User Sentiment Analysis Based on Online Reviews

-Taking the Forbidden City Afternoon Tea Restaurant as an Example

Sijia Zheng[1,a,*]

[1]*Beijing Jiaotong University, Beijing, 100000, China*
a. 13704148627@163.com
*corresponding author

Abstract: Nowadays, Internet consumption has become a familiar name, and there are more and more feedback channels on the production experience of consumers, and the content of evaluation also determines consumers' views on the goods to a certain extent. Therefore, the rational use of online comment contents can not only help consumers understand the basic information of products, but also help enterprises to better obtain the needs of consumers and provide more references for them. This paper selects the comment data of the Palace restaurant in Dianping software, and uses Python software to conduct data mining, emotion analysis, word frequency statistics and word cloud drawing. The conclusion shows that consumers have a positive emotional tendency towards the afternoon tea in the Palace Museum, with few negative comments. At the same time, consumers pay more attention to the dining location and food ingredients. Therefore, businesses can increase the improvement of dishes and environmental transformation through online reviews to better meet the needs of consumers. This research not only provides the basis for the evaluation of the business level, but also helps to enrich the consumer's understanding of the business.

Keywords: Python, online review, text mining, emotion analysis

1. Introduction

On the Internet, consumers' purchasing experience is often displayed in the form of online comments. As consumers' subjective experience when choosing products, online comments provide information support for consumers to understand products and after-sales service of merchants. However, with the increase in the number of online shopping and purchasing experiences, the scale of online comment data also expands, and it is more difficult for consumers to find information conducive to purchase decisions from many comment data. In addition, it is difficult for merchants to obtain information conducive to business decisions from such complicated and huge comments. How to effectively use the comment information and feedback to consumers and businesses has become an urgent issue to be discussed.

In this context, this paper uses Python to analyze and mine the comment data of consumers on the Palace Museum afternoon tea on the platform. Through the emotional analysis of review data, understand consumer emotional attitude to the platform and afternoon tea, obtain the factors affecting consumer emotional tendency, and the afternoon tea concerns and purchasing behavior preferences, on the one hand, help consumers better understand the palace afternoon tea characteristics, provide

purchase decision information; on the other hand, platform and businesses can satisfy the needs of consumers to develop the marketing strategy, provide high-quality products and services, expand market shares, and improve their competitiveness.

2. Literature Review

2.1. Related Research on Online Reviews

Online comments are mainly the comments about products or services left by users on various network platforms, which is the embodiment of consumers' subjective feelings.

The research on online reviews abroad is earlier than that in China. Duan has studied how to improve the box office of films. Among many factors, he found that consumers were interested in movies with a large number of reviews, and the number of comments and content affect the number of viewers [1]. Floh found that consumers' purchasing behaviors are influenced by commodity quality theory, and the quality and quantity of comments will interfere with consumers' purchasing behaviors to some extent [2]. Lee based on a network experiment, examining how the quality of online product reviews affects participants' acceptance of comments and their assessment of sources, and how the product type and the availability of reviewer photos, researches found that high-quality reviews will enhance people's positive evaluation of goods and purchase intention, but consumers make positive evaluations with a premise is that comments have pictures, which shows the importance of visual cues [3]. Barnes taking the hotel as an example, the relationship between online comments and satisfaction was studied, and the satisfaction factors were extracted through the theme model method. The results showed that the score of online comments had the greatest impact on consumer satisfaction, and the influence of other factors such as the number of comments had a small impact [4]. Meng believes that online reviews are divided into quality. It studies the differences between different quality reviews and uses the constructed measurement model in Amazon review data. The results show that high-quality reviews have a stronger influence [5].

To sum up, the research on online comments mostly focuses on comment content itself, and the study of online reviews of consumers buy most after collecting the review data, the empirical part is still based on the traditional questionnaire model, rarely completely based on text mining method, in addition to the existing research angle is relatively single, only for consumers or merchants, so the study of online reviews completely based on text mining method, and the results have realistic significance for consumers and merchants.

2.2. Related Research on Sentiment Analysis

Nasukawa The method of emotion analysis was first proposed. With the continuous maturity of the emotion analysis algorithm, this method has been gradually applied to [6] in various fields. Riloff builds a specific semantic dictionary for the words and categories of specific topics. Users can quickly find the category words and build the core semantic [7]. Whitelaw et al. used fine-grained semantic differences to classify emotions, combined with the standard "package of words" function, to classify movies' reviews, and got a high accuracy [8]. Kim et al. first applied the emotional dictionary method and constructed two dictionaries with different polarities, which laid the foundation for the subsequent analysis [9]. Pang et al. first applied the machine learning method and applied it to movie reviews, giving the classification difference between different machine learning methods [10]. Denil et al. represent the meaning of documents by embedding them in low-dimensional vector space, and hierarchically learning the capture and synthesis of low-level lexical features to high-level semantic concepts. The model can also achieve good results without feature engineering and more compact models [11].

