Sentiment Analysis of Mobile Phone Reviews Using XGBoost and Word Vectors

Zekai Wang

Dalian University of Technology, Dalian, Liaoning, 116000, China

Abstract. Consumer reviews are an important source of data used to judge and examine consumer sentiment, and data mining for reviews of electronic products is an important way to help improve the design of electronic products. The research is based on the consumer reviews of online cell phone e-commerce, The paper constructs a sentiment dictionary in this field based on the Sentiment Oriented Point Mutual Information (SO-PMI) algorithm, and the sentiment weight of the review word vectors. An extreme Gradient Boosting Tree (XGBoost) is used to integrate word vectors and a Large Language Model (LLM) to construct a sentiment recognition model, and finally, a review sentiment index is derived, which unfolds from multiple dimensions to analyze the sentiment tendency in consumer reviews. The empirical analysis shows that the accuracy, recall, area under the curve (AUC), and other validation indexes of the constructed sentiment recognition model are further improved compared with the LLM model, which has a certain application value. When applying the weighted word vector method, the model has been significantly improved compared with the LLM model, the accuracy is increased by 5%, the accuracy is increased by 10%, and the comprehensive accuracy is increased by 2% after the comprehensive application of the two.

1 Introduction

With the advent of the Internet era, e-commerce has shown a booming trend. According to statistics, the transaction volume of the two major e-commerce platforms, Tmall and Jingdong, reached 540.3 billion yuan and 349.1 billion yuan respectively during the "Double Eleven" in 2023. At the same time, as the variety of online products continues to be enriched, as well as the convenience of online shopping, more and more consumers are choosing to conduct transactions on the Internet, and the number of online reviews is also proliferating [1]. For consumers, online reviews have become one of the most important sources of information for making purchase decisions. For merchants, the information embedded in product reviews contains consumers' subjective emotions, and text sentiment analysis allows merchants to understand consumers' feedback on their products and services promptly so that they can optimize their products and services and gain more profits.

The text sentiment analysis research methods are divided into sentiment dictionary-based and machine learning-based [2]. The dictionary-based analysis method first processes the extracted text by segmentation, and then constructs the sentiment dictionary of a certain domain, based on this dictionary, the number of positive and negative sentiment words in the

https://doi.org/10.1051/itmconf/20257003018

text is counted by using linear algebra and statistical analysis, to determine the sentiment category of the text [3]. The machine learning-based analysis method is to manually annotate the text first and divide it into the training set and test set, using the algorithms of machine learning such as Support Vector Machine, K Nearest Neighbors, and Plain Bayes to learn the features of the training set and build a specific classification model, and then apply the classification model to the test set thus making a judgment on the classification accuracy [4].

Although text sentiment analysis and research on online electronic products are currently being hotly carried out, the sentiment analysis of online cell phone e-commerce reviews has not yet been researched on the Large Language Model (LLM). It has not yet constructed a sentiment dictionary for it, if the use of a generalized sentiment dictionary to classify the sentiments will inevitably lead to the situation that the prediction accuracy is not high, it is not insufficient to meet the requirements of improving the accuracy rate [5]. Therefore, this paper randomly selects the cell phone products with 100,000-level comments on Jingdong Mall as the research object, and after constructing a sentiment dictionary in the field of online cell phone e-commerce, proposes a new sentiment analysis model based on the Extreme Gradient Boosting algorithm integrating the weighted word vectors and the LLM, to accurately determine the sentiment polarity of the consumer comments, and constructs a sentiment index based on the analysis results, thus reflecting the current problems in the online cell phone e-commerce. reflect the current problems of online cell phone e-commerce and help merchants grasp the sentiment trends in the field of online cell phone e-commerce.

2 Constructing a lexicon of domain emotions

2.1 Data collection and polarity processing

First, under the Jingdong hot-selling mobile phone category, 5 products with 100,000-level reviews are randomly selected according to sales volume. Python programs are written to crawl more than 2,000 reviews each, with a total of 11,865 Chinese reviews constituting the corpus., and the sentiment polarity of the corpus was discriminated by the star rating generated during the review, with 1, 2 stars for negative reviews, and 3, 4, and 5 stars for positive reviews, to generate a positive sentiment corpus and a negative sentiment corpus, see Table 1.

