Research on online shopping behavior based on Long Short-Term Memory and Latent Dirichlet Allocation

Haochen He[1,a], Liang Xie[2,b,*], Lu Liu[3,c], Zhihan Xu[4,d]

[1]School of Automation, Wuhan University of Technology, Wuhan, 430070, China

[2]School of Science, Wuhan University of Technology, Wuhan, 430070, China

[3]School of Economics, Wuhan University of Technology, Wuhan, 430070, China

[4]School of Management, Wuhan University of Technology, Wuhan, 430070, China

[a]HaochenHe@163.com, [b]whutxl@hotmail.com, [c]liulu0929@whut.edu.cn, [d]783021850@qq.com

*Corresponding author

Keywords: Word2vec, Long Short-Term Memory, Sigmoid function dynamic weight factor evaluation model, Latent Dirichlet Allocation.

Abstract: In this paper, we use natural language processing and machine learning methods to extract product and user characteristics from product descriptions and reviews to establish the sigmoid function dynamic weight factor evaluation model to evaluate the product's success in the market. When establishing the evaluation model, we fully consider the star rating, helpfulness vote, view submitted by the customer. When carrying out quantitative processing of the view, Nltk, Word2vec and Long Short-Term Memory (LSTM) are used to extract the emotional polarity of the text. We compare the results of the model with the actual situation to verify the accuracy of the model. In addition to the above work, we also use the Latent Dirichlet Allocation (LDA) model to investigate whether some words in the view submitted by customers are clearly associated with star rating.

1. Introduction

Nowadays, the e-commerce platform with high-speed development fostered a new kind of business model in the Internet environment–O2O(online to offline), which enables offline sales to be completed online. Through this mode, the trade can be more convenient, moreover, large amount of data are generated as well, such as order information, reviews and soon.

Allowing users to share views freely and acquire some useful information online, the evaluation mechanism of Amazon plays the role of social media rather than a business platform [1]. For sellers, the feedback from the market reflects how satisfied the buyers are with their expected learning, so that they can adjust their marketing strategies, improve their word of mouth, and ultimately make a profit.. Therefore, how to apply these data reasonably and effectively is a problem deserve further study. This paper will use this data to build a model to evaluate product competitiveness to help decision makers make better decisions.

Singh,S.K. and Sachan,M.K. [2] carried out text sentiment analysis through the Lexicon-Based Approach method and the Laplacian Smoothing calculation algorithm which generated a rate that would represent a negative and a positive sentiment, to analyze the sentiments expressed by users replies on the Twitter; Ssumya, Sunil Singh et al. [3] used random forest classifier and gradient enhanced regression to extract features from product review text data, product description data and customer question and answer data, and to predict product help degree scores.

2. Data set introduction

The three data sets cited in this article are records of sales of microwave, baby pacifiers, and hair dryers on the Amazon market. Each record includes the brand of the product purchased, the star rating (1 to 5, with 1 being the lowest and 5 the highest) submitted by the buyer at the time of purchase, and views (views are text-based messages, which express further opinions and information about the product) , helpfulness rating (Other customers can submit ratings on these reviews as being helpful or not - called "helpfulness rating" -towards assisting their own product purchasing decision.) . All three data sets are downloaded from the American college students mathematical modeling contest website. (http://www.comap.com).

3. Sigmoid function dynamic weight factor evaluation model

When establishing a comprehensive model for evaluating the success of products, we fully consider the star ratings, views and helpfulness votes submitted by users in the data set. Among them, star ratings and helpfulness votes are given in the form of quantization, while views are given in the form of text. We first use NLTK to complete word tokenization, and then obtain Word vectors through Word2vec. Finally, we use LSTM to obtain the emotional polarity of the view, which is the quantification of the view. When testing the accuracy of the model, we conducted two kinds of ordering on the products, one was in the order that the R value obtained by the model we built was from large to small, and the other was in the order that the sales volume of products was from large to small, comparing the two kinds of ordering to verify the accuracy of the model.

3.1 Word tokenization

We use NLTK method to implement the word tokenization. NLTK stands for Natural Language Toolkit [4]. NLTK is a Natural Language processing Toolkit for Python, a module developed on Python by Steven Bird and Edward Loper of the university of Pennsylvania. It is widely used in NLP problems. Compared with other classification methods such as jieba, snownlp, etc., NILK is more effective in processing English texts.

