Original Paper

Reputation Analysis of E-commerce Products Based on Online Reviews—Take Amazon as an Example

Lingcong Wang[1*], Kaiwen Tan[1] & Yi Huang[1]

[1] School of Economics, Management and Law, University of South China, Hunan Hengyang, 421000 China

[*] Lingcong Wang, School of Economics, Management and Law, University of South China, Hunan Hengyang, 421000 China

Abstract

Based on the product review information of amazon from 2006 to 2016, this paper measures the correlation between star rating and text review. First, a supervised SVM model was established to classify text emotions. Second, the word2vec model was built to analyze the unsupervised emotion of the comment text. Third, the relationship between the special text and the score is obtained by using the grey metric model. In addition, the article has also carried on the principal component analysis to the commodity reputation, and has determined the representative good commodity and the inferior commodity. Finally, this paper puts forward some Suggestions for merchants, e-commerce platforms and governments.

Keywords

e-commerce, shopping review, SVM, text sentiment analysis, PCA

1. Introduction

Nowadays, with the development of e-commerce and offline logistics, online shopping is popular all over the world. E-commerce platforms, represented by amazon and Alibaba, provide community members with the opportunity to rate and evaluate their products. Online reviews of products are important information resources for consumers, marketing departments and shopping platforms.

For consumers, reviews provide a reference for the quality of the product, while the attributes of the reviewer and the degree to which the review is helpful reflect the quality of the review. The comments of good reviewers tend to be more acceptable, and the comments at the top of the list are considered helpful by most of the members.

For enterprises, only by timely, accurate and reasonable analysis of online reviews, and understanding of consumers' shopping experience and emotional tendency, can they better promote attractive products.

2. Review

In the online comment research of Chinese e-commerce platform, Ye (2016) used neural network to transfer online comment text from traditional document space to high-dimensional vector semantic space, and conducted clustering analysis on mined comment feature words. Cui (2018) calculated the emotional intensity of online comments based on semantic dictionaries. Guo (2019) analyzed the relationship between online reviews and product sales.

In terms of online commentary research on e-commerce platforms outside China, Zhao (2018) analyzed the relationship between the distribution of evaluation resources in Amazon and the usefulness of online customer reviews. Sun (2018) used R language and word cloud analysis technology to conduct text mining on amazon's household products.

It can be seen that there are few text mining articles for online reviews of e-commerce in Amazon, and the application of text mining technology can play a great role in tracking consumers' demands and preferences in real time.

This paper considers the establishment of a supervised support vector machine (SVM) model, then the establishment of an unsupervised text emotion analysis model, and the analysis of the impact of scoring on subsequent comments, and then the establishment of a commodity reputation evaluation model based on principal component analysis (PCA), and the differentiation of successful and failed products.

3. Data Description

3.1 Descriptive Statistics

The data used in this paper includes 30,000 online reviews of three categories of products on Amazon from 2006 to 2016. The basic description is shown in Table 1.

Table 1. Comment Information Description Table

Product	Hair dryer	Pacifier	Microwave oven
type	Beauty	Baby	Major Appliances
Product species	473	5432	55
The total number of reviews	11470	18940	1615
Maximum number of reviews	833	587	394
Minimum number of reviews	1	1	1
verified purchaser	3467	4198	1114

3.2 Handling of Outliers and Missing Values

In order to ensure the scientific nature and accuracy of the research data, this paper took targeted replacement and supplement of abnormal, vacant and wrong data, and deleted the number of wrong comments that could not be counted and analyzed.

1. Delete comments with missing values, as shown in Table 2.

Table 2. Examples of Empty Values

Customer _id	Product _category	…	Helpful _votes	Review _date
34761565		…		8/28/2015

2. Delete the field corresponding to the wrong data.

This paper found 60 misplaced comments in the pacifier data set, as shown in Table 3. Considering that such data only accounted for 0.32% of the total number of comments, it had little impact on subsequent studies, and was also excluded in this paper.

