fig. 3. The hamster information's adding

G. The realization of the function of statistics

In the guide window of information statistics, it can inquire in different methods according to number, gender, birth date and death date and so on, the interface of this window is friendly and smooth, and it can import the content which has been inquired into Excel table. In the statistical management window, it can establish query wizard to carry on multi-positions and multi-choices inquiry to this database. The users can select the fields of number, the pedigree, the copulation time, the Reproductive frequency etc to carry on inquiry. This kind of operation is advantageous for researchers to carry on the statistics and conclusion to Chinese hamster's breeding situation. This system uses precise match inquiry to make the records positioning precisely. The system's statistical interface is shown in Figure 4.

H. The further optimization of the system

The database which is not compressed will take much disk space. The compressed database will reduce obviously and even just one-tenth of the original, and still can call the data directly.

The compression method is that selects "tools" →"database utilities" →"compressing database" from the menu bar, then carries on the compression work.

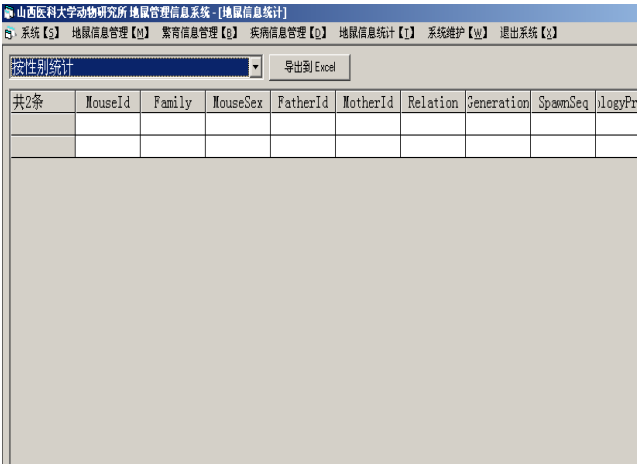


Fig. 4. The statistical interface

In addition, it can make the relevant content of "print report" direct at different requirements and output all of the data or part of the data which meet the requirements in the form of all hamster reports or individual files.

The Chinese hamster information management system that is designed with the above method is easy to be operated and maintained, it can change the various parameters of the system at any time according to the change of hamster relevant information without affecting the material of the original database. It can easily manage thousands of Chinese hamsters' files, and is exactly right.

5 Conclusions

Chinese hamster information management system provides the simplest collocation method and friendly interface for users to be convenient for expansion of the function and upgrading of the system. The system's data processing pattern requests the laboratory member to input the relevant information which is responded by them to the system, the information administrators can review and modify, it reduces the work intensity of relevant persons' collecting and processing data and improves the working efficiency; the system's automatic data processing and statistical function largely improve the data's accuracy. It achieves computer informationization management of laboratory animal management, and has important practical significance to laboratory animal management work's standardization and routinization.

Acknowledgment. Fund project: the youth funding project of Shanxi Medical University (02200935).

References

1. Shao, Y.: Modern Science and Technology Revolution and Laboratory Animal Science. Chinese Comparative Hospital Notes 14(4), 253–256 (2004)
2. Chinese Science and Technology Association. Subject Development Report of Laboratory Animal, p. 6. China Science and Technology Press, Beijing (2009)
3. Zhang, H.: Introduction to Software Engineering, 5th edn. Tsinghua University Press, Beijing (2008)
4. Liu, B.: Visual Basic Programming Tutorial, 3rd edn. Tsinghua University Press, Beijing (2006)
5. Mao, Y.: Chinese Version of Access 2000 Applications and Collection of Examples. People's Posts and Telecom Press, Beijing (2000)
6. Chen, N.: Some Application Skills of VB Database Development. Fujian Information Technology Education (2), 25–27 (2007)

Management System Design Based on Mobile Terminal for Fire Supervision and Inspection

Zhang Hui

Department of Fire Engineering
The Chinese People's Armed Police Force Academy
Langfang, HeBei Province, China
wjzhanghui@yahoo.com.cn

Abstract. Information management plays an increasingly important role in the work of fire supervision and inspection. For the actual needs of fire supervision and inspection, it is proposed to take the equipment of mobile web as the access terminals, adopt B/S architecture and use a combination of desktop and mobile web applications under ASP.Net, thus achieving the system analysis, design and implementation. The system is of easy operation and strong practicability. It provides a real-time, accurate and portable mobile information platform for fire supervision and inspection, thus playing a certain role in promoting the fire informatization construction.

Keywords: Mobile Terminal, Fire Supervision and Inspection, Management System.

1 Introduction

In 2009, The Ministry of Public Security issued "The fire Supervision and Inspection Rules"(The Ministry of Public Security Order No. 107), provided the content and procedure of fire supervision and inspection for the firefighting organ, to further develop the overall combat capability of the public security organ, improve and perfect the law-enforcement supervision of the firefighting organ. It improves the constitution of fire enforcement power, also asks the work of fire supervision and inspection for higher requirements and standards [1]. However, in the current work of fire supervision and inspection, the fire supervisors and inspectors are faced with the problems of little supporting information, low timeliness, poor flexibility and so on, posing a direct impact on the smooth carrying out of the work of fire supervision and inspection.

Traditional fire safety supervision and inspection systems are usually based on general PC. They are not portable, inconvenient to use and influenced greatly by their sites and other factors, to a certain extent affecting the efficiency of fire inspection and enforcement. Mobile terminal equipment wirelessly accesses to system information, which allows the fire inspectors to timely query the information of the place of work, fire laws and regulations and so on in the inspection field. Meanwhile, the mobile device features a unique dial-up function, which helps the supervisor

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 291–296.
springerlink.com

easily contact the relevant places of work and back-office personnel for remote control in a timely manner. In order to let the fire supervisors timely and accurately record inspection during their supervision and inspection on the key places of work, the author develops and designs a management system design based on mobile terminal for fire supervision and inspection, so as to provide the fire supervisors a real-time, accurate, and portable mobile information platform.

2 Mobile Web Development Technology

A. ASP.NET Mobile Designer

The ASP.NET Mobile Designer extends ASP.NET and. NET Framework, which can be used to generate the web applications of mobile phone, PDA and pagers. By using C # language under the VS.NET 2005 integrated development environment, the design and implementation of information query, browse and other modules are achieved through the WAP browser in the mobile terminal.

To develop ASP.NET mobile Web application program, we must reference the namespace System .Mobile.UI provided by .NET Mobile Web SDK (through MobileUI.DLL file). . NET Mobile Web SDK provides three container objects: MobilePage, Form and Panel. MobilePage control is an important container of mobile application program, equivalent to an ordinary Page object under ASP.NET. There can be one or more Form controls in a separate MolilePage. The concept of the Form controls is basically the same with the general one under ASP.NET. There can be 0 or more Panel control in a mobile Form control. The Panel control is used to divide a variety of Mobile controls into groups. Molile controls can be divided into three groups. User interface (UI) controls: the same group of controls as Lable controls which allow the user to control the user interface. Basically the group of controls can correspond with the general controls under ASP.NET one by one. Verify (Validatiln) control: allows users to verify the correctness of input values, such as the Required Field Validator control. These controls validate users' input data before sending data to the server. This group of controls is very similar to the general validation controls under ASP.NET. Function (Utility) control: similar to the class of calendar control [2] [3].

B. WAP system structure

WAP system structure is similar to that of WWW, shown in Figure 1, which gives Web application developers many benefits, for example, the software system structure is familiar, and very effective, and the existing development tools can be used. In order to adapt to the characteristics of wireless application environments, WAP technologies are all tried to make use of the existing technical standards which are taken as a starting point for WAP technology development. WAP control and application are adopted a set of standard communication protocols based on WWW communication protocols to transfer. The micro-browser of wireless terminal, similar to a standard web processor, is taken as a common user interface [4].

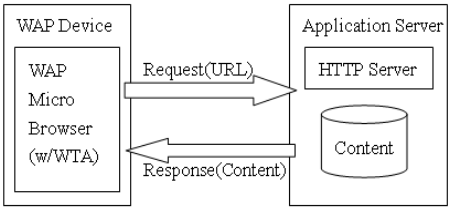


Fig. 1. Diagram for WAP system structure

3 System Analysis

The main functions of this system are: to achieve the records of information about users' supervision and inspection on the front page, view the basic information about the places of work, and visit the related laws and regulations; on the background page, the administrator can add, delete and modify the contents of each module, and other maintenance operation with them. And it can receive timely the information about supervision and inspection on the mobile terminal, thus filling in and printing legal documents. Users can achieve the wireless access through the mobile terminal device, which helps them get rid of constraints of time and space, thus providing fire inspectors a real-time, accurate, and portable platform for the supervision and inspection, also providing the supervision and inspection of the fire fighting forces an efficient platform.

The system uses VS.NET 2005 integration development environment. Microsoft SQL Server 2005 is being developed as a database development tool. By B / S structure, the software can be installed only on the server side, which means the client can achieve a "zero " installation, besides the management and upgrade of all the system only need to be done on the server side, therefore each client does not require any setup and change. The updates and upgrades of the system will not affect the user's data. On the server side, the system administrator can manage the various modules of the system, to achieve the functions of adding, deleting, modifying, uploading and so on. On the client side, the user can use the portable mobile terminal equipment for a visit.

4 System Overall Design

C. Functional Structure Design

The system is mainly for such two groups as the administrators and fire supervisors and inspectors, whose needs are met by planning out the two parts of the system module in system design. In accordance with the requirements of structured programming, the system is composed of such two major components as the mobile terminal and back pages. The mobile terminal achieves the functions of supervision and inspection, while the back page achieves the functions of management and the production and printing of legal instruments. The functional structure design is shown in Figure 2:

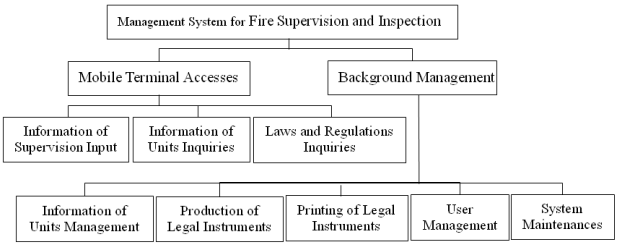


Fig. 2. Diagram for the functional structure

D. Database Design

According to the overall design of the system modules and the specific content and requirements of the fire supervision and inspection, the design of the system include 31 tables and a data table diagram [5]. For example, the supervision and inspection record sheets are used to record some records input by the mobile terminal about key organs on fire supervision and inspection. These records can also be spread to the back page for some relevant management and printing of legal instruments.

5 Detailed Design

The man-machine interface design, including the hardware interface and software interface, are the media between people and computers to pass and exchange information. The quality of human-machine interface affects how the users feel the software. Therefore, the quality of human-machine interface has become one of the criteria to measure the usability of software. Good human-computer interface has also become an important aspect of software design [6]. The system has two main human-machine interfaces: the bake page and the front mobile terminal interface.

E. Back Interface Design

The back page interface is divided into information management of key organs and producing and management of legal instruments, and other parts. It is divided into three levels of permissions of administrators and users. Among them, the organ information management page include some operations like adding, deleting and modifying the basic information on key organs, basic information of construction, fire information, and other 17 categories of basic information. Meanwhile, by the left navigation we can select the specific information item on what we want to manage, which is shown in Figure 3.

F. Front Interface of the Mobile Terminal

The user designs the mobile web form and mobile web controls, to achieve its functions of supervision and inspection and inquiry access. According to the characteristics of mobile terminal devices and taking the needs that the fire supervisors and inspectors can easily query and input information, into account, we should set more multi-select controls or user-based input controls. The specific design is as follows:

(1) Home: After login into the home page, the three functions of the supervision and inspection records, browsing of laws and regulations and inquiries of organ information can be selected.
(2) The supervision and inspection records: the optional filling in of supervision and inspection records is proposed in accordance with the conditions of supervision and inspection, with inspection date defaulted to the current system date.
(3) Organ Information Inquires: according to the names of key organs, such information as unit number, address, telephone number, legal representative and other information can be inquired.

Fig. 3. Interface of adding organ information

Records of supervision and inspection are shown in Figure 4.

Fig. 4. Interface of records of supervision and inspection

6 Conclusion

The system uses VS.NET 2005 integrated development platform, SQL Sever 2005 database and the Photoshop image processing technology, to achieve the management system design based on mobile terminal for fire supervision and inspection. In the back page, administrators can do the follow-up work of dynamic maintenance and the supervision and inspection by ordinary PC; in the front page, the user can use the increasingly popular mobile terminal devices for wireless access to information, information input of supervision and inspection. In this way, it not only meets the maintenance requirements of the large amounts of data on the back page, but also satisfies the fire supervisors and inspectors on their facilitate access to information and information input. The system in this paper is very practical.

References

1. Fire Supervision and Inspection Rules (The Ministry of Public Order No. 107), Ministry of Public Security (2009)
2. Li, C.-D., Zhang, J., Liu, F.Y.: Research and Implementation of Fox ERP Based on Movable Web Application Technology. Computer Science 34(17), 128–130 (2007)
3. Wang, C.: Technology and Application of Web Development, pp. 24–27. Tsinghua University Press, Beijing (2007)
4. Liu, R.-X., Fang, J., Zhang, S.-N., Li, J.-P.: Research and Development of Self-adaptive Mobile Terminal Framework. Computer Engineering 35(18), 266–268 (2009)
5. Sa, S.X.: Database System, pp. 244–247, 147–152. Higher Education Press, Beijing (2006)
6. Zhang, H.: Introduction to Software Engineering, pp. 212–215. Mechanical Industry Press, Beijing (2004)

Analysis and Design of Third-Party Logistics Information System

Ying Jiang[1] and Li-jun Zhou[2]

[1] Department of Foreign Languages, Nanjing Institute of Industry Technology, Nanjing, China
jiangy@niit.edu.cn
[2] School of Economy and Management, Nanjing Institute of Industry Technology, Nanjing
zhoulj@niit.edu.cn

Abstract. The logistics informatization is the core of modern logistics and shows the core competitiveness of third-party logistics enterprise. It is very important to strengthen the construction of logistics information system. The article starts with the connotation of logistics information system, then analyzes its system structure. Based on the analysis, the design of third party logistics information system shall be made.

Keywords: third-party logistics, information system, system design.

1 Introduction

Nowadays, with the development of the information technology and electronic commerce in China, third-party logistics grows quickly and has become an important impetus for the development of national economy. However, because many logistics enterprises arise from the former transportation enterprise or warehouse enterprise, lack of the necessary integration of management information system and business, it is difficult for them to meet the needs of electronic commerce for logistics distribution. Logistics informationization has become the bottleneck of restricting our logistics industry development. So, it is the problem to be solved to improve the level of third-party logistics enterprise informatization and construct the logistics information system which adapt to the logistics information system of third-party logistics infomatization.

2 The Connotation of Third-Party Logistics Information System

The third-party logistics information system, consisting of personnel, equipment and procedures, refers to man-machine interaction one which aims to carry out programs for the administrators of third-party logistics, provide the related information for the operation of such functions as its implementation and control. Besides, according to the special characteristics of third-party logistics, a series of effective control and management activities shall be conducted for the information from the logistics process by modern information technology to be collected, classified, transferred, identified, tracked and inquired, etc. And the system provides the support of information analysis and decision-making with the features of individuality, real-time,

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 297–305.
springerlink.com

systematization, network, large-scale, specialization, integration and intelligence so as to achieve the control on the logistics process, thus reducing the cost and improving efficiency.

The present third-party logistics information system is more dependent upon information technology and network technology than the previous one. The greatest advantage of the third-party logistics information system is to let the third-party logistics enterprise focus on its core business, be committed to its dominant field so as to provide the service of flexible individuality and specialization for customers. Therefore, the third-party logistics enterprise can meet customers' requirements rapidly and effectively with low cost by means of advanced information technology.

3 Features of the Third-Party Logistics Information System

The third-party logistics information system can not only reduce the cost of enterprise's operation, improve its operational efficiency and the level of customers' service, but also gradually enrich and accumulate the logistics management knowledge in the process of using logistics information system. The choice of logistics information system for enterprises is rather the choice of enterprise management mode and market competition strategy than the choice of information technology.

Logistics information system should have the following features:

A. Accessibility

The third-party logistics information system should have easy and consistent accessibility. And the necessary information includes order and inventory condition. When the enterprise is likely to acquire the important data of logistics activities, it is easy to get from a computer system. Rapid accessibility is necessary for customer service and the improvement of management decision because the customer shall ask for frequent accession to the information.

B. Accuracy

Logistics information system should accurately reflect the current logistics service status and regular activities so as to measure the order and inventory levels. When the actual inventory level has the low consistency with the system, the measures of buffer inventory and safe inventory should be taken to adapt to the uncertainty.

C. Timeliness

Timeliness refers to the time delay when an activity happens in the information system. The timeliness of information system includes systematic situation and the timeliness of management control. As for the timeliness of management control, the information can be provided within the time to take the right actions or minimize the loss. Generally speaking, timely information decreases the uncertainty and is helpful to identify various problems so as to reduce inventory requirements and increase the accuracy of the decision.

D. Flexibility

Logistics information system should be flexible in order to satisfy the needs of both the users and the customers. Meanwhile, information system should have the ability

to provide the data which can cater for the special customers. To be specific, information system should have the ability to update and not only to meet the future requirement of the enterprise but also not to weaken in financial investment and scheduling time.

4 Information System Structure

At present, information system structure is mainly composed of two-tier CS mode, three-tier CS mode, three-tier BS mode and three-tier mode of CS and BS integrated. Considering the strong practical processing ability and expansibility of the three-tier mode of CS and BS integrated, the mode has been widely used in third-party logistics. The system of the three-tier mode of CS and BS integrated is shown below in Figure 1:

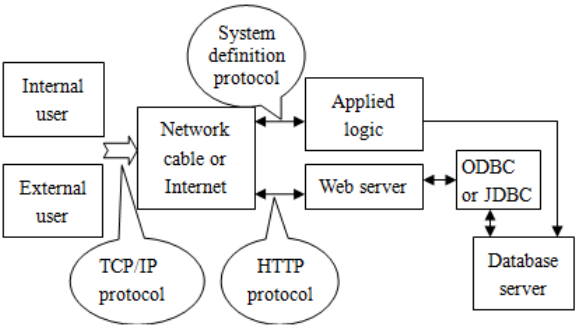


Fig. 1. Structure mode of 3-tier CS and BS integrated

In the system structure, the internal user of the enterprise can directly adopt the TCP/IP protocol of the lower level because of the high requirements of the safety, strong interaction and a large number of processed data, etc. towards system module. According to the actual situation of the enterprise information system, its communication protocol is defined so as to connect the user with applied logic. Then through the data exchange between ODBC or JDBC and database server (DBS), the research results in DBS are returned to the applied logic then to the user. Here the use of three-tier CS system structure is called “enterprise internal mode”.

The external user of the enterprise asks for the compatibility of the different models and operating systems with good openness and universality, strong cross-platform so as to achieve the goal of remote data transmission. Therefore, a large number of users can indirectly exchange data with WEB server through HTTP protocol based on TCP/IP protocol. Here the use of three-tier BS system structure based on Internet and Web can be called “enterprise external mode”. Therefore, based on TCP/IP protocol and different network protocol, through three-tier system structure of CS and BS integrated, all users can have access to internal interface and external interface of the enterprise so as to make full use of the advantages of two modes to adapt to the increasingly complex operation conditions of the enterprise.

5 Design of the Third-Party Logistics Information System

The framework of the third-party logistics information system is shown in Figure 2:

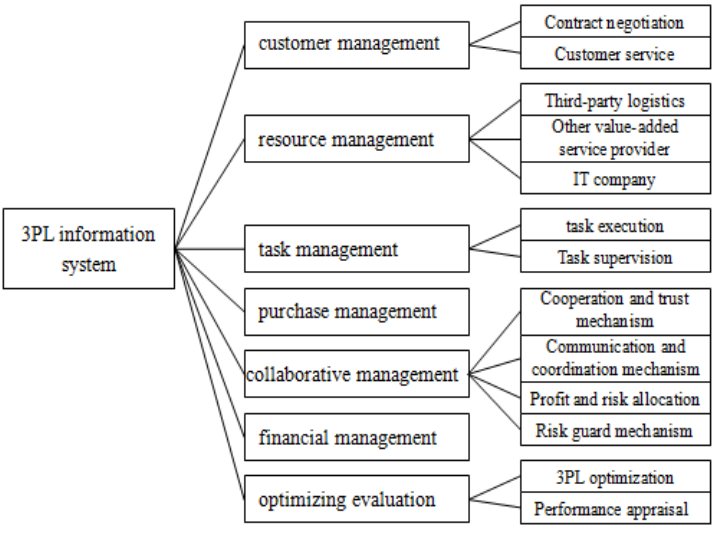


Fig. 2. Framework of third-party logistics information system

E. Subsystem of customer management

The subsystem functions as the communication and consultation between the third-party logistics and customers. And it aims to specify the functions of order management, including customer service and contract negotiation.

1) Customer service module (as shown in Figure 3)

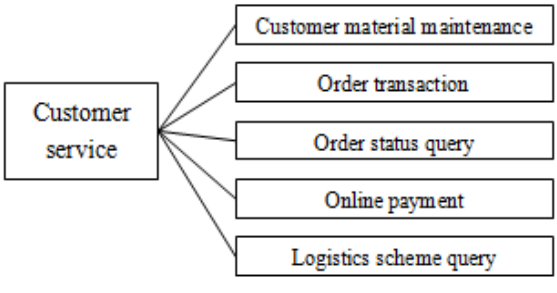


Fig. 3. Customer service module

2) Customer service module mainly includes two aspects: one refers to such tasks as receiving customer orders, verifying the identity of the customer, confirming and managing order information and the inquiry and feedback of logistics service from the customer and the inquiry for the execution of logistics business so as to guarantee that

the customer can receive the goods on time, in the right place, for the right amount and accurately and provide the customer with the instant information and logistics consultation for the goods ordered. The other refers to the management of customer's basic information, namely, customer geographical position, goods type, shipping requirements, orders, storage requirements and other services on his behalf, etc. Contract negotiation module (as shown in Figure 4).

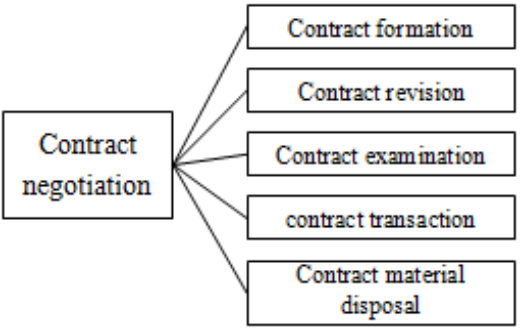


Fig. 4. Contract negotiation module

The module is mainly involved in the negotiation on the adjustment of the order and the contract with the customer, which aims to complete such tasks as the formation, addition, deletion, revision, query, examination and print of the contract. Contract transaction includes the amount details in the contract, the summary table of the contract execution report and the expiring payment list in the contract, etc.

F. Resource management subsystem

The subsystem functions as the following two aspects: one is involved in logistics resources management such as the transportation capacity of transport service provider, the inventory ability of inventory service supplier, the service ability of the third-party logistics and other value-added service provider. The other refers to the management of logistics service supporter of management consulting company and IT company, etc.

1) Third-party logistics service module (as shown in Figure 5)

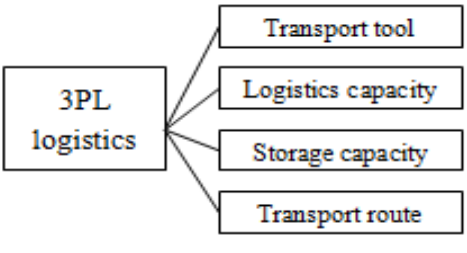


Fig. 5. Third-party logistics service

The module mainly includes 3PL geographical position, the operation business scope, transport tool, transport route, storage capacity, agency service category, etc.

2) IT company service module

The core business of this module focuses on the development of information system and the design of logistics scheme with the scientific and reasonable solutions to the related problems.

3) Other value-added service providers (as shown in Figure 6)

The service providers mainly refer to some professional service providers of marketing, packaging, processing, distribution, transportation and inventory, etc.

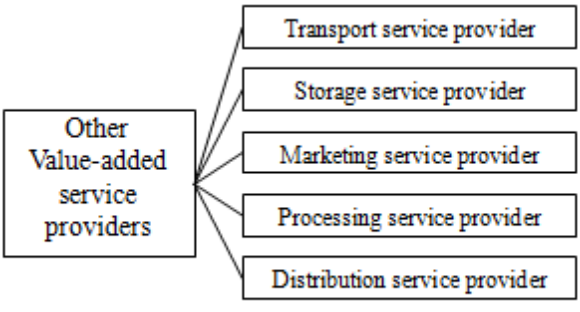


Figure 6 other value-added service providers

Fig. 6. Other value-added service providers

G. Purchasing management subsystem (as shown in Figure 7)

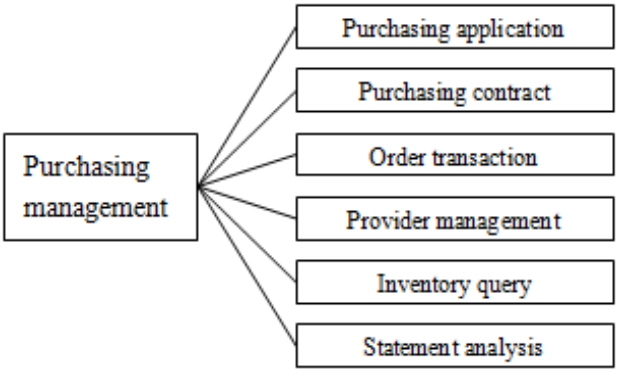


Fig. 7. Purchasing management subsytem

Purchasing management subsystem mainly deals with the interaction business with upstream suppliers and this module aims to provide purchasers with a set of quick and accurate tools for the appropriate suppliers to issue procurement request in the right time and in the right amount so that the goods can be put in storage on time before delivery and the excessive inventory or shortage can also be avoided. Besides,

purchasing management subsystem mainly functions as the statistics of the quantity of the goods needed, the inquiry of the supplier, the negotiation on transaction terms, the calculation of economic order quantity and the purchasing order and tracking behavior for the goods issued by suppliers. Purchasing management module mainly focuses on inventory control, delivery date control and manufacturer information collection.

H. Task management subsystem

Task management subsystem is responsible for the design, decomposition, combination, control and supervision, etc. of logistics tasks. The subsystem is composed of two modules which deal with the execution and supervision of tasks respectively.

1) Task execution module

The module is involved in the assignments to designated transportation service provider, inventory service provider and other logistics alliance. The responsible logistics provider gives feedback and the agreement shall be signed after the confirmation with the beginning of the relevant logistics plan.

2) Task supervision module

The module is to supervise the execution of logistics tasks.

I. Collaborative management subsystem

The subsystem is mainly to coordinate the contradictions and conflicts between enterprises or logistics operations in the third party logistics.

1) Cooperation and trust mechanism

Cooperation and trust mechanism is the premise and basis of the successful operation of the third party logistics. The cooperation based on trust is helpful to increase the transparency of every party of the alliance, reduce transaction cost so as to make every party of the alliance take more active attitude to cooperate, thus realizing the effective integration and utilization of resources.

2) Communication and coordination mechanism

Given the differences of every enterprise in organization scale, management idea, operation pattern and organization culture, the conflicts arising from their cooperation can not be avoided, which requires the effective communication and consultation among the members so as to subordinate the general interests of the alliance.

3) Profit and risk allocation

The successful operation of supply chain collaboration should be based on the formulation of fair and reasonable profit allocation scheme. Profit allocation indicates that every member enterprise is allocated the deserved profit from the overall profits according to its tasks in the logistics operation and its logistics service quality so as to avoid the contradictions and conflicts among members arising from unfair allocation.

4) Risk guard mechanism

The risk of supply chain alliance under the third-party logistics mode is more complex than that of the single enterprise. Besides the risk the common enterprise faces, many risks may arise from alliance. Therefore, risk information and related knowledge should be collected and the risk should be reasonably predicted. The warning and predicting system of risk management should be established so as to eliminate or effectively control risks at source. The previous risk should be assessed and the risk awareness strengthened to increase the control for future risk. After being

warned by the warning system, emergency system should timely handle the emergencies so as to reduce or avoid the serious consequences among the enterprises on supply chain.

J. Financial management subsystem

The subsystem refers to the budget and management of the logistics cost of the operation in the third-party logistics and the allocation of enterprise profits. It mainly includes various expenses, such as warehousing fee, transportation costs, loading and unloading expenses, administrative expenses, the settlement of administrative expenses, the settlement of receivables and payables with customer, and the profit allocation with logistics service provider, IT company and consultation company, etc.

K. Optimizing evaluation subsystem

1) Optimizing the solutions of 3PL supply chain

The module can optimize the design, decomposition, combination and logistics route of the logistics tasks from customers. 3PL provider should establish appropriate network model based on the current logistics information, produce several alternative transportation route according to the actual needs of customers through the decision analysis of intelligent system and expert system and design the optimal solutions of supply chain so as to reduce the unloaded ratio and cost of the transportation.

2) Choice of 3PL provider

On the basis of the evaluation of 3PL providers, the analysis of its logistics cost, time and quality shall be made so as to finally select the appropriate 3PL provider.

3) Performance appraisal

In addition to reflect the correctness of formulation of various operational management strategies and the implementation of plan, performance appraisal aims to provide the foundation for the analysis of policies, the revision of management and implementation method. The performance appraisal of 3PL enterprise is mainly composed of personnel management appraisal, provider management appraisal, order processing performance appraisal, stock service appraisal, transportation service analysis and 3PL performance analysis.

6 Conclusion

With the rapid development of the third party logistics, enterprise logistics information system becomes more and more important. How to design an effective logistics information system becomes a problem to be solved. In fact, perfect third-party logistics management information system shall lead to the advanced and modern management for enterprise, thus making enterprise more competitive in the fierce market competition.

References

1. Zhou, L.: The Application of Supply Chain Performance Evaluation in the Competitive Ability of Enterprise Based on SCOR Model. Journal of Nanjing Institute of Industry Technology (4), 62–65 (2009)

2. Xu, Q.: Third Logistics Information System Function Model Design Analysis. CD Technology (6), 18–19 (2006)
3. Chen, H.-X.: Design and Analysis of Logistics Information System Based on SOA. Logistics Sci-Tech (3), 77–78 (2008)
4. Chen, Y.: Third-party Logistics Information System. Technology Plaza (3), 219–220 (2009)

A Design of Multi-temperature Monitoring System Based on GSM

Jingjing Wu, Luqian Wang, and Jinping Li

Information College
Beijing Union University
Beijing, China
{Xxtjingjing,xxtluxi,xxtjinping}@buu.edu.cn

Abstract. The multi-temperature monitoring system that was designed in the paper constituted by the monitoring center and a number of the terminals. It can monitor simultaneously several terminals. Each terminal is more than 10 meters distance apart and some terminals interval is between the walls. The monitoring temperature data that is from each terminal temperature data to the control center is transmitted through GSM network. What's more, an alarm signal will appear to the control center when the temperature exceeds the value of alert temperature.

Keywords: monitoring system, terminal, monitoring center, GSM.

1 Introduction

With the development of computer technology, particularly internet technology and communication technology extensively in all aspects of people's lives, people's lives were made profound change. From the industrial control field, there are wide variety of monitoring objects, which need a lot of manpower, material and financial resources to maintenance [1]. And there are many conditions for people is not easy to reach, or not always need a place to stay occasionally a number of field data acquisition, a lot of wiring work is uneconomical and unreasonable.

The monitoring system was designed to monitor multiple terminals, in the paper the terminals were designed to detect the temperature data, and the interval distance between terminals is far and not layout. Every terminal can capture the current temperature data at regular intervals, and compared with the alarm value. If the value exceeds the alarm value, it will send short message to the monitoring center and officers on duty through GSM (Global System for Mobile communication) network. Monitoring center also can send control commands to every terminal by the GSM network in order to read the current value of temperature or set the alarm value.

2 GSM Network and MC35i Module

GSM system is based on European Telecommunications Standardization Committee (ETSI) specified in the GSM technical specifications developed from the second

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 307–313.
springerlink.com

generation of cellular mobile communication system. GSM system is a digital cellular mobile communication system, which consists of several sub-systems with a variety of public communication networks such as PSTN, ISDN and PLMN such as interoperability [2].

As the GSM network realize link and roam throughout the country, with more advantage than other traditional network can not match, coupled with GSM data transmission itself has features that allow data transmission applications received spread rapidly based on GSM mobile network. Nowadays GSM networks at home and abroad carried out by messaging applications are very large, such as mobile phone users to receive weather forecasts, financial quotes, news and entertainment, transportation schedules a text message. Short message for its effectiveness, cheapness, can be used by the transmission vector as the remote monitoring system.

In order to communicate through GSM network, Clients must select the appropriate mobile terminal. MC35i is an industrial-level GSM module by Siemens. It has all the features of TC35i and easy integration. MC35i support voice, data, fax, short message and other functions, using the 3V SIM card to provide 9-pin RS232 data interface, controlled by AT commands, mainly for voice transmission, short messaging and data services, wireless interfaces. Its receiving rate can reach 86.20kbps; sending rate can reach 21.5kbps. Compatible with TC35i, MC35i also supports GSM900 and GSM1800 dual-band network. MC35i via RS-232 serial port and connect their wireless applications, and using the standard AT commands to control it, while it collected the information transmitted by the GSM network to the monitoring center [3-4].

3 Overall Design of Multi-temperature Monitoring System

3.1 Overall Design

The multi-temperature monitoring system, studied in the paper, is to achieve the data receive and send instructions through to send or receive short message [5]. MC35i module both the monitoring center or terminal, or the duty officer cell phone, send messages to short message service center and the center is to send, receive messages when they are receiving from the information center. Short message service center is responsible for forwarding the user's information, in terms of its effect is transparent to users, clients only use the appropriate short message service center number according to the SIM card number (Mobile or Unicom). The block diagram of the overall design of multi-temperature monitoring system as shown in Figure 1:

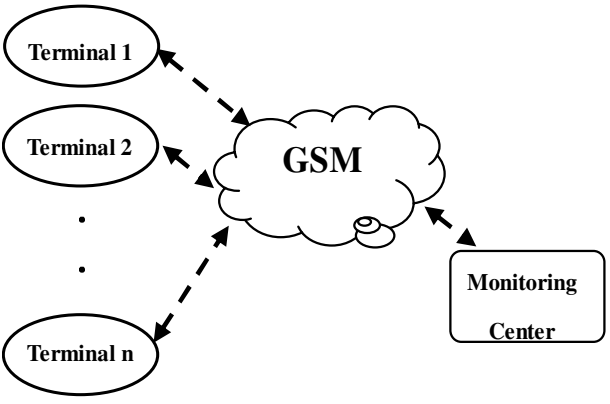


Fig. 1. The design diagram of multi-temperature monitoring system

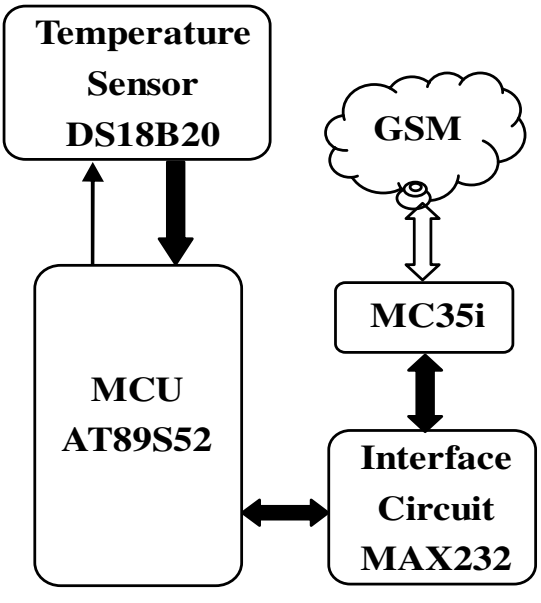


Fig. 2. The design diagram of one temperature acquisition terminal system

In the terminal, MCU collected temperature data at regular intervals, the signal collected by the temperature sensor (DS18B20) directly into a digital signal into MCU, then MCU control the communications module (MC35i) to send the data to the monitoring center via GSM network. In addition, the monitoring center also can send commands to the terminal through the communication module (MC35i). When terminal system receives the command, MCU can analyze the command and complete the content requirements in accordance with the instructions. Also, if the terminal control system wants to monitor multiple or more signals, such as temperature,

voltage, humidity and other signal to monitor, simply add the sensor in the terminal and the data representative the content of the signs, then sent to the monitoring center, and monitoring center can be identified by distinguishing. In addition, MCU in terminal control system can determine whether the value of the monitoring to meet the conditions for sending information such as temperature exceeds the set alarm value. If the condition is met, the notice will send to the control center or duty officer cell phone through the communications module (MC35i).

Monitoring center system is responsible for sending control instructions and receiving the data or abnormal data signal from the terminal, and analysis the data, then show the status of each terminal. If receiving abnormal state data, monitoring center system will provide alarm display, and the officers on duty at the control center can send control commands to the monitoring terminal to change the status of field work. Otherwise, the officers on duty can receive short message via mobile phones or send short message to the command to terminal to control the status of the terminal.

3.2 Format of Transfer Data in Multi-temperature Monitoring System

In the design, sending and receiving data is the entire short message. Short message is divided into three modes: Block mode, Text mode and PDU mode. The message whether sending or receiving was adopted Text mode, as the issue of data transmission is temperature data.

The messages from the terminal to the monitoring center transmit using the following format:

"$ Digital 0 @ Digital 1 Terminator"

In the data format, the "$" is the starting character means that the temperature data will be sent; "Digital 0" indicates that the number of terminal where the current data will be collected ; " Digital 1" indicates the temperature of the specific values, including positive and negative, in which a positive numbers representing the current temperature is higher than zero, negative numbers representing the current temperature is below zero; "Terminator" shall mark the end of a whole string of data that this end of the transmitted data. "@" is the symbol interval between "Digital 0" and "Digital 1".

The commands from the monitoring center to the terminal transmit using the following format:

"* Digital 0 @ Digital 1 Terminator"

In the data format, the "*" indicates the starting character to send command; "Digital 0" indicates that the number of terminal where the commands will be send; "Number 1" indicates the command code, such as "1" represents setting the alarm value of temperature, "2" represents reading the current temperature, etc., the meaning of the figure representing can be adjusted according to user needs; "Terminator" shall mark the end of a whole string of data that this command sent by the end of.

There are no spaces between numbers in these two data formats.

4 The Design of One Temperature Acquisition Terminal System

One temperature acquisition terminal system is composed by the AT89S52 microcontroller, GSM wireless data transmission module MC35i, temperature sensor DS18B20, etc., and its block diagram shown in Figure 2:

As a core part of the terminal system, AT89S52 microcontroller controls the temperature data acquisition and converts it to fit the data transfer mode. Then these data were transmitted to the monitoring center or cell phone on duty through GSM wireless data transmission module MC35i. AT89S52 microcontroller also receives, analyzes and manages the instructions by the control center through MC35i to direct treatment to control the work of terminal itself. The digital sensor DS18B20 is responsible for acquisition the temperature and conversion to suitable format to MCU. GSM wireless data transmission module MC35i is responsible for sending data collected to the monitoring center or cell phone on duty, and receiving the commands from the monitoring center or cell phone.

In the design, the GSM wireless communication modules MC35i as an independent communication module provides the interface to the microcontroller circuit is a standard RS-232 interface, which consistent with EIA RS-232C standard. Logic level of RS-232C uses a negative logic, which logic "1" expresses the-5V to-15V; logic "0" expresses the +5 V to +15 V. In CMOS logic level, "1" expresses 4.99V, "0" expresses 0.01V. In TTL logic level, "1" expresses the 2.4V, "0" expresses the 0.4V. So the converter must between the MCU and the MC35i.

MAX232 is a chip compatible with standard RS232 by Texas Instruments (TI). The device contains two drivers, two receivers and a voltage level generator circuit to provide TIA/EIA-232-F. The devices meet the TIA/EIA-232-F standards, each receiver will transmit TIA/EIA-232-F level into 5V TTL/CMOS level, and each transmitter will transmit TTL/CMOS level into TIA/EIA-232-Flevel. Each transmitter will be TTL / CMOS level conversion to TIA/EIA-232-F level. The external circuit is simple, just four 0.1μF external capacitor can be. [6]

MCU and MAX232 connection method is that TXD and RXD were connected corresponding with the MAX232 TTL / CMOS input and output ports. Figure 3 shows the specific link:

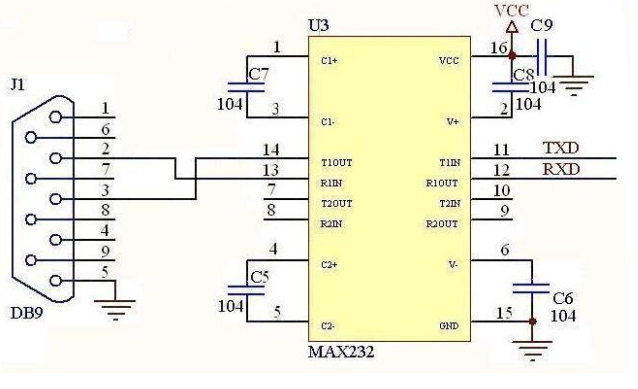


Fig. 3. Interface circuit between AT89S52 and MC35i

The work flow of one terminal was shown in figure 4:

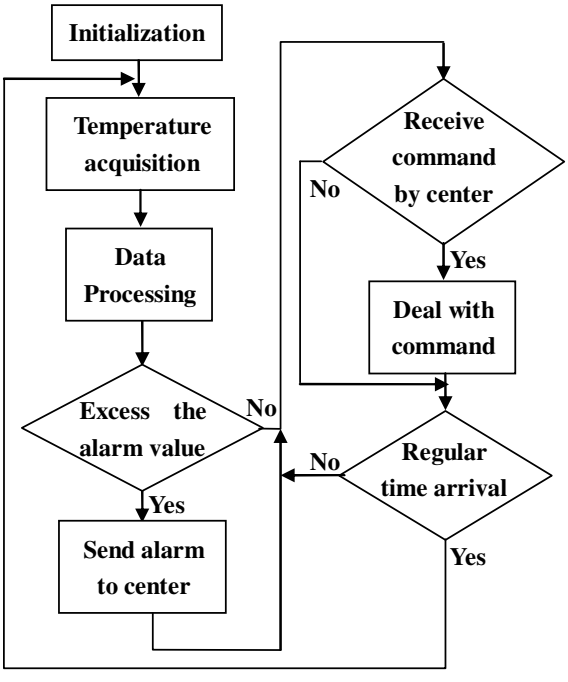


Fig. 4. The work flow of one temperature acquisition terminal system

5 Design of Monitoring Center System

AT89S52 as the core parts of the monitoring center system, with LCD HD44780 to display the temperature data that achieve from the terminal or other information. The monitoring center receives the temperature data or sends commands by wireless transceiver modules MC35i. In order to implement system functions, complete temperature acquisition, real-time display and control functions, AT89S52 microcontroller was designed as the core in monitoring center, supplemented by wireless transmission circuit, display circuit and keyboard. The system hardware design mainly consists of five parts: the MCU module, the wireless transmission module, display module and keyboard module, and its block diagram shown in Figure 5.

Monitoring center can send control commands to any terminal by keyboard, set the temperature alarm value or receive current temperature data of any terminal, which information can be observed through the LCD.

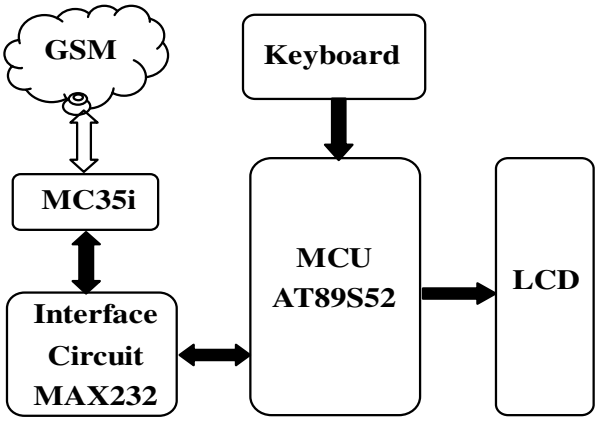


Fig. 5. The Design diagram of monitoring center system

6 Conclusion

The multi-temperature monitoring system is based on the GSM network. It has the advantage of set temperature alarm and detection of more than one terminal at any time, with low cost, small size, transmission distance, and strong anti-interference, and wide application. Meanwhile, according to the requirements of monitoring data, the system can replace sensors and expand the system to adapt to higher and more widespread application of monitoring.

Acknowledgment. This research has been supported by "Virtual Experiment Application Demonstration Project" (A National Science Support Project) under Grant No. 2008BAH29B06 (P. R. China).

References

1. Witten, I., Moffat, A.: Managing Gigabytes, pp. 20–30. Morgan Kaufumann Publishers, San Francisco (1999)
2. European Telecommunications Standards Institute. ETSI GTS GSM, pp.10–12 (May 2007)
3. MC35i Siemins Celluar Engines Hardware Interface Description. SIEMINS
4. MC35i Siemins Celluar Engines AT Command SET. SIEMINS
5. Zhang, Q.M.: Wireless Data Transmission System Based on GSM. Science Mosaic, 32–34 (2008)
6. Lu, S., Li, Z.D.: Short Messages Receiving-sending System Based on AT89S52 and GSM Technology. Journal of Shanxi Agricultural University (Natural Science Edition) 29, 461–465 (2009)

The Design of Deep Web Search Engine Based on Domain Knowledge

Deng Rong, Wang Hao, and Zhou Xin

School of Software,
Jiangxi Normal University,
Jiangxi, P.R. China
chdr_321@163.com

Abstract. A new method of search data on Deep Web is proposed in this paper based on the concept of domain knowledge and the feature of Deep Web data. First they obtain the search interfaces on the Deep Web; and then do feature analysis and domain judge on them; at last, they classify and integrate various interfaces in line with diverse domains. This method has showed up higher degree of correlation of domains from the seek result, via testing several different domains. Compared with other methods, its recall and precision are also satisfactory.

Keywords: Deep Web, domain knowledge, interface, feature analysis, recall, precision.

1 Introduction

Along with the rapid development of Internet, the data on the Internet is becoming more and more, of these data, some may be achieved by traditional search engines, others need for getting through query interfaces. These dynamic pages hidden in the search interfaces are called Deep Web or Hidden Web [1]. According to BrightPlant's statistics, the storage content of data on Deep Web is of 450 to 550 times on the Surface Web, the total number is more than 550 billions[1]. In 2004 a comprehensive estimation was done in UIUC university, inferred that there are 307000 websites providing Web database, 450 Web databases, the a mount of information exceed 200000TB [2], and the statistic has been increasing every year. So, now you pay more and more attention to the research on Deep Web. Therefore, people take more and more attentions to the research of Deep Web now.

In order to improve the recall and precision of Deep Web data, Deep Web search engine (hereinafter referred to as Deep SE) has been intensively studied. According to the present research, the method of obtaining Deep Web pages can be divided into two species: centralized search engine and meta-search engine. The theory of centralized search engine is automatically filling query forms and downloading the pages, and then setting up index for user to inquire. HiWE (Hidden Web Exposer) is the representative[3]. HiWE designed a kind of Crawler which can extract the content on Deep Web. It extracts forms from pages using form processor, then select data to fill in

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 315–321.
springerlink.com

forms automatically from prepared dataset, and then submits the composite URL to Crawler for downloading pages. But HiWE can be only used in specific domain, and also need artificial help. The main idea of the meta-search engine is choose the most relevant sites to client, the representative is the method proposed by G.I peirotis Panagiotis that automatically classify the backend database connected to Web page[4]. First a set of Classifier based on rules are produced using machine learning techniques, and then the Classifier is transformed into query URL. Then, query the backend database and calculate the query structure. Finally, classify the database according to query structure. But this method has poor instantaneity and stability, and its research is aimed at classification of text database, while a large number of content on Deep Web is non-text data [1].

To sum up, a very good method that can obtain the contents of Deep Web is still not appeared till now. A new Deep search engine is proposed in this paper against this background, which is to gain Deep Web content through integrating Deep Web querying interfaces on various fields. We classify these query interfaces according to related domain knowledge, and then generate a generic inquire interface. Experiments show that the results obtained by this method have better domain relativity.

2 Description of Deep Web Search Engine System Based on Domain Knowledge

To obtain the content on Deep Web, users need first to find the inquire interfaces, and then enter keywords through inquiry forms submission to the backend database, and finally the backend database generate dynamic pages to the user. If users need to search a certain information through multi-web pages on the same area, the only way is submitting keywords one by one and then doing comparing and choice in different result pages, which is time-consuming. In this paper a new Deep search engine is proposed, which automatically finds the Deep Web pages containing query forms in different fields (that are querying interfaces); then does feature analysis on these query interfaces; meanwhile, does classification and polymerization on them according to the relevant domain knowledge, and stores them into the interface base; finally, creates a generic quire interface, through which user can input query keywords. As judgment according to specific domain knowledge is done in the process, we can simply choose interface on the certain field to submit queries. Therefore, the result page should be relevant to the certain domain. Figure one shows the integral structure of our system.

Deep SE system includes three main stages: (1) automatically find web pages containing query forms from the Deep Web; (2) analyze the features and attributes of the pages, and do classification and polymerization to the interfaces combining with domain knowledge, (3) establish a general user quire interface.

In the next section, the process of the system is stated in detail, and the simulated experiment result of the system is given too, and the further research in the future is also declared.

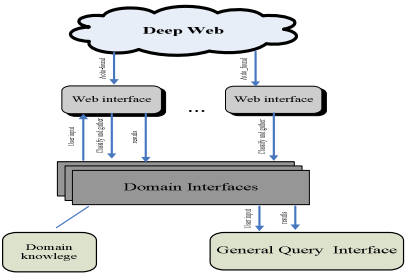


Fig. 1. Overview of Deep SE

3 Discovery of the Deep Web Query Interface Automatically

Deep Web query interface is commonly appeared in the form of Forms, and emerged in HTML, but not all of the forms are Deep Web query form, such as interfaces for user registering or login, and BBs forum for discussion, E-mail and so on. Thus we need to distinguish Deep Web query forms from other forms on these pages. Some text classification algorithms researched comparatively mature can be used for classification of the forms, C4.5 decision tree, simple Bayesian algorithm, KNN algorithm and SVM algorithm, for instance. Among these algorithms, C4.5 decision tree and naive Bayesian method are used more than others[6]. In this system, C4.5 decision tree is adopted.

We use structure characteristics of Deep Web query interface as the judgment criterion. Although the forms of expression or content of interfaces are different, they still have some common architectural features. These forms are expressed as <Form></Form> in HTML, including some attribute name tags, such as Form Name, Input Control Name, Value of Control, Request Path, and so on. Detailed comparison and analysis to the querying interfaces and other interfaces in Deep Web are done in article [6] and [7], and they have some conclusion, from which we mainly select Password Controls, Text Area, Text Boxes, Dropdown Lists, the number of Radio Buttons and Check boxes, and the priority level of these attributes. And C4.5 algorithm is adopted to extract Deep Web query interfaces. The decision tree is shown in figure 2.

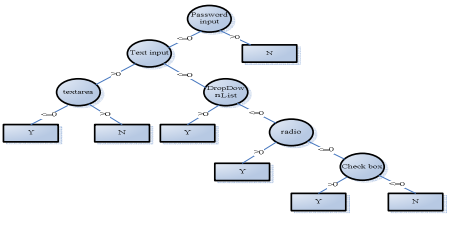


Fig. 2. Decision Tree of Form Classification

In the decision tree, the number of the attributes is judged, less than zero means not found, while greater than zero means there are one or more. Rectangles denote the class, Y means that is a Deep Web query interface, while N is opposed. These nonquery interfaces include registration forms, user login forms, message or complaint forms and empty forms.

4 Clustering of the Deep Web Query Interfaces Based on the Related Domain Knowledge

In the network Deep Web and query interface are diverse and sundry. In order to ensure users query information from multiple Deep Webs in the same field, and compare and analysis conveniently, we classify and integrate various querying interfaces according to the field. In this process, the concept of knowledge is introduced. Knowledge base contains the facts about the data in a certain field, and is self learning and expanding gradually in the process of use [5].

The integration of Deep Web query interface can be ascribed to integration of the feature attribute of the form. This process can be divided into three steps: (1) extract query interface characteristics, (2) judge the field of the interface using knowledge; (3) classify and integrate different interfaces according to the field.

4.1 Extract Query Interface Characteristics

Text messages and spatial information in the form can often provide many information of the form, for example, in the form label the value of form name attribute generally has certain significance; the name attribute value of controls, the default value of lists and the attribute value of drop-down lists in the form can all be extracted as interface features, including the important URL values of pages with the forms and that of the submission pages.

We use Literals{〈L1，k1〉,〈L2，k2〉,...,〈Ln，kn〉} expression to record the text information appeared in the form, 〈Li，ki〉 indicate that character Li has appeared ki times. We use Controls {C1, C2,... Cn} expression to record information of each control in the form, Ci = (name, type,...) record the control information, including name attribute value, control type, the default value, the list items and so on. Then we can express every inquiry Form as F (URL, Name, Literals, Controls), URL is the URL value of the submitted page from the form, and Name is the Name attribute value of the form. F expression is defined as the feature expression of the query interface, and the following work is to analyze these feature expressions.

4.2 Judge the Field of the Interface Using Knowledge

After the feature expressions of all forms are obtained, we analyze these feature expressions in order to judge their fields. The values of Name, Literals and Controls extracted in previous part are all important information, which can provide the field feature of interface. For example if it is a interface page selling computers, the words

like CPU version, size of hard disk and memory are generally appeared; if it is a interface page selling clothes, then the words like color, size, long and short are generally appeared. Therefore, we can judge the domain of the interface according to the knowledge base.

Web form itself is a kind of model that cannot be described as hierarchic, after formalized representation it can be presented as a tree, so match work to the forms and domain knowledge base can be done. Every web form feature attribute is given a weight based on some factors, including name similarity, synonyms, hypernymy, structure and the data similarity in feature expression of the form, which stands for the matching similarity to the concept of the specific field knowledge. Then we can judge the field of the interface according to a given threshold. We use I {URL, Name, D} expression to record the domain of each interface, URL record the URL of page submitted from inquires form, and Name is the Name attribute value of the querying interface.

4.3 Classify and Integrate Different Interfaces According to the Field

After obtaining the domain of each query interface, we classify and integrate the interfaces according to the different domains. D{Domain, $<I_1,URL_1>$, $<I_2,URL_2>$,…, $<I_n,URL_n>$} expression is used to integrate different domain interfaces.

5 Establishment of a General User Inquires Interface

The general user query interface is a unified interface for user to input query information. It is a common interface of all web forms where user can input information. The user unified interface we established is a three-layer model. The lowest layer is the query interfaces on the web, the second is the unified interface of each domain, and the top is the universal interface for users' input.

In the process of matching the interface feature and the knowledge base suggested earlier, we also integrate a general logic feature for the specific logic properties in different interfaces of the same field. We set up a mapping of the concept of knowledge and the attributes which are semantically similar or identical, then integrate the unified interface for each field according to these concepts (general logic properties).

After the user input inquired keywords in the general interface, our system analyzes the keywords and judge the domain with the knowledge base; then the unified interface of such domain is automatically selected; finally, search each form and return result pages to the user.

6 Experimental Results and Analysis

In order to test the validity of the deep search method we suggested, we choose three fields to analyze the test results, daily uses, computers and bags. We extract result pages given by the system, and then judge the domain relevance. Recall and precision is

an important judgment standard for search engine[8][9], so we analyze the average recall and precision of the search result compared with other two kinds of search engine. In the experiment, we analyze the search result pages artificially, and put this manual handing as the benchmark for comparison to analyze and evaluate the three methods effectively. Table 1 shows the statistics of the total number of pages on three fields obtained through our method and the number of pages that belong to the related domain. Table 2 shows the comparison of the recall and precision of the result pages through three different methods.

Table 1. Number of pages on three domains

Domain	Number of Pages	Number of Domain Pages
Daily-uses	51980	49381
Computers	31351	30129
Bags	9855	9760

Table 2. Recall and Precision of three search engines

	Deep SE	WIWE	SemaForm
Recall(%)	92.68	92.80	92.23
Precision(%)	94.32	90.79	94.08

From the experiment data, we can see that the domain relevance of the pages obtained through our search method is very high, and the precision is relatively high compared with the other two methods. The reason is that in the searching process, the query interfaces on Deep Webs are domain classified using domain knowledge base, so the query result has high field relevance and precision. Nevertheless, it might have data omitted in the process of extracting the feature of the interfaces and the process of classifying the fields, the recall of the search results does not make a great change, and this is our main research direction in the future work.

7 Conclusion

Deep web data extraction is still a research field that waiting for continual improvement and perfection. A new search engine based on the domain knowledge is proposed in this paper, and the main theory of the search engine is to classify and aggregate the query interfaces on Deep Webs by the fields. Via comparison and analysis with other two search methods, the experiment has proved the validity of this method, but still needs to be improved.

Acknowledgment. We thank the syndics for carefully proofreading the manuscript and the originate organization. We also acknowledge helpful discussions and comments with classmates, Zhu Kunhong and etc, especially the tutors.

References

1. Bergman, M.K.: The deep Web: Surfacing hidden value. White Paper on the Deep Web (2001), http://www.brightplanet.com/pdf/deepwebwhitepaper.pdf
2. Chang, K.C.-C., He, B., Li, C., et al.: Structured database on the web: Observations and Implications. SIGMOD Record 33(3), 61–70 (2004)
3. Raghavan, S., Careia–Molina, H.: Crawling the hidden web. In: Proc. of the lnternational Conference on Vary Large Data Bases (VLDB), Rome, Italy (2001-2009)
4. Ipeirotis, P.G., Gravano, L., Sahami, M.: Probe, Count, and Classify: Categorizing Hidden Web Databases. In: ACM SIGMODZ 2001, Santa Bathara, Califomia, USA, May21-24 (2001)
5. Walny, J., Barbosa, D.: SemaForm: Semantic Wrapper Generation for Querying Deep Web Data Sources. In: 2009 International Conference on Web Information Systems and Mining (2009)
6. Wang, H., Liu, X., Zuo, W.: Using classifiers to find domain specific online databases automatically. Journal of Software 19(2), 246–256 (2008)
7. Barbosa, L., Freire, J.: Combining classifiers to identify online databases. In: Proceedings of the 16th International Conference on World Wide Web, pp. 431–440. ACM Press, New York (2007)
8. Lu, Y.Y., He, H., Zhao, H.K., Meng, W.Y., Yu, C.: Annotating structured data of the deep Web. In: Proc. of the IEEE 23rd Int'l Conf. on Data Engineering, pp. 376–385. IEEE Computer Society Press, Istanbul (2007)
9. He, H., Meng, W.Y., Lu, Y.Y., Yu, C., Wu, Z.: Towards deeper understanding of the search interfaces of the deep Web. World Wide Web 10(2), 133–155 (2007)

Based on Multi-Agent Systems (MAS) of the Prototype Selection System of Virtual Enterprise Partner

Ke Su[1] and WeiZhou Song[2]

[1] Personnel Department
Chongqing College of Electronic Engineering
Chongqing China 401331
[2] Software Engineering Department
Chongqing College of Electronic Engineering
Chongqing China 401331
{bart.simpson,homer.simpson}@uspringfield.edu,
monkey.king@uhuaguoshan.edu.cn

Abstract. With the development of computer technology and informationization, the progress of the multi-agent system of the virtual enterprise (MAS) selecting research significance. Based on multi-agent system (MAS) virtual information platform construction enterprises, and develop the virtual enterprise partner selection, supply chain management more experimental prototype program, Agent system through practice has certain theoretical significance and practical value.

Keywords: Multi-agent Systems, Virtual Enterprise Partner, Prototype Selection System.

1 Introduction

In recent years, intelligent Agent oriented Agent system technology has become the artificial intelligence and computer science is one of the fastest growing subject. General Agent (Agent) is a kind of in certain circumstances, with goal-driven, autonomy, responsiveness, initiative, sexual intercourse, in order to achieve the objective characteristics, in particular environment can be flexibly, independent activities. But more Agent system (MAS) tissue or society, can be very natural to solve distributed problem solving and cooperation, coordination of similar problems for human society. In many ways, including obtained user interface, personal assistant, mobile computing, information filtering, check and electronic market, telecommunications, network management and computer work, etc, to support the multi-agent system (MAS) of the virtual enterprise partner selection is of great significance.

2 The Concept of Virtual Enterprises

The virtual enterprise is the late 20th century enterprise theory and practice of the latest development. In today's changing market environment, in order to maintain and

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 323–328.
springerlink.com

strengthen the competitive power of enterprises, more and more application of virtual enterprise dedicated to the new organization mode, they put the value chain, and related virtual cooperation of external enterprises establish concentrated, extensive, flexible interactive cooperation relations, as a major strategic countermeasures, as some companies from industrial economy to the knowledge economy transition form competitive action.

The main characteristics of the virtual organization is independent, geographical scattered, with overlapping network task. All members of the core competitiveness of the respective contributions, and cooperation based on stable relationships. The virtual organizations provide products and services rely on innovation and good relationship with customer.

3 The Main Characteristics of Agile Virtual Enterprises

3.1 Agility

Agility, it is the most basic of agile virtual enterprises. From the perspective of organization, agility is to show enterprise organization unit reconfigurability, reusability, scope, namely RRS controllability properties, From the Angle of management, agility is refers to the information connectivity, cross enterprise participation, production flexibility and management. From "virtual enterprise" in the word "dynamic" literally KeTi also shows the virtual enterprise agility. Enterprise established cooperative relationships, different partners can quickly, to bid a new contract. The cooperative enterprises by qualified for the virtual enterprise can be formed with great speed and decomposition. Therefore, agility is about different enterprises core competitiveness, through the combination of dynamic that virtual enterprises with outstanding, flexible, in order to adapt to the rapid changes of the market. And this flexibility should throughout all aspects of the business cooperation process, not just the production process. It is through the core competence of enterprises on different, and select the optimal, can make the virtual enterprise to create "star" ability. Through the integration of different ability, can achieve individual enterprise to reach the effect. For the international competition is indispensable. Agile virtual enterprise with strong adaptability, it has no fixed mode, no fixed size, fixed combinations, the partners may be less than, the cooperation of time can be long or short, but according to the needs of the market, a can meet the needs of the market, once to complete the task, organize immediate disintegration. This makes the agile virtual enterprise to market changes made rapid reaction.

3.2 Virtual

"Virtual" means the existing looks in reality is not really exist. An enterprise looks like traditional enterprise with the equipment, but not really have them, just can use these devices, like oneself really have. Through computer networks, people can work with others, equipment, design tools, software generally together, even if they are in different position, belong to different owners, they can also close cooperation. An

enterprise does not need formal employ many needs talents, also need not have the necessary equipment factory, it can choose and use external resources, to any part of the activity. In extreme cases, the enterprise may be a shell, it fully using the external resources complete product design, production, product marketing planning, marketing, and accounting. Failure orders, As long as the enterprises in the rapidly changing market has a very good idea, it can do so. Facing the opportunity, if the enterprise is still out to build your own design, production, marketing and sales ability, it will be missed opportunities. The only way will seize the opportunity of make full use of the existing equipment. Of course, in reality, the absolute pure "virtual" or "virtual" enterprise is rare, most companies have more or less "virtual" elements.

3.3 Organizational Structure of Network and Flat

Because of using computer and information superhighway "as a representative of the enterprise information model of virtual organizations, gradually formed in the members with other enterprise information network, to make the virtual organization becomes a business network. In this network, each enterprise should contribute some resources sharing, and other enterprises and the virtual organization this net energy and greater than all the energy contributor. Modern enterprise organization theory, the enterprise organization level structure formation is the root cause of effective management of amplitude limit, namely when the organization expanded to a certain extent, must by increasing the management level to ensure effective leadership. But because E-mail, voice mail, Shared database etc. With the development of information technology, middle managers supervise others -- the function, collection, analysis, evaluation and spread of information organization fluctuation layer - largely replaced. Executives can direct communication between the lower and middle and down, thus, the role of the virtual organization of virtual leadership can directly or indirectly to command the virtual team, shorten the distance between the corporate hierarchy, thereby agile virtual enterprise organization presents a "flat" type. This is agility is reflected in the organizational characteristics.

Generally speaking, agile virtual enterprise not only make every member enterprise cost a relatively short period of time, low cost, low risk, and makes the enterprise to provide products and services, and to strengthen enterprise and customer relationship.

4 Constructing Virtual Enterprise Information Platform and the Prototype System

The virtual enterprise is a quick response to market changes, rapid acquisition, share market opportunity of business organization. It has the power, dynamic, initiative, etc. This requires the establishment of virtual enterprise characteristics of information platform. This information platform must have speed, dynamic and heterogeneous, intelligent and socialization. Traditional MIS system cannot achieve this request. Far While many Agent system (MAS) just have independent, flexible, coordination,

dynamic fast problem solving ability. The Agent system (MAS) theory to guide the virtual enterprise information platform construction work. And the development of a virtual enterprise partner selection, supply chain management more experimental prototype program, Agent system has achieved good results prove the Agent system, the feasibility of the application of virtual enterprise and the superiority.

A prototype system includes two kinds of Agent: management Agent, enterprise Agent. Management Agent represents a virtual enterprise alliance, it manages in the league all enterprise Agent. The virtual enterprise information platform of a specified in the computer. Management of enterprise registration Agent accepts Agent, registration, partner, and information to customer orders issued to the ability of the core enterprise. Enterprise Agent is a representative of one enterprise. It has certain aspects of the knowledge, can make in the social environment for the enterprise.

