

Research on Sentiment Analysis of E-commerce Live Comments based on Text Mining

Wei Xiong, Yue Zuo, Menglin Zhang, Chenxi Zhang, Caiqing Guo

Nanjing Audit University, Nanjing, Jiangsu 211815, China

Abstract: With the rapid rise of live e-commerce in the field of e-commerce, the rich emotional information generated by users in live comments has become an important research object. This study takes e-commerce live comments as samples, uses text mining technology combined with machine learning methods, builds sentiment dictionary, trains sentiment analysis model and uses SnowNLP library for sentiment analysis, and deeply digs users' emotional tendency towards goods, services and shopping experience. For negative comments, word cloud is generated through word segmentation and word frequency calculation to visually show users' dissatisfaction and provide substantial support for updating the emotional dictionary. Through this comprehensive analysis, it provides a deeper user insight for e-commerce enterprises, and provides strong support for operational decisions and service optimization. Future research can focus on the application of deep learning techniques in sentiment analysis to further improve model accuracy and adaptability.

Keywords: Text Mining; E-commerce Live Broadcast; Sentiment Analysis.

1. Research Background

With the rapid development of the Internet and the increasing popularity of digital technology, the e-commerce industry has ushered in a rapid development stage. In this process, the e-commerce live broadcast as a new sales model quickly emerged. Through the online live broadcast platform, real-time video is used as the medium to combine commodity display and shopping experience, providing consumers with a more intuitive and interactive shopping method. However, with the rise of e-commerce live broadcasting, the comment information generated by a large number of users has also doubled, which contains rich emotional information.[1]

The emotional information in e-commerce live comments carries consumers' attitudes and feelings towards products, services and shopping experience. For e-commerce enterprises, accurately grasping users' emotional tendencies is of great commercial significance.[2] Through automated analysis of a large number of review texts, users' satisfaction with goods, evaluation of services and attitude towards shopping experience can be mined, and targeted improvement suggestions can be provided for e-commerce enterprises to improve user satisfaction and shopping experience.[3]

However, the characteristics of e-commerce live commentary text include the diversity of language expression, the complexity and timeliness of word-of-mouth information, etc., which bring a series of challenges to sentiment analysis. Therefore, this study aims to use advanced text mining technology, combined with machine learning and natural language processing methods, to in-depth mining and analysis of emotions in e-commerce live comments, in order to improve the understanding of user emotions, and provide more effective decision support and operation optimization strategies for e-commerce enterprises.

2. Literature Review

In the field of e-commerce live broadcasting, sentiment analysis has become a research direction that has attracted much attention, providing profound user insight and

commercial value for e-commerce enterprises.[4] A number of scholars have made positive contributions in this field, through various text mining techniques and sentiment analysis methods, in-depth research on the user emotion information contained in e-commerce live comments.

Wilson et al. (2009) point out that identifying emotional polarity in text is a key task in sentiment analysis.[5] They explored the technology of emotion dictionary construction and emotion classification in sentiment analysis, which provided a theoretical basis for in-depth exploration of e-commerce live comments. In addition, Alaei et al. (2019), through their study of large-scale social media data, emphasized the importance of considering context and context in sentiment analysis, which is crucial for understanding complex emotional expression in live comments of e-commerce.[6]

With the development of deep learning technology, some scholars began to explore the application of neural networks to sentiment analysis of e-commerce live comments. For example, the convolutional neural network (CNN) proposed by Wang (2019) has achieved remarkable results in text classification tasks, inspiring subsequent researchers to explore the application of deep learning methods in e-commerce live commentary.[7] On this basis, Umer et al. (2021) further introduced short term memory network (LSTM) and other models to improve the accuracy of sentiment analysis and provide new ideas for processing long text in live comments of e-commerce.[8]

To sum up, scholars' research in the field of sentiment analysis of e-commerce live commentary provides us with rich theoretical and methodological support to better understand user emotions and optimize e-commerce operations.

3. Data Collection

The methods of data acquisition mainly include two ways: web crawler and user feedback. First of all, through the web crawler technology, a large number of comment text data can be captured from the e-commerce live broadcast platform. For example, using Python to write a crawler program, by directly

accessing the platform API or simulating the user's browsing operation, to obtain the user's comment information left in the live e-commerce program. Secondly, through user feedback, users' emotional experience in e-commerce live broadcasting can be actively collected by means of questionnaires, social media mining and other means. For example, questionnaire surveys for specific live events are designed to collect users' comments on goods, hosts and services to obtain more detailed and in-depth emotional information. The combination of these two methods can build a rich and comprehensive dataset of live e-commerce comments, providing strong support for subsequent sentiment analysis.

