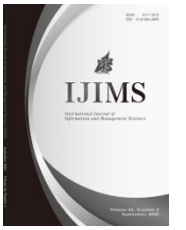
http://ijims.ms.tku.edu.tw/main.php

International Journal of Information and Management Sciences

International Journal of
Information and Management Sciences

37 (2026), 117-138. DOI:10.6186/IJIMS.202606_37(2).0002



Intelligent Classification of E-commerce Platform Reviews Based on Fine-grained Sentiment Analysis Modeling

Hongliang Cheng [1,*], *Qingheng Sun* [2,3,*]
[1] Linzhou College of Architectural Technology, Linzhou 456500, China
[2] Zhengzhou Urban Construction Vocational College, Zhengzhou 451263, China
[3] North China University of Water Resources and Electric Power, Zhengzhou 450045, China

Keywords

Fine-Grained Sentiment Analysis; E-Commerce; Bi-LSTM; Intelligent

Abstract

The rapid progress of the Internet has led to e-commerce platforms becoming one of the main ways for people to shop. How to quickly classify comments based on customer feedback emotions and better provide decision-making references for platform merchants is a new research hotspot. Therefore, the study designs a review sentiment analysis model that integrates a bidirectional long and short-term memory (Bi-LSTM) network and a conditional random field model, and uses it to intelligently classify reviews. The study firstly utilizes the Bi-LSTM network to extract features from the review text; then the output sequence of the Bi-LSTM network is used as the input, and the conditional random field model is used to infer the sentiment labels to achieve intelligent classification of reviews; finally, the dataset is utilized to validate the performance of the model. The experimental data validate that the classification accuracy, classification error rate, PR curve area and F1 value of the comment classification model constructed by the study on the comment data are 92.15%, 5.31%, 0.89 and 0.91, respectively. This study applies a model based on Bi-LSTM network and conditional random field for sentiment analysis of comments, and introduces self attention mechanism and Enhanced Representation through kNowledge IntEgration model to enhance the characteristics of e-commerce comments, in order to accurately capture text sequence features and optimize sentiment classification, providing a new path for intelligent classification of e-commerce platforms.

1. Introduction

Nowadays, e-commerce platforms have become one of the primary ways for shopping due

*corresponding author. Email: lzjzchl@163.com (Hongliang Cheng); 15290866637@139.com (Qingheng Sun)

to the upgrading of the internet. On such platforms, users are free to post comments on goods and services, and these comments have an important influence on other users' shopping decisions. Therefore, sentiment analysis of these comments can help users understand the quality of goods and services more accurately and improve the shopping experience (Tang et al., 2019; Lv et al., 2021; Zeng et al., 2021). Sentiment analysis is an important task in the natural language processing, and its goal is to classify the text in terms of sentiment polarity, i.e., to determine whether the emotion expressed in the text is active, inactive or neutral. Traditional emotion analysis methods are mainly based on feature engineering and machine learning algorithms, but these methods have some limitations when dealing with complex texts (Huang et al., 2020; Wang & Hwan Yun, 2020; Liu et al., 2021). On the one hand, solutions based on feature engineering and machine learning algorithms, such as Support Vector Machine (SVM) and Naive Bayes, overly rely on manual feature extraction, making it difficult to capture deep semantic information when processing complex text, resulting in inaccurate sentiment classification (Annadurai et al., 2023). On the other hand, early deep learning models such as Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) were able to automatically extract text features, but the one-way flow of information prevented them from fully capturing the contextual semantics of the text (Xing et al., 2022). In addition, some pre trained models such as Transformer and BERT perform well in semantic understanding, but their model structures are complex, the training and inference processes have high computational costs, and the pre trained models have many parameters, which are prone to overfitting and difficult to efficiently apply in large-scale comment data scenarios (Mewada et al., 2023). In this context, the study proposes a fusion of Bidirectional Long Short-Term Memory Network (Bi-LSTM) and Conditional Random Field (CRF) model. In sentiment analysis, there is often a certain degree of correlation and dependence between the sentiment labels of comment texts. For example, a sentence expressing a turning point (such as "The phone is great, but delivery is too slow") will contain emotional labels that are consecutive and of opposite polarity inside. The dependency relationship between these labels is mainly reflected in the internal logical structure and emotional tendency of the sentence, which is crucial for ensuring the rationality and consistency of the final sentiment analysis results. Compared with traditional LSTM, Bi-LSTM can capture the contextual features of text more comprehensively by simultaneously processing the forward and backward semantic information of the text, thereby improving the accuracy of sentiment classification (Zhou et al., 2023). In addition, CRF can model and learn the constraint relationships within label sequences, thereby inferring the label sequence that best conforms to language logic at the global level, ensuring the rationality and consistency of sentiment analysis results (Ridderhof et al., 2022). This study aims to combine the deep semantic feature extraction ability of Bi-LSTM with the structured sequence decision-making advantage of CRF, in order to jointly solve the problems of semantic ambiguity, emotional polarity uncertainty, and contextual information neglect encountered by traditional methods in processing e-commerce comments. This cooperation mode effectively combines the expressive learning ability of deep learning with the structural prediction advantage of probability graph model, enabling it to simultaneously achieve high-precision feature capture and highly consistent label output in the emotion analysis task, thus producing better comprehensive performance than a single model.

