

# Product Perspective: A Study of User Demand Priorities in Online Community Reviews

Hong Zhang <sup>a,\*</sup>, Jiayi Nie <sup>b</sup> and Hui Wang <sup>c</sup>

School of Economics and Management, Zhejiang Sci-Tech University, Hangzhou, China

<sup>a,\*</sup> zhh96941@zju.edu.cn, <sup>b</sup>niejoy1030@163.com, <sup>c</sup>wh93hz@163.com

**Abstract.** In this study, we through the data analysis to obtain a list of user demand and priorities to provide different opinions and recommendations for product design and marketing. The study is divided into two parts, one of which is to build a usefulness model in product perspective, and then use product community reviews as an evaluation corpus and experimental data set. The other is to propose a new feature-sentiment extraction method based on the research status of implicit features, and then select the basic algorithm of random forest as the classification model to combine text mining technology with KANO model for user demand research. Based on this, a new user demand priority calculation method is proposed. According to the study, the user's priority list provides real-time and accurate recommendations for product design and improvement, which is more practical in today's information overload era.

**Keywords:** Usefulness classification; Feature extraction; Sentiment analysis; Text mining; KANO model.

## 1. Introduction

The concept of “online review” was officially proposed by Chatterjee for the first time when he was researching whether the user would refer to the buyer’s comments for purchasing decisions [1]. Through online review, users can learn more about the detailed information of products and services to avoid losses because of the information inequalities in the shopping decision process [2]. At the same time, online reviews are also an important channel for companies and businesses to understand the user's real feedback and demand information [3]. Due to the large amount of user-generated content data and easy to access on the online platform, online reviews have gradually become an important data source for user demand research. Therefore, this study is based on the advantages of a large number of online review, combined with text mining technology and KANO model to classify the user demand for sentiment analysis results. A new user demand prioritization method is proposed to obtain a final list of user demand and to provide suggestions for product improvement and design, and also to marketing strategy formulation.

## 2. Related Works

### 2.1 Usefulness Analysis

The usefulness of online reviews is a subjective perception that consumers can get valuable information from reviews to assist them in making purchase decisions. The current research on the usefulness of online reviews is mainly based on the reviewer's perspective, and there are few studies from product perspectives [4]. Some scholars use manual annotations to obtain data analysis samples in product perspectives, and use statistical or machine learning methods to construct review usefulness models [5, 6]. Although the kind of manual labeling can fully consider the user demand, it is relatively time-consuming, labor-intensive and subjective. Therefore, the usefulness of online review based on user demand ought to further exploration.

### 2.2 Feature Extraction and Sentiment Analysis

Feature extraction and sentiment analysis mainly refers to the identification of users' views and opinions on products or services from online reviews. It will lead to the phenomenon of fuzzy sentiment analysis in feature extraction because of the review contains a lot of noise, which will cause

the sentimental direction to be unknown. Rabelo et al. conducted sentiment analysis of user reviews in online communities through graphical modeling and link mining techniques based on user perspectives [7]. In recent years, some scholars have begun to use the deep learning algorithms in machine learning for sentiment analysis [8].

### 2.3 User Demand

For the current fierce market competition environment, the needs of users are increasing with the variety of products and functions. Accurate access to demand has become a key factor in enterprise product improvement, design and marketing strategy development. However, most of the scholars conduct research in user demand and less research on user demand priority, which is still in the exploration stage. Therefore, how to more comprehensively capture the potential needs of users or the type of demand is still the focus of research. Moreover, the analysis results have a “virtual high” phenomenon due to the high praise rate of online shopping platform reviews. In consequence, this study uses product community reviews as a review data set to balance the issue of high praise rate.

## 3. Research Process

### 3.1 Data Preprocessing

In order to ensure that the collected data is valuable for research and reduce the interference of noise data on the experimental results, this study we delete the garbage reviews, duplicate reviews, and treats the special symbols at the same time. In addition, we perform synonym conversion on the text content in the data set before text analysis.

### 3.2 The Usefulness Classification Model

The review data is selected as the training set and the test set, which are manually annotated according to the above usefulness review definition. At the same time, the random forest algorithm is selected to construct the usefulness classification model. In order to make the usefulness model work better, the training set is adjusted. After the adjustment, the experimental verification is carried out on the test set. The optimized random forest model effect and the pre-optimization effect are shown in Table 1.