Table 3. Examples of Error Values

Customer _id	Total_ votes	vine	Verified _purchase	Review _headline	Review _body
38939521	n	y	we love it....	we have …….	8/18/2015

4. Model Building

4.1 The SVM Model

4.1.1 Establishment of TFIDF Model

1. Build text vectors

The reference text is represented as a vector:

$$y_k = \{x_1, x_2, \cdots, x_i \} \tag{1}$$

x_i represents the number of times the ith word in the thesaurus appears in the reference text in clause k. According to the above, all the collected texts are represented in the form of vectors, as follows,

$$\begin{vmatrix} 1 & 0 & 0 & \cdots & 1 & 1 \\ 2 & 0 & 1 & \cdots & 0 & 2 \\ & \vdots & & \cdots & & \vdots \\ 0 & 2 & 1 & \cdots & 2 & 1 \\ 0 & 1 & 2 & \cdots & 1 & 0 \end{vmatrix} \tag{2}$$

2. Tf - idf Calculation

If a word or phrase appears in a text with a high frequency of tf and a small frequency in other texts, it is considered that the word or phrase has a good ability of categorization, so the word is selected as the key word of the text, and the emotion of the word is used as the emotional representation of the text.

4.1.2 Establishment of Support Vector Machine Model

Based on corpus, this paper establishes the model of linear support vector machine.

First, we divided the text review and star rating into training set and test set.

Secondly, based on the training set data, the hyperplane linear equation of the classifier is calculated by using python program: $w^T x + b = 0$, as shown in Figure 1.
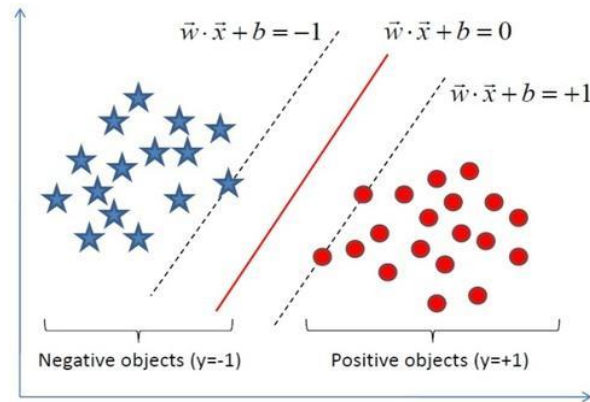

Figure 1. Linear Equation Diagram of Support Vector Machine

4.1.3 Assessment of K-fold Cross-validation

1. In order to better reflect the generalization ability of the model, over-fitting or under-fitting situations are prevented.

The synthesis of the model is evaluated by k-folding cross validation. The basic idea is to divide the total sample set into K points, in which k-1 point is used as the training set and the other 1 point as the test set to obtain a test accuracy.

In order to better reflect the robustness of the model, this operation will be carried out K times in turn, each time different training sets and test sets are selected, and finally the average of all accuracy is calculated, which represents the comprehensive score of this model. Here, 20% fold cross-validation is used.

2. The solution of SVM USES python sklearn library, and tries different kernel functions of SMV, including linear kernel function, polynomial kernel function and gaussian radial basis kernel function. By adjusting parameters, the linear kernel function is fast and has good effect.

3. The first classification is defined: the five levels of star rating are divided into two categories, greater than or equal to 3 for one category, and less than 3 for one category, representing high rating and low rating respectively.

4. The second classification is defined: the five registrations with star rating are divided into three categories: greater than 3 is one category, equal to 3 is one category, and less than 3 is one category, which respectively represent high score, medium score, and low score.

5. The pre-trained SVM model predicted the star rating through text comments. The comprehensive

131

score of the first classification was 0.921, and that of the second classification was 0.862.

Considering that there was a significant difference between low and high scores, the average score was actually more positive.

Therefore, the second division method is used to predict the star rating of some samples according to the text comment, and the data with the predicted result of 3 is analyzed to calculate the difference between the actual star rating greater than 3 and less than 3. The results are shown in Table 4.

