

## Optional Insurance Compensation Rate Selection and Evaluation in Financial Institutions

Xu Zhikun<sup>1</sup>, Wang Yanwen<sup>2</sup> and Liu Zhaohui<sup>3</sup>

<sup>1, 2</sup>College of Finance, Hebei Normal University of Science & Technology, Qinhuangdao, Hebei, P.R.China , 066004

<sup>3</sup>Pupillary workroom Hebei Normal University of Science & Technology, Qinhuangdao, Hebei, P.R.China, 066004

<sup>1</sup>xuvikp@126.com, <sup>2</sup>125721815@qq.com, <sup>3</sup>liying7688@126.com

### Abstract

With the change of market environment, most enterprises have realized the importance of customer resource. In this paper, we use data mining technology to analyze the influence factor of insurance settlement. First, we establish a customer segmentation structure model, and then use decision tree to find out the main factors that affect the amount of settlement and analyze the customer records. According to the result, we can find that the amount of settlement is mainly influenced by the driving experience and the purpose of using. Based on the analysis, the using of the data-mining tech would improve the work efficiency and the consumer management in insurance institutions, and insurance companies can also formulate the corresponding compensation rate.

**Keywords:** Artificial Neural Network, Data Mining method, Software Engineering, Clementine, Financial Institutions

### 1. Introduction

At the beginning of the 1990's, a new business model appeared in the United States and called Customer Relationship Management, CRM is a new management mechanism that aims to improve the relationship between enterprises and customers [1]. CRM provides comprehensive, personalized customer information to sales and service personnel, it also strengthens the tracking service, information analysis capabilities, enabling them to build and maintain the "one-to-one relationship" between customers and enterprise [2], so that enterprises can provide more efficient service, improve customer satisfaction, attract and retain more customers. Customer segmentation as one of the core concept of customer relationship management, is refers to the enterprise which has explicit strategies, business models and specific market, providing products, services according to customer attributes, behavior, preferences, needs and values and other related factors [3]. With the extensive application of management information system, enterprise will accumulate more and more data, the traditional customer segmentation method can not deal with it [4]. Data Mining is a data analysis technology based on artificial intelligence, its main function is to discovery useful information from large amounts of data, which will help enterprises to manage customer resources effectively.

Gannet Group first put forward the concept of CRM (customer relationship management), it consider the customer relationship management is to provide a full range management perspective, give more customer communication skills to enterprises, and finally maximize customer return rate [5]. According to IDC's (International Data Corporation) research, the

mean ROI( return on investment)of company which use CRM can reach 400%,more than 90% enterprises ROI exceeded 40%, 50% enterprises ROI exceeded 160%, 25% enterprises ROI over 600%.

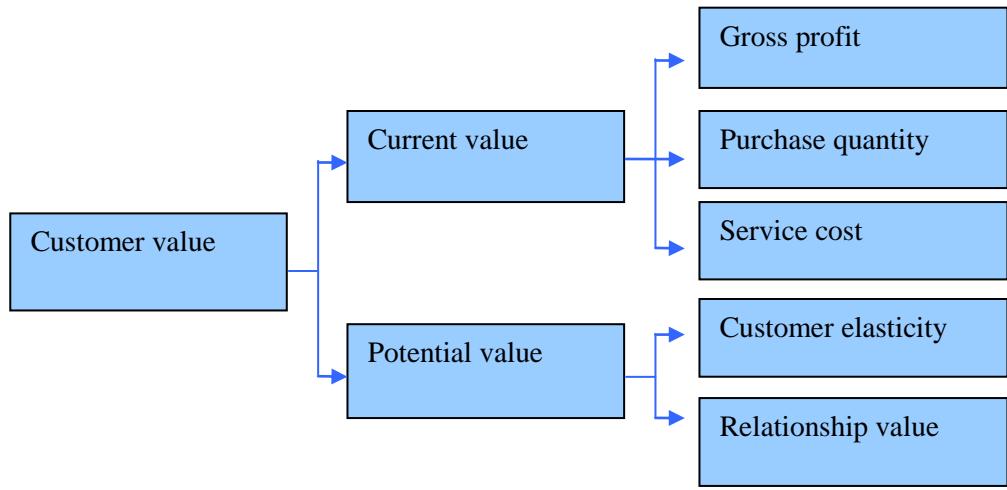
Data mining in customer relationship management applications can contribute significantly to the bottom line, as the development of information technology, customer information increase fast in recent years. So that, data mining is a useful method to discover extract information in large data sets. As customer segmentation is the core function in CRM, data mining can also be used to automatically discover the segments or groups within a customer data set [6]. Usama Fayyad& Gregory P. Shapiro (1996) pointed out that data mining and knowledge discovery in databases have been attracting a significant amount of research and industry, involves six common classes of tasks as: anomaly detection, dependency modeling, clustering, classification, regression and summarization[7].There are also many researchers focused on evaluating the data mining applications in financial industry. Georgios Sarantopoulos (2003) presented a real-world application of a data mining approach to credit scoring, and pointed out that the decision tree showed a great improvement in performance compared to the current manual decisions [8]. Wen-bing Xiao & Qian Zhao (2006) investigated the performance of various credit scoring models and the corresponding credit risk cost for three real-life credit scoring data sets [9]. Xiaohua Hu (2005) presented a data mining approach for analyzing retailing bank customer attrition, and demonstrated effectiveness and efficiency by applying data mining technology in retailing bank [10].