The innovation of this study lies in: (1) systematically applying and optimizing the combination model of Bi-LSTM and CRF for fine-grained sentiment analysis tasks in e-

commerce reviews. (2) Introducing a self attention mechanism layer in the Bi-LSTM network architecture, dynamically adjusting the weight values of data information for text length, to increase the model's focus on key emotional information. (3) The constructed joint extraction model can simultaneously extract entity and relationship information from comment texts, and perform sequence annotation to achieve fine-grained sentiment analysis of comment content.

The first part of the study introduces the current research status of fine-grained sentiment analysis (FGSA) model and online review classification; the second part of the study is based on the intelligent classification of e-commerce platform review sentiment based on FGSA; the third part of the study verifies the performance of the constructed model for review classification through simulation experiments and practical applications; the fourth part of the study summarizes the experimental results and analyzes the strengths and weaknesses of the research methods used.

2. Related Work

One of the core tasks of FGSA is Aspect Based Sentiment Analysis (ABSA), which aims not only to determine the overall emotional orientation of the text, but also to identify specific aspects or attributes discussed in the text and determine the emotional polarity for each aspect. Early research on ABSA often used pipeline methods, which first identified and evaluated aspects, and then judged emotions related to that aspect. This approach is susceptible to error propagation and difficult to handle complex interactions between aspects and emotional words (Zhang et al., 2022). To overcome the limitations of traditional methods, deep learning techniques have been introduced into FGSA tasks, including ABSA, demonstrating powerful capabilities in automatic feature extraction and context modeling (Chen et al., 2024). For the sentiment analysis model, many experts and scholars have carried out in-depth research. D'Aniello et al. (2022) addressed the issue of confusion regarding the basic concepts of ABSA, provided an overview of the latest technologies and methods in ABSA, proposed a new sentiment analysis and ABSA reference model, and pointed out that different tools and indicators should be used to measure every aspect of opinions. Wu et al. (2023) stated that the main challenge of ABSA is how to effectively extract the emotional polarity of specific emotional items. Early research focused on recurrent neural networks, but these models often fail to accurately capture emotional pairs. Therefore, a knowledge aware dependency graph network based on dependency graph is proposed by combining domain knowledge, dependency labels, and syntactic paths. The results indicate that the proposed network is significantly superior to previous state-of-the-art methods in ABSA tasks. Akande et al. (2023) in order to explore the application of Bi-LSTM deep learning technique in the sentiment analysis of new Crown Pneumonia tweets, conducted a retrieval study with COVID or COVID-19 tags as the keywords in the Twitter data. The results show that Bi-LSTM deep learning technology has strong advantages in sentiment analysis, and 49.93% of people have positive attitudes towards New Crown Pneumonia while 50.06% have negative attitudes through predictive analysis, which is positive for alleviating citizens' tensions and providing more professional intervention treatments. Zhang et al. (2021) to solve the multi-word over-slicing problem that exists in sentiment analysis, designed a method based on FGSA and Kano models to extract consumers' demand for product attributes from online reviews. The results show a significant positive correlation between product sales and extracted appealing one-dimensional product attributes. Zhao et al. (2020) in order to extract