Table 1. Comparison of classification models

| Algorithm     | Precision | Recall | F1 score |
|---------------|-----------|--------|----------|
| RF            | 0.82      | 0.82   | 0.82     |
| RF(optimized) | 0.84      | 0.84   | 0.84     |

### 3.3 Feature-sentiment Extraction

This article divides the sentiment-feature rule base into two categories: (1) The 1V1 rule base indicates that such sentimental words only modify one feature word, that is, the product features of the user evaluation can be directly inferred based on the sentimental words; (2)The 1Vn rule base indicates that an sentimental word modifies multiple feature words, that is, the frequency of occurrence of different product features is modified according to statistical sentiment words, as the rule base and similarity calculation as the means of implicit feature and implicit sentiment analysis. The construction rules are shown in Fig.1, where s\_num means the number of times the sentimental word modifies different features, and s\_f\_num means the frequency at which the sentimental word modifies the same feature.

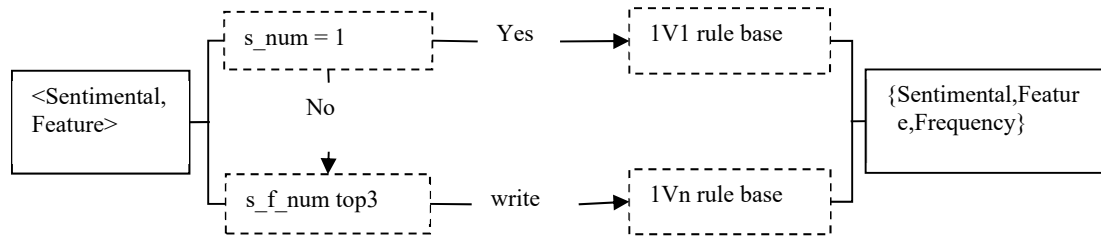


Fig.1 Rule base construction

Among them, the similarity is calculated by using the Doc2vec method [9], which is based on the corpus training similarity calculation model for sentence similarity analysis. When the feature-sentimental word pair is extracted for the implicit feature-sentimental sentence, the Doc2vec similarity model is called to filter the sentence with high similarity degree to it. Based on the previous feature lexicon and the built feature-sentimental rule base, the extraction rules are set for feature-sentimental sentence pair extraction for explicit feature sentiment sentences, implicit feature sentences and implicit sentiment sentences. Finally, the feature-sentimental word pairs obtained through the rules are summarized to obtain the {feature, sentiment} set.

### 3.4 Sentimental Analysis and Demand Mining

#### 3.4.1 Sentiment Analysis

The study combines the sentimental lexicon with the initial scoring of feature words, and obtains a four-dimensional array {features, attributes, sentimental words, initial scores}. For sentiment words with degree adverb modification, the initial sentiment score needs to be weighted according to the polarity value of the degree adverb.

$$Emotional\ score = \begin{cases} Degree\ adverb\ extreme\ value * initial\ score, \\ \quad \quad \quad degree\ adverb\ modification \\ Initial\ score, no\ degree\ adverb\ modification \end{cases} \quad (1)$$

Weighting the sentiment values yields a six-dimensional array of the {features, attributes, sentiment words, initial scores, degree adverbs, sentimental values}. In order to balance the impact of higher praise rate on the final score because of the highly favorable rate of online reviews, the positive and negative sentiment mean is used in calculating the sentimental mean of each feature. The corrected sentiment score is calculated as follows:

$$Score = \overline{positive} + \overline{negative} \quad (2)$$

Where  $\overline{positive}$  and  $\overline{negative}$  represent the positive sentiment mean and negative sentiment mean. After the degree adverb weighting and summary analysis, the final sentiment analysis results {features, attributes, evaluation times, positive evaluation ratios, negative evaluation ratios, initial scores, sentimental values} are obtained.