In these studies, however, did not fully play the basis of customer segmentation function, some is only a simple object with the customers, also there are few articles combine customer segmentation and data mining technology. So that, I use data mining technology in this paper to help insurance company improve CRM by making effective market segmentation and optimal price. CRM maintains and manages customer relationship through the four stages as: identification of customer, analysis of the difference, benign contact and the customized service. Customer relationship management requires "customer satisfaction first" as the core philosophy, so customer service is the core business in CRM.

## 2. Customer Value Evaluation

In this paper, we establish a function structure model to analyze the customer segmentation; this analysis is based on data mining method. The model includes three parts as data, function and method, so this model is also called DFM model. The methods we used in this research includes artificial neural network, enhanced ID3 algorithm and customer segmentation structure model.

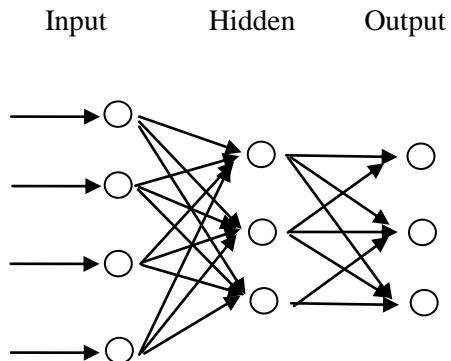
At present, the academic research of customer value mainly concentrated in two methods: One is research of customer lifetime value; and another is customer value evaluation system. In the research of customer lifetime value, we need to identify the customer's life cycle and predict the corresponding currency rate, so that it is difficult to get the accurate value of customer. This also makes more scholars tend to the second method. In this paper, we construct the index system that combined with customer life cycle theory, such as Figure 1. Customer current value determines the enterprise's profit level; and the potential value can predict the changes of customer value in the future. After building the customer value evaluation system, we choose BP algorithm to solve the model. BP algorithm can evaluate the customer value and make customer classification, also it can get the weight set more scientific and simple, so that we choose BP algorithm to evaluate the customer value in the various stages.



**Figure 1. The Index System of Customer Value**

## 2.1. Artificial Neural Network

The inspiration for neural networks came from central nervous systems and has been widely used in information technology. In an artificial neural network, simple artificial nodes called "neurons" are connected together to form a network which mimics a biological neural network [11]. BP neural network has been widely used in training artificial neural network (ANN). The BP neural network has three layers, shown as Figure 2. The first layer has input neurons, which send data via synapses to the second layer of neurons, and then via more synapses to the third layer of output neurons.



**Figure 2. BP Neural Network**

The key to establish a neural network model is how to select the appropriate connection weight matrix, this task is realized by network training, and the basic principle is as follows. Input  $I(n)$  to the corresponding input layer, then output layer will get the output as  $Y(n)$ , then let  $D(n)$  represents the corresponding output that given by the training sample, so the error signal at the  $j$  output neurons can be expressed as:

$$e_i(n) = d_i(n) - y_i(n) \quad (1)$$

Assume the samples number for training is N, then the square error to the total training sample would be:

$$E_r = \frac{1}{N} \sum_{n=1}^N E(n) \quad (2)$$

So the goal is get a set of weights W to make the function Er achieve the minimum value, as:

$$\text{Min} \left[ \frac{1}{N} \sum_{n=1}^N \frac{1}{2} \sum_{i=c} (d_i(n) - y_i(n))^2 \right] \quad (3)$$