The results indicate that the prototype system operation process, high enough information security guarantee, under the situation of virtual enterprise can rely on many Agent system, complete the order at an alarming rate, and production orders. Can a company make imaginary in the night, the enterprise Agent stationed market opportunities, if capture favorable orders can be immediately generated production plan. Then will the submission to the integrated automation system, into the workshop production. If you have not yet reached such information and safety degree of reassuring, also can be done completely supply chain process simulation and contract pre-populated. This kind of situation: at night, the next morning, the garrison Agent can get a general contract. After the examination, the use of digital signature method to generate the contract sent, submit the production department.

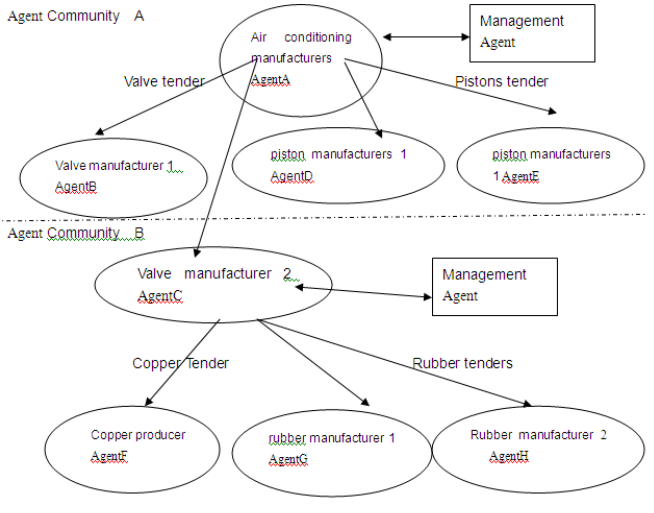


Fig. 1. Multi Agent System using the expression of virtual enterprise

Figure 1 know this much Agent model can support (orders) decomposition layers of the whole supply chain, the complete set (simulated).

According to the structure can be divided into active Agent with reaction type. Active Agent for rational establishes a set of formal justice, similar to person's mental state. Then from the state of mind of justice, according to logical reasoning of decision. But to the actual programming abstractions intelligent model. While using the stimulation reaction type, middle - response behavior does not need logical expression and reasoning, high performance. Our prototype system is a hybrid, reactive, supplemented by certain reasoning. Petri nets, knowledge expression for asynchronous events triggered by reasoning machine (similar to the finite automaton) action. And layered structure in society, said Agent on the layer and layer knowledge expression and reasoning machine. Detailed structure figure 2.

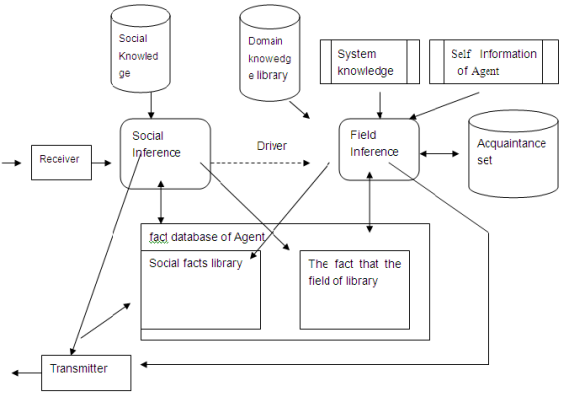


Fig. 2. Implementation of Prototype System structure

Figure 3 shows the underlying structure of communication:

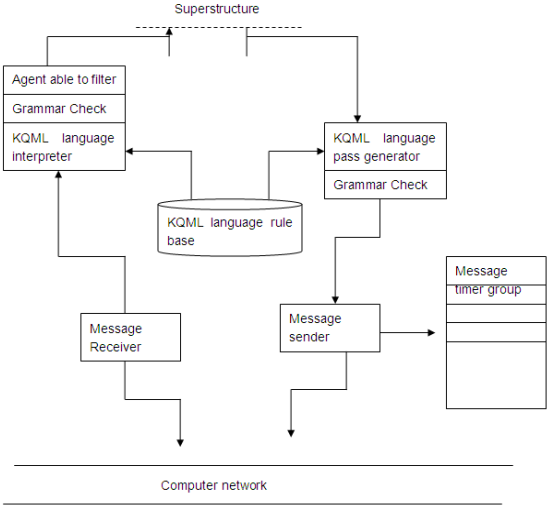


Fig. 3. The underlying structure of communication

References

1. Ii, W., Yang, C.: Extension Information—Knowledge Strategy System for Semantic Interoperability. Journal of Computers 3(8), 32–39 (2008)
2. JADE Home page, Java Agent Development Framework [EB/OL](July 03, 2009), http://jade.tilab.com
3. Ocon, J., Rivero, E., Strippoli, L.: Agents for Space Operations. In: Proc. of AIAA 2008, TresCantos, Spain (2008)
4. Nwana, H., Lee, L., Jennings, N.R.: Coordinationin Software Agent Systems. BT Technology Journal 14(4), 79–88 (2006)

Application of Information Management System in the Marine Environment Monitoring Laboratory

Xiaohui Gao[1], Yanmin Zeng[2], and Xiangyu Zhao[3]

[1] North China Sea Environment Monitoring Center, SOA
Qingdao, Shandong Province, China
[2] BEIHAI Marine Engineering Prospecting Institute, SOA
Qingdao, Shandong Province, China
[3] North Sea Marine Forecast Center of SOA
Qingdao, Shandong Province, China
gxh777@sina.com, ym.zeng@163.com

Abstract. To meet this requirement of Marine environment monitoring business, designing and developing the Marine environment monitoring laboratory information management system. It is mainly used for environmental monitoring of business process management, resources and data of automation management. The acceptance, distribution, analysis, original record, monitoring report, data verifying, sign and issue report for the monitoring sample were realized automatic circulation of whole journey. With the management system which can improve the environmental monitoring work efficiency, the quality management standard, we can get better for environmental management service.

Keywords: Laboratory management, automation, System.

1 Introduction

With the progress of science and technology, people testing laboratory analysis have become increasingly demanding, internationally relevant laboratories have begun to network management direction. As the number of samples, monitoring the project dramatically increased the accuracy of the data requirements more stringent, the sample analysis cycle is getting shorter, man-made general management has not adapted to the development needs. North China Sea environment Monitoring Center through the laboratory information management and information needs analysis, according to the actual situation of laboratory information management system and program design, the structure of the system, the system should have the function described in order to establish a consistent laboratory conditions, with security improved user management [1], a reasonable experimental procedure and experimental examination of experimental data to facilitate efficient management of laboratory information management system to meet the laboratory requirements for advanced scientific research.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 329–334.
springerlink.com

2 System Design

2.1 Construction of Target System

1) Effective use of manpower and equipment resources, reduce staff stress. LIMS system combines business process management, query statistics, testing resources management, quality assurance and control multiple modules into one, to run the laboratory to reach automation, information management and paperless office purposes.

2) Standard laboratory business processes to achieve management of instrument calibration, standard material management, personnel management and inventory management appointment card, so that the entire management system to run more specifications.

3) Improve data accuracy. Using LIMS system to avoid manual copying and calculation error, the data can provide accurate results quickly [2].

4) Data security, centralized storage and sharing. Electronic safe, centrally store all laboratory data, and to provide efficient, flexible data query capabilities.

5) Automatic calculation and automatic report generation, analysis of decision making. Through the LIMS system can generate a large number of reports, through an intuitive graphical report shows data results, can be more convenient analysis for environmental management decision-making support.

2.2 Structural Design

Laboratory Information Management System functional design and laboratory procedures involved in the work related resources, should have the equipment, personnel, business processes and other management functions. System architecture must be designed to monitor business management, monitoring of resource management, system management and query statistics, several modules. Detection of resource management module which is divided into modules and materials, equipment management, reagents and laboratory personnel information management module.

Laboratory Information Management System Function Photos Figure 1.

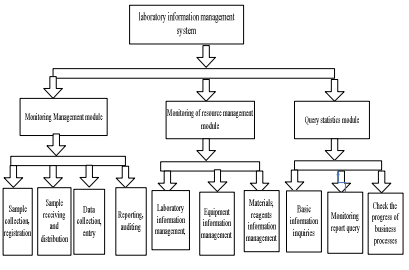


Fig. 1. Design of system structure

3 System Implementation

Laboratory Information Management System can be divided into platform software and applications based on major subsystems. In which the basic software from the mature form of commercial software platforms, including Microsoft.Net technologies, as well as SQL Server, database. Application infrastructure software platform provides the infrastructure needed for subsystem (Infrastructure); application subsystem software modules for business, responsible for completing and testing related to specific business functions. In the development platform on the basis of the convention by the rapid development [3], implementation and North China Sea environment Monitoring Center need to match of applications, including monitoring of business management, monitoring resource management, query statistics, system management.

3.1 Monitoring Management Module

1) Sample collection

The system will request information related to the sample as a delegate tasks assigned to the transfer of authority with the monitoring sample collection department. After receiving the task of monitoring the sector and generate a list of sample print sample container labels, sample collection began.

2) Sample registration

Registration can be automatically imported samples sampling personnel registration information, timely analysis of samples of the task flow to have permission of department. Customers can also automatically import information from on-line registration submission, approval from EXCEL to import sample information form. Shown in Figure 2.

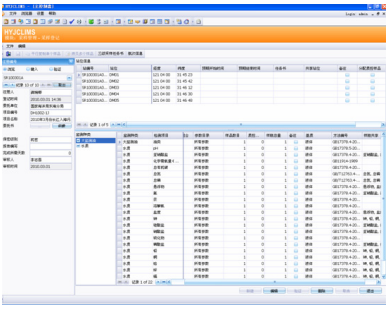


Fig. 2. Sampling Registration

3) Accept the samples to determine whether it is normal

Analysis of the department through the system to accept the task, and sample information to judge whether it is normal. Abnormal returns monitoring department to re-register, then sign it and accept the normal sample.

4) Sample control flow

After receiving the sample of departments, according to different analysis requirements will be distributed to different laboratories analyzing samples of staff, and adjusted according to the appropriate allocation of priorities. Analysis of staff into the system to accept the task was conducted after the sample preparation and analysis, and quality control measures taken.

5) Sample analysis data entry

After the completion of sample analysis will produce large amounts of data, including automatic analysis instrument data system and instrument interface through direct acquisition, automatic import and save, without human intervention, manual analysis of other way to get input data from analysts. Shown in Figure 3.

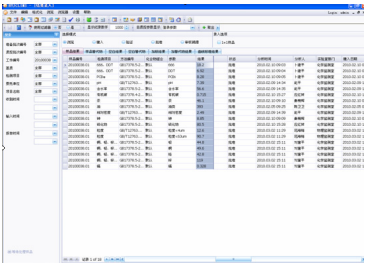


Fig. 3. Data Entry

6) Analysis of data validation

Sample analyzing data needed to analyze audit data for a number of exceptions, error data verified by the system need to review and return the relevant analysts reported change. For quality control data, the system provides for quality control verification, to ensure that analysis of the data is correct [4].

7) Analysis report preparation and examination

After completion of data validation, the system automatically generates an analysis and review of electronic signatures and electronic archiving of all business processes end of the sample.

8) To provide reports

Finally the system will automatically process for the user and the various departments involved in various types of feedback statements, including the completion of relevant information and data, to provide an analysis of decision-making leading statements.

3.2 Monitoring of Resource Management Module

Monitoring of resource management module, including laboratory information management, information management equipment and materials, reagents information management module and so on.

1) Laboratory personnel information module

This module is mainly used for management of laboratory personnel's name, gender, occupation and other basic information. Archives of personal technology

management, including papers, monographs, issues and results, training and certification, as well as continuing education situation in statistics.

2) Equipment information management module

The module is laboratory information management an important content of the module is divided into six sub-function module: equipment management, low product management, consumables management, equipment loan management, the storage management, equipment maintenance management [5].

(a) Equipment management, the functional modules for laboratory management and equipment management executives. Users can access a particular equipment from the storage location, purchase date and the manufacturers, but also in some kind of statistics can be retrieved and the total number of instruments and equipment, the total amount and report output. Consumable materials management and product management of low-value function and this module similar;

(b) Equipment loan management, the function module is designed for laboratory management. If a time to lend an instrument who, for what purposes, just click the instrument on the management, to retrieve the field that;

(c) Out of the warehouse management function module is designed for the equipment officers. Corresponding data table field: instrument number, instrument name, equipment models, unit of measurement, quantity, unit price, manufacturer, date out of storage, access directions, department name, lab name, handling people, notes, examination marks. Equipment such as a library in a certain time the amount of storage and so on, can easily learn from;

(d) Equipment maintenance management module is designed for laboratory management. It is composed of two small module: equipment maintenance and equipment calibration. Equipment maintenance module helps managers to remember each laboratory equipment maintenance history, easy to maintain it better in the future. Instrument calibration laboratory management module will help calibrate instruments on time to ensure the accuracy of measurement instruments.

3) Materials, reagents information management module

This module is mainly used for management of laboratory reagents subscription management personnel, materials, reagents and reagent materials suits to entry.

3.3 Query Statistics Module

The module can carry out the resource center of the query statistics. The resources include the monitoring of personnel, monitoring methods, monitoring projects, health standards, equipment, reagents, standard goods etc. For example, can be checked back to preparation, review, approve the report at different stages of testing. Query to the sample information, original records, the resulting data, evaluation results, final report. For quality control manual and procedure documents, work instructions and other document management, query. Sample properties can be supported by the project, the passing rate, workload query statistics. Equipment utilization statistics can be carried out inquiries. Management may be the work of various departments summary query. Can be found within a period of time the department's workload, including the error check, and print related reports required functions.

3.4 System Management Functions

The modules include database management, system access control module. Database management module can be conducted off-site data backup on a regular basis, with data recovery [6]. System access control module has strict access control. System permissions to each user interface segment, and each function button for each person, and each detail of each application to set permissions. Rights management is logical and assigned a permission to add, remove, as amended, to query the permission settings. Data will be automatically locked after the completion step in front of the data modification operation, only the first step can be carried out behind the operation is complete, data manipulation, I can only modify my registration information.

4 Conclusion

At present, China's environmental protection work has entered a new historical period, the environmental monitoring business has developed very rapidly. The LIMS lab as a modern management tool to improve the efficiency of environmental monitoring, standardized quality control, ensuring data reliability and improve environmental monitoring of the effectiveness of environmental management services. Although LIMS environmental monitoring laboratory in advance is still difficult, but in view of its potential applications, LIMS will be prompted China to enhance environmental monitoring laboratory management, integration and market competition, and international practice, and science the importance of integration of management systems and tools.

References

1. Wang, X., Fu, Q.: Establishment of LIMS For Environmental Monitoring. The Administration and Technique of Environmental Monitoring 19(4), 4–8 (2007)
2. Shen, Y.: Construction and Implementation on Information Management System of Environmental Monitoring Laboratory. The Administration and Technique of Environmental Monitoring 18(4), 4–6 (2008)
3. Shang, F., Wang, Z.: The designing and realization of LIMS. Environmental Monitoring in China 16(4), 1–2 (2000)
4. Tang, L., Li, W.: The Study of Application of Laboratory Information Management Systems (LIMS) in Environmental Monitoring Centr. Jiangsu Environmental Science and Technology 20(2), 69–71 (2007)
5. Ge, Z., Sheng, L.: Design and Realization of a Laboratory Information Management Systems. Computer Teaching and Education Information, 705–707 (2008)
6. Wu, Z., Lou, Z.: Development and Application of Information Management Systems in the Environmental Monitoring Laboratory. Chemical Analysis and Meterage 15(5), 50–53 (2006)

BGA Device Detection System Based on Frame Integral Reducing Noise Method

Wen Zhang and Xin Long Zhang

Engineering & Technical College, Neijiang Teachers College, Neijiang, Sichuan, China
njsywlx@163.com, zhxinlong@126.com

Abstract. For the digital imaging system of solder joint defect detection of BGA device in the printed circuit board, the paper calculate the optimal imaging results by analyzing the effect of the ray source focal size on the imaging results. The superposition display of captured image achieves good real-time effects. By testing, the systemic resolution of optimal magnification achieves 17lp/mm, the minimum resolution of defect analysis accuracy achieves 0.03mm, and show the result picture of collection test.

Keywords: Frame integral Reducing Noise method, BGA, X-ray.

1 Introduction

At present, the printed circuit board develops along the trend of miniaturization and high density. BGA packaged device (Ball Grid Array Package) has been used extensively and it has the advantages of multi-pin and low capacitance and inductance between leads. It is difficult to test the defects of the solder joints because that the solder joints of BGA printed device hide below the devices and it adds the cost in excluding defects and repairing. Fault detection and diagnosis has been a very key problem in the process of PCB manufacturing for all countries. AXI has been the main way to solve the defect problem for invisible solder joint. The paper introduces the X-ray detecting system for printed circuit board and it realizes the function of defect detecting.

2 System Hardware Configuration

The general design of ray detecting machine is as shown in Figure 1. The system is mainly composed of three parts, ray resource controlling part, mechanical controlling part, image acquisition and transmission part. The inner parts of equipment consist of micro-focus X-ray source, 3D moving platform, X-ray image intensifier, image collecting card, CCD industrial camera. The external parts are mechanical controlling console and main controlling computer. It controls the moving state of 3D platform by mechanical controlling console. The main controlling computer can control the X-ray source parameters that are tube voltage, tube current and the state of switch, the moving state of 3D platform, the working state of collecting card.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 335–340.
springerlink.com

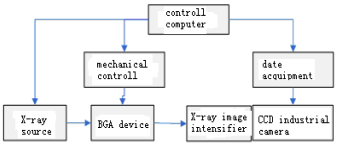


Fig. 1. Graphics of X-ray detecting system

The key equipments are ray source controlling part and image acquisition part. The ray source is inserted to the main controlling computer by serial port RS-232.

The main controlling computer controls the switch of ray source by software. Acquisition part adopts LVDS digital output of 12-bit black\white camera. Sampling card connects to the computer by 32-bit/3MHz PIC bus and transmits the images to the computer to de-noise and be real-time display.

3 The Analysis of System Performance

X-ray digital imaging technology has some parameters to evaluate the performance indices that usually influence the practical key. The system resolution is the most important index among multi-property indices and the system synthesis resolution is affected by every sub-system performance. The effect by the size of ray source focus and the best magnification of the system on the resolution is performed.

A. The size of X-ray source focus

The size of focus is the most important factor and it not only influences the size of detected geometry dimensions but also restricts resolution and image definition. The geometry blurring Ug:

$$Ug = f \bullet \frac{D2}{D1} \tag{1}$$

The size of ray source focus is f, D1 is the distance between ray source and object and D2 is the distance from the object to the intensifier. From equation (1), the way to improve the image quality is to adopt the smaller focus. The geometry magnification.

$$M = \frac{D_1 + D_2}{D_1} \tag{2}$$

From equation (1) and (2)

$$Ug = f(M - 1) \tag{3}$$

The geometry blurring is regarded as the function of magnification and the micro-focus ray source can get the bigger image magnification than usual. Figure 2, (a) and (b) finds the diagram of normal focus ray source imaging and micro-focus ray source imaging.

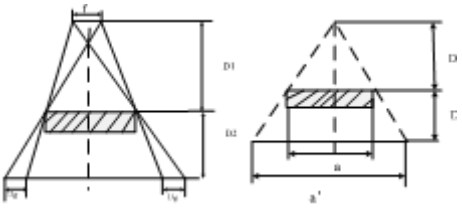


Fig. 2. The diagram of focus imaging (a) normal focus ray source (b) micro-focus ray source

B. minimal resolution

Now, the size of BGA device solder ball achieves 0.3mm, the minimal distance between solder balls achieves 0.5mm. According IPC-7095 standard, the minimal size of solder defect is the ten percent of solder ball. By the principle of resolution, as in Figure3, the size of ray source focus is f, the minimal size of distinguishable object is s, when A, B re-closure and can not discriminate object s. So the minimal resolution dimension of X-ray detecting machine should get 0.03mm (about 16.71p/m). If the ray source focus s=33μm, f=30μm. Figure4 is the curve of geometry magnification related to the minimal resolution.

The results tell us that M can distinguish the object of 30μm (16.7lp/mm) when the micro-focus source f is 0.033 mm under suitable magnification situation. M being equal to 2.14 achieves the limit R being 20lp/mm of resolution, the resolution is approximate to value16.51lp/mm.

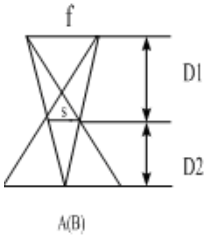


Fig. 3. The minimal resolution

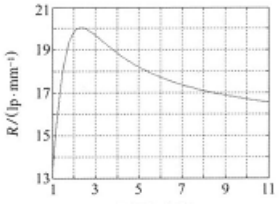


Fig. 4. The cure M related to s (f=33μm)

4 The Software Function and Superimpositon Sampling Algorithm

System software functions are composed of two modules, outside control and image process. The whole software is MFC programming based on VC++6.0. It realizes the control of outside devices by transferring the function SDK from sampling card, cameral and ray source. The main effect of image process module is the treatment and display of sampling image and it is the key part in system analysis.Image sampling is a important part in the system and the following describes frame integral reducing noise method in industrial application.

C. The noise analysis of ray image

In the system of ray detecting, the noise may be introduced in the image in the following steps that are the scattering of X-ray, the noise of image intensifier inverter, the instability of ray source, the noise of photoelectric unit such as CCD and the noise process of digital. Because of the low grey of X-ray image, some detail of the image is annihilated by noise. In order to improve the quality of image, some image enhancement ways are used to reduce the noise and improve the quality of the image. The normal methods of reducing noise are image smooth reducing noise, median filtering, superimposed noise reduction, frequency domain filtering reducing noise and so on. Superimpose noise reduction is

$$f(x,y) = S(x,y) + N(x,y) \tag{4}$$

Estimation error

$$\sigma_f^2 = E\left\{\left[\frac{1}{n}\sum_{i=1}^{n} N_i(x,y)\right]^2\right\} \tag{5}$$

$$\left[\frac{1}{n}\sum_{i=1}^{n} N_i(x,y)\right]^2 = \frac{1}{n}\sigma_N^2 \tag{6}$$

$$\sigma_f^2 = \frac{1}{n}\sigma_N^2 \tag{7}$$

Some frame image averages, the repetitive linear accumulation of the object free-noise image and the stochastic nose makes two accumulations. N frame image accumulates, the signal is N times magnification, the noise is $\sqrt{N}$ times magnification and the ratio of power to noise S/N reaches $\sqrt{N}$ times.

D. Frame integral superimposed reducing noise

The usual superimpose reducing noise is fit to the removal of additive noise when sampling, the computing amount of point by point computing is big, so adopting usual superimposed way is time-consuming and it is not suitable for the requirements of industrial sampling. The set of system improves the original superimposed algorithm and adopts multithreading technology programming avoiding unsuitable real-time request of original algorithm. Figure 5 represents improved software executing process.

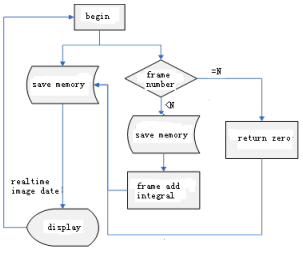


Fig. 5. The software flow graph of improving frame integral algorithm

5 Conclusion

Figure 6 represents the systematic resolution under different magnifications, the practical testing diagram as figure b shows that the systematic most resolution gets 17lp/mm that is increased to the requirement of BGA device defect detecting.

Figure 7 is the solder joint defect detecting result picture. The diagram effect improves obviously, the systematic performance evaluates and the designing result gets the detecting target that is suitable for the requirement of industrial application.

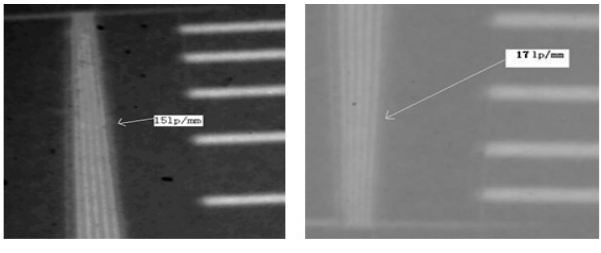


Fig. 6. Resolution detecting test-target (a) The systematic imaging resolution M≈3 (b) The systematic resolution under the best magnification (M=4.1)

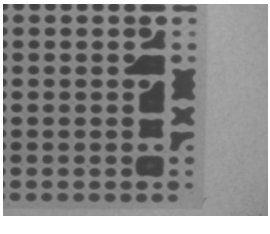


Fig. 7. The practical detecting effect

References

1. Wang, J., Tan, H., Huang, L., Xu, L.-B.: Design of Data Acquisition Transmission. Nuclear Electronics & Detection Technology 28(6), 1087–1090 (2008)
2. Zhang, Y.-Q., Wang, H.: A compact digital X-ray imaging system. Optics and Precision Engineering 16(4), 591–597 (2008)

3. Zhao, C.-Y., Zheng, Y.-G., Su, R.: Algorithm of Reducing Noise Based on Statistical Information of Multi-image. Journal of System Simulation 18(1), 383–384 (2006)
4. Kong, F.-Q.: Development and Application of the Microfocus Radiography. Nondestructive Testing 30(12), 931–933 (2008)
5. Hu, Y.-F.: X-Ray Detection for Soldered Joint Defects in BGA Package. Electronics Process Technology 26(6), 340–344 (2005)
6. Qu, Z.W., Wang, M.Q.: High Resolution Printed Circuit Board X-Ray Exmination System Nondestructive Testing 6, 438–439 (2010)

Design of Structure and Function of Spatial Database on Digital Basin

Quanguo Li[1,2] and Fang Miao[1]

[1] Key Lab of Earth Exploration & Information Techniques of Ministry of Education, Chengdu University of Technology, Chengdu 610059, China
[2] Department of Geography, Xiangfan University, Xiangfan 441053, China
quanguoli@163.com

Abstract. Spatial data of digital basin has many features, such as, Multi-source, different structure, large amount of data, etc. This paper analysis the content of digital basin and spatial database, devise its structure and function, for the basin study, protection and development services. It thinks the basic geographic data, thematic data, remote sensing image data, socio-economic data, hydrology and water resources data, meteorological climate data and metadata is its key components of the digital basin spatial database, spatial data of basin input, search, update, output, sharing and database protection is its important function.

Keywords: digital basin, spatial database, structure, function.

1 Introduction

With the development of digital earth and digital cities, advanced spatial information management technology has been applied to basin data storage and management, basin data of the different sources, standards, professions and formats has been managed and applied to improve the utilization of these basin information, and better services for basin research and resources development, they is the core of digital basin. Spatial data of digital basin has the general characteristics of huge system, that is the wide range of data sources, a huge number, and different data formats, describes the spatial objects of different data formats used in different data models, so data conversion can not accurately convey the original information, regularly cause loss of information, but these data of digital basin construction must be compatible and communication with each other, so the difficulties of building spatial database has serious impeded the development of digital basin. at the same time the service object of digital basin spatial database is a multi-level, multi-types of users, each service object has its own specific service needs [1]. Have to consider the inquiries demand of municipal authorities, professional sector and the social public, but also to meet different professional users demand of the management, evaluation, analysis, scientific research and planning. Hence the need to study for the structure and function of spatial database building of the digital basin, this paper devised structure and function of the digital basin spatial database with advanced spatial information management technology and the characteristics of the data for the basin, and

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 341–347.
springerlink.com

integrated multi-source, heterogeneous basin data, unified organization and management of massive data basin resources in order to realize the scientific and facilitation of basin information management for basin research, protection and development services.

2 Digital Basin and Spatial Database

Digital basin is a platform of information integration, model development, application display, that processing and analysis the geographical, hydrological, meteorological, environmental, ecological and socio-economic statistics data use of computers, network communications and spatial information technology, realization water resources, water environment, ecology management, and it is basin information tool. It can make up the defects in the existing basin, help solve existing problems basin, optimized basin construction, management and operation, and promote sustainable development of basin. Construction of digital basin is a complex giant system, can build the corresponding application modeling by the sub-system division, and distributed to different application servers to improve overall system performance.

Spatial database is the sum of geospatial data, it is stored and applied geospatial date by geographic information system in the computer physical storage media, and is organized by the structure of the document in the storage medium. Large amounts of data, high accessibility and complexity of spatial data model is the characteristics of spatial database digital basin. The composition of spatial database digital basin includes geographic data, thematic data, remote sensing image, socio-economic data, hydrology and water resources, climate, meteorological data and metadata, as shown in Figure 1; it should have the main function of basin spatial data input , output, update, maintain and sharing of basin spatial database, as shown in Figure 2.

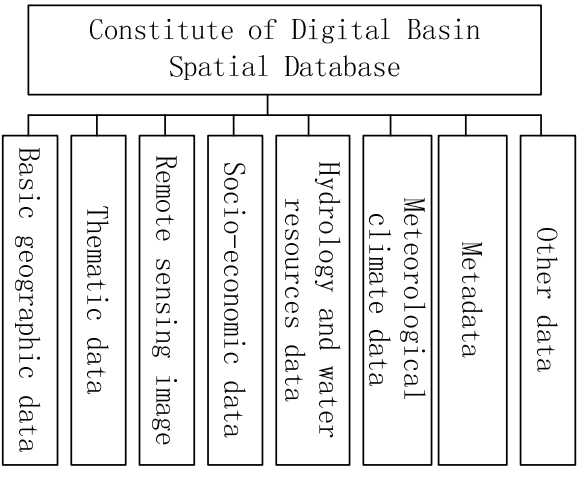


Fig. 1. Design of Digital Basin Spatial Database Constitute

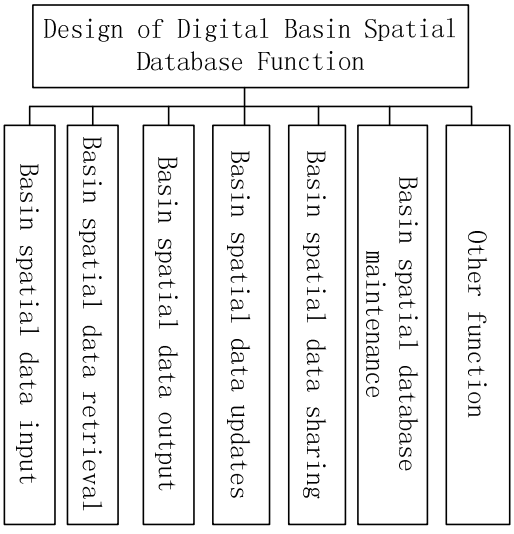


Fig. 2. Design of Digital Basin spatial database function

3 Structural Design of Digital Basin Spatial Database

A. Basin Basic Geographic Data

The mainly source of basic geographic information data is the mapping departments, it is the most universal, the greatest sharing demand, and almost can be used by all related industry with geographic information, as a unified basin basic geographic units of spatial orientation and spatial analysis, it is mainly constituted by various scale basis geographic information in natural geographic information of basin, includes topography, water, vegetation, and including villages, towns, cities, provinces, counties, highway, national highway, provincial highway, railway, special surface features, place names and other factors of social geographic information, and includes basin boundaries, study area boundary, different scales digital contour line, different resolution digital elevation model (DEM) and digital terrain model (DTM), as well as including some basic information of vegetation cover, flooding large-scale flood overflow, high-resolution dynamic monitoring, etc.

B. Basin Thematic Data

Thematic data is important part in the products and service of basin geospatial data. Thematic data of digital basin spatial database is established for geographic information display and analysis in basin, it is played a significant role in controlling the whole map surface of digital basin system, geometric correcting remote sensing image, supporting query and retrieval, statistical analyzing basin historical water resources and ecological environment monitoring, etc [2]. It should include different phase, different scales and multi-temporal land use/cover data, soil, vegetation, geology, topography, land type, forest, grasslands, rivers and other thematic data, and even water and rainfall data, industry intelligence data, hydrological data, water quality

data, soil and water conservation data, indicators of river ecosystem data, morphological structure of the data stream, riparian condition data.

C. Remote Sensing Image Data

Remote sensing images has many characteristics, such as detection range, access speed, short cycle, informative, etc, there are irreplaceable important role by other data in the process of digital basin data source building [3]. Remote sensing image data is part or the whole breadth stored according to raster datasets by the radiometric correction, geometric correction, image enhancement processing and mosaic. Establishment of remote sensing image database not only has multi-source multi-temporal remote sensing image data of basin-wide, but also has very high resolution remote sensing data of key areas within the local basin.

D. Socio-Economic Data

Socio-economic data of basin is an important foundation to management and decision making of the local government within the basin scale. Basin socio-economic data should include the basic organization database, the population database, industrial census database, agricultural census database [4]. The construction of digital basin spatial database can realize space of socio-economic data and unified integrate management to the professional statistical data, it include population, basic organization, industry, agriculture, tertiary industry, and other socio-economic integrated data. And need to collect all kind statistical data. such as policies and regulations, socio-economic, population, population distribution, GDP total, per capita GDP, land area, cultivated area, irrigated area, national output, food production, industrial and agricultural output value, etc. These data is the basic data to realize rational use of water resources and socio-economic, ecological environmental coordinated development.

E. Hydrology and Water Resources Data

Hydrology and water resources data of basin is an important part of national basic scientific data, there are irreplaceable role in the national macro-policy development, technology innovation, national economy sustainable development, national security and social life, etc. and can convenient and efficient services to flood control and drought, optimal allocation of water resources, project planning and design and related scientific research [5]. Hydrology and water resources data of basin include spatial and temporal distribution of surface water hydrology, the average annual runoff, mean monthly runoff, surface runoff distribution, water conservancy, etc, but also include some river sections measured data, reservoir distribution of the whole basin, hydrogeology and groundwater depth chart of basin different regions.

F. Meteorological Climate Data

Meteorological climate data includes climate data and weather data, climate data is been collected, processed, arranged, organized a variety of data integration, they are obtained by conventional measuring instruments, widely refers to the original data collection and processing products about the whole climate system. Weather information is a kind of meteorological data for weather analysis and forecasting

services, and it has very strong real-time characteristic. Meteorological climate data of digital basin spatial database construction mainly includes the recorded been observed climate statistics in different ways for many years, such as the annual average temperature, precipitation, sunshine hours, annual mean temperature, precipitation, etc., and the annual maximum, minimum temperature, relative humidity, average wind speed, evaporation, etc, also includes drought, frost, floods, high winds and natural disasters data, etc.

G. Metadata

Scientific data of digital basin multi-source and management complex, in particular of a large number of hard structured observation data. Therefore, metadata is essential, and plays a central role, we must focus on improving the description of meta-information data in order to ensure a variety data correct interpretation, widespread and sustainable use of digital watershed spatial database, in addition, key data set, the data should be specifically detailed written documentation.

4 Functional Design of Digital Basin Spatial Database

H. Basin Spatial Database Importing Function

Basin spatial data input, including graphics and attribute data input. Graphic data involves map, remote sensing images, measurement data, etc., attribute data is attribute information about spatial entities in basin [6]. According to scale, map sheet numbers, temporal, image type and other means, we can automatically imported remote sensing image data and geospatial data into the database by storage program.

I. Basin Data Search Functions

Query and search is the basic function of basin spatial database, because amount of spatial data is stored in the database, and variety, so the data query function must be developed for the users can quickly find the required data. The purpose of basin spatial data search is the users can quickly and efficiently retrieve the required data from a spatial database, essentially which is the graphics data and attribute data of the spatial entities can be queried and searched from a spatial database according to certain conditions, and to form a new subset of spatial data. We can extract different scope, different formats and different projection data from selected vector data, image and metadata according to the needs of users. We can query and display background data and image data of basin through the map sheet number, latitude and longitude, residential, place names, line buffer, a circular area, rectangular area and polygon, etc, queried background data can real time be switched to user-defined map projection or displayed according to the scale, and it has Hawkeye function, it was able to open the window to zoom, narrow and other operations, the user can query and analysis spatial data, such as inclusion, adjacent, etc, according to guide, or enter the SQL statements.

J. Basin Data Updates Function

Updated design of basin spatial data is an important part of spatial database, along with the application depth of digital basin, the data is considered as a bottleneck to

obstruct the development of digital basin, therefore, the means of data updating and acquisition is urgently required to continuously improve [7]. Completely updated of the entire basin spatial database is rarely used in the update method design of spatial data, but more means is used by partial geometric data or attribute data updates. We can automatically import and delete storage framing, hierarchical data, and can edit metadata. According to scale, sheet numbers, phase, image type and other means, through the storage process to automatically import remote sensing image data, space and geographical data is into the database, and can also create an image pyramid for remote sensing data in order to improve retrieval speed.

K. Basin Data Output Function

Basin spatial data output is defined as a process of spatial operations and analysis of the basin results displayed on screen or print to the drawing according to the practical application requirements and visualization principles, the output function is mainly the result of the query to express by various forms. The expression form of query includes data tables, text boxes, data graphics, distributed graphics, digital images, sound, video and interactive windows, etc.

L. Basin Data Sharing Function

Basin spatial data sharing is information resources of geospatial-related and economic and social development required in the location of the drainage area are integrated according to uniform standards, by the support of computer network systems of through, and can exchange data, sharing data and delivery of public services in the policies and regulations framework [8]. Spatial data sharing has been concerned about the issue in the field of digital earth, but also the constructing purpose of digital basin, but it is not yet fully resolved. There are technical factors affecting it, but also non-technical factors, non-technical factors related to policy and social issues, according to the principle of who will benefit investors, we can consider the commercialization of watershed spatial data to solve the problem. Technical factors share is mainly the standardization of data, and the security processing, massive data concurrency, multi-source spatial data exchange technology and heterogeneous distributed shared technology.

M. Basin Spatial Database Maintenance Functions

Basin spatial database maintenance includes backup and recovery, user rights management, database rights management of basin spatial database. We should develop backup strategies according for system availability, acceptable downtime and data loss, and constantly testing the backup strategies to ensure that the database can timely and effective recovery from a variety of failures. Also includes consistency testing functions and security features of database, mainly to complete the user permissions, passwords, manage permissions [9]. We can import appropriate data into the database according to the user setting certain constraints, and submit error reporting to do not meet the conditions data, we should mainly detect layer integrity, property field effectiveness and normative code.

5 Conclusion

The design of digital basin spatial database composition and function is the one of core part of the construction of digital basin, from basic research point of view, basin science can be seen as specific expression in basin scope of earth system science research methods, which means that the required basin science complex and diverse data [9]. Therefore, we should make full use of all kinds of various branches accumulated and different spatial and temporal scales, and standardized information of earth system science, and assimilate and merge these data to form a higher level of integration of data. Only by this ways, we can provide all the required data for different applications objects within the basin scope, such as soil and water conservation, ecological construction, dynamic monitoring, scientific research, water administration and macro decision-making, and improve the scientific management of basin, resource sharing and information service.

Acknowledgment. This work was supported by the National Natural Science Foundation of china (No. 61071121) and research projects of Xiangfan University (No.2009YB013).

References

[1] Su, T., Liu, B., Zhai, S., et al.: Digital Seabed Database: Study on Integration and Management Method for Multisource and Comprehensive Seabed Data. Advances in Marine Science 23(4), 504–512 (2005)

[2] Li, X., Wu, L., Ma, M., et al.: Digital Heihe River Basin.2: Data Integration. Advances in Earth Science 25(3), 306–315 (2010)

[3] Yan, Z., Chen, Z., Gao, F.: The design and establishment process on remote sensing Image database in Tarim River basin. Journal of Northwest University (Natural Science Edition) 34(6), 735–737 (2004)

[4] Wu, J., An, K., Liang, J.: Study on Metadata Based Socio-economic Statistical GIS. Geo-Information Science 8(3), 17–21 (2006)

[5] Laboratory of Remote Sensing and Geographic Information Science of Cold and Arid Regions Environmental and Engineering Research Institute, Digital Heihe3.2 [EB/OL] (October 26, 2010), http://heihe.westgis.ac.cn/

[6] Li, M., Ren, J., Chen, G., et al.: GIS design and implementation. Science Press, Beijing (2004)

[7] Tang, X., Li, L., Ji, X., et al.: Establishment of 1:10 000 Scale DEM Data Base for Priority Flood Control Areas of China Seven Great River Valleys. Bulletin of Surveying and Mapping (6), 19–22 (2002)

[8] Wan, J., Liu, N., Ma, Z., et al.: Design of spatial database based on digital city. Science of Surveying and Mapping 3(6), 107–108 (2006)

[9] Li, X., Cheng, G., Wu, L.: Digital Heihe River Basin.1: An Information Infrastructure for the Watershed Science. Advances in Earth Science 25(3), 297–305 (2010)

Research on Modeling Design of Numerical Control Machine Tool

Xiaowei Jiang* and Xianchun Cheng

Institute of Mechanical Engineering, Changchun University, Changchun, Jilin, 130022, China
{youyuad11080,bcxc1958@}163.com

Abstract. In view of the design problem of the existing numerical control machine tool, the author states her own personal design idea. So the author makes an exploration on industrial design principle that should be followed by modeling design of numerical control machine tool, the man-machine project, the shape design and the color design. At last, the evolution tendency of the modeling design of numerical control machine tool has been pointed out.

Keywords: Numerical Control Machine Tool, Modeling Design, Man-machine Engineering.

1 Introduction

Nowadays, with the rapid development of the technology, the market competition becomes more and more fierce. Under this kind of social environment, the function of the product is no longer the only factor in determining whether the consumers will buy it or not. The innovation of the product, the personalized appearance and shape and other factors become more and more important, which occupy a commanding position in competition. Numerical control machine is a very important equipment in contemporary manufacture. In recent years, from the market research, we can see the machine tools which are convenient in use, beautifully shaped and well-painted are welcomed by the consumers. While the ones which are complicated in control, clumsily shaped, and jazzy colors are hated by consumers. Therefore, the research on modeling design of numerical control machine tool is an extremely important topic.

2 The Present Development Situation of Numerical Control Machine Tool in Our Country

Numerical control machine tool is one kind of highly effective automated precision machine tool that has been developed by comprehensively applying microelectronic technology, computer technology, automatic control, precision measurement and machine tool structure's newest achievements, is one kind of model integration of machinery product [1].

* Corresponding author.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 349–353.
springerlink.com

In the modern technique of manufacture system, the numerical control machine tool occupies an important status, which has a very important effect on manufacturing industry realizing flexibility automation, integration, intellectualism. The numerical control technology and the numerical control equipment are the important foundation of the industry modernization. And whether this foundation is reliable directly affects a national economic development and the comprehensive national strength, which relates to a nation's strategic status.

Although our country's numerical control machine tool has a very great development, and in the aspect of functional design has gradually close to the world's advanced level. But there still have some problems in the aspect of modeling design of numerical control machine tool. For example, it has been neglecting the more and more high esthetic request and the more and more strong individuality performance of people to the modeling design of numerical control machine tool, making some of our country's numerical control machine tool products' color monotonous and tasteless in the modeling design; the shape is huge and tedious; The control system lacks humanity and is not good for the operation; The display equipment is not suitable for observing, etc. Although many numerical control machine tools have the similar function with the overseas products, cannot achieve the same market competitive power. Therefore, researching the modeling design of numerical control machine tool has became an impending problem for the industry of numerical control machine tool to deal with.

3 Industrial Design Principle That Should Be Followed by Modeling Design of Numerical Control Machine Tool

The modeling design of numerical control machine tool, not like other light industry product's design that can relatively more casual and diverse in color, it must strictly follow the industrial design principle, and especially put more emphasis on the scientific requirement in constitutive property, efficiency and reliability.

3.1 Practicality Principle

"Practicality" is the final goal of the control modeling design of numerical machine tool, that is realizing function. But the load bearing of function is the numerical control machine tool's entity structure. All the means and methods in the process of the numerical control machine tool's design and manufacture which actually aims at attaching to the entity function of numerical control machine tool that can be carried on, function being the essence of the numerical control machine tool. Therefore, the exterior structure of the numerical control machine tool must combine with function.

3.2 The Artistic Principle

"The artistic" refers to the requirement of modeling of the numerical control machine tool expressing artistic function. The artistic function is frequently showed through the product's appearance, the proportion, the color and the surface treatment craft, the material utilization and so on.

The design of numerical control machine tool's artistic function must follow certain esthetics rule and the artistic principles, like change and unification, contrast and well distributed, symmetrical and balanced, proportion and criterion, rhythm and rhythm and so on[2].

3.3 The Efficient Principle

Efficiency, is the concept of one of the material base factors as product art - economic efficiency obtains when producing cost and product is being used. Reducing the production cost is the universal demand for product production, which requires in the entire process of product production by using the least financial resources, the physical resources, the labors and time to obtain the best production effects.

3.4 The Principle of Innovation

"The innovation" as a principle also must be followed by the modeling design of numerical control machine tool, because with the development of time and technology, the modeling design of numerical control machine tool will change the past principle of design "the form performance function", "the form performance of psychology", "the form performance of experience" will become its form principle.

4 The Man-Machine Engineering in the Modeling Design of Numerical Control Machine Tool

Numerical control machine tool is the producer of all mechanical manufacture and the most important constituent in the equipment-making industry. With the develop of time and the improvement of people's esthetic ability, people set a higher request to the numerical control machine tool products, not only should have the high tech function and the fine performance, but also should have the effect to make the human to feel working happily, to reduce weary and the effect of pleasuring the eyes. Therefore, the numerical machine tool's color, the choice of material quality will dominate its whole feeling given to human. The good design of a controller may cause the operator to distinguish rapidly without a misoperation, ensuring the accuracy and highly effection, comfort and convenience. The digital display scope which impose on the normal apparent distance scope can enhance the precision and efficiency of the vision to recognize, and also can reduce the eye strain brought by long-time gazing [3].

5 The Studies in Method of the Modeling Design of Numerical Control Machine Tool

5.1 The Shape Design of the Numerical Control Machine Tool

The shape design of the numerical control machine tool will combine the numerical control machine tool's structure and function and other material technologies with the

artistic content, composing a three-dimensional space modeling, which must conform to the artistic principle, be coordinated, be balanced and stable, taking “the unification” primarily, “the change” as auxiliary, linear being succinctly and natural, which make people feel comfortable, coordinated and static within the feeling of moves[4].

5.2 The Color Design of the Numerical Control Machine Tool

The numerical control machine tool’s color attaches in the physique, but more attractive than physique, affecting human's sense organ before the physique, displaying the products’ function effectively, also playing an vital role in beautifying working conditions. The colors should have something to do with the numerical control machine tool’s type, the structure feature, the use environment, the market for demand as well as the different areas’ customs and hobbies.

6 The Evolution Tendency of the Modeling Design of Numerical Control Machine Tool in the Information Age

6.1 Intellectualization, Network and Flexibility

The numerical control machine tool in the 21st century will have certain intellectualized system. Such as the processing process’s adaptive control, craft parameter automatic production and so on; Simplifing the programming, simplifing the intellectualization in operation aspect, like intellectualized automatic programming, intellectualized man-machine contact surface and so on; also has the intelligence to diagnose, contents in the aspect of intelligence monitoring and so on, in order to facilitate the system’s diagnosis and repair and so on.

In recent year, the network numerical control equipment is a hot spot in the development of the machine tool. The network of numerical control equipment will enormously satisfy the production line, the manufacture system, the manufacture enterprise- s’ demand to the information integration.

With the further development of numerical control machine tool, it will found a new epoch in the flexible automation processing and will have a very strong adaptiveness to satisfy the processing object’s transformation.

6.2 The Recombine

The aim of functional recombine is to further enhance the machine tool’s production efficiency and to reduce the auxiliary non-processing time to the least. It can expand the use scope of the machine tool, enhance the efficiency and realize the multipurpose of one machine in many uses through the functional recombine machine.

6.3 The Ecology

With the development of the information age as well as the massive consumptions to the earth’s resources, the modeling design of numerical control machine tool more and

more pursues the ecology. It is under the ecological philosophy instruction that the ecology thought is utilized, the modeling design of numerical control machine tool is integrated the “human - machine - environment” system, both considers to meet the human's need and pays great attention to the ecological environment protection[5].

References

1. Zhang, D., Chen, S., Lin, B.: Mechanical Manufacturing Equipment and Design. Tianjin University Press, Tianjin (2003)
2. Fu, L.: Research on Modeling Design of Industrial Product. Jilin People Press, Changchun (2002)
3. Zhan, X.: Art Design of Machine. Hunan University Press, Changsha (1999)
4. Li, L.: Introduction of Art Design. Anhui Arts press, Hefei (2003)
5. Chen, W.: Ergonomics, vol. 5, p. 51 (1999)

Transient Sensitivity Computations for Large-Scale MOSFET Circuits Using Iterated Timing Analysis and Adaptive Direct Approach

Chun-Jung Chen

Department of Computer Science
Chinese Culture University
Taipei, Taiwan
teacherchen62@yahoo.com.tw

Abstract. Transient sensitivity information is useful in many EDA tools such as circuit optimizers. In this paper, we investigate the transient sensitivity computations in large-scale MOSFET circuits. Iterated Timing Analysis (ITA) is used as the fundamental algorithm to undertake both timing and sensitivity simulations. A modified Direct Approach for sensitivity computation is proposed, which is called Adaptive Direct Approach that considers the number of design parameters and sizes of subcircuits. All proposed methods have been implemented and tested to justify their outstanding performance better than previous works.

Keywords: Circuit simulation, relaxation-based algorithms, sensitivity computation, Iterated Timing Analysis.

1 Introduction

We know that transient sensitivity information is very useful in many applications in the circuit design community. There are two efficient techniques for computing the transient sensitivity, which are Direct Approach [1] and Adjoint Approach [2]. Adjoint Approach is more complex for computation, but it calculates the sensitivity of a circuit variable with respect to all design parameters at a time. So, it is more suitable for practical optimization programs and is more efficient when the number of design parameters is big. In contrast to Adjoint method, Direct Approach solves for transient sensitivities in the way similar to that for time waveforms, so it is easier to implement. But it is not efficient when many design parameters have been assigned.

This paper designates to investigate efficient transient sensitivity computations for large-scale MOSFET circuits containing various numbers of design parameters and various sizes of subcircuits. It is well known that relaxation-based algorithms [3][4] have lower time complexities in dealing with large-scale MOSFET circuits. Especially that ITA [4] (Iterated Timing Analysis) has outstanding simulation robustness in compared to other relaxation-based algorithms. Also, ITA is popular in practical EDA programs. Due to these reasons, we choose it as the fundamental simulation algorithm. We also choose Direct Approach as the sensitivity subcircuit solver. Moreover, an improved version called *Adaptive Direct Approach* (ADA) is

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 355–363.
springerlink.com

proposed. The affecting factors of using sensitivity computations include the number of design parameters, the sizes of subcircuits, positions of design parameters, and so on. These factors influence the efficiencies of sensitivity computations, and they are considered in this paper. All the proposed methods will be clearly described and real implemented to justify their performance.

In follows, fundamental numerical methods are described in Section 2. Section 3 then follows to describe the Adaptive Direct Approach. Section 4 shows experiments on the implemented program. Finally, a conclusion section is given.

2 Fundamental Numerical Algorithms

2.1 The ITA Algorithm for Time/Sensitivity Simulations

Algorithm 1 shows the pseudo codes of ITA algorithm. At the line labeled with ST, the function *fan_out*(*a*) is the set of all fan-out subcircuits of current simulated subcircuit *a*. This line implements the important *Selective-tracing Scheme* [4] of ITA. A priority queue, at line SEL, is used to store the subcircuits waiting for simulation. These two lines will be used to equip one of the proposed techniques. ITA is lack of the ability to utilize circuits' multi-rate behavior, which is a serious disadvantage in dealing with big circuits. This problem can be alleviated by using the method mentioned in [5] (which "trims" ITA's *fan_out*(*a*) set to break the "global time point simulation" manner of ITA). We use the *Latency-separating Scheme* of [5] to improve ITA's efficiency. By this improvement, big circuits can be simulated much more efficiently.

Algorithm 1 (ITA Algorithm for Circuit Simulation).

```
   // Simulate for T_begin ~T_end
   // E() is an priority queue, whose elements are ordinary queues
   Put subcircuits connected to primary input into E(T_begin);
Sen: // Do graph traversal to get SensMark index array;
   // Put subcircuits containing design parameters into E(T_begin);
SEL: while(E is not empty) {
       t_n+1 = the smallest event time in E;
       for(k = 1; E(t_n+1) is not empty; k++) { // k is the relaxation index
          Clear TMP;   // TMP is a queue
          for(each subcircuit a in E(t_n+1)) {     // E(t_n+1) is a queue
             Solve a at t_n+1 for transient responses;
Sb:             // Solve a for sensitivities with respect to all design
                parameters;
             if(a has been converged) { // converged
                Estimate next solving time t_next ;
Ts:                // Read the t_next from saved time waveforms;
                Add a into E(t_next);
             }
             else { // not converged
                Add a into TMP;
Ps:                // Remove subcircuits not active from fan_out(a);
ST:             Add fan_out(a) into E(t_n+1);
             }
          }
          E(t_n+1) = TMP;
       }
   }
```

Note that ITA is usually used in ordinary circuit simulation that calculates the transient responses. However, it can also be used to solve for transient sensitivity waveforms of circuits [3]. Algorithm 1 has also represented this part of codes for sensitivity simulations. To calculate sensitivities, at lines labeled with Sen, Sb, Ts, and Ps the comment marks need to be removed and replace corresponding lines. In our implementation, the time and sensitivity simulations are separated. The time simulation undertakes at first and stores time waveforms into tables. The sensitivity calculation then goes ahead and utilizes the stored time waveforms (such as to regenerate Jacobian matrices).

2.2 The Direct Approach

We describe the Direct Approach now. The simulated circuit can be represented as follows:

$$F(Y(t),\dot{Y}(t),t)=0 \tag{1}$$

where Y is the vector of circuit variables, t is the time, F is a continuous function and dot means differentiation with respect to time. In relaxation-based algorithm, (1) needs to be partitioned into subcircuits, where the ith one of which is:

$$F_i(Y_i(t),\dot{Y}_i(t),C_i(t),\dot{C}_i(t),t)=0 \tag{2}$$

One subcircuit, a, is rewritten as the abbreviated form:

$$f(y(t),\dot{y}(t),w(t),\dot{w}(t),t)=0 \tag{3}$$

where y (Y_i, a sub-vector of Y) is the vector of n circuit variables in a, w (C_i, the decoupling vector) is the vector of circuit variables not in a, and f is a continuous function. Assume all m design parameters (circuit parameters to be adjusted in circuit design process) are in the vector D, (3) is rewritten as:

$$f(y(t),\dot{y}(t),w(t),\dot{w}(t),D,t)=0 \tag{4}$$

We now consider the transient sensitivity with respect to the jth design parameter, D_j (or d), so we differentiate (4) with respect to d to have:

$$f_y\frac{\partial y}{\partial d}+f_{\dot{y}}\frac{\partial \dot{y}}{\partial d}+f_w\frac{\partial w}{\partial d}+f_{\dot{w}}\frac{\partial \dot{w}}{\partial d}+f_d=0 \tag{5}$$

To solve (5), we need to apply the integral method to transform it into nonlinear algebraic equations. If we use the Trapezoidal Rule (TR), we have:

$$J_y s_{n+1}=-J_w p_{n+1}+(\frac{2}{h}f_{\dot{y}}s_n+f_y\dot{s}_n+\frac{2}{h}f_{\dot{w}}p_n+f_w\dot{p}_n)-f_d \tag{6}$$

in which, $s=\frac{\partial y}{\partial d}$ is the transient sensitivity and $p=\frac{\partial w}{\partial d}$ is the "input" transient sensitivity. The notation "n+1" means at current time point t_{n+1} and "n" means at previous time point t_n. Equation (6) is rewritten as follows for convenience.

$$J_y s_{n+1} = -J_w p_{n+1} + O_n - f_d \tag{7}$$

in which O_n is the vector composed of "old" values at previous time point. This equation is used in sensitivity simulation to derive the sensitivity of a with respect to d, the jth design parameter. Considering entire sensitivities with respect to all design parameters, we have:

$$J_y \hat{s}_{n+1} = -J_w \hat{p}_{n+1} + \hat{O}_n - f_D \tag{8}$$

in which all symbols with hat are matrices for corresponding values with respect to all design parameters (e.g. $\hat{s} = \frac{\partial y}{\partial D} \in R^{n \times m}$, is the matrix for sensitivities of a with respect to D). And, $f_D \in R^{n \times m}$ is a matrix, too.

Equation (8) represents the Direct Approach. Traditionally, Direct Approach is used to solve the entire circuit, but, in this paper, we use (8) to solve a subcircuit at line Sb of Algorithm 1. Following algorithm describe the details of using (8) to solve for sensitivities of a subcircuit.

Algorithm 2 (Direct Approach for Subcircuit Computation):

for(j = 1; j <= m; j++) {
    Calculate f_{Dj}, and O_j;
    Solve (7) for s_{n+1} with respect to D_j;
}

2.3 The Portion-Simulating Method

In the sensitivity computation, transient sensitivity waveforms flow from "source subcircuits" that contain design parameters. Only some portions of the simulated circuits will be affected by these sensitivity waveforms. The *Portion-simulating Method* (PS) is to simulate portion of the circuit that is "active" and views the other portion as nothing. It's appropriate to use PS Method in sensitivity computation.

The *active* status of subcircuits can be statically identified by graph traversal on the *subcircuit signal flow graph* (in which subcircuits are viewed as vertices, and affecting relations are viewed as directed edges). We use an index array *SensMark*[1~m][1~N] (N is the number of subcircuits) to store the "affecting" statuses. For example, *SensMark*[j][i] records whether ith subcircuit will be affected by the sensitivity waveform with respect to the jth design parameter. We call a subcircuit active if it has been affected by at least one design parameter. In Algorithm 1, the priority queue, E, is utilized to install PS Method. Only active subcircuits can be put into the priority queue (see line Ps) to be calculated. Besides this, we can further use the situation that an active subcircuit may not be affected by all design parameters. In Algorithm 3 (in later section), this situation is utilized at line Bp, where some calculations for f_D and O can be saved.

3 The Adaptive Direct Approach

We propose the alternative Direct Approach, called *Row-major Direct Approach* (RDA), which computes the sensitivity matrix in the row-major fashion, while the ordinary Direct Approach (ODA) is in the column-major fashion (Fig. 3 shows these relations). We will prove that RDA is more advantageous in dealing with much design parameters. For a subcircuit *a*, the sensitivity equation for *j*th design parameter can be written as:

$$J_y \hat{s}_{j,n+1} = -J_w \hat{p}_{j,n+1} + \hat{O}_j - f_{D_j} \tag{9}$$

We move the Jacobian to the right-hand-side to get:

$$\hat{s}_{j,n+1} = J_y^{-1}(-J_w \hat{p}_{j,n+1} + \hat{O}_j - f_{D_j}) \tag{10}$$

Extending (10) into matrix form, we have:

$$\hat{s}_{n+1} = J_y^{-1}(-J_w \hat{p}_{n+1} + \hat{O} - f_D) \tag{11}$$

We want to calculate the sensitivity matrix in row-major fashion. Assume *row*(*A*, *i*) takes the *i*th row of the *A*, we have:

$$row(\hat{s}_{n+1}, i) = row(J_y^{-1}, i)(-J_w \hat{p}_{n+1} + \hat{O} - f_D) \tag{12}$$

Equation (11) represents the RDA for a sensitivity subcircuit calculation. The detailed calculations are described in following algorithm.

Algorithm 3 (RDA for Subcircuit Computation):

```
Inverse J_y;
for(j = 1; j <= m; j++) {
Bp:   if(SensMark[j][a] is false) continue;
      Calculate f_Dj, and O_j;
}
for(all circuit variables, v) {
      i = v's order (v is the ith circuit variable);
      Solve (12) for a_{i,n+1};
}
```

We compare the two Direct Approaches. Referring to Algorithm 2, we derive the computation effort formula for solving a subcircuit by ODA:

$$m \times (L + R) \tag{13}$$

in which *m* is the number of design parameters, *L* is the effort for solving linear system, and *R* is the effort to compose the vector on the right-hand-side of (7). Also, we can get the effort formula for solving a subcircuit by RDA:

$$m \times R + I + n \times M \tag{14}$$

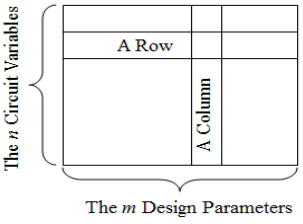


Fig. 1. The entire sensitivity matrix of a subcircuit

Table 1. The Minimum Design Parameter Number for RDA to Beat ODA in MOSTIME[$]

Subcircuit Node Numbers and Design Parameter Numbers									
n	2	3	4	5	6	7	8	9	10
m	22	14	10	7	6	6	5	4	3

[$] If $n < 2$, ODA is suggested. If $n > 10$, RDA is suggested.

Table 2. Specifications of Simulated Circuits

Ckt	**Name**	**Node#**	**MOSFET#**	**Subckt#**	***m***
1	100-stage Inverters	100	200	100	10
2	100-stage Inverters	100	200	10	10
3	4-bit ALU	168	400	44	1
4	4-bit ALU	168	400	44	20
5	64-bit ALU	2688	6400	704	20
6	4-bit Sync. Counter	88	176	12	20
7	16-bit Shift Register	320	640	32	20

in which R is the same as in (13), I is the effort to inverse a matrix, and M is the effort to perform matrix multiplication (on right-hand-side of (12)). Apparently, the effort formulas of these two Direct Approaches have different characteristics. Equation (14) is more advantageous than (13) when the m is bigger. In fact, following equation states the condition that (14) is smaller:

$$m \geq \frac{I(n) + nM(n)}{L(n)} \tag{15}$$

in which n is the number of subcircuit nodes. Due to real implementations, effort variables are hard to state. The Adaptive Direct Approach is the scheme to choose appropriate Direct Approaches. To judge when to use RDA, we devise a heuristic method using real experimental data. Table 1 shows situations that suggest using RDA in our simulator MOSTIME [3][5], in which the minimum m with respect to different subcircuit node numbers are listed. If n is out of the range of Table 1, specific Direct Approaches are suggested (see the bottom line of Table 1). We just implement ADA by using Table 1.

4 Experimental Results

The proposed methods have been implemented in the experimental circuit simulator MOSTIME [3][5]. To emphasize the simulation algorithm's effects, we use the simple analytic model for MOSFET and partition circuits into subcircuits such that small subcircuit feedback loops and strongly-coupled subcircuits do not exist (to emulate the situation of real circuit simulators). The ITA algorithm has been used in both time and sensitivity simulations, but they are separated. The design parameters are all width of MOSFET transistors. We have simulated seven experiments, whose circuit specifications are listed in Table 2. The (sensitivity) running results, including subcircuit calculation counts and CPU times, are recorded in Table 3 and 4. Four versions of algorithms have been performed, which are listed on the first row of tables. The ratios of subcircuit calculation counts and used CPU time compared with those of ITA+ODA (the classical method) are shown to represent the normalized values for our convenience of comparing.

Table 3. Subcircuit Calculation Counts

Ckt	Subcircuit Calculation# (K) / Ratio to ITA+ODA's			
	ITA+ODA	***ITA+RDA***	***ITA+ADA***	***ITA+ADA+PS***
1	2695.3	2695.3/100%	2695.30/100%	1407.3/52.2%
2	1097.4	1097.4/100%	1097.41/100%	666.92/60.7%
3	109.25	109.25/100%	109.254/100%	41.173/37.6%
4	109.90	109.90/100%	109.909/100%	41.651/37.8%
5	283.47	283.47/100%	283.47/100%	83.897/29.5%
6	91.431	91.431/100%	91.431/100%	91.431/100%
7	389.97	389.97/100%	389.97/100%	253.06/64.8%

Table 4. Used CPU Times

Ckt	Used CPU Times[$] / Ratio to ITA +ODA's			
	ITA+ODA	***ITA+RDA***	***ITA+ADA***	***ITA+ADA+PS***
1	108.48	127.12/117%	107.62/99.2%	48.485/44.6%
2	853.53	739.70/86.6%	703.95/82.4%	456.83/53.5%
3	3.807	5.6/147%	3.946/103.6%	1.544/40.5%
4	23.01	14.92/64.8%	15.22/66.1%	5.538/24%
5	64.958	47.25/72.7%	47.751/73.5%	14.274/21.9%
6	87.937	25.59/29.1%	25.413/28.8%	25.179/28.6%
7	534.61	171.28/32%	168.97/31.6%	111.64/20.8%

[$]CPU time is in Pentium 2.2G seconds.

In Table 3, all algorithms spend the same amount of subcircuit calculations except the one with PS Method that bypasses some subcircuits not affected by any design parameters. We note that all simulated circuits have been arranged in the form of "cascaded stages." We just assign design parameters in middle part of these cascaded stages for all circuits except the 6th circuit (whose design parameters are assigned to the first stage). Therefore, the saving ratios on subcircuit calculation counts for all circuits are obvious (about 34% to 71%) excepts that for the 6th circuit is zero. Note that in Table 4, the saving ratios for CPU times are proportional to those in Table 3.

Table 4 shows the used CPU times. We check the first two circuits at first. They are the same but have been partitioned into small and big subcircuits respectively. The 1st circuit has small subcircuits, which is not beneficial to RDA. On the other hand, the 2nd circuit is more beneficial to RDA. The used CPU times clearly justify this situation. We note that ITA+ADA can choose appropriate Direct Approach for subcircuit calculations, so it always keeps the best CPU times.

Next, we check the following two circuits, 3rd and 4th circuits. They are also the same but with different number of design parameters. As we explained previously, RDA is more beneficial in dealing with much design parameters (in the 4th circuit). In contrast to this situation, RDA is much worse than ODA (in the 3rd circuit). We can clearly observe this phenomenon in Table 4. Again, the ITA+ADA version always shows the best performance. We then check the 5th circuit that is big. There are 20 design parameters assigned in this circuit, but in real situation more design parameters might be assigned since this is a big circuit. However, obvious saving in CPU time has been observed already by using ITA+RDA version. The saving ratio would increase if the number of design parameters is enhanced. The 6th and 7th circuits are all feedback circuits, in which many transistors inside feedback loops have been partitioned into same subcircuits and hence big subcircuits emerge. In (13), the value for L increases dramatically when n (node number of the subcircuit) is big, so effort of ODA is much bigger than that of RDA. The saving effects of ITA+RDA version are significant in these two circuits. We note that in the last three circuits, due to big subcircuits and big number of design parameters, ITA+RDA version is already the best version, so ITA+ADA version just performs the same.