Table 1. Emotional corpus [6]

emotional level sex	Comment content	Number of comments
positive	Appearance: A small-screen flagship that you won't be able to put down. Screen sound: The screen is silky smooth and the sound quality is high. Photo effects: It was released two years ago, but the camera is still fully up to date and there is no pressure. Running speed: It is fast, and the speed has increased from 11 to 13. Standby time: You can go a whole day without playing games and still have enough power to charge the phone.	**7451**
	Appearance: A small-screen flagship that you won't be able to put down. Screen sound effects: The screen is silky smooth and the sound effects are high-end. Photo effects: It's been released for two years, but the camera is still fully online and there's no pressure. It's super cost-effective! Haha, a dream phone. It's brand new, so you can charge it with confidence!	

https://doi.org/10.1051/itmconf/20257003018

	The delivery was fast, it arrived the day after I placed the order. This color is pleasing to the eye, it feels good in the hand, it charges quickly, it's done in about half an hour, and it has a long battery life.	

Continue Table 1.

negative	It feels a bit laggy. Sometimes when I'm watching a video and open other apps, the video will automatically hang up. When I use other apps to read comments, I can only see half of the words. Previously, when I used other phones, I could log in to WeChat on the computer and watch someone's video on the computer while watching someone else's video on the phone. Now, when the computer will automatically hang up if the phone is connected to someone else's video. Sometimes when I click on some software, it will often hang for a few seconds before working.	**4414**
	It's the first time I've used a Xiaomi phone, and I'm honestly not impressed. It only took a week of use before it started to lag. It's not as good as the Honor Magic 2 I bought three years ago, which has a Kirin 980 processor and the same amount of RAM. The signal is also not as good as Honor's. Sometimes the internet connection drops, and when I make a call, I can't hear the person on the other end, so I have to replay the call. I won't be buying Xiaomi phones in the future.	
	...	

2.2 Extracting sentiment seed words as well as candidate words

To accurately recognize the multiple entities contained in the comments as well as the complex semantic structure, jieba-paddle based on two-way control gate unit recurrent neural network pre-training was finally used to segment the corpus, and imported into the deactivation word list of Dalian University of Technology (DUT) to remove deactivated words and de-punctuation marks. Then the word frequency was counted, and 20 positive emotion words and negative emotion words each were selected, and fused with some words in the manually screened Chinese emotion vocabulary ontology library of Dalian University of Technology to form the final emotion seed words, see Fig. 1.

(a) (b)

Fig. 1 Seed words word cloud map, (a) positive emotion seed words; (b) negative emotion seed words.

A commonly used method for determining the sentiment polarity of Chinese words is a computational method based on the Point Mutual Information (PMI) algorithm [7]. PMI is used to determine the probability of a word appearing simultaneously with a benchmark word, i.e.

$$\mathrm{PMI} = \log_2 \left[\frac{P(\mathrm{word1},\mathrm{word2})}{P(\mathrm{word1})P(\mathrm{word2})}\right) \quad (1)$$

Where: P refers to the probability of word1 and word2 appearing in the corpus at the same time, if the two words are independent of each other, P(word1, word2) = P(word1) P(word2), which leads to PMI = 0. If PMI > 0, it means that the two terms are related, and the larger the value, the stronger the correlation; if PMI < 0, it means that the two terms are not related.

The SO-PMI algorithm is based on the PMI algorithm to determine the degree of correlation between the new vocabulary and the emotion seed words [8]. If the correlation with the positive sentiment seed word is large, the word can be classified as a positive sentiment word; if the correlation with the negative sentiment seed word is large, the word can be classified as a negative sentiment word. The SO-PMI algorithm is:

$$\text{SO-PMI}(\text{word}) = \sum_{i=1}^{\text{num(pos)}} \text{PMI}(\text{word},\text{pos}_i) - \sum_{i=1}^{\text{num(neg)}} \text{PMI}(\text{word},\text{neg}_i) \tag{2}$$

Where: num(pos) and num(neg) are the total number of positive and negative sentiment seed words respectively. If SO-PMI>0, the word can be judged as a positive sentiment candidate; if SO-PMI=0, it can be judged as a neutral word; if SO-PMI<0, the word can be classified as a negative sentiment candidate. The SO-PMI algorithm is implemented in Python, importing the corpus of Jingdong cell phone reviews to be processed as well as the positive and negative sentiment seed words, and finally outputting the positive and negative sentiment candidates, see Figure 2.

