J Inf Process Syst, Vol.11, No.2, pp.173~183, June 2015
http://dx.doi.org/10.3745/JIPS.04.0012

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Research on the E-Commerce Credit Scoring Model Using the Gaussian Density Function

Xiao Qiang*,**, He Rui-chun*, and Zhang Wei*

Abstract

At present, it is simple to the electronic commerce credit scoring model, as a brush credit phenomenon in E-commerce has emerged. This phenomenon affects the judgment of consumers and hinders the rapid development of E-commerce. In this paper, that E-commerce credit evaluation model that uses a Gaussian density function is put forward by density test and the analysis for the anomalies of E-commerce credit rating, it can be fond out the abnormal point in credit scoring, these points were calculated by nonlinear credit scoring algorithm, thus it can effectively improve the current E-commerce credit score, and enhance the accuracy of E-commerce credit score.

Keywords

Abnormal Point, Credit Scoring, Density, E-commerce

1. Introduction

With the development of information technology and social informatization, E-commerce (electronic commerce) has been rapidly developing. China's E-commerce market turnover rate reached 3.5 trillion Yuan by June 2012 with a wear-on-year growth of 18.6% [1]. The online retail sales market turnover was is 511.9 billion Yuan, which was up 46.6% from the year before. The virtual commerce world is fast becoming a reality and in less than five years, it is estimated that nearly 70% of large business transactions will be conducted and signed on the Internet.

E-commerce is completely focused on electronic data interchanges. In this new E-commerce environment, we can see that product decisions can be made based on online catalogues, which can be customized to the needs of the viewing client. Currently, these types of catalogues are customized based on recognizing a sign-on and secure password. As such, the purchaser has access to customized catalogues that reflect his/her needs and purchasing discounts [2,3].

However, the rapid development of E-commerce has triggered many problems. For example, E-commerce fraud is becoming more and more serious, brush poor commentary and good commentary have become a common means to change the credit score in the present E-commerce sites [4,5]. These factors can easily mislead consumers and increase the number of complaints from E-commerce sites.

Manuscript received November 27, 2013; accepted March 31, 2014; onlinefirst December 31, 2014.
Corresponding Author: Xiao Qiang (lzjt_xq@126.com)
* School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou Gansu 730070, China (lzjt_xq@126.com, hrc@mail.lzjtu.cn, zhangw@mail.lzjtu.cn)
** School of Economics and Management, Lanzhou Jiaotong University, Lanzhou Gansu 730070, China (lzjt_xq@126.com)

They can also hinder the development of E-commerce [6,7]. By researching current domestic E-commerce credit scores and the characteristics on current website credit scores [8], we are proposing a credit rating model based on a density function. This model can effectively reduce the influence of poor commentary and good commentary on network consumers.

In Section 2, we review the previous work that has been done on E-commerce credit evaluation. Section 3 addresses the E-commerce credit score model that is based on density. We describe the model's algorithm, which uses a Gaussian density function, in order to provide an outlier evaluation on E-commerce credit. By taking into account E-commerce credit score rules, Section 4 presents our experimental evaluation of outlier point mining on an E-commerce credit evaluation with the use of a density algorithm.

2. Related Work

Credit scoring is based on a set of techniques for the decision maker on whether or not to grant credit to applicants. Therefore, our aim is to build an appropriate credit scoring model to correctly classify the applicants as 'good' or 'bad' so that the customer with the label of 'good' are trustworthy, while the customer classified as 'bad' is not granted any credit. Establishing the credit system of the virtual market is the main path to solving the credit problems in the virtual markets; the credit evaluation, which informs the credit quality; and it provides an important source of trust for the credit in the abstract system, which has become the key to building this system. According to the past trading information, the purpose of credit evaluation is to determine the quality of the products or services delivered by the seller, as well as by the buyers' behaviors in paying the purchase price. Hence, it is the value label or trademark which reflect overall quality of individuals or enterprises.

