Application Research on E-commerce Credit Evaluation based on Opinion Mining

Xiao Qiang[1,2], He Rui-Chun[1] and Liao Hui[2]

[1]School of Traffic and Transportation, Lanzhou Jiao tong University; Lanzhou; China
[2]School of Economics and management, Lanzhou Jiao tong University; Lanzhou; China

Lzjt_xq@126.com,herc@mail.lzjtu.cn,lzjt_liaohui@126.com

Abstract

To increase the buyers know more about the credit of E-commerce product sellers and the purchase rate of the E-commerce users, E-commerce credit evaluation model based on the opinion mining algorithm was put forward. Extract the feature words and views from the products and user reviews, and then make use of statistical and quantitative way to analyze them. In the mean time, a credit evaluation model with transaction time-frequency can be set up, which can be used to analyze the seller's credit of E-commerce users. Through the experiment, this model was verified to have certain practicability and validity in E-commerce credit evaluation.

Keywords: *Opinion mining, electronic commerce, credit evaluation, point of view*

1. Introduction

The powerful functions of the Internet have promoted the development of E-commerce. In recent years, the product view comments have been an important evaluation when customers choose the goods and the sellers [1]. It is mainly because the comments are based on the experience of buying goods, and the credibility is higher.

The current E-commerce shopping websites have products and user reviews, after E-commerce transactions, the users can publish their own views and opinions of one's own accord within a period of time [2]. However, the users are limited by cognitive ability and the immature information search behavior in the face of the complicated comments. It doesn't necessarily for them to retrieve the information [3], for this reason, many products, users comments and sellers credit are used by the main basis of the credit evaluation of busying E-commerce products [4, 5]. How to effective use of E-commerce credit evaluation, and get the approval of the buyers, which are the only way to improve the purchase rate of E-commerce. Consequently, completing the E-commerce credit evaluation is the only way to solve this problem.

2. Opinion Mining

How to quantitatively extract the effective point-view from the comments on the products and to help consumers better decisions. The process that quantitative extract the opinion of product reviews is named opinion mining [6, 7]. In which there are mainly include three basic elements: view the holder, the individuals or organizations who hold certain view of the particular objects; the object, the role of evaluation view; the view, it is a perception, attitude or evaluation that the view holders make comments on the objects. Opinion mining is to understand the relationship among three basic elements. At present domestic opinion mining research are mainly focus on the comments of the car, bank, movie, electronic products and so on, and there are also some practical opinion mining system mostly by means of the characteristics, and the part of the comments to implement [8].

Foreign research on opinion mining is earlier, from the earliest sentiment analysis tools Review Seer [9] to automatically identify the subjective sentences and the subjective components of the subjective sentences Opinion Finder system [10]; From the online user reviews processing Opinion Observer [11] to a machine learning which is established on the framework of the lexicalization HMMS [12]. In conclusion, it can be seen from the research at home and abroad that opinion mining is gradually a mainstream research in method network analysis.

In order to be able to make use of the electronic commerce of the seller, and effectively help e-commerce consumer to choose and buy products, an E-commerce credit evaluation system based on the product oriented view has been designed. The system can provide the basis for maximum understanding the seller credit through quantitative analyzing the users' reviews in E-commerce sites and quantifying sellers of credit. And it can change the habit of the current credit score to guide consumers by transforming the abnormal brash evaluation.

3. The E-commerce Credit Evaluation System Model based on Opinion Mining

The opinion mining about E-commerce product review opinion is typically included: 1. Establish product reviews feature library and comments view feature library; 2. Extract feature evaluation, judge the point polarity of comment and strength; 3. Evaluate the characteristics of products and quantify the comments. 4. Summarize view mining and statistics; 5. Establish the credit evaluation application. To describe the role and effect of the process in the electronic commerce credit evaluation, this paper designs the model of the E-commerce credit evaluation system based on the opinion mining.
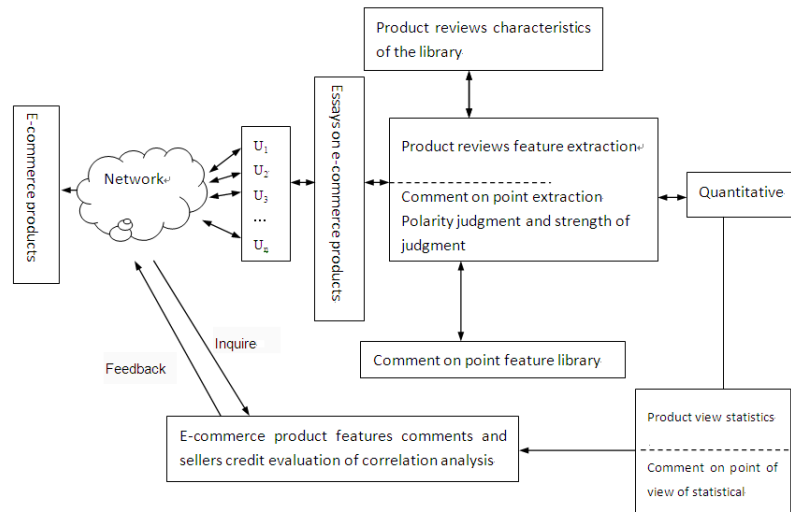
Figure 1. The opinion mining system for E- commerce credit evaluation model

