



Text Mining and Analysis of Meituan User Review Text

Yong-juan Wang^(✉), Guang-hua Yu, Li-nan Sun, and Pei-ge Liu

Heihe College, Heihe 164300, China
641917127@qq.com

Abstract. Based on the current situation of the market, this paper obtains related data of online user reviews on Meituan the online food delivery platform by soft wares and implements preprocessing and mining by language correlation function, and finally draws a conclusion that judging from the mining results of the featured words and the emotions in the reviews, the Meituan platform and its food delivery service have been evaluated by users as being cheap, economical, convenient and fast, and the key elements that users concern regarding to the merchant rating are the merchant's attitude, the delivery man's attitude, the food taste and the food security respectively.

Keywords: Online user reviews · Text segmentation · Keyword extraction

1 Research Design

Based on the empirical study of Meituan food delivery and from the perspective of social psychology, this paper finds out that the basic requirement of users is the key factor for their choices on products to meet their expectations, while the exposure, sales, rating and specific review content of a merchant help them to decide whether to choose it. After a research into the text mining theory applied in the paper, we mined and processed the data as shown in the Fig. 1 below. Firstly, we preprocessed the data source to achieve a denoise effect. We created a user customized dictionary and removed stop words to guarantee the accuracy of the word segmentation, and it demanded repetitions for the best effect. And finally we carried out mining and analysis on the review content after the word segmentation.

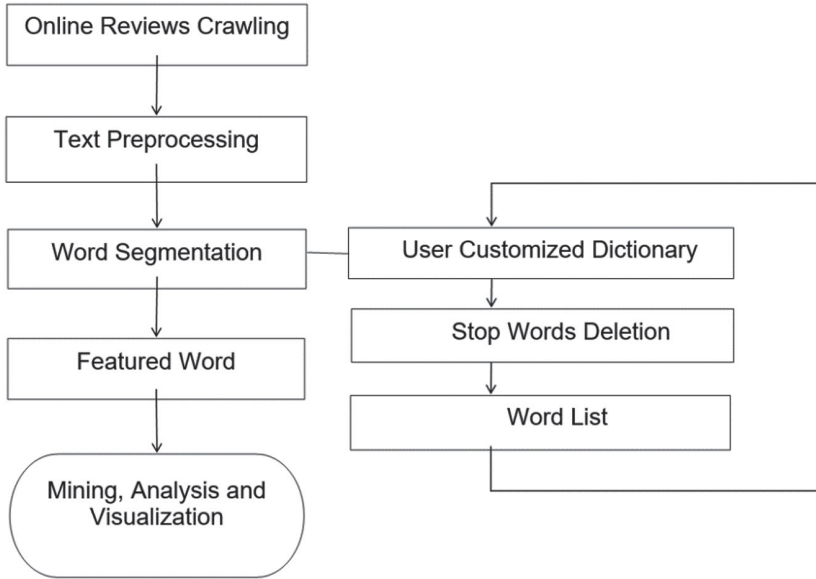


Fig. 1. The analytical framework of empirical study

2 The Empirical Analysis

2.1 Data Crawling and Process

We have crawled 3013 reviews of the biggest food delivery platform in China Meituan integrated on the website Baidu Koubei (<https://koubei.baidu.com/>, a special website integrating users' reviews and merchants' catalogs), and a part of merchants' user reviews on Meituan itself by using a third party software Octopus Crawler. And finally fixed the merchant Xuji Pie as the object of this text mining research considering its high rating and quantity of users' reviews.

There are altogether 597 valid users' reviews crawled by Octopus Crawler on Xuji, a restaurant mainly selling pie in Heihe city. The first step is to take a denoise processing on the text data obtained, and preprocess the data by keeping the useful content while deleting some irrelevant fields including the nicknames, review times and irrelevant reviews. And then deleting poor quality reviews including reviews in default templates and texts in a length no more than 2 characters.

2.2 Keyword Extraction of the Review Texts

The original crawled online reviews should be processed in the form of the structured data for analysis. A part of the review data is set as the training library, and then the review content is processed in the steps of word segmentation and featured word extraction. Partial results are shown as in Fig. 2.

```
> wk<-worker()
> words<-"好吃!米饭有点硬 但是锅包肉超级好吃!海带丝有点咸!不太好吃地三
+   鲜也好吃!"
> segment(words,wk)
[1] "好吃" "米饭" "有点" "硬" "但是" "锅包肉" "超级"
[9] "海带丝" "有点咸" "不太" "好吃" "地" "三鲜" "也" "好吃"
```

Fig. 2. Training result of review text segmentation

The training result of word segmentation reveals that some proper nouns like “地三鲜” (means 3 fresh vegetables growing underground, a dish of stewed fried potatoes, eggplants and peppers) has been separated into two words “地” (ground) and “三鲜” (3 fresh), and the phrase “不太好吃” (not very good) has been separated into two words “不太” (not very) and “好吃” (yummy). Those segmentation results will affect the accuracy of the text feature extraction, so it is necessary to create a user customized dictionary on the basis of merchants’ information for accurate segmentation of review content. And we can also see a large quantity of noise words like “too” or “but” where we still need to denoise the segmented texts by the steps of setting down filter criterias, making a stop word list and deleting useless words accordingly. The stop words are mined by following the principle that only words with the useful information will be adopted, and then added into the stop word list for a further deletion.