In terms of the E-commerce credit scoring method, most E-commerce sites have their own rating rules, the site credit can maliciously brush high or low, because of defects in the credit scoring model. Therefore, many researchers have proposed various algorithms and have worked to improve the credit evaluation model. For example, the scholars Francis and White [9] have attempted to identify the factors that lead to the success of an online store, by placing an emphasis on customer value. Joo [10] investigated the relationship between Korean Internet customer values and repurchase intentions by using personalized, emotion, and trust based variables. Atchariyachanvanich et al. [11] examined why Korean Internet consumers continued making online purchases by focusing on consumer benefits. Hahn and Kim [12] investigated how Korean consumer trust and confidence influence online shopping intentions. Kim et al. [13] identified the critical determinants of Korean online customer satisfaction. By studying E-commerce credit rating mechanisms, Shan et al. [14] pointed out the problems in the current grading methods has put forward a new E-commerce credit scoring model. He has also calculated the user's credit-by-credit rating, frequency, amount, and so on, by analyzing the existing Taobao.com credit evaluation system, Li [15] has pointed out the objective factors, such as trade volume, to credit evaluation model, which shows the effect of trade volume on credit value in the algorithm. At the same time, in order to reflect the change of the transaction time on credit value, this model provides the recent credit value and rate as Wang [16], who conducted research on Taobao.com, proposed the 'reverse evaluation method' to analyze trading history and the various factors that affect a store or individual's credibility which weights in C2C (Customer to Customer) transactions. He established an effective and practical Taobao.com shop (sellers) credit evaluation model and advanced the Taobao.com shop (sellers) credit evaluation based on the AHP model. Hong and He [17] also researched Taobao.com's shop credit evaluation. He analyzed the problems in the credit evaluation model and has put forward some improvement measures for real-name certification, credit scoring, credit evaluation, trading evaluation rules, etc. The improved credit evaluation model is set up on these measures. Xu and Wang [18] have proposed a dynamic C2C E-commerce credit evaluation model. This model is based on transaction history, transaction amount, transaction time, evaluation of credibility, and other factors. It can provide a reliable credit evaluation by means of a dynamic calculation. The model can effectively distinguish malicious users and integrity. As such, it can reduce credit fraud and improve the security of C2C E-commerce.

We know that most of the current study focus on the influence factors of E-commerce credit scores, so I put forward methods of improving credit scores, from these algorithms and models of E-commerce credit. However, brush poor commentary and good commentary problems in E-commerce are still not able to avoid the impact. In this paper, we were put forward E-commerce credit evaluation based on density model. This model can be found abnormal points by use of data mining technology and density detection technology; and it can be avoid the impact on E-commerce credit scores.

3. Model of E-commerce Credit Score Based on Density

3.1 Credit Evaluation Model

Credit is important in an E-commerce context because it mitigates perceptions of uncertainty, decreases perceived risk, and positively affects purchase intentions. In the E-commerce environment, certain cues, such as company reputation, information, or an offline parent brand, may help consumers place their trust in an online retailer [19]. In regards to the internal ethical character and the existing basis of the market economy, credit is the will and capacity to combine on the basis of honesty and trustworthiness. This also includes the formed and developed code of conduct and rules of transactions. As a result, credit, which connotes various elements, such as the need for benefits; psychological confidence; the agreed upon format the required rules; the agreement on practices; value evaluation; and different forms (for example ethical, legal, and political credit) is essentially the integration of words and deeds, qualitative credit and capital credit, and subjective honesty and objective solvency. Credit is not only the foundation of subsistence, but is also the basis for entrepreneurship. In view of the specialty of the virtual market, credit in Internet age has become a scarce moral, economic, and social resource that is more important than in any other time in history [2]. At present, most of the E-commerce credit evaluations use the following method: if you are get high praise, the seller's credit score increases by 1; if you get a medium review, the seller's credit score does not change; and if you are given a bad review, the seller's credit score is reduced by 1. This score is denoted by c :

$$c = \sum_{i=1}^{n} p_i \tag{1}$$

where c is the seller's credit score and p_i is the score value for the credit score that the buyer gave after a transaction [20,21].

Although this evaluation method can reflect the seller's credit in a transaction, if the buyers don't have an accurate evaluation of the sale, thought brush credit's score, the credit of the sellers was not be reacted, high credit score can mislead consumers to buy goods.

3.2 The Density-Based Credit Scoring Model

According to the regulations of credit evaluations in E-commerce transactions, the volume and trading time for transactions were extracted from transaction records. These two items have been named as two variables of the data set S. Set S will be analyzed in our model, and we established set S to be:

$$s = \{(x_1, t_1), (x_2, t_2), \cdots (x_n, t_n)\} \tag{2}$$

where x_n is the transaction volume in every transaction and t_n is the transaction time series in every transaction.