## 2.2. Structure Model

In customer segmentation structure model, the basic work of data mining method is through analyzing the known data and then summed up a forecast model. The data can be either historical data or exogenous data; exogenous data can be obtained by experimental method and research method [12-13]. Data mining method of customer segmentation should have the following abilities:

- a) The dynamic behavior description
- b) The reliability of the data
- c) The noise is resistance and time-varying
- d) Synthesis of various mining methods

The implementation process of customer segmentation is based on sample learning method. When decide the marketing strategy of customer relationship management, managers often use some description of customer characters, such as “high-income customers”, “low-income customers”, “trendy customer”, “conservative clients”, “high risk customer”, the main task of customer segmentation is to ensure the corresponding relations between these concepts and the corresponding customers[14-17]. Customer data contains several discrete customer attributes and continuous customer attributes, each customer attribute as a dimension, every customer as a bit of space, the enterprise customer database of all customers can constitute a multidimensional space, called the attribute space of customers.

Assume  $A=\{A_1, A_2, \dots, A_m\}$  Is A set of properties that describe customer characteristics and behavior, these properties can be either continuous attributes or discrete attributes, these attributes form the m-dimensional space A, each customer value determines the position in space A. We assume  $c \in C$ , the value of c in attribute  $A_i$  is  $c[A_i]$ . we use g as a abstract concept that descript customer,  $f(g)$  is the set of customer. For a set of concepts  $\{g_{i1}, g_{i2}, \dots, g_{ik}\}$ , if  $\{f(g_{i1}), f(g_{i2}), \dots, f(g_{ik})\}$  is k mutually disjoint sets at any moment, and  $C = f(g_{i1}) \cup f(g_{i2}) \cup \dots \cup f(g_{ik})$ , then we can get  $G_i = g_{i1} \wedge g_{i2} \wedge \dots \wedge g_{ik}$ . According to  $\forall c \in C$ , if  $c \in f(g_{ij})$  and  $1 \leq j \leq k$ , it can be recorded as  $\varphi(c, G_i) = g_{ij}$ . In the concept of customer value dimensions, there are three concepts as “valued customer”, “potential value customer”, “no value customers”, these three concepts can summarize all customers[19].

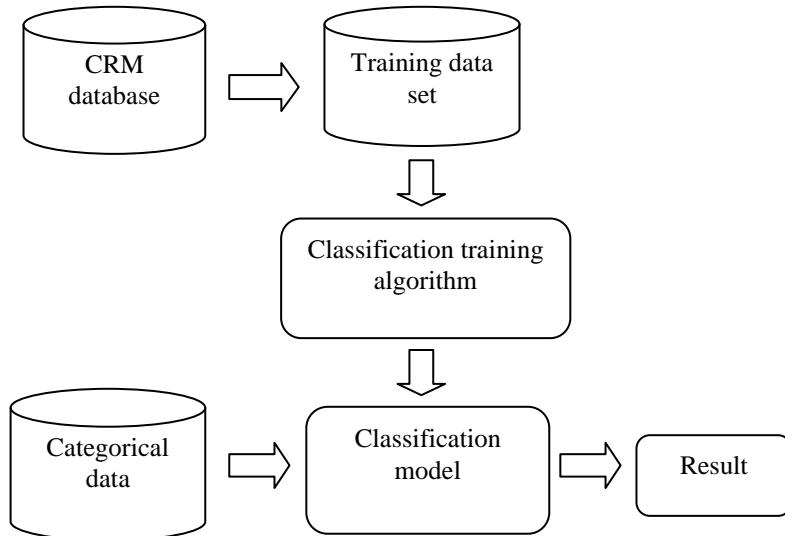
We assume  $B \subseteq G$ ,  $R_B \subseteq C \otimes C$  is a binary relation of C, and  $C \otimes C$  is a Cartesian product of C, the formula of  $R_B$  is:

$$R_B = \{(c_1, c_2) | c_1, c_2 \in C, \forall b \in B, \varphi[c_1, b = \varphi(c_2, b)]\} \quad (4)$$

In this formula,  $R_B$  is an equivalence relation, customers can be divided into several class of the space, and each equivalence class is called a concept class. Customer segmentation is the process that built the mapping relationship of customer attribute space  $A^m$  and concept space  $G^n$  as:  $A^m \rightarrow G^n$ .