#### 3.4.2 User Demand Mining and Classification

Based on sentiment analysis, the KANO model is used to user demand mining and classification. In this study, the initial sentiment scores based on the sentiment lexicon analysis are 1 and -1, and according to the above, the extreme value of the adverb with high intensity is 2. Therefore, the value range of the sentiment value after weighting the degree of adverbs is [-2, 2], and the values of  $\overline{positive}$  and  $\overline{negative}$  are also in the interval [-2, 2], so the value range of score is [-2, 2]. The following we classify the user demand with the value of the sentiment analysis score, and the sentimental value of 1 for user satisfaction. The correspondence between the sentiment value and the KANO model demand type is shown in Fig.2:

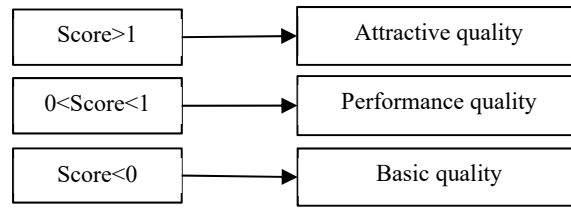


Fig.2 Sentimental value and user demand

### 3.4.3 User Demand Priority

Based on the demand classification, this study comprehensively considers the user attention (the ratio of feature-attribute evaluation times to the total number of evaluations) and the proportion of positive and negative evaluations for product features-attributes to determines the priority of demand. The priority calculation is as shown in the formula.

$$Priority = |\alpha * neg\%| \quad (3)$$

Among them, Priority,  $\alpha$  and  $neg\%$  respectively indicate demand priority, product feature attention and negative evaluation.

## 4. Empirical Analysis

### 4.1 Experimental Setup

The study is based on window10 64-bit operating system, and the experimental platform built by Pycharm and Python 3.6 for text mining and demand analysis. Then we selected the Huawei Sports Bracelet comment review in the Pollen Club, namely the product community review as an experimental data set. It includes the subject of the review, the number of replies, the review ID, the reviewer level, the time, and the text content.

### 4.2 Analysis of Results

We use the usefulness classification model to predict the review data in the dataset and filter the usefulness reviews. The data set has a total of 1212 pieces of text data, and 1095 pieces of useful reviews have been filtered. The usefulness reviews accounted for 90.3%, indicating that the user reviews in the product community are of high quality and can be used as a good source of data for user demand mining. The useful reviews in the data set are segmented and classified. Among them, there are 2438 explicit feature sentiment sentences, 557 implicit feature sentences, and 485 implicit sentiment sentences.

#### 4.2.1 User Demand Mining and Analysis

On the basis of the analysis of sentiment words, authors consider the adjustment effect of degree adverbs on their sentimental tendency, that is, the polarity value of the degree adverbs is used to weight the sentiment scores to obtain the list of {features, attributes, degree adverbs, initial scores, sentimental scores}. The results of the sentimental score of each product feature according to the formula. Through summary analysis, the user in the data set reviews 57 feature attributes of 15 product features. In this way, the user's comprehensive review of 13 product features is obtained, as shown in Fig.3.

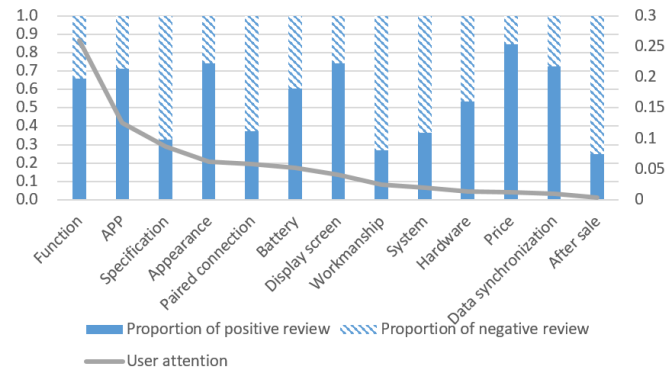


Fig.3 User attention analysis

### 4.2.2 User Demand Classification and Priority

#### (1) User demand classification

The attribute dimensions of the product features are classified according to the conversion rules between the sentiment value and the user demand type. The specific classification and explanations are given in Table 2. Under each type of demand, the priority of feature improvement is determined according to user attention. Product features with higher levels of attention are given priority. In addition to the features already available in the product, users have also introduced new features such as Alipay, cell phone recovery, tapping on the screen, respiratory monitoring and more. For these, further research can be conducted to analyze new market demands and then consider whether to increase the design of next-generation products.

