ISSN 2522-6398 Vol. 3, Issue 7: 68-71, DOI: 10.25236/FER.2020.030720

Research on Online Shopping Comments Based on Text Emotional Comments

Chenhao Ma[1], Shuxian Song[2], Shun Xu[3], Shangrui Xiao[4]

1 College of Information Science and Technology, Dalian Maritime University, Dalian Liaoning 116000, China
2 Maritime Institute, Dalian Maritime University, Dalian Liaoning 116000, China
3 College of Naval Architecture and Ocean Engineering, Dalian Maritime University, Dalian Liaoning 116000, China
4 College of Safety and Ocean Engineering, China University of Petroleum (Beijing), Beijing 100050, China

ABSTRACT. *In order to draw up Sunshine company' s online sales strategy, enhance customers' satisfaction to the products, and predict Sunshine's future sales trends, we do the following work:First, during the performance of data preprocessing, we consider the data labels as variables, process abnormal data, and select meaningful variables based on data redundancy and association principles. Then we regularize the text review, calculate the frequency of emotional words, words-modifiers and negation words and analyze the emotional level of the review body qualitatively. By developing the correlation analysis model, and using the Spearman coefficient to measure the correlation between different variables, we find that there is a strong correlation between helpful votes and the total votes.Secondly, we group and rank emotional words and words modifiers, assign negation values to -1, consider the length of the text-review and the number of times that negation words appear to build the text-review emotion quantification model to calculate the emotional values of each review. And normalizing emotional values which range [0, 5]. After that, with the linear weighted sum method between the star ratings and the emotional values, we obtain a new score called* $score(i)_{new}$*, and refer to the literature for a set of reasonable weight values.*

KEYWORDS: *Text- review emotion quantification model, Correlation analysis, Online shopping*

1. Introduction

With improvements in network technology and electronic commerce in these days, online shopping has become more and more popular. As we all know, customers will face a choice problem due to the large variety of online products and the uneven quality of products when browsing goods, this makes major shopping platforms provide online review mechanisms for consumers [1]. We have witnessed the huge development of online consumer reviews. Online reviews are widely accepted to play an important role in influencing customers' choice. On the one hand, rich user reviews contain abundant useful information to afford references for other users who want to purchase this product [2]. On the other hand, the merchant can use customers' feedback to improve the goods and develop a product sales plan, and they can also provide efficient after-sales service for customers. But the time cost and difficulty are high when we faced with a large and complex number of comments [3]. This paper aims to solve this problem by building a suitable model based on users' data.

2. Data Preprocessing

2.1 Modifying Anomalous Data

There are some abnormal data in the three provided data sets, which may have a disturbing effect to the solution of the mathematical model later. Therefore, we correct the abnormal data in the data set sat first step, the grubbs rules is used to add, delete and supplement data items.

2.2 Text-Review Analysis

We take the hair dryer's data as a representative to start to analyze the text-review which also called review body. At first step, we establish the LDA model to vectorize the review text and remove the numbers and punctuation marks. We utilize Matlab’s tabulate function to count the emotional words, modifier words and

ISSN 2522-6398 Vol. 3, Issue 7: 68-71, DOI: 10.25236/FER.2020.030720

negative words. And make statistics on the results.

We extract a number of emotional words such as great, hard, etc from the text reviews, and take measures to count the frequency of these emotional words. We can get some information: the emotional words great has a highest frequency of occurrence, and amazing and expensive appears least; Lexical words is far outweigh than derogatory words. Hence we can make a preliminary conclusion. In the evaluation of the hair dryer, because the high-score review is more than the low-score review, the performance of the hair dryer is excellent and the sales prospect is good.

After analyzing the emotional vocabulary, we use the same method to analyze the modifier words and negative words, and we calculate the top 10 vocabulary with the highest frequency of occurrence: The negative word “not” appears most frequently, and the modifier word “fairly” which has lowest occurrence. But we all know that due to the grammatical problem, the high number of negative words does not mean that there are many negative comments, so analyzing that alone cannot reach a clear conclusion. Only by combining various parts of speech and judging the position and relationship between words, can a sentiment analysis be performed on a review, but we can roughly get the overall evaluation trend of the hair dryer through the two figures plot, that is, the good score is more than bad rating.

2.3 Correlation Analysis

The first thing we should point out at first is that at this step's correlation analysis, the five variables that we adopt do not include the product parent. This is because we have calculated that the repeat rate of the product parent is less than 1% in the previous steps, and it does not cause much impact on the problem analysis.

The description statistics is as follows:

Table 1 Descriptive Statistics

	N	Minimum	Maxximum	Mean	Std.Deviati	Variance
star rating	11470	1	5	4.116	1.30033	1.691
helpful votes	11470	0	499	2.1791	14.2413	202.815
total votes	11470	0	575	2.5633	15.38253	236.622
vine	11470	0	575	2.5633	15.38253	236.622
verified purchase	11470	0	1	0.8554	0.35175	0.124

At this step we use Spearman correlation model to measure the correlation between variables.

The Spearman formula is:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

Where ρ is spearman rank correlation coefficient; d_i is the rank difference between the observed variables; n is the sample size

We use the SPSS software to get the relationship between the five variables. We made a matrix diagram of these five variables. If the dot clusters in the figure are closer to a circle, it indicates that there is no correlation between the corresponding variables; If slopes are on a straight line to the right, we can get that there is a strong correlation between the variables.

