Research on Credit Evaluation Model of Online Store Based on SnowNLP

Caixia Chen[1], Jue Chen[2], and Chun Shi[1]

[1]Department of E-commerce, School of Information Science and Technology, Hainnan Normal University, Haikou 571158,P.R.China.
[2]School of economic management, Hainan College of Vocation and Technique, Haikou 570105, P. R. China；

Abstract. The online store credit rating is a reflection of the seller's integrity and the quality of the product. The level of the credit rating directly affects the buyer's desire to purchase. Two important factors affecting the credit rating are data and models. The innovation of this research is that the collected data comes from the second evaluation, and the credit evaluation model is improved based on the snowNLP tool, and the malicious brushing filtering function is added. Compared with the credit evaluation system commonly used in current online stores, the evaluation results of the paper are more accurate, detailed and intuitive, and may effectively reduce false brushing and threat review.

1 Introduction

Credit is the foundation of online store survival. For this reason, many online shopping platforms provide credit evaluation systems for both buyers and sellers. When a transaction is completed, both parties can give praise, mid-level review or bad review to the other party according to the satisfaction degree of the transaction.The paper founds that the first review after the completion of the transaction is often completed when the buyer is not particularly familiar with the goods or very casual, and the second review is produced after the buyer's real experience of the goods. If the experience is too good or too poor, the second review is more representative of the actual credit rating of the store than the first review. It is more representative. Therefore, The paper uses sentiment analysis to analyze the text of the second review, filters the malicious brusher review content through the TF-IDF tool, and builds the online store credit evaluation model using SnowNLP sentiment analysis tool.The comparison results show that the credit evaluation model of this study is more accurate, detailed and intuitive than the credit evaluation system commonly used in online stores, and it can effectively improve the phenomenon of false brushing and threat review.

2 Summary of related theories and techniques

2.1 Text sentiment analysis

Text sentiment analysis, also known as Opinion Mining, refers to the process of analyzing, processing, summarizing and reasoning subjective texts with emotional color.

2.2 SnowNLP

SnowNLP is a Python class library that can perform Chinese word segmentation, part-of-speech tagging, sentiment analysis, text categorization, conversion pinyin, traditional simplification, extracting text keywords, extracting abstracts, segmenting sentences, and text similarity. This paper uses the SnowNLP tool to perform text sentiment analysis on the content of the second review. The specific idea is word segmentation—extracting features--feature selection--classification model--recognition results.

2.3 TF-IDF

TF-IDF statistical method used to determine the importance of a word in a certain text set. The traditional method is only TF. TF indicates that the frequency of the word appears is more important, but in the review, "you", "I", "he", etc. often have meaningless words, which leads to simply reading words. Frequently determining its importance is inaccurate. This paper introduces IDF. IDF indicates that the less the text of a word has, the larger the IDF is, the more important the word will be. By combining the two TF-IDFs, it is possible to more accurately determine the importance of words to the text set.

---

[*] Corresponding author: Caixia Chen:914922505@qq.com

2.4 Octopus data collector

Octopus data collector can not only capture the page text data, but also set the automatic page turning and capture multiple pages of text data.This paper uses the Octopus data collector to collect secondary evaluation text data.

3. Credit evaluation model
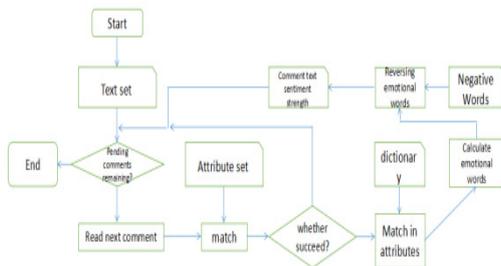
3.1 Sentiment Analysis Framework(Fig. 1)

Fig. 1. Sentiment Analysis Framework

3.2 Sentiment Analysis Process

3.2.1 Emotional factor calculation formula

$$Q_{adv_i} = \begin{cases} Neg_i \times Mod_i \\ \dfrac{Neg_i}{Mod_i} \end{cases} \qquad (1)$$

The negative word sentiment word Neg_i is -1 when there is a negative word, and +1 when there is no negative word.

