Evaluation and Emotional Analysis of Mobile Phone Sales of JD E-commerce Platform Based on LDA Model

Jingfeng Xue*, JingLi, YujiaHan

Big Data College, Qingdao Huanghai University, Qingdao, Shandong, China
* xuejf@qdhhc.edu.cn

Abstract. With the rapid development of the Internet economy, the online transaction volume of e-commerce platforms has soared, and the purchase experience shared by users has become particularly important. How to quickly and effectively extract the required information from a large number of user reviews plays an important role in the decision-making of both consumers and businesses. This article conducts sentiment analysis on JD business platform Huawei mobile phone user comment information. First, this paper uses crawler methods to collect user evaluation text information, and perform data preprocessing such as deduplication and cleaning of the collected raw data; then perform word segmentation on the processed text content, and filter stop words and worthless words; use the method based on emotional dictionary matching to analyze the sentiment tendency of the text. Finally, using the LDA topic analysis model, LDA topic analysis is performed on the sentiment classification text, and the topic words are extracted to obtain valuable internal information, and reasonable marketing strategies are proposed for the business.

1. Introduction

In the fierce market competition, whether a user chooses to buy a certain product is influenced by many factors, advertising and user reviews are very critical. If we can analyze and excavate the voice of users from a large number of user reviews, we will win the advance chance in the fierce market competition. As user text data is unstructured data, it is different from traditional business intelligence data mining. For consumers, they can learn about the products by reading reviews one by one manually. However, for product merchants, it is not realistic to only rely on manual processing in the face of massive amounts of data[1]. Therefore, it is necessary to realize intelligent collection and processing with the help of technical means so as to complete the analysis and decision-making process in a standardized way[2].

In this context, text mining, as a branch of natural language processing, can complete the emotional analysis of user comments and subject extraction through technical processing. On the one hand, businesses can understand the actual needs of users, so as to improve product performance and enhance product competitiveness. On the other hand, business use certain standards to truly display the reputation and brand image of the product in the entire e-commerce user group, and provide a scientific basis for users to rationally purchase a product[3].
2. Research design

2.1. Research frame

![Research frame](image1.png)

2.2. data extraction

First, data crawler. As a well-known domestic e-commerce platform, the JD e-commerce platform is an ideal platform for most users to express opinions and collect evaluation data. This article sets crawler means to complete data crawling, and selects Pycharm software for operation. The evaluation data selects user nicknames, product models, evaluation time, and evaluation content characteristics as analysis objects, and crawls 5000 pieces of data as text analysis objects\(^\text{[4]}\).

2.3. data preprocessing

The data used in real life is often some "dirty" data with missing, duplication, etc., and it is impossible to directly analyze and mine this type of data. Even the information obtained through analysis and mining is not accurate and has no reference value. Therefore, in order to improve the quality of data analysis and mining, data preprocessing technology is also applied. There are many kinds of data preprocessing methods, including: data cleaning, data integration, data transformation, data specification and so on. Preprocessing the original data before data analysis and mining through this kind of data processing technology not only helps to improve the quality of information obtained by data analysis and mining, but also reduces the complexity of the model and the time for data analysis. The results are shown in Table.1.

| number | De-reappraisal                                                                 |
|--------|-------------------------------------------------------------------------------|
| 1      | It's very perfect, it's very powerful, it's got a little bit of a tangle between the rear fingerprint, I didn't expect facial recognition. Unlocking is very convenient, and the rear fingerprint is also very easy to use and practical. |
| 2      | I thought it would be very difficult to get it originally, but I didn't think it would be easy to get it. The color is darker than the picture, and it might be better in the sun. |
| 3      | The screen is just the right size and looks comfortable. Fingerprint unlocking is quick and facial recognition is great, but photos are said to unlock them. |
| 4      | After thinking for a long time, I finally made up my mind to buy it. The last one is still used. It has been used for several years. |
| 5      | The girlfriend is Apple converted to Huawei, the reaction is not card, very smooth. |
| ...    | ...                                                                           |
2.4. comment data segmentation

In text mining, word segmentation needs to be performed on text data. The accuracy of word segmentation plays an important role in subsequent text analysis, especially in the process of feature selection, the effect of word segmentation will affect the selection of the final feature.