In order to prevent the brush poor commentary and good commentary phenomenon from occurring in set S, we used a density-based algorithm to find the abnormal concentration data points (the peaks of the data set density) in set S. When we found the outlier points of the credit evaluation score, we were able to change the score by using the nonlinear calculation. This allowed us to avoid the impact of brush poor commentary and good commentary in an E-commerce credit evaluation.

In order to detect the density of the data sets S, we used the Gaussian density function as a research tool. We did so because the Gaussian function can smooth and imitate any shape and can describe the density of data points [22]. Each data point to the influence of the data space can be formally used as a mathematical function. This represents the influence of the data point in the neighborhood, and it can determine the density of the point of interest by clustering data points. Density is the local maximum global density function. The function can be an arbitrary function decided to the distance between two objects of a neighborhood. Due to the good characteristics of the Gaussian function, we used it as the Gaussian function influence function [23].

x and t are the object of D dimensional data space F^d and the Gaussian function is:

$$f^x(t) = e^{-\frac{d(x,t)^2}{2B^2}} \tag{3}$$

where B is the impact factor and $d(x_i, t_j)$ is distance function.

$$d(x_i, t_j) = [\sum_{k=1}^{p} (x_{ik} - t_{jk})^2]^{\frac{1}{2}} \tag{4}$$

An object x ($x \in F^d$) and its density function is defined as all the data points influence function together given that a few objects are $D = (t_1, t_2, t_3 \cdots t_n) \subset F^d$ in the density of x is then：

$$f_B^D(x) = \sum_{i=1}^{n} e^{-\frac{d(x_i,t_i)^2}{2B^2}} = \sum_{i=1}^{n} e^{-\frac{((x-t_1)^2+(x-t_2)^2+\cdots(x-t_n)^2)}{2B^2}} \tag{5}$$

where $f_B^D(x)$ is the density of data set points.

Set the threshold as $f_B^{D'}(t)$ and then the rules for the calculation are as follows:

$$t_i = \begin{cases} f_B^D(x_i) < f_B^{D'}, & x_i \text{ is Normal point} \\ f_B^D(x_i) > f_B^{D'}, & x_i \text{ is Abnormal points} \end{cases} \tag{6}$$

According to these E-commerce credit score rules, we identified the abnormal data points in a transaction credit evaluation. In order to avoid these points influence to credit score, these points which mean score credit will be calculated again, the rules for the credit score calculation are:

$$c_i = \begin{cases} (\frac{1}{2})^u, u\, is\, abnormal\, number\;\, in\, set\;\, S \\ 1 \;\text{ or }\; -1, normal\, data\, po\text{int}\, in\, in\;\, set\;\, S \end{cases} \tag{7}$$

where c_i is the credit score values.

As can be seen in Eq. (5), it can be found that abnormal data points by the threshold and density function with the nonlinear calculation; it can be avoid that the calculation of E-commerce user credit score inaccuracy which is due to poor evaluation or good evaluation.

The model of the density algorithm is shown in Fig. 1.

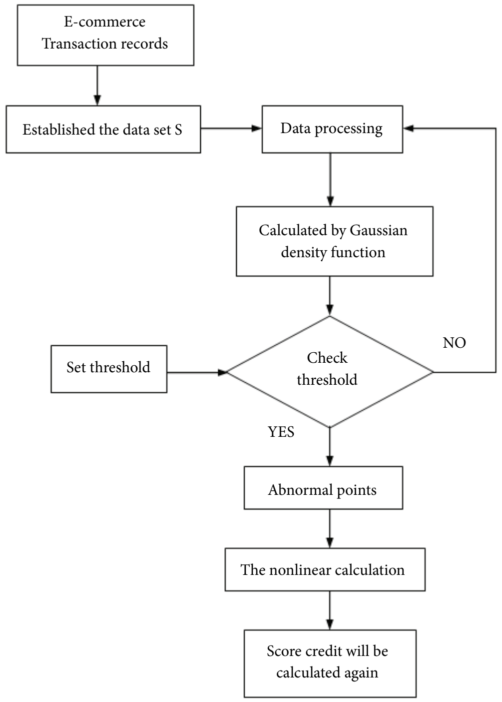

Fig. 1. The credit scoring model with a Gaussian density function.

The basic steps that we followed for calculating a credit score by using the Gaussian density function are as follows:

(1) We selected the transaction records from the E-commerce Web station and find the transaction volume and transaction time in the transaction records.

(2) The transaction volume and transaction time were named aviaries, were named x_i and t_i. They established the data set s.