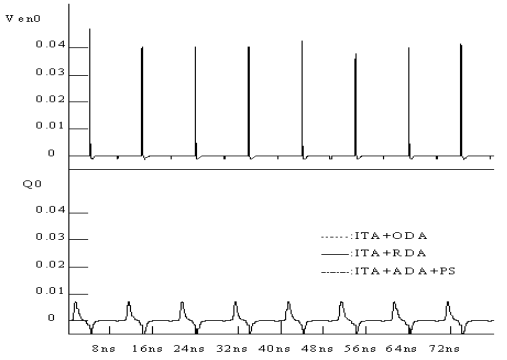


Fig. 2. Sensitivity waveform comparison for the 4-bit Synchronous Counter, where the design parameter is the width of a MOSFET in the 1st stage

The simulated waveforms of proposed methods are just the same as those of classical method. Fig. 2 shows comparisons for the sensitivity waveforms of the 6th circuit.

Various partitioning situations and different numbers of design parameters have been assigned into experimental circuits to test the proposed methods. The running results justify ADA and PS Methods are more efficient than original method. The proposed methods have successfully achieved the designated goal of speeding up the sensitivity computation for large-scale MOSFET circuits.

5 Conclusions

We take into account the diversifying customs of transient sensitivity users, and raise a corresponding adaptive technique to speed up the sensitivity simulation. The elements of this technique including Row-major Direct Approach, which is beneficial in treating bigger subcircuits and much design parameters, and PS Method, which can prune off unnecessary subcircuit calculations. Experimental results justify that the new method is robust and efficient in performing the large-scale transient sensitivity simulation for MOSFET circuits.

References

1. Hocevar, D.A., Yang, P., Trick, T.N., Epler, B.D.: Transient sensitivity computation for MOSFET circuits. IEEE Trans., Computer-aided Design CAD-4, 609–620 (1985)
2. Director, S.W., Rohrer, R.A.: The generalized adjoint network and network sensitivities. IEEE Trans. Circuit Theory CT-16, 318–323 (1969)
3. Chen, C.J., Feng, W.S.: Relaxation-based transient sensitivity computations for MOSFET circuits. IEEE Trans. on Computer-aided Design 14(2), 173–185 (1995)
4. Saleh, R.A., Newton, A.R.: The exploitation of latency and multirate behavior using nonlinear relaxation for circuit simulation. IEEE Trans., Computer-aided Design 8, 1286–1298 (1989)
5. Chen, C.J., Chen, B.C., Lee, C.J., Tsai, C.L., Chou, L.P., Chang, A.Y.: Methods to Enhance the Performance of Iterated Timing Analysis Algorithm. In: IEEE International Conference on Information and Automation, Harbin, China, June 20-23 (2010)

UML-Based Design of University Course-Selective Information System

Shan Peng, Haisheng Li, and Qiang Cai

College of Computer and Information Engineering
Beijing Technology and Business University
Beijing, P.R. China, 100048
ihsh@th.btbu.edu.cn

Abstract. Course-selective information system is accepted by more and more colleges because of its convenience. An excellent system can rapidly relieve the work of staff members, and it is essential in students' study life. In this paper, university course-selective information system is designed and analyzed by UML. As a modeling language, UML is used to analyze requirements of the system by the use case diagram, and used to design the static structure of the system, as well as the dynamic behavior of the system. All works which are realized in the paper are prepared for the future system development and maintenance.

Keywords: UML, university course-selection, system analysis, system design.

1 Introduction

With the development of global informationization, course-selective information system is widely used in colleges and universities to manage their daily work. By executing information system, the universities' managers have reduced their studio load and their work time. Meanwhile, the information system helps students selecting courses more freely and fairly. Along with the extension of courses and students in college, the system should be more elective. Considering about the complexity of the system, we must carefully think about the whole flow path at the process of the system's designing. Based on this, UML is the suitable tool to design the university course information system.

The following of the paper is organized as follows: Section 2 explains the features of UML. The requirements of the information system are discussed in section 3, and then the system is designed by the static model and the dynamic model afterwards. Conclusion is given at the final section.

2 Unified Modeling Language

UML, Unified Modelling Language, is the de facto standard for object-oriented modeling, widely used in intact system development. UML is a family of modeling

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 365–372.
springerlink.com

notations for specifying, visualizing, constructing and documenting artefacts of software-intensive systems. It provides a collection of visual notations to capture different aspects of the system under development [1]. UML has been evolved to describe the structure of software systems, visualize software systems, and establish the documents of systems. UML provides the static model, the dynamic model and the organizational structure of the model, also including the semantic representation and explanation. Generally based on UML, the process of modeling system development is that, firstly, accurately describing the requirements and functions of the system, and then establishing the static structure model with the needs, showing the logical structure, at last, describing the dynamic model by utilizing the dynamic behaviour which is based on the static model in the system. The whole process is shown in Figure 1.

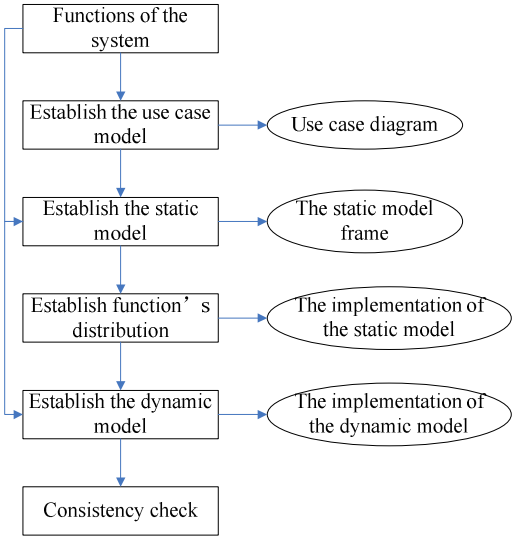


Fig. 1. The process of the system development modeling

Compared with other software modeling method, the main features of UML are:

1. Unified standard. With the integration of the major system development methods' concepts and techniques, UML becomes the standardization in the modeling languages, and terminates the previous modeling languages' inconsistencies and differences.
2. Object-oriented. UML supports object-oriented technology. It provides the graphical elements to concisely express the object-oriented concepts and elements [2].
3. The strong ability of visualization. UML is a graphic language, which can clearly describe the logical model with graphs. UML can be used to model various complex software systems because of its ability of visualization.
4. Independent of processes. UML is independent of the systems' development process. It has been implemented in the object-oriented development process, as well as in the software life cycle method.

3 UML-Based Design of the Information System

College students can choose appropriate courses for themselves when they login into the course-selective information system. The system has three roles: students, teachers and administrators. Students are the main users of the system. At the same time, the works which teachers and administrators do in the system can help students selecting courses. Considering the relationship, this paper is mainly describing the information system from the perspective of the students.

3.1 Requirements Analysis

The use case diagram can easily and completely present the demands and functions of the information system. According to a visual sense, the use case diagram points out the operators of system functions, and shows the needs of the users in the system. The use case diagram of university course-selective information system is shown in Figure 2.

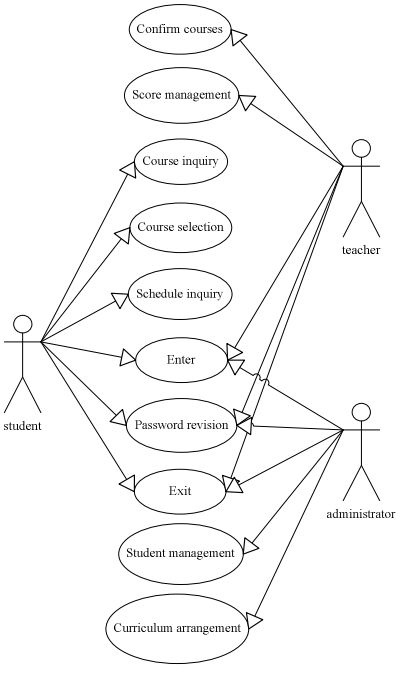


Fig. 2. The use case diagram

From the use case diagram, we can see the functional requirements of students: course inquiry, course selection and schedule inquiry. This three functions have formed a course-selective flow path. Firstly, students can look up the detailed information of courses in the "course inquiry" function, choose some interested courses which are also befitting in time and address in the "course selection" function. Then, when the choice is completed, all information of chosen courses can be seen in the schedule. The functional requirements of teachers are: confirm courses and score

management, but administrators' requirements are student management and curriculum arrangement. The two roles' requirements interact with each other. After teachers confirm courses, administrators need to add other information about courses, such as class time and class place. After this, the information of courses is intact so that students can consider about the selection seriously. While all students complete course selection, the data will be collected by administrators in the "student management" function and teachers can get the data about students. When the course is over, teachers should assess scores for students. The same functions for three roles are: login in, password revision and login out. With the use case diagram, the information system can be designed from the requirements.

3.2 Design of the Static Model

The static model describes the internal structure of systems, and defines objects and classes in the whole demand analysis. The internal and mutual relations of objects and classes are also defined in the static model. There are relations in objects and classes, such as association, polymerization and dependence. The scope of functions in entities is limited tightly in the model. So we can build the static model of systems roughly, and then gradually realize the internal functions of minute-descriptive requirements. There are some diagrams to describe the static structure of systems: Class diagram, Object diagram, Package, Component diagram and Deployment diagram.

One of the static diagrams is Class diagram. We can draw Class diagram to describe the classes and their relations. UML Class diagram describes classifiers (e.g. classes and interfaces) and relationships between classifiers. A class is a classifier that characterizes a family of objects in terms of attributes and operations which are common to the objects [3]. Associations between classes have determined the way of links between class objects. The main classes in the information system are students, teachers, administrators and courses. Figure 3 shows the classes and their relations of Class diagram.

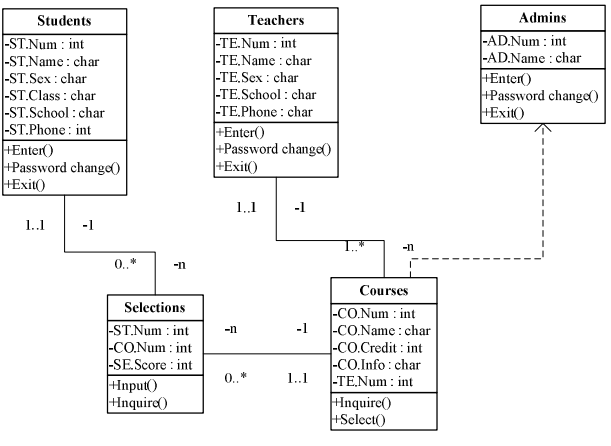


Fig. 3. Class diagram

In Figure 3, including the three roles: students, teachers and administrators, the classes present their properties. They all have primary keys to distinguish themselves. There is a class named courses which is the most important class in the system. The 'courses' class connects the other classes because the system is course-selective. With the connection, the Class diagram has exposed the structure of the system.

In Class diagram, we can see the classes and their relations. Though there are other classes in the system, the classes in the chart are perpetual and frequently-used in the system. The provisional classes are usually created to preserve common information in the system's flows. Therefore, the relations of classes which contain common information can also be described by the provisional classes. While we use Class diagram to represent the classes in the system, we use Object diagram to represent the objects. The objects in the information system are the actual example of the classes. Object diagram is the variant of Class diagram. So the relations in two diagrams are uniform. The class objects and their simple connections in the system are shown in Figure 4.

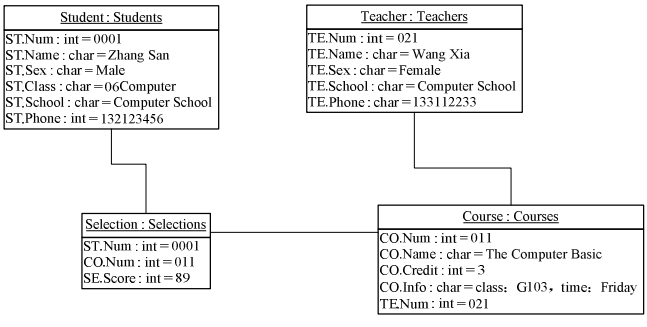


Fig. 4. Object diagram

The objects in the Objects diagram are students, teachers, courses and selections. With the objects, the diagram clearly reflects the internal relations. The 'courses' object is connected by 'teacher' object with the teacher who establish the course, and 'students' object adopts 'selections' object to select the courses. After all, the objects can be connected with each other directly and indirectly.

With Class diagram and Object diagram, we have clearly analysed the static structure of the system. But how are the classes and objects operated in the system designing? The answer is in the dynamic model.

3.3 Design of the Dynamic Model

The dynamic model in systems mainly describes the interactions among objects and the exchanging between messages. According to the dynamic model, we will realize the logical connection of the system. The dynamic model is incorporating State chart, Activity diagram, Sequence diagram and Collaboration diagram.

From the initial to the end, or from the formation to the remove, the system may stay at a series of state changes. State chart is built to show the whole possible states (initialization, finalization, intermediateness and complex state), and the events which trigger all the state changes. State chart describes system behaviour, and it's attached to a class which specifies all behavioural aspects of the objects [4]. Considering Figure 5, we will see the process about students entering into the system.

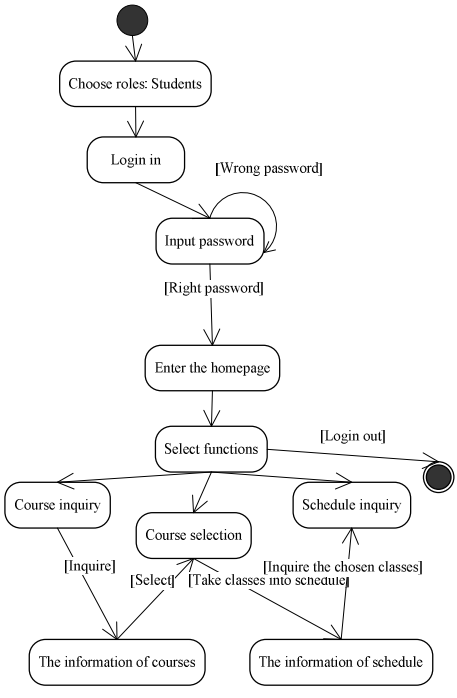


Fig. 5. State chart

In State chart, All behaviours happened in students are recorded. When students enter into the homepage of the system, they can select functions. The presentations will change along with the students' selections. Users should choose their role at the beginning and login out at the last.

According to handling some activities, Activity diagram is conducted to display the process control flow between two objects, as the flow from one activity to another. When the states of objects change, Activity diagram can be acquired the results after changing. We can see about teachers entering into the system in Activity diagram at Figure 6.

However, the two diagrams introduced as before, can hardly express out the time characteristic of changes in systems. We need Sequence diagram built to describe the time feature. Sequence diagram describes how instances interact to accomplish a task. An interaction is expressed in terms of lifelines and messages [3]. The important trait of Sequence diagram is that it emphasizes at the dynamic interaction between objects, and the succession of the message transfer. In Sequence diagram, there are many kinds of lines which mean different purport, like that different horizontal represents different objects, vertical means time, and the arrows of a message indicates its type. In the process of course selecting, there are three objects: students, the WEB server and the Database server. Along with the operations which students deal with the system to select courses, the responses from two servers are different. Figure 7 has shown the sequence process of course selected by students.

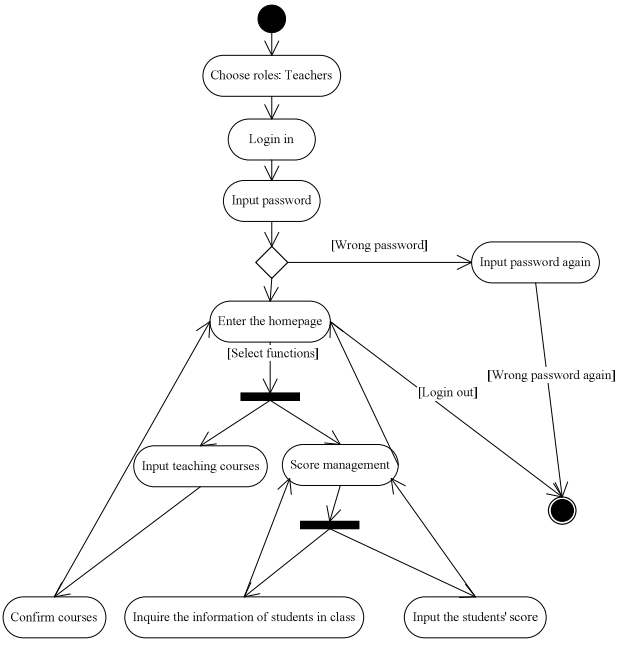


Fig. 6. Activity diagram

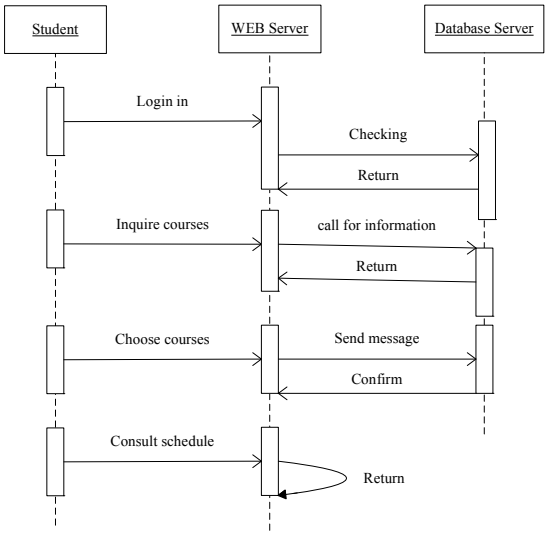


Fig. 7. Sequence diagram

According to three dynamic charts, we have realized the logical connection of the system clearly. State chart and Activity diagram represent the events' and the activities' changes in the dynamic structure, while Sequence diagram indicates the time and sequence features of the dynamic model. In summary, we have finished analyzing the UML models of the information system.

4 Conclusion and Future Work

This paper analysed the functional requirements of the university course-selective information. In the meantime, by using UML modelling, the system is built with the static and dynamic model, which is laying the foundation for the future system development. The standardization of UML not only effectively promotes mutual understanding among the designers, developers and testers, but also makes the development process clearer [5]. The analysis has showed that the UML modeling for the system greatly regulates the whole system development, and this will be helpful for the future work.

Acknowledgment. This research is partially supported by Science and Technology Development Program of Beijing Municipal Education Commission KM200910011007, Funding Project for Academic Human Resources Development in Institutions of Higher Learning under the Jurisdiction of Beijing Municipality PHR20110875 and Beijing Technology and Business University Research and Education Reform Project jg105205.

References

1. Dong, J., Yang, S., Zhang, K.: Visualizing Design Patterns in Their Applications and Compositions. IEEE Trans. Software Eng. 33(7) (July 2007)
2. Kong, J., Sun, Y., Jiang, M., Bi, B.: The Requirement Analysis Based on UML. Computer Eng. and App. 15 (2003) (in Chinese)
3. Robert, B., Kim, D.-K., Ghosh, S., Song, E.: A UML-Based Pattern Specification Technique. IEEE Trans. Software Eng. 30(3) (March 2004)
4. Harel, D., Gery, E.: Executable Object Modeling with Statecharts. IEEE Computer 30(7) (July 2007)
5. Song, J.: Design of the Senate Needs Analysis System Based on UML. Software Guide 8(4) (April 2009) (in Chinese)

Study of Energy Performance Contracting Project Risk Based on Fuzzy Neural Network

Hui Shi, Dongxiao Niu, and Hanmei Wang

Dept. Economics and Management
North China Electric Power University
Beijing, China
beijingwanghanmei@126.com

Abstract. The paper introduced the energy performance contracting (EPC) mechanism and studied the policy, market, financing, operational, benefit and customer risk existing in the process of the development of EPC. On the basis of it, the target system of Risk Evaluation on EPC was discussed. It used the fuzzy neural network to make the evaluation of the risk of EPC project, and fuzzily processed the value of degree and the evaluation standard by adopting the relative membership degree. Through empirical analysis, the results are accurate and rational.

Keywords: energy performance contracting, energy service companies, energy conservation, risk evaluation, fuzzy neural network.

1 Introduction

The energy performance contracting (EPC) is a new energy-saving mechanism which was developed from the western developed countries in the 70s of the 20th century. This mechanism based on the purpose of profit is developing rapidly, and has become a new energy industry. In 1998, the World Bank's Global Environment Fund China Energy Conservation Promotion Project began its operation, which brought the mechanism into China.

The basic mechanism of EPC can be shown as follows: the energy service companies (ESCO) who consider EPC as main business feature sign energy service contract with customers who are willing to conduct energy saving; the ESCO provide the customers with multiple services, such as energy efficiency audits, energy project design, construction, monitoring, operation management and so on; clients pay the cost of products and services every year with the benefits obtained from the energy saving [1].

Chinese enterprises and energy service companies operate the energy-saving project through EPC. This can not only improve their profits and developments, but also enhance the energy efficiency and reduce greenhouse gas emissions, which greatly boost the energy consumption reduction of the whole society. However, since the characteristics of the mechanism which means "enterprise customers have zero risk", energy service companies have to undertake most risks from planning, assessment, decision-making, to implementation. Therefore, it is crucial to conduct risk management to the EPC project, which helps reduce losses and increase revenue [2].

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 373–381.
springerlink.com

From the perspective of ESCO, this paper analyzed various risks of EPC project, established a more comprehensive risk evaluation system. As for the risk evaluation model, considering various defects of previous risk evaluation model, this paper brought in fuzzy neural network optimization evaluation model, that is: fuzzily processed the value of degree and the evaluation standard by adopting the relative membership degree, and then gave the evaluation with BP neural network model. The method can effectively remove the subjective parts of the previous risk evaluation model, and greatly improve the accuracy of the risk evaluation.

2 Risk Analysis of Energy Performance Contracting Project

2.1 Policy Risk

At present, our country is short of the related legal environment which can support the healthy and stable development of the EPC mechanism. Although our country promulgated “The People's Republic of China Energy Conservation Law”, it didn’t adopt coercive measures to the production of companies; the companies that implement the EPC projects can’t attain the country’s policy support in taxation and finance. Therefore, companies have no impetus to save energy.

2.2 Market Risk

Most of the customers don’t understand the EPC mechanism, and they often hold wait-see attitude. In addition, it is difficult to promote the new energy-efficient technology, which results in market risks. Moreover, although the energy management companies in China has just started and their sizes are small, we should know the fact that there are not only competitions among domestic enterprises, some large foreign energy services companies relying on abundant capital and advanced technology are ready to enter China's energy efficiency market. What’s more, the changes of the world energy prices also have an impact on the EPC projects. If energy prices decline, the energy saving projects may have difficulties in cost recovery [2].

2.3 Financing Risk

At present, China's energy service companies have difficulties in financing for the project through banks and other financial institutions, the main reasons are as follows: these companies are considered as SMEs which are quite young, thus they haven’t established credibility in the financial system; energy-saving projects are not like the general projects which can form collateral; credit assessment departments of financial institutions are not yet familiar with the energy services sector, etc.. Thus, although energy service companies have large one-time investment, the sources of funding are not smooth. In addition, fluctuations in interest rates also increase financing risk.

2.4 Operational Risk

Operations include project implementation and post-implementation operation and maintenance. In the operation process of energy-saving project, the project management, quality and cash flow are crucial to the success of energy saving projects. Meanwhile, in the view of the implementation results of the EPC projects in China these years, the post-implementation operation and maintenance of energy-saving projects have great impact on the success of the projects and the long-term cooperation with customers [2].

2.5 Benefit Risk

When the potential energy savings decrease, the expected profit margins of energy-saving projects will reduce, leading to the loss of investment attractiveness to users or policy makers. Expected risk will increase borrowers' borrowing costs, thus threatening the internal costs of energy projects and reduce the overall level of the sources of funding [3].

The energy-saving forecast contains uncertainties, many factors can affect the results of energy-saving investment, thus producing instability. In fact, energy savings results often can't be fully measured, which greatly increases the risk of future investment returns.

2.6 Customer Risk

There are big credit risks in the client side as follows: in the process of contracts implementation, customers conceal or false the energy-saving earnings of the project through various means; customers delay in paying the energy-saving earnings of EMCO; the credit mechanism in China is still imperfect, and poor credit is widespread; there are default risks, for example, once other energy companies offer more favorable terms, customers may break a contract and cooperate with other companies; customers engage in illegal business or other major problems which led to suspension or closure, etc.. These issues can cause the clients' breach of contract, resulting in the loss of EMCO. In addition, whether the customer actively support and facilitate the project will affect the project schedule and budget completion.

3 The Establishment of Risk Evaluation System of Epc Project

Through analyzing the risk of the EPC projects in the implementation process, and combining with the characteristics of EPC style, we established the risk evaluation system of the EPC projects (see Figure 1).

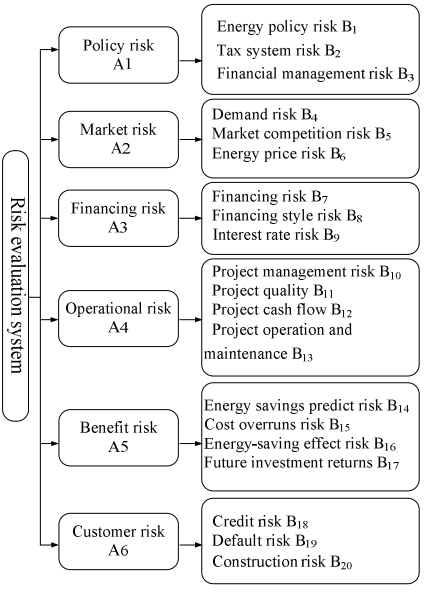


Fig. 1. The target system of risk evaluation on EPC

4 The Establishment of the Fuzzy Neural Network Model

4.1 Establish the Indicators Relative Membership Degree Matrix

1) Determine the Evaluation Grades

This paper sets five evaluation grades, namely lower risk, low risk, general risk, high risk, higher risk.

2) Determine the Sstandard Value Matrix of the Risk Evaluation

In this paper, there are 20 indicators of the risk evaluation, it is difficult to quantify these indicators. Therefore, this paper uses the way of expert evaluation to determine the standard value matrix. That is, based on the characteristics of this project and the expectations of the level of risk, experienced experts determine the expected value of every indicator relative to the above five risk levels, the relative grades of this 20 indicators constitute a risk evaluation standards value matrix $X = (x_{ij})_{20\times5}$.

In the above formula, "x_{ij}" is the evaluation criteria value of the leve "j" corresponding to evaluation factor "i" (i =1, 2,⋯,m; j =1,2,⋯,n).

3) Determine the Test Sample Value Matrix of the Project

According to the actual risks faced by the project, the experts determine the actual risk values of the project indicators on the basis of the above principles. This can constitute a test sample value matrix $Y= (y_{i1})_{20\times1}$.

In the above formula, "y_{i1}" is the value of risk indicator "i" in the set of data (i =1,2,⋯,m).

4) Determine the Relative Membership Degree Matrix of the Standard Value of the Risk Evaluation Indicators

Compare the data between matrix "X" and matrix "Y". Score each indicator in the way of expert evaluation. The relative membership degree of the first grade standard value of indicator "i" which is relative to fuzzy set "low" is "p_{i1}", set "$p_{i1}=0$". The relative membership degree of the "n" grade standard value is "p_{in}", set "$p_{in}=1$". The relative membership degree of the "j" grade standard value of indicator "i" is "p_{ij}". Calculate "p_{ij}" according to the linear interpolation formula:

$$p_{ij} = \frac{y_{ij} - y_{i1}}{y_{in} - y_{i1}} \tag{1}$$

5) Determine the Relative Membership Degree Matrix of the Testing Samples Indicators of the Actual Project

Transfer the value of every risk indicator to the relative membership degree relative to the evaluation grades set. When the value is bigger, the risk is smaller, that is, when $y_{ij} \leq x_{i1}$, $r_{ij} = 0$; when $y_{ij} \geq x_{i1}$, $r_{ij} = 1$; when $x_{i1} < y_{ij} < x_{in}$:

$$r_{ij} = \frac{y_{ij} - y_{i1}}{x_{in} - x_{i1}} \tag{2}$$

According to formula (1) (2), we can construct the relative membership degree matrix "R" of the standard value of the risk evaluation indicators, and the relative membership degree matrix "T" of the testing samples indicators of the actual project: $R = (r_{ij})_{20\times5}$; $T = (t_{ij})_{20\times1}$.

4.2 Establish Fuzzy Neural Network Model

Through Careful analysis of fuzzy neural networks, this paper established a multi-input and single-output network model which has three-tier structure: the first layer is input layer, each neuron represents a fuzzy model input variables, input the evaluation criteria and test samples, and the input of the current layer is equal to the output of the previous layer; the second layer is hidden layer, do weighted calculation of the input, then get the evaluation vector "e" and the evaluation index weight "W"; the third layer is the output layer, take the maximum value of the evaluation vector, then the corresponding grade of the maximum degree is just the risk evaluation grade of the EPC project [4].

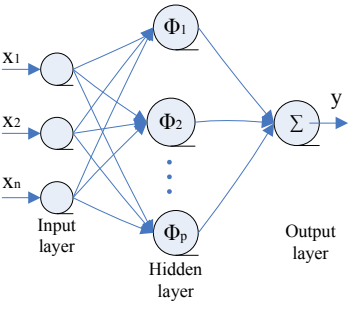


Fig. 2. Neural network structure

The operation procedures of fuzzy neural network model are as follows [4]:

Firstly, normalize the sample vectors, and the input data should be in the range (0, 1). Set random initial value which in the range (-1, 1) to weights and thresholds, and provide a set of inputs and objectives of the samples for the network.

Secondly, calculate the input and output of each unit of the hidden layer and output layer.

Thirdly, the output layer error is calculated based on the output and hidden layer error.

Fourthly, use the error adjustment value to adjust the weights and threshold levels.

$$v_{sj} = v_{sj} + \alpha\delta_s h_j, \omega_{ji} = \omega_{ji} + \alpha\eta_j x_i$$
$$\gamma_s = \gamma_s + \alpha\delta_s, \theta_j = \theta_j + \alpha\eta_j \tag{3}$$

Fifthly, provide the next sample vector for the network, return to the second step, until the global error "E" is less than pre-set value, then the study will be over.

4.3 Network Training

According to matrix "R" and "T", train the network. To improve the training accuracy, it is required a limited time interpolation to matrix "R". The relative membership degree of index "I" of the interpolation sample "K" relative to the evaluation grade "j" is r_{kj}, and $\Sigma r_{kj} = 1$. Set the relative standard grade value of the interpolation sample "K" as "P_k".

$$r_{kj}=r_{ij}+(r_{i(j+1)}-r_{ij})q/c\,, \quad p_k = j + q/c \tag{4}$$

In the above formula, i =1,2,⋯,m; j =1,2,⋯,n; q = 1,2, ⋯,c – 1; the value of "c" can be adjusted to the number of the interpolated samples. Select some samples. The selected samples and the standard sample can be seen as learning samples, the corresponding grade value can be seen as the output sample. The remaining samples can be seen as the network test samples.

Through learning and training of the samples, BP network will have the expert knowledge included in the sample mode, which lays in the weights. Therefore, the trained BP network can be used to make a comprehensive evaluation of the object system. The process is as follows:

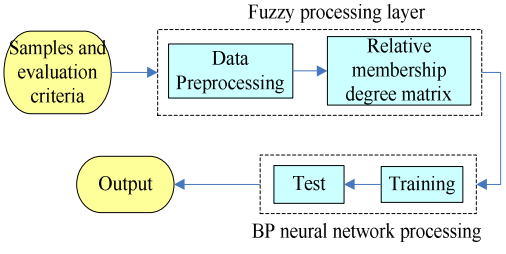


Fig. 3. Process of fuzzy neural network model

5 Empirical Analysis

A machinery company constructs a green lighting transformation project. The company is in great need of electricity for lighting, and its annual lighting consumption of electricity is more than 4,000 million kWh. The company generally uses incandescent, mercury vapor lamps, tungsten light which are of low light effectiveness, high energy consumption and high pollution. The company plans to use advanced energy efficient lighting to replace the original light source, eliminating unnecessary lighting facilities. The following are the risk analysis of its EPC mechanism.

Due to the specificity of the EPC projects, different regions and different enterprises have different evaluation criteria. According to the actual situation of the company, establish the risk evaluation criteria for this project, and get the actual values in the basis of the provided information and its practical operation. As for the qualitative indicators, determine their values through expert scoring method.

Table 1. The standard of risk evaluation

Indictors	Lower Risk	low Risk	General Risk	High Risk	Higher Risk	Actual Value
B1	10.0	9.9	9.8	9.6	9.0	10.0
B2	10.0	9.9	9.8	9.5	9.0	10.0
B3	10.0	9.9	9.8	9.5	9.0	10.0
B4	10.0	9.5	9.0	8.8	8.5	9.9
B5	9.5	8.5	7.0	6.0	5.5	9.3
B6	8.0	6.0	5.5	5.0	4.5	6.4
B7	6.4	6.6	6.8	7.0	7.3	6.5
B8	9.0	8.0	7.0	6.0	5.0	8.6
B9	5.5	4.8	3.7	3.5	3.0	5.1
B10	8.5	8.0	7.8	7.7	7.5	7.3
B11	8.8	7.9	7.8	7.7	7.6	7.5
B12	9.5	8.9	8.5	8.4	8.3	8.2
B13	9.0	7.2	7.0	6.7	6.5	8.1
B14	9.5	8.5	7.5	6.5	5.5	9.1
B15	9.0	8.0	7.0	6.0	5.5	8.4
B16	9.5	8.5	7.5	6.5	5.5	9.3
B17	9.5	8.5	7.5	6.5	6.0	9.3
B18	10.0	9.0	8.0	7.0	6.0	10.0
B19	10.0	9.0	8.0	7.0	6.0	10.0
B20	9.0	8.0	7.0	6.0	5.0	8.6

Through formula (1) and (2), we can get the standard membership matrix “R” of the project success indictors, and the indictor value membership matrix “T” of the actual sample.

$$R = \begin{bmatrix} 0 & 0.1 & 0.2 & 0.5 & 1 \\ 0 & 0.1 & 0.2 & 0.5 & 1 \\ 0 & 0.1 & 0.2 & 0.5 & 1 \\ 0 & 0.333 & 0.666 & 0.8 & 1 \\ 0 & 0.25 & 0.625 & 0.875 & 1 \\ 0 & 0.572 & 0.713 & 0.875 & 1 \\ 0 & 0.236 & 0.517 & 0.648 & 1 \\ 0 & 0.25 & 0.5 & 0.75 & 1 \\ 0 & 0.281 & 0.708 & 0.8 & 1 \\ 0 & 0.5 & 0.7 & 0.8 & 1 \\ 0 & 0.75 & 0.834 & 0.916 & 1 \\ 0 & 0.5 & 0.832 & 0.918 & 1 \\ 0 & 0.72 & 0.787 & 0.92 & 1 \\ 0 & 0.25 & 0.5 & 0.75 & 1 \\ 0 & 0.286 & 0.572 & 0.857 & 1 \\ 0 & 0.25 & 0.5 & 0.75 & 1 \\ 0 & 0.286 & 0.572 & 0.857 & 1 \\ 0 & 0.25 & 0.5 & 0.75 & 1 \\ 0 & 0.25 & 0.5 & 0.75 & 1 \\ 0 & 0.25 & 0.5 & 0.75 & 1 \end{bmatrix}, T = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0.067 \\ 0.056 \\ 0.46 \\ 0.1 \\ 0.235 \\ 0.1 \\ 0.142 \\ 0.21 \\ 0.217 \\ 0.243 \\ 0.1 \\ 0.17 \\ 0.05 \\ 0.056 \\ 0.167 \\ 0 \\ 0 \end{bmatrix}$$

In matrix “R”, the relative membership degree of "low risk" standard value is zero, and the relative membership degree of "high risk" standard value is one. In order to generate more training samples and reflect the significance of the standard value of all risk indicators, do interpolation in matrix “R” with formula (4) to generate more samples. Take c = 5, and generate twenty-one samples, the samples 8, 14 and 17 are test samples, and the others are training samples.

Through the calculation of BP neural network, the values of the three testing samples are 0.3012, 0.3605 and 0.4202, and the errors relative to the actual value are 0.368%, 0.112% and 0.024%, which can fully meet the evaluation grade. Through the above evaluation method, the final result is 0.2148, and the membership grade is one, which means project is of lower risk.

6 Conclusion

In this paper, fuzzy neural network model was used in evaluating the EPC projects, its accuracy and speed are better than the general methods. Therefore, fuzzy neural network model can be a new method in evaluating the EPC projects and is of great practical significance. If modify the evaluation criteria appropriately, the method can also be used in other EPC projects.

The developing process of the EPC mechanism in western countries proves that: it is not only a mature technology, but also has great development potential and market. At present, the EPC projects in China indeed have some problems in the implementation process. Through specific analysis of these problems, we can establish the risk evaluation system and risk evaluation model of EPC project, which

can be seen as the decision making basis when the energy service companies implement the EPC project in China, thus providing development space for the EPC mechanism and energy service companies.

Acknowledgment. Shi Hui, Niu Dongxiao and Wang Hanmei thank the predecessors whose research results pointed out the direction for our paper and provided us with a wealth of reference materials. Meanwhile, thank the teachers and friends who gave a lot of guidance and help for our paper writing.

References

1. Shang, T., Pan, Z.: Energy Performance Contracting Project Risk in Modern Business Administration. Journal of Tianjin University (Social Sciences), 215–217 (May 2007)
2. Sri-Lanka, Model ESCO Performance Contracts, USAID-SARI Energy Program, pp.32-35 (November 2002)
3. Zhou, X., Mo, J.: The Use of the Fuzzy Neural Network in the Post-evaluation of Success Degree of Power Plant Construction Project. China Management Informationization, 91–92 (May 2009)
4. Jang, J.-S.R.: ANFIS: Adaptive-Network-based Fuzzy Inference System. IEEE Trans. on Systems, Man, and Cyberneties, 665–685 (March 1993)

The FPLP Evaluation Model Based on FNN Optimization Design Algorithm

Xinfa Lv

School of Business, Agricultural University of Hebei
Department of Society Science, Hebei College of Finance
Baoding City, China
doctor_lee04@126.com

Abstract. As farm produce occupies a pivotal position in the daily lives, the farm produce logistics accompanied by it has become a rising industry in the logistics industry. Evaluating the farm produce logistics performance (FPLP) and finding the existed problems can improve the farm produce logistics management. To evaluate the FPLP, this paper establishes an effectiveness evaluation system combined with Kirkpatrick model and describes the evaluation mechanism based on fuzzy neural network algorithm. The performance evaluation of 10 cities in Henan province shows that the results given by this model are reliable, and this method to evaluate the FPLP performance is feasible.

Keywords: fuzzy neural network, FPLP, Kirkpatrick model, performance evaluation, indices system.

1 Introduction

Fresh farm produce logistics is the control and management of fresh farm produce, including all the product life-cycle of the procurement, production, processing, storage, distribution, and dispatching. At present, although the transport problems is a key link to solve the poor farm produce logistics, the production and marketing of the fresh farm produce would depend on the efficient operation of fresh farm produce logistics system ultimately.

The development of fresh farm produce logistics in China is rapid, in 2006, our country vegetables output is 582,330,000 tons, the fruit output is 170,500,000 tons; the meats output is 81,000,000 tons, the aquatic product output is 52,500,000 tons, and the milk output is 32,900,000 tons. The partial fresh farm produce is used by the farmers themselves, the major fresh farm produce becomes the circulation of commodities, has formed the giant quantity fresh farm produce. Sell to the places such as Japanese, European Union, American and Hong Kong. But still has the numerous problems during the process of development.

The standardized degree of the logistics facility and the equipment in our country is lower, and the special-purpose technical equipment for the fresh farm produce logistics is short, the technology is backward. In circulation, it is also the primarily

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 383–391.
springerlink.com

shape by normal temperature and natural logistics; lack the cold chain logistics facility and the equipment. According to the statistics, at present, the of the awning truck in the motor transport is 70% in our country, and only then the proportion of the sealed theater box type automobile is 30%, the proportion of the sealed theater box type automobile, the refrigeration machinery, the heat preservation box-type refrigeration vehicles is only 10%. The facilities of the fresh farm produce logistics processing, refrigeration, heat preservation, warehousing are not perfect, cause the rotten deterioration of fresh farm produce for the circulates impeded, increase the cost and the logistics operational risk greatly.

The FPLP performance evaluation is a systematic evaluation process, and a scientific and quantitative argumentation. There are many methods about the performance evaluation have been widely applied, such as: Delphi, the analytic hierarchy process, etc. However, these methods are subject to stochastic factors in the evaluation, and the evaluating results are influenced by subjective experience and knowledge limitations easily, which often with personal bias and one-sidedness. In recent years, with the rapid development of the neural network that has the unique advantages—self-learning, self-organizing and self-adapting ability, it can overcome the influence of subjective factors and has been applied widely.[1][2] This paper will use the fuzzy neural network (FNN) algorithm based on the Kirkpatrick model to evaluate the farm produce logistics performance (FPLP) synthetically.

2 The Fuzzy Neural Network Principle

2.1 The Development of Fuzzy Neural Network

Since the emerging of the fuzzy neural network that integrated the neural network and the fuzzy logic system, its research and the application obtained the swift development.

The international seminar of the neural network and the fuzzy system was sponsors by the NASA in the USA in 1988. Afterward, the related research about the fuzzy neural network was launched in the US, Japan, France, Canada, Singapore, has had the massive achievements. B. Kosko of University of Southern California is the renowned expert who studies the neural network and the fuzzy logic. He proposed the important concept of fuzzy associative memory, the fuzzy cognition chart. The fuzzy neural network's research contents and the structural style are very widespread; the typical fuzzy neural network structural styles include the fuzzy associative memory neural network, fuzzy enormous minimum neural network, fuzzy relationship neural network, fuzzy Hopfield network, regularization fuzzy neural network and so on.

The common union way of the fuzzy logic system and neural network includes: Substitute for the node excitation function in the neural network with the fuzzy logic operator: Neural network weight, input fuzzy and so on. In the fuzzy neural network, the fuzzy neuron's design should enable it to have the approximate same function with the non-fuzzy neuron. But it should be able to reflect the neuron's fuzzy nature simultaneously, has the fuzzy information handling ability. The fuzzy neural network system structure is shown as figure 1.

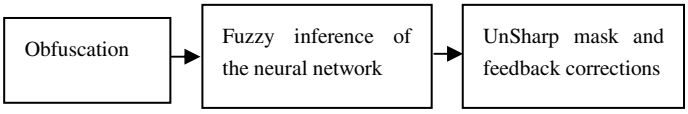


Fig. 1. Fuzzy neural network system structure

2.2 The Parameters Determination of the Fuzzy Neural Network

The fuzzy system of n inputs and single output, if $X \in R_j$, $y=f_j(X)$, among them, $X=(x_1,x_2,\ldots,x_n)^T$, R_j is the part space after input space separated. In this paper, the FNN system is composed of five neural networks, each one uses back-propagation neural network, the input layer of each network has the same input neurons, the number of hidden layer and its neuron can be adaptive to change. Among them, the four BP networks show four fuzzy rules of the function $f_j(X)$ in the conclusion, NN_1 shows "very good effect", NN_2 shows "good effect", NN_3 shows "general effect", and NN_4 shows "poor effect", the four fuzzy rules are corresponding to the different training effect. Another BP network NN_{mf} calculates the membership grade of each input for the four fuzzy rules, so it is four outputs. The output of FNN system is:

$$y = \sum_{j=1}^{4} \mu_j \bullet g_j \tag{1}$$

In this formula: g_j shows the output of four BP network; μ_j shows the membership grade of the four fuzzy rules.[3]-[11]

2.3 The Learning Process of the Fuzzy Neural Network

The study process of the FNN can be divided into two stages: information toward pass and error back pass. In the toward pass, the state of every layer neurons only affects the next layer neurons, if we can't acquire the expected output value in the output layer, then transfer to the back pass. Through revising the weight of every layer neurons, the error signals can return along the former connect pathway, transmit to the input layer by the level and then compute, and return to toward pass process again. The two processes can be applied repeatedly to realize the smallest error. When the error achieves the expected requirements or the prescriptive training numbers, the study process is end.

Suppose given N samples, which are $p_1,p_2, \ldots p_N$, the goal output of network is $d_1,d_2,\ldots,d_N$, the actual output is$y_1,y_2,\ldots,y_N$. When input the i sample, we can acquire output y_{ji} (j=1, 2,…, m), and its error is the sum of every output unit error, the square error indicator function is:

$$E_i(x) = \frac{1}{2} e_i^2(x) = \frac{1}{2} \sum_{j=1}^{m} (d_{ji} - y_{ji})^2 \tag{2}$$

In the formula, e(x) shows error, grads $\nabla E(x)$ and Hessian matrix $\nabla^2 E(x)$ are:

$$\nabla E(x) = J^T(x) e(x) \tag{3}$$

$$\nabla^2 E(x) = J^T(x)e(x) + S(x) \tag{4}$$

$$S(x) = \sum_{i=1}^{N} e_i(x)\nabla^2 e_i(x) \tag{5}$$

J shows jacobian matrix:

$$J(x) = \begin{bmatrix} \frac{\partial e_1(x)}{\partial x_1} & \frac{\partial e_1(x)}{\partial x_2} & \cdots & \frac{\partial e_1(x)}{\partial x_n} \\ \frac{\partial e_2(x)}{\partial x_1} & \frac{\partial e_2(x)}{\partial x_2} & \cdots & \frac{\partial e_2(x)}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial e_n(x)}{\partial x_1} & \frac{\partial e_n(x)}{\partial x_2} & \cdots & \frac{\partial e_n(x)}{\partial x_n} \end{bmatrix} \tag{6}$$

Suppose $x^{(k)}$ show the vector composed of weight value or threshold value through the k iteration, $x^{(k+1)}$ show the vector through the (k+1) iteration, the change of both them is ∇x, and then there is:

$$x^{(k+1)} = x^{(k)} + \Delta x \tag{7}$$

$$\Delta x = -[J^T(x)J(x) + \mu I]^{-1} J(x)e(x) \tag{8}$$

In this formula, proportion factor μ (>0) is constant, I show unit matrix. In practical application, μ is a trial parameter, for a given μ, if Δx can reduce error indicators function, and then μ reduce; Otherwise, μ increase.

Choosing samples and training neural network, the trained neural network can be keep into the repository for evaluating comprehensively the others. We only input the indicator attribute value vector (Matrix) of the object about to evaluate, will receive comprehensive evaluation results (Vector). Therefore, at this time that the neural network can be called neural network comprehensive evaluation model, the connecting weight and internal threshold of the neural network is the model parameters. This paper provides training samples through using fuzzy evaluation samples and evaluation results as model, and evaluates comprehensively manager training effectiveness using the trained model.

3 The FNN Evaluation Model Construction

3.1 The Evaluating Indices System

When evaluating the fresh farm produce logistics performance, we should keep to the systemic, consistency, comparability and feasible principles, build a comprehensive performance evaluation indices system, which are shown as the table 1. The following evaluation indices are taken into account when we evaluate the performance of farm produce logistics performance.

Table 1. The Indices System of the FPLP Evaluationn

One Class Indices	Two Class Indices	Three Class Indices
FPLP performance evaluation	External environmental indices	Market demand index (X_1)
		Infrastructure index (X_2)
		Logistics node index (X_3)
		System environment index (X_4)
		Logistics standards index (X_5)
		Logistics education and training index (X_6)
	Internal flow evaluation indices	Transport cost index (X_7)
		Fully loaded rate index (X_8)
		Loss ratio of transportation (X_9)
		Storage utilization ratio index (X_{10})
		Inventory turnover rate index (X_{11})
		Warehousing costs (X_{12})
		Loading, unloading and handling index (X_{13})
		Circulation processing index (X_{14})
		Delivery index (X_{15})
		Logistics Information index (X_{16})
		Flexible logistics system index (X_{17})
	Overall efficiency indices	Total value of fresh farm produce logistics (X_{18})
		Logistics cost of fresh farm produce (X_{19})
		Loss ratio of fresh farm produce logistics (X_{20})

3.2 The FNN Evaluating Model Construction

We take the 20 indices of describing the FPLP performance as the input vector, and take the corresponding comprehensive testing results as the network expectation output. We take enough samples to train the network, make the relative error to meet the scheduled accuracy after ceaseless learning process. At this time the weight value and the threshold value hold by the neural network is the correct internal denotation acquired by the self-adaptive learning. Once the network has been trained, it could serve as an effective tool to evaluate the performance.

(1) Normalizing the training sample.

(2) Given the training allowable error values ε, β, μ_0, the primary weight value and threshold value vector, cause k=0, $\mu=\mu_0$.

(3) Calculate network output and error indicator function $E(x^{(k)})$.

(4) Calculate jacobian matrix J(x).

(5) Calculate Δx.

(6) If $E(x^{(k)}) < \varepsilon$, then to (8).

(7) Calculate $E(x^{(k+1)})$ using $x^{(k+1)}$ as weight value and threshold value, if $E(x^{(k+1)}) < E(x^{(k)})$, then update weight value and threshold value, cause $x^{(k)} = x^{(k+1)}$ and $\mu = \mu/\beta$, return to (3); otherwise, don't update weight value and threshold value, cause$\mu = \mu \times \beta$, and return to (5).

(8) Stop. [12]-[17]

4 Simulation Experiment

We take the FPLP performance evaluation based on the FNN of 10 enterprises as an example, which are shown in table 2. We use the MatLab to realize the software program, establish the three-layer FNN structure of performance evaluation, the given study accuracy $\varepsilon = 0.0001$, and we select 7 network neurons for the hidden layer. We take 1-6 group evaluation data and evaluation results as the training set, train the network, and carry through the simulation evaluation using the evaluation indices data of the four residual groups and the trained network. In table 3, the network training results and the actual evaluation results are shown. The results in the table 3 show that not only all the training samples is very close to the actual evaluation value, but the results of the four simulation test sets is also very close to the actual evaluation.

To better understand the evaluation results visually, we staple the evaluation results. The evaluation scores between 0.9-1.0 show that the manager training effectiveness is very high; the scores between 0.8-0.9 show that the effectiveness is high; the scores between 0.6-0.8, we say that the effectiveness is general; the scores below 0.6, we say that the effectiveness is low.

Table 2. The FPLP Evaluating Data

No.	X_1	X_2	X_3	X_4	X_5
1	0.658	0.669	0.580	0.352	0.242
2	0.478	0.544	0.596	0.546	0.458
3	0.569	0.742	0.554	0.432	0.752
4	0.966	0.865	0.746	0.478	0.723
5	0.421	0.245	0.687	0.698	0.354
6	0.254	0.247	0.254	0.244	0.852
7	0.147	0.687	0.457	0.225	0.574
8	0.247	0.855	0.655	0.544	0.784
9	0.406	0.565	0.577	0.145	0.856
10	0.556	0.856	0.745	0.856	0.568

CONTINUED TABLE

No.	X_6	X_7	X_8	X_9	X_{10}
1	0.955	0.963	0.965	0.752	0.962
2	0.245	0.457	0.632	0.441	0.245
3	0.755	0.354	0.431	0.622	0.651
4	0.684	0.675	0.852	0.745	0.774
5	0.752	0.752	0.654	0.574	0.455
6	0.445	0.569	0.258	0.856	0.963
7	0.698	0.254	0.552	0.475	0.478
8	0.855	0.689	0.698	0.633	0.528
9	0.963	0.412	0.756	0.885	0.654
10	0.421	0.698	0.741	0.475	0.447

CONTINUED TABLE

No.	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
1	0.241	0.454	0.700	0.147	0.789
2	0.524	0.744	0.440	0.144	0.447
3	0.368	0.568	0.458	0.158	0.496
4	0.225	0.247	0.365	0.855	0.478
5	0.455	0.687	0.544	0.456	0.355
6	0.869	0.753	0.405	0.753	0.456
7	0.456	0.357	0.800	0.951	0.903
8	0.887	0.473	0.863	0.406	0.700
9	0.445	0.823	0.456	0.905	0.520
10	0.897	0.445	0.850	0.753	0.420

CONTINUED TABLE

No.	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}
1	0.500	0.801	0.756	0.780	0.805
2	0.412	0.230	0.445	0.496	0.426
3	0.665	0.756	0.265	0.963	0.478
4	0.774	0.450	0.478	0.536	0.921
5	0.569	0.604	0.330	0.753	0.479
6	0.452	0.853	0.520	0.745	0.862
7	0.863	0.741	0.541	0.420	0.575
8	0.412	0.805	0.753	0.564	0.632
9	0.880	0.632	0.963	0.756	0.441
10	0.963	0.500	0.456	0.541	0.802

Table 3. The Actual Evaluation Results Compared with the Network Training Results and Classification

No.	Actual evaluation Results	Network training Results	FPLP
1	0.5445784	0.54414558	Low
2	0.8562222	0.8624445	High
3	0.8962211	0.8962712	High
4	0.4125555	0.4125789	Low
5	0.9621111	0.96214785	Very high
6	0.4558111	0.4558261	Low
7	0.7511416	0.7511478	General
8	0.7569554	0.7569357	General
9	0.9135544	0.9135789	Very high
10	0.8354419	0.8354415	High

5 Conclusion

FNN is an organic combination of the neural network and fuzzy logic systems, which can centralize the both advantages of learning, association, identification, adaptation and fuzzy information processing, and improve the learning and expression ability. In terms of FPLP performance evaluation, this method enhances convergence rate and prediction accuracy of the network system, and has a very good application prospect in the other evaluation field. In addition, the FPLP performance evaluation not only relies on the network model, but also depends on the FPLP performance parameters, especially in terms of the FPLP complexity, which are the difficulties of evaluating the FPLP performance. Therefore, FNN model should be further improved and perfected in the practical application and research.

References

1. Chen, Y., Ye, J.: Development situation and countermeasure research of fresh farm product logistics in Henan province. Economist (11), 271–273 (2006)
2. Zhao, Y.: The farm product logistics evaluation indices system construction in China. Commercial Research (1), 211–213 (2007)
3. Wang, P., Li, Y.: Fuzzy neural network research development. Journal of Rizhao Polytechnic 2(3), 31–34 (2007)
4. Mamdani, E.H.: Application of fuzzy algorithms for simple dynamic plant. Proc. IEEE (121), 1585–1588 (1974)
5. Maiers, J., Sherif, Y.: The application of fuzzy control system. Automat. 11(2), 136–142 (1972)
6. King, P.J.: The application of fuzzy control system to industrial processes. In: IFAC World Congress, Mitt, Boston, vol. 13(3), pp. 235–242 (1975)

7. Mamdani, E.H.: An experiment in linguistic synthesis d1a fuzzy logic controller. Int. J. Man Math. Studies (70), 1–13 (1975)
8. Det, D.: Fuzzy logics and the generalized modus pones revisited. Lybem Syst. (15), 3–4 (1984)
9. Takagi, T., Sugeno, M.: Fuzzy identification of system and its applications to modeling and control. IEEE Trans. on SMC (15), 116 (1985)
10. Buckley, J.J., Hayashi, Y.: Fuzzy neural networks: a survey. Fuzzy Sets and Systems (66), 1–10 (1994)
11. Jang, J.R.: Selflearning fuzzy logic controllers based on temporal back-propagation. IEEET Trans. Neural Networks 3(5), 714–723 (1992)
12. Wang, L., Mende, J.M.: Generating fuzzy rules by learning from examples. IEEE Trans., Neural Networks 40(12), 1320–1336 (1992)
13. Jang, J.R.: ANFIS: adaptive network based fuzzy inference system. IEEE Transactions on Systems, Man, and Cyemetics 23(3), 665–682 (1993)
14. Numberger, A., Nauck, D., Kruse, R.: Neural fuzzy control based on the NEFCON model: recent developments. Soft Computing (2), 168–175 (1999)
15. Nauck, D., Kruse, R.: Neural fuzzy systems forfunction approximation. Fuzzy Sets and Systems (10), 261–269 (1999)
16. Machado, R.J., Rocha, A.F.: Evaluative fuzzy neural networks. In: Proc. First IEEE International Conference on Fuzzy Systems, RJZZ-IEEE 1992, San Diego, pp. 493–500 (1992)
17. Wang, P., Li, Y.: Fuzzy neural network research development. Journal of Rizhao Polytechnic 2(3), 31–34 (2007)

Evaluation on Multi-path Routing Protocol of Ad Hoc Networks

Yang Yan

Electronic and Electric Engineering Institute
Nanjing Institute of Industry Technology, Nanjing, China
yangy@niit.edu.cn

Abstract. Mobile Ad hoc network is a mobile, multi-hop wireless network; its main features are dynamic network topology, limited bandwidth, no pre-construction of basic network equipment etc. In such network, the design of network routing protocol is the key issue. Ad hoc network protocols requires the rational use of bandwidth, reduce the cost of establishing and maintaining routes. Quickly find the path to reduce the transmission delay; prevent the routing loop occurs. This article focuses on the realization of multi-path routing methods in ad hoc network and their impact on network performance changes. In the study, we selected currently representative multi-path routing protocol to study and compare, carried on in-depth analysis characteristics and the change of several different methods to achieve multi-path network performance.

Keywords: Ad Hoc, Multipath Routing Protocol, AODMV, AODV－BR.

1 Introduction

Ad Hoc network is a mobile network without wired infrastructure to support, nodes in the network is constitutes by the mobile host. Ad Hoc network was originally used in military field, its research originated in the battlefield environment, data communication packet radio network project funded by the DARPA and subsequently, in 1983 and 1994 tested to destroy the network of anti SURAN (Survivable Adaptive Network) and Global Mobile Information Systems GloMo (Global Information System) research project. As wireless communication and terminal technology continues to evolve, Ad Hoc Networks in civil environment has also been developed, such as the need for wired infrastructure in the absence of the temporary communication area; it is easy to set up Ad Hoc Network.

In the Ad Hoc network, two mobile hosts within range of each other's communication, they can communicate directly. However, due to the mobile host's communication range is limited, if the two are far apart to communicate with the host, you need to move between them to achieve the forwarding host B. Therefore, in Ad Hoc network, the host is also a router, responsible for routing and forwarding packets to find work. In the Ad Hoc network, each host communication range is limited, so usually by the multi-hop routing components, data forwarding through multiple hosts to reach the destination. Therefore, the network was also known as the Ad Hoc Multi-hop Wireless Network.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 393–400.
springerlink.com

2 Ad Hoc Network Routing Protocols

Ad Hoc Network Design is a key issue in the development of two nodes can provide high quality and efficient communication between the routing protocols. Mobile network nodes make changing network topology. Traditional routing protocols based on Internet can not adapt to these features, you need to have a special Ad Hoc networks used in routing protocols. According to earlier on the Ad Hoc network architecture and features described, the design of the routing protocol must meet the following conditions.

1) The need for dynamic network topology with a rapid response capability, and try to avoid routing loops from occurring, to provide easy and convenient location in the network nodes.

2) The need to efficiently utilize the limited bandwidth resources, as compression of unnecessary overhead.

3) Implementation of multi-hop communication is limited number of intermediate relay, generally no more than 3 times.

4) The need to minimize the launch time and launch of data, limited work to save energy.

5) Under possible conditions, it makes the routing protocol with security, reduce the likelihood of attack.

In general, ad hoc networks can be applied to the three basic types of routing protocols: table-driven; demand-driven; mixed type.

In table-driven routing protocols each node constantly updated to any other node in the network routing information, so that the establishment of routes, establishment of latency will be smaller. However, frequent topology changes for the network; this mechanism requires a lot of resources to maintain this correctness and reliability of the routing table.

On the demand-driven routing protocols, nodes only need to maintain the route to reach the destination node can be active. When the node needs to communicate with the new destination node, the routing protocol will start a routing process. Thus communication used to update the topology changes reduced, the route search establishment delay will be increased. Another fast-changing topology may break an existing route and the route selection process leading to new.

Ad hoc wireless networks using simple table-driven or on-demand routing protocols can not completely solve the routing problem. Therefore, many scholars have put forth with the table-driven or on-demand routing protocol advantages of hybrid routing protocols such as ZRP protocol. ZRP protocol is a table-driven or a combination of on-demand routing protocols, all nodes within the network to them as the center has a virtual area. Number of nodes in the region depends on the zone radius, therefore areas are overlapped, and this is the difference between clustering routing.

Reduce the frequency of initiating the development of routing for Ad hoc network routing protocol design is important. Rational design of multi-path routing protocols can reduce the frequency of initiating the development of routing. Moreover, the

rational design of multi-path routing protocol can effectively improve the throughput and reduce network transmission delay. Therefore, the study is of great significance to many routing protocols.

3 Ad Hoc Network Multi-Routing Protocol

Now most routing protocols are designed based on an ordinary single-path routing protocol. The majority of multi-path routing protocol in this article also uses a single path routing protocols as the basic established method. AOMDV agreement protocol and AODV-BR is generated based on the AODV protocol. For multi-path routing, the link disjoint and node-disjoint is two important concepts. The introduction of these two concepts is established between two points to show the number of different paths of independent (separate) extent. Figure 1 described in detail the path of isolation and separation of the node respectively. The two paths shown in Figure 1 ABCDE and AGFE addition to the source and destination node, there is no common parts, saying the path for the node separating the two paths, and for two paths in the figure ABCDE and ABCFE, although they have a common node C, but did not share the link, saying that the path to link the two separate paths.

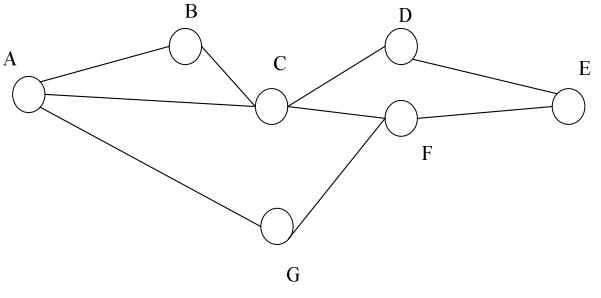


Fig. 1. Isolation and separation of the link node diagram

In a multi-path, the path of separation is a very important aspect, because the more independent the path is, the more effective resources it may provide, that is, the performance influence of a path to another is less. The stronger the independence of the path is the more full use of network resources by multi-path routing. Increase the effective bandwidth between nodes, and reduce the degree of network congestion to reduce packet loss rate. Node must also be independent of the multi-path.

Be involved in the routing protocol to determine the two paths are link-disjoint or node-disjoint problem, for the use of source routing for multi-path routing methods, the use of source routing approach, the source and destination node stores all paths routing information, only eleven could be compared to determine the path of the separation of the two, but did not use source routing for multi-path routing method, the source and destination node can not access the path information, you need to

establish the route by the node to take certain measures to ensure the path to the degree of separation. Specific methods of the relevant agreements are described in detail in CHAMP agreement AOMDV narrative.

The establishment of multi-path routing is the first multi-step routing protocols, establish the route, the route maintenance and data transmission is also a multi-path routing protocols are important to design the content. The specific design of these two multi-path routing for different methods and the establishment of different, but the basic idea is nothing more than the following two. One is after the establishment of multi-path routing data can be transmitted only in one path, the path of the other as a backup path, the path of the data to be transmitted error in the link, you can continue through the backup path transmission, the other is a few shall establish a parallel path to transmit data, but this transmission has a problem to be solved - different paths how to share the load of data transmission, data transmission by the different ways that multi-path routing protocols are different methods were proposed.

3.1 AOMDV Protocol

AOMDV established AODV routing the same way. In the process of establishing the routing, AODV retain only one route, other routes may be established, AODV protocol to the serial number and stored in the RREQ's hop count information in the block out. In AOMDV in order to establish a multi-path routing, but set some conditions to meet the conditions of the route is preserved. These include two aspects: first, loop-freedom; second, link-disjoint.

Similarity with the AODV protocol to prevent loops, AOMDV protocol uses two values as criteria to determine whether to establish routes, including: the purpose of the serial number (destination sequence number) and the advertised hop count. At the same time using the next hop and last hop to determine the link disjoint.

If a path entry in the routing table for each node all the nodes in the next hop and last hop are not in any of the same, then this path is the path links with other separation.

When a node receives a message need to establish the route, first compare the serial number, serial number when the message is large, then writes the routing information, if you need to compare equal to the value advertised hop count, if the message more small, it will not constitute a loop, then compare the next hop and last hop, if not the same value, establish the route.

Establishment of each node routing a message is received, it will carry out such steps. The above mentioned steps are key steps for the establishment of multi-path routing.

Following is the AOMDV to establish the main aspects of the routing process. It establishes the basic principle and process of routing and AODV protocol is similar, but in order to achieve the purpose of establishing a multi-path routing and meet the requirements of some unique ways.