The sampler learning method is mainly through data mining process about the customer data, this data mining process was shown as figure 3. We assume  $B=\{G_{i1}, G_{i2}, \dots, G_{ik}\}$ , so that  $L=\{L_1, L_2, \dots, L_k\}$ ,  $c \in C$ .  $c$  is the customer set that concept class has been known, customer segmentation should follow two steps as:

- Define a mapping as  $p: C \rightarrow L$ , which makes  $\forall c \in C$ , if  $c \in L_i$ , so  $p(c)=L_i$ ;
- $\forall c \in C$ , determine the class by calculate  $p(c)$ .



**Figure 3. The Implementation Model of Customer Segmentation**

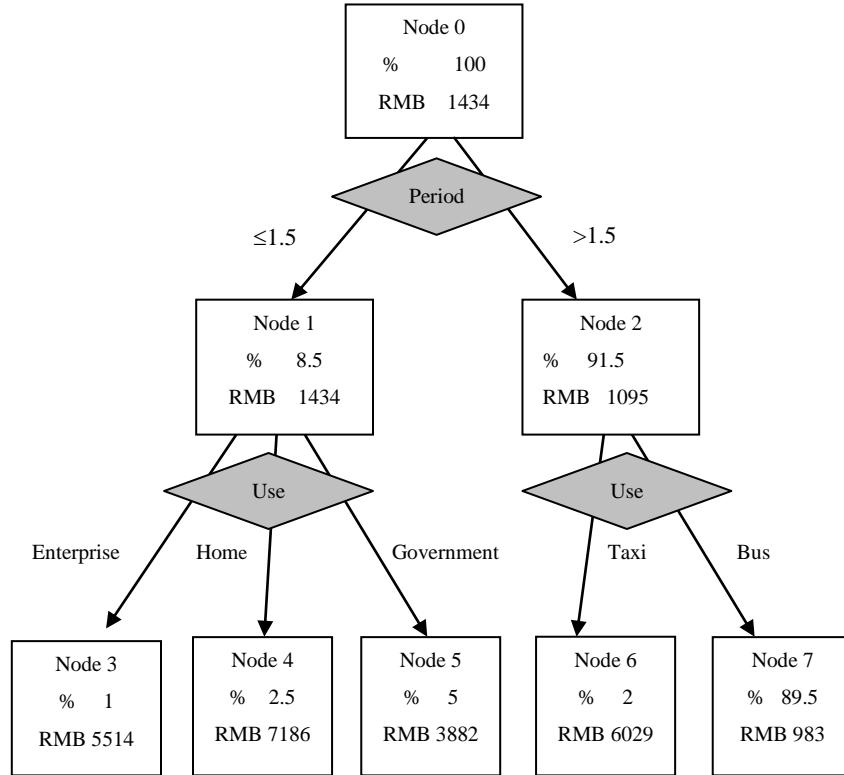
### 3. Empirical Analysis

In this paper, I use data mining method to find out the influence factor of settlement in insurance companies. The insurance company paid most attention to the amount of settlement, how to reduce the amount of settlement and how to formulate the corresponding settlement rate is important for insurance company survive. The factors which influence the amount of settlement, is the primary solution to the insurance company. In fact, the relationship between these factors is very complex, so we use data mining technology and decision tree method to make a further research on the influence factors of the amount of insurance settlement.

#### 3.1. Decision Tree

The data is collected from China insurance company; it includes the settlement records of car owners. First, I put the data into Clementine, and then make standardization processing of the data by using data discretization and data generalization to remove out the redundant information. Then, I use feature selection to choose the main factors that affect the amount of settlement and analyze the customer records. The result shows that: the time period of driving experience and the purpose of using a car is the main factors affect the amount of insurance settlement.

Using the Clementine decision tree model and enhanced ID3 algorithm [20], we can get the decision tree as shown in Figure 4.