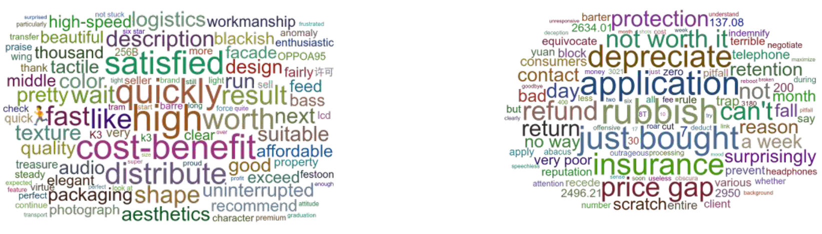

Fig. 2 Candidate word cloud, (a) positive emotion candidate words; (b) negative emotion candidate words.

2.3 Consolidation of the universal sentiment dictionary

The obtained emotion candidates were manually eliminated from the extremely unreasonable words and were de-emphasized and merged with the three most widely used general emotion dictionaries (HowNet Emotion Dictionary, NTUSD Simplified Chinese Emotion Dictionary, and Dalian University of Science and Technology Chinese Emotion Vocabulary Ontology) to form the final emotion dictionary for online cell phone e-commerce domain. A total of 6451 positive and 6451 negative emotion words were counted.

3. Emotion recognition model based on XGBoost integrated weighted word vector and LLM

This study proposes a method that combines sentiment lexicon and machine learning algorithms to construct a sentiment analysis model. First, after processing the text data using the word vector technique, LLM is introduced to determine the sentiment polarity of the comments at the same time. In the field of natural language processing, LLM has been proven to be very effective in recognizing the sentiment of text [9]. First of all, LLM performs better than traditional models in understanding context, and can accurately recognize textual sentiment [10]. Secondly, LLMs are usually able to handle more complex sentence structures and recognize finer levels of sentiment. It has been shown that LLM can capture subtle emotional differences, even in texts with implicit or complex emotional expressions [11].

To achieve accurate judgment of online mobile e-commerce consumer reviews for sentiment classification, the weighted word vectors based on the sentiment dictionary of mobile e-commerce and the sentiment polarity results of LLM judgment are used as the input layer of XGBoost, which is used for integrated training to obtain the optimal sentiment judgment results.

3.1 Weighted word vectors

In training the sentiment recognition model, the first problem faced is to convert the comment text into word vectors that can be effectively recognized by the model. Using the Bi-GRU-based jieba-paddle model of the word segmentation technique, through this precise method, the key sentiment information can be effectively extracted from a large number of comments, and at the same time, the dimension of the word vectors is significantly reduced, providing a solid foundation for the subsequently constructed sentiment analysis model.

After preserving the entity core, the comments are converted into word vectors using the word frequency-inverse document frequency technique TF-IDF. The score of each entity is first calculated by TF-IDF. The TF-IDF algorithm first calculates the word frequency and inverse document frequency of each word and then multiplies the TF value with the IDF value and the final value represents the importance weight of the entity in the document. The TF-IDF, TF, and IDF are calculated as

$$TF - IDF \ = TF \ x \ IDF \tag{3}$$

$$TF = \ NAW/TNW \tag{4}$$

where:TF means Word Frequency, IDF means Inverse Document Frequency, NAW means Number of a word appears in a text, TNW means Total Number of words in the article.

After obtaining the word vector mentioned above, the word vector is weighted based on the SO-PMI score of the affective candidate words in the field of online mobile e-commerce constructed earlier.

