Application of Data Mining Technology in the Analysis of E-commerce Emotional Law

Shuyue Zhang¹,*

¹Guangzhou College of Technology and Business, Guangzhou, Guangdong, 510850, China

*Corresponding author E-mail: zsyue601@ctb.deu.cn

Abstract. The rapid development of the Internet has brought opportunities for enterprises. Data mining is becoming more and more important in e-commerce websites. In e-commerce, data mining is helpful to discover the development trend of the industry and help enterprises make correct decisions. This paper summarizes the current data mining methods of e-commerce, describes the application of e-commerce website in some data mining applications, classifies its objects, and analyzes the functions of network data mining. It is an important tool to support the experimental research and design of e-commerce system oriented to emotion. The technology of Intelligent Emotion Recognition Based on modern mathematical model and the simulation model of emotion oriented intelligent e-commerce system based on computational intelligence are proposed. For the future of e-commerce practice to provide a reference, especially the development and application of software. The progress in this field is outstanding.

Key words: Data Mining Technology, E-commerce Emotion, Online Shopping Evaluation, E-commerce

1. Introduction

With the continuous improvement of computer technology, e-commerce has become indispensable in our life. E-commerce has been growing and has reached the forefront of a new motivational approach to exploring marketing. These new approaches include sub personality based marketing, which deals with many aspects of human personality. In the field of commercial retail, emotion plays a fundamental role. When we choose or buy goods, our choices have a profound emotional dimension. We are social people. We are influenced by other people's opinions. We have our own opinions. We like to bargain for everything. Emotion is the basic element of network service and the key factor to make the final consumption decision. In e-commerce, positive emotions are the driving factors of sales.
Due to the continuous improvement of information technology, many experts have studied e-commerce technology and data mining technology. For example, some domestic teams have studied data mining technology in e-commerce, introduced the current application of data mining technology in e-commerce, focused on the application of fuzzy clustering algorithm in customer group division, predicted and analyzed the sales volume and turnover of goods by time series, The visualization algorithm is used to map the time series data to the complex network. This paper introduces the related knowledge of e-commerce, and expounds the types of mainstream personalized systems and their commonly used recommendation algorithms. Focus on the fuzzy clustering algorithm, according to the user evaluation of goods and commodity classification information and other data, the commodity clustering analysis [1]. In order to improve the customer relationship, the paper uses RFM to analyze the customer relationship, and uses the method of customer relationship analysis and data mining, Reduce enterprise customer maintenance costs. The cluster analysis method is used to analyze the behavior results in the e-commerce operation strategy, and the internal correlation between the behavior results is discussed. Based on the research and analysis of the association marketing operation strategy of the products in the store, it is concluded that data mining technology can help e-commerce enterprises to formulate operation strategies and provide data support [2]. Some experts have studied the data mining of online consumer value, established a comprehensive analysis model of network consumer value combined with clustering and decision tree algorithm, proposed the selection of evaluation index of network consumer value based on RFCP model, and divided online consumers into three groups by cluster analysis method, which maximized the similarity of consumer value within the group. This paper puts forward the necessity of building differentiated services and the marketing strategies of enterprises for different value network consumers. It analyzes the value of online consumers by quantitative and qualitative methods, analyzes the characteristics and laws of different value consumers' purchasing behavior, and provides appropriate marketing strategies, This paper combines theory with practice, uses clustering, decision tree and rule-based classification methods to model and research online consumer value [3]. Although the research on data mining technology is quite fruitful, there are still deficiencies in the research of data mining technology in the analysis of e-commerce emotional laws.

In this paper, in order to study the application of data mining technology in the analysis of e-commerce emotional law, through the research of data mining technology, we found the method of named entity recognition. The results show that data mining technology plays an important role in the analysis of e-commerce emotional law.