The word segmentation in this article is based on the stuttering word library in the python language. The comment text is segmented by stuttering segmentation, and the words after segmentation are marked (see Appendix 2).

| number | word segmentation |
|--------|-------------------|
| 1      | ("very","d")("perfect","a")("","","x")("performance","n")("very","zg")("powerful","a")... |
| 2      | ("originally","n")("thought","c")("be","v")("than","d")("difficult","a")("get","x")("","","a")... |
| 3      | ("screen","n")("size","b")("just","d")("right","a")... |
| 4      | ("think","v")("end","u1")("for a long time","m")("end","u1")... |
| 5      | ("girlfriend","n")("Apple","n")("convert","v")("Huawei","nr")("of","uj")("","","x")... |

After the data is segmented, the text data is initially segmented. Because the punctuation and some stop words are meaningless for text analysis, they are selected to be removed.

| number | word segmentation |
|--------|-------------------|
| 1      | Perfect a |
| 2      | Performance n |
| 3      | Powerful a |
| 4      | Before f |
| 5      | Tangle v |
| 6      | Rear n |
| 7      | Fingerprint n |
| 8      | Convenient a |
| 9      | ... |

The goal of this text analysis is to analyze the advantages and disadvantages of the product. Although comments such as "good, very good product", "very good, support", express emotional inclination towards the product. In fact, it is impossible to extract which product features are satisfactory to users based on these reviews. The comment is meaningful only when there are clear nouns in the comment, so it is necessary to extract the comments containing nouns according to the part of speech.

2.5. Emotional tendency analysis

2.5.1. Emotional words matching

In order to analyze the sentiment tendency of comments, firstly, the sentiment words are matched, and the matching method based on dictionary is mainly used. The positive and negative emotion words in
the reviews are matched through the dictionary, and the positive emotion words are assigned a value of 1, the negative emotion words are assigned a value of -1, and the non-emotional words are assigned a value of 0.

| Number | Word  | Part-of-speech | Emotional score |
|--------|-------|----------------|-----------------|
| 1      | Presell | vn             | 0              |
| 2      | Expensive | a             | -1             |
| 3      | Blue    | n              | 0              |
| 4      | Sound   | v              | 1              |
| 5      | Imagine | n              | 0              |
| 6      | Support | v              | 1              |
| 7      | Recharge | v             | 0              |
| 8      | Quickly | d              | 1              |
| 9      | ...     | ...            | ...            |

2.5.2. Correcting emotional tendencies
Emotional direction correction is mainly based on whether there are negative words in the first two positions of the emotional word to judge the correctness of the emotional value. Due to the multiple negation phenomenon in Chinese, it means that when the negative word appears an odd number of times, it means negation; when the negative word appears an even number of times, it means affirmative. According to Chinese habits, search for the first two words of emotional words, and if a negative word appears once, adjust the opposite emotional tendency.

| Table.5 Emotional score table

| Number | Word  | Part-of-speech | Emotional score |
|--------|-------|----------------|-----------------|
| 1      | Presell | vn             | 0              |
| 2      | Expensive | a             | -1             |
| 3      | Blue    | n              | 0              |
| 4      | Sound   | v              | 1              |
| 5      | Imagine | n              | 0              |
| 6      | Support | v              | 1              |
| 7      | Recharge | v             | 0              |
| 8      | Quickly | d              | 1              |
| 9      | ...     | ...            | ...            |

| Table.6 Dictionary of Negative Words (part)

Dictionary of Negative Words
never, no, no, non, not, unable, cannot, disuse, no...

| Table.7 Principle table of computational affective score

| Word | Thing | Good | Purchase | No | Regret |
|------|-------|------|----------|----|--------|
| affective score | 0 | 1 | 0 | 1 | -1 |
| computational affective score | 0 | 1 | 0 | 0 | 1 |

2.6. LDA Emotional thematic analysis
After the text emotion classification analysis is completed, the frequency of text characteristic subject words is analyzed from a statistical perspective. Semantic associations can be made through the LDA model[^5]. A document often contains multiple topics, and the frequency of some specific words that can represent different topics will be greatly increased. At this time, the LDA theme model can be used to find the regular features of words in the article, and summarize these rules. Finally it is presented in an unstructured form[^6].