Figure 3 shows the sentiment entity words selected from more than 10,000 comments and their polarity judgments made through the sentiment dictionary. Considering the wide range of SO-PMI scores of sentiment candidate words, the Z-score standardization method is used to map the SO-PMI score of each sentiment candidate word to the interval of -1 to 1, which is used as the weighting coefficient of the sentiment word. It can be observed from Figure 3 that the word weight is mainly concentrated in the range of 0-0.2, which also reflects the accuracy of the polarity judgment of sentiment words. Finally, the corresponding weighted coefficients are multiplied by their TF-IDF scores to adjust the final weights.

In feature selection, only the top 388 most representative features are retained, and a minimum document frequency of 1 is set. This reduces the complexity of the word vector as much as possible while retaining the most informative lexical features, and ultimately constructs weighted word vectors with high value.

 https://doi.org/10.1051/itmconf/20257003018

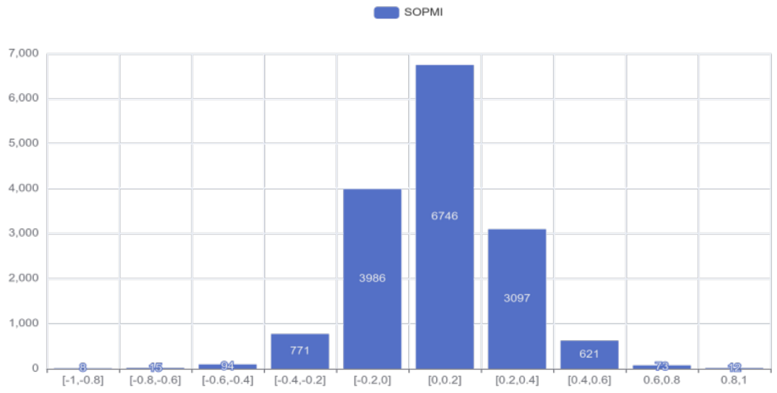

Fig. 3 Distribution of weighted coefficients after SO-PMI standardization (Photo/Picture credit: Original).

3.2 Recognizing emotional polarity based on LLM

LLM models usually cover rich language samples and scenarios and are trained on large and diverse text datasets, so LLMs can deal with multiple languages and expressions, in the field of natural language processing, LLMs have proven their excellent performance and application value, whether in understanding complex language structures or in processing large-scale text data, LLM has demonstrated its critical role in both understanding complex linguistic structures and processing large-scale text data.

To make accurate sentiment polarity judgments, three LLMs applicable to Chinese are used, including Baichuan2-7B-Base, llama-3-Chinese-8b, and Chinese-Mistral-7B-Instruct-v0-1. To select the model with better performance, 400 randomly chosen labeled sentiment comments, including 200 positive and 200 negative comments. The comparative analysis in Fig.4 shows that the model performs better in judging the sentiment polarity of the comments, so the subsequent study will be carried out based on using Baichuan2-7B-Base. A direct questioning approach is used when using LLM to judge the sentiment polarity of comments. First, explicit questions were asked to the model to guide it in judging the sentiment polarity of the comments, and the answers were restricted to "positive" or "negative". This approach effectively reduces the risk of the model deviating from the topic during the response process and ensures that the model focuses on judging the sentiment polarity of the comments. At the same time, by specifying response options, computational resource consumption is significantly reduced. The direct questioning approach can more efficiently and accurately utilize the Baichuan2-7B-Base model to determine the sentiment polarity of the comments, which improves the work efficiency and ensures the quality of the sentiment analysis at the same time.