Degree adverbs are classified as positive and negative Mod_i according to the degree of classification.

Emotional factor Q_{advi}

3.2.2 Emotional value polarity formula

$$Score\left(f_{t_i}\right) = Q_{adv_i} \times P_{w_i} \qquad (2)$$

f_{ti} is an attribute in the comment.

3.2.3 Positive formula of emotional value intensity

$$Sentiment\left(f_{t_i}\right)_+ = \sum_{j=1}^{cp} w_j Score\left(f_{t_j}\right)_+ \qquad (3)$$

3.2.4 Affective value strength negative formula

$$Sentiment\left(f_{t_i}\right)_- = \sum_{k=1}^{cn} w_k Score\left(f_{t_k}\right)_- \qquad (4)$$

The *cp* in Formula 3 represents a sentence whose text attribute obtained after filtering corresponds to a positive polarity;

The *cn* in Formula 4 represents a sentence in which the text attribute obtained after the screening corresponds to a negative polarity;

w represents the weight value of the positive and negative polarity of the text analysis, W_j is positive and W_k is negative.

3.2.5 Keyword extraction construction formula

$$WS(V_i) = (1-d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j) \qquad (5)$$

(1) The sentence segmentation is given for a given text set *T*, which is divided into several small sentences *S*, that is, the set *T = [S1, S2, S3, ..., S4]*.

(2) Separating, segmentation, and part-of-speech of *S*, filtering words with no information meaning, leaving words *t*, verbs, nouns, adjectives, etc. with specified part of speech, forming a set *S=[ti,1,ti,2,... Ti, j]*, *ti, j* are selected candidate keywords.

(3) A graph of the candidate keywords *G = (V, E)*,Here we have *V* as the node set, which consists of the specified keywords selected in the above 2, and uses the co-occurrence relationship to construct the edges of any two nodes in the node set. The existence edge between the two nodes will be co-occurring only when the length of the corresponding vocabulary is *K* (the size of this window), indicating that it has a maximum of *K* words on its surface.

(4) Then, according to the formula of the weight, the weights of the respective nodes are iteratively calculated until the convergence is finally displayed.

(5) Reverse the weighted weights to get the *N* words with the highest importance. The collected keywords are marked in the original text set, and if they are adjacent, keywords containing multiple words are formed.

4 Model Verification

4.1 Collecting data

The Octopus Data Collector software was used to analyze the product review of Taobao's home textile shop.

(1) Select the web address and set the page turning cycle in the process interface.

(2) Select all the reviews with review in the review. Each review selects 5 fields, which are user name, user review, review time, whether the review is valid, and additional review, including the failure to complete the review according to the time. The system defaults to praise, but later there is additional review.

(3) Import the filtered data into the excel table.

4.2 Data preprocessing

(1) Segmentation of the text:
  For example:
  s1.snownlp.SnowNLP ('Look at it, it's not so good,

general')
Print(s1.words)
The result of the word segmentation is:
'Look', 'de', 'also', 'can', 'bar', ', ', 'and', 'no', 'yes', 'then', 'good', ', ', 'general '
(2) Mark the part of speech:
For example:
Tags=[x for x in s1.tags]
     Print(tags)
The result of the labeling is:
('Look', 'v'), ('at', 'u'), ('also', 'd'), ('may', 'v'), ('bar', 'y'), (',', 'w'), ('and', 'c'), ('no', 'd'), ('yes', 'v'), ('then', 'r'), ('good', 'a'), (', ', 'w'), ('general', 'a')
(3)Break the sentence in a big sentence:
For example:
Print.(s1.sentences)
The result of the sentence:
 'Look at it, ', 'not so good', 'general'
(4)Scoring the sent comments to the sentiment analysis
All points are positive, scores greater than 0.5 are positive emotions, and scores less than 0.5 are negative emotions.
For example:
 Print.(s1.sentiments)
The score result is:
 0.6105560234212057
The result of sentiment analysis here is a value in the interval [0,1]. The closer to 1, the more positive the emotion, the closer to 0, the more negative the emotion or it can be understood as the probability of positive.
Probability of *s.sentiments#positive*
0.8463107097139686
(5) Obtain the emotional positive result graph and the emotional negative result graph