2. Method

2.1 Data Mining Technology

Data mining is to combine the information collected by traditional data mining methods and technologies with the information collected through the world wide web. Data mining refers to the process of extracting unknown information hidden in massive, fuzzy, disordered and random data, mining the potential valuable information and knowledge hidden in these huge and complex data, and summarizing these information into knowledge. Data mining has also become knowledge mining in data, but these expressions can not accurately explain the significance of mining from a large number of data [4]. It extracts hidden information and process knowledge from a large number of incomplete and fuzzy random designs, and people have not found valuable information and process knowledge. Knowledge is the premise and constraint condition, and only in a specific domain environment can it have specific practical value. Data mining is not only to discover the causal relationship between things, but also to explore the correlation between things. With the increasing business needs of enterprises, data mining has been extended to medical, commercial, retail and other fields. At present, most domestic researchers focus on the algorithm and details of data mining, rather than the application deployment of data mining[5]. With the continuous innovation of technology, many enterprises attach great importance to the establishment of their own data warehouse and the
application of technology. It is not a simple list of data. It needs to describe the data between classes or concepts by summarizing, comparing, etc. Data mining technology is to discover and extract massive incomplete, noisy, fuzzy and random data. For the shops of e-commerce enterprises, the most important part of establishing data mining system is to collect customer information. For the past shops, most of them rely on experience summary, and subjective judgment often has large errors and lacks objective basis. Data mining technology has a strong ability to identify data information, which makes it more and more widely used in e-commerce operation strategy [6].

2.2 Named Entity Recognition Method

Let \( g = (V, e) \), the node set is \( V \), and the edge set is \( e \). It is a set of random variables based on the node \( V \) in the undirected graph \( G \) [7]. Taking the text sequence \( x \) as the condition, each of them depends on \( X \) and conforms to the Markov characteristic, such as formula (1), where \( V \sim w \) means that two nodes have connected edges and are adjacent nodes, then \( (x, y) \) constitutes a conditional random field.

\[
P(Y_v | X, Y_u, v \neq w) = P(Y_v | X, Y_u, v \sim w)
\]  

(1)

In theory, if the label sequence has conditional independence, then the structure of undirected graph \( G \) is arbitrary, which can be obtained by linking the output node sequence [8].

When defining the characteristic function, we also define the \( \{0,1\} \) binary feature \( B(x, I) \) based on the input sequence \( x \) to represent some distribution characteristics. For example, when \( I \) position of sequence \( x \) is a specific word, the value of \( B(x, I) \) is 1, otherwise it is 0. The transition function is defined as \( B(x, I) \) when and satisfies some matching condition, otherwise it is 0. The state function and characteristic function can be expressed as formula (2-3):

\[
s(y_i, X, i) = s(y_{i-1}, y_i, X, i)
\]

(2)

\[
F_j(Y, X) = \sum_{i=1}^{n} f_i(y_{i-1}, y_i, X, i)
\]

(3)

The local characteristic function refers to the state feature. Then the conditional probability of conditional random field is defined as formula (4):

\[
p(Y | X, \lambda) = \exp(\lambda \cdot F_j(Y, X)) \big/ Z(X)
\]

(4)

The principle of named entity recognition based on conditional random field is the same as that of Chinese word segmentation, which is the process of tagging text sequence [9]. The symbols of the general feature model are represented by the corresponding strings and tags at two to three positions before and after the current position.

3. Experience

3.1 Experimental Data

The data in this paper come from the user's evaluation of water heater. After data cleaning, 21537 negative comments and more positive comments were obtained. In order to ensure the balance of the samples, we randomly selected 5421 positive evaluations from all positive evaluations as experimental data, and randomly selected positive evaluation data to reduce the interference of human factors. 2354 comments were randomly selected from 15236 data sets, including 432 positive comments and 432 negative comments. The polarity of comments in the experimental dataset was manually marked.