4. First, A Calculation Method of Emotion Value based on the Combination of Dictionary and Machine Learning

4.1. Construction of Emotional Dictionary

The construction of sentiment dictionary is one of the important steps of sentiment analysis based on text mining. Firstly, a corpus containing positive, negative and neutral emotions is collected, which can be screened from the general corpus. Then, using manual annotation or automatic annotation, each word is assigned an emotion polarity value, such as positive emotion is assigned a positive value, negative emotion is assigned a negative value and neutral emotion is assigned a zero value. It is also possible to choose appropriate emotional dictionaries according to the characteristics of e-commerce live commentary. Common emotional dictionaries include the emotional dictionary of CNK.com, the Chinese emotional word NTUSD organized by Taiwan University, the Dictionary of Students' Praise and Criticism, and the Emotional Dictionary of Dalian Institute of Technology.

4.2. Training the Sentiment Analysis Corpus

Training emotion analysis models requires a corpus with emotional labels as a training set. This corpus should contain a variety of e-commerce live commentary texts, covering a variety of emotional expressions and contexts. First, a part of the corpus was manually labeled to label each comment with its emotion, with positive, negative and neutral as the main labels. Then, using machine learning algorithms, such as support vector machines (SVMS), Naive Bayes, or deep learning models, the corpus is trained so that the model can learn the relationship between text features and emotions. The trained model can perform sentiment analysis on unlabeled live e-commerce comments and predict their emotional polarity.

5. E-commerce Live Emotion Analysis

5.1. Analyze the Emotional Tendency of Comments

Sentiment analysis aims to determine the emotional tendencies expressed in the comments, whether they are positive or negative, subjective, or objective. Specifically, the SnowNLP library in Python can be used to conduct sentiment analysis on the text of live comments on e-commerce. The sentiment function of SnowNLP is used to analyze the sentiment value of each comment and return a probability value in the range of [0,1], indicating the sentiment tendency of the comment. If the probability value is greater than 0.5,

the emotion tends to be positive; otherwise, it tends to be negative. This method effectively quantifies the emotion of the comments and provides specific emotional values for subsequent analysis.

5.2. Analysis of Negative Comments

Based on the results of sentiment analysis, the negative comments are extracted, and the word segmentation and word frequency calculation are carried out using the Jieba module of Python. By counting high-frequency words in negative comments and generating word clouds, we can intuitively understand users' dissatisfaction in webcasts. These high-frequency words not only provide specific expressions for negative emotions, but also provide substantial data support for the constant updating and improvement of emotion dictionaries.

In the word cloud, font size reflects the distribution of word frequency, providing a visual tool for further analysis. Through the word cloud image, we can find the problems that users generally dislike in the live broadcast environment of e-commerce. This provides guidance for e-commerce live streaming platforms to regulate content and improve services to enhance user experience.

6. Conclusion

As an innovative sales model, e-commerce live broadcasting provides consumers with a more intuitive and interactive shopping experience, and brings new business opportunities and challenges for e-commerce enterprises. Through the emotional analysis of e-commerce live comments, we can deeply understand users' attitudes, preferences and dissatisfaction, and provide powerful decision support for enterprises.

This study constructed a comprehensive text mining technology, including sentiment dictionary construction, machine learning model training and sentiment analysis method based on SnowNLP library, to analyze e-commerce live comments in a comprehensive and in-depth way. Through the analysis of negative comments and the generation of word cloud, we can better understand the dissatisfaction points of users, and provide substantial data support for updating the sentiment dictionary.

Sentiment analysis of live comments on e-commerce is still a growing field, facing challenges such as linguistic diversity and emotional complexity. Future research can further explore the application of deep learning techniques in sentiment analysis to improve the accuracy and generalization ability of the model.

Acknowledgments

Fund Project: Innovative Training project for college students in Jiangsu Province in 2023 "Research on the" Double-edged sword "effect of dialect narration on the effect of live broadcast in assisting agriculture from the perspective of Big Data". (202311287019Z).

References

- [1] HU Y, OTHERS. Research on the commercial value of Tiktok in China[J]. Academic Journal of Business & Management, 2020, 2(7): 57-64.
- [2] ROSA J A, MALTER A J. E-(embodied) knowledge and e-commerce: How physiological factors affect online sales of

- experiential products[J]. *Journal of Consumer Psychology*, 2003, 13(1-2): 63-73.
- [3] DEBORTOLI S, MÜLLER O, JUNGLAS I, et al. Text mining for information systems researchers: An annotated topic modeling tutorial[J]. *Communications of the Association for Information Systems (CAIS)*, 2016, 39(1): 7.
- [4] KARN A L, KARNA R K, KONDAMUDI B R, et al. Customer centric hybrid recommendation system for E-Commerce applications by integrating hybrid sentiment analysis [J]. *Electronic Commerce Research*, 2023, 23(1): 279-314.
- [5] WILSON T, WIEBE J, HOFFMANN P. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis[J]. *Computational linguistics*, 2009, 35(3): 399-433.
- [6] ALAEI A R, BECKEN S, STANTIC B. Sentiment analysis in tourism: capitalizing on big data[J]. *Journal of travel research*, 2019, 58(2): 175-191.
- [7] WANG R, LI Z, CAO J, et al. Convolutional recurrent neural networks for text classification[C]//2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019: 1-6.
- [8] UMER M, ASHRAF I, MEHMOOD A, et al. Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model[J]. *Computational Intelligence*, 2021, 37(1): 409-434.