specific target sentiments from text, an E2E framework was constructed using deep learning, and using this framework, the study was trained to annotate textual sentiments in relevant legal domains. The framework performance is significantly greater than the existing methods for extracting textual sentiment in the legal domain. Tang et al. (2019) to address the effective classification of polarity in FGSA in online reviews, a maximum entropy-based sentiment analysis model was constructed, using which viewpoints, sentiment polarity, and granularity were jointly extracted. The results show that the model has better applicability in the comprehensive evaluation of real-world reviews of electronic devices and restaurants. Huang et al. (2022) constructed a contextual positional weight-based model to handle the matter of correlation between contextual sentences in FGSA by the traditional model. The model can adjust the weights of context words in given the position of body words in the sentence and alleviate the interference of the difference in numerous words on both sides of the body words on the judgment of sentiment polarity. This shows that the model accuracy is 4.27% higher than ASGCN, while the F1 value is improved by 4.31% compared to ASGCN.

Huang et al. (2020) designed a text dictionary based on consumers' language usage behavior in order to study the potential factors affecting consumers' purchase of goods in the context of e-commerce. The dictionary is capable of recommending effective information based on consumers' different shopping comments. The study integrates the dictionary into a typical e-commerce recommender system to predict consumers' purchasing behaviors. The system integrated by the dictionary can significantly modulate the impact of online lifestyle on consumer preferences. Hou et al. (2022) in order to explore whether the reviews of reviewers with rich reviewing experience have an impact on the consumption of others, experience expertise and experience diversity were defined by evaluating the number of user's historical reviews in a given category, and the homogeneity of this distribution across categories, respectively. The results suggest that experienced reviewers articulate some influence on other consumers and that their reviews can provide insights into the management of online review platforms and e-marketing.

In summary, existing research has demonstrated the effectiveness of deep learning models in sentiment analysis tasks, particularly in capturing contextual information and processing complex patterns, outperforming traditional methods. However, existing methods still face challenges in FGSA for comments on e-commerce platforms. For example, a single LSTM or Bi-LSTM has limitations in accurately modeling the dependency relationships between label sequences. Moreover, existing methods perform poorly in adapting to different datasets and tasks, requiring significant adjustments and optimizations. Therefore, this study proposes the Bi-LSTM+CRF model, which can integrate the automatic feature learning ability of deep learning with the structured prediction advantage of probability graph models, filling the gap in existing research in processing e-commerce comments, and thus achieving more accurate and efficient fine-grained intelligent classification of e-commerce comments.

3. Design of an Intelligent Classification Model for E-Commerce Platform Review Sentiment Based on FGSA

FGSA refers to the subdivision of sentiment analysis into more specific sentiment categories than just positive or negative. In e-commerce platform reviews, users' comments can relate to multiple aspects, including product quality, customer service, logistics speed, etc. (Bouras et al., 2022; Zhang et al., 2023). Therefore, intelligent categorization of these comments can help e-

commerce platforms better understand users' needs and feedback.

3.1 Review data collection and review sentiment analysis model design

When classifying e-commerce platform comments, it is first necessary to collect relevant comment data, which is the prerequisite for intelligent classification; and then after obtaining the corresponding data, it is necessary to perform sentiment analysis, and data preprocessing within the comments, this is because user-generated comment data has a high level of noise, which leads to the inability to meet the grammatical and content specifications, which negatively affects the accuracy of the analysis model. Therefore, it is crucial to perform cleaning operations after collecting the data, and after completing these steps, it is also necessary to construct a sentiment comment analysis model, which is designed using a deep learning-based approach. LSTM is obtained by optimizing and improving the original recurrent neural network, which includes adding the concept of temporal dimensions and introducing the gating unit mechanism, and these improvements allow the LSTM model to gain the selective forgetting ability. It can be seen that for the problem of sentiment analysis, it is an excellent solution to utilize the LSTM model, and the structure of LSTM is Figure 1 (He et al., 2023).