3.2 Obtain the word vector

Word2vecor was proposed by Tomas Mikolov et al. [5]. basis on NNLM. Word2vetor is a neural network with only one hidden layer. Map the word to an n-dimensional space, using a low-dimensional, dense word vector to represent the word. It is widely used in natural language processing, classification feature processing and other fields. The following Figure 1 shows the topology of the Word2vector.

The input layer of Word2vec is in the form of one-hot encoder. After inputting a one-hot encoder of a word, we can get a weight vector from the input layer to the hidden layer, and this weight vector is the corresponding word vector.

3.3 Obtain the polarity of each view

Recurrent Neural Network (RNN) is a kind of Neural Network used for processing sequence data [6]. Compared with ordinary network, it can process the data of sequence change. Long short-term memory (LSTM) is a special RNN. The structure of LSTM is very complex, composed of weighted input, activation function and so on. LSTM is mainly proposed to solve the problems of gradient disappearance and gradient explosion in the process of long sequence training. LSTM performs better when dealing with longer sequences.

Figure 1 shows here is long short-term memory. It was named LSTM because the program uses a structure based on short-term memory processes to create long-term memory. First notice that LSTM connect the new word sequence value x_t to the previous output of cell h_{t-1}.

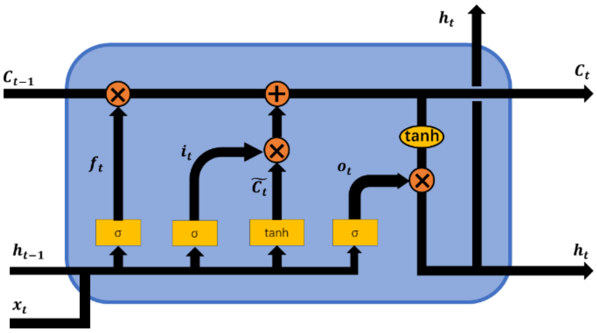

Figure 1. Graphical model representation of LSTM

The first step in LSTM is to compress the input h_{t-1} and x_t and decide which information to discard from the cell state. This step is realized by a forgetting gate. When the output is 1, it means to keep completely, and when the output is 0, it means to discard completely. The sigmoid function outputs values between 0 and 1 and gives the output value to the number in state C_{t-1} of each cell, so the input weights connected to these nodes can be trained to "discard" (or "retain" the output value as close to 0).Here we have:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

The next thing LSTM needs to do is decide to keep that new information in the cellular state. The operation here involves two steps: first, by a entering gate layer (which is also the sigmoid function), determine which values will update. Second, a new candidate value vector C_t is created by a tanh function, and C_t will be added to the state. There are two formulas involved:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = \tan g(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

By multiplying the old state C_{t-1} by f_t, LSTM have done the job of throwing away some of the information. The product is then added to the $i_t * \tilde{C}_t$ to retain the new candidate values. To complete the cell status update. Here it is:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$

Through the previous preparation, the LSTM will determine what value to output, which will depend on the cell state C_t.First, LSTM use a sigmoid layer to determine which part of the cell's state will be output. Next, LSTM processed the cell state by tanh, multiplying the output value by the sigmoid output. You get the output value. Here we have:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

3.4 Sigmoid function dynamic weight factor evaluation model

Through the previous work, we have obtained the text polarity of all the comments. Next, we will establish the Sigmoid function dynamic weight factor evaluation model to calculate the Comprehensive competitiveness of different brands of microwave oven, hair dryer and baby pacifiers. The calculation formula of R is as follows:

$$R = [1-\lambda(hv_i)] \cdot \frac{1}{n_1}\sum_{i=1}^{n_1} s_i + \lambda(hv_i)\frac{1}{n_2}\sum_{j=1}^{n_2} p_j \tag{7}$$

Where R represents the comprehensive competitiveness of a certain type of commodity, s_i is the star rating of the i user of the brand, p_j is the text polarity of the j normal user comment of the brand, $\lambda(hv_i)$ represents the variable weight function, n_1 represents the total number of star rating of the brand, and n_2 represents the total number of user comments of the brand.