Table 4. Reviews on Forecast Results

Sample Size	8767
The Number Predicted to Be 3	947
Star Rating>3	37
Star Rating<3	7

Conclusion: It can be seen that the number of star rating greater than 3 is significantly greater than the number of star rating less than 3. Comparatively speaking, under the above training conditions, the SVM model is more inclined to positive emotion when the result of star rating training is 3.

4.2 Emotional Evaluation Model of Reviews

4.2.1 Modeling Ideas

SVM is a supervised machine learning model that can monitor the differences between text comments and star ratings in real time. In order to carry out emotion analysis on text comment alone, we need to choose an unsupervised model to classify the intensity of text emotion. Therefore, this paper establishes an emotion analysis model based on word2vec.

Word2vec is a correlation model used to generate word vectors, which is based on a two-layer neural network, and is used for training to reconstruct linguistic word text. After training, the word2vec model can be used to map each word to a vector, which can be used to represent the relationship between word pairs, which is the hidden layer of the neural network.

After analysis, it can reflect the emotional words usually associated with adjectives and verbs, so only to extract text comments adjectives, verbs, and negative word is analyzed, and then use word2vec skip the similar degree between the words - gramm model training, in order to determine the direction of sentiment analysis, select "great" and "disappointed" for the root node, the two words of breadth-first Search BFS (Breadth First Search), establish a thesaurus of positive affection and negative affection. Based on these two word bases, a comprehensive evaluation model was established to score the emotional intensity of each comment.

4.2.2 Model Establishment

1. Text preprocessing

Using the python NLTK toolkit, extract adjectives, verbs, and negatives from the comments. The word

132

frequency of text words is counted and the most frequent words are selected to form a vocabulary. Generate a one-hot vector for each word, which represents whether the word has ever appeared in the vocabulary.

2. Establish a skim-gram neural network structure

Let the current word be w_t, then the context is represented as w_(t-1), w_(t-2), w_(t+1), w_(t+2), and the neural network model including the input layer, the hidden layer and the output layer is established to predict the context when the current word is known.

3. Use gradient descent method to solve the updated formula.

4. Divide the positive and negative word bases. With "great" and "disappointed" respectively as the root node, the word similarity matrix to perform breadth-first search, search count set to 100 layers, each time to extract the maximum correlation between parent and more than ten words, run continually add thesaurus lists, ensure no repeat words appear at the same time, at the same time, many times to adjust parameters, makes the number of words in two words library, the resulting word library size 223 and 220 respectively, partial results as shown in Table 5.

Table 5. Comment Word Property List

Positive comment	Negative comment
wonderful	frustrating
fantastic	annoyingto
good	irritating
awesome	badly
amazing	scary
excellent	sturdily
perfect nice	pleasant

5. Establish comprehensive evaluation operator

Let the words in the positive thesaurus be P_i, the words in the negative thesaurus be N, the function F(x,y) be the degree of similarity between word x and word y, remember each comment Q_n, N is the number of words under the comment, $len0$ is the total number of words in the comment, len1 is the total number of positive thesaurus, $len2$ is the total number of negative thesaurus, $len1 \approx len2$.

Then calculate the number of negative and positive words contained in the comment text. If the number of positive words is greater than the number of negative words then the comment tends to be positive and less than negative. This paper USES Emotion to measure the degree of Emotion tendency, and defines the following evaluation methods:

$$Emotion = \frac{\alpha|sum1 - sum2|}{len} \tag{3}$$

Then standardize all comments into {-1, -0.5, 0, 0.5, 1}, five emotional bias, positive means positive, negative means negative.