3.1 review of data acquisition and the establishment of the characteristic library

1. Comment data collection

The current E-commerce sites have evaluation system, so the users can evaluate the seller products and credit ratings after the purchase with it. The later buyers can decide whether to buy on the basis of previous buyers' comments and rating. In order to get effective E-commerce comment, this model establishes the buyer evaluation data set by collecting comment data [10]. And then make the following definition.

Define P as the customer evaluation data set, whose data items are p1, p2, p3...pn. Buyers comments on the goods are expressed as p, p= [x1 x2 x3], x1 expresses that the characteristics of the product evaluation, which means that find out evaluation for the product characteristics from the product reviews. X2 expresses the user's subjective attitude, i.e. which means that find out the user's evaluation of subjective attitude from the product review. The x3 said user transaction time.

2. The establishment of the characteristic library

Establish product reviews feature library and comments view feature library. Product characteristics mainly include product type library and product view-word library. Product type library refers to the type of E-commerce products (such as: electronic products, home appliances, clothing, *etc*.) including primary and secondary directory, etc. Each type of products can be found in the library, which is defined as the data set T = [t1 t2, t3... tn]. View word library refers to the evaluation content of E- commerce product (such as: the poor quality, the long standby time, the good fabric, *etc*). The features library of comment view is that the user set point to evaluation of goods, mainly including two categories, namely the polarity (positive and negative), and the strength (adverbs like common ,quite, very, and *etc*.) of the word. Define the data set as L = [l1 l2, l3,.... ln].

3.2 The characteristics extraction of the comment data

The extraction of the review data is the key link of opinion mining, is also the core part of E-commerce credit evaluation System application. This model adopted Chinese Lexical Analysis System ICTCLAS2009 software developed by Institute of Computing Technology, Chinese academy of sciences. The software function includes Chinese word segmentation; the part of speech tagging, named entity recognition, new word recognition; and at the same time support the user dictionary. It may realize the extraction of the characteristics and opinion words of the user data set P.

To extract key words compare to the product reviews feature library and comments view feature library, respectively. Steps are as follows:

1. Firstly, extract the key words and ideas of the users' comments, and then establish the data set S. Define S = [s1 s2, s3... sn], sn = [y1 y2, y3... yn], yn said software extract key words and opinions from each user comments.

2. Establish a "FOR" loop process, compare t and y with l, then draw to the same key words and view words set up data set T1 and L1, respectively, until the end of pumping.

3.3 Quantitative and process the extracted comment data.

To clearly grasp the effect of buyers evaluation for the E-commerce product, and to better provide the basis for E-commerce credit evaluation, the quantitative analysis of the extracted key words and ideas will be carried on. Specific steps are as follows:

1. Define the quantitative standards of the product characteristics values. Viewpoint in the key word are divided into five grades: very good, good, normal, bad, worse, worst. The corresponding score is 6, 5, 4, 3, 2, and 1. After extracting view, quantify the view based on the principles of meaning similar. If extract key words is "product" and view word is "product damaged", which is quantified as 2 points; If the extracted opinion word is "product easy to use", quantified as 5 points.

2. Define the quantitative standards of the user reviews. The standard is the same as the quantitative standards of the product characteristics value. After extracting view s, quantify the view based on the principles of meaning similar. If the extracted review opinion is "normal", qualified as 4 point, if the extracted review opinion is "inconvenient", and get 2 points.