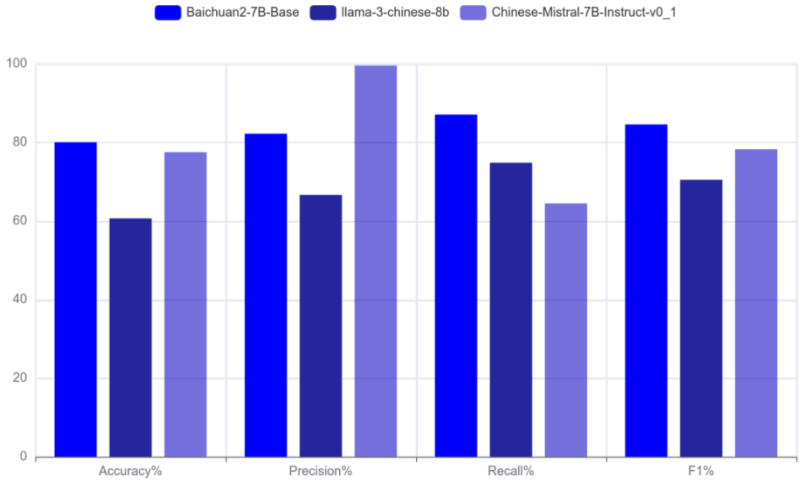

Fig. 4 LLM model performance (Photo/Picture credit: Original).

3.3 Integration of training XGBoost models

As shown in Table 2, the weighted word vectors and the sentiment polarity results of the LLM judgment are combined as the input layer of XGBoost for training, and the output results are assigned a value of 1 if they are judged as positive and 0 if they are negative.

Through this integration approach, not only does it utilize the word vectors' ability to capture sentiment features in a detailed way, but it also gives full play to LLM's advantages in understanding complex semantics and sentiment tendencies, so that the model's predictive ability is optimized.

4. Results and analysis

4.1 Comparison of model performance

Three models are constructed separately for sentiment analysis of reviews in the online mobile e-commerce domain. Without dividing the test set at the earliest, all the data are used to train the models and then the performance of the models on the data set is calculated. As shown in Fig. 5, the determination of sentiment polarity using only the weighted word vector method has already shown excellent results with an accuracy of 96.01%, which is much better than the model relying only on LLM. Subsequently, the results of integrating the weighted word vectors and LLM sentiment polarity determination into the XGBoost model show that the accuracy, recall, and area under the curve (AUC) values are further improved.

AUC, as a key indicator of the performance of a binary classification model, not only marks the effectiveness of the model in distinguishing between positive and negative sentiment comments but also highlights its comprehensive accuracy in sentiment analysis. The accuracy of the model in sentiment analysis is also emphasized. The bar chart in Fig.5 shows that the model is trained with the remaining data after dividing the test set and then used to predict the test set. Overall, the performance of the model after dividing the test set is not as good as before dividing the test set, which indicates that the weighted vector and LLM prediction can be a good feature representation of the comments, which makes XGBoost, a relatively complex model, overfitting phenomenon.

Based on the above analysis results, it can be seen that the model training effect based on the XGBoost algorithm integrating weighted word vectors and LLM confirms the feasibility of the proposed method in discriminating the sentiment polarity of online cell phone e-commerce reviews. This integration strategy significantly optimizes the generalization ability and robustness of the model and shows powerful analytical ability in practical applications.

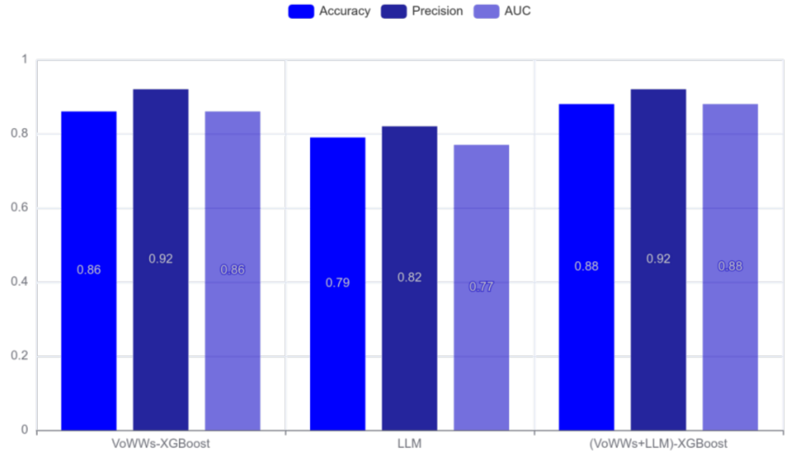

Fig. 5 Comparison of model performance (Photo/Picture credit: Original).