4.3 Influence of Cumulative Special Scores

First, by using the word2vec emotion analysis and breadth-first search emotion analysis model, the degree of emotional inclination of all text comments was quantified, ranging from -1 to 1. In order to reflect the relationship between cumulative star rating and review of goods, the product parent with large amount of data was selected for analysis. After the statistics of all product parent, it was found that the amount of data with high rating was far more than that with low rating. After the accumulation of the quantity, there is no situation that the data volume of low score is higher than that of high score, so the influence of high score accumulation on the emotional tendency of the current text comment is analyzed here.

1. Extract the data of product_parent 246038397 and 392768822.

2. Combine the time into months, count each value of star rating respectively, and sum up the emotion of each day.

3. Since there is no cumulative value for the goods that have just been put on the shelves in the previous months, the cumulative value of five stars in the first nine months of other months can be calculated without considering these data.

4. considering the low grade and high grade will influence each other, such as although sometimes low score accumulation value is bigger, but because of high grade cumulative value far more than the low score accumulation, leading to the text of the overall comments this month emotions tend to be positive, so the analysis of star rating greater than or equal to 4 and cumulative value of star rating less than or equal to 2 cumulative value of the difference and relationship between emotional tendencies.

5. Reduce the cumulative difference of monthly star rating in the unit 1 range, because the emotional range is -1 to 1, so this range is defined as -1 to 1.

6. Compare the relationship between cumulative score and text emotion, as shown in Figures 2 and 3.
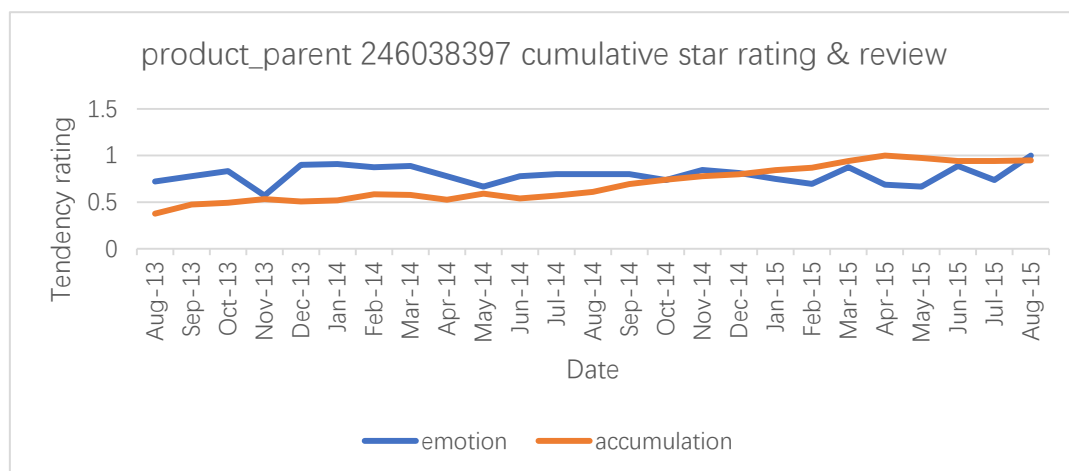

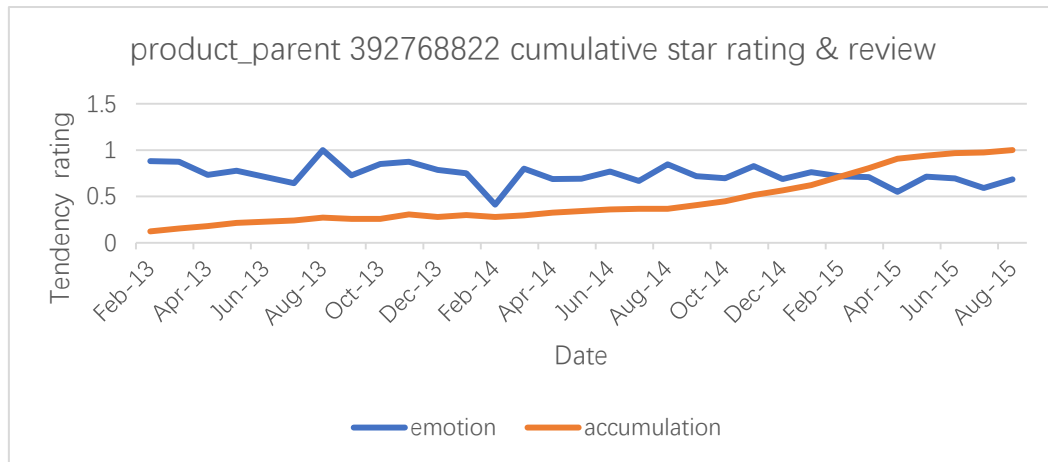**Figure 2. Cumulative Score and Emotion Analysis**