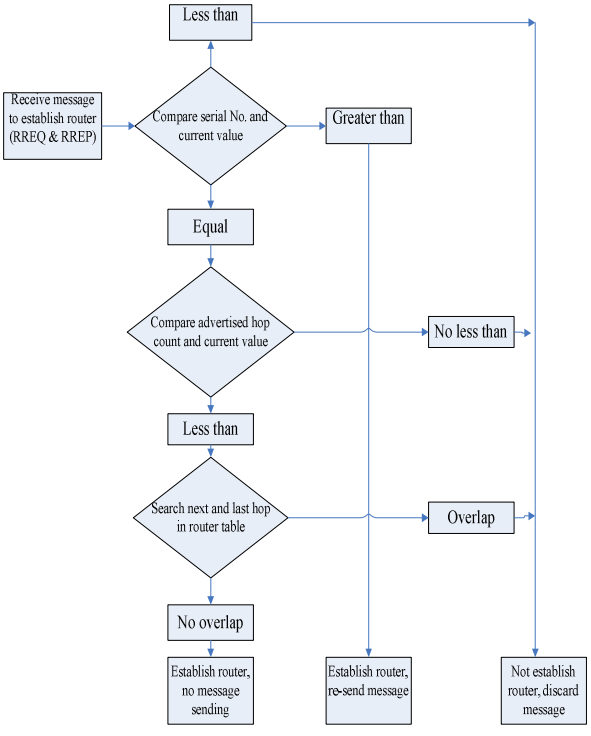


Fig. 2. Process of AOMDV protocol node receives reverse routing message

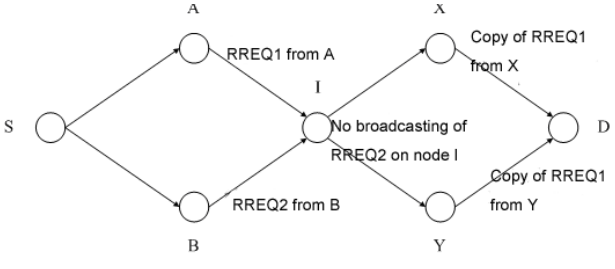


Fig. 3. Diagram of routing cut effect

3.2 AODV-BR Protocol

AODV-BR and the establishment of the basic AODV routing protocol process of the same, but in addition to establishing a path from the source to the destination of the main routes, should also create a series of backup routes. Specifically, the path is a node not in the primary monitor (overhear) of its adjacent nodes (the nodes on the path should be the primary node) issued by the RREP, if listening to the record for the

purpose of the node adjacent to the next hop node in order to build a backup route. There is a problem, a node is likely to listen to more than one or as many nodes with a main path sent to the RREP, the node under such circumstances choose the best quality RREP packet to send a backup route node. By this method, establish the main routes, while also set up the backup route.

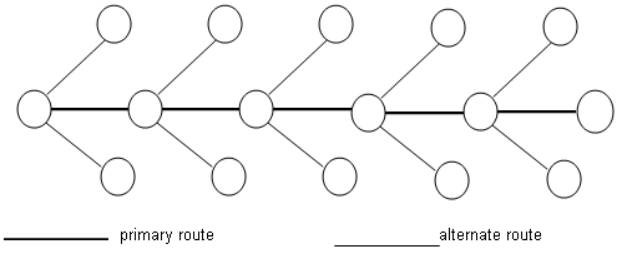


Fig. 4. Diagram of AODV-BR main route and replace route structure

When the primary path of some of the link failure, link down at the node broadcasts a hop limit the transmission of data packets, the header that the link is broken, when the establishment of a backup route to the destination node receives the node When the data packet to hop through the establishment of a backup route, and the back of the main path to send packets to the destination node. Nodes on the main route the data packets are not retransmitted.

In fact, we can see that the maintenance of routing and save the package (salvage packet) the corresponding process and DSR are very similar thought processes. And DSR are not used just because the same source routing, different specific methods to achieve the effect are basically the same. AODV-BR by AOMDV and a way of protecting the comparison of routing, we can see that the rescue package used in DSR also used the idea of these two protocols, this rescue package may well be thinking of a mature and reliable thinking of the rescue package.

4 Analysis of Multipath Routing Protocol

In front of this section will attempt several routing protocols to make a systematic comparison. In fact, the process described above has been described in different protocols and a simple evaluation of some aspects of several agreements made some rough comparison. Table 1 compares the similarities and differences between the two methods of various routing.

Multi-path concept and meaning of separation have been described before. Routing protocol discussed in the resolution establishing the multi-path is different, AOMDV agreement to establish the multiple paths the minimum separation is the separation of the link. Although AOMDV agreement establishing a multi-route when the route is a link established to ensure the separation, but these were likely to be nodes in the path separation, because separation of the two path nodes must be separated from the link.

Table 1. Comparison of multi-path routing protocols

	complete multi-path	*source routing*	*loop-free*	*independence*	*transmission*
AOMDV	Yes	No	Yes	separation of link	single path,backup
AODV－BR	No	No	Yes	uncertain	single path,backup

Implementation of the multi-path routing process, due to multiple paths to ensure separation between the need to judge the node, the node can not just know the next hop information, but also need to know the path to more comprehensive information for the use of source routing protocol for multi-path routing, the source and destination node know all the information of all the paths, the path to judge the degree of separation is relatively easy; the AODV protocol for the use of a similar establishment in the routing table entries in the node, and do not know the whole path of a node routing approach to route discovery packets required for additional information in the record, the link is used to determine the degree of segregation, to be carried out to determine more accurate, more information is needed in AOMDV agreement records the only other path to the final hop information, to judge is limited to ensure that the path is link-disjoint.

Transmission of data on backup path is small, relatively easy to apply some. The multi-path parallel transmission of the program will encounter in the application of more complex problems. For example, how to distribute the traffic between the paths, a path failure to resume its path after the other cached data, the destination node receives the data to ranking. To solve these problems, need more information on the network, not only is the path of structural information, including QoS parameters.

Multi-path routing is maintained after the establishment of important factors affecting network performance. Broken link failure at the node processing the packet buffer method used to relatively fixed, the agreement basically several DSR protocol using a similar rescue package of programs, first in the routing node can be used to find replacement route, if there is, through its transmission line in the packet, not the route to re-initiate the process of building. Multiple paths should be followed in the handling of an error, is the only path to remove the broken, the use of other paths to transfer, or re-initiate route discovery process, all the routing information update is to discuss the issue. This paper analyzes the routing protocol, AODV-BR SMR agreement protocol and can be disconnected on a path to re-initiate route discovery.

5 Conclusion

Ad hoc network through a variety of single-path routing protocol, it is found to establish multiple paths as independently as possible, and at the same time using the appropriate route maintenance and data transmission strategies to improve the network performance is very obvious. Frequently initiated for the establishment of mobile Ad hoc routing networks, multi-path routing method should pay attention to make full use of the routing path discovery process that information, a number of paths is established, it should make full use of multiple routes established full use of network bandwidth to achieve better performance.

References

1. Kourkouzelis, D.: Multipath Routing Using Diffusing Computations. Master Thesis (1997)
2. Lee, S.-J., Gerla, M.: Split Multipath Routing with Maximally Disjoint Path in Ad hoc Networks
3. Valera, A., Seah, W.K.G., Rao, S.: Cooperative Packet Caching and Shortest Multipath Routing in Mobile Ad hoc Networks
4. Ad hoc On-Demand Distance Vector (AODV) Routing IETF- rfc3561
5. The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR) draft-ietf-manet-dsr-10
6. Marina, M.K., Das, S.R.: On-demand Multipath Distance Vector Routing in Ad Hoc Networks. In: Ninth International Conference on Network Protocols, November 11-14 (2001)
7. Wu, K., Harms, J.: On-Demand Multipath Routing for Mobile Ad Hoc Networks
8. Nasipuriand, A., Castañeda, R.: Performance of multipath routing for on-demand protocols in mobile ad hoc networks
9. Chen, W.-P., Hou, J.C.: Dynamic, ad-hoc source routing with connection-aware link-state exchange and differentiation; Global
10. Alleyne, A.: Physical insights on passivity-based TORA control designs. IEEE Transactions on Control Systems Technology 6(3) (May 1998)

A Novel Network Communication Model Utilizing Windows Sockets

Quanyong Yu

Mudanjiang Normal University

Abstract. The network communication model described in this paper is different from the general model of communication network introduced in many materials. According to the MES's function requirement the transmission of information is accomplished by the Cmsg class. The message transferred in General network communication model is normally a character string, while more abundant information is transferred by means of the Cmsg class in the network communication model of the MES simulator, including the communication purposes of the both sides of communication.

Keywords: Windows Socket, network communication, model.

1 Introduction

Using the simulator for teaching is an important step to improve the students' practice operation ability in maritime college, the mobile earth station (MES) simulator plays a more and more important role in the teaching of GMDSS (global maritime distress and safety system) experiment and operation.

2 Simulator Network Communication Function and Communication Process

MES simulator of network communication function is divided into two parts: the sending and receiving of the data file and Interactive communication between two MESs involved.

- Editing message or opening the programmed message.
- Select "Send FILE" in the "FILE" menu for Sending and receiving data between a MES and a land user, two MESs or select "conversation" in the "FILE" menu for interactive communication between two MESs.
- Input business code, ocean code and subscriber identifier in the popup dialog. And press "OK" button to pop-up call dialog box (see Fig.1), or press "look up" button to popup "Select Subscriber" Dialog box, press the "select" button to Select MES users.
- The data terminal begin to call subscriber through the LES (land earth station). The information of the LES and the answer back information of the subscriber

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 401–407.
springerlink.com

being called are displayed if the Calling process is successful. Then the communication window is opened. Otherwise, the calling failure window is displayed.

- If for sending and receiving data the communication process is shown on the display screen, the message is converted into uppercase characters and transmitted. If for interactive communication between two MESs, the source and destination MESs can make an interactive dialogue via keyboard. The communication process is shown on the display screen.
- Press" ... "(five dots) to end communication after sending messages or finishing a dialogue. Then the answer back of the other side and the duration time being transmitted by LES are displayed on the screen.

3 The Function of the Windows Sockets

The Socket [1] is the key of network communication and is the basic operation unit of network communication being supported by TCP/IP protocol [1]. We can think socket as two-way communication terminals of different hosts' processes. Windows Socket supported by multiple protocols is a network programming specifications and interfaces in the windows operation system. It is open and widespread used. Windows Socket includes not only the Berkeley Socket style's library functions which are familiar to people but also a group of windows expansion library functions. Programmers can make full use of the Windows driving mechanism to program.

According to the characteristics of network communication, Windows Sockets can be divided into two classes: the Stream socket and the Datagram Sockets. The network information transfer mode of the Stream socket is connection-oriented, error-fee, data stream length unlimited, TCP protocol families use these interfaces.

The network information transfer mode of the Datagram socket is No connection and the message is divided into some independent packets before transferring Data cannot ensure the same order during the transferring period, and reliability and repeatability could not guarantee.

The Stream socket is selected in the MES simulator network communication mode described in this paper. The data and telex transmission, the information broadcast of LES can be made identical to actual situation.

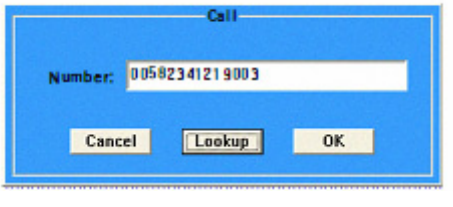


Fig. 1. Call dialogue box

4 Windows Sockets Program

Microsoft Foundation Class provides two Windows Sockets [2] encapsulation classes: CasyncSocket and Csocket, CSocket is the derived class of CasyncSocket and provides more advanced abstraction to apply socket for CArchive object. The use of CSocket class is easier than CasyncSocket class, Sending and receiving data operation become simple and clear because transferring data is managed by CArchive class and CsocketFile class. A CSocketFile object is a CFile object of sending and receiving data used by Windows Sockets in the network. Once the CSocketFile object establishes connection with CSocket object and CArchive object, the programming method of a CSocketFile object is similar to that of a general CFile object, in other words, to send and receive data by means of CArchive object.

The general programming model to utilize CSocket object for communication[2] between customers and server is:

- Constructing the client and the server socket object respectively.
- Invoking the Create function of the object to Create a socket, and the create function call the Bind function and to bound the socket to the specified address. Note to create a socket for the server is needed to designate a port number.
- After creating the socket, the server calls the member function of listen and starts detecting customer's connection request, and the customer can call the member function of connect to ask the server for link.
- When the server detects the client's connection requests, a new socket will be created and be transferred to the member function of accept for receiving client's connecting requests, specific error code will be returned when the function operation is unsuccessful.
- Creating a related CSocketFile object for the client and the server respectively.
- Creating a Carchive object associated with CSocketFile to send and receive data for the client and the server socket object respectively.
- Transferring data between the client and server socket utilizing the CArchive object.
- Destroying CArchive, CSocketFile, and CSocket objects after the completion of the task execution.

5 Inmarsat Mes Simulator Communication Mode

5.1 The establishment of Network Communication Model

Programming timing sequence is shown in Fig. 2.

1) On the server: First, creates a CServerDoc: mySocket for ClisteningSocket's object, and invokes member fnction of m ySocket-> listen 0 to detect customer's connect requests. When the server monitors the client (i.e. MES simulator) connection requests, executes CListeningSocket OnAccept (int nErrorCode), invokes CServerDoc::ProcessPendingAcceptO, Creates a new socket CClientSocket pSocket

* and be transferred to the member fnction of Accept for receiving customer's connect requests(m_pSocket - > Accept pSocket (*» , calls CClientSocket::InitO, Creates CSocketFile object and two Carchive objects, m _pArchiveIn and m yArchiveOut, and to realize the sending and receiving of data.

2) On The client: When the MESSIM (the main program of the MES simulator) starts-up, Retrieves the server IP address from the database, executes ConnectSocket(m _From, to, strServer, nChannel), Creates client Socket object m _pSocket, invokes connect member fnction and requests connect to server. Afer the successfl connection to the server, the CSocketFile object, CArchive object: myArchiveln and myArchiveOut be created, and the successfl link message be transferred to the server immediately.

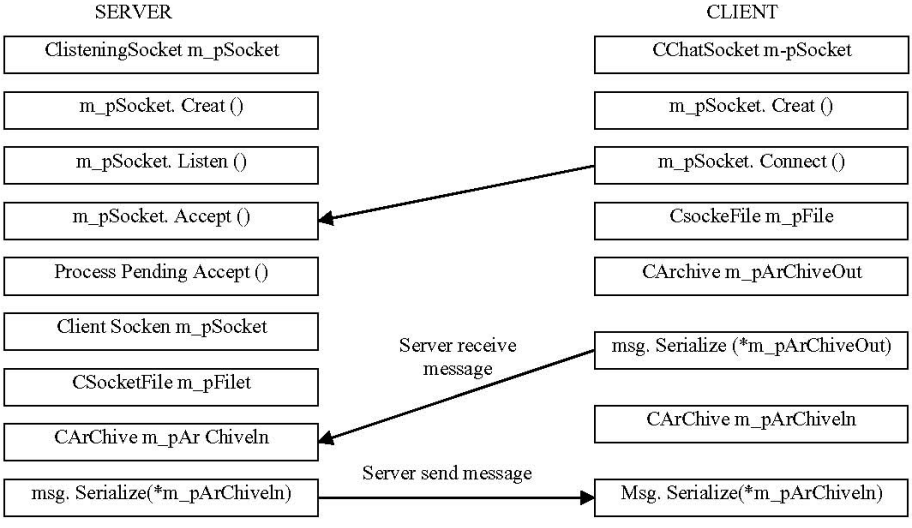


Fig. 2. Diagram of network communication program timing sequence

5.2 The Implementation of the Main Communication Process

The m communicatingstatus is a global variable defned in MESSIM.h, its Data structure is:

```
typedef struct tagCOMST    ATUS    {
BOOL conversation;
BOOL active;
BOOL telex;
BOOL distress;
} COMST ATUS;
} COMST ATUS;
```

It describes the state of the MES. Four Boolean quantities are Real-time dialogue, receiving and sending, telex or data and distress. The relations between the document

and socket must be established because the MES opens a document in real-time dialogue, telex and distress. The socket object defned in MESSIM.h, CSocketFile object and two CArchive object pointer variable be assigned to the member variable of C MESSIMDoc to implement the storage and display of the communication documents when the main program initiates.

The message class CMsg is defned in the project, some member variables be defined as:

```
Public:
    CString m_strText;
    BOOL m_bClose;
    CStringList m_ msgList;
    BOOL m_Telex;
    CString m_To;
    CString m_From;
```

Its member variables define the abbreviation of the source and destination MESs, the transmitting information and Boolean quantity of telex state, etc.

3) The data sending between two MESs: After building the connection between the source MES and the sever socket, edits message or opens the programmed message, select "Send FILE" in the "FILE" menu call dialog box, press the "select" button to Select the MES user, the call initiates and the call progress dialog box popup.

The MES simulator makes use of the access database management system, adopts OOBC [3] (open database connectivity) for Database access. OOBC is standard interface used in a relevant or irrelevant database management system to access data. It is an element relevant with the database in the Microsoft Windows open service architecture, WOSA. OOBC provides a standard interface for client applications to visit different database, builds a group of database access specification, provides a standard OOBC API for different database and supports the SQL language. Users can transfer SQL statement directly to the OOBC API. The OOBC class of MFC packages complicated OOBC API, and offers simplified invoking interface and facilitates the development of database application greatly.

The CRecordset object of MFC OOBC class indicates a group of records chosen from a data source and be called the recordset. The CStationsSet class is the Crecordset's derived class, the member variables defined by the CStationsSet class is:

```
public:
    int m_Region; // the ocean region of the satellite
    CString m_Mark; // selected mark
    CString m_Code; // code of satellite LES
    CString m_ Name; // name of satellite LES
    CString m_ Capabilities;
    CString m_RCCAnswerBack; //answer back
```

After Popup the Call Progress dialog box, data sending begins. LES in the database will be found by means of Crecordset object three seconds late. The ocean region, code and name of satellite LESs displayed on the call progress dialog box, this means

the MES be connected with the satellite. The CterrestrialSubscriberSet record set make a note of the MES subscriber's information, the member variables defined by the CterrestrialSubscriberSet class is:

```
public:
    CString m_Number; //MES abbreviation
    CString m_ AnswerBack; //answerback
    CString m_ComputerName; //name of MES
    CString m _ ComputerIP; // IP address of computer
```

timing 2 seconds again, the answerback of the source MES and the name of the destination MES in the database will be found by means of CterrestrialSubscriberSet, and be transferred to the server via the StartCommuniacting (CString& to, CString& answerback) member function of the Global object the App. Wait for the destination MES response in five seconds and to set up the MES's states, telex and active, are true; If the destination MES doesn't response in 5 seconds, communication failures dialog box displayed and communication process terminated. After receiving the message from the source MES, the server broadcasts the news to all sockets. The destination MES will send response signals to the source MES when it receives messages belongs to itself.

The source MES does the next processing according to the different state in which it is situated after receiving the server's message. The real-time dialogue communication between the two MESs will be processed if the m_communicating status. Conversation is truth. If the conversation and the m_communicating status.active are false, the MES data receiving will be processed. The source MES is situated in the state of the transmitting data If the telex and the m_communicating status.active are truth, and makes the process of transmitting data when receiving the destination MES's answerback. The ReceiveMsgO member function of the global object, the App, implements the above functions.

The answerback of the destination MES will be displayed on the Call Progress dialogue box when receiving the answerback of the source MES. The connection be made between the two MESs involved and the Call Progress dialogue box closed. The contents of active windows and the data generated in the communication be write to the CMsg object and sent to the server by the socket.

4) The data receiving between to MESs: When the socket connection between the source and the destination MESs be made and the source MES receives the message from the server, a new window be created and dispays a new file. The new document receives the data sent from the calling MES, creats the link between the documents and the socket. The Socket object, CSocketFile object and two pointer variable of the CArchive objects be assigned to the member variable of the CMESSIMOoc document class. Implementation of the FirstReceiveMsg (CMsg MSG) member function of the CMESSIMOoc document class, the function displays the received information (i.e. the answerback of the source MES), and sends itself answerbak to the source MES and ready to receive data.

The ReceiveMsgO function of the CMESSIMOoc responses when the receiving MES receives the message the second time from the server. One character be displayed in the window in every 100 milliseconds if the message received from the server is data, and the duration time and the answerback of the satellite LES be

displayed in the window if the received message from the server is the break information (i.e. Five " ... "). Then the communication window is closed, the data receiving program is over and the files generated in the receiving process are stored.

5) The conversation communication between two MESs: When the socket connection between the two MESs involved be made, selects "Conversation" in the "FILE" menu, the call dialog box popup, presses the "select" button to Select the MES subscriber and the call progress dialog box popup. After timing 3 seconds the satellite LES in the database is found and the link between the MES and the satellite LES is made, the ocean region, the code and the name of the satellite LES are displayed in the Call Progress dialogue box. The answerback of the calling MES and the receiveing MES's name are sent to the server 2 seconds late, the active and the conversation state of the calling MES's are set true, and waits the answer from the receiving MES in five seconds. The communication failures dialog box popup and the communication process terminated If no response made from the receiving MES in 5 seconds. The receiving MES send itself answerback to the calling MES when it received the answerback of the calling MES. The document for communication and the window are established when the two MESs involved receive the answerback from each other. The OnCharO function of view class CMESSIMView responds the user's keyboard input and the characters typed from the keyboard send to each other. The function of the conversation communication [4] realizes.

6 Conclusion

The communication between a MES and a land user, two MESs, are the core contents of the MES simulator. The communication model discussed in this paper uses the client server mode.

References

1. Wang, X.P., Zhong, J.: Visual c++ network communication protocol analysis and application realization, pp. 22–130. People post and telecommunications press of China (February 2003)
2. Liao, G., Zhou, J., Liu, R.: Sockes program realized by Visual C++6.0. Electronic Computer Program Technique and Maintain, 18–22 (March 2000)
3. Tang, S., Gou, X., He, Z.: Applications of Dynamic Link Library to Simulator Program Designing. Jounral of System Simolation, 81–83 (January 2002)
4. Yin, Y., Jin, Y.-C., Li, Z.-H.: Network communication in distrubution navigation simulation system. Journal of System Simolation, 621–624 (June 2000)
5. Gou, X., He, Z.: Network Communication Design for Simulator Based on Windows NT. Jounral of System Simolation, 78–81 (March 1999)

A Novel Background Dynamic Refreshing in Road Traffic Monitoring

Zhu Cheng, Meichen Zhou, and Fei Zhu*

School of Computer Science and Technology, Soochow University
Suzhou, China, 215006
{0827402057,0827402084,zhufei}@suda.edu.cn

Abstract. Traffic is necessary in modern society. Road traffic monitoring is especially fundamental. Unfortunately, road traffic jam is so common nowadays. As a result, road traffic monitoring is getting more and more important. In the research and application of monitoring, a method called background subtraction is often used. This article aims to optimize the approach of refreshing background in road traffic monitoring while using background subtraction method. We study typical methods for refreshing background and make comparisons as well. And after that, you will find different method to update the background that different traffic conditions ask for. The approach chosen to refresh the background decides the accuracy of monitoring. For improper methods will cost unnecessary time and memory. And we can construct a counter affected by subtraction between frames of pictures to decide the best way to refresh the background.

Keywords: component, background dynamic refreshing, road traffic monitoring, Gaussian model, background subtraction.

1 Introduction

Background subtraction in the Video Surveillance Technology is faced with two problems—one is how to divide the moving object from the picture, and the other is how to extract its background. It's obvious that single-picture background model is not conducive to target monitoring. So there should be a method that can dynamically process the background picture by dynamically refreshing it. Through background subtraction, the moving objects in the foreground can be extracted. In background refreshing, lots of difficulties have to be solved:

1. sunlight changing
2. moving objects in the background like waving branches or water
3. moving objects resulting in changes in the background
4. noise disturbance
5. reflection of the rearview mirrors

* Corresponding author.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 409–414.
springerlink.com

2 Conventional Bakground Updating

2.1 Median Filter

The median filter method can be applied to obtain the median gradation level of each point in the first N frames of historical images. Points of this median gradation level can thus be used to form the initial template. The equation 1 is shown as follow:

$$B_{t+1}(x, y) = median(I_t(x, y), ..., I_{t-L}(x, y)) \tag{1}$$

However, the median filter method, requiring the gradation level of the background to be over 50%, is featured by the either-or choice of its leak hunting rate and algorithm execution time. In other words, with more frames of historical images, the execution time spent would increase; and with fewer frames less time would be taken, but the leak hunting rate would increase or even hole, especially when slowly moving objects are processed.

2.2 Adaptive Background Model

Adaptive background model is very good in that it is the extension of the basic background subtraction method. Its reference background is refreshed dynamically using Infinite Impulse Filter (IIR), as shown in equation 2.

$$| I_t(x, y) - B_t(x, y) | > T \tag{2}$$

T is the threshold value $B_t(x, y)$ estimated by $I_t(x, y)$ ……, and $B_0(x, y) = I_0(x, y)$.

And its changing trend of background can be learned by a value called Background learning rate shown in equation (3):

$$B_{t+1}(x, y) = \begin{cases} \alpha B_t(x, y) + (1-\alpha) I_t(x, y), & (x, y) \in BG \\ \beta B_t(x, y) + (1-\beta) I_t(x, y), & (x, y) \in FG \end{cases} \tag{3}$$

In equation (3).Bt is the background picture and Lt the current picture. Firstly we should judge whether the point is in the foreground or background and then deal with it according to its category. In the chart above, α and β are the background learning rate, and their specific value should be set depending on the actual situation. There is another method of refreshing background pictures similar to the adaptive background model.

$$B_{t+1}(x, y) = \begin{cases} B_t(x, y) & f(x, y) \in IB_1 \\ I_t(x, y) & if (x, y) \in IB_2 \\ \alpha B_t(x, y) + (1-\alpha) I_t(x, y) & otherwise \end{cases} \tag{4}$$

In equation(4), the idea of this method is to turn the points that remain unchanged or cease moving into part of the background pictures; as with points that have changed,

they can become background points through formulas. In the formulas above,α (0≤α≤1) is called background update rate. As the value of α increases, the background updating would be slowed down; when α=1 the background model would cease updating; and when α=0, it means the object that was marked to be moving has suddenly come to a stop. At this time, corresponding areas of the background model should be replaced by points of current frame of the picture to obtain the complete background image.

Simple as it is, these methods of refreshing background image have shortcomings. For example median filter method has to store historical images to obtain the median gradation. And Adaptive background model provides poor detection accuracy and cannot construct accurate background model. Compared to the above methods, some complex methods show unmatched accuracy.

2.3 Gaussian Model

There is another method to refresh the background—we build a Gaussian model for each pixel to form an accurate background model consisting of parameters of every pixel. It's more accurate than the background model made of median gradation from historical images in the median filter method. The equations 5 are shown as follows:

$$\begin{aligned} B &= [\mu(x,y), \sigma(x,y)] \\ \mu(x,y) &= \frac{1}{n}\sum_{i=1}^{n} f_1(x,y) \\ \sigma^2(x,y) &= \frac{1}{n}\sum_{i=1}^{n} (f_i(x,y) - \mu_i(x,y))^2 \end{aligned} \tag{5}$$

In refreshing the background, in order to prevent negative effect caused by pixels of moving or radically changing objects, we judge from the variance whether the pixel should be updated. Therefore, as the probability of the new pixel belonging to the background increases, its percentage in the refreshing will also become greater; and vice versa. Equations (6) for the refreshing are shown as follows:

$$\begin{aligned} \mu_t &= (1-\alpha)\times\mu_{t-1} + \alpha\times f_t \\ \sigma_t^2 &= (1-\alpha)\times\sigma_{t-1}^2 + \alpha\times(f_t - \mu_t) \\ \alpha &= \begin{cases} 0 \\ \dfrac{K}{\sqrt{2\pi}\sigma_{t-1}} \exp\left(-\dfrac{(f_t - \mu_{t-1})^2}{2}\right) \end{cases} \\ K &= 256^{\frac{|f_t - \mu_{t-1}|}{3\sigma_{t-1}}} \in [0,1] \end{aligned}$$

$$|f_t - \mu_{t-1}| \notin [-3\alpha_{t-1}, 3\alpha_{t-1}]$$
$$|f_t - \mu_{t-1}| \in [-3\alpha_{t-1}, 3\alpha_{t-1}] \quad (6)$$

In equation (6).α is the update factor, which means the probability of the point belonging to the background. Bt-1 is the current background model. $f_t(x, y)$ is the current frame of picture. B_t is refreshed background image. K means certain-time multiplier, which can further strengthen the positive effect of high-probability pixels in the refreshing and reduce the negative effect caused by intensive noise and low-probability pixels.

3 Design of A Dynamic Refreshing Aproach

To sum up, there are two categories of methods of refreshing background. One is to base on historical values and generate a reference image by way of value distribution, namely an approximate background image similar to the current static scene except the moving area. This method works when objects move continuously and the background picture is visible most of the time. However, it is not reliable when there are too many moving objects, especially when they are moving at quite a slow speed. In this case, "ghost" phenomenon will appear, which means ghost still exist even when objects have already left. The other kind of method is to build background models like the Gaussian model according to the distribution of pixels. And these methods are of great adaptive ability and accuracy. But they need large amount of calculation and run slowly in PC, which will result in delay to some extent and is not conducive to real-time monitoring. As with real-time monitoring, higher requirements of monitor equipment would lead to high cost of monitoring. So we can choose proper method for refreshing background accordingly—for fewer moving targets and faster moving scenes, the first kind of method is recommended; and for slower moving scenes with more disturbance and more targets, the latter is proper.

We can get a value computed by the subtraction between current frame of picture and background picture in certain time. If the subtraction is larger than a threshold value, a fixed value should be added to a counter which decreases in certain time in fixed value. Then the counter becomes the judger which can decide the method we use to refresh background. The bigger counter is, the faster scene changes. And Gaussian model can be used when the counter is larger than a fixed value. But if the counter is smaller than the fixed value, median filter method or adaptive background model are proper for updating background. The brief process is shown as Figure1:

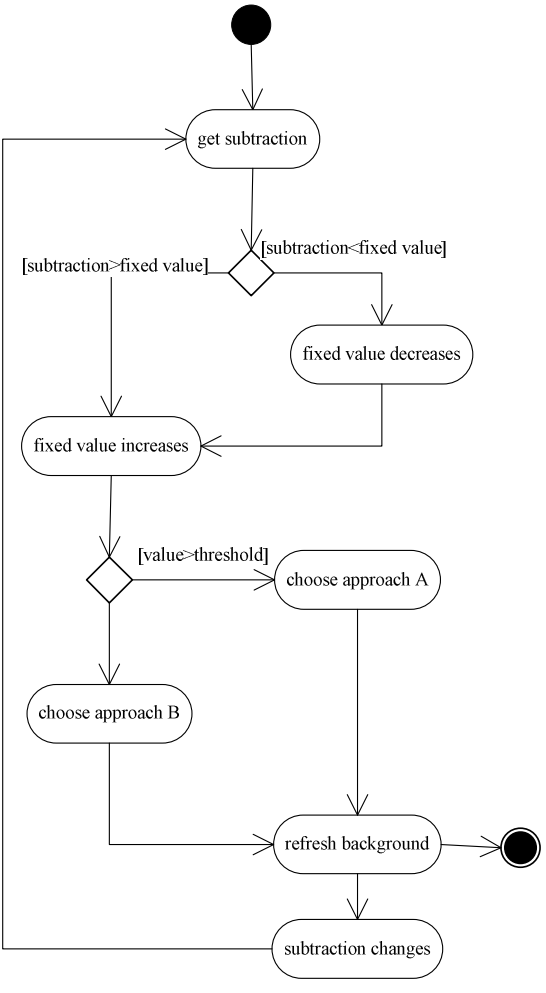


Fig. 1.

And this way of choosing approach to update background will not affected by sudden change in traffic. For example, a few cars passing in a short time will never cause changing in refreshing. So it only changes for steady condition of traffic.

4 Conclusion

Nowadays, traffic directly affects the development in cities. So road monitoring plays an important role in the future. Background subtraction method is often used in road monitoring. Updating background method in background subtraction makes it more accurate in monitoring. But now the way of updating background is always wrongly chosen. As different traffic conditions need different approaches to update background.

We come up with a method to choose the proper way to update the background based on subtraction between frames. This method saves much time and memory. And monitoring will be more accurate.

Acknowledgment. This work was supported by School of Computer Science and Technology, Soochow University, and Soochow University NO. KY2010292B.

References

1. Hou, W.-X.: Object tracking with a background subtraction and inter-frame subtraction, pp. 1008–0570 (2009), 02-3-0242-02
2. Gan, X.-S., Zhao, S.-B.: Comparison on Background Subtraction Algorithms for Moving Target Detection, 1673–3819 (2008), 03-0045-06
3. Wang, X.-H., Wang, B.: An efficient method for refreshing background, pp. 1001-9146 (2009), 03-0056-04
4. Zhang, H.-B., Huang, S.: Comprehensive dynamic background updating method for real-time traffic visual surveillance, TP391.41
5. She, H.: An Improved Method of Updating the Background Dynamically Used in ITS, pp. 1000-3428 (2004), 14 -0152-03
6. Ai, H., Le, X.: Dynamic Background Updating Method for Real-Time Tracking in Visiual Surveillance, pp. 1002–8331 (2001), 19-0104-03
7. Xu, S.: Dynamic Background Modeling for Foreground Segmentation, 978-0-7695-3641-5
8. Ying, Z., Comaniciu, D., Pellkofer, M., Koehler, T.: Passing vehicle detection from dynamic background using robust information fusion, 10.1109/ITSC.2004.1398962
9. Gallego, N., Mocholi, A., Menéndez, M., Barrales, R.: Traffic Monitoring:Improving Road Safety Using a Laser Scanner Sensor, 978-0-7695-3799-3
10. Liu, J., Wang, M.: An Approach for Lane Segmentation in Traffic Monitoring Systems, 978-0-7695-3843-3
11. Chen, J., Kim, M., Wang, Y., Ji, Q.: Switching Gaussian Process Dynamic Models for simultaneous composite motion tracking and recognition, 978-1-4244-3992-8
12. Wang, C.-X., Li, Z.-Y.: Face Detection Based on Skin Gussian Model and KL Transform, 10.1109/SNPD.2008.145

The Accessibility Assessment of National Transportation Network and Analysis of Spatial Pattern

Xiao Fan[1], Shufang Tian[1], Tiyan Shen[2], and Jinjie Zhang[2]

[1] School of Earth Sciences and Resources
China University of Geosciences
Beijing, China
[2] School of Government
Peking University
Beijing, China
{fxdy39,zjjslr}@163.com

Abstract. The accessibility assessment is an effective way to qualify the performance of transportation network. This paper assesses China's transportation system with many impact factors, which are calculated using gravity model respectively, and then calculate the indicator of transport convenience with the method of weighted superposition. For better understanding of the spatial distribution pattern of transport accessibility and its relationship with socioeconomic status, statistical analysis and spatial autocorrelation analysis are made. The results show that accessibility of the national transportation presents an "east high, west low" spatial distribution pattern and has a positive correlation to regional economic development and population density. Besides, the spatial pattern of transport accessibility has a significant spatial aggregation, especially evident in some eastern cities, while the western part of the China doesn't have an apparent spatial aggregation, which also shows the considerable differences between the east and the west, revealing economic development and transport accessibility drive and promote each other synchronously. So accelerating the development of regional transport infrastructure construction and boosting local economies are good ways to reduce regional differences.

Keywords: transport accessibility, factor evaluation, gravity model, statistics analysis, spatial distribution pattern.

1 Introduction

Accessibility generally denotes the ease which any land-use activity can be reached from a location, using a particular transport system [1]. It has been widely used in many areas and the definition of this concept varies with the understandings of people in different fields. Lynch took accessibility as an ability of residents traveling, carrying out activities and accessing to resources, services, and information [2]. It reflects the ease of communication overcoming obstacles between spatial entities, so

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 415–423.
springerlink.com

it has a close relation to a set of spatial concepts, including spatial location, interaction among entities (physical size, attractiveness and so on) and the scale of studying areas. Accessibility now has been used in broader fields, such as land use, transportation system, economic, financial and individual psychology. It's an important reference indicator in urban planning, land use and transport planning. Scholars in different fields, focusing on different subjects, analyze accessibility at different scales through different spatial framework. Accessibility can be divided into destination-based accessibility and personal-based accessibility. The former is based on the destination, defined as the opportunity from any other locations to the destination. In other words, it's the potential of a region to be got close to. The latter refers to the opportunity of individuals to reach other regions, which reflects the quality of life or available benefits [3,4]. Both the destination-based and personal-based accessibility reflect many-to-one and one-to-many relationship between starting point and destination.

Accessibility used in regional transportation network is hot subject of research. Accessibility now is widely used in evaluating the efficiency of urban traffic system, which reflects the relationship among urban structure, land use, urban planning and infrastructure. Space-distance-accessibility has become a precondition for the linkage and development of urban regional integration [5], and accessibility of transportation network directly affects urban development. Zhang Zhilin and Jiang Hairong carried on comparison among the six provincial capitals from points of transportation distance, accessibility coefficient of the city, the position potential and the convenience of railroad and aviation [6]. Guo Lijuan and Wang Ruyuan calculated the railway and highway comparative accessibility of 18 cities in Sichuan Basin Urban Agglomeration by the Shortest Time Distance Model (STDM) and established Integrated Accessibility Model to measure integration accessibility of cities. They also established the Shortest Time Distance Matrix to analyze the spatial contact intensity among the 18 cities [7]. Xu Xu analyzed the current accessibility in the landway transport network in Guangzhou-Hong Kong Urban Corridor, using the shortest time distance index and the weighted average travel time index [8]. Cao Xiaoshu and Yan Xiaopei analyzed the impact of road and railway network on the spatial structure of accessibility in the past 20 years using Dongguan city in Guangdong province as an example [9]. Liu Chengliang taking Wuhan as an example, constructed shortest-distance matrix, introduced temporal and spatial accessibility models, built the evaluation index system of the traffic network with higher hierarchy development, and quantitatively investigated the changes and spatial structure of the urban accessibility [10]. Feng Zhiming established the transportation ability index model and evaluated the transportataion ability of China in 2005 [11].

2 Methodology

In this paper, we choose some indicators that can reflect the transport efficiency, including indicators measuring the distribution of traffic lines, convenience of transport link among different regions or transport capacity etc, to establish an

indicator system. As the data acquisition limitations, we choose distance from the central cities, national roads, provincial roads, county and township roads, highway toll stations, airports, harbours and railway stations as evaluation indicators. Then take the gravity model proposed by Hansen as the calculation model [12], which is written as

$$C_i^H = \sum W_j d_{ij}^{-\beta} \quad (1)$$

H represents the calculated indicators, j is the supplied points, which represents the number of central cities or road lines etc. that influence region i. C_i^H is H's accessibility in region i ,W_j is the weight of point j in the region i, d_{ij} is the distance between i and j, and βis a resistance coefficient. Considering the calculated amount, we set the value of n as 1. In other words, the convenience of each region is only defined by the nearest city, road node and road line. βis also set to 1, meaning distance is a linear attenuation.

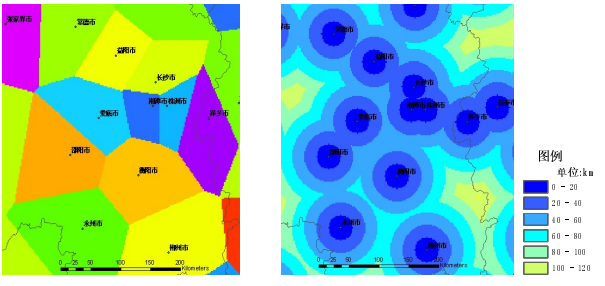


Fig. 1. Influence range and influence distance of central cities(a part of Hunan province)

Based on how indicators impact the transportation system, we select the appropriate value or other relative indicators as weights. Road indicators (national roads, provincial roads, county and township roads) take different levels of roads' annual average day and night traffic volume as weight, railway stations indicator takes the level of station as weight, highway toll stations indicator has a certain value as weight due to lack of related data, airport indicator takes the city's civil aviation passenger capacity as weight, ports indicator takes the cargo handling capacity of the ports as weight, Central cities indicator takes the population of the city as weight. Because the weight of different indicators varies greatly, logarithmic transformation is needed to reduce the differences.

$$B_i = \ln(w_j) — \beta \ln(d_{ij}) \quad (2)$$

Then normalized each indicator

$$A_i = (B_i — \min(B)) / (\max(B) — \min(B)) \quad (3)$$

The transport accessibility is calculated by overlaying all the indicators A^i. It is worth mentioning that we take grid sampling before evaluation to improve accuracy. Conduct a nationwide sampling and take every cell as the basic unit for calculation. And then use the bottom-up method to summarize all the cells in every administrative division and get the transport accessibility in different administrative division level. In this paper, we take two kilometers as a sampling unit to make sure the smallest county can be explained by at least one unit.

3 Evaluation of the National Transport Accessibility

Traffic data used in this article is from national vector database of 2007 and the socio-economic data come from the national statistical yearbook of 2008. Grid sampling and evaluating indicators are all completed using ARCGIS software. Sampling is executed using the raster-vector transfer function and gravity model using the spatial analysis module.

Firstly, we set the influence range of every indicator and calculate the distance from the cells contained to the indicator. Taking central city indicator as an example, calculate the distance from each cell to the nearest central city and assigned the population value of the nearest central city to the cell, then the central city indicator of the cell is calculated using the gravity model(Fig. 1). Likewise, the same method is used to calculate each cell's road indicator, highway toll stations indicator, airports indicator, ports indicator, and railway stations indicator.

Then calculate the summation of the eight indicators using raster calculation function of ARCGIS software. After that, summing up all the cells in every county and province and take the average of all the cell values in each administrative division as its transport accessibility, so we get the national transport accessibility in provincial and county level. For the convenience of comparison, normalize accessibility ranges from 0 to 1.

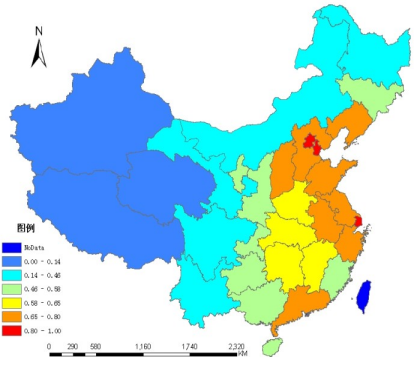


Fig. 2. China transport accessibility in provincial level

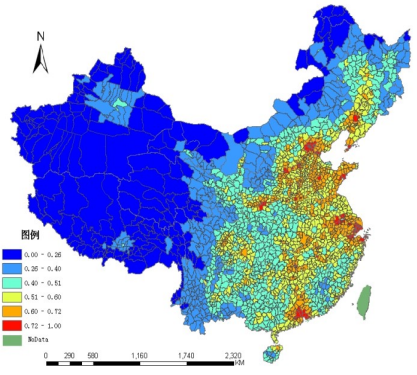


Fig. 3. China transport accessibility in country level

4 Regional Analysis of Transport Accessibility

The transport accessibility has significant differences across the county, it's much higher in the east is than the west as a whole. Regions having higher transport accessibility mostly distribute in north, east and southeast of China (Fig. 2), among that Beijing, Tianjin, Shanghai, Hong Kong and Macau are highest. Other high accessibility regions include coastal cities and cities role as transportation hubs, such as Xuzhou and Baoding (Fig. 3). Xinjiang, Tibet and Qinghai have the lowest transport accessibility (Fig. 2). Some regions in Western China have relatively high accessibility and reflect the accumulative effect, such as Lanzhou-Yuzhong region in Gansu province, Chengdu-Shuanliu region in Sichuan province (Fig. 3).

Table 1 shows the regional differences of transport accessibility in China. There is a great accessibility gradient exists and each partition of accessibility distributes unevenly. Most provinces in East China have convenient transportation, transport accessibility mostly above 0.66 except Guangxi, Hainan and Fujian. Provinces in Middle China have a wide range of accessibility from 0.14 to 0.80. All the Provinces in West China have the transport accessibility lower than 0.57.

Table 1. Regional differences of transport accessibility

Access-ibility	**Regions of China**		
	East China	***Middle China***	***West China***
0.00-0.14	--	--	Qinghai, Xinjiang, Tibet
0.14-0.46	--	Heilongjian, Inner Mongolia	Gansu, Sichuan, Yunnan
0.46-0.57	Guangxi, Hainan, Fujian	Jilin	Ningxia, Guizhou, Chongqing

Table 1. (*continued*)

0.57-0.66	--	Henan, Hunan, Hubei, Jiangxi, Shanxi, Shanxi	--
0.66-0.80	Liaoning, Hebei, Jiangsu Shandong, Zhejiang	Anhui, Guangdong	--
0.80-1.00	Beijing, Tianjin, Shanghai	--	--

4.1 Correlation Analysis of Transport Accessibility

To find the relationship between transport accessibility and socioeconomic status, we choose two indicators representing regional social and economic condition separately to make empirical analysis. The two indicators are population density and GDP of every province in 2007 (except Hong Kong, Macao and Taiwan). Spearman correlation coefficient is used to measure the relation between transport accessibility and the two indicators respectively. Spearman correlation coefficient is a non-parametric measure of statistical dependence between two variables. It assesses how well the relationship between two variables can be described using a monotonic function. The raw scores are converted to ranks and the differences between the ranks of each observation on the two variables are calculated. The Spearman coefficient is denoted with the Greek letter rho (ρ).

The Spearman correlation coefficient of transport accessibility and population density is 0.828 (sig<0.001), representing a significant positive correlation (Table 2). The Spearman correlation coefficient of transport accessibility and GDP is 0.706 (sig<0.001), also representing a significant positive correlation (Table 2). The above analysis shows that transport accessibility has a close connection with regional economic development and population distribution. Areas having convenient transportation are often densely populated and economically developed. Regions having lower accessibility are mostly undeveloped and sparsely population distributed.

Table 2. Spearman correlation coefficient

	Spearman correlation coefficient	**Sig. (2-tailed)**	**N**
population density and accessibility	0.828	0.000	31
GDP and accessibility	0.706	0.000	31

4.2 Spatial Distribution Pattern of Transport Accessibility

In order to analyze the spatial distribution pattern of national transport accessibility, the spatial autocorrelation statistic Moran's I is used to measure the spatial autocorrelation of transport accessibility.

In this paper, we calculate the global Moran's I in provincial and county level respectively based on first-order neighbour-joining standard using Geoda software. The results are 0.5206 and 0.8732 respectively, which show that spatial distribution pattern of national transport accessibility has a significant autocorrelation and spatial aggregation both in provincial and county level. We also find that Moran's I is lower at the larger scale. This is because the filtering features when averaging the data at the thicker scale and Moran's I's nonlinear characteristics for distance.

As it is difficult to find local spatial relevance pattern by using global Moran's I, so we use Moran's LISA to estimate the local spatial autocorrelation in county level (Fig. 4). Red regions in Fig.4 are high accessibility agglomeration areas and blue regions are low accessibility agglomeration areas($p<0.05$). The white regions have no significant regional agglomeration effects. Aggregation trend of accessibility is clear, and regional disparity is great. High accessibility agglomeration areas mainly concentrated in Bo Hai coastal region, including south part of Liaoning, Beijing, Tianjin, northwest of Shandong and Shandong peninsula, yangtze river delta, pearl river delta and capital cities in the east. Besides, the urban agglomerations centered at the transportation hubs also have high autocorrelation. Low accessibility agglomeration areas are mainly distributed in West China, northwest of Northeast China and west of North China. The results in this section are consistent with the previous ones.

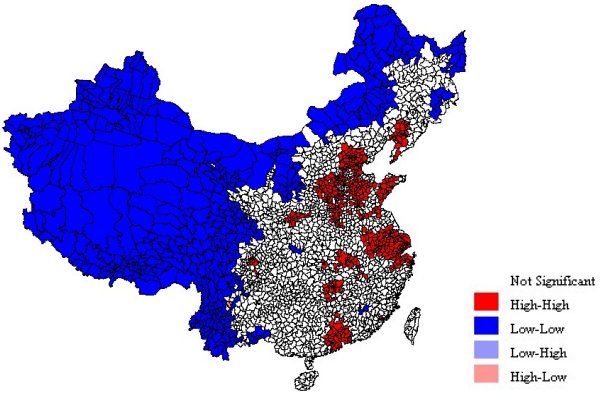


Fig. 4. Moran's I LISA in county level

5 Conclusions

Geographical information system (GIS) is an information system that integrates, stores, edits, analyzes, shares, and displays geographic information. Recent advances

in GIS make it possible to enhance our understanding of spatial relationships. Exploratory spatial data analysis (ESDA) is used to explain the phenomena of spatial dependence, spatial correlation and spatial autocorrelation which are relevant with spatial locations. As an important field in spatial econometrics, its application is becoming more and more widespread internationally. This paper chooses eight impact indicators and evaluates the transport accessibility using gravity model. To enhance the accuracy, we sample the whole nation at two-kilometer scale, and take the bottom-up method to calculate the transport accessibility in province and county level. Then spearman correlation coefficient is calculated to get the relation between accessibility and local socioeconomic situation. To gain a better understanding of spatial distribution pattern of the whole nation's transport accessibility, global Moran's I and Moran's LISA is calculated using Geoda software. Conclusions are obtained as follows:

1 In summary, transport accessibility in East China is higher than the west. Big cities and transportation hubs in the east are highest.

2 The transport accessibility has a significant positive correlation with regional population density and GDP in provincial level.

3 Spatial distribution pattern of transport accessibility in the east represents very strong spatial cluster characteristics, especially the three biggest economic circles.

4 Areas which are not well economically developed also have low transport accessibility, revealing that transportation, economy and population promote mutually and develop synchronously.

5 Based on the above conclusions, we propose speeding up the construction of transportation infrastructure in less developed areas and boost local economies to reduce regional differences.

There are still many problems that require further research, including:

1 The gravity model used in this paper is simplified to a linear attenuation model, but the model is more complex actually. The evaluation model which is more consistent with the facts needs future research.

2 Topography is also an important impact indicator to transportation, especially in mountainous areas, but it is not taken into consideration in this paper due to the limitation of data. So the selection of indicators needs to be further improved.

References

1. Dalvi, M.Q.: Behavioural modelling accessibility, mobility and need: concepts and measurement. In: Hensher, D.A., Stopher, P.R. (eds.) Behavioural Travel Modelling. Croom Helm, London (1978)
2. Lynch, K.: Good city form. MIT Press, Cambridge (1981)
3. Yang, T., Guo, X.: New concept of urban travel accessibility and its application. China Journal of Highway and Transport 8, 25–30 (1995) (in Chinese)
4. Li, Y., Qin, Y., Pan, S.: Study on the transportaion network accessibility measures based on GIS. Yunnan Geographic Environment Research 19, 99–104 (2007) (in Chinese)

5. Liu, C., Ding, M., Zhang, Z., Zhang, H.: Accessibility analysis on the spatial linkage of Wuhan metropolitan area. Progress in Geography 26, 96–107 (2007) (in Chinese)
6. Zhang, Z., Jiang, H.: Comparative research on transportation accessibility of six procincial captials in central China. Journal of Shenyang Normal University (Natural Science) 24, 495–498 (2006) (in Chinese)
7. Guo, L., Wang, R.: Studying on accessibility and spatial connecting among the cities in Sichuan basin urban agglomeration. Human Geography 3, 42–48 (2009) (in Chinese)
8. Xu, X., Cao, X., Yan, X.: Potential accessibility and its spatial pattern in Guangzhou-Hong Kong urban corridor. Geographical Research 26, 179–186 (2007) (in Chinese)
9. Cao, X., Yan, X.: The impact of the evolution of land network on spatial structure of accessibility in the developed areas: the case of Dongguan city in Guangdong province. Geographical Research 22, 305–312 (2003) (in Chinese)
10. Liu, C., Ding, M., Zhang, Z., Zhang, H.: Accessibility analysis on the spatial linkage of Wuhan metropolitan area. Progress in Geography 26, 96–107 (2007) (in Chinese)
11. Feng, Z., Liu, D., Yang, Y.: Evaluation of transportation ability of China: From county to province level. Geographical Research 28, 419–429 (2009) (in Chinese)
12. Hansen, W.G.: How accessibility shapes land use. Journal of the American Institute of Planners 25, 73–76 (1959)

Public Traffic Intelligent Dispatch Algorithms Based on Immune Genetic Algorithm

Xu Hai Yan

LinYi University, LinYi, China
xuhaiyan1997@163.com

Abstract. Taking into account the current transit vehicle operation scheduling problems using genetic algorithms and hybrid genetic algorithm for static scheduling of public transport vehicles were studied. Intelligent features of the genetic algorithm, effectively improved the public transport vehicles, static scheduling. On the basis of integrated biological immune mechanism interdisciplinary theory, this paper into the genetic algorithm immune algorithm which constitutes an improved genetic algorithm, and applied to solve the bus scheduling management optimization, and use of simple genetic algorithm The calculation process is simulated and compared. Application results show that the algorithm is simple, efficient, stable, traditional methods and can better overcome the shortcomings of existing genetic algorithms, performance has been significantly improved, with satisfactory results, improving the operational efficiency of public transport vehicles.

Keywords: APTS, PTIDS, Genetic Algorithms(GA), Immune Genetic Algorithm (IGA), static dispatch, dynamic dispatch.

1 Introduction

There is no doubt that public transport is indispensable to daily life an important part, especially in the populous cities, public transport vehicles are the preferred way people travel. Public transport operators dispatch management is usually divided into three stages: planning, scheduling and control. Scheduling is a key intermediate links, vehicle operation scheduling problem is a biodegradable transportation tasks for a certain time constraints and the sequence of how the arrangement worked to get the transportation costs or time optimization. Do a good job of scheduling bus transportation for the perfect urban environment, improve people's transportation situation and improve the bus company's economic and social benefits, are of great significance. Bus Dispatching, including static scheduling and dynamic scheduling, this paper focuses on the static scheduling, namely: operating vehicle departure timetable scheduling process, which is the basis for dynamic scheduling. At present the team's scheduling of the bus company mainly depends on the experience of staff manually, although it has some relevance, but it is obvious lack of scheduling with a lot of blindness, it is difficult to ensure that the results of scheduling In terms of operational efficiency is optimal or nearly optimal. To meet the needs of practical applications, the use of intelligent algorithms to solve vehicle

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 425–430.
springerlink.com

scheduling problems, within the limited steps of the algorithm to find all the programs to meet the scheduling constraints of the optimal solution or near optimal solution. Bus companies can improve the management and service levels.

2 Genetic Algorithms

A. The synopsis of Genetic Algorithms

Genetic Algorithms (Genetic Algorithms, GA) is an adaptive global optimization probability search method based on the principle of natural selection and genetics. There are two objectives of founding Genetic Algorithms, the one is to solve natural adaptation processes abstractly and critically , the other is to put the important mechanistic in natural biology system to the project system. Genetic Algorithms begins to operate from many way , and then heuristics search efficiently in the solution space, so this algorithms can avoid search process effectively to convergence to local best. Genetic Algorithms simulates biological evolutionary process and genetic manipulation in computer, and calculate the goodness of fit statistics with object function, it doesn't need specific knowledge of objects ,so it has optimizing ability of overall situation and can solve hard and complex problem, it is used widely to control automatically ,graph manipulation and power switching and so on.

B. The process of basic operation

The steps of Basic Genetic Algorithms are:

1) Generate initialize population randomly;

2) Whether or not it demand stop condition? If it demands , and then go to the eighth stage.

3) Or else calculate each current individual fitness function

4) Go to choose and generate intermediate group with each current individual fitness function.

5) Go to choose two individual with P_c *probability then exchange their chromosomes, generate new individual to replace the old one; then insert the new one into the group.*

6) Choose one chromosome with P_m *probability to change, then generate new individual to replace the old one;*

7) Go to step 2;

C. Build mathematical model of dispatch

The mathematical model that be used in the dispatch system simplify the run environment, such as: constant vehicle speed, keep an even speed, no specific events happening; the minimal time unit is minute, these are reasonable to arrange timetable;

Suppose that the model of passenger movement can reflect daily passenger flow volume of this bus line (we suppose that the arrival passengers obey uniform distribution, and have different distribution density in different time). first train departs

at six o'clock in the morning, final train departs at 22 o'clock in the evening, all buses depart at minute of integer, total train frequency one day are m, all runtime are 16 hours, namely 960 minutes. Use x_m to denote the m bus running time between departing and depart firstly, the unit is minute. Decision variable are $X = [x_1, x_2, \cdots, x_n]^T$; chromosome X is a complete departure time-table, each gene is departure time of one bus. The constraint conditions are as following:

$$\begin{cases} x_i \in Z and x_i \geq 0, and i = 1,2,\cdots,m; \\ x_1 = 0, x_m = 960; \\ x_1 < x_2 < \cdots < x_n < x_{n+1} < \cdots < x_m \end{cases}$$

$f(X) = \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}(T_{ij} - t_{ijk})$ denotes objective function, the number n denotes total floor of buses lines, total train frequency one day are m, T_{ij} denotes the runtime of number j bus arrives the i stop in the bus line, t_{ijk} denote the time of the k passenger in the j-1 bus arrives to the I stop, f(X) denotes total passengers waiting time in one day in one train line. It will be seen that least the objective function value is, shorter the passenger waiting time is, better the service level is. Because the variation range of gene in algorithms and the searching space is larger, and encode with true value, so that it simplifies the process of encoding and improves the computational efficiency. The following will introduce simply genetic manipulation:

1) Choosing operator: use proportional selection method, namely roulette selection. The probability of each individual be chosen is proportional to sufficiency.

2) Crossover operator: at first, put m individuals make up M/2 pairs of pairing individual groups, and then put each individual groups to go to crossover operator with P probability. Set one crossing point randomly in the individual encoded string firstly, then exchange two pairing individual chromosome at this crossing point. The crossing point choosing should be ensured new individual demand constraint conditions.

3) *Mutation operator: Use uniform variant operation. Then designate in turn individual gene in the individual encoded string into change point, take one uniformly distributed random numbers with little change probability p from value range of corresponding gene to replace the old gene for each change point.*

3 Lead in GA to Immune Mechanism

GA is used extensively to many fields owning to operation simply and effective capacity of problem solving. GA can find out randomly optimal solution of problem. Using GA will appear some problems, for example, it will happen premature convergence easily and weak local optimizing ability and so on. Generally speaking, solving ability of basic GA isn't always the best; and GA also can't avoid the same feasible solution with many searching, which also impact operational efficiency of GA. But some optimizing algorithms have strong local searching ability, so we can estimate to mix these

optimizing algorithms into GA, then form a Hybrid Genetic Algorithm to improve operational efficiency and the solving quality of GA.

Immune Algorithm emulates Biological Immune System, it puts foreign antigens and immune antibody correspondent partly objective function of problem solving actually and problem solution, if antibody can exclude antigens effectively, that means find out optimizing solution; antibody which has good affinity to antigens is gone to recall can accelerate solving. It has some features as following:

D. Diversity of antibody: Immune System can generate antibody which has diversity to resist various antigens.

E. Self Regulation. Immune System has keeping immunologic balance and can self regulation generate right quantity and necessary individual. If find out optimizing individual once among the evolutionary process of Immune System, the individualization will also breed in abundance in condition that it gives consideration to diversity of group.

Immune System Algorithm develops on the basis of individual; but species macroevolution has important impact on individual Immune System macroevolution. Immune System evolutes slowly along species evolution, on the other hand, it evolutes fast to adapt pathogens environment. It overcomes the defect of out of control convergence direction with GA, convergence direction can be controlled well. IGA can be seen to an improved gene algorithm, a new intelligent computation algorithms and a gene algorithm having immunologic function gene algorithm. It has GA the feature of global searching, overcomes the defect of local search inefficiently in the solution spaces because to Cross searching with GA, and can avoid largely immaturity convergence, it shows to surpass GA and Immune Algorithm from many sides.

4 IGA Apply to Dynamic Dispatch of Public Train

F. Design philosophy of IGA

Optimal individual of IGA is the feasible solution of the highest sufficiency each generation. In the probability speaking, the distance between optimal individual and globally optimal solution is less than the distance between optimal individual and other individual, therefore, optimal individual reflects feature information of problem solving directly. In terms of Immunology, when antigen invade, antibody going with it is excited to make useful antibody save. It can be seen from those studying, the key that whether new algorithm is successful is whether it is carried saving essence strategy and make the best of optimal individual information each generation. Use for reference of biological immune mechanism, the reproduction of child individual with GIEA is as following:

$$\begin{cases} x_i^{t+1} = x_{i,best}^t + \sigma_i^t \times N \\ \sigma_i^{t+1} = \sigma_\xi + \sigma_i^0 \times e^{-\frac{A \times t}{T}} \end{cases} \quad i = 1,2,\cdots,n$$

x_i^{t+1} denotes the I component of child individual; $x_{i,best}^t$ denotes the I component of parent optimal individual; σ_i^{t+1} denotes standard deviation of the I component of child individual; σ_i^t denotes standard deviation of the I component of parent individual; A denotes standard deviation of dynamic coefficient adjusting, usually $A \in [1,10]$; T denotes total evolving algebra , t denotes evolving algebra; σ_ξ denotes cardinality of standard deviation, usually is zero in using; n denotes the component numbers of x; N denotes random number obeying uniform distribution generated among -1 and 1; σ_i^0 denotes standard deviation of the I component of initialize population, usually $\sigma_i^0 \in [1,3]$, A and σ_i^0 dereference according to research question.

In summary, the key of IGA is that making the best of optimal individual information, make optimal individual evolution to replace group evolution, and integrate local search and global search dynamic with adjustment of standard deviation. The measure of realizing algorithm is that put emphasis on embodying optimal individual reservation, reproduction and the static adjustment of standard deviation, compared with current other evolutionary algorithm, it has excellences of high search efficiency and local optimal solution uneasily.

C. Operating steps of IGA

We still use above mathematical model, initial scale N is still 200 and doesn't change with evolving algebra, so that it will be convenient to compare IGA with GA. The steps of using IGA to optimize are as following:

1) Generate initial group randomly in the solution spaces, and calculate its sufficiency, confirm optimal individual x_{best}^0, and give the value of σ^0, A is chosen to 1, σ_i^0 is 3 and σ_ξ is 0.

2) According to the 4.7 formula, generate child group in the solution spaces, the scale is N.

3) Calculate sufficiency of child group, confirm optimal individual x_{best}^{t+1}, if it demand the condition of $f\left(x_{best}^{t+1}\right) > f\left(x_{best}^t\right)$, we choose x_{best}^{t+1} to be optimal individual, or else choose x_{best}^t.

4) Carry out the second and the third steps repeatedly until achieve the conditions of ending, T is chosen to be 100, we choose the last generation to be optimizing result.

5 Worked Example and Results Analysis

We do lots of experiments to confirm crossing probability, probability of mutation, evolving algebra and so on, we confirm finally effective parameters are as following: the

crossing probability is 0.0005, probability of mutation is 0.6. in addition, size of the group is 200, genetic algebra is 600, so that we can give attention to operand and render.

Relationship with evolving algebra of each generation individual objective average and optimal individual objective average in GA and IGA, it can be seen obviously. Convergence speed of IGA is better and faster than GA. Use IGA to management system of public traffic intelligent dispatch will get better result, and can get optimal solution rapidly. IGA is an improvement to GA, and has large advancement in capability.

6 Conclusion

In this article, we use a new hybrid genetic algorithm namely Immune Genetic Algorithm to study static dispatch of public trains on the basis of Genetic Algorithm, and compare with application outcomes of both, the results show that using IGA can overcome the deficiency of GA, and improve the service level of public transport enterprises, have practical significance on improving the city traffic problem and saving citizen travel time. Because the train is influenced by lots of objective factors in operational process, the train may not arrive the intended station as planned, we should do dispatch to the train operation, so the dynamic dispatch needs further researches.

References

[1] Ni, C.: The research of GIEA and application in water problems, the doctoral dissertation of sichun university, pp. 50–51 (2003)
[2] Zhou, M., Sun, S.: The principle and application of GA. National Defence Industrial press, Beijing (1999)
[3] Li, Y., An, T.: The study of public train computerized Dispatch based on GA. Journal of Transportation Systems Engineering and Information Technology 3(1), 41–45 (2003)

Analysis of the Over-Current Phenomena While the EMU Passing Neutral Section

Xin Li, Qun-zhan Li, Fu-lin Zhou, and Yan-kun Xia

School of Electrical Eng., Southwest Jiaotong University, Sichuan Chengdu 610031

Abstract. This paper analyzes the theory and characteristic of the over-current phenomena while the EMU passing neutral section and builds the mathematical expression of the flux linkage when several parallel transformers closing at the same time. It also builds the simulation model based on matlab/simulink and investigates the relationship between the number of transformers closed while the EMU passing neutral section and the value of the inrush current.

Keywords: Passing neutral section, over-current, inrush current, closing in parallel, EMU.

1 Introduction

Electrified railway power supply system take the way of sectional power supply, the phase of adjacent section is different and isolated by a non-power neutral section. Locomotives experience a process from power to non-power, and then from non-power to power while passing neutral section. The world main technology of automatic passing neutral section include ground-switching automatic mode, pillar-switching automatic mode and on-board automatic mode. Because of the less investment, better performance and reliable, The on-board automatic mode is wildly used in China. Using on-board automatic mode, locomotives firstly cut off the main circuit breaker to enter the neutral section, after passing neutral section the locomotive close the breaker in order to restore power supply into next supply arm.

With the vigorous development of passenger dedicated line, the train runs faster and the distance becomes longer. The EMU need to operate the traction transformers on-board frequently. The reclosing inrush current while passing neutral section can caused electric shock, thermal effects and mechanical shock to influence the traction transformer, reduced its service life. In recent years, the failure of closing main circuit breaker have also occurred due to the reclosing inrush current.

Currently the research of reclosing over-current is focused on the process of ordinary locomotive with single transformer, such as literature[1] pointed that the inrush current doesn't appear if electric locomotive can avoid the impact of the closing angle and transformer remanence. Literature[2] analyzed the impact factor of single traction transformers over-current and its influence to public grid. Compared to ordinary locomotive, EMU adopt power distributed structure. It has a larger number of

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 431–438.
springerlink.com

transformers so that the mechanism of multi-transformer reclosing over-current is more complicated. The study on it shows the importance and necessity. This paper analyzes the theory and characteristic of the over-current phenomena while the EMU passing neutral section, builds the simulation model based on matlab/simulink to investigate the analysis result.