4.2 Affective index

Online cell phone e-commerce sentiment index refers to the calculation of the overall sentiment intensity of all consumer reviews of the product using a comprehensive evaluation algorithm from the dimension of independent stores of cell phone e-commerce based on quantifying the sentiment polarity of all user reviews of the merchant [12]. After referring to the research results of the literature, the following sentiment index calculation model is constructed [13, 14]:

$$I_i = 100 \times \left[\begin{matrix}\frac{(M_{\text{pos}}-M_{\text{neg}})}{M_{\text{all}}} \\ 2+0.5\end{matrix}\right] \quad (5)$$

where: I_i is the overall sentiment index of the ith store; M_{pos} and M_{neg} are the number of positive and negative reviews, respectively; M_{all} is all the consumer reviews of store i.

In the Jingdong e-commerce platform to select the hot cell phone classification under the brand's stores with the highest sales of commodity reviews, a total of five online cell phone Jingdong self-owned stores crawled to the June-August 2024 consumer reviews, and with the previous sentiment recognition model to automatically determine the review sentiment polarity as well as calculate the store sentiment index, the data is shown in Table 2.

Table 2. Shop Sentiment Index

brand	name	pos	neg	point
comment10 0026667910	Xiaomi 13 Leica Optical Lens Snapdragon 8 Gen 2 Processor 12+512GB Black 5G Phone	2000	1000	66.666
comment10 0078549401	Xiaomi (MI) Redmi K70 Ultra 9300+ IP68 Xiaomi Dragon Crystal Glass 12GB+256GB	1479	1241	54.375
comment10 0049486733	Apple iPhone 13 (A2634) 128GB Starlight Blue 5G Dual SIM Dual Standby Mobile Phone	1714	1286	57.133

comment10 0077950816	Honor 100 Single-lens reflex-level main camera Honor Oasis eye protection screen Snapdragon 730G AI mobile phone 16+512 Moonlight white 5G	1193	652	64.661
comment10 0128967368	OnePlus Ace 3 12GB+256GB Starry Black 1.5K AMOLED 2nd Generation Snapdragon 8 flagship chip 5G ultra-long-lasting gaming phone	1065	235	81.923

From Table 2, it can be observed that the number of bad reviews of the Yiga cell phone is less, while the number of bad reviews of Apple 13 is more, and the sentiment index of the Yiga cell phone is relatively high, able to reach more than 80, while the sentiment index of Xiaomi K70 and Apple 13 is relatively low.

5. Defect and improvement

The limitations of this study are mainly reflected in the unidirectionality of the dataset and the construction of the sentiment lexicon. The dataset is mainly derived from mobile phone reviews, which may affect the generalization ability of the model in other fields. In addition, the construction of the sentiment lexicon requires manual intervention, which affects the recognition accuracy of the model for new words and complex expressions. The distinction between the positive and negative nature of the comments is not accurate in today's distinction mode, and a more reasonable classification method should be adopted.

To further improve the generalization ability of the model, future research can be improved in the following directions:

1. To address the limitations of the sentiment analysis model in short text processing, Transformer-based models such as BERT can be introduced. According to Zhou, BERT performs well in processing short texts and implicit sentiment expressions, especially when dealing with long-distance dependencies [15]. By incorporating BERT's pre-trained embeddings in sentiment lexicon construction, the model's ability to capture complex sentiment can be improved.

2. To address the shortcomings of Model 3 in dealing with complex emotions, adversarial training can be considered to improve its robustness in dealing with highly variable data. Studies have shown that adversarial training can reduce overfitting in the model and provide better sentiment classification results when faced with highly variable comments [16].

3. To reduce manual intervention, automated tools can be used to screen and optimize emotional words. For example, emotional polarity can be automatically labeled by introducing multi-task learning, and combined with existing sentiment dictionaries (such as LIWC, NRC Emotion lexicon, etc.) for optimization, thereby further reducing subjective intervention and improving the accuracy of sentiment classification.