4.3 Effective keyword extraction research

By crawling the review data, it is found that there is a malicious phenomenon, and it is necessary to filter the review with the following characteristics:
(1) There are a certain number of similar or identical reviews.
(2) The review in the front of the comment area is long and the review is comprehensive.
(2) The text is short, but there are too many similar words.
(3) A large number of anonymous ratings appeared in a short period of time, especially those with higher levels of Taobao users.
(4) The length is moderate, but the text content is almost identical.
(5) Common high frequency words for network brushing.
(6) Default praise

5 Conclusions and significance

The currently widely used evaluation system is compared with the evaluation system of this paper:

(1) Through the comparison of pictures, we can clearly find that the scores of the analyzed comments are more detailed. In the original interface, only some of the key short sentences are displayed, and the consumers are not given a more intuitive visual form.After the improvement, specific words will be presented, and the label of the keyword language will be more completely supplemented. This avoids users who are threatened by praise or their own review habits. Although they are not satisfied with the quality of the products, they still give Praise. The content contained in these reviews may not be a compliment to the product, but more dissatisfaction. It is difficult to notice this review only through the original praise evaluation system.

(2) The current society is developing rapidly, and people's time is getting more and more tense. Many consumers may not have enough time to browse the review of this product. After the extraction, they can quickly find the main review characteristics of this product. With more detailed guidance, in addition to the key words of all comments, we can clearly understand the weight of them, can more accurately and succinctly understand the consumer's feelings of this product, increase the overall user viscosity.

(3) According to the weight of these keywords evaluated by consumers, merchants can also find some advantages and disadvantages of their own products, and can further improve their own advantages, and can also be highlighted in the promotion, and the description is more concentrated. Disadvantages can be improved.

Acknowledgement

References

1. J.Lu,B.Zhuang,Chinese scientific papers online, *Fine-grained chinese text sentiment analysis based on deep learning*,**12**:712-717(2017)

2. X.Zhang,Economic Research Guide,*Research on Taobao Credit System Evaluation,*No.**19**:4-5+26(2017)

3. X.Sun,DALIAN MARITIME UNIVERSITY NEWS CENTER,*Research on Credit Evaluation System of Taobao Merchants*,2017

4. X.H.TANG,Capital University of Economics And Business,*Research on the Construction of E-commerce Credit Evaluation System*,2016

5. Y.Yang,Central School of Communist Youth League of China ,*A Study upon the Act of Unfair Credit Evaluation under the Mode of C2C*,2016

6. Y.Luo,S.L.Zhao,Journal of Computer Applications,*Text keyword extraction method based on word frequency statistics*,**36**(03):718-725.(2016)

7. H.H.Chen,F.L.Huang,C.D.Pang,Journal of Fujian Normal University(Natural Science Edition),*Mining User Satisfaction Based on Product Review*

Sentiment Analysis,**33(01)**:14-21.(2017)

8. X.M.Wang,Productivity Research,*Analysis and Improvement of C2C Website Store Credit Evaluation System——Taking Taobao as an Example*,**(08)**:104-105.(2012)

9. K.F.Chen,Economic & Trade,*Analysis on the development status and future development trend of Taobao*,**(13)**:130..(2016)

10. Xi Wu,Li Yan,Ting Zhong. *How the dimension of some GCF ϵ sets change with proper choice of the parameter function ϵ (k)*[J]. Journal of Number Theory,2017,174.

11. Alexandra Piryatinska,Boris Darkhovsky,Alexander Kaplan. *Binary classification of multichannel-EEG records based on the ϵ-complexity of continuous vector functions*[J]. Computer Methods and Programs in Biomedicine,2017.