3.2 Experimental Process

When the similarity value is greater than the threshold, the word is similar to the reference word and inserted into the emotion dictionary. After the emotional words are added, the corresponding weight is set for the emotional words in the emotion dictionary. When the emotional word is composed of several parts of speech, the corresponding weight is accumulated to get the final weight of the emotional word. It is difficult to determine the weight of some emotional words, which should be
determined according to experience. The first step is to preprocess user comments [10]. In this process, the method of comparative deletion is used to delete duplicate text. Chinese word segmentation tool is rwordseg, which uses the traditional stop word list to delete stop words. The second step is to match the emotion words in the emotion dictionary with the segmented words. If the match is successful, the user comment text will be copied, and then the weight of the word item will be recorded in the emotion dictionary as its corresponding weight value. If the match fails, the weight of the word item is marked as 0. The third step is to repeat the second step until all the words in the comment match the emotion dictionary in turn[11]. Then the weight values of all words in the comment are accumulated to get the final score of the comment. If the score is greater than or equal to 0, the comment is marked as positive, otherwise it is marked as negative. Step 4, repeat until all comments are marked. The method of sentiment analysis based on emotion dictionary is to construct the emotion Dictionary of related fields, calculate the score of emotional polarity value with the help of emotional dictionary, and finally judge the emotional tendency according to the score of emotional polarity value.

4.Discussion

4.1 Analysis of Experimental Results

In e-commerce product reviews, due to the colloquialism of language, many new network words with emotional tendency will be produced. Because the machine has no idea, it can not distinguish the new words with important emotional tendency through the traditional emotional dictionary, which should be retained. Affective analysis, also known as opinion mining, is one of the most researched fields, aiming at analyzing people's opinions. E-commerce sites allow users to share their views on products/services by providing text reviews and digital ratings. These views have a great impact on consumers' future purchase decisions. If there are no emotional words in the emotion dictionary, it is difficult to accurately calculate the emotional polarity of the sentences corresponding to the emotional words. The constructed sentiment Dictionary of e-commerce product reviews can greatly improve the classification performance of sentiment analysis of e-commerce product reviews, which verifies the effectiveness of custom domain sentiment dictionary. Taking water heater reviews as experimental data, this paper compares the classification effect of four emotion dictionaries based on user-defined e-commerce product review sentiment dictionaries myemotion, ntusd, HowNet and Qinghua, as shown in Table 1.

| Emotion dictionary | Accuracy | Sensitivity | Precision | Specificity | Number match | match Ratio |
|--------------------|----------|-------------|-----------|-------------|--------------|-------------|
| HowNet             | 0.233    | 0.413       | 0.675     | 0.546       | 0.546        | 0.345       |
| NTUSD              | 0.355    | 0.134       | 0.346     | 0.466       | 0.453        | 0.654       |
| Qinghua            | 0.132    | 0.254       | 0.634     | 0.245       | 0.436        | 0.654       |
| My Emotion         | 0.132    | 0.256       | 0.245     | 0.566       | 0.565        | 0.435       |
As can be seen above, number is the dimension of emotion dictionary, and matching ratio is the matching rate of emotion dictionary. The matching rate is limited by the delicacy of the emotion dictionary, that is, if there is no emotion word in the emotion dictionary, it is difficult to accurately calculate the emotional polarity of the sentence corresponding to the emotional word. Specificity represents the probability of positive reviews in the total number of positive reviews, which indicates that positive emotion words in Chinese positive and negative dictionaries are suitable for emotional analysis of positive evaluation of e-commerce products. The use of affective words in the two traditional dictionaries is far more than that in this paper, especially in the Chinese Dictionary of commendatory and derogatory meanings.

4.2 Construction of Evaluation Information Extraction Model

Finally, the evaluation information is extracted by using the conditional random field model of syntactic and semantic relations to extract evaluation objects and evaluation words respectively. Evaluation words refer to the words that evaluate the evaluation objects. Evaluation objects refer to the entity itself or entity attributes modified by evaluation words, as well as other aspects beyond product attributes. The experiment of semantic based evaluation information extraction is: training and testing. Firstly, the template and labeled training set are trained to generate template file; then, the test data set is verified by template file; finally, the evaluation results such as accuracy rate, recall rate and F value are evaluated by using evaluation standard. The evaluation objects and evaluation words are extracted from the comment text at the same time, so four kinds of evaluation tags are set up, which are SK, sy, FD and FH. Four groups of comparative experiments were designed by using different combinations of four features: word, part of speech, syntactic relationship and semantic relationship. The selected feature combination forms were represented by D1 ~ D4. As shown in Table 2.