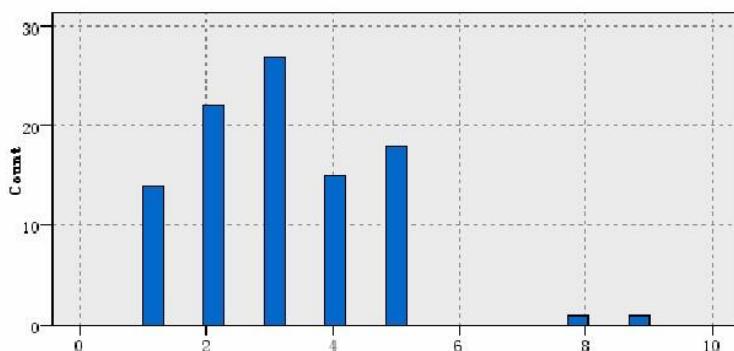


**Figure 4. Decision Tree**

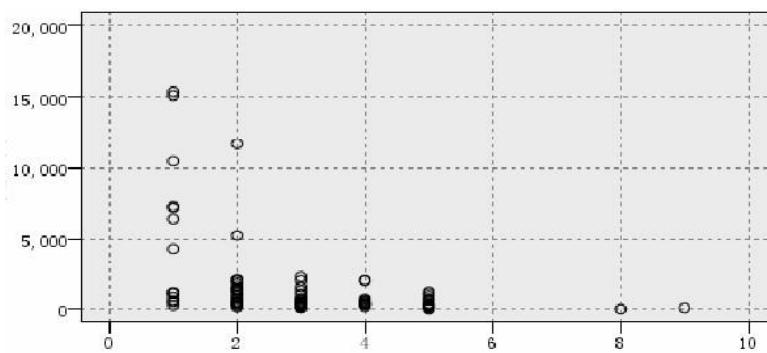
From the decision tree model, we can get that: first, the average settlement amount is 1434 RMB, so that we should increase the consumer ratio that gets average settlement less than 1434 and make a higher insurance price for those who get settlement amount more than 1434; second, we can find that the person who has the driving experience less than 1.5 year needs a average settlement as 5040 RMB, this part of the customer has brought a heavy burden to the company. Therefore, the company should give this part of customers a relatively low settlement rate or increase the insurance price to these customers; third, we can find that using cars for enterprise, home, government and taxi often has a high settlement than 1434 RMB, however, using cars as bus often has a especially low settlement, so that the insurance company should pay more attention to those bus insurance.

### 3.2. Influence Factor Analysis

To sum up, we can know the amount of settlement is mainly influenced by the driving experience and the purpose of using, and then I will make a further analysis on these influence factors. I use the Histogram function in Clementine to analyze the influence factor as driving experience, the result were shown in Figure 5 and Figure 6. From Figure 5, we can find that settlement events occurred mainly in those customers who have the driving experience between 1-5 years, but less occurred in customers who has driving experience more than 5 years. From Figure 6, we can find that the customers who have a driving experience between 1-2 years often have a higher settlement amount.



**Figure 5. Time Period of Driving Experience**



**Figure 6. Settlement Amount**

According to the data analysis, the insurance company should make corresponding strategies: first, make a relative low settlement ratio to the customers who have driving experience in 1 ~ 5 years, especially in 1 ~ 2 years; second, make a higher settlement ratio to the customers that have driving experience more than 5 years, the final purpose is to keep the existing customers, and be able to attract some potential customers.

#### 4. Conclusions

This paper first introduced customer relationship management theory and customer segmentation, and then used data mining method to find out the influence factor of settlement. First the essay puts forward the principle of customer segmentation on the base of existing problems of customer segmentation in customer relationship management. Then from the logical model of customer segment, and puts forward the model of data - function - methods of customer segmentation.

In the case study, I used the decision-making tree and the visual function of the Clementine to analyze the influence factor of insurance settlement. The results showed that the time period of driving experience and the purpose of using a car is the main factors affect the amount of insurance settlement. Then I used that data to make a further analysis, and find out that the consumers who have the driving experience between 1-2 years have a higher risk to get a car accident, these consumers got an average settlement as 5040 RMB, and it is much higher than the average settlement of total consumers. So that the insurance company should

make a relative low settlement ratio to these customers, and make effective customer segmentation based on data mining method.

## Acknowledgements

Where there is great love, there are always miracles. Love is like a butterfly W. E. Scott and R. E. Smalley, J. Naco. Nanotedfdl. 3, 75 (2003).