2 The Characteristic of Single On-Board Transformer Over-Current

The over-current of single on-board transformers is mainly caused by transformer inrush current. Transformer core magnetization curve is nonlinear. When the transformer internal flux is small, the inrush current is nearly zero in general. But if the transformer core flux reaches a certain level, the inrush current increases rapidly with the increase of the flux in the core. Due to the flux of transformer can't mutation while single on-board transformer unload switching, a non-periodic transient flux is produced in side transformer. The synthetic flux of internal transformer is equivalent to the superposition of the steady-flux and transient-state flux. The existence of transient flux caused the magnetizing slip saturation, resulting in inrush current phenomenon. The theoretical analysis of inrush current phenomenon caused by single on-board transformer unload switching is followed:

Assuming the input voltage is $u=U\sin(\omega t+\alpha)$, according to electromotive force balance principle the Equation 1 is listed as follows:

$$\frac{d\psi}{dt} + iR = Usin(wt + a) \tag{1}$$

Ignoring the voltage drop caused by resistance can simplify from Equation 1 to Equation 2 :

$$\frac{d\psi}{dt} = U\sin(\omega t + \alpha) \tag{2}$$

Assuming that there is no remanence in transformer core, the Equation 3 is as follows:

$$\psi = \psi_m[\cos\alpha - \cos(\omega t + \alpha)] \tag{3}$$

As seen from Equation 3, there is no non-periodic component when α equals 90°, transformer operate to the steady-state directly. When α equals 0°, there are both periodic component and non-periodic component, the peak value synthesized by the two component can be $2\Psi_m$. Transformer core is designed to run near the inflection point, due to transformer saturation and the nonlinear relationship between excitation current and magnetic flux, the excitation current can be very big if the flux reaches $2\Psi_m$, that's the inrush current phenomenon.

Single on-board traction transformer reclosing inrush current is similar to the power transformer inrush current. The inrush current curve is spire slope and its value is big, the maximum value can be 8 to 10 times of the rated value. The current contains a lot of higher harmonics. Because inrush current is the result of transformer core saturation, the inrush current is nearly zero when the transformer does not reach saturation, so the inrush current tend to the timeline side and has dead angle apparently. The attenuation of inrush current is related to the saturation of transformer core, the greater degree of saturation and the larger of the transformer reactance, the quicker of the attenuation. So the attenuation speed of inrush current is gradually slowed down and finally to steady-state operation.

3 The Analysis of Over-Current While EMU Passing Neutral Section

The process of EMU passing neutral section reclosing is equivalent to multi-transformer unload closing in parallel, over-current is mainly the inrush current caused by multi-transformer unload closing in parallel, Therefore the characteristic of over-current while EMU passing neutral section can be got through the analysis of multi-transformer inrush current characteristics.

In order to simplify the study, it is assumed that the EMU is 8 grouping of 4M4T. Its reclosing circuit model is showed in Figure 1. R_s, L_s is traction network equivalent resistance and reactance in Figure 1. R_i, L_i is the resistance and reactance of transformer i.

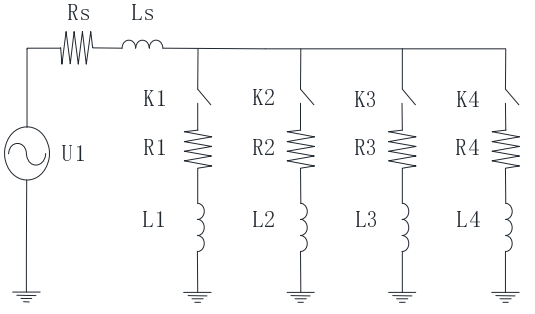


Fig. 1. Reclosing circuit model of passing neutral section

Through the reclosing circuit model in Figure 1, it is easy to get flux formula of any transformer as Equation 4 shows:

$$\frac{d\psi_{pi}}{dt} = u_s - R_s i_s - L_s \frac{di_s}{dt} - R_i i_i \tag{4}$$

Integrated Equation 4, assuming the initial flux $\psi_{pi}(0)$ is zero excluding the impact of transformer remanence. Equation 5 can be obtained considering a period changes of flux:

$$\begin{aligned}&\psi_{pi}(2\pi)-\psi_{pi}(0)\\&=\int_0^{2\pi}u_s(\theta)d\theta-R_s\int_0^{2\pi}i_s(\theta)d\theta-L_s[i_s(2\pi)-i_s(0)]-R_i\int_0^{2\pi}i_i(\theta)d\theta\end{aligned}\tag{5}$$

In Equation 5, i_s is the summation of the multi-transformer inrush current. Due to the dead angle of inrush current so $i_s(0)\approx i_s(2\pi)\approx 0$ [4], the changes of flux can be obtained as equation (6) shows:

$$\Delta\psi_{pi}=-\int_0^{2\pi}(R_sI_{SD}+R_iI_{iD})d\theta=-\int_0^{2\pi}[(R_s+R_i)I_{iD}+R_s\sum_{k=1,k\neq i}^{4}I_k]d\theta\tag{6}$$

In Equation 6, I_{SD}, I_{kD}, I_{iD} is respectively the non-periodic component of current i_s, i_k, i_i. Assuming that the inrush current generated by reclosing is positive direction, it can be obtained $\Delta\psi_{pi}<0$ by Equation 6.Because the main flux of each on-board transformer reclosing at the same time in parallel is $\psi_{pi}>0$, the changes of flux is $\Delta\psi_{pi}<0$, that is the magnetic circuit is mitigated due to the existence of the system resistance R_s and the transformer magnetizing resistance R_i. Each transformer magnetic circuit is mitigated mutually due to system resistance R_s, the more transformers reclosing at the same time, the stronger weakening effect of the system R_s. The inrush current value of each transformer is smaller than single transformer when multi-transformer reclosing at the same time.

The reduce of the main flux ψ_{pi} can also influences the inrush current non-periodic component I_{iD} of each on-board traction transformer, it makes I_{iD} become small resulting that the changes of each transformer flux $\Delta\psi_{pi}$ mitigates. With the increasing number of on-board transformers reclosing at the same time, each additional trend caused by an on-board transformer is reduced gradually. The flux can be decreased as long as the existence of inrush current, so the inrush current reduction can never be zero.

4 The Simulation of EMU Passing Neutral Section

This paper analyzes the over-current of the EMU passing neutral section process by using matlab/simulink. The on-board traction transformer capacity is 6100KVA in the simulation, rated voltage is 25000V/1350V, rated current of primary side is 240A, leakage inductance is 1.6mH. The simulation circuit model is showed as Figure 2:

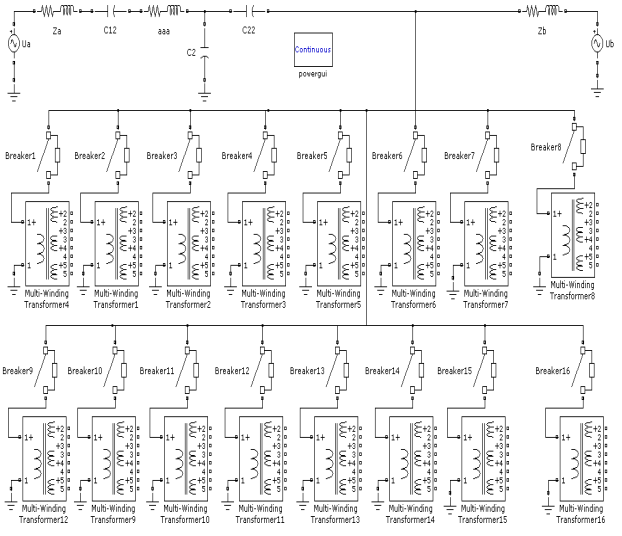


Fig. 2. Over-current simulation model of EMU passing neutral section

The process of multi-transformer reclosing is simulated by using simulation model in Figure 2.The maximum of inrush current can be obtained by setting the closing angle to 0°.The simulation result of different numbers of transformer reclosing at the same time is as followed:

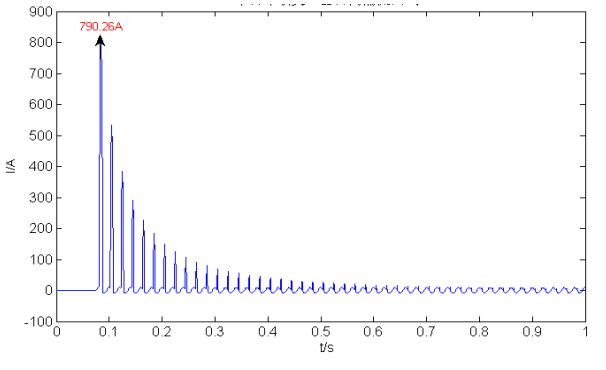


Fig. 3. Inrush current curve of single traction transformer

Figure 3-6 is the inrush current of multi-transformer, It can be seen from the figure that the inrush current value of multi-transformer is not integer times to the single traction transformer but far less than the linear superposition value. The main reason is the mutual demagnetization between each transformer as mentioned before. Table 1 reflects the relation between the number of transformers and the inrush current value.

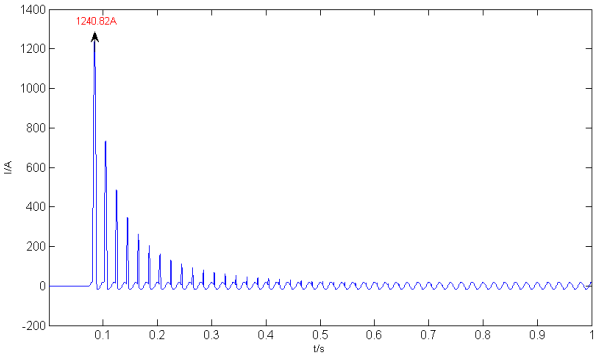


Fig. 4. Inrush current curve of two traction transformers

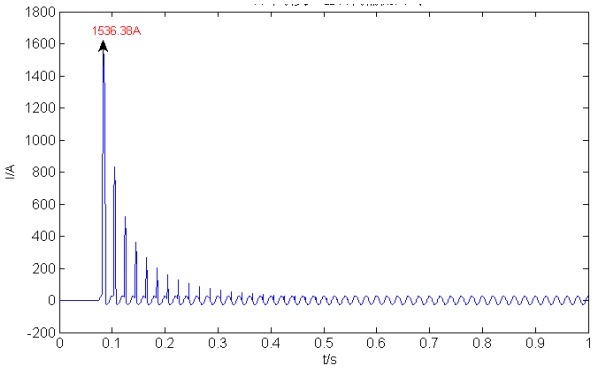


Fig. 5. Inrush current curve of three traction transformers

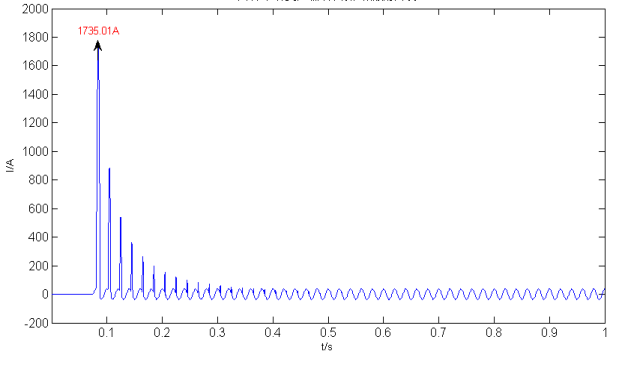


Fig. 6. Inrush current curve of four traction transformers

Table 1. The data of inrush current while EMU passing neutral section

Number of transformer	Peak value of each transformer (A)	Number of transformer	Peak value of each transformer (A)
1	790.2603	2	620.4120
3	510.7439	4	433.7541
5	377.6419	6	333.6711
7	298.8209	8	270.4516
9	247.1621	10	227.3595
11	210.2958	12	195.8737
13	183.1378	14	171.5603
15	161.8683	16	52.8928

Figure 7 is the reduction curve of reclosing inrush current, the abscissa is the number of transformers reclosing once and ordinate is the inrush current changes when increasing an on-board transformer. It can be seen from the figure that the inrush current value of each transformer reduced with the increasing number of transformers reclosing at the same time. The reduction tendency is firstly speed up, then the transformer main flux reduced with increasing number of transformers, so the reduction tendency slow down finally.

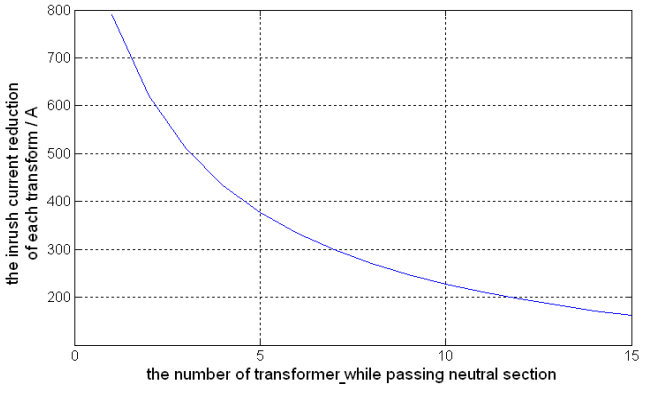


Fig. 7. The reduction tendency curve of reclosing inrush current value

5 Conclusions

This paper analyzes the theory and characteristic of the over-current phenomena while the EMU passing neutral section, builds the mathematical expression of synthetic flux while multi-transformer of EMU reclosing in parallel. The characteristics of over-current is analyzed when passing neutral section, then the simulation model is set up by using matlab/simulink to investigate the result of theoretical analysis, the conclusions are as followed:

(1) The value of inrush current is not just linear superposition of the value of single transformer when EMU passing neutral section. Due to the reduction effect between traction transformer, the degree of transformer saturation is reduced, so the inrush current value of each transformer is reduced.

(2) The degree of inrush current reduction is not a constant but a non-linear variable. The decrease degree of single inrush current is related to the value of the other transformers by the mathematical derivation. With the increase of transformer number in parallel, the flux saturation degree of each transformer is reduced. The tendency of single transformer inrush current reduction is also slow down gradually. So in the practical application of engineering, the number of transformer reclosing at the same time should be arranged under a reasonable circumstances so as to achieve the best operation result when passing neutral section.

References

1. Xiong, J., Yang, X.-L.: Analysis of the flash current of electric locomotive′s no-load turn-on. Electric Locomotives & Mass Transit Vehicles (06) (2005)
2. Ma, G., Wu, G.-N., Wang, T.: Simulation of Excitation Surge Current in Electric Locomotive and Analysis of Influence to Public Grid. Electric Railway (01) (2010)
3. Zhou, F.-L., Li, Q.-Z., He, J.-M.: Research on Simulation, Practical Measurement and Mechanism of Locomotive Over-voltage and Passing Neutral-section Based on Probability. Electric Drive for Locomotives (06) (2008)
4. Zhang, X.-S., He, B.-T., Zhang, J.-S.: Principle and Influencing Factors of the Transformer Sympathetic Inrush. Automation of Electric Power Systems 29(6), 15–19 (2005)
5. Huang, S.P., Li, Y.J.: MATLAB-based simulation of instantaneous change process when transformer no-load switching. Relay 32(8), 91 (2004)
6. Hao, Z.-G., Zhang, B.-H., Chu, Y.-L.: Trend and Situation of Distinguish Technology of Transformer Magnetizing Inrush Current. Transformer 42(7), 23 (2005)

Investigation of the Sympathetic Inrush Influence While the EMU Passing Neutral Section

Xin Li*, Qun-zhan Li, and Fu-lin Zhou

School of Electrical Eng., Southwest Jiaotong University, Sichuan Chengdu 610031

Abstract. This paper investigates the phenomena of the sympathetic inrush influence on two cars while the EMU passing neutral section. The mathematical expression of the change of the flux linkage is built when two cars running on the same line. Using the mathematical expression, the theory and characteristic of the sympathetic inrush is analyzed. At last, the theory is proved by the result of the simulation model which is built based on matlab/simulink.

Index Terms: passing neutral section, over-current, sympathetic inrush, flux linkage, EMU.

1 Introduction

In order to satisfy the increasing need of passengers traveling, the distance between two trains on the traction network begin to decrease with the departure gap of passenger dedicated line is cut down gradually. It makes the interaction between two running train become a hot issue which railway dispatchers most concerned. Due to the own characteristics of Chinese high speed railway, the on-board automatic mode is wildly used when EMU passing neutral section. This process requires to reclose the main circuit breaker while EMU passing neutral section in order to restore power supply into next supply arm. However, the magnetizing inrush current caused by the following train can lead to the inrush current of the train in normal operation. This is the sympathetic inrush current phenomenon.

The mechanism and influencing factors have been investigated in the literature [1~4], the simulation model of sympathetic inrush current has been built and the way has been proposed using the changes of transformer internal flux to analyze sympathetic inrush current. The research also analyzes the attenuation characteristic of sympathetic inrush current phenomenon. A general method analyzing sympathetic inrush current has been proposed in literature [5] by using matlab/simulink. However, these research are all focus on the power transformer. There is no research on the sympathetic inrush current caused by EMU passing neutral section. The sympathetic inrush current caused by EMU has its specificity due to its own characteristic of on-board traction transformer. This paper investigates the phenomena and

* Corresponding author.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 439–445.
springerlink.com

characteristic of the sympathetic inrush influence on two cars while the EMU passing neutral section, analyzes the impact factor of sympathetic inrush current. Finally, the simulation model is built to investigate the analysis result.

2 The Analysis Model of EMU

The magnetizing inrush current caused by the following EMU may lead the transformer magnetic bias of front EMU, This is the reason of the sympathetic inrush current phenomenon. In order to analyze the influence to on-board traction transformer when EMU passing neutral section simply, The two EMU model is equivalent to load and transformers in parallel. Equivalent circuit is shown in Figure 1. Transformer 1 and load is equivalent to front one(EMU in normal operation) , Transformer 2 is equivalent to the following one(EMU ready to reclose).

According to the equivalent circuit shown in Figure 1, the flux of the EMU should satisfy the equation (1), (2), and $R_s = R_s' + R_k$.

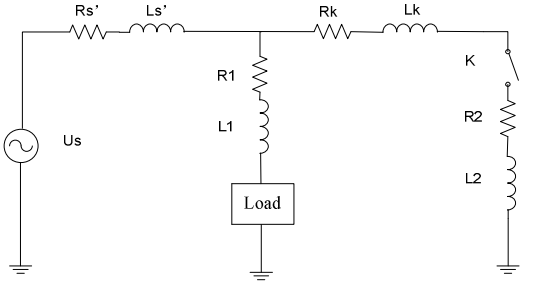


Fig. 1. Equivalent circuit of EMU reclosing in parallel

$$\frac{d\psi_{p1}}{dt} = u_s - (R_s - R_k)i_s - L_s' \frac{di_s}{dt} - R_1 i_1 \tag{1}$$

$$\frac{d\psi_{p2}}{dt} = u_s - (R_s - R_k)i_s - L_s' \frac{di_s}{dt} - R_k i_2 - L_k \frac{di_2}{dt} - R_2 i_2 \tag{2}$$

Equation (1), (2) is integrated in a cycle. Assuming power u_s is ideal sine wave, the integrating value is zero. Because of the dead angle of magnetizing inrush current, so $i_s(0) \approx i_s(2\pi) \approx 0$ [2]. Removing the periodic component of and integrating the non-periodic component of current i_1, i_2, i_s, the changes of transformer flux is as follows:

$$\Delta\psi_{p1} = -\int_0^{2\pi} [R_S' I_{SD} + R_1 I_{1D}] d\theta = -\int_0^{2\pi} (R_S I_{SD} - R_k I_{SD} + R_1 I_{1D}) d\theta \tag{3}$$

$$\Delta\psi_{p2} = -\int_0^{2\pi} [R_S' I_{SD} + (R_k + R_2) I_{2D}] d\theta = -\int_0^{2\pi} (R_S I_{SD} + R_k I_{1D} + R_2 I_{2D}) d\theta \tag{4}$$

In Equation 1, $I_{SD} = I_{1D} + I_{2D}$, $R_S = R_S^{'} + R_k$, current I_{1D} , I_{2D} , I_{SD} is the non-periodic component of i_1 , i_2 , i_s . At the beginning of the following one reclosing, the front one is in normal operation, so $I_{1D} \approx 0$. Assuming the magnetizing inrush current is positive direction caused by the following one. It can be seen that the magnetic bias of the front one is negative caused by the non-periodic component of the following. The inrush current of the front one begin in a opposite direction to the following current when the negative magnetic bias reaches to a certain level. The influence to sympathetic inrush of different distance and load of front one is investigated below.

A. The distance influence to sympathetic inrush when load of front one is certain

The value of R_k is smaller with the decrease of the two EMU distance, the absolute value of $\Delta\psi_{p1}$ becomes larger. The negative magnetic bias increase resulting in the inrush current increase of the front one. That is the front one is more vulnerable to the impact of following one to produce the sympathetic inrush current phenomenon. Because I_{1D} and I_{2D} are in opposite direction, so the front one inrush current have also an influence on the following one. It makes the inrush current of following one decrease.

B. The front load influence to sympathetic inrush when distance is certain

The current i_s become larger with the increase of the load, it leads to the decrease of transformer voltage. The transformer internal flux decrease in Equation 1. The magnetizing inrush current non-periodic component I_{1D}, I_{2D}, I_{SD} decrease as a result, and this makes $\Delta\psi_{p1}$ decrease. Due to $\Delta\psi_{p1}$ is the magnetic bias caused by inrush current of following one, it makes the front one inrush current become smaller. The sympathetic inrush current phenomenon would disappear with growth of the front load.

The current I_{SD} decrease with the increase of the front one inrush current because of the opposite direction of the two EMU. It can be seen from Equation (3) and (4) that the attenuation of network resistance R_S to the inrush current is weakened. It leads the $\Delta\psi_{p1}$ and $\Delta\psi_{p2}$ decrease and the attenuation speed slows down of sympathetic inrush current phenomenon. Due to the magnetizing resistance of on-board traction transformer is far more lager than the network resistance, so the attenuation speed is faster compared to ordinary power transformer.

3 The Simulation While EMU Passing Neutral Section

The simulation model is built by matlab/simulink, It is assumed that the EMU is 4M4T .All the on-board traction transformers close at the same time, and the loading rate of front one is 30%. In order to get the maximum value of inrush current, the closing angle is 0°. The simulation result is shown in Table 1.

Table 1. Simulation data of the sympathetic inrush current

situation of EMU / Dictance (km)	Peak value of following one inrush current (A)	Peak value of following one inrush current (A)
2.5	1586. 88	-494.66
5.0	1598. 16	-443.93
7.5	1610. 76	-380.39
10.0	1624. 81	-350.64
12.5	1640. 18	none

The simulation data shows that the tendency of the two EMU is opposite. With the increase of the distance, the following one current become larger and the front one become smaller. While the distance reaches a certain level, the sympathetic inrush current phenomenon disappears. The result proves the theory before. The inrush current wave of two EMU is shown as Figure 2 to 7.

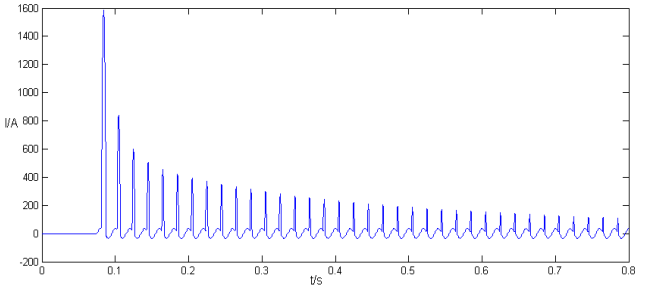


Fig. 2. The following EMU inrush current wave at 2.5km

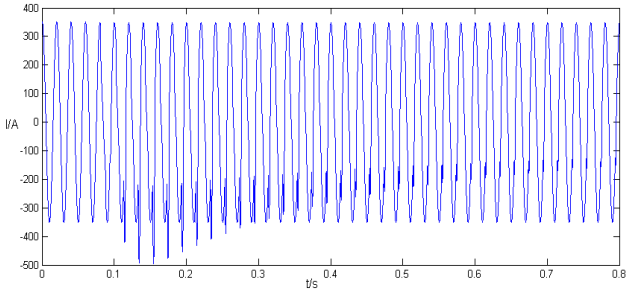


Fig. 3. The front EMU inrush current wave at 2.5km

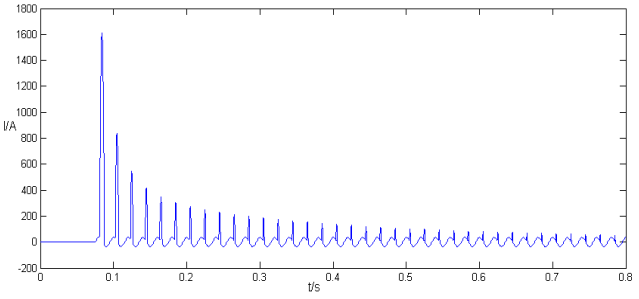


Fig. 4. The following EMU inrush current wave at 7.5km

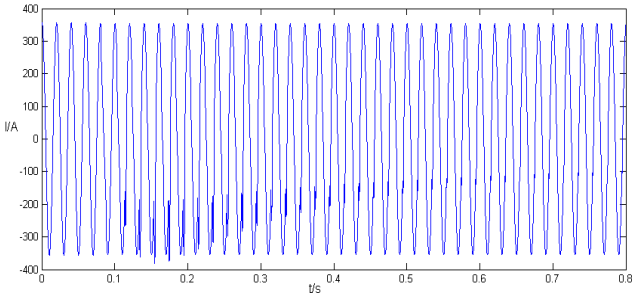


Fig. 5. The front EMU inrush current wave at 7.5km

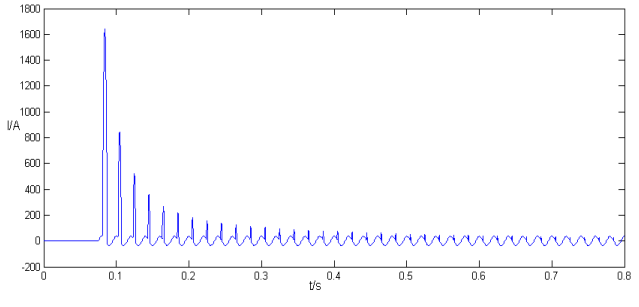


Fig. 6. The following EMU inrush current wave at 12.5km

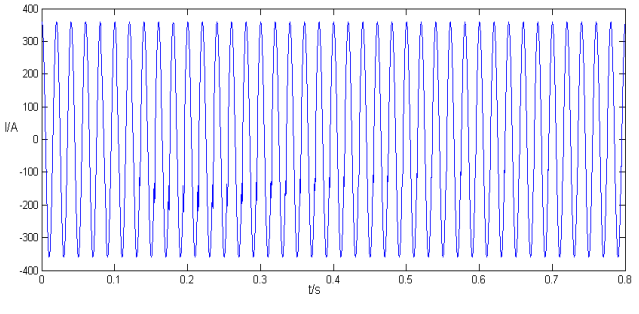


Fig. 7. The front EMU inrush current wave at 12.5km

It can be seen from the front EMU inrush current wave that compared to ordinary power transformer the attenuation speed is faster. As analyzed before, the reason is the magnetizing resistance of on-board traction transformer is far more lager than the network resistance, so the attenuation speed is mainly determined by transformer internal magnetizing resistance.

The distance of two EMU is set up to be 2.5km. Changing front EMU loading rate to be 20%, 30%, 40%, the wave of front EMU inrush current is shown as Figure 8 to 10:

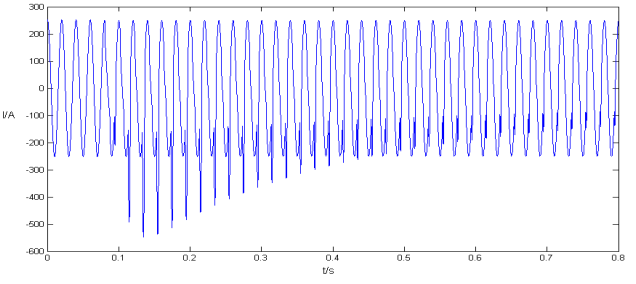


Fig. 8. The front EMU inrush current wave at loading rate 20%

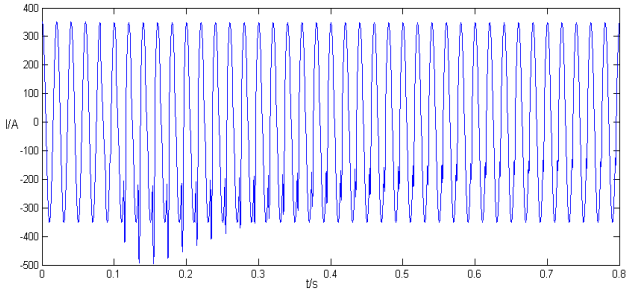


Fig. 9. The front EMU inrush current wave at loading rate 30%

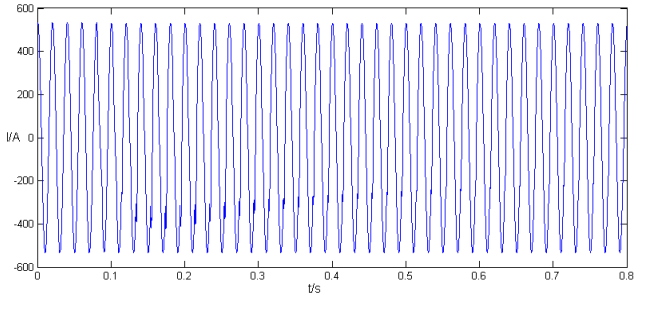


Fig. 10. The front EMU inrush current wave at loading rate 40%

The simulation result shows the front inrush current decrease with the growth of the load. The result proves the analysis before is correct.

4 Conclusions

This paper investigates the phenomena of the sympathetic inrush current influence on two cars while the EMU passing neutral section and builds the mathematical expression of the change of the flux. The characteristic of the sympathetic inrush current and relation between two EMU are analyzed, then the simulation model is built to prove the theory. The conclusions are as followed:

(1) When the distance of two EMU is close, the following EMU inrush current can influence the front one in normal operation. That is the sympathetic inrush current phenomena. The longer the distance, the smaller the inrush current is. The sympathetic inrush current phenomenon would disappear with increase of distance.

(2) When the distance is certain, the lower the front EMU loading rate, the easier the sympathetic inrush current occurs. The inrush current of the following one decrease with the growth of the front one load.

(3) Due to the magnetizing resistance of on-board traction transformer is far more lager than the network resistance, so the attenuation speed is faster compared to ordinary power transformer.

In summary, the relationship between EMU loading rate and sympathetic inrush current should be considered comprehensively before the departure distance is determined, so as to arrange the operation map scientifically and make comprehensive benefits achieve optimal.

References

1. Sun, X., Shu, H., Yu, J.: Comparison of sympathetic inrush influence on differential protection between parallel and series transformers. Electric Power Automation Equipment (03) (2009)
2. Zhang, X.-S., He, B.-T., Zhang, J.-S.: Principle and Influencing Factors of the Transformer Sympathetic Inrush. Automation of Electric Power Systems 29(6), 15–19 (2005)
3. Shu, H.-C., He, X., Li, L.-X.: Research on sympathetic inrush in operating transformer. Electric Power Automation Equipment (10) (2006)
4. Yuan, Y.-B., Li, D.-J., Lu, Y.-P.: Physical Mechanism of Sympathetic Inrush of Transformer and Its Influence on Differential Protection. Automation of Electric Power Systems (06) (2005)
5. Lu, S.-S., Du, J.-W., Chen, B.-L.: Study on Simulation of Sympathetic Current for Three Phase Transformer Based on MATLAB. Electric Switchgear (02) (2007)
6. Bi, D.-Q., Wang, X.-H., Li, D.-J.: Theory Analysis of the Sympathetic Inrush in Operating Transformers. Automation of Electric Power Systems (06) (2005)

A Vehicle Identification System Based on the Ultrasonic Technology

Wenhong Lv, Jifen Zhang, Youfeng Chen, Anliang Li, and Yanxia Wang

Shandong University of Science and Technology, Qingdao, Shandong, 266510, China
xbgjcl@126.com

Abstract. With the increasingly serious problems of traffic, the intelligent traffic system (ITS) has became a research hotspot, vehicle type identification is the basic of ITS. The paper studies on a vehicle type identification system based on ultrasonic technology, designs a vehicle type identification system which has sensor arrays with higher veracity and widespread applications. In this thesis, the research actuality of vehicle type identification technology at home and abroad in recent years was introduced, the composing of the system, the working theory of the sensor terminal and the processing method of collected data were particular introduced.

Keywords: ultrasonic technology, vehicle type identification, intelligent traffic system, method of least squares Introduction.

Vehicle type identification technology in recent years has been developing with the development of intelligent transportation system (ITS) [1] and highway no-waiting toll collection system. The existing vehicle type identification methods[2] mainly as follows: Classification based on ultrasonic detection, radar identification method, classification based on infrared detection, classification based on pressure sensors detection[3], loop-coil identification method[4], classification based on video detection using image processing technology[5], classification of RFID based on wireless communication technology, etc. Considering the advantages and disadvantages of these models, combined with actual situation and the requirements of measurement precision, this paper adopts ultrasonic sensors for research of vehicles classification systems. Ultrasonic sensor has advantages of simple structure, small volume, low cost, simple and reliable handling information processing, etc.

1 Ultrasonic Detection Principle and the System Structure

1.1 Ultrasonic Detection Principle

Ultrasonic ranging calculates the distance between the object and the ultrasonic probe by echo time difference and the known speed of sound, the specific formulas is

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 447–454.
springerlink.com

$$d = \frac{c \cdot t}{2} \tag{1}$$

Where d is the measured distance, c is real-time speed of sound; t is the time from sending ultrasonic pulses by the sensor to receiving the echo.

The mathematical model used for the velocity of sound in air is

$$v = \sqrt{\frac{\gamma \lambda T}{\mu}} \tag{2}$$

Where γ is gas specific heat capacity, λ is molar gas constant, μ is gas molar mass, T is thermodynamics temperature. In terms of air, specific heat capacity, molar gas constant and gas molar mass are constant, so the velocity of sound in air is only related to temperature, that is to say, real-time speed of sound is affected by the temperature, through the derivation of the approximate ultrasonic velocity speed in air formula is

$$c = 331.45\sqrt{1 + \frac{T_C}{273.16}} \approx 331.45 + 0.607\, T_C \tag{3}$$

Where T_C is centigrade. In this way, we can get more accurately results of vehicle parameters in the detection.

1.2 The System Structure

Although the current vehicle identification method can detect the parameters of moving vehicles and identify the vehicle type, accuracy of detection parameters and identification will be reduced by a certain extent when the vehicle in high speed or bad weather interference. Therefore, to find a vehicle parameter detection method with stability and accuracy is the key to achieving vehicle type identification. In this study, we adopt the way to distribute the sensors array on double gantry, as shown in Figure 1.

The double gantry is erected on a one-way lane according to the defined distance with 5 m high, 4 m wide, and 10 m between the two gantry (In practical applications the distance between the double gantry can also be set according to the site condition and need of test).The digital temperature sensors and infrared sensors will be erected on roadside of the first gantry, multiple detection terminal will be erected on the two columns and beams of the gantry, each test terminal is controlled by a microcontroller. In order to avoid the interference of the detection devices, we should use different operating frequency of ultrasonic transducer. In a computer terminal server, the data from detection terminals will be processed, so we can get a more accurate vehicle length and width, vehicle truck data matching will be operated, and then identify the model.

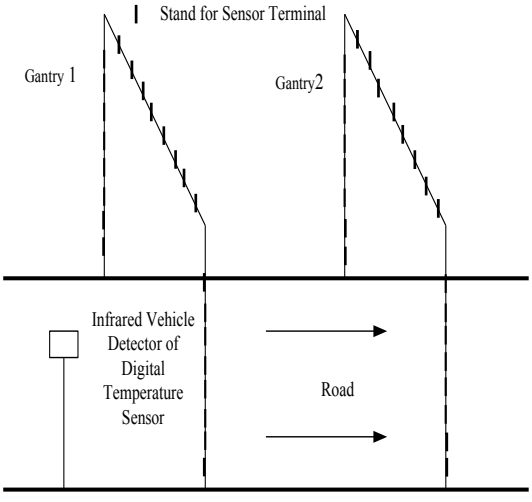


Fig. 1. Distribution of the system detecting terminal

The block diagram of the system is shown in Figure.2.

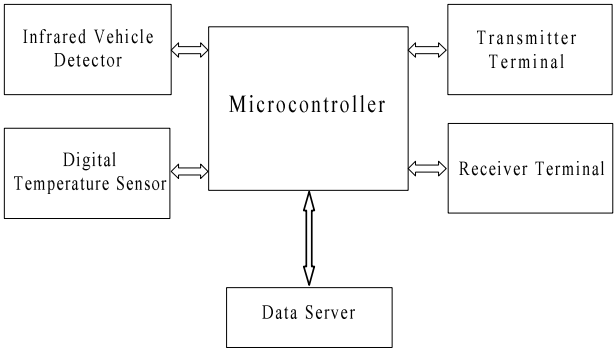


Fig. 2. Frame of the system

1.3 Detection Methods of Physical Quantities

Vehicle width: when the system begins to work, the detection terminals on the gantry 1 and 2 of the column send co-rotating ultrasonic signals at the same time to form a wave-front and receive the back wave. The time interval of the transmit wave and echo is t in formula (1), because the width of gantry is fixed, once c is determined by formula (3), the vehicle width can be carried out. So as the principle of detecting vehicle height.

Vehicle speed: the MCS of the double gantry record the enter time (t_1) and exit time (t_2) of the vehicle and the distance between the two gantries is D, then we can calculate the average speed of the vehicle according to the formula (4).

$$v = D/(t_2 - t_1) \tag{4}$$

Vehicle length: The average speed has been obtained, the time difference of the vehicle passing the each frame is rectifiable respectively, and then the vehicle length approximation can be obtained.

2 Ultrasound Transceiver Circuit of Detection Terminala

According to the characteristics and requirement, we choose piezoelectric ultrasonic sensors in this system. We designed the detection terminal in which the receiver and the sender are separate. That is to say, placing in a test terminal send and receive ultrasonic sensors which executive single function respectively, in order to avoid blind detection to the largest extent. Each beam and column is installed with 8 groups of terminal controlled by a STC89C52, 8 groups' terminals occupy two I/O port of the microcontroller to receive and transmit ultrasonic.

Ultrasonic transmission circuit (Figure 3): The microcontroller outputs pulse signal, then the signal will be processed by zoom-driven ultrasonic sensor to launch ultrasonic to realize generating detection signal.

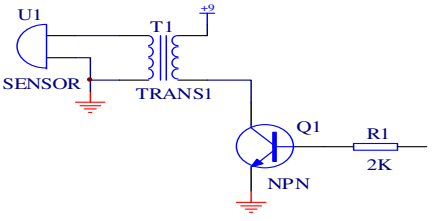


Fig. 3. Ultrasonic emission circuit

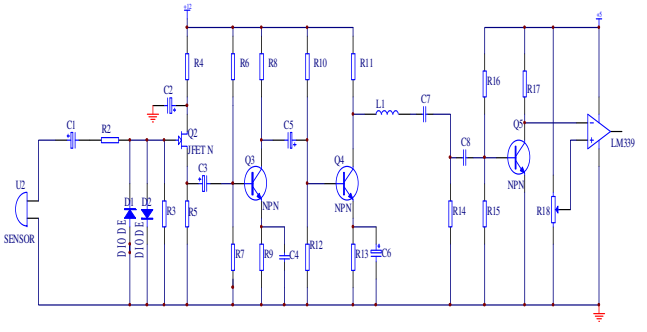


Fig. 4. Ultrasonic signal receive circuit

Ultrasound signal receiving circuit (Figure 4): Because when ultrasonic spread in the air there will be some attenuation, so the reflection signal is very weak and noise seriously. Therefore, when the ultrasonic sensor receives the signal we should first

carry out amplification, filtering, analog-digital conversion before it can be sent to microcontroller for data processing.

3 Terminal Software Design

3.1 Work Process of Detection Terminal

Infrared vehicle detectors and digital temperature sensor begin to work after the system powers on, when the infrared vehicle detector detects a vehicle passing, the detection terminals on double gantry start working. Take the detection terminal on beam (40 kHz) as an example, microcontroller sends pulse signal of 40kHz to drive the 8 groups terminals to launch ultrasonic and starts the timer, when echo is detected,

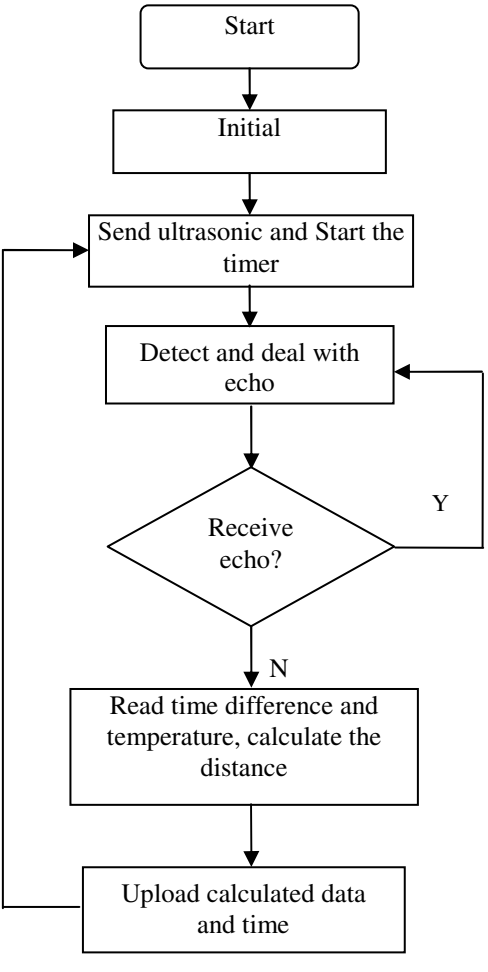


Fig. 5. Terminal software flowchart

read the time difference and temperature, calculate real-time velocity and the distance between the roof and the beam of double gantry. Then the time difference and the calculated distance will be transmitted through the RS485 bus to the server which will do further data process. Column detection is similar, but SCM need to send the pulse signal with different frequencies, because the ultrasonic sensors' working frequencies is different.

3.2 Terminal Software Flowchart

Terminal software flowchart is shown in Figure 5.

4 Experimental Data Processing

During the process of the system working, server will receive large amounts of data, the data processing is another key to identify vehicle accurately. Server receives data mainly include the time difference and distance uploaded from columns and the beam of double gantry. There are multiple sets of data, in which we need is the minimum distance to identify vehicle, in other words, what we need is the maximum of vehicle height, length and width. Due to the temperature compensation of sound velocity is not accurate, the air humidity influence to velocity is not be considered, detection terminal execute instruction will take time, so there's deviation between test data and actual value. Table 1 lists the deviation between test value and the actual value (Listed value is the distance between the object and the beam).

Table 1. The Measurement and the fact

cm

Actual value(y_i)	90	120	150	200	250	300	350	400
Test value (x_i)	98.0	126.8	159.2	207.4	256	307.6	357.8	405.8

From Table 1, we can see that the deviation between test value and actual value is about 7 cm, in addition, the 8 groups of data in the rectangular coordinate system is roughly linear. In order to make more accurate measurements, we used least-squares method for revising the data [8].

4.1 Data Processing

$$\text{Make } F(a,b) = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2 ,$$

To make $F(a,b)$ minimum, by the extremism principle: $\frac{\partial F}{\partial a} = \frac{\partial F}{\partial b} = 0$,

namely $\begin{cases} \frac{\partial F}{\partial a} = -2\sum_{i=1}^{n} x_i(y_i - ax_i - b) = 0 \\ \frac{\partial F}{\partial b} = -2\sum_{i=1}^{n} (y_i - ax_i - b) = 0 \end{cases}$, then we can get the formula

$\begin{cases} a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \\ b = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{a}{n}\sum_{i=1}^{n} x_i \end{cases}$, combine the 8 group data with the formula and

gain the result $\begin{cases} a \approx 1.0045 \\ b \approx -8.4042 \end{cases}$, finally the equation is available: $y = 1.0045x - 8.4042$.

Substituting the measured values into the equation above, the revised measurement value and the actual value are shown in Table 2.

Table 2. The revised value and actual value

cm

Actual value	90	120	150	200	250	300	350	400
Revised value	90.0368	118.966	151.512	199.929	248.748	300.580	351.006	399.222

As we can see from Table 2, after revised the maximum error between measurement values and actual value is 1.5 cm, which can meet the requirement of vehicle identification accuracy.

When the server gets revised data of vehicle width, length and height, it will compare the data with model data to judge vehicle type, then save the data and the vehicle type code into the database, so as to execute subsequent operations such as calculating tolls, recording data of passed vehicles type, allocating traffic cost, etc.

5 Summary

Through ultrasonic sensor array, the system detects vehicle parameter of height, width and length and uploads the parameters to the server. Then the server revises the vehicle data and compares it with the standard vehicle type data in the database so as to achieve the purpose of identifying vehicle type. This method is better than other methods on accuracy and the speed of identification, it has a great potential for development and wide space of application.

Through experiments, the accuracy of the vehicle parameter collected by system can meet the accuracy requirements for vehicle identification, in terms of speed detection, data transmission and recording speed, the system can all meet the design requirements. Otherwise it has good electromagnetic compatibility [9] and lays a good foundation for subsequent operations. But the system can not identify trailer and retrofit car very accurately, so we need a more accurate algorithm to improve it, and this will be the key work in future.

References

[1] Ou, D.: Transportation Information Technologise. Tongji University Press, Shanghai (2007)
[2] Li, H.: The Design of Vehicle Detection Identification System. Changchun University of Science and Technology, Changchun (2004)
[3] Zhang, J., Zhang, J., Dai, Z.: Vehicle Classification System Based on Pressure Sensor Array. Journal of Highway and Transportation Research and Development 3(7), 125–129 (2006)
[4] Zhu, H.: A Vehicle Classification System Based on Inductive Loop Detector. Southwest Jiaotong University, Chengdu (2003)
[5] Cao, Z.: Vehicle Detection and Classification Based on Video Sequences. Zhe Jiang University, Hangzhou (2004)
[6] Yin, X., Zhang, J., Gao, S., et al.: Application of Ultrasonic Technique in Measurement. Modern Electronic Ttechnology 5, 100–102 (2003)
[7] Zhang, J., Li, G.: Study and design of ultrasonic measuring system. Journal of Hefei University of Technology 27(6), 640–643 (2004)
[8] Liu, R., Mei, X.: A Data Processing System of Least Square Method. Computer Knowledge and Technology 5, 1345–1346 (2007)
[9] Lv, W., Guo, Y., Kang, F., et al.: Principle and Application of Electromagnetic Compatibility Tutorial, 2nd edn. Tsinghua University Press, Beijing (2008)

Computing Stability for Autopilots Based on Vector Margin

Jiang Wang, De-fu Lin, and Jun-fang Fan

School of Aerospace, Beijing Institute of Technology, Beijing, China
School of Automation, Beijing Information Science & Technology University, Beijing, China
Wjbest2003@163.com, wyhffjf@bistu.edu.cn

Abstract. A novel stability computation approach for tactical missile autopilots was detailed. The limitations of traditional stability margins were exhibited. Then the vector margin was introduced and examined. All available linear autopilot topologies were given and evaluated on the basis of vector margin. The analysis and computation results show that the presented method could give well description for control system stability.

Keywords: stability computation, autopilot, vector margin.

1 Introduction

Longitudinal autopilots for tactical missiles have been successfully employed for over decades, and the three-loop autopilot from U.S. Raytheon has been the design topology of choice [1]. Recently, requirements for the new generation air-to-air missile and the advanced missile interceptor, particularly with respect to the capability to engage highly agile fighter and tactical ballistic missile, and achieve precision end-game trajectory in seconds or less, have prompted a revision and research of the way in which the guidance and autopilot design is undertaken [2-4]. Therefore, it is well worth analyzing the new design and performance evaluation approach for most tactical missile autopilots based on the three-loop topology and all available feedback variables. Due to the limitations of common gain margin (*GM*) and phase margin (*PM*), the vector margin (*VM*) approach is analyzed and applied in this paper to examine the flight control stability.

2 Vector Margin and Stability

The frequency-domain stability specifications of gain margin (*GM*) and phase margin (*PM*) for an open-loop system have been adopted widely, as shown in Fig. 1.

However, both *GM* and *PM* could only describe the system stability partially. Given open-loop system *PC*, the frequency-domain stability can be expressed by φ_m and *gm* as defined in Fig. 1, which actually describes the distance between system Nyquist diagram and critical point indirectly. Similarly, the vector margin (*VM*) can be

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 455–460.
springerlink.com

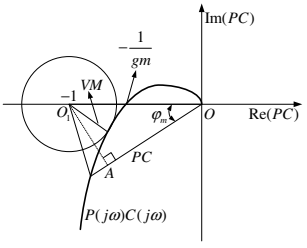


Fig. 1. Stability evaluation using GM and PM

defined as a measurement of nearness to the critical point (-1, 0) on the Nyquist diagram given by

$$VM = \min_{\omega} \left| 1 + P(j\omega)C(j\omega) \right| \tag{1}$$

From Fig. 1, inequalities can be obtained as

$$\begin{cases} VM \le \left| O_1 A \right| = \sin \varphi_m \\ -1/gm - (-1) \ge VM \end{cases}$$

which can be rewritten into

$$\begin{cases} \varphi_m \ge \arcsin(VM) \\ gm \ge 1/(1 - VM) \end{cases} \tag{2}$$

Such that both *GM* and *PM* could only give an approximate boundary of *VM*.

For the Sensitivity Function *S*

$$S = \frac{1}{1 + PC}$$

the maximum of *S* is defined as M_S and can be written as

$$M_s = \max_{\omega} \left| S(j\omega) \right| = \frac{1}{VM} \tag{3}$$

Therefore, a greater M_S means *PC* is closer to point (-1, 01) and the system is more unstable, which is equal to a smaller *VM*. Some control systems have well *GM* and *PM* values but fail in *VM*. Compared with *GM* and *PM*, *VM* presents the essential robust stability.

The maximum of sensitivity function for a well-designed control system is usually less than three, such that the vector margin should be over 1/3.

Besides, for many control system, even though the closed loop characteristics are the same, the open loop transfer functions are not. These are multi-loop systems which close their loops in different ways, and thus the open-loop systems can be different, even though they all lead to the same closed loop system. The particular robustness measure for tactical missile autopilots is an unstructured perturbation at the

input. The measure of robustness to this perturbation can be measured by the *VM* of the open loop system broken at the input to the plant.

3 Autopilot Topology

The longitudinal (pitch axis) missile dynamics can be described using the short period approximation of the longitudinal equations of motion [5]. Written in differential equation notation the basic plant is

$$\begin{aligned} \dot{\alpha} &= -b_\alpha \alpha + \dot{\vartheta} - b_\delta \delta_z \\ \ddot{\vartheta} &= -a_\alpha \alpha - a_\omega \dot{\vartheta} - a_\delta \delta_z \\ \dot{\theta} &= b_\alpha \alpha + b_\delta \delta_z \\ a_y &= V\dot{\theta} \\ a_{ym} &= a_y + c\ddot{\vartheta} \end{aligned} \tag{4}$$

where, the measured normal acceleration (a_{ym}) from an accelerometer combines both of the translational acceleration and rotational acceleration.

The topology of the classic three-loop autopilot is shown in Fig. 2.

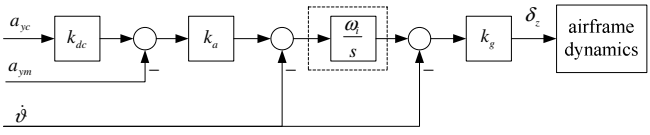


Fig. 2. Three-loop autopilot topology

and the control law can be expressed as

$$\delta_z = k_{dc}k_a\omega_i k_g \int a_{yc}dt - k_a\omega_i k_g \int a_{ym}dt - \omega_i k_g \int \dot{\vartheta}dt - k_g\dot{\vartheta} \tag{5}$$

There is also the two-loop acceleration autopilot from engineering experiences, as cancelling the dashed part in Fig. 2. Besides, the attitude autopilot can be obtained by substituting the acceleration command (a_{yc}) with attitude command (ϑ_c) from Fig. 2.

There are several three loop topologies that can be shown to be equivalent from a closed loop standpoint. It should review the possibilities firstly before we dive into all of these topologies. As shown in (5) there are potentially three measured quantities that can be used to form feedback quantities, including the acceleration (a_{ym}), the pitch rate ($\dot{\vartheta}$) and the fin deflection (δ_z). The last is not really measured, but since the fin command is output from the autopilot and there are no fin dynamics, a lag can be used instead of real feedback. Also, since the fin rate is the control, an integral of the variables could also be used for feedback. That is, $\int a_{ym}dt$, $\int \dot{\vartheta}dt$, $\int \delta_z dt$ are also available.

Table 1. All autopilot topologies

Topology 0	Topology 1	Topology 2	Topology 3	Topology 4	Topology 5
$\int \alpha dt$	$\int a_{ym} dt$	$\int a_{ym} dt$	$\int \dot{\vartheta} dt$	$\int a_{ym} dt$	$\int a_{ym} dt$
$\int \dot{\vartheta} dt$	$\int \dot{\vartheta} dt$	$\int \dot{\vartheta} dt$	$\int \delta_z dt$	a_{ym}	$\int \dot{\vartheta} dt$
$\int \delta_z dt$	$\dot{\vartheta}$	$\int \delta_z dt$	$\dot{\vartheta}$	$\dot{\vartheta}$	a_{ym}

Topology 6	Topology 7	Topology 8	Topology 9	Topology 10
$\int a_{ym} dt$	$\int \dot{\vartheta} dt$	$\int \dot{\vartheta} dt$	$\int \delta_z dt$	$\int a_{ym} dt$
$\int \delta_z dt$	a_{ym}	$\int \delta_z dt$	a_{ym}	$\int \delta_z dt$
a_{ym}	$\dot{\vartheta}$	a_{ym}	$\dot{\vartheta}$	$\dot{\vartheta}$

There are six measured variables for feedback, as following.

$$a_{ym},\ \dot{\vartheta},\ \delta_z,\ \int a_{ym} dt,\ \int \dot{\vartheta} dt,\ \int \delta_z dt$$

Table 1 presents the possibilities. Note that Topology 0 is full state feedback and is included only for comparison since angle of attack α is not measured. Table 1 does not include δ_z because those feedback formulations result in singular matrix difficulty; thus, these are not listed and topology 10 will not be discussed further.

Topology 1 is equal to the classic three-loop autopilot, and topology 4 equal to the two-loop acceleration autopilot with PI compensator. Note that the topology 4 has one additional zero point more than topology 1. Topology 3 uses only rate gyro and equal to the attitude autopilot.

Rewritten the longitudinal missile dynamics in state space notation the basic plant is

$$\dot{x} = A_m x + B_m u \quad y = C_m x + D_m u$$

where

$$x = \begin{bmatrix} \alpha & \dot{\vartheta} \end{bmatrix}^T \qquad y = \begin{bmatrix} a_{ym} & \dot{\vartheta} \end{bmatrix}^T \qquad u = \delta_z$$

Thus there exists a nonsingular transformation Γ_{ij} that makes autopilot shown in Table 1 have identical closed-loop characteristics [6], $i, j = 1, 2, \ldots, 10$.

The question then is which of the topologies are "better"? The closed loop responses are identical. However, the feedback mechanisms are different and therefore the open loop properties will be different. This means that different topologies will have different robustness properties for a given level of performance. This will be examined in the next section.

4 Stability Computation and Analysis

The response curves of previous ten autopilots could be identical when neglecting the actuator dynamics. However, when adding a second-order actuator dynamics into the closed system, topology 5 diverges seriously, and other curves show different features more or less [6], as shown in Fig. 3 and Fig. 4. Therefore, it should recheck the system robustness using a more precise approach.

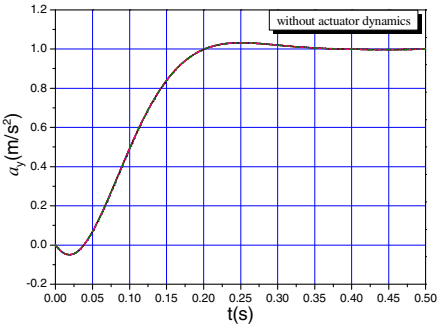


Fig. 3. Autopilot response without actuator dynamics

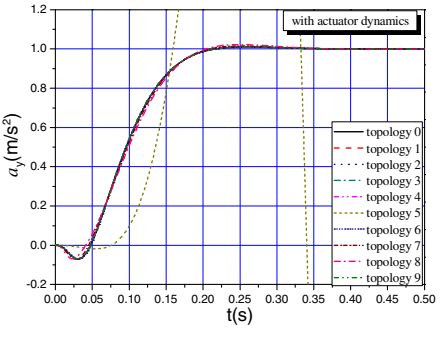


Fig. 4. Autopilot response with actuator dynamics

On the basis of vector margin, the autopilot stability is calculated. Wherein, for the topologies with direct feedback δ_z, two cases included the direct feedback and a lag loop substituted for δ_z are analyzed respectively.

It is obviously that topology 5 has the least *VM* value and could suffer from disturbance, which is identical to the time domain case.

The *VM* value for topology 0 will decrease from 1 to 0.8733 when using lag loop instead of direct feedback δ_z, which seems that topology 0 does not belong to a true full state feedback case.

Topology 1 known as the classic three-loop autopilot exhibits the best robustness properties. While the topology 4 equal to the two loop acceleration autopilot with PI compensator is next to the topology 1. Topology 1 uses only the measurable quantities and the acceleration integral instead of acceleration self, which thus

belongs to a true full state feedback control and has the *VM* value as 1. The classic three-loop autopilot provides a best choice for engineering application.

For topology 2, 3, 6, 8 and9, their *VM* values could drop greatly when using the lag loop as alternative feedback. Further analysis shows that all topology 5~9 has poor robustness properties and unworthy of investigation.

Table 2. All autopilot topologies

	Topology 0	Topology 1	Topology 2	Topology 3	Topology 4
feedback	1	-	1	1	-
δ_z	0.8733	1	0.8849	0.7240	0.9447
lag loop					

	Topology 5	Topology 6	Topology 7	Topology 8	Topology 9
feedback	-	0.9473	-	0.5591	0.9502
δ_z	0.0512	0.8533	0.6865	0.5388	0.5539
lag loop					

5 Conclusion

The vector margin based stability evaluation approach for control system particular multi-loop closed system has been presented in details. Ten three-loop autopilot topologies have been examined. It is demonstrated that given the same performance, the robustness characteristics are not the same. The calculation and analysis results show that the three-loop classic autopilot has the best robustness properties, and the two-loop acceleration autopilot with PI compensator is the second choice. This analysis is identical to engineering experiences.

References

1. Mracek, C.P., Ridgely, D.B.: Missile longitudinal autopilots: connections between optimal control and classical topologies. In: AIAA Guidance, Navigation and Control Conference and Exhibit, pp. 1–29 (August 2005)
2. Fleeman, E.L.: Technologies for future precision strike missile systems. RTO-EN-018, 5-1–5-15 (2001)
3. Zarchan, P.: Tactical and Strategic Missile Guidance, 5th edn. AIAA, Reston (2007)
4. Tekin, R., Ateşoğlu, Ö.: Modeling and Vertical Launch Analysis of an Aero and Thrust Vector Controlled Surface to Air Missile. In: AIAA Atmospheric Flight Mechanics Conference, pp. 1–17 (August 2010)
5. Qi, Z.-K.: Guided weapon control systems, pp. 143–153. Beijing Institute of Technology, Beijing (2003)
6. Fan, J.-F.: Research on autopilot design for static unstable missiles, Ph.D dissertation. Beijing Institute of Technology (2009)

Research on Recycling-Oriented Automotive Design

Y.Z. Yu

Mechanical Engineering Department,
Zhejiang Institute of Mechanical and Electrical Engineering, No. 528 of Binwen Road,
Binjiang District, Hangzhou City, 310053, P.R. China
cvn79@hotmail.com

Abstract. With the rapid increasing number of automobiles, China will face the growing problem of resource consumption and environmental pollution due to automobile manufacturing and waste management. For the purpose of sustainable development of automotive industry, the paper analyzed the reasons of inefficient recycling, and then illustrated recycling-oriented design concept and design process, which are based on eco-material selection, DFD principle, and feasibility of using remanufactured components.

Keywords: automotive industry, recycling-oriented, eco-material, disassembility, DFD.

1 Introduction

In august of 2011, Wardauto reported that the total number of automobiles in use on the earth exceeded 1 billion. China has 78 million registered automobiles, which makes China be the second large automobile owner after America. Referring to the statistic of CATARC (China Automotive Technology & Research Center), in China mainland, the number of registered automobile jumped from 20.53 million to 59.33million between 2002 and 2009 [1]. Meanwhile, the quantity of automobiles at end of life had a terrible increase as well. Fig.1 illustrated that the number of annual waste automobiles takes about 5.3% of automotive ownership of the last year though the proportion fluctuated. It was reported that China had 3 million waste automobiles for managing in 2011. The tendency predicts that after a decade, China will be facing a serious problem of waste automobiles management.

Automobiles consume great amount of natural resource and energy during manufacturing and operating. Waste automobiles are also potential environmental polluters if the materials such as heavy metal, fluids and plastics contained in the components were not properly managed. In recent years, Chinese automotive industry has made progress in energy saving and emission control, but automotive waste management has negligibly been improved since traditional solutions were still applied, such as reuse after rough repair, cutting into scraps as raw materials, etc. In some regions, waste vehicles were crushed and buried, or even left in the open-air without any processing. That was a waste of recyclable resources, and requires large area of land. More seriously, soil and underground water were in the risk of permanent pollution.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 461–466.
springerlonk.com

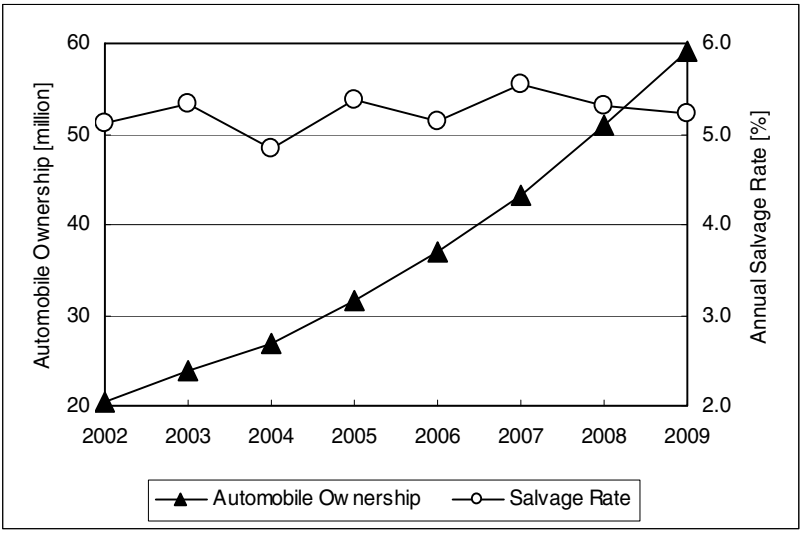


Fig. 1. Statistic of registered automobiles and waste rate in China between 2002 and 2009

2 The Concept of Design for Recycling (DFR)

Traditional automotive design was focusing on function, characteristics, safety, emission and cost, rather than the management of automotive wastes. As results, the materials lacked of recyclability, and nondestructive disassembly became complex and time consuming. Those barriers have made recycling inefficient and costly, and hardly be a business.

Automotive recycling should be taken into considering during design stage, rather than the waste problem became uncontrollable. A new design concept for sustainable development of automotive industry is known as DFR (design for recycling), which carefully evaluates raw material selection, application of DFD (design for disassembly), and feasibility of components reuse [2]. Moreover, low energy consumption and zero pollution are always be emphasized during the process of design and manufacturing.

3 Eco-material Selection

To minimize environmental pollution of automotive waste, eco-material should be considered. Ferrous metal, non ferrous metal, plastics, rubber, glass are commonly used materials for commercial vehicles. Fig.2 illustrated material weighting percentage of a family car. Even though the materials are safe and non-toxic, energy consumption and pollution may be unavoidable during manufacturing, and disposal. Therefore, eco-material selection became a challenge since it is concerned with environmental impact through the life cycle of the product [3]. Besides, quality and cost of the product are still vital for the manufacturer.

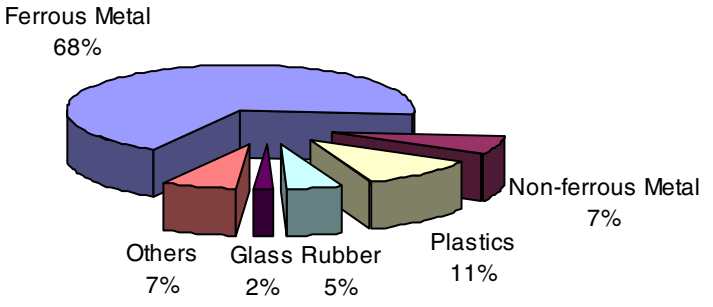


Fig. 2. Material weighting percentage of a family car

Sustainable development followed '3R principle (standing for reduce, reuse, recycle)' in terms of raw material utilization to save resource and to minimize pollution. Downsizing of family car was a typical application of 'reducing'. Besides controlling the consumption of hazardous and non-renewable materials, alternates should be developed and applied. Material selection also determined the management of components at end of life. Moreover, in an assembly that met functional requirements, homogeneous or compactable materials were prior to others. Take usual automotive textile as an example: polypropylene and polyamide (known as nylon) are commonly used as face fabric to wrap and bond around polyurethane foaming rubber by heat fusion method, which was difficult to separate when recycling. If using polyester fiber as the face fabric and its 3D woven as lining instead, the recycling became much convenient [4].

4 Design for Disassembly

Disassembly is defined as a process that releases constraints between parts by certain tools and methods to separate an assembly into independent parts, components or sub-assemblies. For the sake of profit, if the salvage value of a component could not afford the cost of disassembly, it is not likely to be recycled. Therefore, disassembility determined the possibility of recycling of a product at end of life.

Traditional automotive design put little attention on disassembility. In China, to save the waste managing cost, the main form of automotive disassembly was manual deconstructive separating aiming for raw materials, whose recycling required much economical and environmental expense because of bad purity [5]. If disassemblility of a product was adequately considered in design stage, recycling profit would be maximized due to significantly simplified disassembly.