| number | SK(%) | SY(%) | FD(%) | FH(%) |
|--------|-------|-------|-------|-------|
| C1     | 97.54%| 92.48%| 94.45%| 95.65%|
| C2     | 95.78%| 90.76%| 96.89%| 95.46%|
| C3     | 94.87%| 94.45%| 97.55%| 94.79%|
| C4     | 97.43%| 87.54%| 95.56%| 93.65%|
Figure 2. Accuracy of evaluation objects and evaluation words extraction

From the above analysis, the effect of extracting evaluation words SK and FD is better than SY and FH. The result of extracting FH evaluation words is better than that of SK evaluation words.

5.Conclusion
There are many websites on the Internet that provide a variety of product information, allowing users to comment on various products and rate products from 1 to 5. This paper introduces the concept of data mining technology and the architecture of data mining technology. In the field of e-commerce product review, the method of constructing emotion dictionary is: firstly, some emotion words are extracted from e-commerce product review corpus by hand, and then the similarity between reference words and other words in corpus is calculated by semantic similarity method. This article reports on an e-commerce application to analyze users’ feelings for specific products. The results take into account the reduced workload of creating data mining applications using the proposed integration framework and existing data mining tools, as well as qualitative analysis of the quality of development applications, such as the impact on evolution.

Acknowledgments
Young innovative talents project in Guangdong Province in 2020（No.2020KQNCX109）

Reference
[1] Zhou F, Lim M K, He Y, et al. What attracts vehicle consumers' buying[J]. Industrial Management & Data Systems, 2019, 120(1):57-78.
[2] Peng S L, Wang S J, Balas V E, et al. [Advances in Intelligent Systems and Computing] Security with Intelligent Computing and Big-data Services Volume 733 || Automatically Generating Aspect Taxonomy for E-Commerce Domains to Assist Sentiment Mining[J]. 2018, 10.1007/978-3-319-76451-1(Chapter 13):126-137.
[3] Liu X, Wen Y. Analysis of Consumer Appraisal of China's Rural Specialty E-Commerce under Data Mining Method[J]. Open Journal of Stats, 2018, 08(3):401-415.
[4] Saragih M H, Girsang A S. Sentiment analysis of customer engagement on social media in transport online[J]. 2017:24-29.
[5] Yuan H, Xu W, Li Q, et al. Topic sentiment mining for sales performance prediction in e-commerce[J]. Annals of Operations Research, 2018, 270(1):553-576.
[6] Liu Y, Lu J, Shahbazzade S. Sentiment Classification of E-Commerce Product Quality Reviews by FL-SVM Approaches[C]// 2018:292-298.
[7] Tseng K K, Lin F Y, Zhou H, et al. Price prediction of e-commerce products through Internet sentiment analysis[J]. Electronic Commerce Research, 2018, 18(1):65-88.

[8] Vo A D, Nguyen Q P, Ock C Y. Semantic and syntactic analysis in learning representation based on a sentiment analysis model[J]. Applied Intelligence, 2020, 50(3):663-680.

[9] Liang R, Wang J Q. A Linguistic Intuitionistic Cloud Decision Support Model with Sentiment Analysis for Product Selection in E-commerce[J]. International Journal of Fuzzy Systems, 2019, 21(3):963-977.

[10] Ji P, Zhang H Y, Wang J Q. A Fuzzy Decision Support Model With Sentiment Analysis for Items Comparison in e-Commerce: The Case Study of PConline.com[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018, PP:1-12.

[11] Pradeep I K, Bhagyasri I, Praneetha P. E-Commerce With Backbone Of Data Mining[J]. International Journal of Engineering and Technical Research, 2018, 2(7):460.