The independent variable of the variable weight function $\lambda(hv_i)$ is the helpful vote quantity of the brand products, which is essentially the sigmoid function.

$$\lambda(hv_i) = \frac{1}{1+e^{-hv_i}} \tag{8}$$

It can be seen that $\lambda(hv_i)$ can map the number of valid comments to the semi-probability space of [0.5,1).

4. Further study of consumer behavior

In addition to build a model to evaluate the product competitiveness in the market, the enterprises tend to focus on star ratings, views, helpfulness vote, and so on factors between whether there is a certain relationship between the quantitative or qualitative, if we can find some relations, will provide decision-makers with more help. Here we examine whether there is a pure relationship between certain words in the view and star rating. We take out two extreme emotional polarity texts (star 1 and star 5) and mix them, and then use LDA to investigate the effect of words on extreme star texts in an unsupervised environment.

4.1 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a probability generating model in which review d obeys K-dimension theme distribution θ, and the probability that the first word in the review belongs to the topic k is φ. The dimension reduction of vector space is realized by introducing theme dimensions between the text and words[7]. In this paper, LDA model is applied to the analysis of review information to find potential topics in consumer's review for the products. the general flow is shown in Figure 2.

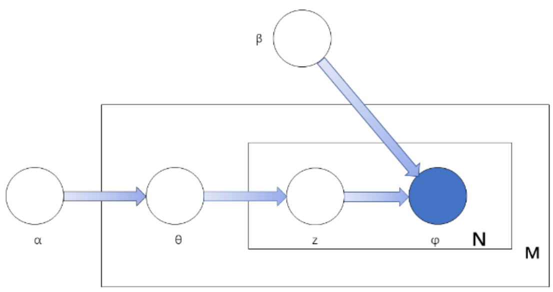

Figure 2. Graphical model representation of LDA

As shown in Figure 2, LDA reveals the hidden dimensions in the review text. Each text is considered as a sequence of n words, M denotes the number of the text in the text set, z denotes the theme distribution of the words, α and β denote the super parameter of the theme distribution

θ and word distribution φ in the text, respectively and obey the prior Dirichlet distribution. Given the word distribution φ_k and the theme distribution $z_{d,j}$ for each word, the possibility of belonging to the specific text corpus can be expressed as:

$$p(T \mid \theta, \psi, z) = \prod_{d \in T} \prod_{j=1}^{N_d} \theta_{z_{i,j}} \psi_{z_{i,j}, w_{i,j}} \tag{9}$$

Where, the parameter $\psi = \{\theta, \varphi\}$, N_d denotes the number of words in the review d, $w_{d,j}$ denotes the numbered j word in the review numbered d of product numbered i. Here, we multiply all the text in the corpus and all the words in each text.

5. Experiment

5.1 The experiment of sigmoid function dynamic weight factor evaluation model

To train LSTM, we divide all comments into three subdata sets:

- Trainset: Use 60% of the dataset as training data.
- Validset: Use 20% of the dataset as valid data.
- Testdata: Use 20% of the dataset as testing data.

Where, the word vector obtained by each comment is the input, and the star rating is the label.

Figure 3 shows the change of loss value with iterations in the training process when Adam optimizer is adopted. Through training, it can be found that after 40,000 iterations, the loss function value decreases to about 0.012.

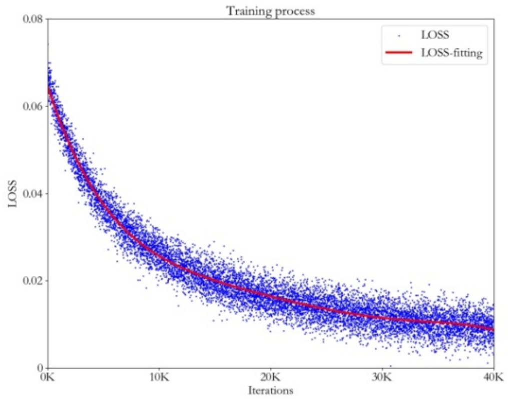

Figure 3. Training process

Figure 4. The relationship between star rating and review polarity

When the test set was brought into the trained LSTM network, the accuracy was 92.6%. At The same time, the experimental results show that there is a strong correlation between the rating and The polarity of The review text.

