

Analysis and Application of User Comments on a Brand of Mini Washing Machine on the Jingdong E-commerce Platform

Zhang Wang

Wenzhou Polytechnic, Wenzhou 325035, China

Abstract

With the rapid development of e-commerce, a large number of user reviews have been left on e-commerce platforms. These data contain a large number of user experience feelings, which have great data value for enterprises to quickly mine users' concerns and needs. . This paper takes a brand mini washing machine of JD.com as an example, uses machine learning, deep learning and other methods to analyze and mine user comments, through data collection and cleaning, index keyword phrase extraction and sentiment analysis, and analysis and application of comment data. A link to study and analyze the application value of user reviews to enterprises.

Keywords

Comments from E-commerce Users; Opinion Mining; Sentiment Analysis; Text Mining; User Reviews.

1. Introduction

With the popularization of the mobile Internet, e-commerce technology has become more and more convenient for people's lives, and has now become an inseparable part of people's life. There are a large number of user reviews for almost every product in mainstream e-commerce platforms. User reviews are formed after users purchase related products on e-commerce platforms. The content of the reviews is mainly for products or services. The feeling of the whole process of purchasing the product. Enterprises can quickly understand users' concerns and needs for products through user reviews, as well as the shortcomings of the products themselves and problems related to store services, and through closed-loop management, adjust and improve products and stores to improve user satisfaction. . However, based on the massive user comment data, enterprises need a set of methods that can quickly realize the value mining in the massive user comments.

This project intends to conduct text analysis and mining of user reviews by means of machine learning, deep learning and other methods through the review analysis of a brand mini washing machine in Jingdong, analyze the user's concerns and needs from the reviews, and develop products according to user needs. Optimize products and improve services according to user complaints, so as to achieve the purpose of enterprise product optimization and service improvement.

2. Research Status at Home and Abroad

Product review analysis generally includes two tasks: review sentiment analysis and opinion mining.

Natural Language Processing (NLP) is an important application area of deep learning. In the development of the past few decades, the mainstream research direction of natural language processing is based on statistical models. Since the beginning of the 21st century, some papers related to natural language processing based on deep learning have been published one after another.

2.1. Research Status of Opinion Extraction

The development of English opinion extraction has developed from single text or sentence processing to user-oriented research, which is analyzed and processed by natural language processing technology. In 2003, Yu et al. used polarity combination to analyze semantic polarity from sentence level and document level, and used statistical methods to identify sentence polarity and calculate document polarity[1]. In 2005, the IBM Almaden Research Center developed the WebFountain system, which uses grammatical relations to find out the relationship between topics and modifiers[2]. The Opinion Observer system led by Liu of the University of Illinois in the United States can obtain online user comment information, and then conduct statistics on the advantages and disadvantages of each feature, and conduct a visual comprehensive quality comparison.

The earliest research on Chinese opinion extraction can be traced back to 2005, Benjamin K. of City University of Hong Kong, Y. Tsou and Raymond W. M. Yuen et al. conducted part-of-speech analysis on 600 news reports in four cities of Beijing, Shanghai, Hong Kong, and Taipei, and analyzed the reputation evaluations of four famous politicians from the document-level[3]. The method adopted is to use the marked corpus to obtain the polar elements of the text, and then use the three measures of density (Density), breadth (Spread) and intensity (Intensity) to divide the polar elements (Polar Elements) in the article. Hu and Liu proposed to perform part-of-speech tagging on the review corpus to extract the thing files composed of nouns and noun phrases, and then use association rules to filter candidate product features[4]. Li Shi et al. improved the method proposed by Hu et al. for the characteristics of Chinese reviews, and achieved good results in mining product features expressed in Chinese[5]. In review data, there is usually a specific relationship between features and views, and some scholars use this feature sequence and syntactic dependency to extract product features and views. In 2012, Kumar et al. used syntactic dependencies to identify product features and opinions[6]. Qiu et al. used the bidirectional propagation algorithm to extract features and opinions, and improved the accuracy by constructing product features. Hao Zhifeng[7] proposed in 2018 a sequence tagging model that integrates semantic relations and parts of speech to extract viewpoint dimensions.

So far, opinion extraction systems have emerged in the English domain, which can target different corpora, domains and users. In contrast, the system in the Chinese field is relatively simple, involving only some industries and fields, and there is still huge room for development.

2.2. Status Quo of Sentiment Analysis Research

At present, the main application areas of sentiment analysis are e-commerce and social networking sites, such as the analysis of comment information in e-commerce platforms, and the user's comment guides in social media platforms, all of which belong to the category of sentiment analysis. Taking Taobao website as an example, improving the sentiment analysis of users' comment data can dynamically push products for users, or adjust the order in which users view products, so as to improve users' desire to buy.

