

Analysis of text mining methods and comparison of differences between different e-commerce platforms based on text mining

Wenjie Fan

School of Business, Macau University of Science and Technology, Macau, China,
999078

1909853nb011002@student.must.edu.mo

Abstract. With the further popularity of online shopping, people can buy the goods they want on online platforms without leaving home. However, due to the asymmetry of information, it is difficult for consumers or companies to compare the advantages and disadvantages of the same goods on different e-commerce platforms. This paper uses the word cloud method for text mining and Python to crawl and analyze customer reviews of the same product on two e-commerce platforms, which are Taobao and Jingdong, in order to compare the similarities and differences of the same product on different platforms. Through the comparison, we found that customers on the Jingdong platform think that the Jingdong platform provides good logistics services, while customers who consume on Taobao more often think that the products on that platform have high cost performance.

Keywords: Customer Reviews, Text Mining, E-Commerce Platforms, Comparative Analysis.

1. Introduction

The booming development of e-commerce in the past decade and the further increase in Internet pervasiveness, it has profoundly affected the shopping habits of consumers [1]. Consumers can select the products they want online without leaving home.

However, while online shopping brings convenience to consumers, due to the asymmetry of information, it is difficult for consumers to compare the advantages and disadvantages of products and services by different e-commerce websites [2]. Currently, several major shopping platforms in China are currently open to the function of user reviews of products. The use of big data text mining to extract key information from consumers' product reviews is important to help consumers and merchants understand the differences between different sales platforms.

There have been many scholars who have conducted research on text mining. Usually, the texts studied by scholars include user reviews, journals, and articles. For user reviews, common analysis methods include using R or Python to extract data and using word clouds or other statistical tools to organize and analyze word frequency [3][4][5]. In the case of texts such as journal articles, the common method is to use R statistical tools and to use Latent Dirichlet allocation (LDA) for topic modeling [6]. In this paper, we use Python to crawl the data and visualize the results as a word cloud, referring to previous approaches to analyze customer review data.

Despite the fact that many scholars have studied the topic of customer reviews, there have been few studies on the consistency of reviews on e-commerce websites and by consumers after shopping [7]. The reason for this problem may be, on the one hand, that scholars have focused their research on text mining methods. On the other hand, it may be due to the high degree of difficulty for scholars to obtain review data from multiple platforms. However, with the development of related technologies, the data is becoming more accessible.

By analyzing the opinions of consumers across multiple platforms, similarities and differences in reviews can be identified. By comparing these similarities and differences, it is possible to discover the true characteristics of the goods and their features on different platforms, which is important for both consumers and companies.

2. Research method

2.1. Data collection

Getting real and valid data from online stores is a key part of getting the real needs of users. The method to obtain user reviews is to use crawlers to get review data from product pages of shopping platforms. To facilitate comparison, some brands of products from the two major Chinese e-commerce sites Taobao and Jingdong are selected here for analysis.

Requests and Beautiful Soup are excellent third-party open-source libraries for Python [8]. This paper crawls data using the Requests library and Beautiful Soup library to achieve the crawling and parsing of online store web data.

The design idea of the crawler is to use the Requests library to start with an initial URL and get the initial URL. The design idea of this crawler is to use the Requests library to get the initial URL starting from an initial URL. This program continuously extracts new URLs from the current page in the process of crawling the web page to form a queue, parses the web page by applying Beautiful Soup's data parsing module and downloads and stores the crawled user reviews of a particular product [3]. Until the stopping conditions of the system are met, certain analysis, organization and indexing are performed to facilitate queries.

In this paper, some products that sell well on both platforms are mainly selected as the object of crawling data, and from which the last hundred comments are analyzed. The following article takes the user review data of the Shenzhou (HASEE) Warrior Z8-DA7NS New 12th Generation i7-12650H RTX3060 15.6-inch gaming laptop as an example.

2.2. Data processing

Since users' comments are highly arbitrary, they interfere with the truly valuable parts of users' comments. Therefore, before analyzing the text, it is necessary to delete the comments that do not reflect users' needs to reduce the influence of useless comments on the text mining effect. The organized user comments are saved under the same text for word cloud analysis.

3. Results and analysis

3.1. Results of word cloud analysis of text data

The word cloud analysis can well identify the important keywords in the text [9]. Using these keywords, we can infer the real user experience of the product, understand the user's consumption preference and adjust the marketing strategy on the user's focus.

Word cloud analysis is a good way to get the important keywords in the text. Using these keywords, we can infer the user's real experience of the product, understand the user's consumption preference and adjust the marketing strategy on the user's focus.



Figure 1. Jingdong user evaluation word cloud map (photo credit: Original).

The user evaluation text in this paper was put into the word cloud analysis tool, and the word cloud diagram of user evaluation was generated as shown in Figure 1. The size of the keywords on the word cloud map reflects the frequency of these keywords in the user reviews of the Shenzhou Zhangshen laptops sold on the Jingdong platform in the last year. The more keywords appear in the word cloud chart, the larger the head will be, the closer to the center of the word cloud chart.

After that, the Taobao platform of the same model of laptop sales for user evaluation analysis, the word cloud as shown in Figure 2.