In the mobile phone sales review text analyzed in this article, words representing mobile phone performance, such as pixels, price, cost performance, have relatively high frequency after text segmentation. Combining these relatively high-frequency features will provide an in-depth analysis of the focus of most users on mobile phones.

If the text sentiment is unclassified, doing LDA topic analysis will confuse the text, so the topic analysis should be done on the basis of the positive comments and negative comments in advance. In the previous step, the sentiment classification has been performed, and the result is "Positive sentiment result" and "negative sentiment result", and then perform LDA analysis on the two texts separately. This method will get higher precision analysis results[^7].

After segmenting positive and negative emotion comments, the LDA topic model is analyzed. The comment texts are finally clustered into two major topics. There are 8 most likely words and their cor-
responding probabilities under each topic. The following table shows the two potential themes analyzed in the mobile review text, which are the potential themes of positive reviews and the potential themes of negative reviews.

Table.8 Table of potential topics for positive comments

| Word     | P   | Word     | P   |
|----------|-----|----------|-----|
| Great    | 0.088 | Good     | 0.046 |
| Question | 0.072 | Screen   | 0.043 |
| Speed    | 0.066 | Satisfy  | 0.033 |
| Photo    | 0.064 | Battery  | 0.029 |
| Support  | 0.064 | Domestic | 0.028 |
| Love     | 0.058 | Worthy   | 0.024 |
| Run      | 0.055 | Function | 0.022 |
| System   | 0.049 | Appear   | 0.011 |

Table.9 Table of potential topics for negative comments

| Word     | P   | Word     | P   |
|----------|-----|----------|-----|
| Buy      | 0.076 | Garbage  | 0.032 |
| Service  | 0.061 | Speak    | 0.029 |
| No       | 0.057 | No       | 0.026 |
| Photo    | 0.056 | Question | 0.021 |
| Signal   | 0.051 | Apple    | 0.019 |
| Quick    | 0.050 | Membrane | 0.016 |
| Exchange | 0.048 | Store    | 0.014 |
| Network  | 0.042 | Speed    | 0.011 |

3. Analysis results and evaluation

3.1. Analysis of LDA model results
From the analysis results of the above LDA model, there are two potential themes of Huawei mobile phone praise. From the high-frequency words in theme 1: very good, camera, speed, support, system, etc., it can be concluded that Huawei mobile phones have the characteristics of fast running speed and good pixels, which are reflected in the performance of the mobile phone; the high-frequency words in theme 2: good, screen, battery, domestic, worthy, etc., mainly reflect Huawei's mobile phone battery life, high cost performance, and Chinese support for domestic mobile phones.

There are two potential themes for negative reviews. High-frequency words in topic 1: customer service, signal, network, etc., which reflect the poor service attitudes of Huawei and mobile phone network stalls. This objectively reflects that some customers no longer only pay attention to the performance of mobile phones, and put forward higher requirements for service attitudes and after-sales service of merchants; high-frequency words in topic 2: trash, apple, problem, physical store and other words reflect the possible deficiencies in comparison with Apple phones. There are also some problems with offline physical store operations. Although there may be some slander, it also shows that Huawei mobile phones are still at a disadvantage compared with some well-known foreign brand mobile phones. It may not only be caused by performance. We should learn from others’ advanced business models and service experience.