We use the Spearman coefficient to perform a significance test on the selected variables to gain the data. The data is the spearman rank correlation coefficient p, where $\rho \in [-1, 1]$. The closer the absolute value is to 1, the stronger the correlation between the variables, the superscripted number in the numerical value also indicates strong significance. We initially assumed that ρ is 0 (the null hypothesis). After performing a significance test, we found that except for the weak correlation between vine and star rating, the values of the other variables are 1 or have two stars. The value of the number was found to have a strong correlation after the significance test.

3. Emotional Value Quantification

3.1 Grouping Words

●Grouping Emotional Words

ISSN 2522-6398 Vol. 3, Issue 7: 68-71, DOI: 10.25236/FER.2020.030720

Emotional words are words that express positive or negative attitudes, we need to group them according to their nature and degree of emotions. we have extracted the emotional words in each review by reference. The scores' range of the emotional words is clearly. It is obvious that the range of scores for emotional words depends on the accuracy required for sentiment analysis, the more detailed the analysis required, the wider the score should be.

●Grouping Modifiers-word

There are two main groups of words modifiers:

The first modifiers-words which increase emotional assessment, for example the words like (very, fairly, highly, very, even, etc.);

The second modifiers- words which decrease emotional assessment, the words like happened, (happened, not really, not really, etc.);

For example, if the word “good” refers to a positive assessment, the word-modifier “very” increases the positive assessment, the result is' very good”. In turn, if we put the word not” before this phrase, we get a restrained negative assessment of “not very good”. That is way it is important to take into account the influence of words-modifiers on the emotional colour of the review.

3.2 Review on the Calculation of Emotional Value

Based on the above analysis, we have obtained the formula for calculating the emotional score of modifiers-words and negatives as follows:

$$degree(s_i) = degree(ad) * (-1)^n$$

Where $degree(s_i)$ shows intensity values for words-modifiers and negatives; $degree(ad)$ shows polarity intensity of words-modifiers; n is negatives’ occurrences number.

Table 2 the Intensity of Feature Words

Grade	Description of word characteristics	Examples of words in the dictionary
*	**emotional characteristics**	
1.5	vocabulary for highly rated products	perfect,excellent
1.3	vocabulary for positive product evaluation	great,favorite,good,love
1.1	vocabulary for positive reviews	nice,like
1	words that objectively describe product characteristics	soft,clean
0.9	glossary of minor dissatisfaction with the product	expensive,complicated
0.7	extremely negative wordsnfor products	worthless
0.5	vocabulary for product authenticity/integrity	fake,broken
*	**degree adverbl characteristics**	
1.6	express extremely strong words	definitely,absolutely,especially
1.4	more prominent words	very,highly
1.2	words indicating slightness	a little,a few
0.8	weaker words	just,only

The emotional value of a review is mainly related to emotional words, modifiers-words and negative words. In addition, the length of the comment also affects the emotional value of the comment.

The longer the comment length, the deeper the emotional value. We get the calculation formula

$$degree(f_i) = degree(e_i) * degree(s_i) * length(i)$$

Where $degree(f_i)$ is emotional value of review I; $degree(e_i)$ means the score of emotional words; $degree(s_i)$ is the score of words-modifiers and negative words; $length(i)$ is the length of a comment. The rating of the product is a five-star system.

We need to normalize the emotional value of the review body and the value ranges from zero to five. We get the formula

$$degree(f_i)_{normal} = \frac{degree(f_i) - degree(f_i)_{min}}{degree(f_i)_{max} - degree(f_i)_{min}} * 5$$

In question, by a significant test we found that there is a strong linear correlation between helpful votes and total votes, so we choose the number of helpful votes to modify the score in the new scoring model; the vine and discount purchase indicators affect the credibility and accuracy of reviews. Among the three indicators we have mentioned, the helpful votes have a fuzzy relationship with the ratings of the reviews and the review body, and

ISSN 2522-6398 Vol. 3, Issue 7: 68-71, DOI: 10.25236/FER.2020.030720

the vine indicator and verified purchase affect both the ratings and reviews of a review. We build a model of these five variable indicators in two steps.

First we create a new scoring model for each review.

$$score(i)_{new} = \omega_1 * score_i + \omega_2 * degree(f_i)_{normal}$$

$$\omega_1 + \omega_2 = 1$$

Where i is the i_{th} review; $score_i$ is star rating value range [0,5]; $degree(f_i)_{normal}$ is the emotional value of the review body; ω_1 and ω_2 is the weight of the star rating and the emotional value of the review body respectively; $score(i)_{new}$ is the new rating of the i_{th} review combining the star rating with the emotional value.

3.3 The Consequence

Through reading references [3], we get the weight of ratings ω1 and the weight of reviews ω2, and a new rating model for each review is as follows:

$$score(i)_{new} = 0.458 * score_i + 0.542 * degree(f_i)_{normal}$$

References

[1] Lutfullaeva, Malika, Marina Medvedeva, Eugene Komotskiy (2018). “Optimization of Sentiment Analysis Methods for classifying text comments of bank customers.” IFAC-PapersOnLine, vol.51, no.32, pp. 55-60.

[2] Xiaoyun Wang, Lingling Shi (2016). Commodity comprehensive scoring model based on sentiment quantification of online reviews. Journal of Hangzhou Dianzi University (Social Science Edition), vol.12, no.3, pp.8-15.

[3] Wang Longge, Wang Min (2019). A quantitative emotional scoring model based on user comments. Digital Technology and applications, vol.6, pp.71-72. .