Six principles should be followed through the process of DFD (design for disassembly). (1) Appling constraints tending to release, especially where the parts were in different materials and were expected to be thoroughly separated. (2) Minimized number of total parts. Considering to combine a group of parts into one if it was possible. (3) Using homogeneous or compatible materials could simplify sorting and improve recycling purity. Renewable material was in prior consideration. (4) Standardization and modulization design would improve efficiency of automatic disassembly because a series of products might have similar disassembly procedure. (5)

Explicit construction was prepared for parallel disassembly, i.e. most parts could be removed in any sequence. (6) Efficient disassemblydesign. That includes applying simple connection and clear marking for parts with high salvage value [6].

In a computer-aided design work, details of those principles would be stored and called as a construction evaluation knowledge base (KB) that helped engineers to make decisions. Fig.3 showed the guidelines of design for disassembly.

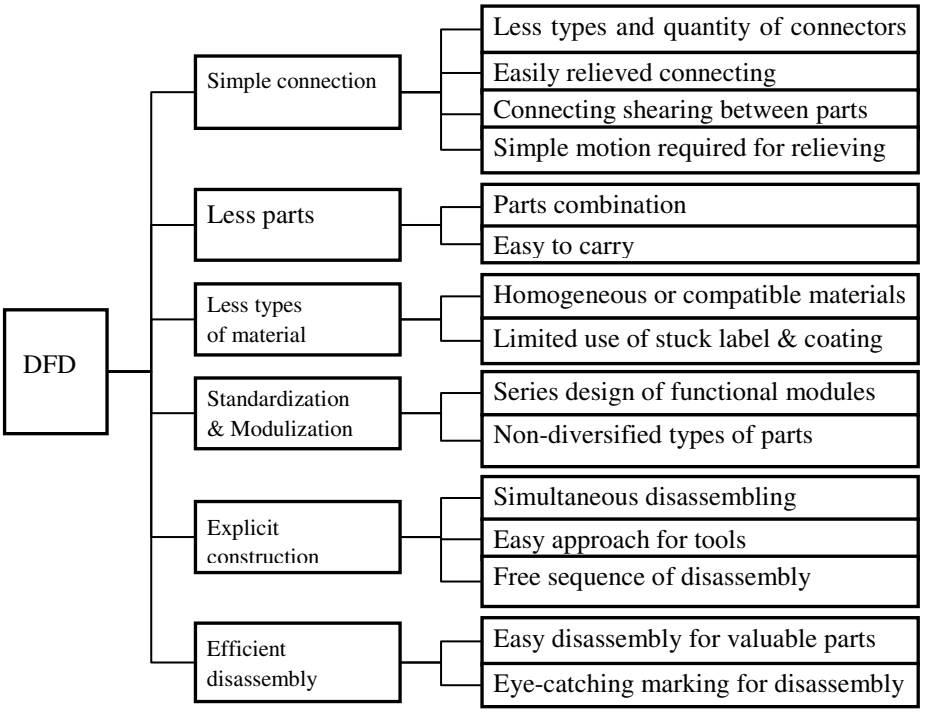


Fig. 3. Guidelines of DFD

5 Utilization of Remanufactured Components

The purpose of remanufacturing is to recover a salvage product by thorough disassembly, cleaning, repair, inspection, and reassembly so that it can work in the manner which it was intended. Most automotive components are renewable by remanufacturing, which results in much less energy consumption and pollution than that of manufacturing from raw materials. Recently, remanufacturing rates of salvage automotive components in developed countries are usually over 80%.

Remanufacturing was based on salvage components, which were in random conditions, and often had various constructions and specifications. In terms of technology and management, remanufacturing was much more complex than that of conventional production. A policy known as EPR (Extended Producer Responsibility) that came from Europe provided a solution. The policy required manufacturers to be responsible for their products through life cycle [7]. Since the manufacturers always

had the best understanding of their products, they were likely to have the greatest ability and responsibility for their salvage products. Therefore, by financial incentives, manufacturers were expected to develop environmentally-friendly products that have minimized pollution during manufacturing, operating and disposal.

It is recommended that the producers evaluate the remanufacturing possibility of components such as engines, transmission assemblies, braking systems, etc. when performing design work, and use remanufactured functional modules or parts as much as possible.

6 Conclusions

Recycling-oriented automotive design was focusing on low energy consumption during manufacturing and minimizing scrap during life cycle. The main tasks of recycling-oriented automotive design included eco-material selection, DFD, and using remanufactured components. Those tasks should be taken into consideration together since they were not independent to each other. Fig.4 illustrated the procedure of a recycling-oriented design.

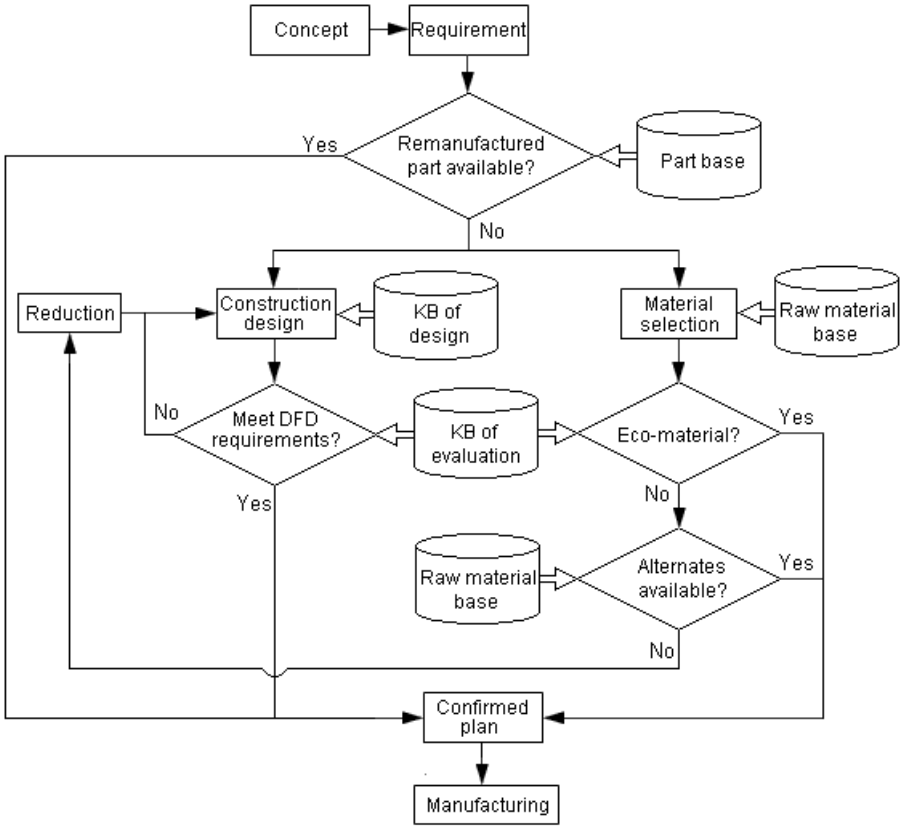


Fig. 4. Procedure of recycling-oriented design

Automotive industry plays an important role in Chinese national economy since it has contributed 13% of tax and 6.13% of GDP in 2010. With the rapid increasing ownership of automobile, looking for recycling solutions became urgent. Environmentally friendly design, manufacturing, operating, and recycling are the tendency of sustainable development in the near future.

References

1. Information on, http://www.catarc.ac.cn/ac/index.htm
2. Schiavone, F., Pierini, M., Eckert, V.: Strategy-based approach to eco-design: application to an automotive component. International Journal of Vehicle Design 46(2) (2008)
3. Yamamoto, H., Shibata, S., Neijenhuis, H.: Inverse factory for end-of-life-cars: Complete dismantling system. In: EcoDesign 1999: First International Symposium, pp. 942–945 (1999)
4. Chen, R., Liao, Q.: Development Trends of Automobile Textile under the Policy of Automotive Products Recycling. Technical textiles 8 (2010) (in Chinese)
5. Ferrao, P., Amaral, J.: Design for recycling in the automobile industry: new approaches and new tools. Journal of Engineering Design 17 (2006)
6. Fu, H., Cai, J.: Guidelines for Disassembly and Recycling-oriented Design. Mechanical Science and Technology 20(4) (2001)
7. Lindhqvist, T.: EPR in Europe. In: 2nd National Extended Producer Responsibility, Canada, pp. 14–18 (2002)

Influence Factors Analysis and Study of Emotion Absence in Distance Education

Yahui Sun[1], Jian Zhang[2], Jing Li[3], Guoliang Ma[1], and Lisheng Zhang[1]

[1] College of Light Industry, Hebei United University, TangShan, HeBei, China
[2] Tangshan College, TangShan, HeBei, China
yahui_2008@yeah.net, jianzhang1997@163.com, queen_jj@126.com

Abstract. With the rapid development of computer technology and network technology, the distance education has been a new height step by step. But the problem in distance teaching has become increasingly obvious. Athymia is an urgent problem in Virtual Environment. This paper analyses the influence factors and gives the Resolution Strategy from the teaching theory, Instructional Design, Learning Community and affective computing.

Keywords: Distance Education, Emotion Absence, Emotion Compensation, Affective Computing.

1 Introduction

With the rapid development of computer technology and network technology, the man-machine relationship of the computer and the human is more and more close. The integration of the computer and network technology into distance education has largely made up for the defects of traditional education, but it can't achieve the face-to-face interaction between teachers and students in the emotional comunication.

2 Influence Factors of Emotion Absence in Distance Education

In the environment of distance education, the physical space separation between teachers and students results in the loss of facial expression, pronunciation, intonation, emotional information brought about by facial expression during a transmission. The lack of emotional information will have a certain effect on the emotional interaction between teachers and students.

2.1 Teachers' Emotion Absence

Distance education is a form of education. Students and teachers, students and educational organizations mainly adopt various media ways in the way of systematic education and communication links. One of its key factors is the separation including the physical space separation of teachers and students, and the separation of schools and students. The impact on the distance education, referring to the separation of

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 467–470.
springerlink.com

classroom and campus cultural atmosphere, is far greater than the one of the physical separation. Under the environment of the distance education, the teacher-student interaction is virtual. But the emotional communication that occurred is real. We can't deny the true feelings between teachers and students because of the virtualization. The signal of true feelings is weakened greatly in the course of transmission. However, under the environment of the traditional classroom, techers can give a sigal to students and have certain impacts on their learning by an eye contact, a word, an action, the speed or the intonation. In the distance education, students are diffult to feel the concerns from the teachers. By contrast, they are easy to become confused and lazy.

2.2 Students' Emotion Absence

The reason from the students can be analysed from the students' physiology and psychology. On one hand, computer is the main teaching media in the network teaching. Compared with traditional teaching, students are no longer facing the blackboard and the teacher. As the long teaching time, students are easy to develop physical senses of fatigue and difficult to concentrate, owing to computer monitors and sound on students' visual, auditory stimulation. Without teacher's supervision and learning tips, students are easily absent-minded. On the other hand, pressures of study from outside parties, coupled with long time facing the unemotional network classroom, and lack of emotional communication make the students be under great psychological burden. These are the reasons of students' emotion absense.

2.3 Interactivity Difference between Teachers and Students

Interactivity difference between teachers and students is also the crucial reason of students' emotion absence. Compared with the traditional teaching, network classroom lacks the visual and vivid language of instruction such as postures, gestures and facial expressions of teachers. Interactive design focuses on the interactive of media operation, and less really considers to produce the helpful media contents, the strategies and interactions of teacher-student and student-student. So that it can't meet the needs of students' emotional expression and teacher-student' emotional commuciation.

2.4 Network Education Resource

The network education resource is the prerequisite and basis of the distance education. With the gradual development of the network education, the network education resource is more and more abundant. However, it is just the reason that the contents are uneven. If network education resource lacks reasonable resource management and information filtering, it will make the students not quickly find the resource they need, or even the "information getting lost" and the "information overload" will appear. Thereby it makes learners feel confused and helpless. Therefore, the effective management of network education resource is the key to carry out distance education.

3 Resolution Strategy of Emotion Absence in Distance Education

Many educators and psychologists' researches show that emotion greatly influences the choice of perception of, memory and thinking activity in the human cognitive process. Therefore, in distance education, we must reconstruct the connection of teaching behavior and learning behavior by applying Constructivism. Due to the students' emotional support to the research is relatively weak. We should use the educational theory and modern information technology to compensate the emotion absence.

3.1 Education Thought Blended into Constructivism and Humanism

Constructivism advocates the learning which is under the guidance of the teacher and learner-centered. Teachers are the helper and facilitator of the significant construction, rather than the people imparting and infusing knowledge to learners. Students are the main body of information processing, and the active constructor of meaning, rather than the passive recipients stimulated by the external and the objects infused. Blending the learning idea of Constructivism into the distance education, and using various techniques to construct the emotional support subsystem of the learning support system are the major constructions now.

Distance education bears the mission of lifelong education. In strongly promoting quality education's today, humane education and humane care should become the center of distance education. Humane care which is people-oriented pays more attention to pelple's development and the quality life's improvement. Therefore, in network education, we should establish the comprehensive development of humanistic idea, so that the network learners will be in an intelligent, humane, cultural realm.

3.2 Create Teaching Situation

By designing certain kinds of teaching situations, creating teaching situations make the teachers and students exchange their ideas and souls, activate students' prior experience, result in the desires of cognitive conflicts and exploration. They also let textbooks approach to students' life and psychology and enable their thoughts to constantly sublimate. Due to the telepathic communication between teachers and students is forgotten in distance education, we should create teaching situation which can trigger students souls. Before that, we not only know the important role of situation in teaching but also learn to operate according to the concrete teaching contents, the level of students the environmental conditions, so as to create vigorous teaching situation. Distance education demands teachers should have basic skills and can put the advanced educational ideas such as intuitive teaching, heuristic teaching and entertaining teaching into practice. In the limited time, the teaching situation can fully stimulate students desire for knowledge and enthusiasm for learning, so the network class will be vigorous.

3.3 Construct Learning Community

Learning community is a group composed by learners and their mentors (including teachers, experts, facilitators and peers).In the learning process, they often exchange

their ideas, share common learning resources and accomplish learning tasks. Thus the members form the interpersonal relationship that members influence and promote each other. This makes the construction of learning community be more significant to distance education which the learners disperse to each other and to learners' emotion compensation. Learning community not only achieves the learners to communicate, share experience and discuss to each other, but also plays an important role in stimulating interest, keeping motivation, having sense of belonging and reducing learning anxiety.

3.4 Emotion Compensation

We can use the emotional processing technology of emotional calculation to compensate emotion, and construct the embedded emotion interaction of distant education system by the related theory and technology of emotional calculation. For example, according to extract learner's vision, expression gestures and other emotional information, we can judge and know his emotional state, then gives him corresponding emotional encouragement or emotional compensation strategy to help him to solve the prlblems of emotion absence in distance education. We can also construct a virtual teacher with specific personality as the teacher of the distance education system by emotional modeling technology. So the human-computer interaction is more harmonious. Constructing a individual, optimistic and virtual teacher is the key point in this study.

Emotion absence in distance education is the long-term problem we are facing. For the distace education workers, how to establish a sound supprot service system for students, develop humanized curriculum materials, strenghten interaction of students and enhance the emotional experience of students is still an uphill journey.

References

1. Zhao, Q., Liu, F.: Influence Factors Analysis and Study of Emotion in Modern Distance Education. Journal of Xianning College (December 2009)
2. Che, Q.: Emotiona Absence Reasons of Emotional Distance Learners in Modern Distance Education. China Education Informatization (April 2010)
3. Wang, J., Ma, X., He, J.: Investigation and Counter measure of emotion. Absence in Modern Distance Education (April 2007)

Research of Mobile Location Estimation with Reducing NLOS Errors

Ning Liu[1,2], HaoShan Shi[2], Jie Wang[2], and ShuXia Guo[1]

[1] National Key Laboratory of UAV Specialty Technique
[2] School of Electronics and Information, Northwestern Polytechnic University, Xi'An, China

Abstract. With the rapid development of mobile communication networks, MS location accuracy is subjected to a concern increasing, and NLOS errors is one of key reasons that influence the accuracy of MS location. A hybrid TOA/AOA method which improves the [6] method is put forwarded. The method inherits the good performance of reference [6]'s method and uses AOA measurement to elevate the capability of location. The simulation results show that the method can improve the precise of MS location to some extend.

Keywords: NLOS, location, positioning accuracy, TOA/AO.

1 Introduction

In the mobile communication network environment, because of the build and terrain's effect, the signal can not only be transmitted through LOS between the mobile station and the base station, but also can reach the receiver through NLOS that is reflection, scattering, etc. So the energy and arrival angle of the signal will produce a certain bias and losses during the transmission. Because NLOS error exists commonly, it has become a key factor that affects the location accuracy of the mobile cellular network platform.

Many literatures have studies the location method in the Gaussian and NLOS environment. Chan method basing on TDOA and having higher location accuracy in Gaussian environment is proposed [2].The MS Location method also are introduced at the non-line sight environment [1][3][7]. The GLE method is introduced in [1]. The method improved the location accuracy's in the NLOS environment through limiting the location results within the framework of overlap by combining the geometric constraints between MS and BS with two step LS algorithm. The TOA\AOA method can improve the location accuracy in NLOS environment through supposing that some BS have a main reflective point exist [3]. The TDOA/AOA location method improves the location accuracy through adopting a Kalman filter to eliminate the error of TOA measurement [4]. It can also improve the location accuracy in the NLOS environment through processing the TOA measurement [5]-[7].

Location error would be produced in the NLOS environment because the transmitted signal will form excessive delays when it encountered obstacles. In view of the approach of dealing with TOA measurements in [6], this paper proposes a new TOA/AOA location method by means of suing the characteristics whenever TOA

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 471–477.
springerlink.com

measurements are always positive. This method eliminates the TOA measurements which have large error and retains the best TOA measurements; combines the AOA measurements and adopts WLS method through iteration gradually to decrease the error of location estimation.

2 The Proposed Hybrid TOA/AOA Location Scheme

Deal the TOA Measurement

In practical cellular mobile communication system, the actual transmitting time of signal is longer than the theory time because signal will be reflected and scattered by the buildings and terrains. If the location method in LOS environment is employed in the NLOS environment, it will produce considerable error. Analyzing from the time measurement model, the error caused by NLOS delay is proportional to the real distance between the base station and the mobile station. In another word, the longer distance the base station and mobile station is, the larger error produced in the NLOS is .So the location accuracy can be improved through selecting the optimum 3 TOA measurements to participate the calculation.

How to select the TOA value of the smaller error, the main ideas are as follows [6]: 1) In the same conditions, choosing the smallest measured value of TOA; 2) At least retaining the measured value of three base station for TOA\AOA hybrid location; 3) If there is one or more location round circle covered by one location round circle, the larger coverage is excluded; 4) If there are two separate round circles, both are retained; 5) If there is one location round circle covered completely by all the other location round circle, measured TOA values are retained.

TOA\AOA Hybrid Location Algorithm

Supposing the position of MS is (x, y) and the position of BS is (x_i, y_i), the function between them can be constructed, that is:

$$\begin{aligned} l_i^{\,2} &= (x_i - x)^2 + (y_i - y)^2 \\ &= x_i^{\,2} + y_i^{\,2} - (x^2 + y^2) - 2x_i x - 2y_i y \end{aligned} \tag{1}$$

$$l_i^{\,2} - K_i = -(x^2 + y^2) - 2x_i x - 2y_i y \tag{2}$$

Where $K_i = x_i^{\,2} + y_i^{\,2}$, t_i is TOA measurements. Supposing server BS can always supply the AOA measurements, the function can be constructed based on AOA measurements, that is:

$$\frac{y - y_1}{x - x_1} = tg\beta \tag{3}$$

$$\text{So} \quad y_1 - x_1 tg\beta = y - xtg\beta \tag{4}$$

Supposing $z_a = [x \quad y \quad R]^T$, where $R = x^2 + y^2$.

$$h = \begin{bmatrix} l_1{}^2 - K_1 \\ l_2{}^2 - K_2 \\ \vdots \\ l_m{}^2 - K_m \\ x_1 tg\beta - y_1 \end{bmatrix} \quad G_a = \begin{bmatrix} -2x_1 & -2y_1 & 1 \\ -2x_2 & -2x_2 & 1 \\ \vdots & \vdots & \vdots \\ -2x_m & -2y_m & 1 \\ tg\beta & -1 & 0 \end{bmatrix}$$

Using the function (2) and (4), establish the linear equation based on the z_a

$$h = G_a z_a \tag{5}$$

So the error function corresponding MS is

$$\varepsilon = h - G_a z_a{}^0 \tag{6}$$

Where $z_a{}^0$ is the real position of MS.

Defining $\psi = c^2 BQB$, $B = diag(t_1{}^0, ..., t_m{}^0, 1)$, $t_i{}^0$ is TOA measurements between MS and BS, Q is covariance array of each measurement including the distance error caused by TOA measurement and distance error caused by AOA measurement, the standard error corresponding AOA measurement is

$\delta \approx l * tg(\alpha) \approx l * \alpha$, where $l = ct$ is the measurement distance of MS and server BS. The z_a is pre-estimated by WLS.

$$\begin{aligned} z_a &= \arg\min((h - G_a z_a)^T Q^{-1} (h - G_a z_a)) \\ &= (G_a{}^T Q^{-1} G_a)^{-1} G_a{}^T Q^{-1} h \\ &= (G_a \psi^{-1} G_a)^{-1} G_a{}^T \psi^{-1} h \end{aligned} \tag{7}$$

To restrain the influence caused by NLOS error to location accuracy, the estimate value of MS position is taken as a pre-estimate value. The optimized location of MS $z_a{}'$ is estimated by WLS which corresponding parameter $B' = diag(t_1{}', ..., t_m{}', 1)$ is amended basing on the pre-estimated position of MS. $t_i{}'$ is the TOA measurements which had been amended. Above process supposes that the element of z_a is independent, but in fact R is related with MS position that $R^2 = x^2 + y^2$. The more accuracy location can be got by using this term. At the first estimate the covariance array of $z_a{}''$, that is

$$\text{cov}(z_a{}'') = (G_a{}^T \psi'^{-1} G_a)^{-1} \tag{8},$$

then construct new error vector $\varepsilon'' = h'' - G_a''z_a''$. where $h'' = \begin{bmatrix} z_{a1}^2 \\ z_{a2}^2 \\ z_{a3}^2 \end{bmatrix}$, z_{a1}, z_{a2}, z_{a3} are the elements of z_a', $G_a'' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$, $z_a'' = \begin{bmatrix} x^2 \\ y^2 \end{bmatrix}$.

Defining $\psi'' = 4B''Q''B''$, where $B'' = diag(x^0, y^0, 0.5)$, x^0, y^0 is the estimate position of MS, Q'' is $\mathrm{cov}(z_a'')$ which get from above process. Then z_a'' is estimate using ML method. $z_p = \pm\sqrt{z_a''}$, the value can be positive or negative. We can take the value in the interesting area as the final estimate position of MS.

3 Simulation and Analysis

To test and validate the location capability of the method, the simulation condition is as following: the radius of cell network is 1000 meter, it's distribution is as Fig 1; the calibrate error is independent and obey the Gauss distribute that its mean is 0 and variance is 0.1us; the delay caused by NLOS obey the exponent distributing whose parameter is $\tau_{i,rms} = T_1 d^{\sigma}\xi$, where $10\lg\xi$ obey Gauss distribution that its mean is 0 and variance is σ_ε。 Table1 shows the parameters of different channel. Monte Carlo is 10000 times.

Table 1. The parameter in the different channel

Channel	$T1(us)$	ε	$\sigma_\varepsilon(dB)$
hive	1.0	0.5	4
urban	0.4	0.5	4
suburb	0.3	0.5	4
exurb	0.1	0.5	4

Simulate 1: under the situation that MS moves by some track and leave the coordinate original gradually in the normal city, the location capabilities are compared between the method of this article, the method of [6] and normal TOA/AOA method. The result is as figure 2. The figures shows that the distance of MS and original have an effect to the method, the capability become worse with the farer distance, but the total capability is better than [6] and the normal TOA/AOA method.

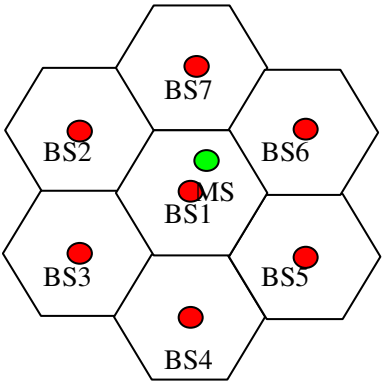


Fig. 1. The distribute of cell network

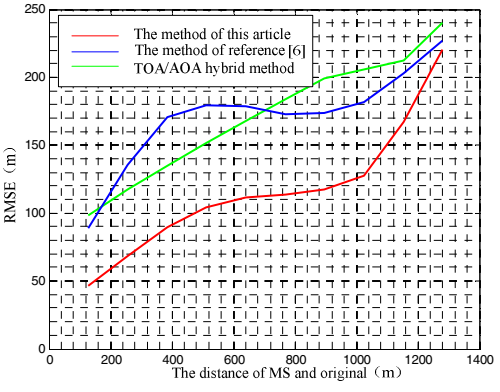


Fig. 2. The result of different method

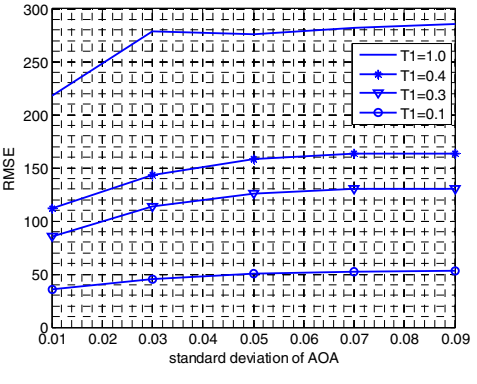


Fig. 3. The effect of different channel circumstance

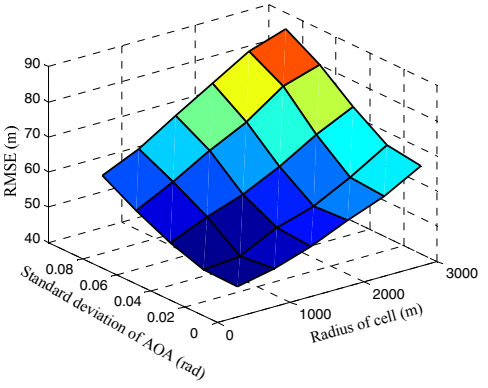


Fig. 4. The effect of the cell radius

Simulate 2: the capability of the method is tested through changing the standard deviation and the channel circumstance. The standard deviation are [0.01, 0.03, 0.05 0.07, 0.09], the channel circumstance are [1.0, 0.4, 0.3, 0.1]. The simulation result is as figure3. The figure shows that the capability of the method become various with the change of precision of AOA measurement and the channel situation. The precious is better, the RMSE of error is litter; the channel situation is better, the RMSE of error is litter. And if the AOA measurement error is within in some range, the location precise can meet the requirement.

Simulation 3: the effect of radius cell. The simulation result is as figure4. it shows that the smaller cell radium, the more precious of MS location.

4 Conclusion

NLOS error is one of the main impact on the cellular network positioning accuracy. It is a hot research for mobile communication networks that how to suppress or eliminate the impact of NLOS on the location accuracy. The digital simulation results show that the TOA/AOA hybrid location methods in this paper can weaken the effect of NLOS and improve the positioning accuracy.

References

1. Chen, C.-L., Feng, K.-T.: An efficient Geometry-Constrained Location Estimation Algorithm for NLOS Environment. In: International Conference on Wireless Network, Communication and Mobile Computing, pp. 244–249 (2005)
2. Chan, Y.T.: A simple and efficient estimate for hyperbolic location. IEEE Trans. on Signal Processing 42(8), 1095–1115 (1994)
3. Al-Jazzar, S., Ghogho, M., McLernon, D.: A Joint TOA/AOA Constrained Minimization Method for Locating Wireless Devices in Non-Line-of-Sight Environment. IEEE Transactions on Vehicular Technology 58(1), 468–472 (2009)

4. Liu, J., Li, J.: TDOA/AOA hybrid wireless location method in NLOS situation. Journal on Communications 26(5), 63–68 (2005)
5. Tian, X., Liao, G.: An Efective TOA-Based Location Method for Mitigating the Influence of the NLOS Propaga-Tion. Acta Electronica Sinica 31(9), 1429–1432 (2003)
6. Chang, R., Lü, S., Wang, P.: Improvement of the position location accuracy in NLOS environment by TOA data process. Chinese Journal of Radio Science 22(1), 12–16 (2007)
7. Wang, X., Wang, Z., Liu, S.: A TOA-based location algorithm reducing the errors due to non-line-of-sight propagation. Journal of China Institute of Communications 22(3), 1–8 (2001)
8. Aszetly, D.: On antenna arrays in mobile communication system: Fast fading and GSM base station receiver algorithms. [Ph.D. dissertation], Royal Institute Technology, Sweden (March 1996)
9. Klukas, R., Fattouche, M.: Line-of-sight angle of arrival estimation in the outdoor multipath environment. IEEE Trans. on VT 47(1), 342–351 (1998)

Effect of Car Length in the Biham-Middleton-Levine Traffic Model*

Wei Huang, Rui Jiang, Mao-Bin Hu, and Qing-Song Wu

School of Engineering Science, University of Science and Technology of China Hefei 230026, China
rjiang@ustc.edu.cn

Abstract. The Biham-Middleton-Levine (BML) model is the simplest cellular automaton model of traffic in idealized networks of streets in cities. Three phases (i.e., free flow phase, jam phase, and intermediate stable phase in which free flow coexists with jam) are observed in the model. This paper studies the influence of car length on the average velocity $\overline{v}$ and on the structure of the intermediate stable phase. It is shown that when the car length $\varpi > 1$, there appears a density range $\rho_{c2} < \rho < \rho_{c1}$ in which all the three phases could coexist. More interestingly, when the car length $\varpi > 1$, the structure of the intermediate stable state qualitatively changes. There appear more than two stripes in each direction. However, the average velocity of the phase essentially does not change.

Keywords: BML model, free flow, jam, phase coexistence.

1 Introduction

In the last few decades, traffic problems have attracted considerable attention [1,2]. Various types of traffic flow models have been proposed, e.g., macroscopic hydrodynamic models based on partial differential equations, and microscopic car-following models based on ordinary differential equations.

Recently, Cellular Automaton (CA) approach has become a well-established method to model, analyze, understand, and even forecast the behavior of real traffic [2,3]. The BihamMiddleton-Levine (BML) model is the simplest CA model of traffic in idealized networks of streets in cities [4].The model describes traffic flow on a $L \times L'$ square lattice, in which there are two species of cars, eastbound ones and northbound ones. Hence, each lattice site can be in one of three states: empty, occupied by an eastbound car, or occupied by a northbound car. The cars are initially distributed randomly among the lattice sites and they move based on the regulated traffic signals: In the odd (even) time step, traffic lights for the eastbound (northbound) cars are green, hence the eastbound (northbound) cars move forward to the next site provided the target site is empty.

* This work is partially supported by National Basic Research Program of China (No. 2006CB705500), the NSFC (No. 10872194 and 11072239), and the FANEDD.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 479–487.
springerlink.com

In the BML model, when the periodic boundary condition is adopted, the number of cars N is conserved. The traffic density is defined as $\rho = \frac{N_e+N_n}{LL'}$ Here Ne and Nn are numbers of eastbound cars and northbound cars, respectively. For simplicity, we focus on the symmetric situation $N_e = N_n$ in this paper.

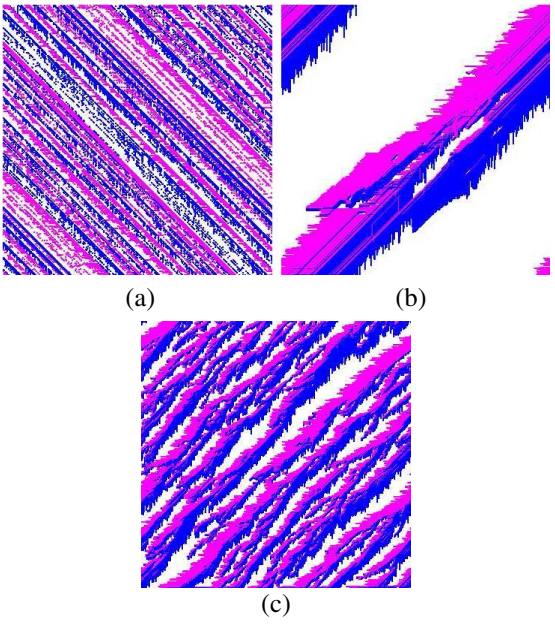


Fig. 1. (Color online) Typical configurations of the BML model on an L × L square with L = 360. (a)The free flow phase at ρ =0.2. (b) One global jam at ρ =0.36. (c) Many random local jams at ρ =0.6. See http: //home.ustc.edu.cn/~huangshi for the spatiotemporal evolution of the three configurations. The red points denote eastbound cars and blue points denote northbound cars.

Despite its simplicity, the BML model exhibits interesting phase transition behavior. For more than ten years, it is generally believed that there are only two phases in the model, separated by a critical density ρ_c. When the density is smaller than ρ_c, the free flow phase exists in which all cars can move freely (the average velocity of cars $\overline{v}$ =1), see Fig.1(a); When the density $\rho > \rho_c$, the jam phase appears in which no car can move at all ($\overline{v}$ =0). The jam phase can be further classified into two sub-phases, one global jam when $\rho - \rho_c$ is small (Fig.1(b)) and many random local jams when ρ is much larger (Fig.1(c)). The transition from free flow phase to jam phase is believed to be a sharp one (see Fig.2) and efforts are devoted to find ρ_c analytically [5-7].

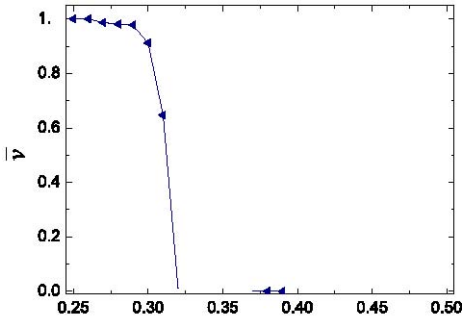
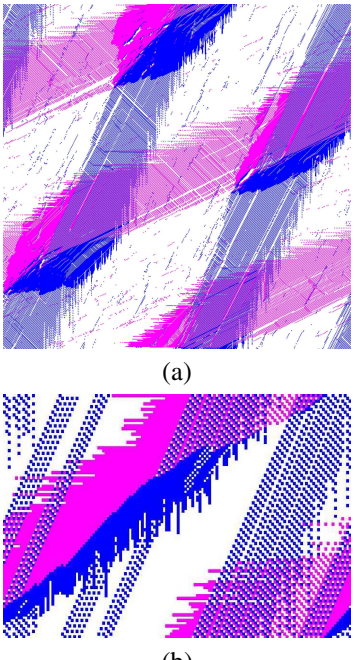


Fig. 2. The average velocity $\overline{v}$ as a function of the density ρ. System size L× L' = 512 × 512, car length ϖ =1. The plot is obtained by averaging over many realizations.



(a)

(b)

Fig. 3. (Color online) Typical configurations of the BML model on an L× L' lattice.(a) The intermediate stable phase with average velocity $\overline{v}$ ~ 2/3 obtained at ρ =0.34, L= L' = 600.(b) The intermediate stable phase with average velocity $\overline{v}$ ~2/5 obtained at ρ =0.46, L = 144, L' =89. See http: //home.ustc.edu.cn/~huangshi for the spatiotemporal evolution of the two configurations.

The BML model has attracted the interests of scientists. At the time of this writing, Ref.[4] has received over 380 citations in the scientific literature. However, until 2005, D'Souza [7] discovered that there exists an intermediate stable phase, where jams and freely flowing traffic coexist in a way that bands of free flowing traffic intersecting at jammed wave fronts. The structure propagates smoothly through the space. Hence, instead of a phase transition as a function of density, bifurcation is observed.

Fig.3 shows two typical configurations of the intermediate stable phase. In Fig.3(a), the system is a square with L= L' = 600 and the average velocity $\overline{v}$ ~ 2/3. In Fig.3(b), the system size L = 144, L' = 89 and the average velocity $\overline{v}$ ~ 2/5.

In previous studies, the car length ϖ is always set to be one, i.e., each car occupies only one site. This paper considers the case that car length is larger than one. We study the effect of car length in the BML model. In next section, the simulation results are presented and discussed.

2 Simulation Results

For simplicity, we consider only square lattice with L = L' in this paper. The rectangular lattice with L$\neq L'$ will be studied in future work. We implement the simulations for a range of system size L, car length ϖ and density ρ. Here the density ρ is defined as $\rho = \frac{(N_e+N_n)\varpi}{L^2}$.

We perform many realizations for each set of initial conditions. All realizations were simulated until converged or for times out to at least t =3×10^6 time steps.

Figure 4 shows the plot of the average velocity $\overline{v}$ versus the initial density ρ for different car length ϖ with system size L = 1200. We denote that the free flow phase could occur in the density range $\rho < \rho_{c1}$, the jam phase could occur in the density range $\rho > \rho_{c2}$, and the intermediate stable phase could occur in the density range $\rho_{c3} < \rho < \rho_{c4}$. It can be seen that

(i) With the increase of car length, the four critical densities $\rho_{c1},...,\rho_{c4}$ decrease. This is somewhat counter-intuitive because in a 1D system, by increasing the car length, the critical density increases (The critical density in 1D system is $\rho_c = \varpi / (\varpi+1)$. The average velocity $\overline{v}=1$ when $\rho < \rho_c$, and $\overline{v} = \varpi\ (1 - \rho)/\rho$ when $\rho > \rho_c$). This difference is easy to understand because in 1D system, average velocity is mainly determined by average distance between cars. However, in 2D system, the average velocity mainly depends on the traffic flow of cars of the other direction. When car length increases, a car can block more cars in the other direction. Hence, the critical densities decrease.

(ii) The density range of intermediate stable phase $\rho_{c4} - \rho_{c3}$ decreases when car length increases from ϖ =1 to ϖ =2. However, the range essentially remains a constant when car length further increases.

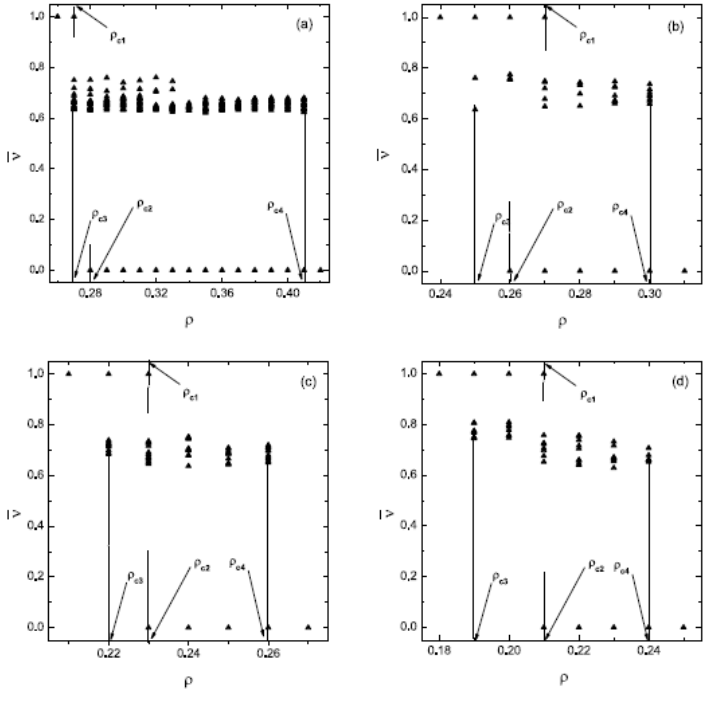


Fig. 4. Plot of the average velocity $\overline{v}$ for each individual realization versus the density ρ. The system size L = 1200. The car length ϖ = (a) 1, (b) 2, (c) 3, (d) 4.

(iii) With the increase of car length, the order of the four critical densities changes. When ϖ =1, we have

$$\rho_{c3} \approx \rho_{c1} < \rho_{c2} < \rho_{c4}$$

When ϖ =2, it becomes

$$\rho_{c3} < \rho_{c2} < \rho_{c1} < \rho_{c4}$$

Thus, in the density range $\rho c2 < \rho < \rho c1$, all the three phases coexist. Which phase occurs depends on the initial condition. When ϖ =3, $\rho c2$ becomes approximately equal to $\rho c1$, we have

$$\rho_{c3} < \rho_{c2} \approx \rho_{c1} < \rho_{c4}$$

When ϖ increases to 4, the situation are similar to that of ϖ =3.

(iv) The average velocity $\overline{v}$ of the intermediate stable phase essentially remains independent of the car length.

Next we study the effect of system size L with car length fixed at ϖ =4. The results are shown in Fig.5. It can be seen that (i) When L is small, there is no intermediate stable phase (Figs.5(a) and (b)). When L is large enough, intermediate stable phase could occur (Figs.5(c) and (d)); (ii) The density range of intermediate stable phase also essentially remains a constant as L increases; (iii) The average velocity $\overline{v}$ of the intermediate stable phase essentially remains independent of the system size. The three results could also be observed under other values of ϖ. However, with the increase of ϖ, the system size at which the intermediate stable phase begins to appear increases.

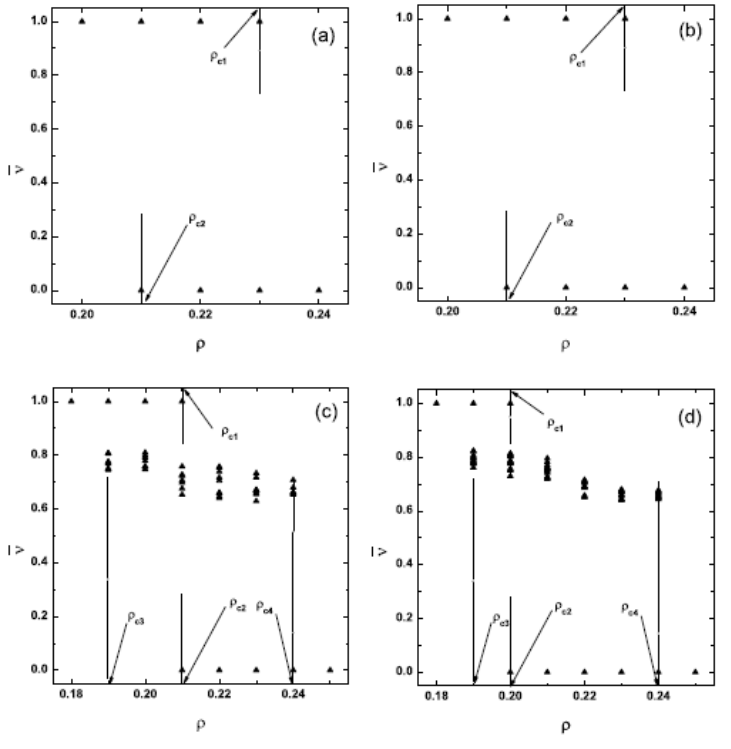


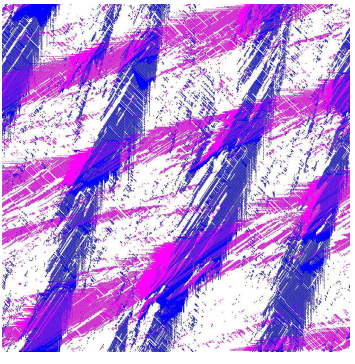
Fig. 5. Plot of the average velocity $\overline{v}$ for each individual realization versus the density ρ. The car length ϖ =4. The system size L = (a) 360, (b) 600, (c) 1200, (d) 2400.

More interestingly, we find that car length has nontrivial influence on the structure of the intermediate stable phase. Fig.6 shows two typical snapshots of the intermediate stable phase with car length $\varpi > 1$. One can see that there are more than two stripes in both directions. Moreover, in most realizations, the number of eastbound stripes equals to the number of northbound stripes (Fig.6(a)). However, they might not equal in other realizations (Fig.6(b)). We have executed many

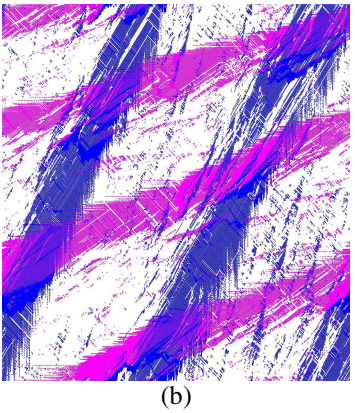
simulations with $2 \leq \varpi \leq 4$, and found that the number of stripes does not exceed $\varpi + 1$ in each direction.

Comparing the intermediate stable phase shown in Fig.6 with that shown in Fig.3(a), one can see that when the car length $\varpi > 1$, the stripes become less compact. This is because with the increase of stripes, the number of intersections of stripes increases remarkably. For instance, it increase from 3 in Fig.3(a) to 8 in Fig.6(a). Hence, the jam region is much smaller (see Fig.7). Due to the random distribution of cars, a jam region may appear and disappear in the evolution. As a result, the stripes are less regular and they become sparse.

If the numbers of stripes in the two directions equal and both equal to $\varpi + 1$, then the pattern are easy to understand as in the case of $\varpi = 1$ which could be classified into jam region, cross region, and free flow region (see Fig.7).



(a)



(b)

Fig. 6. (Color online) Two typical configurations of the intermediate stable phase with car length $\varpi > 1$. In both (a) and (b), $\varpi = 2$, $\rho = 0.27$, L = 1200, but they starts from different initial conditions. See http : //home.ustc.edu.cn/~huangshi for the spatiotemporal evolution of the two configurations.

In the jam region, cars cannot move. In the cross region and free flow region, cars move with velocity 1. Since the stripes are formed as outflow from the jam region, the northbound stripes have slope ϖ +1 and the eastbound stripes have slop 1/ (ϖ + 1). The area of jam region could be calculated. Hence, the average velocity could be obtained. More details will be reported elsewhere due to page limit.

However, when the numbers of stripes in the two direction are not equal or they are not equal to ϖ +1, then the situation is not so easy to understand, which needs further investigations.

3 Conclusion

This paper studies effect of car length in the two-dimensional BML traffic model. We have investigated the influence of car length on the average velocity $\overline{v}$ and on the structure of the intermediate stable phase. It is shown that with the increase of car length, the four critical densities ρ c1,..., ρ c2 decrease. Moreover, when the car length $\varpi > 1$, there appears a density range $\rho_{c2} < \rho < \rho_{c1}$ in which all the three phases could coexist.

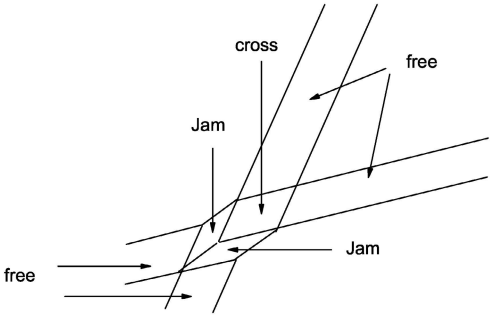


Fig. 7. Pattern sketch where two stripes intersect

More interestingly, when the car length $\varpi > 1$, the structure of the intermediate stable state qualitatively changes. There appear more than two stripes in each direction. In most realizations, the number of eastbound stripes equals to the number of northbound stripes. However, they might not equal in other realizations. We have also found that the number of stripes does not exceed ϖ +1 in each direction. We would like to mention that although the structure of the intermediate stable phase changes, the average velocity of the phase essentially does not change. In the future work, we need to investigate the details of the situation where the numbers of stripes in the two direction are not equal or they are not equal to ϖ +1.

In the future, the work could be extended in many aspects. For instance, (i) The rectangular system with L $\neq L'$ needs to be studied; (ii) The asymmetric situation with Ne ≠ Nn will be investigated; (iii) The effect of cars with mixed length should be considered.

References

1. Helbing, D.: Traffic and related self-driven many-particle systems. Rev. Mod. Phys. 73, 1067–1141 (2001)
2. Chowdhury, D., Santen, L., Schadschneider, A.: Statistical physics of vehicular traffic and some related systems. Phys. Rep. 329, 199–329 (2000)
3. Maerivoet, S., De Moor, B.: Cellular automata models of road traffic. Phys. Rep. 419, 1–64 (2005)
4. Biham, O., Middleton, A.A., Lenive, D.: Self-organization and a dynam¬ical transition in trafffic-flow models. Phys. Rev. A 46, R6124–R6127 (1992)
5. Wang, B.H., Woo, Y.F., Hui, P.M.: Improved mean-field theory of two-dimensional traffic flow models. J. Phys. A-Math. Gen. 29, L31–L35 (1996)
6. Wang, B.H., Woo, Y.F., Hui, P.M.: Mean field theory of traffic flow problems with overpasses and asymmetric distributions of cars. J. Phys. Soc. Jpn 65, 2345–2348 (1996)
7. D'Souza, R.M.: Coexisting phases and lattice dependence of a cellular automaton model for traffic flow. Phys. Rev. E 71, 066112 (2005)

Spin Current in a GaAs 2DEG with the Coexistence of Rashba Spin-Orbit Coupling and Magnetic Field

Xi Fu[1,2*]

[1] Department of Electronics, Hunan University of Science and Engineering, Yongzhou, 425100, China
[2] Key Laboratory of Low Dimensional Quantum Structures and Quantum Control (Hunan Normal University), Ministry of Education, Changsha 410081, China
Fuxi1980613@126.com

Abstract. Spin current in a two-dimensional electron gas (2DEG) with the coexistence of Rashba spin-orbit coupling and a perpendicular magnetic field has been investigated theoretically. It is found that there exist five linear and three angular spin current density elements which present more oscillations and larger strength with the increasing of Landau level number, and their transverse distributive ranges can be shifted and compressed with the increasing of magnetic field strength. Moreover, strength of seven elements show a linear relation to the Rashba constant indicating that spin currents of the 2DEG are controllable.

Keywords: 2DEG, spin current, Rashba spin-orbit coupling, magnetic field.

1 Introduction

Spin dependent transport phenomena in low dimensional systems have attracted considerable attentions in recent years because of potential application in information processing and storage devices [1,2]. A paradigmatic proposal is the spin field-effect-transistor which utilizes the gate-controllable [3] Rashba spin-orbit coupling (SOC) [4] in two-dimensional electron gases (2DEGs) to control the spin rotation of electrons as they propagate across the device [5-7]. The Rashba SOC results from the structural inversion asymmetry of microscopic confinement potential formed at the interface of semiconductor hetero-structures [4]. There have much interests recently in 2DEG systems with the SOC and external magnetic field [8,9]. By applying a perpendicular magnetic field to a 2DEG system, the SOC competes with the Zeeman splitting and this interplay leads to further modification of band structure and other interesting results.

Spin current in ballistic systems have been investigated in several years [10-18]. However, there still exists a lot of debates over the correct definition of spin current

* Corresponding author.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 489–497.
springerlink.com

[10,12,15,16]. Rashba first suggested that conventional definition $\mathbf{I}_s = \mathrm{Re}\{\Psi^+ \hat{v}\hat{s}\Psi\}$ should be modified to eliminate the nonzero spin current [10]. After Rashba's work, Sun *et al* bring forward a definition by using the conventional (linear) spin current and a angular spin current describing spin motion and rotation respectively, and further they discussed the reasonableness of this definition in details [13-17]. On the other hand, when a magnetic field and Rashba SOC are appended Wang *et al* have studied magnetotransport and intrinsic spin-Hall conductivity of a 2DEG respectively [9]. Nevertheless there do not have researches on spin current in the presence of SOC and magnetic field, therefore, in this paper using the definition by Sun *et al* spin currents in a 2DEG with the coexistence of Rashba SOC and a perpendicular magnetic field have been investigated, and nonzero spin current density elements are calculated and demonstrated analytically.

2 Mode and Formalism

We consider a two-dimensional electron gas (2DEG) in the *x*-*y* plane subjected to a perpendicular magnetic field, and in the Landau gauge $\vec{A}$=(B_zy,0,0) the effective mass single-electron Hamiltonian including Rashba SOC reads

$$H = \frac{(\vec{p}+e\vec{A})^2}{2m^*} + \frac{\alpha_R}{\hbar}[\vec{\sigma}\times(\vec{p}+e\vec{A})]_z + \tfrac{1}{2}g\mu_B B_z\sigma_z . \tag{1}$$

In the above Hamiltonian, m^*, α_R, g and μ_B denote effective electron mass, Rashba constant, Zeeman factor and Bohr magneton respectively. The electron wavefunctions for the *n*th Landau state in the 2DEG can be written as a spinor [9]

$$\Psi_n^\sigma(x,y) = \begin{pmatrix}\psi_{n\uparrow}^\sigma(x,y)\\ \psi_{n\downarrow}^\sigma(x,y)\end{pmatrix} = \begin{pmatrix}e^{ik_x x}\varphi_{n\uparrow}^\sigma(y)\\ e^{ik_x x}\varphi_{n\downarrow}^\sigma(y)\end{pmatrix} = \frac{e^{ik_x x}}{\sqrt{L_x}}\begin{pmatrix}\phi_{n-1}(y-y_c)\sin\theta_n^\sigma\\ \phi_n(y-y_c)\cos\theta_n^\sigma\end{pmatrix} (n=1,2,3,\ldots) \tag{2}$$

with $\sigma = \pm$ denoting two branches of Landau levels, while $\tan\theta_n^+ = \sqrt{2n}\alpha_R / l_c\sqrt{\varepsilon_0 + \sqrt{\varepsilon_0^2 + 2n\alpha_R^2/l_c^2}}$ with $\varepsilon_0 = \hbar\omega_c/2 - g\mu_B B_z/2$ and $\theta_n^- = \pi/2 + \theta_n^+$. Furthermore the harmonic oscillator function is

$$\phi_n(y-y_c) = \sqrt{\sqrt{\pi}2^n n!}e^{-\frac{(y-y_c)^2}{2l_c^2}} H_n\left(\frac{y-y_c}{l_c}\right), \tag{3}$$

where $l_c = \sqrt{\hbar/m^*\omega_c}$ is the radius of cyclotron orbit centered at $-y_c = -l_c^2 k_x$ with $\omega_c = eB_z/m^*$ the cyclotron frequency, and $n = 1,2,3,\ldots$ is Landau-level number respectively. Moreover, the associated electron energies are $E^\sigma = \hbar^2 k_x^2/2m^* + \varepsilon^\sigma$ with the sublevels $\varepsilon_n^+ = n\hbar\omega_c + \sqrt{\varepsilon_0^2 + 2n\alpha_R^2/l_c^2}$

and $\varepsilon_n^- = n\hbar\omega_c - \sqrt{\varepsilon_0^2 + 2n\alpha_R^2/l_c^2}$ [9], therefore, one can find there exists four longitudinal wavevectors (that is, $k_x^{\sigma,\pm}(E^\sigma)$ with σ denoting two branches of Landau levels and $\pm$ positive and negative sign respectively)

$$k_x^{+,\pm}(E^+) = \pm\sqrt{2m^*(E^+ - n\hbar\omega_c - \sqrt{\varepsilon_0^2 + 2n\alpha_R^2/l_c^2})/\hbar^2}\,, \tag{4}$$

$$k_x^{-,\pm}(E^-) = \pm\sqrt{2m^*(E^- - n\hbar\omega_c + \sqrt{\varepsilon_0^2 + 2n\alpha_R^2/l_c^2})/\hbar^2}\,. \tag{5}$$

In order to calculate spin current density, we use the definition brought forward by Sun *et al* [15,16]

$$\mathbf{j}_s(r,t) = \mathrm{Re}\{\Psi^+(r,t)\hat{v}\hat{s}\Psi(r,t)\}\,, \tag{6}$$

$$\vec{j}_\omega(r,t) = \mathrm{Re}\{\Psi^+(r,t)\hat{\omega}\times\hat{s}\Psi(r,t)\}\,, \tag{7}$$

where $\hat{v} = \dfrac{\vec{p}+e\vec{A}}{m^*} + \dfrac{\alpha_R}{\hbar}(\hat{z}\times\vec{\sigma})$ and $\hat{\omega} = \dfrac{2}{\hbar}\{\vec{B} + \dfrac{\alpha_R}{\hbar}[(\vec{p}+e\vec{A})\times\hat{z}]\}$ are linear and angular velocity operators. Note that linear spin current density $\mathbf{j}_s$ is a tensor while angular spin current density $\vec{j}_\omega$ is a vector. We use these symbols to represent spin current density elements, e.g. $j_{s,xy}$ represents linear spin current density element of an electron moving along the longitudinal *x*-direction with its spin in the transverse *y*-direction and $j_{\omega,x}$ represents electron spin precession in the *x*-direction.

After substituting Eq.(2) into Eqs.(6) and (7), the linear and angular spin current density can be obtained straightforwardly. There have five nonzero linear spin current density elements

$$j_{s,xx}^{n\sigma} = \mathrm{Re}\{\frac{\hbar^2 k_x^\sigma}{m^*}\varphi_{n\uparrow}^{\sigma*}\varphi_{n\downarrow}^{\sigma} + \frac{eB_z y\hbar}{2m^*}\varphi_{n\uparrow}^{\sigma*}\varphi_{n\downarrow}^{\sigma} + \frac{\hbar^2 k_x^\sigma}{m^*}\varphi_{n\downarrow}^{\sigma*}\varphi_{n\uparrow}^{\sigma} + \frac{eB_z y\hbar}{2m^*}\varphi_{n\downarrow}^{\sigma*}\varphi_{n\uparrow}^{\sigma}\}/L_x\,, \tag{8}$$

$$j_{s,xy}^{n\sigma} = \mathrm{Re}\{-\frac{\alpha_R}{2}\varphi_{n\uparrow}^{\sigma*}\varphi_{n\uparrow}^{\sigma} - \frac{\alpha_R}{2}\varphi_{n\downarrow}^{\sigma*}\varphi_{n\downarrow}^{\sigma}\}/L_x\,, \tag{9}$$

$$j_{s,xz}^{n\sigma} = \mathrm{Re}\{\frac{\hbar^2 k_x^\sigma}{m^*}\varphi_{n\uparrow}^{\sigma*}\varphi_{n\uparrow}^{\sigma} + \frac{eB_z y\hbar}{2m^*}\varphi_{n\uparrow}^{\sigma*}\varphi_{n\uparrow}^{\sigma} - \frac{\hbar^2 k_x^\sigma}{m^*}\varphi_{n\downarrow}^{\sigma*}\varphi_{n\downarrow}^{\sigma} - \frac{eB_z y\hbar}{2m^*}\varphi_{n\downarrow}^{\sigma*}\varphi_{n\downarrow}^{\sigma}\}/L_x\,, \tag{10}$$

$$j_{s,yx}^{n\sigma} = \mathrm{Re}\{\frac{\alpha_R}{2}\varphi_{n\uparrow}^{\sigma*}\varphi_{n\uparrow}^{\sigma} + \frac{\alpha_R}{2}\varphi_{n\downarrow}^{\sigma*}\varphi_{n\downarrow}^{\sigma}\}/L_x\,, \tag{11}$$

$$j_{s,yy}^{n\sigma} = \mathrm{Re}\{-\frac{\hbar^2}{2m^*}\varphi_{n\uparrow}^{\sigma*}\varphi_{n\downarrow}^{'\sigma} + \frac{\hbar^2}{2m^*}\varphi_{n\downarrow}^{\sigma*}\varphi_{n\uparrow}^{'\sigma}\}/L_x \tag{12}$$

and three nonzero angular spin current density elements

$$j_{\omega,x}^{n\sigma} = \mathrm{Re}\{-\alpha_R k_x^{\sigma} \varphi_{n\uparrow}^{\sigma^*} \varphi_{n\uparrow}^{\sigma} - \frac{eB_z y \alpha_R}{\hbar} \varphi_{n\uparrow}^{\sigma^*} \varphi_{n\uparrow}^{\sigma} + \alpha_R k_x^{\sigma} \varphi_{n\downarrow}^{\sigma^*} \varphi_{n\downarrow}^{\sigma} + \frac{eB_z y \alpha_R}{\hbar} \varphi_{n\downarrow}^{\sigma^*} \varphi_{n\downarrow}^{\sigma}\} / L_x, \quad (13)$$

$$j_{\omega,y}^{n\sigma} = \mathrm{Re}\{B_z \varphi_{n\uparrow}^{\sigma^*} \varphi_{n\downarrow}^{\sigma} + B_z \varphi_{n\downarrow}^{\sigma^*} \varphi_{n\uparrow}^{\sigma}\} / L_x, \quad (14)$$

$$j_{\omega,z}^{n\sigma} = \mathrm{Re}\{\alpha_R(-\varphi_{n\uparrow}^{\sigma^*} \varphi_{n\downarrow}^{'\sigma} + \varphi_{n\downarrow}^{\sigma^*} \varphi_{n\uparrow}^{'\sigma}) + \frac{eB_z y \alpha_R}{\hbar}(\varphi_{n\uparrow}^{\sigma^*} \varphi_{n\downarrow}^{\sigma} + \varphi_{n\downarrow}^{\sigma^*} \varphi_{n\uparrow}^{\sigma}) + \alpha_R k_x^{\sigma}(\varphi_{n\uparrow}^{\sigma^*} \varphi_{n\downarrow}^{\sigma} + \varphi_{n\downarrow}^{\sigma^*} \varphi_{n\uparrow}^{\sigma})\} / L_x \quad .(15)$$

In the above equations, each spin current density element has two branches j^{n+} and j^{n-} corresponding to two branches of Landau levels.

From Eqs.(4) and (5) one can find in the equilibrium the four occupied states $k_x^{\sigma,\pm}(E^{\sigma})$ ($k_x^{+,+}$, $k_x^{+,-}$, $k_x^{-,+}$, $k_x^{-,-}$) are provided with the relation $k_x^{+,+} + k_x^{+,-} + k_x^{-,+} + k_x^{-,-} = 0$ due to same contributions [13,16]. Then spin current density elements which including k_x can be obtained as

$$j_{s,xx}^{n\sigma} = \mathrm{Re}\{\frac{eB_z y\hbar}{2m^*} \varphi_{n\uparrow}^{\sigma^*} \varphi_{n\downarrow}^{\sigma} + \frac{eB_z y\hbar}{2m^*} \varphi_{n\downarrow}^{\sigma^*} \varphi_{n\uparrow}^{\sigma}\} / L_x, \quad (16)$$

$$j_{s,xz}^{n\sigma} = \mathrm{Re}\{\frac{eB_z y\hbar}{2m^*} \varphi_{n\uparrow}^{\sigma^*} \varphi_{n\uparrow}^{\sigma} - \frac{eB_z y\hbar}{2m^*} \varphi_{n\downarrow}^{\sigma^*} \varphi_{n\downarrow}^{\sigma}\} / L_x, \quad (17)$$

$$j_{\omega,z}^{n\sigma} = \mathrm{Re}\{-\alpha_R \varphi_{n\uparrow}^{\sigma^*} \varphi_{n\downarrow}^{'\sigma} + \frac{eB_z y \alpha_R}{\hbar} \varphi_{n\uparrow}^{\sigma^*} \varphi_{n\downarrow}^{\sigma} + \alpha_R \varphi_{n\downarrow}^{\sigma^*} \varphi_{n\uparrow}^{'\sigma} + \frac{eB_z y \alpha_R}{\hbar} \varphi_{n\downarrow}^{\sigma^*} \varphi_{n\uparrow}^{\sigma}\} / L_x, \quad (18)$$

while for the other elements, their expressions have not been changed because they are independent on k_x. From the Eqs.(8)-(18) one can find that the presence of magnetic field B_z directly induces the formation of $j_{s,xx}^{n\sigma}$, $j_{s,xz}^{n\sigma}$ and $j_{\omega,y}^{n\sigma}$, and $j_{s,xy}^{n\sigma}$ and $j_{s,yx}^{n\sigma}$ present a inverse symmetry relation. Moreover, it has been pointed out that each spin current density element possesses two branches, therefore spin current density element for the *n*th Landau level $j^{n,sum} = j^{n+} + j^{n-}$. Furthermore, total spin current density elements $j_{s,ij}^{T}$ and $j_{\omega,i}^{T}$ ($i, j = x, y, z$) can be calculated by summing $j_{s,ij}^{n,sum}$ and $j_{\omega,i}^{n,sum}$ from the 1th level to the *n*th level accordingly [16].

3 Results and Discussion

In the following, numerical examples of linear and angular spin current density elements will be presented. The electron effective mass is taken as that for GaAs quantum well $m^* = 0.038m_e$ and the Rashba constant $\alpha_R = 0.5 \times 10^{-11}$ eVm, moreover, the Bohr magneton $\mu_B = e\hbar / 2m_e c$ and the factor g =-4.38 [19]. The initial length and transverse wide of 2DEG are taken as L_x =200 nm and d =300 nm

while $E^+ = E^- = 70$ meV. It should be noted that in order to study transverse distribution of spin current the transverse wide which is taken as 300 nm is changeable.