3.2. Evaluation of LDA model results
By analyzing the final results of the model, the following enlightenment suggestions are made for the merchants:
In the “80-20 rule”, the “80-20 business rule” pointed out that operators should conduct penetration marketing on 20% of key commodities and key customers. The results of sentiment analysis in this article show that nearly 20% of users have high sentiment scores. Combined with the "80-20 Marketing Principle", these 20% of customers are key customers that businesses should focus on. For these customers, businesses can conduct follow-up investigations and provide personalized services to continuously optimize their user value and maximize user loyalty and corporate benefits.

After all, marketing strategy is just a business method, and the core of the enterprise is still the product. For products, a good marketing strategy should be the icing on the cake, not to send carbon in snow. Through the analysis of the subject and emotional tendency of user reviews, we can understand the user's evaluation information on the advantages and disadvantages of Huawei's mobile phones. This is the customer's demand and the direction for the subsequent optimization of mobile phones by merchants. Through the analysis results, we can see that users generally have great disputes about the size of mobile phones. Large screens have many benefits, but they are not suitable for everyone, especially in one-handed operation. Therefore, for mobile phone research and development, businesses can consider adjusting the size, or developing several mobile phones of different sizes to meet the needs of different users. In terms of cost performance, there is a difference between product positioning and user positioning. As far as the product itself is concerned, unlike the Huawei Honor series, which focuses on cost-effectiveness, Huawei mate9 is positioned as the business flagship, focusing on the high-end market, and its price is destined not to be too low. However, most users do not consider the so-called positioning when purchasing a phone. Therefore, for some users, the price-performance ratio of the mobile phone is low, which also reflects that the market positioning and pricing of Huawei mobile phones still need to be improved. At the same time, Huawei can also focus on mobile phone positioning when conducting mobile phone marketing. In terms of service experience, although the overall evaluation is good, there are still some users who are unhappy in the process of receiving the service. Merchants should also pay attention to this aspect. Logistics and customer service play an important part of mobile phone marketing. Once they perform poorly, users will have a bad impression of mobile phones.

Build a user comment mining and display platform to enhance the realizable value of data. The data mining results of user reviews can not only be used as a data basis for merchants to formulate marketing strategies and determine the direction of product development and optimization. Appropriate display of the mining analysis results on the platform can also help consumers get the information needed in the reviews faster. At the same time, merchants can have a more comprehensive understanding of the comprehensive evaluation of the various attributes of the goods by the purchased users. Some measures can even be taken to encourage online shoppers to post detailed product experience reviews or display platform improvements to help consumers obtain a better shopping experience.

Acknowledgments
This work was supported by “the Construction team project of the introduction and cultivation of young innovative talents in Colleges and universities of Shandong Province of China, 2019( Project Name: Big data and business intelligence social service innovation team)”.

References
[1] Ankit S, Vijendra S and Gurdeep D 2019 Sentiment analysis of Twitter data: a hybrid approach. *J. International Journal of Health-care Information Systems and Informatics* IJHISI 14 p2.
[2] Yanhui Z, Zongwei L and Yicheng Z 2017 Influence of information quality based on review data of Taobao.com on the usefulness of online comments *J. Journal of management* 14(01) pp77-85.
[3] Yin L 2018 P. A LDA based analysis for News coverage difference of Chinese Portals *DEStech Transactions on Computer Science and Engineering*.
[4] Min L, Xiangqian W, Huizong L and Baolong Z 2018 Emotional analysis of Online commodity
comments based on text mining J. Journal of Liaoning University of Technology (Natural Science edition) 38(05) pp 330-335.

[5] Nianfeng W, Kunyi L, Xinhao Z, Jinfan L and Xianmin Z 2019 The recognition of grasping force using LDA J. Biomedical Signal Processing and Control 47.

[6] Chenxi L and Luodi X 2017 Text classification and view point mining based on LDA Model J. Electronic Technology and Software Engineering 4 pp 209-210.

[7] Qingfu W and Xingguo W 2016 Research on topic discovery of Network comments based on LDA J. Wireless Internet Technology 11 pp 103-104.