(3) We changed the data set S to the data series time in data processing and input data into the algorithm.

(4) We calculated the data series times by using the Gaussian density function. The density value was then obtained.

(5) Then we set the threshold. (Note: if the density value is smaller than threshold then go back to Step 3, if the density value is bigger than the threshold, go onto Step 6.)

(6) We found the abnormal point by contrasting the density value with the threshold, and decided upon the credit score in this transaction.

(7) The credit score is entered as a re-accumulated credit evaluation.

4. Experimental Simulations

4.1 Simulation Data Test

In order to validate the algorithm, it can effectively reduce the 'brush' impact on credit, through the virtual online trading system and 50 participants, simulation of random transaction, and credit evaluation, it arranges a brush credit score by five participants in the transaction, and the simulation data is generated by the credit score calculation.

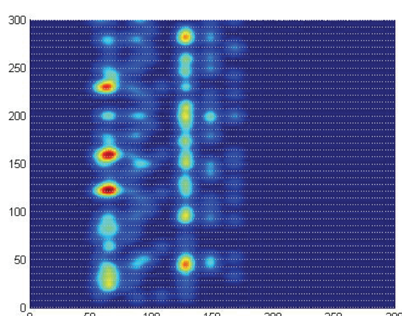

Fig. 2. The density distribution plans of a transaction data set in a virtual trading system.

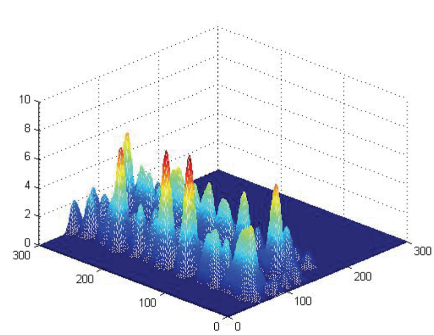

Fig. 3. The transaction data set of the density distribution of a three-dimensional figure in a virtual trading system.

We ran the experiments on a Pentium (R) Dual - CPUE5300 core computer with 2.6 GHz and 1.98 GB of memory. The experiments were conducted on R2007b MATLAB software.

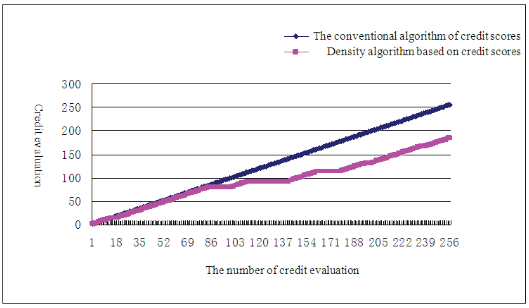

Fig. 4. A credit scoring comparison chart on conventional credit scores and credit scores that are based on a density algorithm in a virtual trading system.

As can be seen in Figs. 2 and 3, in the months of trading,the simulation data generates a density peak in the credit evaluation model. Some scores are red, which shows that the abnormal data points in the data and the data points are bigger than the threshold. The blue scores show that the normal data points in the data and the data points are smaller than threshold. We removed the abnormal data and the credit score was calculated by the nonlinear cumulative. This allowed us to obtain the new credit scores. As shown in Fig. 4, we can see that this algorithm reduces the credit score as compared to the original algorithm. We can also see that it prevents the abnormal evaluation which led to the credit score change. It is effective in improving the authenticity of an E-commerce credit score evaluation.

4.2 Real Data Test

User's transaction data was downloaded from Taobao.com. It was then calculated and verified to the data by using the model for the credit scores. From the trade credit evaluation on Taobao.com, we selected part of the transaction evaluation records, as shown in Fig. 5.

Fig. 5. The transactions and evaluations for Taobao.com's user network records.

As can be seen in Fig. 5, the transaction evaluation records include transaction volume and transaction time. With the establishment of data set S1, we were able to find the abnormal data points. This allowed us to calculate the credit score by using this credit score evaluation model. The results are as shown in Figs. 6 and 7.

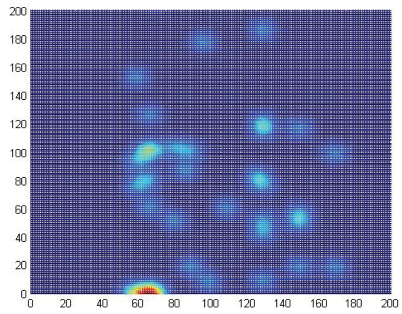

Fig. 6. The density map of Taobao.com's user transaction data sets.