Figures 1-4 show the plots of five linear and three angular spin current density elements $j^{n\sigma}$ and $j^{n,sum}$ as a function of dimensionless coordinate $u = x/d$ for the 1th to 4th level respectively, moreover, the unit of $j^{n\sigma}_{\omega,y}$ is taken as $\alpha_0/\hbar$ while the unit of other elements are chosen as α_0 ($\alpha_0 = 1.0\times10^{-12}$ eVm). Among eight nonzero elements, $j^{n\sigma}_{s,xy}$ and $j^{n\sigma}_{s,yx}$ are special because they are spin current density flowing along one direction with its spin pointing to another direction, and the

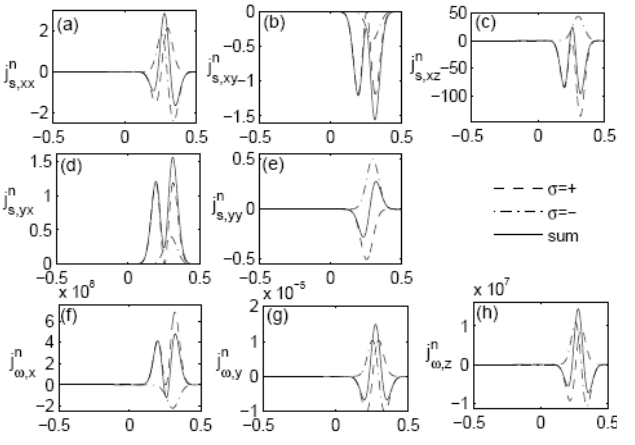


Fig. 1. The plots of spin current density elements $j^{n\sigma}$ and $j^{n,sum}$ as a function of *u* for the *n*=1 case

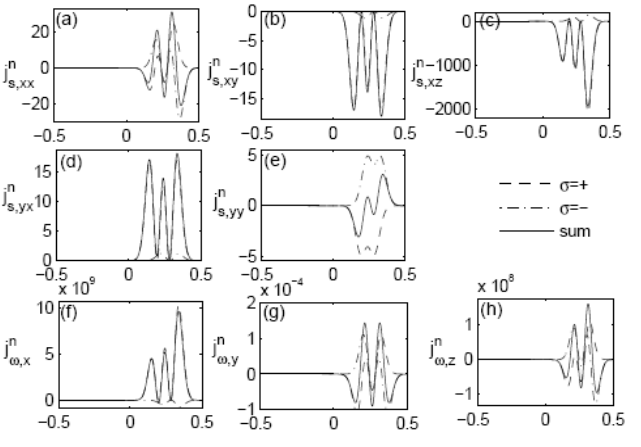


Fig. 2. The plots of spin current density elements $j^{n\sigma}$ and $j^{n,sum}$ as a function of *u* for the *n*=2 case

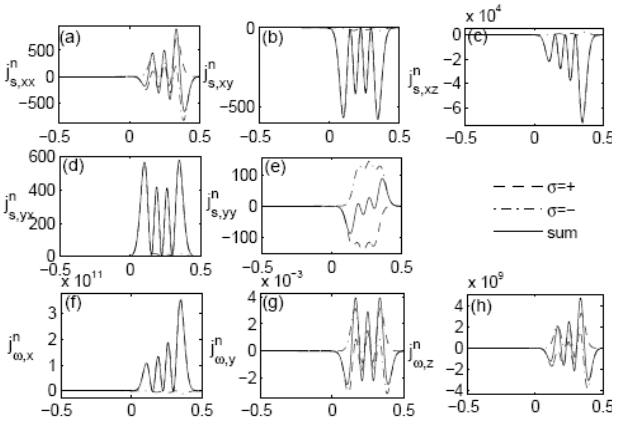


Fig. 3. The plots of nonzero spin current density elements $j^{n\sigma}$ and $j^{n,sum}$ as a function of u for the n=3 case

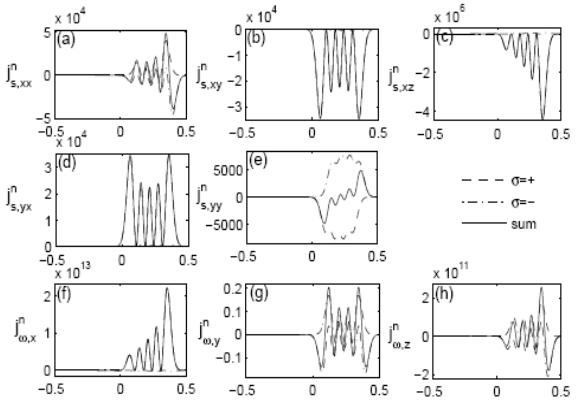


Fig. 4. The plots of nonzero spin current density elements $j^{n\sigma}$ and $j^{n,sum}$ as a function of u for the n=4 case

existence of two elements is due to spin precession in accompany with the electron motion [13]. From the figures, due to the existence of B_z the center of all elements shift towards the left side of 2DEG along the transverse direction which is relate to the harmonic oscillator function in Eq.(3), and strengths of angular elements are larger than that of all linear elements. Furthermore, when the level number increases from n=1 to n=4 strengths of eight elements present larger augments and more oscillations along the transverse direction accordingly, and this result is consistent with that for no magnetic field case [17]. Moreover, the altitude of $j^{n-}_{s,xy}$, $j^{n-}_{s,xz}$, $j^{n-}_{s,yx}$, $j^{n-}_{\omega,x}$ is lower than that of $j^{n+}_{s,xy}$, $j^{n+}_{s,xz}$, $j^{n+}_{s,yx}$, $j^{n+}_{\omega,x}$, and with the level

number n increasing $j_{s,xy}^{n-}$, $j_{s,xz}^{n-}$, $j_{s,yx}^{n-}$, $j_{\omega,x}^{n-}$ become much more smaller indicating that $j_{s,xy}^{n+}$, $j_{s,xz}^{n+}$, $j_{s,yx}^{n+}$, $j_{\omega,x}^{n+}$ is dominant to the summation $j_{s,xy}^{n,sum}$, $j_{s,xz}^{n,sum}$, $j_{s,yx}^{n,sum}$, $j_{\omega,x}^{n,sum}$ respectively. From the figures one can also find that $j_{s,xy}^{n\sigma}$ and $j_{s,yx}^{n\sigma}$ have inverse symmetry relation.

Figure 5 shows the influences of transverse wide to the transverse distribution of spin current, and when the wide increases from 300 nm to 600 nm distribution of eight elements have not been changed anymore indicating that spin currents are independent on the transverse range of 2DEG, however, when no magnetic field exists our calculations show that for a quantum wire its spin current density elements will distribute at all transverse range [17]. In Fig.6 and Fig.7 the influences of perpendicular magnetic field B_z and Rashba constant α_R on total spin current density elements j^T for four levels case ($j^T = j^{1,sum} + j^{2,sum} + j^{3,sum} + j^{4,sum}$) are demonstrated. In Fig.6 when the strength of magnetic field B_z increases from 2 T to 4 T, the center of all elements also shift along the transverse direction of 2DEG and further their transverse ranges are compressed accordingly, which indicates that the transverse distribution of spin current in 2DEG can be tuned by the magnetic field. In Fig.7 when the Rashba constant α_R increases from 0.5×10^{-11} eVm to 1.0×10^{-11} eVm, the altitudes of seven elements increase for twice accordingly and a linear increasing relation with α_R exists which still presents for no magnetic field case [17], while the element $j_{s,xz}^T$ does not change any more due to its independence on the α_R.

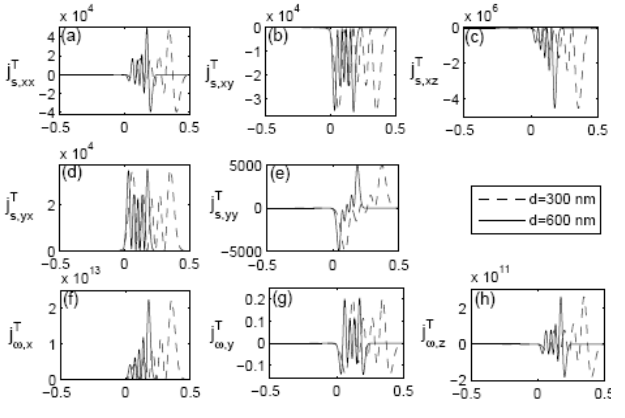


Fig. 5. The plots of total spin current density elements j^T as a function of *u* for the *n*=4 case with different transverse ranges

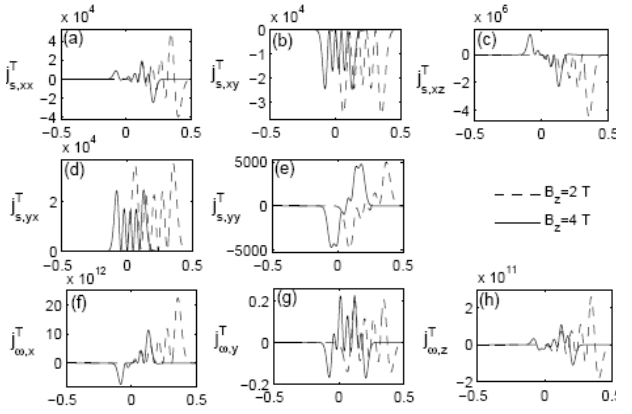


Fig. 6. The plots of total spin current density elements j^T as a function of u for the n=4 case with different magnetic field strengths

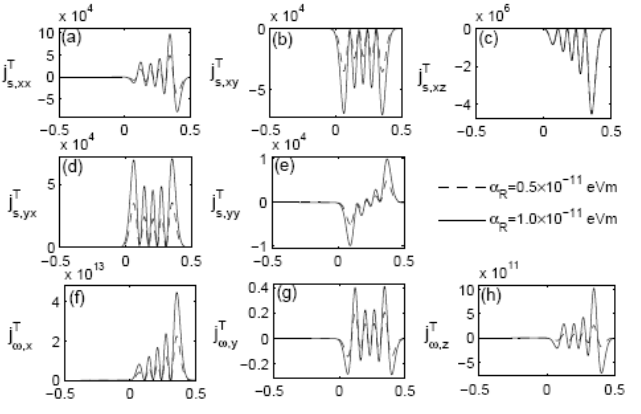


Fig. 7. The plots of total spin current density elements j^T as a function of u for the n=4 case with different Rashba constants

4 Conclusion

In this paper, we have studied spin current in a 2DEG with the coexistence of Rashba SOC and a perpendicular magnetic field. It is found that there exist eight linear and angular spin current density elements which take on more oscillation peaks and larger strength with the increasing of Landau level number. Furthermore, the center and transverse distributive area of eight elements present a transverse shift which is depend on the magnetic field, and the distributive range is compressed with the increasing of magnetic field strength. Our consequences also show that the strength of seven elements have a linear increasing relation to the Rashba constant.

Acknowledgment. Project supported by the Open Project of Key Laboratory of Low Dimensional Quantum Structures and Quantum Control (Hunan Normal University), Ministry of Education of China under Grant No QSQC1002.

References

1. Wolf, S.A., Awschalom, D.D., Buhrman, R.A., Daughton, J.M., Molnár, S.V., Roukes, M.L., Chtchelkanova, A.Y., Treger, D.M.: Science 294, 1488 (2001)
2. Zutic, I., Fabian, J., Sarma, S.D.: Rev. Mod. Phys. 76, 323 (2004)
3. Nitta, J., Akazaki, T., Takayanagi, H., Enoki, T.: Phys. Rev. Lett. 78, 1335 (1997)
4. Rashba, E.I.: Sov. Sov. Phys. Solid State 2, 1109 (1960)
5. Datta, S., Das, B.: Appl. Appl. Phys. Lett. 56, 665 (1989)
6. Mireles, F., Kirczenow, G.: Phys. Rev. B 64, 024426 (2001)
7. Fujita, T., Jalil, M.B.A., Tan, S.G.: J. Phys.: Condens. Matter 20, 115206 (2008)
8. Lucignano, P., Raimondi, R., Tagliacozzo, A.: Phys. Rev. B 78, 035336 (2008); Wang, Z.G., Zhang, W., Zhang, P.: Phys. Rev. B 79, 235327 (2009); Erlingsson, S. I., Egues, J.C., Loss, D.: Phys. Rev. B 82, 155456 (2010)
9. Wang, X.F., Vasilopoulos, P., Peeters, F.M.: Phys. Rev. B 71, 125301 (2005); Wang, X.F., Vasilopoulos, P.: Phys. Rev. B 72, 085344 (2005); Wang, X.F., Vasilopoulos, P.: Phys. Rev. B 75, 075331 (2007)
10. Rashba, E.I.: Phys. Rev. B 68, 241315 (R) (2003)
11. Ambegaokar, V., Eckern, U.: Phys. Rev. Lett. 65, 381 (1990)
12. Lee, M., Choi, M.S.: Phys. Rev. B 71, 153306 (2005)
13. Sun, Q.F., Xie, X.C.: Int. J. Mod. Phys. B 21, 3687 (2007)
14. Sun, Q.F., Guo, H., Wang, J.: Phys. Rev. B 69, 054409 (2004); Sun, Q.F., Xie, X.C., Wang, J.: Phys. Rev. Lett. 98, 196801 (2007)
15. Sun, Q.F., Xie, X.C.: Phys. Rev. B 72, 245305 (2005)
16. Sun, Q.F., Xie, X.C., Wang, J.: Phys. Rev. B 77, 035327 (2008)
17. Wang, Y., Zhou, G.H.: Chin. Phys. Lett. 23, 3065 (2006)
18. Nita, M., Marinescu, D.C., Manolescu, A., Gudmundsson, V.: arXiv, 1012.4952 (2010)
19. Rodriguez, M.V., Nazmitdinov, R.G.: Phys. Rev. B 73, 235306 (2006)

Application of Switched Reluctance Motor to a Mechanical Press

Wanfeng Shang

School of Mechanical Engineering, Xi'an University of Science & Technology, Xi'an, Shaanxi of China, 710054
shangwanfeng@gmail.com

Abstract. The paper presents a mechanical press driven by Switched Reluctance Motor (SRM) and researches on its tracking performance and operating noise. A position servo control system is proposed with a hybrid closed-loop structure to achieve flexible process of servo press. In addition, the hardware configuration and software interface of working bench is illustrated in detail. Based on the working bench, experiments are made to measure the tracking performance and operating noise of the press. Experimental results indicate the press driven by SRM has better tracking performance and lower working noise.

Keywords: Mechanical press, switched reluctance motor, flexible process.

1 Introduction

Switched Reluctance Motor (SRM) has emerged and gained increasing popularity for general purpose industrial drives [1,2] with the development of power electronics technology and high-speed processors. SRM presents additional advantages such as a simple compact construction, good capability of fault tolerance, lower cost, high torque-to-inertia ratio, and high torque output at low speed [1-4]. The most attractive feature of SRM drive is the series connection of the converter phase-leg switches with the motor phase winding, which eliminates the possibility of any shoot-through fault caused by the converter switches [5]. In addition, the drive systems of AC servo motor and bushless DC motor are also widely employed in servo system, but for a greater torque or power application over a wide speed range they incur a rather higher cost, which limits their applications on a certain degree. Hence, SRM becomes competitive with DC, AC and brushless DC motors. It can be found SRM applications of variable-speed or servo-type drive in different industries, such as robots [6], electric vehicles [7], aerospace [8], and hydraulic pump[9]. In the paper, variable speed control of SRM for the application of a mechanical press is presented and the performance characteristics of flexible process with the press are discussed.

The paper is organized as follows. Section 2 describes a control scheme for a mechanical press driven by SRM. In section 3, the configuration of working bench for the SRM-based press is introduced. Furthermore, Experiments are made to measure

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 499–507.
springerlink.com

tracking performance and operating noise of flexible process with the press in section 4. Finally, section 5 concludes this work.

2 Control Scheme of System

2.1 Control Scheme of SRM

Assuming that mutual inductances between stator phases are negligibly small, we can write the generalized m-phase SRM model [1], as follows,

$$\frac{d\psi_j}{dt} = u_j - i_j \mathrm{r}_j \qquad j = 1,...,m \tag{1}$$

$$\frac{d\theta}{dt} = \omega \tag{2}$$

$$\mathrm{J}\frac{d\omega}{dt} = \sum_{j=1}^{m} T_j - T_L - \mathrm{D}\omega \tag{3}$$

where u_j is the voltage applied to the stator terminals of the jth phase and i_j is the current flowing in the jth phase, r_j, ψ_j and T_j are the resistance, flux-linkage, and electromagnetic torque of the jth phase, respectively. The angular position of the rotor and its velocity are denoted by θ and ω. T_L is the load torque, and D and J are the viscous friction coefficient and the inertia moment of the load, respectively.

The torque produced by a single phase can be derived from the D'Alembcrt principle [18] as

$$T_j(i_j,\theta) = \frac{\partial \int_0^{i_j} \psi_j(\theta, i_j) di_j}{\partial \theta} = -\frac{1}{2}{\psi_j}^2 \frac{dR_j(\theta)}{d\theta} \tag{4}$$

where $R_j(\theta)$ is the air-gap reluctance for the phase j and its derivative with respect to θ is

$$\frac{dR_j(\theta)}{d\theta} = -\frac{f_j^{'}(\theta)}{\psi_{sat} f_j^2(\theta)} \tag{5}$$

We get a linear relationship between torque and flux-linkage using Eq.(5).

$$\psi_j(T_j,\theta) = \sqrt{\frac{-2T_j}{\frac{dR_j(\theta)}{d\theta}}} \tag{6}$$

From the equation (4), a variable flux linkage must be generated to compensate nonlinearities of the plant and hence keep a constant torque independent of the angular position. Hence, a suitable inner controller is required for high performance torque

control in SRM drive system because nonlinearity electromagnetic characteristics of SRM incur the disadvantage of torque ripple.

As shown in Fig.1, a cascade speed control scheme is introduced based on a speed controller and an inner torque controller. The torque controller is composed of a flux linkage controller [10] and a torque feedforward compensator by TSF [4] and the translation function of Eq.(6).

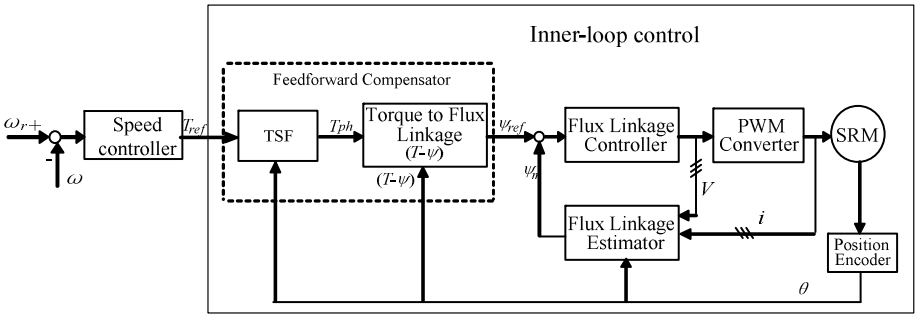


Fig. 1. Speed control scheme with inner torque controller for SRM

2.2 Control Scheme of Press

The flexible process of servo press not only relies on the servo control of SRM, but also the control of mechanical transmission system. When half closed-loop structure is adopted, a position servo system is easily set, but it is difficult to compensate position errors caused by mechanical transmission, so leading the position control accuracy can not meet requirements. On the contrary, when the closed-loop structure is adopted, the position servo system is hardly tuned. It makes the control system increasingly complexity because the oscillation of the stable point caused by various nonlinear factors is not easily eliminated. Therefore, in this paper, a hybrid closed-loop structure is proposed for a position servo system, as shown by Fig.2.

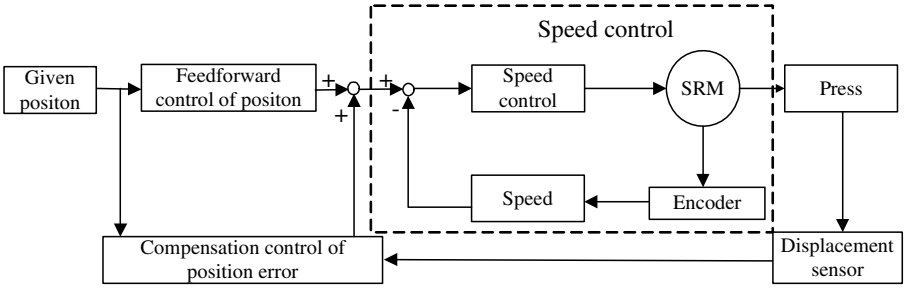


Fig. 2. Position servo system with hybrid closed-loop structure

In Fig.2, a half closed loop and a closed loop simultaneously exist in the drive System. For SRM, the speed control is a closed-loop system, while the slider position control is a half close-loop system. In a half-closed loop, the action of executed mechanism is relatively independent of the action of the electrical control. So in the half closed-loop, the control of slider position can be translated into the control of SRM speed via a position feedforward controller without close coupling oscillation. This operation enables the drive system to more rapid response. Furthermore, the closed-loop only can be used for the compensation of steady-state error, where the lower position gain can be chosen in order to ensure the system stability. Hence, higher tracking precise of position and speed is gained in the hybrid closed-loop structure.

3 Configuration of Mechanical Press Driven by SRM

SRM is installed on JH23-63 crank press involving a crank-slider mechanism, a pulley decelerator and a gear decelerator. As displayed in Fig.3, the working bench is mainly divided into four modular: state monitoring, detecting of slider position, variable speed control of SRM, and measuring forming force. The modular of state monitoring serves for switch signal of start and stop, and state acquisition of SRM and press. The slider position is detected and analyzed in real time for feedback control. Comparing a feedback actual speed and a given reference speed, the speed controller in SRM drive system is allowed for tracking a given speed curve such that the press can satisfy different requirements from various process. The forming force is measured using a strain apparatus for estimation load Torque.

The hardware configure of the control system of working bench is presented in Fig.4. The modular of SRM servo control and the modular of detecting slider position are executed in parallel thread, simultaneously in charge of the slider speed. The slider

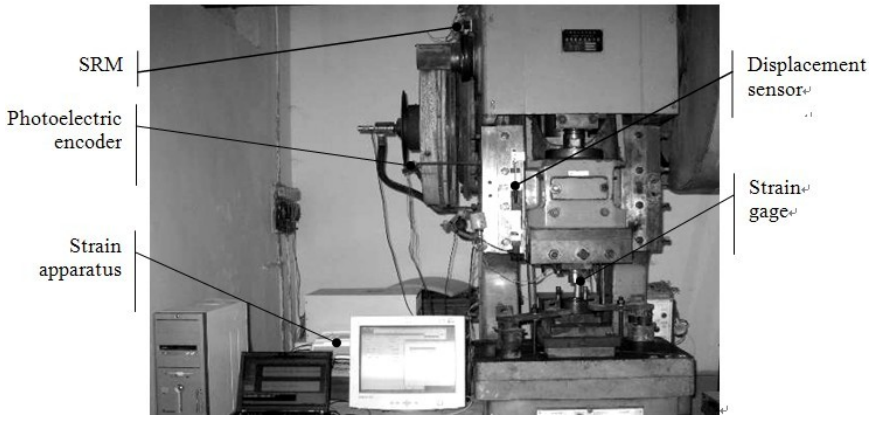


Fig. 3. Picture of working bench

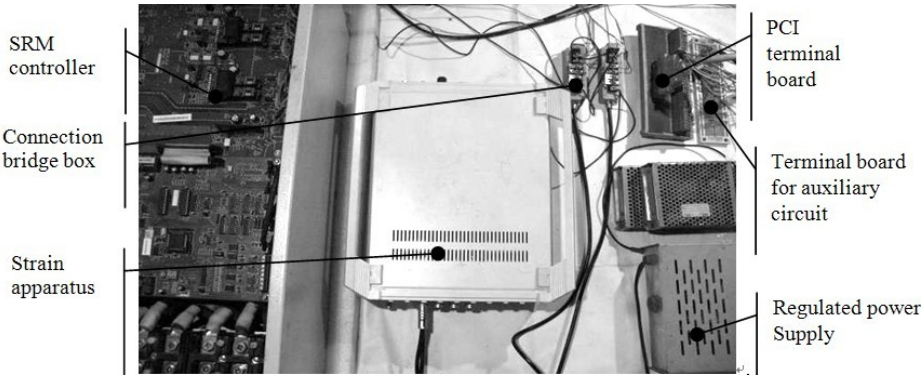
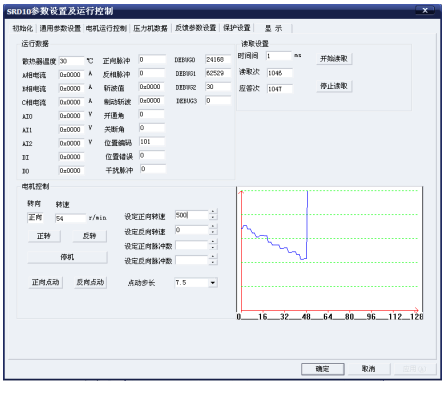
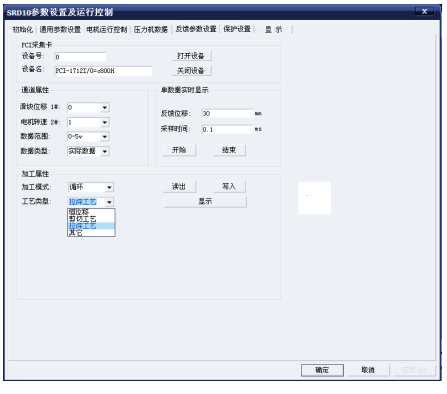


Fig. 4. Configuration of control system of working bench



(a) Operation interface of SRM



(b) Operation interface of machanical press

Fig. 5. Operation interface of the control system

position detected as feedback signal compensates for the SRM speed in order to further achieve the higher tracking precise. Hence, as Fig.5 shows, the software interfaces of the control system are developed using Microsoft Vision C++ for its advantage of multithread programming and extensive Microsoft foundation classes.

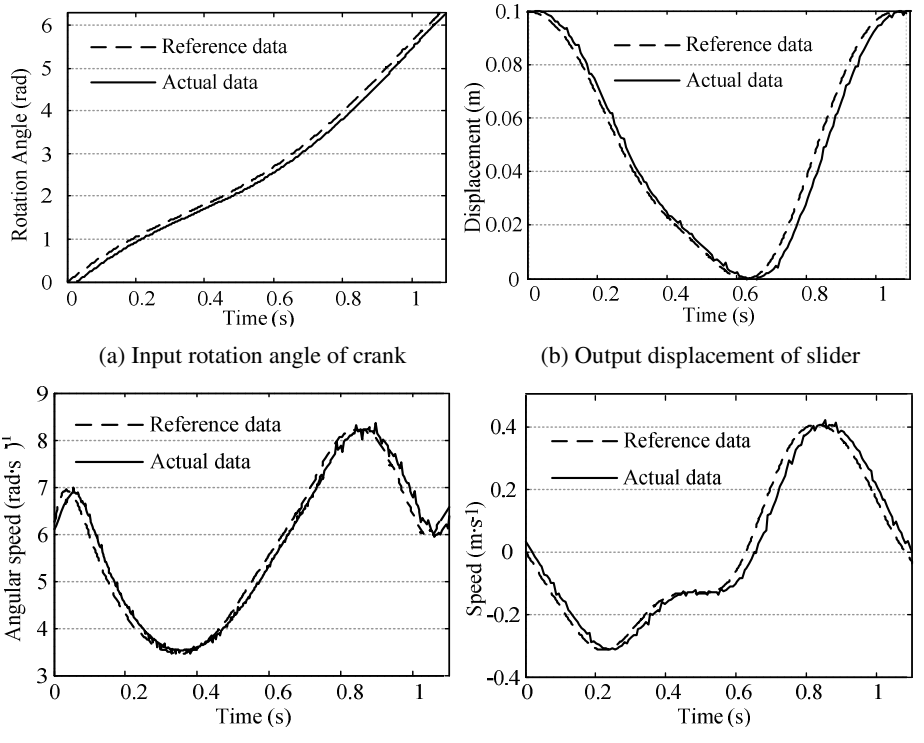
4 Experimental Results

4.1 Tracking Performance

The mechanical press in Fig.4 utilizes a SRM as the power input. By properly designing the input speed, the output motion of press can pass through a desired trajectory. This paper plans the input motion characteristics with Bezier curves [11] to make the crank's

motion suitable for the drawing depth of 0.02m. Experiments are to verify the tracking performance of a flexible process with the press driven by SRM.

In the experiment, the angular speed of crank in Fig.6(c) and the displacement of slider in Fig.6(b) are sampled in real time. The actual data in Fig.6(a) is a smooth curve for it is estimated by integral operation of Fig6.(c), while the actual data in Fig.6(d) fluctuates around the reference data because it is estimated by differential operation of Fig.6(b). Also, as Fig.6 shows, the actual measured data lags behind the reference data, which causes the tracking error of system. That is because of a limited communication baud rate of RS232, a time lag between the control operation and the measurement operation, and the inherent accuracy of hardware setup in the experiment system.



(a) Input rotation angle of crank

(b) Output displacement of slider

(c) Input angular speed of crank

(d) Output speed of slider

Fig. 6. Tracking performance of press for deep drawing

4.2 Operating Noise Performance

There are two major reasons for the noises reduction of servo presses machine. First, the slider of the press machine is only controlled by the servo motor since the traditional press brake clutch is removed. Thus, the exhaust noise and friction noise can be avoided during the starting and braking process. Second, the speed of the slider is controlled by the servo motor in real time. The operating speed can be set to a relative low value when the slider touches the sample and hence the noise can reduce greatly.

Fig. 7. Experimental setup for operating noise measurement

In this paper, the mechanical press driven by SRM can flexibly work and then theoretically has obvious reduction of operating noise for it meets these two reasons. For the experimental setup shown in Fig.7, a precision sound level meter (HS5721) is used to measure the operating noise of the press driven by SRM during blanking process with the constant and variable speed s respectively. The blanking process always generates a lager noise than the deep drawing for the same material, and the noise is always larger if the material has a higher shear or tensile strength. Therefore, the noise produced by the blanking process of a stainless steel sheet is studied in this paper.

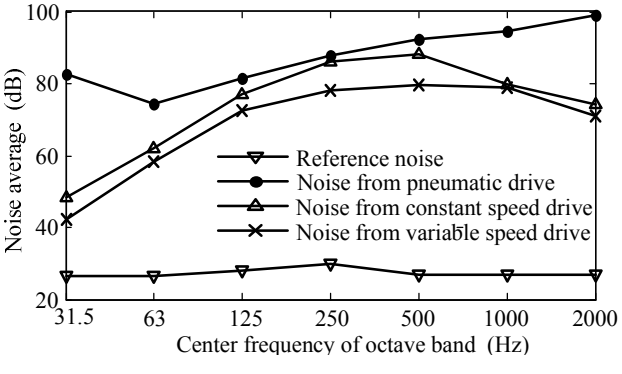


Fig. 8. Maximum values of sound pressure level at various frequencies

The "Weighting Network" switch is set to "filter" on the sound level meter (HS5721) for the noise measurement. Since there is an octave filter between the input and output amplifier, the noise spectrum analysis can be realized directly by rotating the octave filter's switch. As Fig.8 shows, the maximum values of the sound lever meter are measured at 31.5, 63, 125, 250, 500, 1000, 2000 Hz respectively.

The line with "▽" indicates the reference noise that is defined as the environmental noise when the press is not working. The line with "●" expresses the noise of the press

when the pneumatic drive working at a pressure 0.24 MPa. The noises of the press for the stainless steel blanking with constant and variable speed driven are also indicated by "Δ" line and "×" line respectively. The results show that the noise can reduce greatly when the press using constant speed control by SRM comparing with the pneumatic clutch. Moreover, the noise has the lowest level when using the variable speed control by SRM. Therefore, the press driven by SRM can reduce the noise pollution greatly.

5 Conclusion

In this paper, a working bench is presented for the application of SRM to a mechanical press, and its hardware configuration and software interface are described. In addition, a control system on the working bench is composed of state monitoring, detecting of slider position, variable speed control of SRM, and measuring forming force in a hybrid closed-loop structure. It forces the press to track flexible process curves and satisfy different motion requirements. The tracking performance of the SRM-based press is tested for a deep drawing with the flexible input curves planed by Bezier curves. Furthermore, the operating noise of the SRM-based press is measured when blanking a stainless steel workpiece. Experimental results indicate the flexible mechanical press driven by SRM has better tracking performance and lower operating noise. The flexible press is a suitable candidate to compete commercially with existing servo press for low-cost applications

References

1. Krishnan, R.: Switched reluctance motor drives: modeling, simulation, analysis, design, and applications, Boca Raton, FL, USA (2001)
2. Filicori, F., LoBianco, G.C., Tonielli, A.: Modeling and control strategies for a variable reluctance direct-drive motor. IEEE Trans. Ind. Electr. 40(1), 105–115 (1993)
3. Shang, W., Shengdun, Z., Yajing, S.: Application of LSSVM with AGA Optimizing Parameters to Nonlinear Modeling of SRM. In: IEEE Conf. Ind. Electr. Applicat., pp. 775–780 (2008)
4. Sahoo, S.K., Panda, S.K., Xu, J.: Indirect torque control of switched reluctance motors using iterative learning control. IEEE Transactions on Power Electronics 20(1), 200–208 (2005)
5. Lee, S.H., Kang, F.S., Park, S.J., et al.: Single-stage power-factor-corrected converter for switched reluctance motor drive. Electr. Pow. Syst. Res. 76(6-7), 534–540 (2006)
6. Cao, J.Y., Chen, Y.P., Zhou, Z.D.: Robust control of switched reluctance motors for direct-drive robotic applications. International Journal of Advanced Manufacturing Technology 22(3-4), 184–190 (2003)
7. Balaji, M., Kamaraj, V., Ramkumar, S.: Optimum Design of Switched Reluctance Machine for Electric Vehicle pplications Using Chaotic Particle Swarm Optimization. International Review of Electrical Engineering-Iree 6(2), 770–776 (2011)
8. Forrest, S.J., Wang, J., Jewell, G.W.: Analysis of an AC fed direct converter for a switched reluctance machine in aerospace applications. In: 5th IPEMC 2006, pp. 977–982 (2006)

9. Lu, K., Rasmussen, P.O., Watkins, S.J., et al.: A New Low-Cost Hybrid Switched Reluctance Motor for Adjustable-Speed Pump Applications. IEEE Transactions on Industry Applications 47(1), 314–321 (2011)
10. Shang, W., Zhao, S.: A Sliding Mode Flux-Linkage Controller with Integral Compensation for Switched Reluctance Motor. IEEE Transactions on Magnetics 45(9), 3322–3328 (2009)
11. Yan, H.S., Chen, W.R.: A variable input speed approach for improving the output motion characteristics of Watt-type presses. International Journal of Machine Tools and Manufacture 40(4), 675–690 (2000)

Determining the Suitability of FPGAs for a Low-Cost, Low-Power Underwater Acoustic Modem

Ying Li[1,*], Lan Chen[1], Bridget Benson[2], and Ryan Kastner[2]

[1] 7th Research Group, Institute of Microelectronics of Chinese Academy of Sciences
Beijing 100029, P.R. China
[2] Department of Computer Science and Engineering, University of California, San Diego
La Jolla, CA92037, USA
liyjennifer@yahoo.com.cn, chenlan@ime.ac.cn,
{b1benson,kastner}@cs.ucsd.edu

Abstract. Few dense underwater wireless sensor networks exist because commercial acoustic modems are designed for sparse, long range, applications and are thus too expensive and power consuming for small, dense, sensor nets. In order to enable the proliferation of dense underwater sensor networks, a new low-cost, low-power acoustic modem must be designed. This paper investigates whether an FPGA is a suitable platform for such a modem design. We present the design and in-water test results of a complete FPGA implementation of a Frequency Shift Keying underwater digital transceiver and compare its cost and power consumption to other research underwater digital transceivers implemented on different hardware platforms.

Keywords: underwater acoustic modem, FPGA, FSK, sensor networks.

1 Introduction

Small, dense underwater sensor networks (containing 10s to 100s of nodes spaced 10s to 100s of meters apart) have the potential to greatly improve environmental (pollution, coral reef, seismic, ocean current, etc.) and structural (oil platform, pipeline, undersea tunnel, etc.) monitoring by providing high temporal and spatial resolution data of a given region of interest. This data can provide increased insight into episodic and periodic processes localized in a region of interest leading to greater understanding of our earth's bodies of water and the increased safety of mankind. Few dense networks currently exist because commercial off-the-shelf (COTS) modems' power consumption, ranges, and price points are all designed for sparse, long-range, expensive systems rather than small, dense, and cheap sensor-nets [1]. For example, Linkquest underwater modems all cost > $8000 [2] and require a minimum of 4W transmit power. Therefore, a new low-cost (to allow for the deployment of 10s

* Corresponding author.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 509–517.
springerlink.com

to 100s of nodes), low-power (to allow for long deployment) underwater acoustic modem must be designed.

We have designed a low-cost modem for small, dense underwater sensor networks by starting with the most critical component from a cost perspective – the transducer. The design substitutes a commercial underwater transducer with a home-made underwater transducer (<$50) and builds the rest of the modem's components (its analog and digital transceiver) around the properties of the transducer to extract as much performance as possible. This paper focuses on determining whether a field programmable gate array (FPGA) provides a suitable hardware platform for the digital transceiver of the low-cost, low-power modem.

Most existing research underwater modem designs [3-5] make use of digital signal processors for their digital transceivers. As many control and signal processing applications can often be implemented quickly on DSPs, designers often do not even consider implementing the design on an FPGA due to the specialized knowledge and increased design time of a hardware implementation. And, although studies have shown that FPGAs have evolved into highly valued DSP solutions platforms that reduce overall systems costs and power consumption for high throughput applications, there is still little evidence as to whether an FPGA provides power and cost benefits for simpler, lower-throughput applications [3].

Thus, design time aside, this paper investigates whether an FPGA provides a suitable digital hardware platform for the low-cost, low-power underwater modem design. The major contributions of this paper are:

- A fully implemented and tested FPGA implementation of an underwater FSK digital transceiver
- A hardware software (HW/SW) co-design test platform for real-time functional verification of the digital transceiver
- A power consumption comparison between different underwater digital transceiver implementations

Section 2 is a brief overview of our complete acoustic modem system. Section 3 describes the details of the FPGA digital transceiver design including the digital down converter, symbol synchronizer, and modulator/demodulator. Section 4 describes the HW/SW co-design used for accurate control and I/O. Section 5 presents in-water test results of the full modem design and compares the power consumption and cost of digital transceiver to existing research underwater digital transceiver designs. Section 6 is the conclusion.

2 Acoustic Modem Design

Underwater acoustic modems consist of three main components: 1. an underwater transducer, 2. an analog transceiver, and 3. a digital transceiver for control and signal processing. The most costly component is the underwater transducer as commercially available underwater omni-directional transducers (such as those as seen in existing research modem designs [4-6]) cost on the order of two to three thousand dollars. To

substantially reduce costs in our complete acoustic modem design, we substituted a commercial transducer with a lab-made transducer made from cheap piezoelectric ceramic material and potting compound. Our transducer has a center frequency of 40KHz, a narrow bandwidth, an input power capacity of about 50W, and costs <$50. To substantially reduce power consumption, our analog transceiver, under development, maximizes power efficiency in the transducer's frequency range, and contains a power management circuit to lower the output power of the transmitter when the actual distance between transmitter and receiver is small. The digital transceiver should be both low-cost and power efficient to compliment the full modem's design. The analog components' narrow operating frequency range governs the selection of the physical layer protocol for the digital transceiver implementation.

Frequency shift keying is a modulation scheme that has been widely used in underwater communications (especially in the shallow water, short distance environment we target) over the past two decades due to its resistance to time and frequency spreading of the underwater acoustic channel [4, 7]. Due to its relative simplicity and narrow bandwidth requirements, FSK is a suitable physical layer protocol for our low-cost, low-power underwater acoustic modem design as it can be implemented in a small, low power device and meet the frequency requirements of the modem's analog hardware. Table 1 shows the time and frequency parameters used in the FSK design.

Table 1. Modem parameters

Properties	**Assignment**
Modulation	FSK
Carrier frequency	40 KHz
Mark frequency	1 KHz
Space frequency	2 KHz
Symbol duration	5 ms
Sampling Frequency	192 KHz
Baseband Frequency	16 KHz

3 Design of UWSN FSK Digital Transceiver

3.1 Digital Down Convertor

The digital down converter is responsible for converting high resolution signals to lower resolution signals to simplify subsequent processing. It takes the incoming signal *adc_in* and multiplies it with a locally generated 40kHz signal. The mixed signal then passes through a low pass filter to filter out the high frequency components. Then the signal is downsampled from 192kHz to 16kHz to reduce processing power. The low-pass filter is a small 20-tap FIR filter designed using Spiral tool [8].

3.2 Modulator/Demodulator

The modulator/demodulator is responsible for translating a bit stream into a waveform and vice versa by shifting the frequency of a continuous carrier to the 'mark' or 'space' frequency each symbol period.

The modulator takes a binary input and selects to generate a sinusoidal wave using a cosine look up table. The phase angle offset is calculated using the formula:

$$\text{Offset} = \text{round}\ (\text{size} * F / Fs) \tag{1}$$

Where *size* is the number of elements in the look up table, *F* is the mark or space frequency and *Fs* is the sampling rate.

The demodulator uses the classic 'matched' filter structure, which is optimal for FSK detection with white Gaussian noise interference. It works by sending a symbol duration of the received signal through two add-and-shift band-pass filters. An energy detection block is applied to determine the relative amount of energy in each frequency band.

3.3 Symbol Synchronizer

Symbol synchronization, the ability of the receiver to synchronize to the first symbol of an incoming data stream, is the most critical and complex component in our digital transceiver design. When the modem receiver obtains an input stream, it must be able to find the start of the data sequence to set accurate sampling and decision timing for subsequent demodulation. Without accurate symbol synchronization, higher bit error rates incur thus reducing the reliability of the wireless network.

Our symbol synchronization approach relies on the transmission of a predefined sequence of symbols, often referred to as a training, or reference sequence. The transmitter sends a packet that begins with the reference sequence and the receiver correlates the received sequence and the known reference sequence in order to locate the start of the packet (and start of the first symbol). When the reference and receiving sequence exactly align with each other, the correlation result reaches a maximum value and the synchronization point can be located as the maximum point above a pre-determined threshold. We use a 15-bit Gold code as our reference sequence and perform a correlation with a 15-bit orthogonal Gold code to set a dynamic threshold. Due to space constraints, details of our symbol synchronization implementation and design considerations can be found in [9].

4 HW/SW Co-design

Xilinx Platform Studio 10.1 is applied to build a HW/SW co-design for accurate control and I/O of the digital transceiver. The co-design consists of the digital transceiver, a UART (Universal Asynchronous Receiver Transmitter) to connect to serial sensors or to a computer serial port for debugging, an interrupt controller to

process interrupts received by the UART or the transceiver, logic to configure the on board ADC, DAC, and clock generator, and MicroBlaze, an embedded microprocessor to control the system (Figure 1).

The MicroBlaze processor interfaces to the digital transceiver through two fast simplex links (FSLs), point-to-point, uni-directional asynchronous FIFOs that can perform fast communication between any two design elements on the FPGA that implement the FSL interface. The MicroBlaze interfaces to the interrupt controller and UART core over a peripheral local bus (PLB), based on the IBM standard 64-bit PLB architecture specification.

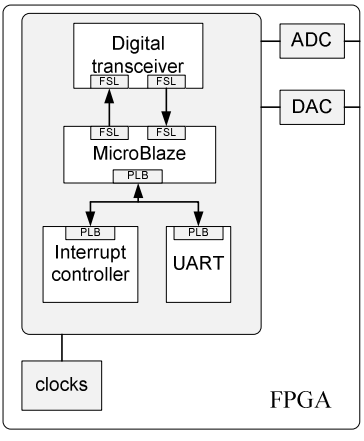


Fig. 1. HW/SW Co-Design for the digital transceiver

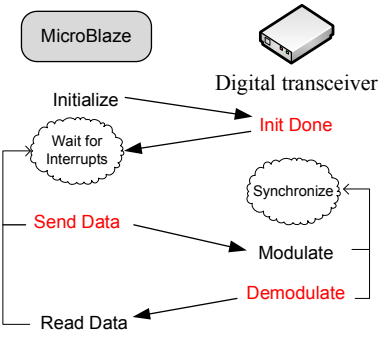


Fig. 2. Modem Control Flow. Interrupts are shown in red

Upon start-up, the MicroBlaze initializes communication with the digital transceiver through sending a command signal through the FSL bus signaling the transceiver to turn on. When the transceiver is ready to begin receiving signals, it sends an interrupt back to MicroBlaze to indicate initialization is complete. The transceiver then begins the down conversion and synchronization process, processing the signal received from the ADC and looking for a peak above the threshold to

indicate a packet has been received. If the transceiver finds a peak above the threshold, it finds the synchronization point, and demodulates the packet. The demodulated bits are stored in the FSL FIFO. When the full packet has been demodulated, the transceiver sends an interrupt indicating a packet has been received and the MicroBlaze may retrieve the packet from the FSL. The transceiver then returns to synchronization, searching for the next incoming packet.

After initialization, the MicroBlaze remains idle, waiting for interrupts either from the transceiver or UART. If it receives an interrupt from the transceiver indicating that a packet has been demodulated, the MicroBlaze reads the bits from the FSL FIFO and sends the bits over the UART to be printed on a computer's Hyperterminal for verification. If the MicroBlaze receives an interrupt from the UART, indicating that the user would like to send data, the MicroBlaze sends a command to the transceiver to send the bitstream the MicroBlaze places in the FSL. The transceiver then modulates the data from the FSL and sends the modulated waveform to the DAC for transmission. The MicroBlaze then returns to waiting for interrupts from the transceiver or the UART and the transceiver returns to synchronization, searching for the next incoming packet. This control flow is depicted in Figure 2.

5 Results

The digital transceiver design has been fully tested on a FPGA prototype platform, the DINI DMEG-AD/DA [10]. The test packet consisted of the 15 bit Gold Code ,011001010111101, followed by a 385 bit packet of randomized signals. Using our full modem design (the homemade transducer, analog transceiver, and digital transceiver), we sent packets in a 0.5m tank of water, a 50m swimming pool, and in Westlake, a freshwater lake in Westlake Village, CA. The results of bit error rate vs. SNR, are shown in Figure 3 [11]. All tank and lake tests achieved a bit error rate of less than 5% for distances up to 380 meters. The current modem design, without channel equalization, did not perform well in the pool's high multipath environment.

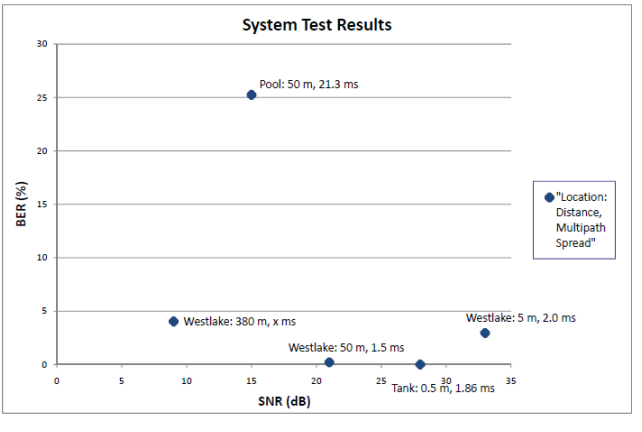


Fig. 3. Test results

We obtained a power estimate of the FSK digital transceiver design on various FPGA devices by entering the resource values of the HW/SW co-design into the Xilinx XPower Estimator 9.1.03 and the Altera Cyclone IV PowerPlay Early Power Estimator. The devices reported (except for the Virtex IV XC4VLX100 which is the device we used for prototyping) are in device families known for their low power consumption (the Xilinx Spartan 6 and the Altera Cyclone IV being some of the newest FPGA device families on the market). The particular devices reported are the smallest devices in their family that fit the total modem design. The letters 'Q', 'D', and 'T' in Table 3 stand for 'quiescent,' 'dynamic', and 'total' power respectively.

Table 2. FPGA power consumption

Device	Q Pwr (W)	D Pwr (W)	T Pwr (W)
XC4VLX100	0.775	0.2	0.975
XC3S4000	0.274	0.105	0.379
XC6SLX150T	0.212	0.021	0.233
EP4CE30	0.087	0.06	0.147

Table 3 compares the total digital processing power and cost of the digital transceiver design with existing underwater digital transceiver designs using various modulation schemes and platforms. From Table 3 we notice our FSK design on an FPGA provides comparable cost and power to other FSK designs.

Table 3. Modem design comparison

Mod	Device	Platform	T Pwr (W)	*Cost ($)
FSK[4]	Fixed DSP	TMS320C5416	0.180	45
FSK[12]	MCU	Blackfin 533	0.280	25
PSK[4]	FP DSP	TMS320C6713	2.0	25
DSSS[5]	FP DSP	TMS320C6713	1.6	25
Ours	FPGA	XC6SLX150T	0.233	14
	FPGA	EP4CE30	0.147	40

6 Conclusion

This paper investigates whether an FPGA provides a suitable digital hardware platform for the low-cost, low-power underwater modem design. We describe a

complete FPGA implementation of a FSK underwater acoustic digital transceiver and compare its cost and power consumption to other research underwater digital transceiver designs. Although an apples to apples comparison cannot be made, the cost and power estimates suggest that the FPGA provides comparable cost and power consumption to a fixed point DSP and offers the advantage of a relatively easy transition to ASIC once volume dictates.

Our anticipated cost and power estimates for the full underwater acoustic modem prototype (without housing or batteries) are shown in Table 4.

Table 4. Cost and power estimates for the underwater modem

	Cost ($)	**Power (W)**
Transducer	50	N/A
Analog Transceiver	150	TX:1.2-6.9 RX:0.275 Idle:0.240
Digital Transceiver	150	TX: 0.097 RX: 0.147 Idle: 0.087

Acknowledgment. This material is based upon work partially supported by the china scholarship council and partially supported under national science foundation grant #0816419 and a national science foundation graduate research fellowship.

References

1. Heidemann, J., Li, Y., Syed, A., Wills, J., Ye, W.: Research Challenges and Applications for Underwater Sensor Networking. In: Proceedings of the IEEE, WCNC 2006 (2006)
2. Link Quest, Inc. Underwater acoustic modems, http://www.link-quest.com/html/uwm_hr.pdf
3. BDTI "FPGAs vs. DSPs: A look at the unanswered questions" DSP Design Line (November 1, 2007)
4. Freitag, L., Grund, M., Singh, S., Partan, J., Koski, P., Ball, K.: The WHOI Micro-Modem: An acoustic communications and navigation system for multiple platforms. In: Proceeding of OCEANS Conference (2005)
5. Ronald, A.I., Lee, H., Kastner, R., Doonan, D., Fu, T., Moore, R., Chin, M.: An Underwater Acoustic Telemetry Modem for Eco-Sensing. In: Proceedings of MTS/IEEE Oceans (September 2005)
6. Yan, H., Zhou, S., Shi, Z., Li, B.: A DSP implementation of OFDM acoustic modem. In: Proc. of the ACM International Workshop on Under Water Networks (WUWNet), September 14 (2007)
7. Kilfoyle, D.B., Baggeroer, A.B.: The State of the Art in Underwater Acoustic Telemetry. IEEE Journal of Oceanic Engineering 25(1) (January 2000)
8. Spiral, http://spiral.net/hardware/filter.html

9. Li, Y., Benson, B., Zhang, X., Kastner, R.: Hardware Implementation of Symbol Synchronization for Underwater FSK. In: IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (2010)
10. DINI Group, http://www.dinigroup.com/index.php?product=DNMEG_ADDA
11. Benson, B.: Design of a Low-cost Underwater Acoustic Modem for Short-Range Sensor Networks, PhD thesis. University of California, San Diego (2010)
12. Vasilescu, I., Detweiler, C., Rus, D.: AquaNodes:An Underwater Sensor Network. In: Proceedings of ACM International Workshop on Underwater Networks (September 2007)

The Application of Multi-media Surveillance Systems in the Open Nursing Training

Haiyang Zhang[1], Guanghui Li[2], Junlei Zhang[3], Fengxia Wang[3], and Rong Li[3,*]

[1] Pediatric Surgery of First Affiliated Hospital of Xinxiang Medical University, Weihui, Henan Province, China
[2] Management College of Xinxiang University, Xinxiang, Henan Province, China
[3] Nursing college of Xinxiang Medical University, Xinxiang, henan Province, China
huimin000@126.com, 35290915@qq.com, 471784045@qq.com, 625475834@qq.com, lr2665@yahoo.com.cn

Abstract. Objective To explore the open experimental training model using multi-media control system for nursing students to develop their core competencies. Methods Different model designs of multi-media control system were applied in the study group. After the one year training program, a self-designed survey was used to collect the students' evaluation. A comprehensive test was carried out and the scores of it were compared to the traditional training group. Results Test scores of the study group was significantly higher than the control group's ($P<0.01$), and the survey results showed that using multi-media control system in the open experimental training program was highly accepted by the nursing students (97.1%). Conclusions Using multi-media control system in the open experimental training program can improve the nursing students' competencies and comprehensive quality.

Keywords: Multi-media control system, Nursing skills training, Open experimental training.

1 Introduction

At present nursing is one of the six short professional in our country. Nursing is a strongly practical comprehensive area, practice teaching can not only deepen the understanding of nursing students knowledge, but also develop standardized operation ability and the innovation ability, which plays an important role and position [1]. Skill training is an important segment to cultivate students' ability. Nursing skills operation is the basic technology of clinical nurse, directly affect nursing practice work.

Nursing teaching practice which we experienced: skill training is a systematic, timely, effective and comprehensive, elaborate, to select a skilled training method timely, can arouse the enthusiasm of nursing students training, innovation ability and practice ability. For years, medical education use the basic courses, professional class, clinical practice "tasting" teaching mode widespread, with disciplines for unit ,teaching and clinical work is apart, so the study pressure is big, the study effect is bad, no

* Corresponding author.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 519–524.
springerlink.com

interest, adapt ability is poor. In the traditional care experiment teaching, "demonstration- practice-guide", the simulation method and role playing method are used [2]. First commonly, the instructor teach the purpose, things, steps, and the notice of an operation, then the teacher teach and explain, then student practice groups, teacher guide , and finally correct the student feedback error. That in years of nursing practice teaching, we found that "the independent learning ability", "hands-on ability" seriously short, facing a nursing procedures do not know how to begin. In nursing practice, mechanical imitation of the teacher's demonstration, lacking the ability of learning and management [3]. The high-speed development of modern science and technology, nursing profession is considered as a life-long learning. Nursing practice, not only can deepen the theoretical knowledge, but can train students' standardized operational ability and the adaptational ability.So how to train the ability is the important problem for nursing teachers.

Open nursing laboratory is now widely used in nursing teaching, providing all-exercises places, training students' practical and innovative ability, and improve the students' comprehensive, to be an important auxiliary form of nursing experiment teaching [4]. Therefore, to enhance reform and innovation of nursing laboratory management, fully exploit and develop existing teaching resources, improve the utilization rate, better service for experimental teaching, which is worthy of attention [5-6]. Through 1 year the system which has obtained a better effect.

2 Objects and Methods

2.1 Objects

The level of 2009 nursing students' 110, for the observation group, level 2008 nursing students' 110, for the control group.

2.2 Methods

The control group which uses the routine management and teaching methods. The observation group uses the multi-media monitor system management teaching mode.

Timely Mode: teachers and experimental personnel through the multi-media monitor guidance system to operate the students, discover the problems and correct timely. Because the open experiment takes the students' self management as the model, the teacher or the researcher who can't in at that moment. Therefore, students in the skilled training questions such as not meeting the corrective and feedback. The teacher and the researcher who do not know the process of problems and reasons, and thus pointed to solving it, can make the students feel badly, reduce learning enthusiasm and effect learning effect. Using closed-circuit TV system, to solve the problem, the teacher and the researchers who if not in, also can through the multi-media monitor guidance system of information and feedback, control training process. At the same time, through the closed circuit television dynamic observation, strengthen management, avoid the unwinding of the phenomenon, ensure public training quality.

Record Mode: operation skills from master to skilled, depends on teacher. Teaching standard and student's initial imitate. Therefore, the teaching process weather be understanding is crucial. To solve the teach angle deviation and numbers limitations, through the multi-media monitor guidance system camera functions, the whole process of teaching record down, for special complex operating procedures can conduct action decomposition, the explanation is the main point, video production by CD or keep it down, is useful for students' future study . At the same time, students can record their operation exercises process, information feedback to teachers, so that establish communication platform between teachers and students.

Playback Modes: the biggest characteristic of multi-media records is repeatability. Students in operating practice, can use it playing the teacher's demonstration teaching process, when viewing and operation, review and consolidate the content of the class to and operating procedures, and master the methods and steps, enhance memory. Meanwhile, students can also by watching his operation exercises, and compared with teachers teach action , found the problem, the gap, analyze the reasons and deepen our impression, improve the operation skills at the same time, develop their own observation, analysis and independent thinking ability.

Video Mode: electronic computer operation is the basic and the general skills of the students. Teachers' teaching process recorded CD or video, based on the computer and network resources, the way which students are more ready to accept. The maximum advantage of the teaching model is making the teaching and learning is limited in OR out class no longer, but everywhere can learn more flexible and diversified.

The Evaluation Method: compared two groups the comprehensive quality of nursing students and performance. At the end of the experiment, the observation group students use self-made questionnaire to survey multi-media monitor system of open experimental training, which derived from the domestic and foreign relevant literature material, and modify through the Delphi experts consultation method, the internal consistency, content validity, structural validity evaluation, internal consistency test results Cronbaeh 's q coefficient is 0.86, binary reliability coefficient > 0. 80, content validity index 0.83, factor analysis showed that the basic idea and consistent, structural validity is better. Show that the questionnaire of the high reliability and validity. After the experiment, No-named questionnaire, unified time to fill out, to complete the questionnaire by objects, questionnaire was 110 totally, 104 of the recovery, efficient rate is 94.5%.

Statistical Methods: SPSS 17.0 statistical soft ware.

3 Results

The Two Groups of Nursing Students Comprehensive Quality Assessment: the control group is (87.02 ± 3.003), the observation group is (90.38 ±4.96), the difference was statistically significant ($t = 5.900$, $P < 0.01$). The result is table l.

Table 1. The comparison of the two groups of nursing students comprehensive quality

Group Number	80~		86~		90~		95~100		
	Num	per	Num	per	Num	per	Num	per	
Observation	104	31	29.81	54	51.92	19	18.27	0	0
Control	104	22	21.15	33	31.74	32	30.77	17	16.34

The Evaluation Multi-media Monitoring System: the observation group students use multi-media monitor system, table 2.

Table 2. The effect of multi-media supervisory system of the open laboratory training

Training effect	totally agree	agree	neutral	disagree	totally disagree
Deepen understanding	36	56	12	0	0
Favorable team	43	58	3	0	0
Improve critical thinking	44	51	7	2	0

4 Discussion

Opening laboratory increases students' extracurricular operation time, strengthen and improve the content [7]. Therefore, management should enhance the students' extracurricular practice and effectiveness of the teaching quality as[8]. Multi-media monitor system is not only in order to check students' participation, class order, but to observe the correct rate, the action is in point, behavior is standard, to find out problems timely, and the student feel extracurricular exercise is not blind. In management, adds a humanistic care, make more effective management, more humanistic. The control group and the observation group experiment examination results were statistically significant ($P < 0.01$). Nursing open experiment combined with the multi-media teaching mode can obviously improve the quality of teaching and the teaching effect. Set up information feedback platform, realize the teaching interaction. The teaching quality, depends on teaching process, and teaching process is controlled by the transmission of information and feedback. 94.2% of the students agree with nursing skills training in extracurricular monitoring system which can promote information communication, with the closed-circuit TV and video material recording function, teachers can understand and master students nursing skills operation process, between the teachers and students builds a information feedback bridge.

To explore various flexible learning forms, stimulate students' interest in study. Open laboratory and the effective use of multi-media teaching resources, independently

of time and space for students, is also more vivid and lively [9]. Through the multi-media monitor system, the active exchange between the classmate, to get the information, and teach consult discussion, thus enhance the learning interest and entertaining, become boring for vivid image, fully mobilize students' learning enthusiasm and initiative, arouse their learning enthusiasm and interest, and improve the ability to study independently, and learn to communicate with others and get along. The students who use multi-media monitor system for nursing skill ,the rate of deepen understanding, good in teamwork, critical thinking for improving were 95.2%, 97.12% and 91.34% respectively, multi-media in extracurricular monitoring system, not only improve the comprehensive ability of nursing students,but more solidate the patient communication.

Nursing is a strong practical comprehensive area, the key is in the ability, including the beginning ability, innovation ability, thinking ability, strain capacity, coordination ability, communication skills, and so on, have elegant appearance and good psychological quality. Opening laboratory is the ability to implement effective approaches to train [10], the results of the survey showed that 98.07% of nursing students think through effective use of the multi-media monitor system, enhance the analysis and the ability to solve problems, and improve the practice operation, to develop students' potential, actively guide the student active exploration, diligently to begin, all-round, multi-level exercise students' comprehensive ability, and makes teaching and the self development, so as to promote the students' knowledge, theory, attitude, emotion, skills, the development of the art of sublimation, utmost make students' knowledge and clinical standards.

Acknowledgment. This paper is a teaching reform task supported by Association of Social Sciences, which is the subject of Top three hospitals in the undergraduate nursing students needs and abilities and quality requirements (SKL-2011-2317). Rong Li is the corresponding author.

References

1. Chen, J.X., Xiao, S.F., namely earth Chen., et al.: Construction nursing professional experimental ideas. Fujian Institute of Traditional Chinese Medicine Journal (1), 57 (2004)
2. Liu, C., Ou, S.K., Lampe, M.: The basic nursing experiment report copies of the basic nursing care in the teaching application field. Handan Medical College Journal (3), 200–201 (2004)
3. Niu, G., Li, X.H.: Basic nursing teaching reform nursing students of cultivating the ability of autonomic learning evaluation. The Chinese Modern Care Med. 12(17), 1655 (2006)
4. Lan, Z.T.: Open nursing laboratory management and teaching mode is discussed. Journal of Nursing Science 37(7), 18–20 (2007)
5. Hu, Y., Jia, H.L., Wang, J.Q., et al.: The united network laboratory open platform of the teaching reform of management nursing skills. Nursing Journal (12), 14 (2009)
6. Zhang, J.: Nursing opening laboratory management practice. Nursing Journal 21(4), 65–66 (2006)

7. Ge, X.R., Li, H.: Nursing skills training and guidance model open the exploration and practice. Nursing Research 37(5), 1397 (2008)
8. Ma, J.P., Li, Y.L.: Play exiting in nursing experiment teaching experience of main body role. Modern Care (12), 964–965 (2005)
9. Li, J., Zhang, L., Wan, X.Y.: For nursing the open laboratory investigation. Modern Nursing 13(21), 1990–1991 (2007)
10. Chen, X.: Nursing the opening experiment teaching improvement measures. Nursing Research 23(2), 547–548 (2008)

PDM Technology and Its Application in Manufacturing

Junming Zhang

Shandong Transportation Vocational College, Shandong Weifang, 261041

Abstract. Product Data Management, PDM is a new technology in recent years, its main features include electronic storage and document management, product flow / design process management, product structure and configuration management, lifecycle management and integrated development interface. He will be the entire enterprise as a whole, across the entire engineering community to quickly develop products and business process transformation, but also in the distributed enterprise management model, based on the application and other important tools to establish direct contacts. In the application process, a comprehensive and correct understanding of PDM, the development needs of the market dynamics to create an object model, the emphasis consulting, focusing on cooperation.

Keywords: PDM, Machinery Manufacturing, workflow, product flow, electronic warehouse.

1 Introduction

PDM is a product-centric, by calculating the machine network and database technology, the production process of all product-related information and process integration together, unified management, product life cycle data in its consistency, the latest and safety, engineering and technical personnel to provide a collaborative work environment to shorten product development cycles, reduce costs, improve quality, gain competitive advantage for enterprises. PDM technology appeared in the early eighties, when the purpose is only to solve a large number of engineering drawings, documents.

Management difficulties, the early development path are: management design drawings eleven management design drawings, electronic documents (electronic Sub-document is generated with Word and other word processing software, technical specifications, material specifications, etc.) eleven management design drawings, electronic files, a management change a single design drawings, electronic documents, change orders, bills of materials (BOM). A mature PDM system enables all involved in creating, sharing, maintaining design intent throughout the product life cycle in freely shared with all product-related heterogeneous data, including drawings and digital documents, CAD files and product structure [1].

In short, PDM can be defined as software technology, product-centric, to achieve product-related data, processes and resources for integrated management of technology integration.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 525–529.
springerlink.com

2 The Necessity and Feasibility of PDM Applications in Manufacturing

In manufacturing companies, products, materials, manufacturing activity is not only a conversion process, but also a complex transformation process of information transmission of information is correct, a direct impact on the normal production process. Therefore, the orderly management of information in the enterprise has a pivotal position. Many companies are plagued by a large number of design data, especially in the 20 since the 1960s and 1970s, companies in the design and production process to start using CAD, CAM technology, a large number of electronic data and find difficult to manage. In the manufacturing of computer integrated manufacturing environment, product development and design in the modern demand-driven process automation, to produce a set of database data management, network communications and process control in one, able to achieve the distributed ring.

Environment to design activities and information exchange and sharing of the design process for dynamic adjustment and monitoring of product data management (Product Data Management, PDM) technology [2].

PDM is mainly used in the initial application of engineering change management, and engineering fields such as CAD drawings, document review and approval, but today, according to the emphasis on sharing of common information point of view, has expanded the product development model meaning, its scope included the allocation of resources, manufacturing, planning and scheduling, procurement, sales, market development and other areas; and the entire manufacturing enterprise as a whole to consider some issues.

PDM product data efficiently from concept to design, calculation and analysis, detailed design, process design, manufacturing, sales maintained until the demise of the entire product life cycle and various stages of the data, according to some data model to define the learning organization and management, product data throughout the life cycle consistent with the latest, sharing and security. He will be a good business-oriented organization of production, to enable enterprises to improve product quality, shorten development cycles, improve efficiency, accelerate speed to market, thereby enhancing the competitiveness of their products.

3 PDM Technology Upgrading in Manufacturing, the Prospects for Management

3.1 PDM Application Status

PDM technology in the 1990s has seen rapid development, PDM can be seen as an enterprise information integration framework (Framework). Various applications such as CAD / CAM / CAE, EDA, OA, CAPP, MRP ... ... through a variety of ways, such as application interface development (package), etc., directly as one by one, the object (Object), and is integrated, so that the distribution of the various parts of the enterprise, used in various applications all product data to be highly integrated, coordinated, shared, all the product development process can be highly optimized or restructured. Currently, many large international companies are gradually restructuring its business

processes to support (BPR), concurrent engineering (CE), ISO9000 quality certification, so as to maintain the competitiveness of enterprises of key technologies. In recent years, PDM is a product of the fastest growing industry in a technology. PDM applications to the enterprise of extraordinary achievements. PDM technology has been widely used abroad, the U.S. company, CIMdata survey, 98% of companies have implemented PDM. Currently there are many well-known foreign PDM software, such as: SDRC's Metaphase, EDS company IMAN, PTC's Windchill, IBM's Product Manager, CV company Ortega, etc., which basically represent the highest on today in the PDM level. Numerous software vendors have also developed their own PDM products, such as: Zhejiang Ruifeng software Tianchuang developed GS-PDM system, Wuhan Tianyu company IntePDM, Tongfang software company TFPDMS, Huazhong open mesh integration Technologies KMPDM, Lei technology companies in Europe SmarGroup, Beijing Daheng company DHPDM, information technology company on the Hai Sipu SIPM / PDM software, Chinese Academy of Sciences of the Keith Corporation, Nanjing Genesis Corporation, Northeastern University, Qing Hua Gaohua companies and the Alpine's PDM software, etc. [3].