Table 2. User demand classification

| Type                | Product features   | Explanation  |
|---------------------|--|--|
| Basic quality       | APP: function;<br>...<br>workmanship.                    | User standards have not been met and need to be improved to meet the basic needs of users.   |
| Performance quality | APP: installation, software;<br>...<br>intelligent mode. | It is necessary to fully study the user's psychology to improve the features of such products and establish a competitive advantage. |
| Attractive quality  | Specifications: material;<br>display: protective film.   | It exceeds user expectations and needs to be sustained   |

#### (2) User demand priority

According to the user priority calculation method, the user demand under the three demand types are prioritized, and some results are shown in Table 3.

Table 3. User demand priorities

| Feature        | Attribute        | Sentimental score | Type of demand | Attention | Priority |
|----------------|------------------|-------------------|----------------|-----------|----------|
| APP            | Features         | -0.09             | 1              | 10.91%    | 0.03     |
| Battery        | Battery capacity | 0.08              | 2              | 1.60%     | 4.70     |
| ...            | ...              | ...               | ...            | ...       | ...      |
| Specifications | Wristband        | 0.13              | 2              | 7.62%     | 55.50    |
| Appearance     | Appearance       | 0.04              | 2              | 6.21%     | 15.99    |

From the experimental results, although the user reviews of the product community are mostly positive reviews, the gap between the positive and negative reviews is significantly smaller than that of the e-shopping platform. Users in the product community can more objectively review the product,

but mining research may result in incomplete analysis results based on user reviews of the e-shopping platform. Therefore, in the following research on user demands and product innovation, the features of the electronic shopping platform and the high quality of the product community data can be utilized to conduct research and analysis, so as to obtain more comprehensive and accurate user demands.

## 5. Conclusion

The study builds a useful model from a product perspective to capture user demand based on the advantages of online reviews with large amounts of data and easy access. Through analysis, the theoretical and practical significance of this research are summarized as follows: (1) From the product perspective, establish a good review classification model. (2) Based on the study of the current implied features, a new extraction method is proposed to study the implicit emotions of comments. At the same time, this study establishes an effective implicit feature extraction and implicit sentiment analysis method, which fills the research gap in the field of implicit sentiment analysis in sentiment analysis. (3) Using the KANO model to classify the user needs of sentiment analysis results, a new user demand priority calculation method is proposed to obtain a detailed list of user needs. In the next study, deep machine learning and semantic analysis can be combined to automatically judge the usefulness of comments for greater accuracy and efficiency.

## Acknowledgements

This work was supported by grants from the National Social Science Foundation of China (15BSH107), Youth Innovation Foundation of Zhejiang Sci-Tech University (2019Q095) and the National Natural Science Foundation of China (71702170, 71602179).

## References

- [1]. Chatterjee P. Online reviews: do consumers use them? *Advances in Consumer Research*, Vol. 28 (2001), p.1-21.
- [2]. Lis B. In eWOM we trust a framework of factors that determine the eWOM credibility. *Business & Information Systems Engineering*, Vol. 5 (2013) No. 3, p. 129-140.
- [3]. Assimakopoulos C, Papaioannou E, Sarmaniotis C, et al. Online reviews as a feedback mechanism for hotel CRM systems. *Anatolia*, Vol. 26 (2014) No. 1, p. 5–20.
- [4]. Lee SG, Trimi S, Yang CG. Perceived Usefulness Factors of Online Reviews: A Study of Amazon.com. *Journal of Computer Information Systems*, Vol. 58 (2017) No. 4, p. 344–352.
- [5]. Liu Y, Jin J, Ji P, Harding JA, Fung RYK. Identifying helpful online reviews: A product designer's perspective. *Computer-Aided Design*, Vol. 45 (2013) No. 2, p. 180–194.
- [6]. Guo Y, Zhang X, Zhou M, et al. Analysis of the Usefulness of Online Reviews Based on Different Moderating Effects. 2016 9th International Symposium on Computational Intelligence and Design (ISCID). Hangzhou, 2016.
- [7]. Rabelo JCB, Prudencio RBC, Barros FA. Using link structure to infer opinions in social networks. 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Seoul, 2012.
- [8]. Ghiassi M, Skinner J, Zimbra D. Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, Vol. 40 (2013) No.16, p. 6266–6282.
- [9]. Le QV, Mikolov T. Distributed representations of sentences and documents. *International Conference on International Conference on Machine Learning*. Beijing, 2014, p. 1188-1196.