To sum up, through the literature it can be seen that the study of emotion analysis often tends to choose a research method, the single and emotion analysis method has been limited in the research field, so this paper applies the emotion analysis method to the Forbidden City afternoon tea review data, and the emotional dictionary method and machine learning method combine the consumer emotional tendency to products and platform, and then the comments are set into positive negative two dimensions.

3.   Methodology

3.1.   Study Design

This paper mainly analyzes the online comments on afternoon tea in the Palace Museum, and digs out valuable information for consumers, businesses and platforms. Using the python network crawler method collection platform of the Forbidden City afternoon tea review data and pretreatment, using the dictionary analysis of consumers' emotional tendency of goods and services, from the two dimensions of positive negative mining factors affecting consumer emotional tendency, influencing factors also reflect the consumers in the electric business platform to buy aquatic products of main concern, consumer preferences.

In the part of data processing and empirical analysis, this paper mainly uses web crawlers, natural language processing, emotion analysis and other methods. Internet crawlers are mainly used in the crawling of comment data, while natural language processing is mainly used for the pre-processing of data. In emotion analysis, emotion dictionaries and machine learning are used to judge the emotional tendency of comments, so as to get the factors affecting consumer tendency.

3.2.   Data Acquisition

This paper uses Python software to write code to collect the comment data. First, it obtains the URL of the Forbidden City afternoon tea online review, sets the headers, cookies and referer request heads, sends the requests to the server, and saves the obtained results as a CSV file, that is, to complete the data collection.

The data collected in this paper mainly includes user ID, evaluation time, comment content, rating, preference degree, etc., and finally obtained 4,231 comments. Some reviewed data are shown in Table 1.

Table 1: Crawled partial data display.

ID	Time	Contents	Grade	Like
Amy	2023/7/16 20:28	Dragon Boat Festival before clocking in the Palace Museum tea, fortunately, not many people, nor very hot.	4	0
Sam	2023/7/16 20:18	There are so many people that we have to fight for numbers every day.	5	0
Lucky	2023/7/16 19:37	It was true that having a bowl of this kind of afternoon tea in the Hot Palace Museum instantly made people feel that everything was worth it.	5	0

Table 1: (continued).

xixi	2023/7/16 19:15	Both chocolate and matcha mousse are delicious but not very sweet.	5	0
lemonnie	2023/7/16 11:29	The taste is very good, but the seats are few, the experience is affected.	4.5	0
wx	2023/7/16 3:41	Taste: good taste, very delicious. Environment: There are really many people, the seats are few, need to reserve enough time in advance to comment, in a queue is very important.	3.5	0
SUM	2023/7/15 19:35	The drinks are all good, especially this concubine drink. Have a good snack!	4.5	0
meng000	2023/7/15 16:36	It's worth it. Once you're in Meridian Gate, go straight there, get a number, and then hang out.	4.5	0
sunshine	2023/7/15 13:06	Rushed into the queue, more people, less seats, "Rich red plum lichee drink" delicious explosion.	4.5	0

3.3. Data Preprocessing

One is word segmentation. In linguistics, words have rich emotional connotations as the smallest independent emotional unit. The second is the part of speech annotation, part of speech refers to the basis used to divide the class of words according to the characteristics of words. Including nouns, verbs, AIDS, adjectives, adverbs and conjunctions. The third is to stop words and special characters. Stop words refer to some functional words that can be filtered out in natural language processing. These words appear frequently in the text, but have no practical impact on the emotional analysis of the text, such as pronouns, auxiliary words, prepositions, tone and auxiliary words, etc.

3.4. Sentiment Analysis

Using the SnowNLP library to analyze the emotion polarity of online comments, the training text of SnowNLP is the comment data, so the accuracy is high in the emotion analysis of online comment text. Read each comment and analyze the emotion value in turn. When the value is greater than 0.5, the emotion polarity of the sentence is positive. When the score is less than 0.5, the emotional polarity is negative. The more biased the value is to both sides, the more extreme the emotion is. Using the results of the analysis of the SnowNLP library draw an emotional scatter plot, visualization of data, the figures seen from most consumers of the Forbidden City afternoon tea are extremely praised, present positive emotional characteristics, and neutral or not express emotional polarity comments are relatively less, comments present polarization, but enthusiasm is more than negative, as shown in Figure 1.
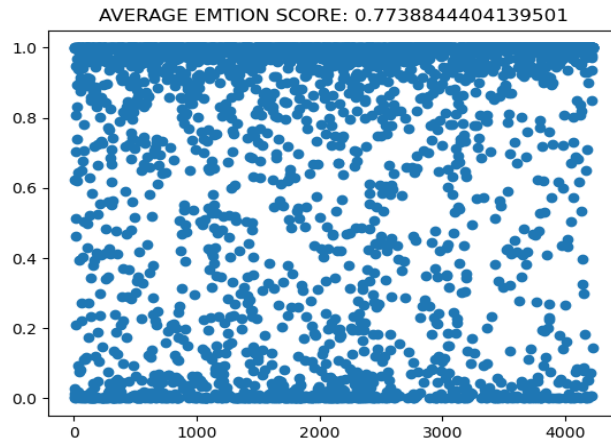
Figure 1: Emotional scatter plot.