3.2 PDM Trends

With the application of PDM technology, it's becoming more and more concern for everyone. From the current trend, mainly the following aspects.

1) The establishment of a common development platform for all PDM PDM PDM software developed by software developers have their own characteristics, that is suitable for which industries and which companies, but its function is concerned, there are many similarities, the development of similar waste development time, increase the cost of the software, delayed time to market the software.

2) PDM software development industries will not like the CAD software, you can use to take over, use the PDM software tailored for the enterprise needs, the need for the implementation of PDM, which increases the risk and cost of application. For industry development, then the implementation of personalized, not only can reduce the implementation risk, but also can reduce the implementation costs. PDM applications with the industry characteristics of areas: auto assembly, auto parts processing, aerospace, survey and design, medical equipment, hospitals, machine tools, appliances, furniture and so on.

3) PDM and Web technology combined with the trend of economic globalization, the production company with a product may not be the same area, not even the same country, that is, there are some virtual company or affiliate companies, to the remote transmission of data and sharing, without network technology. Now the development of three-tier application model is the integration of Web services and C / S database application advantages. Through the Web platform, E-mail, electronic bulletin boards, newsgroups and other advanced means of communication can quickly between employees in the enterprise, between enterprises and facilitate access to information, saving business expenses. Therefore, PDM to the three-tier architecture (browser / Web server / database server) will be the development trend of the future.

4) The integration of PDM with other systems to play well with other applications to have affinity, that is, data exchange capabilities. PDM system integration support, first in should have minimal integration, that is, integrated CAD / CAM products and MRP

Ⅱ / ERP software; followed by the integration on the depth and robustness, not only in the PDM system startup applications, but also for A fully integrated, it should be able to directly PDM PDM system operations such as check (Check in), detection (Checkout) etc., and can exchange data between two systems and messaging. CAD / CAPP / CAM / PDM / ERP / MRP Ⅱ integration is a major trend in the development of CIMS [4].

4 PDM in CIMS Application Examples

PDM in the CIMS (Computer Integrated Manufacturing System) integration to occupy an important position, not only the general management of data, and has become a platform for data exchange. PDM system's basic function is the product from design to the demise of the entire life cycle of data, according to some abstract model to be defined, organization, delivery and management, product data throughout its life cycle consistent with the latest, share and security. Product life cycle, including product data from concept design, calculation and analysis, detailed design, process design, manufacturing, sales maintenance until the demise of all stages of product data.

The first PDM created a single data source, design, technology, needed for the production of uniform preservation and maintenance of data, avoiding data redundancy and confusion. Second, PDM provides a data interface with other systems, making the other systems to access a single data source is very convenient and fast. PDM system integration is key for data communication, product structure information, routing information, drawings, information transmission. Transfer process is as follows:

5 PDM Applications in Manufacturing Problems and Solutions

Traditional data management PDM system information model and its essence is the product structure tree and independent documents associated with the decomposition that is open to two modules, namely, product structure management and document management plans. They are the two main traditional PDM system modules, the traditional document management systems usually map directly manage physical storage of computer files, and product structure management, product structure data is independent of the map there is a separate document and data collection. This processing mode using the computer model is relatively simple, easy to understand, but this model with the data management information management information technology advances, more and more difficult to meet the actual needs of large enterprises [5].

6 Conclusion

Currently, manufacturers have been the main application sectors of PDM, PDM software vendors with growth and fierce competition, the team, PDM will be more wide range of applications, PDM software, the price will come down, as accepted by most businesses. Then, for an enterprise, the implementation of PDM is no longer a problem or not, but the implementation When did the problem. Successful implementation of manufacturing PDMis a successful manufacturer of computer

integrated manufacturing system the best choice, the successful implementation of PDM systems will significantly enhance the market competitiveness in the global economic competition in an invincible position.

References

1. Zhang, Z.-G., Di, W.: PDM information technology in manufacturing a number of key technology research. Computer and Digital Engineering (3), 119–123 (2008)
2. Li, F., Kim, Y.: PDM Technology Analysis and implementation strategies. Computer Applications (1), 41–45 (2001)
3. Zhou, M., Wanchuan, G., et al.: On the PDM of the machinery industry, the impact of enterprise information technology. Machinery (34), 116–117 (2007)
4. Rong, Z.: PDM technology status and development trend. Machinery (34), 114–115 (2007)
5. Has, F.: Traditional PDM system problems. Chinese Manufacturing Informatization (3), 44–46 (2004)

Research of Analysis Method Based on the Linear Transformation of the Hinge Four Bar Mechanism Motion

Ming-qing Wu*

Shandong Transport Vocational College, Shandong Weifang, 261206

Abstract. Use the analysis method based on the linear transformation to research the sports of hinge of four-bar motion. Through the use of "blind spots" analysis method, according to the active thing of hinge four-bar is located in or out of the "blind area" it can be divided into two kinds of circumstances for analysing. Establishe parametric equation of movement points, compare and test this results and the results of using ADAMS to simulate mechanism of the motion, perfectly fit hinge four-bar sports practice.

Keywords: Linear transformation, hinge four-bar mechanism, motion analysis, parameterization.

Hinge four-bar mechanism is one of commonly used sports organizations. It is convenient to realize the known motion law and represent the known trajectory. It is widely used in mechanical industry, mining production, construction and installation projects such as the vehicle doors and Windows openning, realizing the automatic device of boiler, the lifting of tipper, and the arm institutions of excavator. Its movement analyzing is one of the typical mechanism movement analysises.

1 Classification of the Hinge Four-Bar Mechanism Based on the Movement in the Blind Hinge

As shown in figure 1, when the hinge four-bar ABCD moves to the position of $AB''C''D$, B'', C'',D are in the same line. There are two area that $\overline{AB}$ stem can't reach. As shown in the picture the α dihedral area and the symmetry region corresponding with X axial. The two areas that can't reach are "blind area" of the active stem $\overline{AB}$ in the hinge four-bar mechanism. As shown in the graph the 2 area is the second "blind spots"of active stem $\overline{AB}$. We can understand it according to the triangle theorem, such as the hinge four-bar from position $AB''C''D$ along the counterclockwise rotation. In the process of turning use the solid line showing the position of $ABCD$ arrivving the position of $AB'C'D$. In this position, B', C',D three in the same line again, that is connecting rod $\overline{BC}$ and side link $\overline{DC}$ move again in the same line. At this time, the distance between

* Author: Wu Mingqing (1977 -), male, Han nationality, Shandong Weifang Shouguang people, Shandong Transport Vocational College, research for the mechanical engineering.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 531–536.
springerlink.com

B' and D Equals to the sum of the rod $\overline{B'C'}$ and rod $\overline{C'D}$. If the pole $\overline{AB'}$ continue rotating along the counterclockwise, the distance between B' and D will continue to increase, and the distance between rod B' and rod D will be greater than the sum of the length of $\overline{B'C'}$ and $\overline{C'D}$, this violates the triangle relations among theorem. So the area that active stem can reach will be divided into two parts, two "blind area" and two normal area of movement.

For hinge four-bar mechanism if two side links all have blind area, it is double rocker organization. If two side links don't have "blind area", it is double crank mechanism. If one has "blind area", and the other dosen't have"blind area", it is the crank rocker organization.

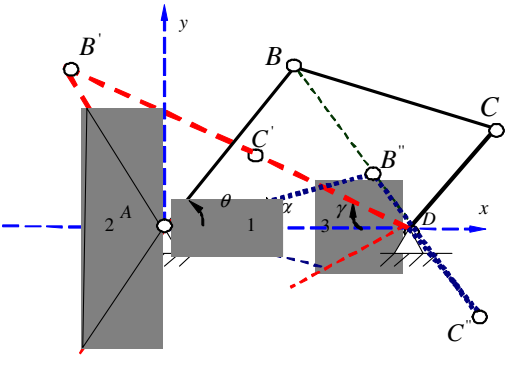


Fig. 1. Four-bar movement

2 Analysis Method of Four-Bar Mechanism Based on Linear Transformation

According to the right-hand rule establish the local coordinate system $x'Dy'$ and the global coordinate system xAy of hinge four-bar $ABCD$ as shown in figure 2 and figure 3. φ_1 is the angle that the local coordinate system $x'Dy'$ relative to the global coordinate system xAy relative to turned from the Z axis, θ is the angle between $\overrightarrow{AB}$ and the positive direction of x axis in the global coordinate system . l_1、l_2、l_3 is respectively the length of $\overline{AB}$、$\overline{BC}$、$\overline{DC}$.

According to above analysis, hinge four-bar mechanism can't enter the "blind area", so we should make sure the hinge four-bar mechanism is in "blind area" or out .It is the first task of the analysis and calculation to hinge four-bar mechanism: If $\overline{AB}$ is active pole, θ is the input of movement mechanism, the coordinates of B、C change constantly when mechanism moving, and the coordinates of hinged point B in every movement position can be determined through the input Angle θ :

$$\begin{Bmatrix} x_B \\ y_B \end{Bmatrix} = l_1 \begin{Bmatrix} \cos\theta \\ \sin\theta \end{Bmatrix} \tag{1}$$

$$l_4 = \sqrt{(x_B - x_D)^2 + (y_B - y_D)^2} \tag{2}$$

if $l_4 \geq l_2 + l_3$ (a)or $l_4 \leq |l_2 - l_3|$ (b), the hinge four-bar mechanism is in extreme position or has entered the "blind spots".

If they do not conform to the above conditions, it need to analyse and calculate the hinge four-bar mechanism as follows:

The coordinates of the hinged points C in each sport position can be got by the method based on the linear transform:

$$\begin{Bmatrix} x_C \\ y_C \end{Bmatrix} = \begin{Bmatrix} x_D \\ y_D \end{Bmatrix} + \begin{Bmatrix} \cos\varphi_1 & -\sin\varphi_1 \\ \sin\varphi_1 & \cos\varphi_1 \end{Bmatrix} \begin{Bmatrix} x_C^{'} \\ y_C^{'} \end{Bmatrix} \tag{3}$$

Among them, $\{x_C^{'}, y_C^{'}\}^{\mathrm{T}}$ is the coordinate of the point C in the coordinate local coordinate system $x^{'}Dy^{'}$:

$$\begin{Bmatrix} x_C^{'} \\ y_C^{'} \end{Bmatrix} = l_3 \begin{Bmatrix} \cos\varphi_2 \\ \sin\varphi_2 \end{Bmatrix} \tag{4}$$

$\{\cos\varphi_1, \sin\varphi_1\}^{\mathrm{T}}$ and $\{\cos\varphi_2, \sin\varphi_2\}^{\mathrm{T}}$ is respectively the orientation cosine in the global coordinate system xAy of vector $\overrightarrow{DB}$ and the direction cosine in the local coordinate system $x^{'}Dy^{'}$ of vector $\overrightarrow{DC}$, among them $\{\cos\varphi_1, \sin\varphi_1\}^{\mathrm{T}}$ gets by next type:

$$\begin{Bmatrix} \cos\varphi_1 \\ \sin\varphi_1 \end{Bmatrix} = \frac{1}{l_4} \begin{Bmatrix} y_B - y_D \\ x_B - x_D \end{Bmatrix} \tag{5}$$

Among them, l_4 is the distance between B and D.

In the triangle ΔDBC, we can get according to law of cosines:

$$\cos\varphi_2 = \frac{l_3^2 + l_4^2 - l_2^2}{2 l_3 l_4} \tag{6}$$

It should be divided into two kind of situations for the value of $\sin\varphi_2$:

(1) when planar four-bar moves to as shown in figure 2 the position $\overrightarrow{DB} \times \overrightarrow{DC}$ is positive, it is:

$$\sin\varphi_2 = \sqrt{1 - \cos^2\varphi_2} \tag{7}$$

(2) when planar four-bar moves to as shown in figure 3 the position of $\overrightarrow{DB} \times \overrightarrow{DC}$ is negative, this time it is:

$$\sin\varphi_2 = -\sqrt{1 - \cos^2\varphi_2} \tag{8}$$

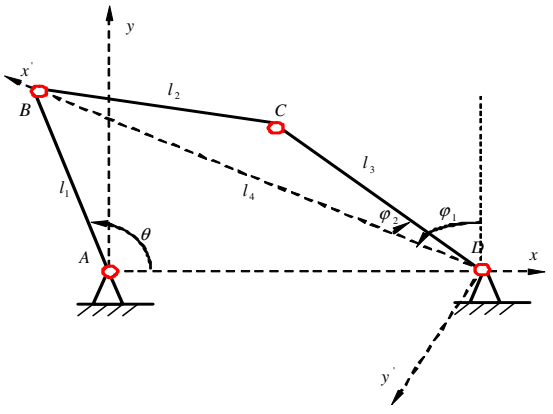


Fig. 2. Four-bar mechanism analysis and calculation chart (1)

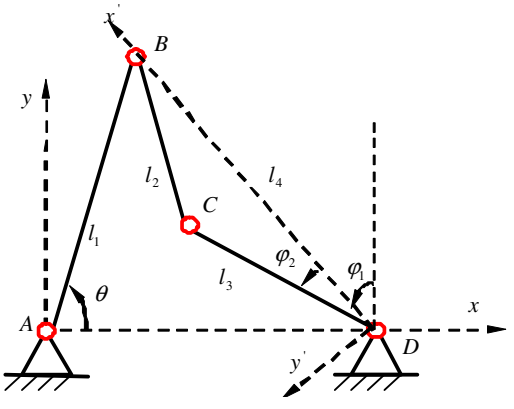


Fig. 3. Four-bar mechanism analysis and calculation chart (2)

3 Verify

Table 1 is initial conditions of a hinge four-bar. The corresponding picture is this hinge four-bar motion analysis with MATLAB draw after the point C coordinate change curves.

This illustration shows the MATLAB draws the above several institutions in the sport of the halfway point C coordinate change curves.

In order to validate MATLAB program is correct or not, we simulate movement by ADAMS. Then compare the movement curve painted by MATLAB and the curve simulated by movement, as results the two groups curve perfectly fit. So it perfectly fit the practical sports of hinge four-bar using the analysis method based on the linear transformation of hinge four-bar. It is a simple and accurate analysis method.

Table 1. Hinge four-bar movement analysis example

Coordinates of the Hinge four bar mechanism When in the initial position (unit/mm)					Active rod rotation direction	the angle range of Hinge four bar mechanism active rod rotation
	A	B	C	D		
x	0	-200	200	400	Clockwise	13° ~ 120°
y	0	300	250	0		

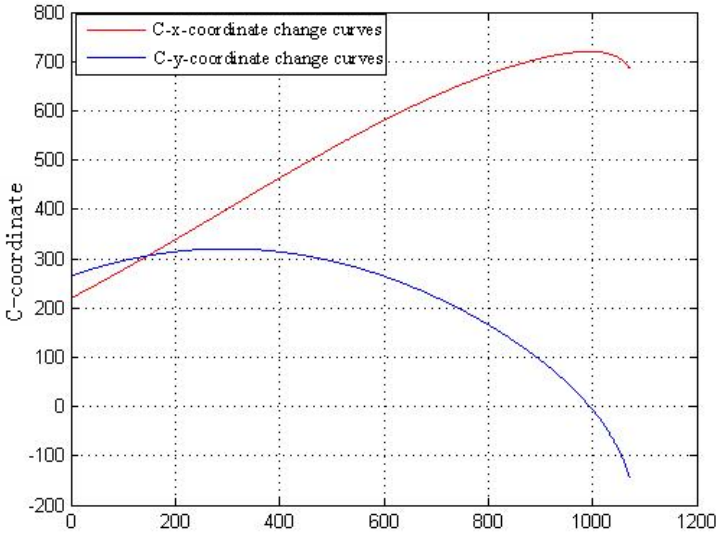


Fig. 4. C point horizontal y-coordinate change curves

References

1. MaRui: Automobile self-discharging mechanism design. The Master Degree from the University of Science and Technology, Beijing (2007)
2. Akin, J.E.: Computer-Assisted Mechanical Design. Prentice-Hall, New Jersey (1992)
3. Eliahu, Zahavi: The Finite Element Method in Machine Design. Prentice-Hall, NewJersey (1992)
4. Cao, C.: Our country a special car market analysis and prediction. Commercial car (November 1997)
5. Liu, M., Liu, J.: Several of the lifting mechanism structure and performance analysis. Qingdao Construction Engineering Institute (November 1999)
6. Jiang, C., Ming, H.: A special car design. Wuhan Industrial University Press (1994)

7. Zheng, D., et al.: Special automobile structure and maintenance. Shanghai Science and Technology Publishing House (1993)
8. Lin, Y.: Dump the lifting mechanism design research, Qingdao Construction Engineering Institute Master's Degree Thesis (1999)
9. hu flags. Dump the simulation of lifting mechanism design research, Qingdao construction engineering institute master's degree thesis (2000)
10. Wang, Y., Chang, S., Huang, W.: Automobile self-discharging mechanism stress analysis method. A Special Car (April1994)
11. China automobile corporation. Car technology parameters manual. Cars and Driving Maintenance Magazine Publishers (1993)

Study on Model Building for Fast Seismic Disaster Assessment Based on the Habitation

Dongping Li[1] and Yao Yuan[2]

[1] Earthquake Monitoring and Forecast Center
Seismological Bureau of Zhejiang, Hangzhou, China
llgis@163.com
[2] Regional Cooperation Department
Hangzhou Domestic Economic Cooperation Office, Hangzhou, China
honeyroger@hotmail.com

Abstract. Earthquakes in both Zhejiang Province and nearby districts, attempts to analyze firstly the historical damages, then study the distribution of an earthquake possibility and the characteristics of the loss according to the geological conditions, economy and population of this province, so that the method of approximate estimation can be established to improve the governmental abilities in dealing with the quake and provide information as the basis of official relief decisions. Firstly can produce an evaluation model according to the recorded earthquake, stastistic lines with same force in Zhejiang Province, and may thus attain the relation model between the long and short axes of the impact scope with different intensity. Then by comparing this province with nearby districts or those that have similar geological conditions on the occurrence of earthquakes, we can produce the model of the impact scope. This new method adopts the local total output value (GDP) and the population data to replace the traditional house building classification. By combining with the GIS to build a model, this new method may get an efficient and feasible result. Comparing to the traditional evaluation methods, the rapid evaluation of earthquake loss is a valid method which predicts the big dimension of earthquake disaster.

Keywords: Earthquake Emergency, Habitation, Emergency Command, GIS.

1 Introduction

With the development of Zhejiang province's social economy accesses to a steady stage, people's living standard enhances with steady step, all kinds of construction work unfold completely, the general public and the community demand higher standard of life and security. Therefore, the impacts of nature disasters are more and more notable. The ruinous and sensible earthquake is one of these disasters. Once the earthquake happens, the concerned departments starve for the information of earthquake area to direct disaster rescues. Thereby, each relative department should work swift, accurately and efficiently during the disaster rescues.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 537–543.
springerlink.com

The rapid evaluation of earthquake loss uses experiential models to evaluate the loss of earthquake in a short time. The information about the scale of earthquake may help the government and the relative department to make countermeasures, dispose disaster rescue action and strive for foreign aid. If we can estimate the distribution and degree of the earthquake in a short time, the government may make a scientific decision to rescue disasters, then more people who lived in the earthquake area may rescued, the loss of economy may reduced.

At present, the dominating method of the rapid earthquake hazard evaluation system based on the class list of building. It considers the economic loss of earthquake equals to the multiplication of the dangerous of earthquake, the building vulnerability and the loss rate of building and financial affair. Accordingly, the evaluations of the other kinds of loss are also based on the building's class list. All the evaluations are based on the concern databases of buildings. Nevertheless, the estimations of seismic hazard which based on the buildings' class listing have some deficiencies. Firstly, the collection of data is cockamamie. It needs to collect the exhaustive catalogs of buildings and establishments which are built in the earthquake areas. At present, the relative departments still can not fulfill the database. However, if they finish the whole database now, it is very hard to renew the data in the future. Secondly, it is not scientific to calculate the loss and breakage of establishment in the way of money. It is because the values of buildings and establishments will fluctuate when the value of money is changed. If we use this way to calculate the value of buildings and establishments, the value changes very quickly in the province whose economy develops very quickly, such as Zhejiang province. Thirdly, although the complicated computer programs and the perfect databases are used in the evaluation, the misestimating still exists. Therefore, the rapid evaluation of seismic hazard which is based on the macroeconomic index is attractive. Since it is easy to acquire the database of macroeconomic index, and the database renews in time.

2 The Rapid Estimation of Seismic Loss Based on the GDP Index

2.1 The Rapid Estimation of Seismic Loss Based on the GDP Index Is Proposed

The assumption of the rapid earthquake hazard evaluation system which based on the GDP index is: the hazard which is caused by the earthquake has direct correlation with the social finance of this area. The social finance usually can be expressed by the macroeconomic index, for example the GDP or GNP. It directly calculates the economic loss of earthquake, and then we do not need to not consider the detail of loss, and therefore it can not provide the countermeasure of disaster defense. For the macroeconomic index and vital statistics are renewed very quickly, it is a feasible and efficient way to evaluate the seismic hazard.

2.2 The Acquisition of the Data of the Rapid Evaluation of the Economic Loss of Seismic Hazard

The database which is used for the rapid evaluation of seismic hazard is based on the GDP and the vital statistics of villages and towns. The macroeconomic index and the

vital statistic database can be acquired easily from the statistic department. In the countries, especially the mountainous area, the density of population is small, we can use 1:50000 base geographic data to dispose it again. In order to use the vital statistic data adequately, we depend on the residential distribution chart and allot the information of census and the GDP data according to the proportion of the occupied acreage of residential buildings. At last, we can use the data and the divinable model of earthquake to make the rapid evaluation of seismic hazard. In the calculation, we use GIS to read the acreage of residential building in different villages and the proportion of residential building in each village. Then we use Attenuation Relations for Ground Motion which is commended by the Forth-Generation Zoning Map of Earthquake Motion in East of China to calculate the infection of earthquake, and judge the distribution of population and the Gross Indices of Macroeconomic Operations which may belong to corresponding area of seismic intensity.

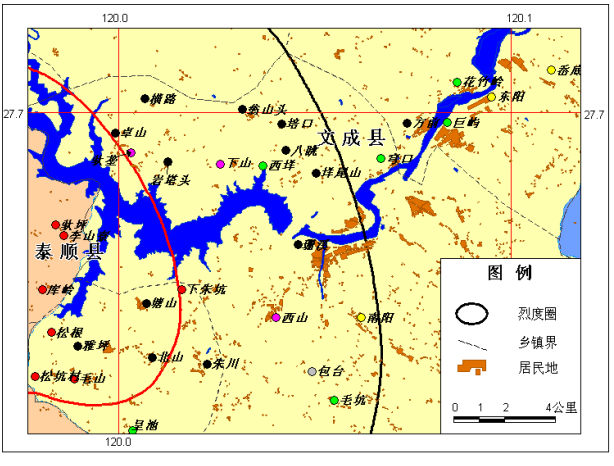


Chart 1. 1:50000 Residential Distribution, Seismic Intensity distribution Chart

2.3 The Model of Seismic Hazard Evaluation

1) The Calculation of Economic Loss

In order to objectively measure the differences of seismic macro vulnerability of class index in these areas which have different levels of economic and social development, we use population density, GDP per capita, GDP per unit area and the GDP per capital and per unit which are calculated through the constant price of 2000, and so on five indicators for references to analysis the loss rate and the death rate of seismic intensity zoning. Through these calculations, we can measure the possible catalogue value of different indicators. In order to get an objective and rational class index and a scientific outcome, we decided to use the GDP per capital which are calculated by the constant price of 2000 as the class index of seismic macro vulnerability. (We will call it GDP per capita 2000) According to the GDP per capita 2000, we set 2700$ and

Table 1. The seismic economic loss rate which is based on the macroeconomic index (GDP)

GDP per Capita（Constant Price of 2000）/Yuan	Intensity *I*				Remark
	VI	VII	VIII	IX	
Above 10000	0.0285	0.1647	0.7525	2.8739	$A=4\times10^{-11}$, B=11.377
2700-10000	0.2070	1.2345	5.7987	22.6950	$A=2\times10^{-10}$, B=11.585
Below 2700	0.8576	3.8897	14.4118	45.7542	$A=2\times10^{-8}$, B=9.8082

10000$ as the references of different societies. And then we can build the relationship of seismic macro vulnerability in different societies.

Then we do the analysis according to the intensity area allocation methodand the GDP per capita 2700 Yuan and 10000 Yuan which are calculated by the constant price of 2000. (Look at chart 2) Then the relationship of macroeconomic loss can be expressed as:

$$MDF = A \bullet I^{B} \tag{1}$$

MDF represents the loss rate of GDP, I represents the seismic vulnerability, A and B are coefficients. The value will be filled in Table 1.

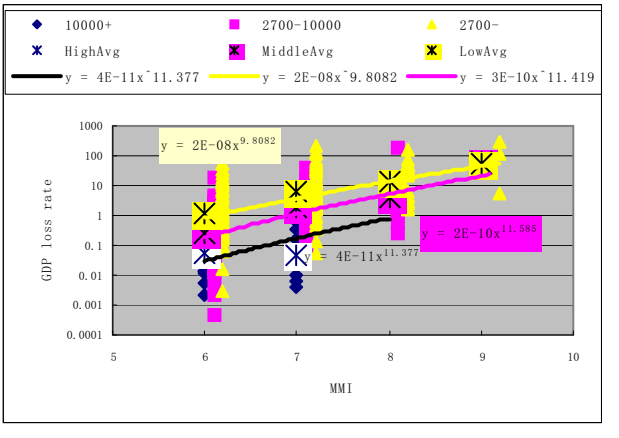


Chart 2. The relationship between the GDP loss rate and the vulnerability

2) The Calculation of Casualties

Casualties prediction include: calculate number of the casualties and number of the injured in the affected areas in a given earthquake condition.

$$M_d(I) = c\eta(A_1 r_{d1} + A_2 r_{d2} + A_3 r_{d3}) \tag{2}$$

$$M_h(I) = c\eta(A_1 r_{h1} + A_2 r_{h2} + A_3 r_{h3}) \tag{3}$$

In the formulas (2) is number of the casualties and (3) is number of the injured, C is for the personnel percentage indoors in the earthquake, A_1 is for the destroyed houses' area, A_2 is for severely damaged houses' area, A_3 is for medium damaged houses' area, η is for personnel density indoors. Units: people/m^2; Rd1, rh1 are respectively the mortality rate and the seriously injured rate in the destroyed house; r_{d2}, r_{h2} are the mortality rate and seriously injured rate in the severely damaged house; r_{d3}, r_{h3} are respectively the mortality rate and seriously injured rate in medium damaged houses. In the forecast, time can be divided into the day and the night, now suppose during the day in the earthquake indoors personnel percentage is for 40%, and during the night in the earthquake indoors personnel percentage is for 100%. The day time is from 8:00 to 18:00, and the night time from 18:00 to 8:00.

In Blind estimate system casualties calculating is similar to property losses calculating, using space superposition analysis method to calculate the casualties. First the overlying town's clustered houses' area and separate house's area can be obtained by the space superposition method, then put the result into the casualties calculation formula to get the number of the dead, and the injured and the homeless.

3 The Instance of Seismic Hazard Analysis

In order to test the applied effect of the method of seismic hazard evaluation that is based on the GIS and GDP, we will choose an earthquake which happened between Taisun county and Wencheng county in Zhejiang province on February 9th, 2008 to analysis. The parameters of earthquake will use the actual epicenter position and magnitude of earthquake. Then we will use the model mentioned above and consult the seismic hazard loss rate and the cost of building to do the calculation. This calculation can be separated into two parts:the first part uses the traditional method which based on the breakage of building to evaluate the hazard of earthquake; the second part uses the method which based on the distribution of GDP to do the rapid evaluation of seismic hazard.

Table 2. The comparison of two methods of seismic hazard evaluation

Method of Evaluation	Casualty of People (Per Person)		Economic Loss (Per Ten Thousand Yuan)	Corresponding
	Be Injured	Death		
Construction of Building	0	0	4521	46%
The Distribution of GDP	0	0	6652.5	67%
The Result of Investigation	0	0	9936.1	100%

Analysis the Table 2, we will find that: when we use the data of buildings' construction which acquired from the 5th census in 2000 and the traditional method of seismic evaluation, the outcome of calculation has large difference with the result of actual investigation. However, when we use the method of GDP distribution and residential area model which based on the 1:50000 base geographic data, we get the

85%value of actual economic loss. The new method is correspondingly better. The new method which uses the GDP distribution to evaluate the economic loss of seismic hazard is a feasible way. Therefore, we can get the conclusion that: the traditional way which uses the constructions of building to evaluate the economic loss of seismic hazard is not as good as the new method which uses the distribution of GDP.

4 Conclusion

Therefore, we can get the conclusion that: the traditional way which uses the constructions of buildings to evaluate the economic loss of seismic hazard is not as good as the new method which uses the distribution of GDP. This is because the data of the evaluation method is quite hard to acquire and uneasy to renew. However, the data of macroeconomic index is quite easy to acquire and they are renewed in time. The result of the new method of seismic evaluation which based on the GIS and GDP is scientific and efficient. In the future more researches can be made on the new methods of seismic hazard evaluation.

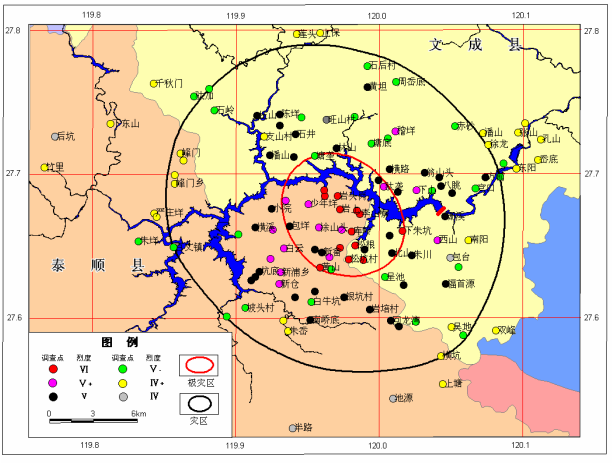


Chart 3. The distribution of intensity of M_L4.6 earthquake of Taishun

Acknowledgment. This study is supported by Science and Technology Project of ZheJiang (No.2011c23060). The authors would like to thank the assistant of Ms Mingzi Wang.

References

1. Chen, Z.-R., Li, Z., Liu, T.: Spatial information grid service model for mobile devices. Journal of Zhejiang University (Science Edition) 37(5), 577–582 (2010)
2. He, X., Gao, Y.-Y.: WSRF-Based Spatial Information Services State Management. Journal of Geomatics Science and Technology 27(5), 375–378 (2010)

3. Liu, X.-F., Meng, L.-K., Huang, C.-Q.: Z GIS-Based Reconstruction of Basin Paleotectonics: An Example from Paleo-Central Uplift Belt, Northern Songliao Basin. Earth Science-Journal of China University of Geosciences 28(3), 346–350 (2003)
4. Zhao, J.-H., Li, D.-P.: The Variation of Caco_3 Content in Four Profiles from the North of China and the Reconstructed Paleo-Precipitations. Marine Geology & Quaternary Geology 24(3), 117–122 (2004)

Detection and Location of Transient Power Quality Disturbances in White Noise Using Wavelet Techniques

Ronghui Liu, Erbin Yang, and Xiu Yang

School of Power and Automation Engineering, Shanghai University of Electric Power, Shanghai 200090, China
{Liuronghuiyzy,yangeb507}@126.com, yangxiu72@263.net

Abstract. In this paper, the principles of wavelet transform, discrete stationary wavelet transform and wavelet packet transform were presented. Transient power quality analysis method based on wavelet techniques was then porposed. Finally, detection of noisy power quality disturbances such as voltage sag, voltage swell, voltage interruption, transient harmonics, impulsive transient and oscillatory transient were simulated with MATLAB emulator. The perfect results of simulation showed that the presented method could denoise and detect the transient disturbances accurately, which provided support for improving power quality.

Keywords: Transient power quality, denoising; translation invariant, wavelet techniques, disturbances localization.

1 Introduction

Due to the widespread use of power electronic devices and nonlinear loads, voltage and current waveform distortions have caused the significant deterioration in power quality [1]. Both power suppliers and electricity users have focused on transient power quality problems. In order to improve the power quality, it is necessary to detect and classify transient disturbances accurately. Traditional power quality analysis method (Fourier transform) is based on the effective value theory, which is not suitable for non-stationary signals processing. Wavelet transform has good time-frequency localization property; thus it is suitable for time-varying signals analysis [2].

The primary characteristics of transient disturbances are the magnitude and duration of the disturbances voltage. In transient power quality disturbances analysis, it is very important to remove noise, localize disturbances time and classify the disturbances fast and accurately.

Transient power quality disturbance signals often submerged in noise signals; hence the first step is to remove noise. Conventional soft and hard thresholding wavelet de-noising methods proposed by Donoho et al. are all accomplished by discrete wave transform (DWT). Since DWT is not translation invariant, the artificial interferences called Pseudo-Gibbs phenomena exist in the neighborhood of discontinuities. In order to suppress such artifacts, for a range of shifts, cycle spinning is to shift the data, de-noise the shifted data, and then un-shift the de-noised data.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 545–553.
springerlink.com

After de-noising, the second step is to detect, localize and classify transient power quality using wavelet transform and wavelet packet transform [3,4].

In this paper, after introduction of the principles of wavelet transform, discrete stationary wavelet transform and wavelet packet transform, transient power quality analysis method based on wavelet techniques is porposed. The noisy transient disturbances denoised by translation invariant de-nosing, detection and location of the de-noised power quality disturbances such as voltage sag, voltage swell, voltage interruption, transient harmonics, impulsive transient and oscillatory transient are finally simulated to verify the effectiveness of the presented method with Matlab wavelet toolbox.

2 Wavelet Techniques

As a powerful tool for signal analysis, wavelet analysis is widely used in signal processing.

Continuous Wavelet Transform (CWT)

The continuous wavelet transform of any finite energy signal $x(t) \in L^2(R)$ located at position b and of scale a. is the projection of $x(t)$ on the corresponding wavelet atom:

$$W_x(a,b) = < x, \psi_{a,b} > = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^*(\frac{t-b}{a}) \mathrm{d}t \tag{1}$$

It is given by the inner product of $x(t)$ and the specific mother wavelet $\psi_{a,b}(t)$. The scale parameters *a* and translation parameter *b* are continuous variables; the transform is therefore called continuous wavelet transform.

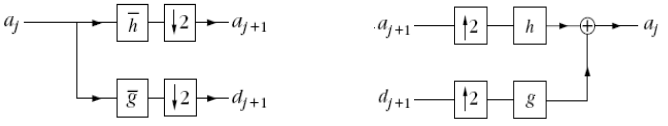
Discrete Wavelet Transform (DWT)

Discrete wavelet transform (DWT) is widely used in practical applications. DWT transforms a time domain signal into time-frequency domain simultaneously.

For any discrete signal input $x(n)$, the wavelet decomposition coefficients are computed by successive convolutions with the low-pass filter $\overline{h}$ and the high-pass filter $\overline{g}$ followed by a factor 2 downsampling. Starting from the original signal $x(n) = a_0$, the approximation coefficient a_1 and detail coefficient d_1 are obtained by convolving $x(n)$ with the low-pass filter $\overline{h}$ for approximation, and with the high-pass filter $\overline{g}$ for detail, followed by dyadic decimation. The approximation coefficient a_1 is then split into two parts using the same scheme, replacing $x(n)$ by a_1 and producing a_2 and d_2, and so on.

The wavelet decomposition and reconstruction of a given signal at level *j* are illustrated by Fig. 1. The symbol $\downarrow 2$ is used to indicate the signal down sampled by 2 whereas the symbol $\uparrow 2$ represents the signal up sampled by 2. The low-pass filter

$\bar{h}$ removes the higher frequencies of the inner product sequence a_j whereas the high-pass filter $\bar{g}$ collects the remaining highest frequencies. The frequency divisions of wavelet decomposition are not uniform, and the higher frequency bands cover more frequency components than the lower frequency bands do. WT is an efficient mathematical tool for singularity detection

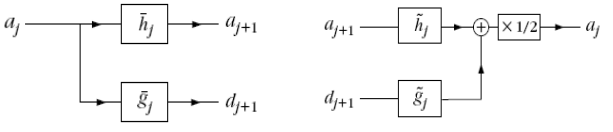


(a) decomposition (b) reconstruction

Fig. 1. Diagram of the DWT decomposition and reconstruction

Discrete Stationary Wavelet Transform (SWT)

Since the DWT is not a time-invariant transform, even with periodic signal extension, the DWT of a translated version of a signal is not the translated version of the DWT of the signal. The stationary wavelet transform (SWT) averages some slightly different DWT. This property is useful for translation invariant de-noising.



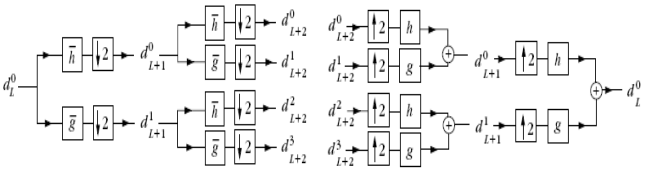
(a) decomposition (b) reconstruction

Fig. 2. Diagram of the SWT decomposition and reconstruction

SWT is very simple and is close to DWT. The approximation a_1 and detail coefficient d_1 of a given signal can be obtained by convolving the signal with the appropriate dilated filters as in the DWT case but without downsampling. The SWT decomposition and reconstruction of the given signal at level j are shown in Fig. 2. The original signal is reconstructed through convolutions with $\tilde{h}$ and $\tilde{g}$. A multiplication by 1/2 is necessary to recover the next finer scale signal a_j [5].

Wavelet Packet Transform(WPT)

The low-frequency part and the high-frequency part are subdivided in wavelet packet transform whereas only the low-frequency part is subdivided in wavelet transform; hence, wavelet packet transform provides uniform frequency decomposition of a given signal.



(a) decomposition (b) reconstruction

Fig. 3. Diagram of the WPT decomposition and reconstruction

The wavelet packet decomposition and reconstruction of a given signal at level L are illustrated by Fig. 3. The coefficients d_{L+1}^0 and d_{L+1}^1 are obtained by down sampling the convolutions of d_L^0 with a dual low-pass filter $\overline{h}$ and a dual high-pass filter $\overline{g}$. The coefficients d_{L+1}^0 and d_{L+1}^1 are then split into two parts using the same scheme, replacing d_L^0 by d_{L+1}^0 and d_{L+1}^1, and producing d_{L+2}^0, d_{L+2}^1, d_{L+2}^2 and d_{L+2}^3. Wavelet packet filter bank decomposition of the given signal is with successive filterings and downsamplings, while reconstruction is performed by inserting zeros and filtering the outputs.

3 Simulation of Transient Power Quality Disturbance Signals Detection

Transient power quality disturbances signals are easy to be contaminated by random electrical noise that can be described as white noise. The original signal $x(t)$ includes the disturbance $f(t)$ and white noise $w(t)$:

$$x(t) = f(t) + w(t) \tag{2}$$

In this section, denoising and detection of noisy transient power quality disturbances signals were simulated with Matlab's Wavelet Toolbox [6]. Six different conditions were discussed to illustrate transient disturbances analysis method based on wavelet techniques. Transient disturbances with white noise were denoised by stationary wavelet transform; and Daubechies 1(db1) was selected as mother wavelet. Transient disturbances were detected, localized and classified by wavelet transform and wavelet packet transform.

Voltage Sag

The mathematical expression of voltage sag can be expressed as

$$f(t) = \sin(2\pi f_0 t) - \alpha \sin(2\pi f_0 t)[u(t-t_1) - u(t-t_2)] \tag{3}$$

where f_0 is power frequency, the value of α is from 0.1 to 0.9; the duration being from 0.5 times power frequency cycle to one minute.

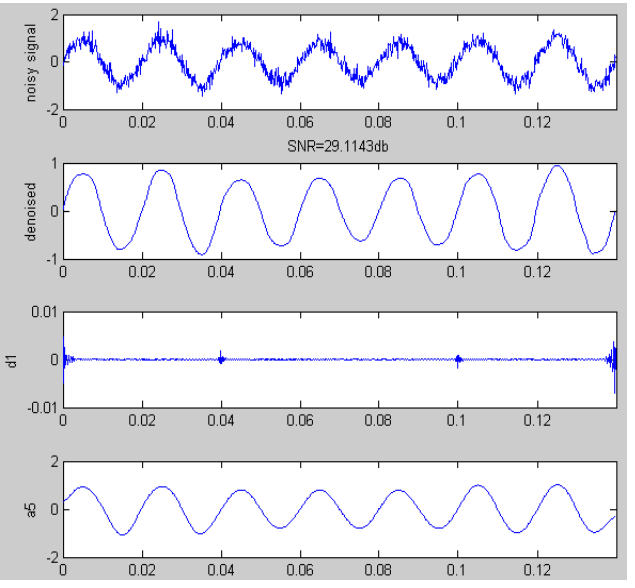


Fig. 4. Detection and location of noisy voltage sag disturbance signal

In this case, $t_1 = 0.04\text{s}$, $t_2 = 0.1\text{s}$, $\alpha = 0.2$. Detection and location of noisy voltage sag disturbance signal is shown in Fig. 4. The first waveform is the noisy signal; and the second one is the denoised signal. According to the third waveform, the starting time and the ending time are 0.04s and 0.1s, respectively. The last waveform shows the reconstructed waveform of power frequency waveform from which transient disturbance is classified as voltage sag.

Voltage Swell

The expression of voltage swell is given as

$$f(t) = \sin(2\pi f_0 t) + \alpha \sin(2\pi f_0 t)[u(t - t_1) - u(t - t_2)] \tag{4}$$

where the value of α is from 0.1 to 0.8; the duration is from 0.5 times power frequency cycle to one minute.

In this case, $t_1 = 0.04\text{s}$, $t_2 = 0.1\text{s}$, $\alpha = 0.4$. Detection and location of noisy voltage swell disturbance signal is illustrated by Fig. 5. Voltage swell disturbance starts at time 0.04s and ends at time 0.1s, as shown in Fig. 5. According to the reconstructed waveform of power frequency waveform, transient disturbance is classified as voltage swell.

Voltage Interruption

The voltage interruption disturbance signal is expressed as

$$f(t) = \sin(2\pi f_0 t) - \alpha \sin(2\pi f_0 t)[u(t-t_1) - u(t-t_2)] \tag{5}$$

where the value of α is from 0.9 to 1; the duration is less than one minute.

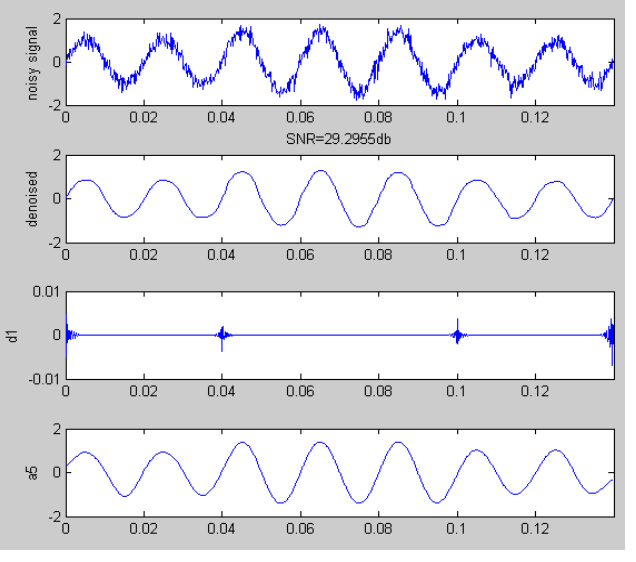


Fig. 5. Detection and location of noisy voltage swell disturbance signal

In this case, $t_1 = 0.04\text{s}$, $t_2 = 0.1\text{s}$, $\alpha = 1$. Detection and location of noisy voltage interruption disturbance signal is illustrated by Fig. 6. Voltage interruption disturbance begins at time 0.04s and ends at time 0.1s, as shown in Fig. 6. According to the reconstructed waveform of power frequency waveform, the transient disturbance is classified as voltage interruption.

Transient Harmonics

The expression of transient harmonics disturbance is expressed as

$$f(t) = \sin(2\pi f_0 t) + \sum_{k=1}^{N} \alpha_k \sin(2\pi k f_0 t)[u(t-t_1) - u(t-t_2)] \tag{6}$$

In this case, $t_1 = 0.04\text{s}$, $t_2 = 0.1\text{s}$, transient 3rd and 5th harmonics are added to power frequency voltage. Detection and location of noisy transient harmonics disturbance signal is shown in Fig. 7. The starting and ending point of transient harmonics disturbance are time 0.04s and 0.1s, respectively. The reconstruction waveforms of the fundamental, 3rd and 5th harmonic components are reconstructed by wavelet packet transform, as shown in the last three waveforms of Fig. 7.

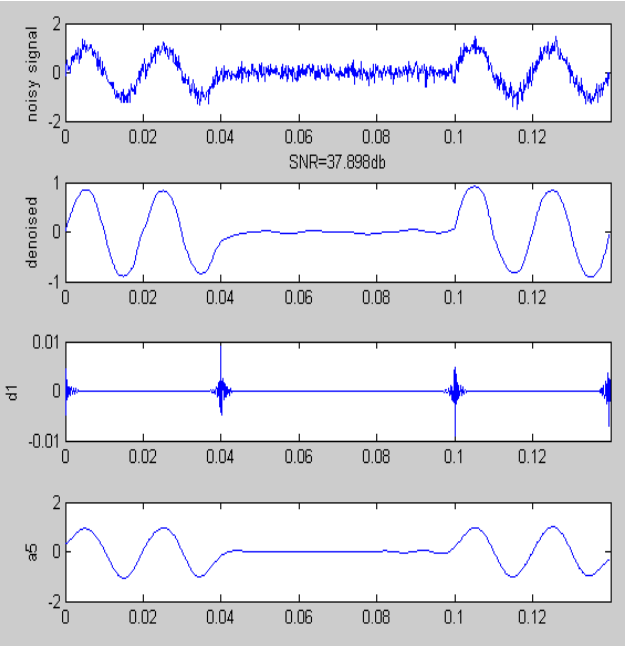


Fig. 6. Detection and location of noisy voltage interruption disturbance signal

Impulsive Transient

The expression of impulsive transient disturbance is given as

$$f(t) = \sin(2\pi f_0 t) - \alpha \sin(2\pi f_0 t)[u(t-t_1) - u(t-t_2)] \tag{7}$$

where the duration is very short.

In this case, $t_1 = 0.042\text{s}$, $t_2 = 0.044\text{s}$, $\alpha = 0.8$. Detection and location of noisy impulsive transient disturbance signal is shown in Fig. 8. The starting and ending time of impulsive transient disturbance are 0.042s and 0.044s, respectively.

Oscillatory Transient

The oscillatory transient disturbance signal is given as

$$f(t) = \sin(2\pi f_0 t) + \alpha e^{-p(t-t_1)} \sin(2\pi k f_0 t) u(t-t_1) \tag{8}$$

In this case, $t_1 = 0.04\text{s}$, $\alpha = 0.4$, $p = 150$, $k = 11$. Detection and location of noisy oscillatory transient disturbance signal is illustrated in Fig. 9. Transient oscillation begins at time 0.04s. Reconstruction waveform of transient oscillation is reconstructed by wavelet packet transform, as shown in the last graph of Fig. 9.

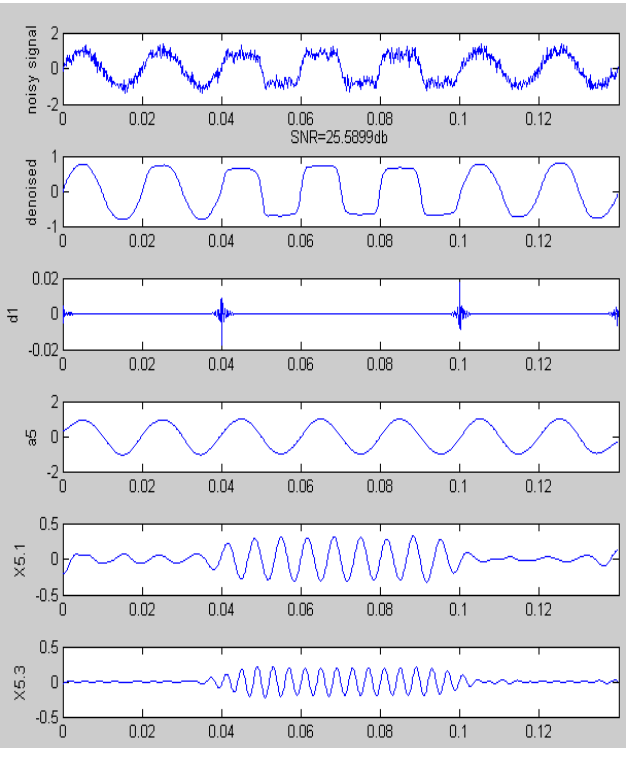


Fig. 7. Detection and location of noisy transient harmonics disturbance signal

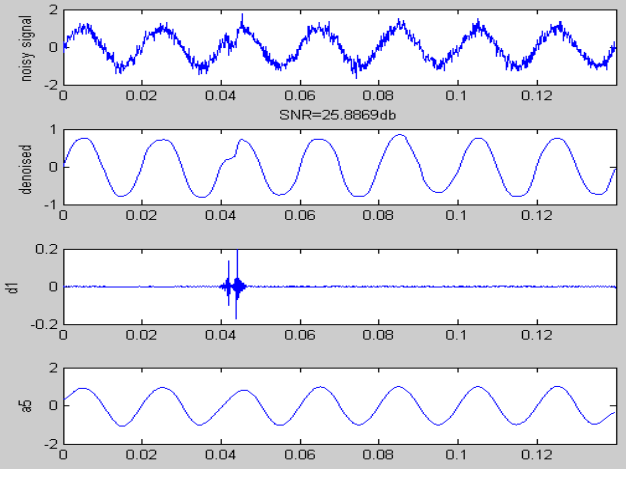


Fig. 8. Detection and location of noisy impulsive transient disturbance signal

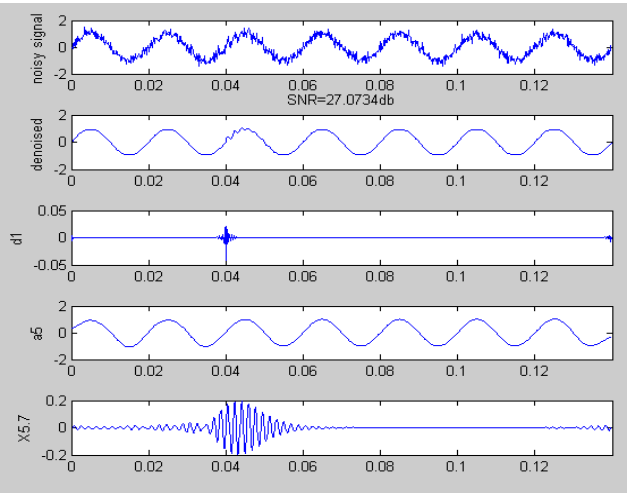


Fig. 9. Detection and location of noisy oscillatory transient disturbance signal

4 Conclusion

This paper presents transient power quality analysis method using wavelet techniques. The perfect results of simulation show that transient disturbances signals can be detected effectively and accurately, which provides a reliable basis for improving power quality.

Acknowledgment. This work was supported by leading academic discipline project of Shanghai Municipal Education Commission (J51301) and Shanghai Municipal Natural Science Fund (09ZR1412900).

References

1. George, J.W.: Power Systems: Harmonic Fundamental Analysis and Filter Design, pp. 36–58. Mechanical industry press, Beijing (2003)
2. Mallat, S., Hwang, W.L.: Singularity detection and processing with wavelets. IEEE Transactions on Information Theory 38(2), 617–643 (1992)
3. Donoho, D.L.: De-noising by soft-thresholding. IEEE Transactions on Information Theory 41(3), 613–627 (1995)
4. Donoho, D.L., Johnstone, I.M.: Adapting to unknown smoothness via wavelet shrinkage. Journal of American Statistical Association 12, 1200–1224 (1995)
5. Mallat, S.: A Wavelet Tour of Signal Processing, pp. 168–189. Academic press, Burlington (2009)
6. Hu, C., Li, G., Zhou, T.: System Analysis and Design based on MATLAB 7.x, pp. 368–380. Xidian university press, Xian (2008)

Application of Pocket PC in Marine Fishery Survey

Yangdong Li[1,2,3,*], Zhen Han[1,2,3], Guoping Zhu[1,2,3], and Siquan Tian[1,2,3]

[1] College of Marine Science & Technology, Shanghai Ocean University, Shanghai 201306, China
[2] The Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources (Shanghai Ocean University), Ministry of Education, Shanghai 201306, China
[3] The Key Laboratory of Shanghai Education Commission for Oceanic Fisheries Resources Exploitation, Shanghai Ocean University, Shanghai 201306, China
lyd911@163.com

Abstract. At present, during marine fishery survey, the field work heavily depends on the traditional recording fashion, i.e. recording data on paper sheets by hand. Obviously, it is very inefficient and easily lead to bad data quality. To address this problem, a solution of marine fishery survey which is based on Pocket PC is put forward. This solution, making use of mobile devices, records the field data in electronic form directly. It can assure the data consistency, improve the efficiency of field staff and make them more productive. The electronic collected data can transfer to desktop computers or database servers directly over the wired or wireless network outdoors, or by direct connection indoors, and so the indoor workload is reduced. Finally, a prototype system is designed and developed for squid survey data collection. Its test result shows that the system can make up the shortages of the traditional work mode obviously, and improve the efficiency of field work. It also proves that the solution of marine fishery survey based on Pocket PC is feasible.

Keywords: Marine Fishery Survey, Pocket PC, Data Collection, Windows Mobile.

1 Introduction

The researches on the fishery ecology, the development and evaluation of fishing ground, and deep-sea fishing, are closely related with field fishery investigation (K.Mu, 2003; Y.Tang et al, 2009). The investigation content includes fish forecasting, fish biological characteristics and marine environment, so a large amount of investigating data will appears. Such data includes fish biological characteristic data (e.g., the body size and weight of fish, etc.), marine environmental data (e.g., temperature of sea surface, salinity, etc.) and fishing data (such as fishing position, catch and so on). Much of that is geographical. Some of data is collected in the field, and others are produced by experimental analysis. For those on-site survey data, it is necessary to record them timely. At present they are mainly recorded by means of

* Corresponding author.

W. Deng (Ed.): Future Control and Automation, LNEE 173, pp. 555–560.
springerlink.com

paper hand-books, and minority of them is inputted into laptop computers directly. In the previous two modes, the former is more traditional and is adopted dominantly. But there are some obvious disadvantages in the paper form mode, such as operating efficiency is low, the data recorded is difficult to use and save, and in the room the data recorded on paper forms needs to be transferred to computer by manual, increasing the labour intensity of workers and leading to poor data quality (Kanstrup, 2009). The latter approach also has two disadvantages. On the one hand, the laptop computer battery power is not durable; On the other hand, laptops are more bulky and inconvenient to carry. Therefore, it is essential to explore one feasible solution to carry out highly efficient, good-quality and rapid survey data collection in the marine fisheries survey.

Pocket PC platforms (such as Personal Digital Assistants (PDAs), smart phones with PDA functions) have the following characteristics, such as some features of desktop computers, small size, low power consumption, powerful functions, easy to develop and so on (H.Chen & X.Mei, 2008). Pen and paper are being replaced by them, to help people in some daily management, mainly for calendar, address book, tasks manager and notes. And with the development of science and technology, PDAs are integrating some functions such as, computing, communications, network, storage, entertainment, e-business, and so on. And they have become an indispensable tool in our mobile life. Especially, at present, almost all of Pocket PC devices integrate a GPS positioning module, you can get to location information timely (C. Ciavarella and F. Paternò, 2003; X. Zhang & Y. Li, 2006). Some Pocket PC-based platforms for mobile data collection have been widely used in digital land (Y. Zhu et al, 2006), agriculture, forestry investigation (X. Zhang & Y. Li, 2006) and socio-economic fields (X.Wang et al, 2009), and so on, but there has not found any reports of applying the Pocket PC platform to investigations of marine fisheries.

This paper contributes to research on how the Pocket PC platform used in the investigation of marine fisheries. The paper is organized as follows: In the following section, a Pocket PC-based solution for marine fishery survey was presented. Next, a Squid Survey Data Collecting System was designed. Then the system was developed.

2 Pocket PC-Based Solution for Marine Fishery Survey

Fig.1 shows the Pocked PC-based solution for marine fishery survey. First, the field personnel utilizes handheld device (e.g. Pocket PC), which runs a marine fishery survey data collection system (MFSDCS), to download the necessary data from the server side, marine fishery survey server, to the device side, marine fishery survey data collecting system, in the wired connection (e.g. USB) or wireless connection (such as infrared, Bluetooth, WiFi, wireless communications, etc.) environment. Then, the field user collects fishery survey data by the device under the offline environment. The collected data is cached at the device side first. When the data collection is fully or partially completed, the cached data can be transferred to the server side, through utilizing the wireless communication ability of mobile device to connect to the server through the WWW network under the WAP protocols, or connect to the physical network of desktop computer (desktop or laptop) through the USB cable or Bluetooth module first, then transfer the cached data to the server side through the Intranet or the

Internet, or update the data cached at the device side to the server side through WLAN, using the WiFi ability of the mobile device. The device database can adopt Oracle Lite database, Microsoft SQL Mobile database, the Adaptive Server Anywhere of Sysbase, Linter, or some open source database, e.g. SQLite, Berkeley DB, McObject's Perst, etc. These embedded databases or mobile databases almost support multi-platform, a variety of development languages oriented, and multiple interfaces and so on. The server-side database can use one of the common database management systems like Oracle, SQL Server, MySQL and so on.

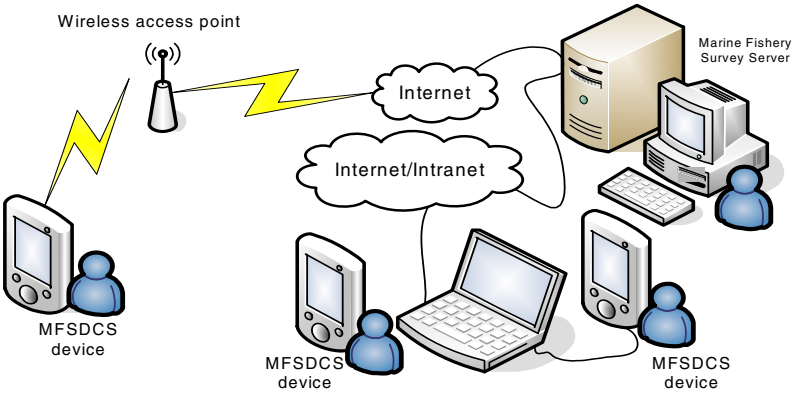


Fig. 1. Pocked PC-based solution for marine fishery survey

3 Design and Development of Pocket PC-Based Marine Fishery Survey Data Collection System

3.1 Design of Marine Fishery Survey Data Collection System

As an example, a squid survey data collection system has been designed. The target platforms of this system are the Pocket PC devices on which Windows Mobile operating system runs. The main functions of the system include fishing data

Table 1. Fishing data table

Field code	Field caption	Data type
Fishing_date	Date	datetime
Company	Fishery Company	varchar(40)
Longitude	Longitude	double
Latitude	Latitude	double
N_fishing_vessel	Number of Fishing vessels	integer
Catch	Catch	decimal(4,2)
Fishing_type	Fishing type	varchar(20)

Table 2. Biological data table

Field code	Field caption	Data type
Fishing_date	Date	datetime
Vessel_name	Vessel name	varchar(12)
Call_sign	Call sign	varchar(4)
Longitude	Longitude	double
Latitude	Latitude	double
Mantle_size	Mantle size	decimal(4,1)
Body_weight	Body weight	decimal(4,1)
…	…	…

collecting, biological data collecting, ocean current and weather data collecting and temperature, salinity and depth data of seawater collecting. The system database adopts Windows SQL Mobile. To manage those data, four tables have been designed, i.e., fishing data table (see table 1), biological data table (see table 2), current and weather table, and temperature, salinity and depth data table.

3.2 Implementation of Marine Fishery Survey Data Collection System

According to the requirements and system design of fishery survey, a marine fishery survey data collection system has been developed to satisfy squid fishery survey (see fig.1 and fig.2). The development environment was Microsoft Visual Studio 2005

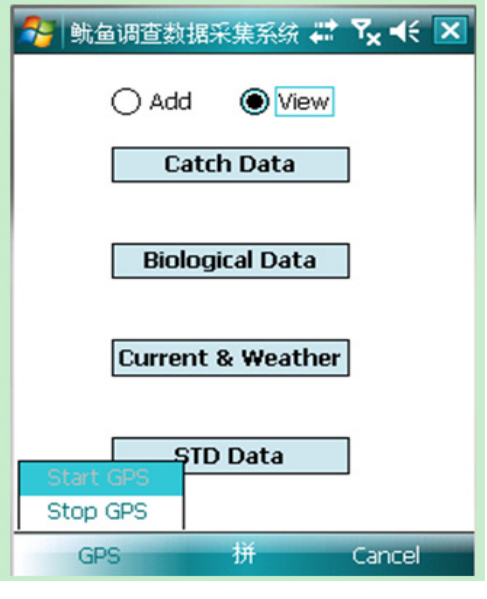


Fig. 2. Main interface of squid survey data collection system

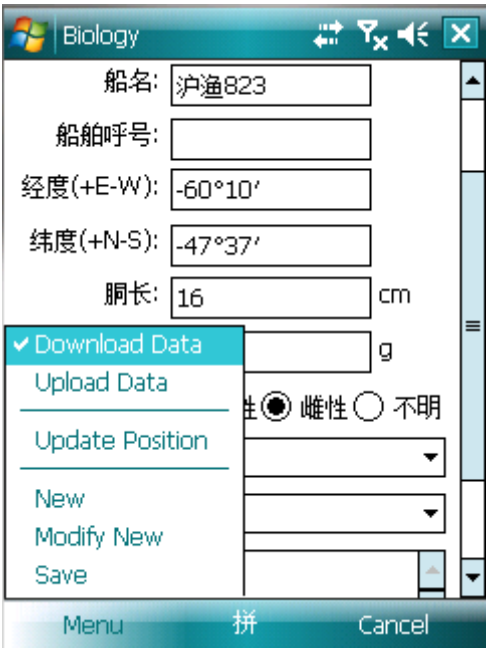


Fig. 3. Biological data collection interface

Integrated Development Environment which ran on the Microsoft Windows XP SP2 Professional Edition operating system. The database environment run on device-side and server-side was Microsoft SQL Mobile 2005. The system can run on the Pocket PC mobile devices whose operating system is Windows Mobile. The system has been tested on an Pocket PC device, Dopod T2222, with Windows Mobile 6.1. The test result shows that the system runs well and all developed functions fully meet the requirements of the system design.

4 Conclusion

Pocket PC devices are bringing a very positive impact to people's work and life. It's a very promising to take advantages of Pocket PC devices to serve marine fishery investigation. In this paper, a Pocket PC-based solution for the marine fishery survey was proposed and that can solve the disadvantages of traditional data recording mode. Furthermore, a data collection system for squid survey was designed and developed. Using a Pocket PC mobile device, which was installed the system, for marine fishery survey data collecting, it can guarantee the integrity of records, obtain location information automatically, improve operational efficiency, and reduce labour intensity.

Acknowledgment. The work described in this paper was substantially supported by the Specialized Research Fund for Selecting and Fostering Outstanding Young

Teachers in Higher Education Institutions of Shanghai (Project No. B-8101-09-0236), Open Fund for the Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources (Project No. B-8211-09-0003-7), the Program of Science and Technology Commission of Shanghai Municipality (Project No. 08230510700), and the grants from the Doctoral Starting Foundation of Shanghai Ocean University (Project No. B-8202-08-0288).

References

1. Chen, H., Mei, X.: The Development of Mobile GIS on PDA. Geomatics & Spatial Information Technology 31(6), 75–76 (2008)
2. Ciavarella, C., Paternò, F.: Design Criteria for Location-Aware, Indoor, PDA Applications. In: Mobile HCI 2003, pp. 131–144 (2003)
3. Kanstrup, A.M., Stage, J.: From Paper to PDA: Design and Evaluation of a Clinical Ward Instruction on a Mobile Device. In: Processing of International Federation for Information 2009, pp. 670–683 (2009)
4. Mu, K.: Resources Change and Fishery Management of Engraulis Japonicus in Area of Eastern Yellow Sea. Marine Sciences 27(9), 30–31 (2003)
5. Tang, Y., Zou, W., Hu, Z.: An Analysis of Utilization Status and Management of Marine Fisheries Resources in China based on Statistics Data. Resources Science 31(6), 1061–1068 (2009)
6. Wang, X., Tang, J., Ma, W., et al.: Design and Realization of Information System of Supervision and Management for Harbour Governance Based on PDA. Computer Applications and Software 26(5), 101–103, 125 (2009)
7. Zhang, X., Li, Y.: Application of PDA and GPS in Investigation of Forest Resources. Forest Investigation Design (3), 82–82 (2006)
8. Zhu, Y., Zhang, S., Xiao, H.: Research on Land Change Survey System on Palm. Geo-Information Science 8(3), 24–28 (2006)

Author Index