At present, a large number of researchers have used machine learning for sentiment classification. In 2002, Pang et al. used supervised machine learning Support Vector Machine (SVM) for sentiment classification of movie reviews[8]. Turney used the unsupervised method PMI-IR (Pointwise Mutual Information and Information Retrieval) for sentiment classification of consumer reviews[9].

Domestic research is relatively late, mainly analyzing Chinese comment data, such as e-commerce platforms, news, and Weibo. There is a big difference between English and Chinese. English has spaces as separators, while Chinese has no clear boundaries. Only punctuation marks are used as sentence segmentation. Therefore, Chinese word segmentation is more difficult.

However, the research methods in Chinese and English are similar. There are currently three types of mainstream Chinese sentiment analysis methods. One is the method based on the sentiment dictionary. This type of method mainly relies on the sentiment dictionary and uses artificially designed rules to classify sentiment. The second type is the traditional method based on machine learning. This method needs to first mine the features of words, and then use classification algorithms such as Support Vector Machine (SVM), Naive Bayes (NB), and random forest to classify the words. The text carries out the judgment of the sentiment tendency. The last category is the method based on deep learning, that is, using different neural network models, mapping the text to the vector space to obtain the numerical representation of the word, and then inputting the vector into the classifier.

However, in practical applications, sentiment classification of Chinese comments also faces challenges in at least two aspects. First of all, online user comment information covers a wide range of fields. Take Taobao and online news as examples, including electronic products, food and daily necessities, etc., while online news generally involves entertainment, sports, and politics. Furthermore, the labeling of the dataset required for training is time-consuming and labor-intensive.

3. Research Process and Methods

The overall research process of this paper is shown in Fig.1. The specific steps are divided into data acquisition, text preprocessing, construction of commodity index system, sentiment analysis and result analysis and application.

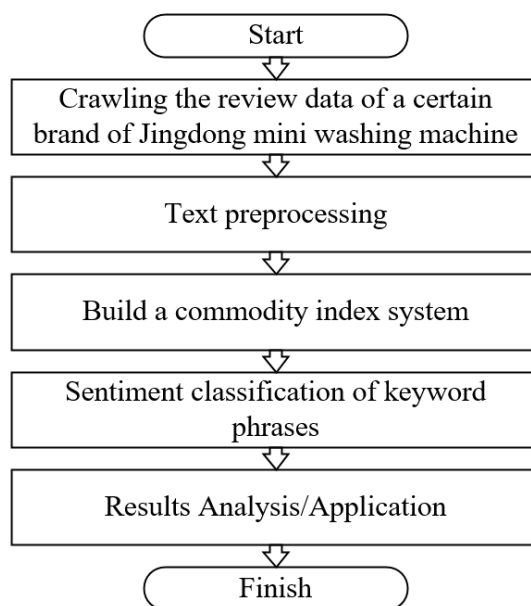


Fig 1. Research flow chart of e-commerce review analysis and application

3.1. Get Data

Collect data related to user comments from e-commerce platforms by means of web crawler. This article captures the user evaluation data of a certain brand flagship store on the Jingdong e-commerce platform and some two mini washing machines in various specialty stores, as shown in Fig.2. The collected data is mainly about product-related dimensions, as follows: store name, product model, user evaluation, and evaluation time.

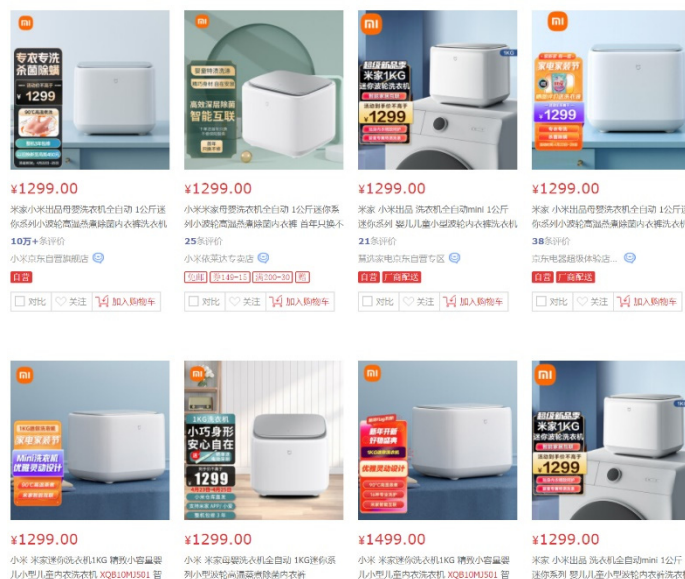


Fig 2. A brand of mini washing machine on JD.com

According to the above process and measures, the web crawler is performed on the Jingdong platform, and the obtained data is saved in excel. The web crawler collected about 3749 pieces of data this time, and the data obtained is shown in Table 1.