134

Figure 3. Cumulative Scoring and Emotion Analysis

As can be seen from the figure, on the whole, the text emotion of the two commodities is positive, and the emotional tendency of each month is greater than 0, indicating that consumers have been very fond of the two commodities. With the increase of cumulative star rating, the emotion of each month is always inclined to the positive side, indicating that the previous accumulation of high rating will cause consumers in the current month to be inclined to the positive emotion of the product. Therefore, this paper also concludes that the low score based on the historical accumulation will make consumers tend to have negative comments on the product in the current month.

4.4 The Relationship between Rating and Special Text

After the analysis of the influence of cumulative score on text comments, this paper further analyzes the relationship between score and special text. In order to ensure the validity and representativeness of data, the comments of verified buyers and green label reviewers are selected. Then, positive emotional correlation analysis was conducted for comments with a star rating of 4 or 5, negative emotional correlation analysis was conducted for comments with a star rating of 1 or 2, and correlation analysis was conducted for comments with an overall sample of star ratings and emotions. The results are shown in Table 6.

Table 6. Results of Correlation Degree

TYPE	Grey Relative Correlation	Grey Absolute Correlation	Comprehensive Correlation
review sentiment and overall star rating	0.6938	0.7581	0.7259
negative emotion reviews and star rating	0.6806	0.8464	0.7635
positive emotion reviews and star rating	0.7086	0.7421	0.7244

From the Comprehensive Correlation degree, it can be seen that both positive emotion and negative emotion have a high correlation degree with star rating, and the correlation degree between negative emotion and star rating is relatively higher than that between positive emotion and star rating. The Gray Relative Correlation reflects the correlation degree of the sequence change trend, from which it can be seen that the change of positive emotion is more closely related to the change of star rating.

4.5 Product Reputation Measurement Model

4.5.1 Principal Component Model

The principle of principal component analysis is to try to recombine the original variables into a new set of several integrated variables that are unrelated to each other, and at the same time, according to the actual needs, a few less sum variables can be taken out to reflect the information of the original variables as much as possible. The specific steps are shown in Figure 4.
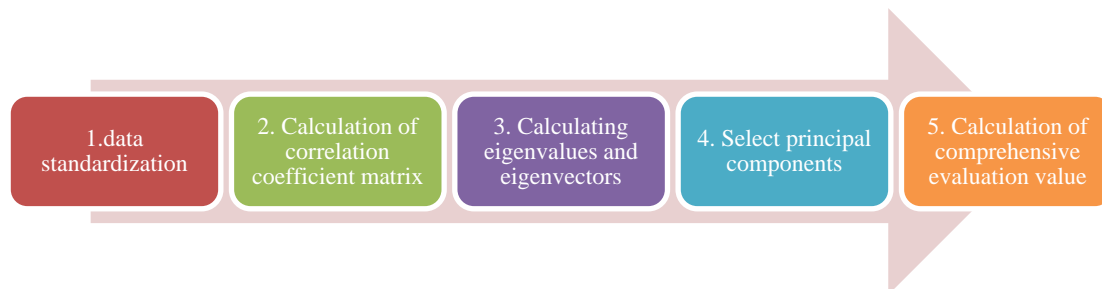

Figure 4. Flow Chart of Principal Component Analysis

Based on the comprehensive evaluation method of principal component analysis (pca) mentioned above, this paper establishes a principal component model to evaluate the reputation of commodities by using the eight variables of 5 levels of star rating, number of helpful comments and emotion analysis. In this paper, the first few characteristic roots of the correlation coefficient matrix and their contribution rates are obtained based on the blower index data, as shown in Tables 7 and 8.