Through the previous work, we have obtained the text polarity of all the comments. Next, we respectively calculate the R value of all brand data of each commodity and sort them according to the R value. At the same time, we give the average star of its sales volume. In the Table.1. We list eight typical types of pacifier.

Table.1. Indexs of ten types of pacifier

Brand number	Value of R	The sales amount	The average star rating
106521272	10.2	183	5.3
52203997	9.8	81	4.6
28435092	7.2	53	3.9
6450021	6.4	44	3.4
241607501	5.1	26	2.9

It can be seen that after the product are arranged from large to small according to R value, the sales also show a trend of decreasing from top to bottom, shows that our evaluation method can well the success of said products, at the same time, we found that the high R value of types of products: 1, their polarity text comprehensive index is more big, popular commodity, their comments will be more "active" performance.2. They all have high average stars.

5.2 The experiment of LDA

Experimental environment: The numpy (array operations) module and sklearn (machine learning) module of Python.

Data Set: All the comment text in the valid comment is 1 star and 5 star after mixing the three data sets.

This task focuses on the weight of words rather than natural language, therefore, the Bag-of-Words is constructed and the words are quantified as the input of LDA. In the Super parameter, we set the number of categories to 2 and the number of iterations to 100.

The output results are shown in Table.2. and Table.3..

Table.2. The output results of the top five in the model

Weight sorting	word	weight	category
1	Great	528.08	5stars
2	Love	454.56	5stars
3	One	375.61	5stars
4	Good	315.2	5stars
5	really	320.71	5stars

Through the experiment result, it can be found that LDA could effectively identify the difference of the text in the emotional topic, and the words output can better reflect the category of the text after classifying the text according to the theme. In the category of 5 star, "great" and "love" could highlight the positive polarity of emotion, but "one", as a neutral word, has no emotional color. This is due to the extent to which other statements can be used to express positive emotions. In the same way, "really" is a better way to enhance your emotions. What is surprising is the verb "waste" is the

No.1 in 1 star category, that maybe results for a bad product experience may lead to a waste of time, money, energy and many other situations. In addition, "disappointed" ranked second in the title, and the following words also reflect consumers’ negative shopping mentality. Besides, the "enthusiastic" didn’t appear in the top ten. After analyzing, we have found that consumers always tend to use short commonly used colloquial words to describe products, so that their comments on products are usually colloquial. It may be also the reason of that complex words, despite their strong emotional polarity, still do not appear in the list of featured words.

Table.3. The output results of the worst top five in the model

Weight sorting	word	weight	category
1	Waste	476.12	1stars
2	Disappointed	434.27	1stars
3	Useless	386.24	1stars
4	Defective	338.06	1stars
5	poor	310.97	1stars

6. Conclusion

The sigmoid function dynamic weight factor evaluation model established in this paper, using Nltk, word2vec, LSTM method extract text polarity, star ratings and helpfulness rating for each brand products, to mark size sorting, and with sales as an index to the size of the order for the sorting and compared to verify the effect of the comprehensive evaluation model, the sorting result is very close, our evaluation model effect is good. When studying whether reviews are closely related to star rating, we found a strong relationship between them and found the words with the strongest correlation between the highest and lowest stars.

References

[1] Danescu-Niculescu-Mizil C , Kossinets G , Kleinberg J , et al. How opinions are received by online communities: A case study on Amazon.com helpfulness votes[J]. 2009..

[2] Singh S K , Sachan M K . SentiVerb system: classification of social media text using sentiment analysis[J]. Multimedia Tools & Applications, 2019, 78(22):32109-32136.

[3] Saumya S , Singh J P , Baabdullah A M , et al. Ranking online consumer reviews[J]. Electronic Commerce Research & Applications, 2018:S1567422318300358.

[4] Wagner W . Steven Bird, Ewan Klein and Edward Loper: Natural Language Processing with Python, Analyzing Text with the Natural Language Toolkit[J]. Language Resources & Evaluation, 2010, 44(4):421-424.

[5] Mikolov T . Distributed Representations of Words and Phrases and their, Compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26:3111-3119.

[6] Greff K , Srivastava R K , Jan Koutník, et al. LSTM: A Search Space Odyssey[J]. Neural Networks and Learning Systems, IEEE Transactions on, 2017, 28(10):2222-2232.

[7] Blei D M , Ng A Y , Jordan M I , et al. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003, 3:993-1022.