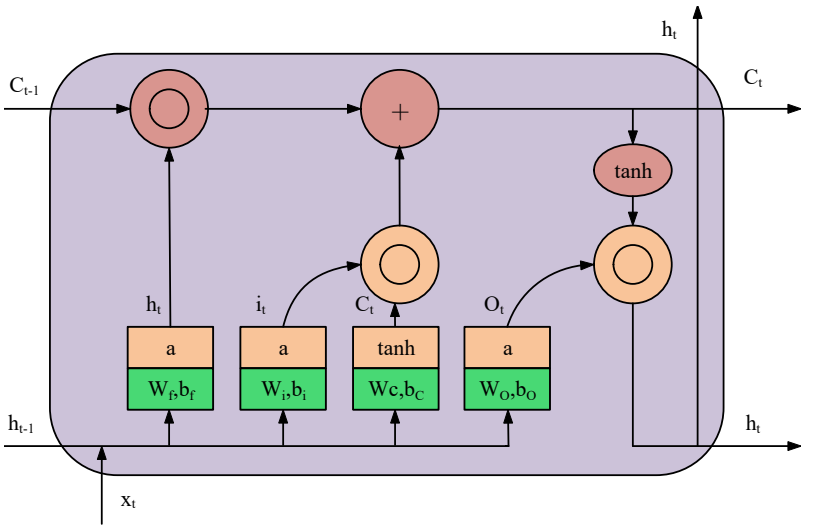


Figure 1 *Sketch of the structure of LSTM*

The key reason why LSTM can be used for sentiment analysis is that it has a fiducial state, which has a conveyor belt-like function that runs through the entire timing operation. Since the overall gradient flow condition of the LSTM model is equal to the sum of the gradient flow of each long distance, so as long as one of the paths of the model is guaranteed not to have the problem of gradient vanishing, it is not possible to make the model not to have the phenomenon of gradient vanishing. As for the problem of gradient explosion, since the computation of the other long-distance gradients of the LSTM model involves multiple activation functions, the possibility of the model experiencing gradient explosion is reduced. Synthesizing the research on LSTM, it is found that when using LSTM network to get the semantic representation of the text, the model does not understand the information expressed in the reverse of the text sequence due to the limitation of its structure. For example, when a product review in an e-

commerce platform reads "This parcel was dirty when I received it", the LSTM model does not understand that "not" is a modification of "dirty", and the model will have semantic problems. When the LSTM model does not understand that "no" is a modification of "dirty", the semantic representation of the model will be incomplete to deal with it, the study adopts the Bi-LSTM structure. This network is also improved on the basis of LSTM, the difference is that it utilizes bidirectional LSTM, and its structure is shown in Figure 2 (Liu et al., 2023).