## References

- [1] R. Jayanti and B. Visha, "Application of data mining techniques in the financial sector for profitable customer relationship management", International Journal of Information and Communication Technology, vol. 2, no. 4, (2010).
- [2] T. Yamaguchi, "Marketing strategy with the introduction of customer relationship management: in the case of financial institutions", Artificial Life and Robotics, vol. 14, no. 4, (2009).
- [3] A. Tuzhilin, "Customer relationship management and Web mining: the next frontier", Data Mining and Knowledge Discovery, vol. 24, no. 3, (2012).
- [4] G. E. Fruchter and S. P. Siguré, "Social Relationship and Transactional Marketing Policies—Maximizing Customer Lifetime Value", Journal of Optimization Theory and Applications, vol. 142, no. 3, (2009).
- [5] L. Yu, S. Wang and K. K. Lai, "Developing an SVM-based ensemble learning system for customer risk identification collaborating with customer relationship management", Frontiers of Computer Science in China, vol. 4, no. 2, (2010).
- [6] G. Monika, V. Rajan, "Applications of data mining in higher education", International Journal of Computer Science Issues, vol. 9, no. 2, (2012).
- [7] F. Usama, P. S. Gregory and S. Padhraic, "From data mining to knowledge discovery in databases", AI Magazine, vol. 17, (1996).
- [8] G. Sarantopoulos, "Data mining in retail credit", Operational Research, vol. 3, no. 2, (2003).
- [9] W. Xiao, Q. Zhao and Q. Fei, "A comparative study of data mining methods in consumer loans credit scoring management", Journal of Systems Science and Systems Engineering, vol. 15, no. 4, (2006).
- [10] X. Hu, "A Data Mining Approach for Retailing Bank Customer Attrition Analysis", Applied Intelligence, vol. 22, no. 1, (2005).
- [11] G. Crapanzano, V. Müller and D. Nelles, "Application of artificial neural networks in network calculations", Electrical Engineering, vol. 83, (2001), pp. 5-6.
- [12] M. A. Emmelhainz and C. B. Kavan, "Using Information as a Basis for Segmentation and Relationship Marketing: A Longitudinal Case Study of a Leading Financial Services Firm", Journal of Market-Focused Management, vol. 4, no. 2, (1999).
- [13] C. V. L. Raju, Y. Narahari and K. Ravikumar, "Learning dynamic prices in electronic retail markets with customer segmentation", Annals of Operations Research, vol. 143, no. 1, (2006).
- [14] Y. Chen, G. Zhang, D. Hu and C. Fu, "Customer segmentation based on survival character", Journal of Intelligent Manufacturing, vol. 18, no. 4, (2007).
- [15] X. Wang, P. Zhao, G. Wang and J. Liu, "Market segmentation based on customer satisfaction-loyalty links", Frontiers of Business Research in China, vol. 1, no. 2, (2007).
- [16] N. H. Chen, S. C. Huang, S. T. Shu and T. S. Wang, "Market segmentation, service quality, and overall satisfaction: self-organizing map and structural equation modeling methods", Quality & Quantity, vol. 47, no. 2, (2013).
- [17] K. Tomak, K. Altinkemer and B. Kazaz, "Market segmentation using service levels in data networks", Information Systems and e-Business Management, vol. 1, no. 1, (2003).
- [18] L. Bulysheva and A. Bulyshev, "Segmentation modeling algorithm: a novel algorithm in data mining", Information Technology and Management, vol. 13, no. 4, (2012).
- [19] X. Hu, "A Data Mining Approach for Retailing Bank Customer Attrition Analysis. Applied Intelligence", vol. 22, no. 1, (2005).
- [20] Y. Yasami and S. P. Mozaffari, "A novel unsupervised classification approach for network anomaly detection by k-Means clustering and ID3 decision tree learning methods", The Journal of Supercomputing, vol. 53, no. 1, (2010).

## Authors



**Xu Zhikun**, he received his bachelor's degree of Economics in Inner Mongolia Agricultural University, (1993) and master's degree of management in North China Electric Power University, (2010), now he work in Hebei Normal University of Science and Technology Institute of Finance and Economics, Qinhuangdao, Hebei. He is a Master's Supervisor. Her major in the fields of study is practical economics.



**Wang Yanwen**, she received her bachelor's degree of economics in China University of Mining & Technology, Xuzhou, Jiangsu. (1999) and master's degree of economics in University of International Business and Economics (2010) , Now she is a lecturer in Hebei Normal University of Science & Technology, Qinhuangdao, Hebei. Her major fields of study are economics, trade, and education.



**Liu Zhaohui**, he received her bachelor's degree of education in Hebei Normal University of Science & Technology, Qinhuangdao, Hebei. (1993) and master's degree of management in North China Electric Power University, Baoding, Hebei. (2009), now he is a lecturer in Hebei Normal University of Science & Technology, Qinhuangdao, Hebei. His major fields of study are practical economics, vocational education, and information education.