3.5. The Word Frequency Statistics of Word Segmentation by Jieba and Word Cloud Combination

In order to further analyze the above-crawled text, this paper combines Python jieba and word frequency analysis with Wordcloud. From the perspective of word frequency, the 30 words with the highest frequency of online comments are listed and the word cloud map is drawn, so as to achieve more intuitive emotion analysis and display (See Table 2 and Figure 2).

Word frequency through numerical expression, you can see the word frequency ranking in the top 12 words, in addition to expressing emotional vocabulary, there are some nouns related to location, food, you can see that consumers for "imperial palace", "litchi" afternoon tea related factors such as high attention, and their comments and views on these elements, so businesses can increase through online comments for the improvement of food, environment, etc.

Table 2: Word frequency statistics.

Words	Number
Forbidden City	3579
Lychee	2734
Bayberry	2374
Taste/Flavor	2307
Afternoon Tea	2060
Pretty Good	1625
Queuing	1591
Delicious	1517
Mousse	1231
Chocolate	1206
Service	1136
Environment	1063

Figure 2: Word Cloud.

4.   Conclusion

With the development of the Internet and the increase of online comments, consumers and merchants' platforms cannot find the potential and valuable information in the comments only by browsing and reading. Therefore, text mining technology is playing an increasingly important role in solving such problems. Using text mining technology can not only provide potential consumers with purchase decisions, but also find consumers' preferences from the comments, so as to constantly meet consumer needs. At the same time, it can also help e-commerce platforms to differentiate competition and achieve economic benefits. In addition, the use of text emotion analysis and word frequency statistics can not only enable merchants and consumers to deepen their understanding of products and services, but also enable businesses to focus on the areas that consumers are more concerned about, which has a positive impact on both merchants and consumers.

This study expands the application field of online comments and text mining methods. This paper applies the online review and promotes the relevant researches in the review software, which further expands the research field. In addition, from the research results of domestic scholars, there are few practical application researches, and most scholars pay more attention to the improvement and optimization of mining methods. This paper enriches the practical research of text mining to some extent.

What's more, this study enriches the research methods in the field of theme restaurant research. Previous researches in the field of theme restaurant uses questionnaire, traditional model methods, such as structural equations, number according to a long time and subjective, in this paper, the first use of Python climb review information online comments are mostly after consumers' using the product, which can more directly reflect consumers' attitude towards the product, using emotion analysis method can further mine comments valuable information, and provides a new way of thinking for the palace afternoon tea of the theme tea research.

References

[1] B, D.W.G.B.W.A., Do online reviews matter? An empirical investigation of panel data [J]. Decision Support Systems, 45(4)1007-1016 (2008).

[2] Arne Floh, M.K.A.Z., Taking a deeper look at online reviews: The asymmetric effect of   valence intensity on shopping behaviour [J]. Journal OF Marketing Management, 29(56):646-670 (2013).

[3] Lee E. J., Shin S. Y. When do consumers buy online product reviews? Effects of review quality, product type, and reviewer's photo [J]. Computers in Human Behavior, 31:356-366 (2014).

[4] Guo Y., Barnes S. J., Jia Q. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation [J]. Tourism Management, 59(APR.): 467-483 (2017).

[5] Meng Yuan, Hongwei Wang, Lijuan Zheng. Impact of online word-of-mouth on sales: the moderating role of product review quality [J]. New Review of Hypermedia & Multimedia, pp.1-27 (2018).

[6]   *Agarwal B., Mittal N. Prominent Feature Extraction for Sentiment Analysis [J]. Springer International Publishing, pp.77-88 (2015).*

[7]   *Riloff E. M., Shepherd J. A Corpus-Based Approach for Building Semantic Lexicons [C]// In proceedings of the second conference on empirical methods in natural language processing (1997).*

[8]   *Whitelaw C., Garg N., Argamon S. Using appraisal groups for sentiment analysis [C]// Acm International Conference on Information & Knowledge Management. ACM (2005).*

[9]   *Kim S. M., Hovy E. H. Automatic identification of pro and con reasons in online reviews [C]// International Conference on Acl. Association for Computational Linguistics (2006).*

[10]   *Pang B. Thumbs up? sentiment classification using machine learning techniques [J]. Proc. EMNLP (2002).*

[11]   *Denil M., Demiraj A., Kalchbrenner N., et al. Modelling, Visualising and Summarising Documents with a Single Convolutional Neural Network [J]. Computer Science (2014).*