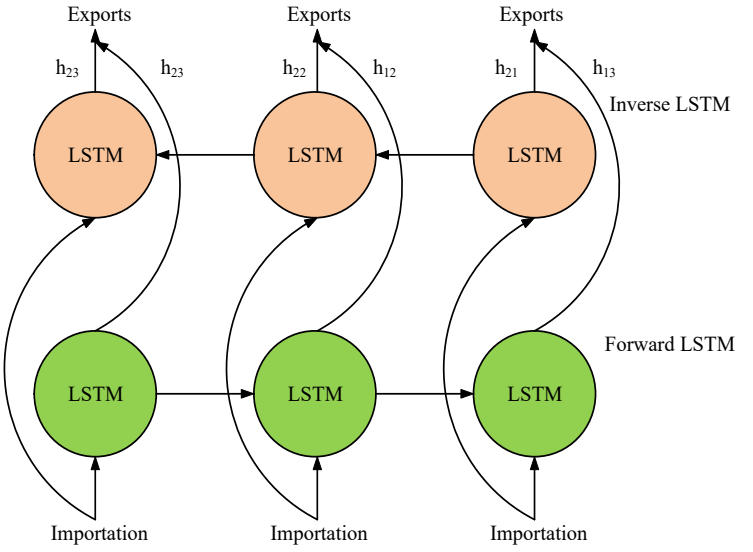


Figure 2 *A sketch of the network structure of Bi-LSTM*

Bi-LSTM networks have an advantage over ordinary LSTMs in obtaining complete semantics due to their ability to simultaneously encode the forward and backward semantics of textual data. When dealing with the task of sentiment analysis, the common practice of Bi-LSTM model is to splice the output vectors of LSTMs in different directions at the last moment, instead of simply taking out the results of individual neurons spliced at the same moment, because the results that can be obtained by this splicing method contain both the forward semantic representation information of the text and the reverse semantic information (Fu et al., 2021; Mannepalli et al., 2023). Combined with Fig. 2, the Bi-LSTM network structure is the splicing of the three output vectors in the figure, at this time, if we set the data input sequence of the Bi-LSTM, the forward LSTM output form can be expressed by Eq. (1).

$$\overrightarrow{h_n} = \overrightarrow{LSTM}(x_n, \overrightarrow{h_{n-1}}, \overrightarrow{C_{n-1}}) \tag{1}$$

In Eq. (1), x_n denotes the word vector corresponding to the input at the moment of n during the input sequence; h_{n-1} denotes the representation vector corresponding to the output at the moment of $n-1$; and C_{n-1} denotes the cell state corresponding to the moment of $n-1$. The reverse LSTM output form can be expressed by Eq. (2).

$$\overleftarrow{h_n} = \overleftarrow{LSTM}(x_n, \overleftarrow{h_{n-1}}, \overleftarrow{C_{n-1}}) \tag{2}$$

Combined with the derivation of Eqs. (1) and (2), the splicing of the output forms in the forward and reverse directions can be expressed by Eq. (3).

$$h_n = concat(\overrightarrow{h_n}, \overleftarrow{h_n}) \tag{3}$$

In Eq. (3), $concat(\cdot)$ denotes the function used to splice the positive and negative output forms. To rise the relevance of the comment text, the study introduces a self-attention mechanism (SAM) layer into the Bi-LSTM network structure. The main job of the SAM layer is to perform optimization operations on the sentence vector output of the previous layer with the goal of increasing the relevance of the internal feature text. This layer is similar to the multi-head attention mechanism in the vectorized representation layer, although the multi-head concept is applied in a different way, both belong to the category of attention mechanism in essence. Since the attention mechanism layer, the data information weight value of the text length at the corresponding moment can be calculated by Eq. (4).

$$\theta_n = \frac{\exp(q_i k_n / \sqrt{d_k})}{\sum_{n=1}^{d} (\exp(q_i k_n / \sqrt{d_k}))} \tag{4}$$

In Eq. (4), q_i denotes the result of sentence vector being transformed by weight matrix product; k_n denotes the result of weight matrix product transformation; d_k denotes the value corresponding to processing the k th text. The final representation of the final word vectors obtained after the self-attentive wit layer can be expressed in Eq. (5).

$$result = \sum_{n=1} v_n \theta_n \tag{5}$$

In Eq. (5), v_n denotes the value of matrix weights after the processing of the self-attention with layer. Through the construction of the above model structure, the final fully connected layer (FCL) is still missing, and the function of the FCL is for the judgment of emotional tendency in the comments. It is usually used with the normalized exponential function, and its main role is to perform the downgrading and classification operations on the computed sentence vectors. The specific process is as follows: when the sentence vector passes through the FCL, the layer will map it, and the complete sentence vector is mapped into a sequence of real numbers equal to the number of categories. Subsequently, after the processing of the normalized exponential function, the problem of emotional tendency in the comment sentence is finally transformed into the problem of the size of the proportion of each emotional category, the formula of which is shown in Eq. (6).

$$\hat{y} = soft\max(W_x + b) \tag{6}$$

In Eq. (6), $\hat{y}$ denotes the kind of comment sentiment obtained from the model predicted comments; W_x denotes the weight matrix related to the x th comment in the FCL; b denotes the bias term of the activation function in the FCL. In summary, after the SAM sentiment analysis model based on normalized exponential activation function and Bi-LSTM has met the requirements of sentiment analysis in comments.