Figure 2. Taobao user evaluation word cloud chart (photo credit: Original).

After sorting these high-frequency words statistically, the following conclusions were obtained

In Taobao platform user reviews, the high-frequency words appearing in the one hundred most recent reviews are presented in Table 1.

Table 1. The main high-frequency words of Taobao platform user reviews.

characteristics:	frequency:
appearance	79
material	45
performance	45
office	45
boot	30
smooth	25
cost performance	25
clear	20
speed	15
fluency	10

In the Jingdong platform user reviews, the high-frequency words appearing in the one hundred most recent reviews are shown in Table 2.

Table 2. The main high-frequency words of Jingdong platform user reviews.

characteristics:	frequency:
speed	56
picture	40
appearance	38
game	29
clear	25
screen	21
fluency	20
performance	19
quality	19
logistics	15

3.2. Analysis of word cloud analysis of text data

Because the selected laptops ranked among the top products of their kind in terms of sales on both platforms, most of the neutral words appearing in the statistical results indicate positive meanings. The analysis of the high-frequency words and their word frequencies in both tables shows that users on Taobao care more about the appearance, materials, performance, and ease of working of the laptop, while users on Jingdong care more about attributes such as the speed, graphics, appearance, and ability to run games on the laptop. As can be seen, in the above example, the similarity in user reviews on both platforms is that customers who consume on both platforms consider the quality of the product itself to be relatively good. The difference between the two platforms' consumer reviews is that a significant number of Jingdong users mention that they are satisfied with the logistics, while a larger number of Taobao users mention that the product has a high-cost performance.

Through the above analysis, it can be found that consumers on Taobao and Jingdong platforms hold the same views on the quality of the above products themselves, but the comments on consumer logistics and value for money are not consistent. This shows that the difference in cost performance and logistics is more influenced by the different policies of the platforms and other aspects, and also shows the different characteristics shown by the two platforms in this case.

Consumer reviews are a very important source of information for both companies and consumers [10]. On the one hand, the consistency of consumer reviews can help companies and consumers analyze the reliability of information. On the other hand, the differences in consumer reviews can also show the differences between companies and businesses, so that consumers can find out the strengths and weaknesses of different platforms and make the best decision in their own situation. Companies can also

study the differences in consumer reviews to find out the differences between themselves and their competitors, so that they can make changes to improve their competitiveness.

4. Conclusion

This paper uses Python techniques to crawl data from Taobao and Jingdong platforms and uses word cloud methods to statistically analyze the high frequency words of customer reviews. By comparing the similarities and differences between consumer reviews on different e-commerce platforms, we propose the impact on consumers or companies. And the user's needs were clearly stripped out from the huge text. Customers on the Jingdong platform believe that the platform provides better logistics services, while those who consume on Taobao are more likely to believe that the platform's goods are cost-effective. Thus, by using this approach, I argue that different platforms and consumers have access to the information they need to optimize their decisions and thus drive the economy. The shortcoming of this paper is that there is less data used for analysis when using case studies. In the future, if platforms are more open to allowing customers to access other customers' reviews and more efficient analysis methods are available, it will be easier, faster and more accurate for such case studies to help companies or consumers to make decisions.

References

- [1] Beiyun Wang, Jiansong Huang. Research on online consumer psychology and online marketing strategy[J]. Modern Economic Information,2016(03):142-143.
- [2] Ruiyuan Wu. Text mining analysis of online user evaluation information [D]. Tianjin University of Finance and Economics, 2019.DOI: 10.27354/d.cnki.gtcjy.2019.000046.
- [3] Guerreiro, J., Rita, P. & Trigueiros, D. (2016). A Text Mining-Based Review of Cause-Related Marketing Literature. J Bus Ethics 139, 114. <https://doi.org/10.1007/s10551-015-2622-4>
- [4] Younis, E. M. (2015). Sentiment analysis and text mining for social media microblogs using open source tools: an empirical study. International Journal of Computer Applications, 112(5).
- [5] Qing Cao, Wenjing Duan, Qiwei Gan, (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach, Decision Support Systems, 50(2), 515. <https://doi.org/10.1016/j.dss.2010.11.009>.
- [6] Alexandra Amado, Paulo Cortez, Paulo Rita, Sérgio Moro, (2018). Research trends on Big Data in Marketing: A text mining and topic modeling-based literature analysis, European Research on Management and Business Economics, 24(1)2018, 3. <https://doi.org/10.1016/j.iieden.2017.06.002>.
- [7] Yili Zou. Research on the factors influencing the usefulness of online reviews on shopping websites [D]. Yunnan University,2019.
- [8] Xiaoxu Du, Xiaoyun Jia. Analysis of Python-based Sina Weibo crawler[J]. Software,2019, 40(04):182-185
- [9] Weisong Huang, Yuzhu Zeng, Senlin Wei. Research on the application of text mining in enterprise network opinion analysis[J]. Computer Programming Skills and Maintenance,2017(22):5-8+17.DOI: 10.16184/j.cnki.comprg.2017.22.001.
- [10] Guoliang Shi, Qiaofeng Shi. Research on the consistency of product reviews of different shopping websites based on text mining[J]. Modern Library and Information Technology,2011(12):64-68.