Table 7. Component Matrix

	composition				
	1	2	3	4	5
Zscore(r1)	-.372	.339	.619	-.505	.327
Zscore(r2)	-.465	.400	-.595	.397	.332
Zscore(r3)	.099	.441	.547	.625	-.324
Zscore(r4)	.750	.387	-.352	-.395	-.074
Zscore(r5)	-.232	-.965	.016	.024	-.115
Zscore (helpful comment rates)	.532	-.262	.211	.298	.716
Zscore (emotional)	.983	-.108	.056	.080	-.045

Table 8. Total Variance Interpretation

composition	Initial eigenvalue			Extract the sum of the squared loads		
	total	Percentage of variance	cumulative %	total	Percentage of variance	cumulative %
1	2.229	31.848	31.848	2.229	31.848	31.848
2	1.632	23.318	55.165	1.632	23.318	55.165
3	1.209	17.264	72.430	1.209	17.264	72.430
4	1.055	15.078	87.508	1.055	15.078	87.508
5	.856	12.230	99.738	.856	12.230	99.738
6	.018	.262	100.000			
7	-1.951E-16	-2.787E-15	100.000			
Extraction method: principal component analysis.						

It is known that the cumulative contribution rate of the first four indexes has reached 87.50%, so the first four indexes are selected as the main components in this paper to obtain the following expression:

$$A_1 = -0.210R_1 - 0.170R_2 - 0.184R_3 - 0.077R_4 - 0.358R_5 + 0.095R_6 + 0.383R_7$$
$$A_2 = -0.310R_1 - 0.088R_2 - 0.113R_3 + 0.692R_4 - 0.305R_5 + 0.150R_6 + 0.124R_7$$
$$A_3 = 0.434R_1 - 0.698R_2 + 0.250R_3 + 0.066R_4 - 0.019R_5 + 0.332R_6 - 0.003\ R_7$$
$$A_4 = -0.468R_1 - 0.159R_2 + 0.691R_3 - 0.207R_4 - 0.053R_5 + 0.465R_6 + 0.038R_7$$

Then, according to the contribution rate, the total expression of reputation is as follows

$$F = 0.36A_1 + 0.20A_2 + 0.16A_3 + 0.14A_4$$

Similarly, by the same method can be obtained ID number of the microwave oven for 423421857, ID for 450475749 pacifier reputation expression.

Reputation value of microwave oven $F = 0.34A_1 + 0.27A_2 + 0.16A_3 + 0.11A_4$

Reputation value of pacifier $F = 0.38A_1 + 0.18A_2 + 0.16A_3 + 0.15A_4$

4.5.2 Product Reputation Analysis

At the same time, according to the principal component measurement model, it can also be found that the reputation value of goods is positively correlated with the positive score, the number of helpful comments and the text emotion, and negatively correlated with the negative score.

According to the above expressions, the reputation values of three products in recent years are calculated respectively, and the change of reputation of commodities in the future is analyzed with grey prediction, and the size of its average reputation value is used to determine whether it is a successful commodity or a failed commodity.

This paper analyzes the data of many products in the three series of hair dryer, pacifier and microwave oven. The results are shown in Table 9.

Table 9. The Results of Reputation

Product	Product parent	Average reputation	Reputation value trend
The microwave oven	423421857	-0.14391	rising
Hair dryer	758099411	0.099073	falling
The pacifier	450475749	0.0873	rising

According to the above analysis, this paper found that although the microwave oven reputation value is low, but its reputation value of the trend of change in the rise. So it has a certain appreciation of space. For hair dryer, although the average reputation value is larger, but it presents a downward trend. Therefore, it can not be considered a success. However, for the pacifier, the paper found that its average reputation value is large, and presents an upward trend, is a typical successful commodity.