6. Conclusion

In this study, an effective dictionary for the field of online mobile e-commerce was constructed based on the SO-PMI algorithm. An effective analysis model for reviews was built by using the integrated weighted word vector and the affective polarity output of LLM as the input layer of XGBoost. The empirical analysis with the online reviews of goods under the hot-selling cell phone category of Jingdong as the research object shows that the constructed online cell phone e-commerce field emotion dictionary has better performance in judging the text emotion polarity, and the established emotion index model can dynamically monitor the changes in customers' emotions and help brand merchants grasp the emotion trend of the whole online cell phone e-commerce industry on time.

The model performed well in terms of AUC and accuracy, and it showed strong adaptability, especially in the affective analysis of long texts. However, it is slightly lacking when dealing with short texts, which may be due to its weak ability to capture local features.

In contrast, the LLM model is weak when dealing with complex emotional expressions, especially when capturing implicit emotions, which shows that traditional machine learning algorithms are limited when dealing with complex emotional expressions.

The experimental results show that the sentiment recognition model proposed in this study exhibits great potential in the field of sentiment analysis, which is not only innovative in theory but also demonstrates excellent performance in practical applications, providing a new direction for future research and practice.

References

1. Z. Zhu, M. H. Shan, X. Y. Fang. "Statement" or "Rhetoric": The influence of online review language style on consumers' purchase intention. Finance and Economics, 11, 92-102 (2023).
2. S. Liu, J. Zhao, H. Yang, et al. A review of text sentiment analysis. Software Guide, 17(6), 1-4 (2018).
3. X. Wang. Research on sentiment analysis of tourism web evaluation based on sentiment dictionary and machine learning. Computer and Digital Engineering, 44(4), 578-582 (2016).
4. H. Luo, Q. Yang. Emotional analysis of bi-directional long and short-term memory networks based on emotional lexicon and stacked residuals. Computer Applications, 42(4), 1099-1107 (2022).
5. B. Su, Y. Zhou, X. Liang. Text sentiment recognition model for e-commerce reviews based on XGBoost algorithm. Internet of Things Technology, 8(1), 54-57 (2018).
6. Millet (MI) Redmi K70 ultimate edition 9300 + phegda IP68 millet dragonshard 12 gb + 256 gb fine white glass Millet red rice K70 Ultra 5 g (N). https://item.jd.com/100128967368.html.
7. X. Ye, J. Cao, F. Xu, et al. Adaptive learning method for Chinese domain sentiment dictionary. Computer Engineering and Design, 41(8), 2231-2237 (2020).
8. C. Huang, S. Wen, X. Yang, et al. Research on the interaction between individual investor sentiment and stock price behavior. China Management Science, 28(3), 191-200 (2020).
9. J. Devlin, M. W. Chang, K. Lee, et al. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint, arXiv:1810.04805 (2018).
10. Z. Zhang, D. Robinson, J. Tepper. Detecting hate speech on Twitter using a convolution-GRU-based deep neural network. The Semantic Web: 15th International Conference, ESWC 2018, 745-760 (2018).
11. C. Sun, X. Qiu, Y. Xu, et al. How to fine-tune BERT for text classification? Chinese Computational Linguistics: 18th China National Conference, CCL 2019, 194-206 (2019).
12. J. Jin, L. Wu, T. Zhang, et al. Research on takeaway merchant scoring based on sentiment analysis. Journal of Management, 33(2), 66-75 (2020).
13. Y. Liu, L. Kan. E-commerce online review data mining based on text sentiment analysis. Statistics and Information Forum, 33(12), 119-124 (2018).
14. M. Liu, W. Li, S. Zhu, et al. Construction of consumer sentiment index based on internet text sentiment analysis. Statistics and Information Forum, 33(8), 31-38 (2018).
15. R. Zhou, G. Qian. A study on the classification of satirical texts based on BERT-BiGRU-CNN. Intelligent Computer and Applications, 1-6 (2024).

16. Y. Zhou, H. J. Ma, Y. N. Xu, et al. A model for extracting bidirectional sentiment triplets at the segment level based on adversarial training. Software Engineering, 27(09), 73-78 (2024).