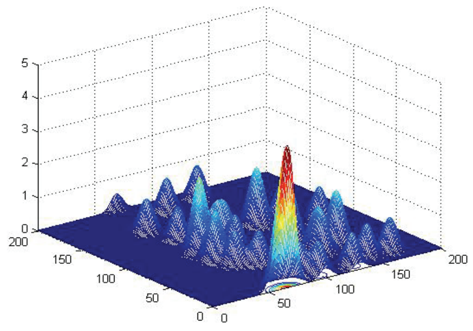

Fig. 7. Taobao.com user transaction data sets density three-dimensional figure.

As can be seen in Figs. 6 and 7, the data sets appear many data dense point. This is because this data is within the scope of the threshold. The points in the graph that are pale blue show that there is no brush evaluation phenomenon in most of these transaction records. The red points the figure show that there are abnormal points in some of the transaction records. As this point isn't within the scope of the threshold, it shows that there is a brush bad/good commentary phenomenon.

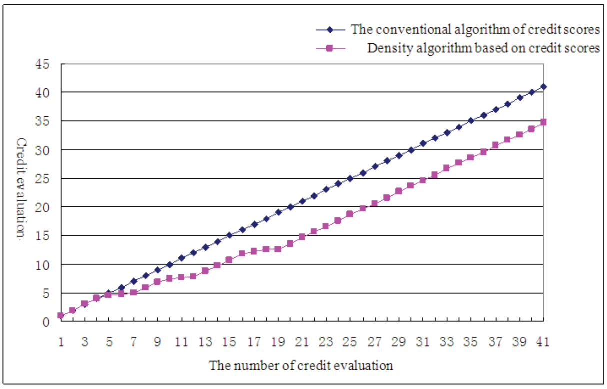

Fig. 8. Credit score comparison between a conventional credit score and a credit score that is based on a density algorithm.

As can be seen in Fig. 8, there is a model algorithm in normal trading; the credit accumulation score are same between convention algorithm and density algorithm in early. However, if an abnormal point was appeared, the credit score model will be lower than the usual Taobao.com credit score. It can then be concluded that this model can effectively prevent the occurrence of brush evaluation phenomenon.

5. Conclusion

E-commerce can be developed with health; it can be favorable to economic growth, so we must avoid bad factors to influence the E-commerce development. In this paper, we show how the factors of a brush score can prevent the E-commerce transaction and that it can mislead consumers to buy goods.

As such, we have proposed this model as a mean to find the brush score and to come up with a reasonable credit score. All of which will allow the customer to feel at ease when making an E-commerce transaction.

Through the introduction and the analysis of the experimental verification about credit score method based on density model, this model can effectively improve the E-commerce credit score evaluation of the results, it can make the credit real to reflect the reality of the E-commerce sellers in E- commerce, and avoid credit score may mislead the consumers' behavior.

However, as seen from our experiments, the influence of the density function factor B and the density threshold setting are the keys to this model, which can accurately ensure abnormal data points. How to select reasonable parameters for different data sets, it will have great influence to the applicability of the model, will be the key research question to this model.

Acknowledgement

This paper is supported by the National Natural Science Foundation of China (Grant No. 6106402 and 61364026) and by the young scientific research Foundation of Lanzhou Jiaotong University (Grant No. 2011044).

References

[1] China Electronic Commerce Research Center, "2012 (on) business data of market of Chinese electron detection report," 2012.

[2] J. Liu and J. Yu, "On the credit alienation and countermeasures of E-commerce", *Asian Social Science*, vol. 5, no. 8, pp. 38-42, 2009.

[3] Z. H. Lei, Z. H. Li, and N. Wu, "Study on E-commerce credit evaluation system based on C2C," *Popular Business*, vol. 109, no. 1, pp.147-149, 2010.

[4] S. R. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TinyDB: an acquisitional query processing system for sensor networks," *ACM Transactions on Database Systems*, vol. 30, no. 1, pp. 122-173, 2005.

[5] Z. Fuguo and X. Shenghua, "Research on recommendation diversification in trust based E-commerce recommender systems," *Journal of the China Society for Scientific and Technical Information*, vol. 29, no. 2, pp.350-355, 2010.

[6] Z. Li, "A species-coexistence model defending against credit cheating in E-commerce," *Computer and Information Science*, vol. 2, no. 3, pp. 154-160, 2009.