5. Policy Suggestions

5.1 The Businessman

1. Businesses can track users' comments in real time, timely follow up and deal with consumers' negative comments, and minimize the impact of negative comments and negative comments. According to the analysis in this paper, the accumulation of high ratings can greatly affect consumers' impression of the product, while the accumulation of low ratings can lead to more reviews, similar to the Matthew effect.

2. Merchants can establish a commodity reputation evaluation system and a product sales forecast system to analyze consumers' shopping preferences and timely push marketable products. Based on the product's historical sales information and reputation information, you can make inventory preparation in advance to deal with the situation of out of stock.

3. In any case, producing high quality products at low prices is the only way to win over consumers. Besides commodity quality itself, producer still can establish after sale service system to dispel the doubt of consumer, win the favour of consumer. In the era of mobile Internet, merchants can take advantage of the traffic dividend, establish a social economy model, and enhance customer loyalty through the interaction with consumers.

5.2 Platform

1. Establish a mechanism to identify malicious comments and false comments. Avoid merchants to pursue the ranking of stores and goods to make false comments, disturb the market order.

2. Establish a scientific three-level credit evaluation system. For each transaction on the platform, the platform can enrich the evaluation indicators, such as product quality, merchant attitude, logistics satisfaction, etc. For the validity of the comment, the transaction situation, transaction value and the recognition degree of other users can be considered comprehensively. For the credit of the business, the comprehensive consideration of its sales, praise rate.

3. Appropriately invite senior customers to comment on the product for evaluation, and provide reference evaluation for the product. Independent from merchants and platforms, commentators have strong objectivity and authenticity in their comments, which helps platforms and consumers to identify the quality of goods.

4. Establish the entry threshold of the mall, conduct qualification test for the settled merchants, regularly check the business license of the merchants, and timely clear the unqualified merchants.

5.3 The Government

The country should face up to the rapidly developing electronic commerce and make perfect and effective laws and regulations. The government can strengthen the real-name system and the management of credit files to eliminate the hidden dangers caused by the virtual and anonymous nature of online transactions, so as to safeguard the legitimate rights and interests of consumers, producers and platforms.

6. Conclusion

To sum up, this paper mines consumers' text comments based on online comment data, and measures consumers' text emotions by combining supervised and unsupervised classification models. Based on the emotion analysis model and the scoring data of commodities, the evaluation model of commodity reputation is established, which provides a basis for consumers to identify superior products and inferior products, provides a basis for producers to track consumers' preferences to develop marketing strategies, and provides constructive Suggestions for online shopping platforms to build a perfect trading environment.

References

Chen, C. (2012). A study on the impact of contradiction on consumers' purchasing behavior—Based on consumers' perceived usefulness and product search. *Fujian Tea*, *42*(02), 54-56.

Cui, Y. S. (2018). Research on the impact of online comment text mining on e-commerce. *China Business Theory*, *33*, 17-23.

Guo Q. Q. (2019). *Research on the influence of online review characteristics on product sales*. Shanxi University.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. ICLR Workshop.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). *Distributed Representations of Words and Phrases and their Compositionality*. NIPS.

Si, S. K., & Sun, Z. L. (2016). *Mathematical modeling algorithm and application* (Version 2). Beijing: national defense industry press.

Sun, H. F., & Yu, S. L. (2018). Review text mining of online home goods online shopping based on R language—A Case study of amazon. *Popular Literature and Art*, *22*, 248-249.

Ye, K. (2016). *Research on online comment topics and feature extraction based on word vectors*. University of Electronic Science and Technology.

Zhao, F. F. (2018). *Analysis on the usefulness of online customer reviews in amazon*. Guangdong University of Foreign Studies.

Zhou, Z. H. (2016). *Machine learning*. Beijing: Tsinghua University Press.