3.4 Products and comment view statistics

1. The product and user reviews statistics. According to the product quantitative processing, the results of the product and user reviews, statistics of the proportion quantitative indicators in products,

$$\frac{\sum x1}{xn}, \frac{\sum x2}{xn}, \frac{\sum x3}{xn}, \frac{\sum x4}{xn}, \frac{\sum x5}{xn}, \frac{\sum x6}{xn}$$

x1 said that the quantitative value is number of 6, x2 said quantitative value is the number of 5, x3 said quantitative value is the number of 4, x4 said quantitative value is the number of 3, x5 said quantitative value is the number of 2, x6 said quantitative value is the number of 1, xn is the total number of the

comments. The calculation method that it statistics the proportion of each view of user comment is the same as the above methods.

2. The word frequency statistics. The view words that extracted from the product and the comments are clustered statistic. Calculate the proportion of the high frequency view words in total, respectively, and then give the review on the produces and comments.

3. The time and frequency statistics. Density analysis was carried out on the trading time of the products. Find out the trading time, and to establish a trading time series $T = \left[0, x_{32} - x_{31}, x_{33} - x_{32}, ... x_{3n} - x_{3(n-1)}\right]$, in which $x_{3n} - x_{3(n-1)}$ said the time difference between the first comments and the second one. Using Gaussian density function is to determine the time frequency. Specific calculation method is as follows [11];

If x and t as the object of D dimensional data space F^d, and the Gaussian function of the data object t and x is as follows:,

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

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

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

and its density function in an object x $(x \in F^d)$ is defined as the sum of the influence function of all data points. Given the density of a few objects $D = (t_1, t_2, t_3 \cdots t_n) \subset F^d$ in x :

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

3.5 Product reviews view statistics and analysis of electronic commerce credit evaluation

Since most of the current e-commerce sellers obtained the credit score from the trade comments. The quality of credit shows the buyer's credit. In the mean time, the consumers can only know the sellers by the comments. As the comments is so multitudinous that it is difficult for consumers to judge the seller credit, however, the review quantitative statistics can help consumers know the seller's credit and make decisions. The main analyses are as the following aspects:

Through statistical and analysis the product reviews, we can see consumer feedback on the product quality in the E-commerce, and find a whole concept of most consumers using products at the same time. To help customers better understand the product and e-commerce credit.

Through trading frequency statistics, buyers can clear find the products concerned by other consumers, by the overall evaluation of the usage of product use, buyers can be better understand goods of consumption.

Through trading frequency statistics, buyers can understand consumer purchase frequency of the goods for a period of time in the past, and to better understand the product sales rate and the formation of credit score.

To establish the following type of E-commerce credit evaluation model based on opinion mining:

$$C = C_1 + C_2 + C_3$$

C_1 means the score of the product view-word quantitative statistics, $C_1 = \sum_{i=1}^{6} i \times b_i$;

$C2$ means the score of users view-word quantitative statistics, $C_2 = \sum_{i=1}^{6} i \times b1_i$;

which i means quantitative score (6,5,4,3,2,1), b_i and $b1_i$ mean the ratio of each quantitative. C_3 means the approximation cycle number of the sellers trading density $f_B^D(x)$ sampling and statistical.

4. Experimental Analyses

In order to validate that this application model can effectively help buyers to understand E-commerce the sellers credit status, we choose one online score in taobao.com, 451 records and 156 comments for nearly a month as the experiment analysis data, use the aspect of mining algorithm to evaluate statistics overall word frequency, and the density calculation of evaluate the time series. Specific results are as follows:
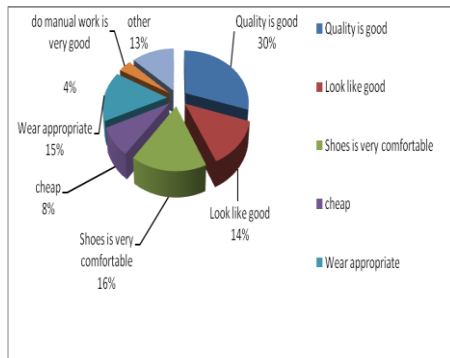

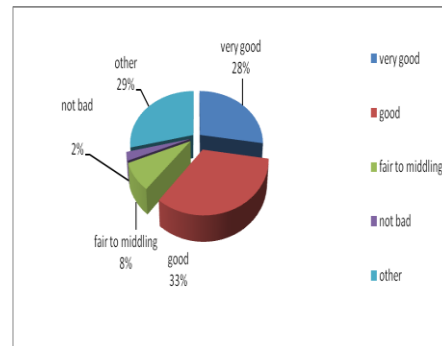**Figure 2. Product features view word statistical figure**