Basically, we create a user customized dictionary based on the product, cut reviews into sentences, finish segmentation by using the customized dictionary, and then mark the part of speech of each word. Because the product characteristic values are mainly nouns, we filter the nouns by filtering criteria when extracting featured keywords, and also delete the stop words in the list of marked texts.

All the crawled review texts are processed by word segmentation in terms of the training results. The processed texts are used for word frequency statistics and keyword extraction. After processing the featured words by algorithm, we finally selected 20 key words for the use of Meituan’s user review analysis. The 20 key words selected from Xuji Pie’s user reviews for analysis are listed in Fig. 3 below.

```
> keys<='mydata.txt'  
31226.3 1668.3 986.093 682.193 526 399.133 363.915 356.982  
" " "好吃" "吃" "味道" "不错" "菜" "地三鲜" "骑手"  
328.698 302.484 293.48 292.16 278.563 270.002 270.002 246.523  
"肉" "馅饼" "量" "米饭" "砂锅" "锅包肉" "盖饭" "饼"  
246.523 245.773 234.784 228.49  
"做" "小哥" "点" "实惠"
```

Fig. 3. Keyword extraction result of Xuji Pie’s user reviews

3 Review Mining and Result Analysis

A further visualization is conducted on those segmentation words and extracted key words. The cloud picture as shown in Fig. 4 made by the words can directly show the core content of the user reviews.



Fig. 4. The cloud picture of Meituan’s user review words on Baidu Koubei

The word “food delivery” has been mentioned many times among all reviews, which can prove that the food delivery is the featured business of Meituan. The key words standing for the comprehensive assessment are “economical”, “cheap” and so on. And we can also infer from Fig. 4 that quite a large number of user’s order food delivery by using Meituan, and the most commonly words to describe their experience are “not bad”, “cheap”, “economical” and “love it”. And another word “Islamic” is also a high

frequent word, but according to the research, the specific reviews related to “Islamic” topic are almost passive reviews such as “discrimination”, “rubbish”.

Considering that the words reflecting the comprehensive assessment are “economical”, “cheap” and so on, Meituan and its food delivery team should aim to provide low prices and convenient and fast service in order to acquire more users, which can also help to improve users’ satisfaction on both the service and the food, and hence to increase their dependence and tendency.



Fig. 5. The cloud picture of review words of Xuji Pie

Figure 5 finds that most users prefer to describe their user experience by mentioning the words “deliveryman”, “deliveryman buddy”, “attitude”, “volume”, “salty”, “on time”, “taste” and “clean” in their reviews, which means the factors that users concern in their experience lie in the punctuality of deliveryman, the attitude of deliveryman, the attitude of the merchant, food taste, food volume and food security.

For merchants, to attract users or improve their satisfaction degree, they need to serve with a good attitude and meanwhile guarantee food security and taste. For delivery men, they should remember to keep a good attitude and deliver the food on time in safety so that they can help to improve the user loyalty to Meituan, and hence to increase both the user number and business sales of Meituan food delivery.

References

1. Li, T.: Data Mining Where Theory Meets Practice. Xiamen University Press, Hainan (2013)
2. Li, Z.: Research on the Influence of O2O Take out Catering Platform Online Comments. E-commerce Press (2017)
3. Dong, S., Wang, X., Ge, Z.: Analysis of content characteristics of B2C shopping website online comments based on text mining. Library Theory and Practice (2017)
4. Li, X.: Feature word extraction in Chinese text classification. Comput. Eng. Des. (2019)
5. Tao, Q., Ge, T.: Research on big data mining technology based on THDS. J. Mudanjiang Norm. Univ. (2017)
6. Zhang, L.: Language Data Analysis and Mining Practice. Mechanical Industry Press, China (2015)
7. Hao, Y.: An Empirical Study on the Influence of Online Reviews on Consumers’ Perception and Purchase Behavior. Harbin Institute of Technology, Harbin (2010)
8. Fang, N., Liu, X.: Business process change domain analysis based on data flow constraint of petri net. J. Mudanjiang Norm. Univ. (2018)