Table 1. Sample user comment data

Store Name	Product model	User Comments	Evaluation time
Xiaomi Jingdong self-operated flagship store	XQB10MJ501	The master is very nice and patient. like	2022-04-15 21:10
Xiaomi Jingdong self-operated flagship store	Xqb10mj501	Jingdong's self-operated flagship store is very popular.	2022-04-15 20:23
Xiaomi Jingdong self-operated flagship store	XQB10MJ501	Satisfied after installation	2022-04-15 12:10
Xiaomi Jingdong self-operated flagship store	XQB10MJ501	If you have children at home, or female friends, you should really prepare a Xiaomi underwear washing machine, which can be sterilized at high temperature, and can also sterilize and remove mites.	2022-04-15 09:42
Xiaomi Jingdong self-operated flagship store	XQB10MJ501	The Xiaomi washing machine looks amazing, and it is convenient to use. I will not wash underwear by hand in the future.	2022-04-14 21:18
Xiaomi Jingdong self-operated flagship store	XQB10MJ501	Oops, my hands are finally freed. The underwear is clean. Simple and clean.	2022-04-14 19:36
Xiaomi Jingdong self-operated flagship store	XQB10MJ501	The installer is very patient.	2022-04-14 13:10
Xiaomi Jingdong self-operated flagship store	XQB10MJ501	The washing machine is good and has a good texture. Xiaomi's things are still good.	2022-04-14 12:26
Xiaomi Jingdong self-operated flagship store	XQB10MJ501	It's even smaller than expected, just the size of an ordinary washbasin. It's okay to wash	2022-04-13 21:15

3.3. Build a Commodity Index System

Through TF-IDF keyword extraction algorithm, comment segmentation, and regular expressions, commodity indicators and their index keyword phrases are formed, and the sentiment classification of key phrases is carried out through a variety of machine learning classification algorithms, deep learning algorithms LSTM and CNN. Form a commodity index system.

In the user comment content, there are often multiple comment objects contained in the same user comment, and it contains the user's emotional tendency when writing the comment content. In order to more quickly understand the comment object in user comments and the emotional tendency of users when commenting on the object, we usually use indicators to measure, so this paper designs a set of commodity indicators system, which consists of commodity indicators, indicator keywords and keyword phrases constitute.

Combined with the content of user comments, the user comments are segmented through python, and the commodity indicators are adjusted according to the word frequency statistics to form commodity indicators. The TF-IDF keyword extraction algorithm in the python stuttering word segmentation package is used, and the importance of the vocabulary is calculated by using the TF-IDF value of the vocabulary in the user comments as the weight. After processing the user comment keyword extraction algorithm, the number of extracted comment keywords has been significantly reduced, and the number has been reduced to 70% of the amount before the adoption of the measures. In the extracted user comment keyword table, the comment keywords and commodity indicators are mapped one by one by manual screening, so as to form commodity indicators and their keywords. This time, 140 words were screened from the comment keywords, and formed a mapping relationship with the commodity index. The mapping table is shown in Table 2.

Table 2. Commodity indicators and their keyword examples

First-level indicator	Secondary indicators	Three-level indicator	Key words
product experience	Exterior	Size	area
product experience	Exterior	Size	space
product experience	Exterior	Size	size

3.4. Sentiment Classification of Keyword Phrases

By building a text classification model to classify keyword phrases, this paper tries the SnowNlp algorithm, LSTM and CNN algorithms under Keras. Among them, in the default state, the accuracy of the SnowNlp algorithm is 67.8%, and the LSTM model and CNN model in deep learning are further used for sentiment classification of keyword phrases. First, we preprocess the data according to the requirements of the LSTM and CNN models, and adjust the parameters corresponding to the models to classify the sentiment of keyword phrases. Through the effect evaluation of the training model, it can be seen that LSTM is better than CNN in the training set, but slightly worse than CNN in the validation set and test set. As shown in Table 3, CNN is finally selected for sentiment classification of keyword phrases.

Table 3. Model effect evaluation of LSTM and CNN

Model	Accuracy	F1
LSTM	93.01%	94.67%
CNN	93.80%	94.33%

3.5. Results Analysis and Application

Analyze and mine the data after product indexation, draw corresponding conclusions, and analyze users' concerns and needs, which can be applied to a brand's product development and network marketing strategy optimization for these two mini washing machines.