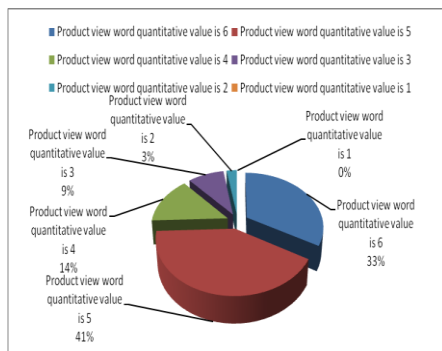**Figure 3. The user review opinion word statistical figure**

Figure 4. The product view word quantify value analysis statistical figure
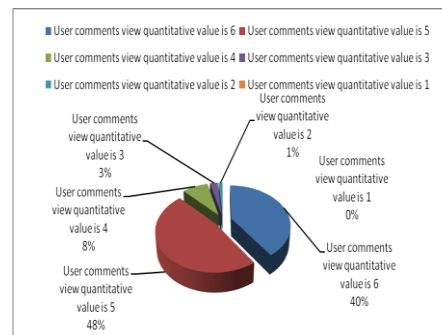

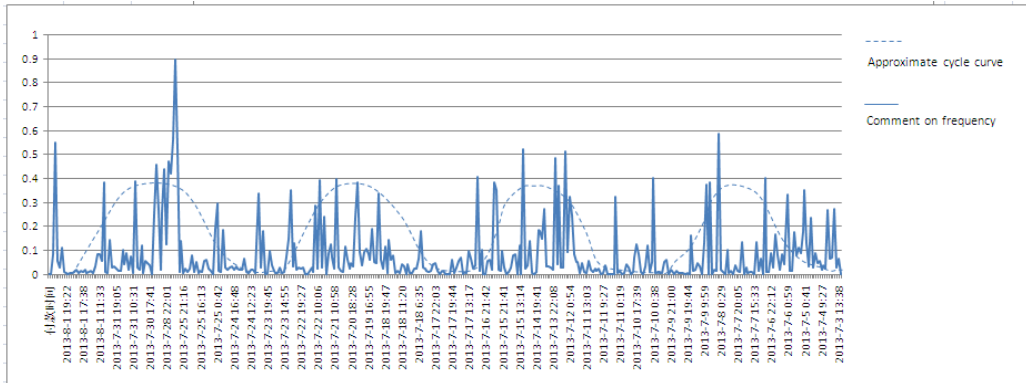**Figure 5. User review opinion word quantitative statistical figure**

Figure 6. User transaction frequency density and approximate cycle graph

Figure 2 and Figure 3 illustrate the user's evaluation for shoes:, the feedback from the products and comments tend to bright side, the "good quality" for product quality, and user comments "best" comment amount for the larger proportion of the overall comment, It suggests that the seller credit value is higher, the site can be trust, and the consumers are satisfied in overall.

From the Figure 4 and Figure 5, we can see that the quantitative value distribution of product and comments, respectively. It is can be seen from the distribution of comments in this figure that most buyers are satisfied, and the rate is high. The conclusion is consistent with the quantitative statistics the products and user comments reviews.

Figure 6 is transaction frequency statistical figure of E-commerce the seller's. It can be seen from the figure that density distribution for nearly a month is reasonable, the dotted line means transaction cycle curve of density, it is shown that the sellers trading is normal, and have a stable customer source. The merchant's credit is very good, and have no deal short cycle (transaction cycle fluctuations faster within a certain amount of time, mean sales too quickly) or long period (transaction cycle fluctuations slow within a certain amount of time, poor sales), which tell us that the business have take no other means to improve the credit in a period, the products can be assured to buy.

According to the experimental data, and the credit evaluation model described earlier:

$$C = C_1 + C_2 + C_3$$
$$C_1 = (6 \times 33\% + 5 \times 41\% + 4 \times 14\% + 3 \times 9\% + 2 \times 3\% + 1 \times 0\%) = 4.92$$
$$C_2 = (6 \times 40\% + 5 \times 48\% + 4 \times 8\% + 3 \times 3\% + 2 \times 3\% + 1 \times 0\%) = 5.23$$
$$C_3 = 4$$
$$C = 14.15$$

The formula of the model algorithm show that if the product and user reviews are high, the proportion is greater that the score will be higher, but it can be seen from the C1 and C2, user reviews and quantitative score products are limited, not always increase. And the C3 is dynamic; it will gradually increase with the transaction time and transaction frequency. Thought the different sales in credit model, we can better understand the credit status of the store.