[7] D. Liu, "Study on improvement of dynamic reputation Evaluation model in C2C market," *Journal of Hangzhou Dianzi University (Social Science)*, vol. 6, no. 2, pp. 13-16, 2010 .

[8] J. Fraser, N. Fraser, and F. McDonald, "The strategic challenge of electronic commerce," *Supply Chain Management: An International Journal*, vol. 5, no. 1, pp. 7-14, 2000.

[9] J. E. Francis and L. White, "Value across fulfillment-product categories of internet shopping," *Managing Service Quality*, vol. 14, no. 2/3, pp. 226-234, 2004.

[10] J. Joo, "An empirical study on the relationship between customer value and repurchase intention in Korean Internet shopping malls," *Journal of Computer Information Systems*, vol. 48, no. 1, pp. 53-62, 2007.

[11] K. Atchariyachanvanich, N. Sonehara, and H. Okada, "What are the benefits of continued purchasing through

the Internet? A study of South Korean Consumers," *Journal of Service Science and Management*, vol. 1, no. 3, pp. 101-110, 2008.

[12] K. H. Hahn and J. Kim, "The effect of offline brand trust and perceived internet confidence on online shopping intention in the integrated multi-channel context," *International Journal of Retail & Distribution Management*, vol. 37, no. 2, pp. 126-141, 2009.

[13] M. Kim, H. M. Oh, H. and R. D. McNiel, "Determinants of online shoppers' satisfaction in Korea," *Applied Economics Letters*, vol. 15, no. 10, pp. 805-808, 2008.

[14] X. L. Shan, T. Xian, and Y. Yin, "Study on electronic commerce credit scoring method," *World Standardization Quality Management*, no. 6, pp. 32-36, 2007.

[15] L. Li, "Research on Improvement of Credit Evaluation Model for C2C E-Commerce Based on Taobao.com," *Modern Computer (Professional)*, no. 28, pp. 30-32, 2011.

[16] X. M. Wang, "Research and improvement on Taobao.com: as an example to analyze system website store credit evaluation C2C," *Research Productivity*, vol. 8, pp.104-105, 2012.

[17] Q. Hong and G. He, "Analysis and research on the model of the C2C electronic commerce website credit evaluation," *China Management Informationization*, vol. 11, no. 11, pp.96-98, 2008.

[18] H. R. Xu and F. Y. Wang, "Dynamic trust evaluation model for C2C electronic commerce," *Journal of Shangdong University of Technology (Natural Science Edition)*, vol. 26, no. 1, pp. 74-77, 2012.

[19] M. Brengman and F. P. Karimov, "The effect of web communities on consumers' initial trust in B2C e-commerce websites," *Management Research Review*, vol. 35, no. 9, pp. 791-817, 2012.

[20] F. Musau, G. Wang, S. Yu, and M. B. Abdullahi, "Securing recommendations in grouped P2P e-commerce trust model," *IEEE transactions on Network and Service Management*, vol. 9, no. 4, pp. 407-420, 2012.

[21] G. Tian, H. Peng, C. Sun, and Y. Li, "Analysis of reputation speculation behavior in China's C2C E-commerce market," *Journal of Computers*, vol. 7, no. 12, pp. 2971-2978, 2012.

[22] N. Chen, "Automatic image annotation method based on Gaussian mixture model," *Journal of Computer Applications*, vol. 30, no. 11, pp. 2986-2993, 2010.

[23] X. S. Li and D. Y. Li, "A new method based on density clustering for discretization of continuous attributes," *Acta Simulata Systematica Sinica*, vol. 15, no. 6, pp.804-813, 2003.

Xiao Qiang

He received the master degree from the school of information and electrical engineering, Lanzhou Jiaotong University, in 2007. He is currently working toward the Ph.D. degree in the school of Traffic and Transportation, Lanzhou Jiaotong University, at Lanzhou in china. His research interests include data mining, E-commerce and information system.

He Rui-chun

She is currently a professor and Ph.D. adviser at the School of Traffic and Transportation, Lanzhou Jiaotong University. Her major research focuses on Analysis and optimization of transportation system, Analysis of traffic network complexity, Management decision analysis.

Zhang Wei

She received the master degree from the school of traffic and transportation, Lanzhou Jiaotong University, in 2007. She is currently working toward the Ph.D. degree in the school of traffic and transportation, Lanzhou Jiaotong University, at Lanzhou in china. Her research interests include analysis and optimization of transportation system and management decision analysis.