4. Experimental Verification

We will collect the user comment data from the Jingdong e-commerce platform, through data cleaning, index construction, keyword extraction, keyword phrase extraction and sentiment classification, to form indexed user comment data, as shown in Table 4.

Table 4. Example of application results of user review model

Model	Store	Comments	First-level indicator	Secondary indicators	Three-level indicator	Positive or Negative	Match
XQB10MJ501	Xiaomi Jingdong self-operated flagship store	The customer service attitude is good	service quality	customer service	Customer service attitude	P	service attitude is good
XQB10MJ501	Xiaomi Jingdong self-operated flagship store	The washing machine is good, very textured	product experience	quality	product material	P	textured
XQB10MJ501	Xiaomi Jingdong self-operated flagship store	Very small, does not take up space	product experience	size	product size	P	Very small
XQB30MJ101	Xiaomi Jingdong self-operated flagship store	good. still mute	product experience	volume	mute	P	still mute
XQB30MJ101	Xiaomi Jingdong self-operated flagship store	the wire is too short	product experience	quality	Effect	N	the wire is too short

Through user comments and different dimensions, we can analyze the product model indicators through product analysis. We can see which indicators of the two mini washing machines perform poorly and need to be improved urgently. Table 5 shows the product analysis results.

According to the data analysis results, the overall customer feedback of the brand's two mini washing machines is good, and the negative evaluation rate is low. Negative reviews are mainly dissatisfaction with the after-sales service of the product and dissatisfaction with the quality of the product. Based on the above conclusions, it is recommended that the merchants improve the quality of after-sales service. After the customer receives the goods, they should arrange for the installation personnel to provide on-site service as soon as possible, and the installation must be in place; and the after-sales service should be online for 24 hours, and the automatic

reply should be set up, and the manual reply should also be timely. In addition, optimize product performance and quality, use more durable materials, reduce the volume of laundry, reduce the amount of water used, and improve the effect of laundry.

Table 5. Product Analysis

Product model	Total number of comments	Total negative rating	Price	Quality	Serve	Exterior	Overall product
XQB10MJ501	1821	2.63%	0.26%	0.56%	1.22%	0.14%	0.45%
XQB30MJ101	1697	2.38%	0.24%	0.48%	1.30%	0.28%	0.18%

5. Conclusion

This paper collects user comments and related data by means of web crawler, and analyzes and mines user comments by formulating a commodity index system, finds out users' concerns and needs, finds the advantages and disadvantages of the product, and conducts research on the shortcomings. Improve optimization and achieve closed-loop management through continuous monitoring. At present, there are still many deficiencies in this paper. First, there is still a lot of manual workload based on human experience in the expansion of index keywords and the extraction of keyword phrases. In addition, in the preprocessing, fake reviews cannot be fully identified, which reduces the reliability of the analysis results. The method applied in this paper still has room for further improvement.

Acknowledgments

This research is supported by the school-level general scientific research project of Wenzhou Polytechnic (WZY2021026).

References

- [1] Yu H, Hantzivassiloglou V. towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences[C]//Proceeding of the 8th Conference on Empirical Methods in Natural Language Processing, 2003: 129-136.
- [2] Yi J, Niblack W. Sentiment mining in Web Fountain[C]//Proceedings of IEEE Computer Society Press, 2005: 1073-1083.
- [3] Benjamin T, Raymond Y, et al. Polarity Classification of Celebrity Coverage in the Chinese Press[C]//Proceedings of the International Conference on Intelligence Analysis, 2005: 102-108.
- [4] Hu M, Liu B. Mining Opinion Features in Customer Reviews[C]//Proceedings of 9th National Conference on Artificial Intelligence, 2004: 755-760.
- [5] Li Shi, Ye Qiang, Li Yijun. Research on Product Feature Mining Method of Chinese Online Customer Reviews [J]. Journal of Management Science, 2009, 12(2): 142-152.
- [6] Kumar R, Raghuvver K. Dependency Driven Semantic Approach to Product Features Extraction and Summarization Using Customer Reviews[C]//Proceedings of the 2nd International Conference on Advances in Computing and Information Technology, 2012:225-238.
- [7] Hao Zhifeng, Huang Hao, Cai Ruichu. Fine-grained opinion analysis based on multi-feature fusion and bidirectional RNN [J]. Computer Engineering, 2018, 01(07): 199-204.
- [8] Pang B, Lee L, Vaithyanathan S. Thumbs Up: Sentiment Classification using Machine Learning Techniques[C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 2002: 79-86.

- [9] Turney P. Thumbs Up or Thumbs Down: Semantic Orientation Applied to Unsupervised Classification of Reviews[C]// Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 2002: 417-424.