5. Conclusions

With the method of using opinion mining, this paper analyzes and statistics the comments, and calculates the score model of the E-commerce credit from the perspective of quantitative analysis, to lay the foundation for improving the purchase rate online, which have a certain practical value. From this paper, the keys of the model review quantification are to build key words and view library, but only for a commodity in E-commerce. So, how to build a full library is the key of that whether the system can used in the practice. Next step is aimed at the research of the establishment and perfection of the key repository, and to design application the system model of commerce credit evaluation of with greater practical.

Acknowledgements

References

[1] L. Hui and L. Dongsu, "E-Commerce Reputation Model Based on Elimination Differences of User Subjective".
[2] Y. Li and W. Li-ya, "Modeling and Evaluation for Product Usability Based on Opinion Mining", Computer Engineering, vol. 38, no. 16, (2012), pp. 270-274.
[3] L. Zhiyi and R. Jinfeng, "User Cognitive Information Retrieval Model in B2C E-commerce - A Case Study of Dangdang.com and Joyo.com", Library and information service, vol. 56, no. 12, (2012), pp. 144-147.
[4] S. Zhang and H. Zhong, "Trust Network and Small World Trust Community Clustering for E-Commerce", International Journal of Hybrid Information Technology, vol. 6, no. 2, (2013) March.
[5] S. Zhang, J. Chen and H. Zhong, "Trust Network and Trust Community Clustering based on Shortest Path Analysis for E-commerce", International Journal of u- and e- Service, Science and Technology, vol. 5, no. 2, (2012) June.
[6] C. -m. Yu, "The formal definition of op inion mining in product review", Information Studies: Theory& Application, vol. 32, no. 7, (2009), pp. 124-128.
[7] D. Bhattacharyya, S. Biswas and T. -h. Kim, "Trust Network and Small World Trust Community Clustering for E-Commerce", International Journal of Smart Home, vol. 4, no. 2, (2010) April.
[8] M. Qian, J. Yue and H. -L. Hu, "An Opinion Mining System of Stock Recommendations Based on Web", Computer system application, vol. 21, no. 12, (2012), pp. 38-420+52.
[9] K. Dave, S. Lawrence and D. M. Pennock, "Mining the Peanut Gallery Opinion Extraction and Semantic Classification of Product Reviews", Proc. of the 12th International World Wide Web Conference (WWW2003), Budapest, Hungary, (2003).
[10] T. Wilson, P. Hoffmann, S. Somasundaran and Patwardhan, "Opinion Finder: A system for subjectivity analysis", Demonstration Description in Conference on Empirical Methods in Natural Language Processing, Vancouver, (2005).
[11] B. Liu, et al., "Opinion observer: analyzing and comparing opinions on the web", Proc. of WWW'OS, the14th International Conference on WWW. Chiba, Japan, (2005).
[12] W. Jin, H. Ho and R. Srihari, "OpinionMiner: a novel machine learning system for web opinion mining and extraction", Proc. of KDD, (2009).
[13] X. Wu, Z.-s. He and Y. -w. Huang, "Product review mining: A survey", Computer Engineering and Applications, vol. 44, no. 36, (2008), pp. 37-40.
[14] X. -S. Li and D. -y. Li, "A New Method Based on Density Clustering for Discretization of Continuous Attribute", Journal of system simulation, vol. 15, no. 16, (2003), pp. 804-806+813.
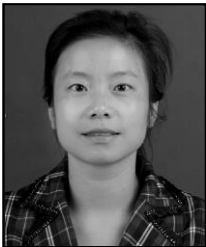
Authors

Xiao Qiang

He received the master degree from the school of information and electrical engineering, Lan Zhou Jiao tong University, in 2007. He is currently working toward the PhD degree in the school of Traffic and Transportation, Lan Zhou Jiao tong 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 Jiao tong University. Her major research focuses on Analysis and optimization of transportation system, Analysis of traffic network complexity, Management decision analysis.

Liao Hui

She received the BS degree from the school of Economics and Management LanZhou Jiao tong University, in 2011. She is currently working toward the master degree in the school of Economics and Management; Lan Zhou Jiao tong University, Lanzhou. Her research interests include business management, E-commerce and information system.