3.2 Construction of fine-grained sentiment review classification and analysis model by fusing Bi-LSTM and CRF

The need for joint extraction for categorizing comments and sentiment views can be converted into a named entity recognition task, where the named entity recognition is defined

as extracting the actual representation of a concept in textual data. The granularity of this analysis is finer than the overall sentiment analysis, which focuses not only on the overall sentiment tendency of consumers towards a product or service, but also on their preferences for individual attributes of the product or service. The study, in order to better utilize sentiment classification for intelligent classification of e-commerce platform reviews, then introduces the CRF model, which is a probabilistic graphical model commonly used in sequence annotation tasks such as named entity recognition, lexical annotation, and sentiment analysis. It can improve the accuracy of classification by considering contextual information for sequence labeling of labels (Ortega, 2022). In the process of using a CRF model for named entity recognition tasks, a set of text sequence data is usually input into the model first. Then, the model performs calculations based on the categories in the labeled sequence to derive the category to which the maximum conditional probability of the current comment word belongs. The output of the computational linear chain CRF model can be expressed using Eq. (7).

$$P(L/c)=\frac{1}{Z(c)}\exp\left(\sum_{a}\sum_{j=1}^{n}\lambda_a t_a+\sum_{b}\sum_{j=1}^{n}\mu_b s_b\right) \tag{7}$$

In Eq. (7), L denotes the labeling sequence; c denotes the input text sequence; $Z(c)$ is the specification function that ensures that the result conforms to the probability distribution; a denotes the transferred features; λ_a denotes the weight value of the number of features; t_a denotes the transferred feature function of the same feature function; μ_b denotes the weight of another feature function; s_b denotes the state feature function; and b denotes the number of state features. The CRF in the task of naming entity comment classification recognition is Figure 3.

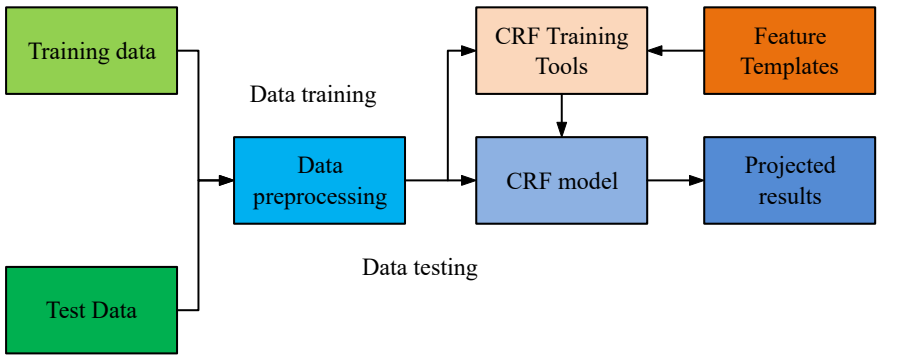


Figure 3 *Flowchart of the CRF model for the named entity comment classification and recognition task*

During the model training phase, relevant review data often needs to be collected in advance and the weights of different features need to be determined. It is common to find these weights using maximized likelihood estimation, and during the testing of the CRF model, the study used the Viterbi algorithm to determine the final sequence categories, which can be represented by Eq. (8).

$$L=\arg\max_{L}(P(L/c)) \tag{8}$$

In Eq. (8), $\arg\max(\cdot)$ denotes the parameterization function, which aims to obtain the

weights of the corresponding sequences before the model output. By studying the CRF model, it is found that it needs to be provided manually in advance to the training tool feature template parameters, which are based on the content of the training dataset, selecting representative words and rules. Therefore, the accuracy of recognition using the CRF model depends entirely on whether the content of the provided feature templates is comprehensive and complete. To address the accuracy problem, the study fuses Bi-LSTM and CRF to form a joint extraction model (Nimrah & Saifullah, 2022; Wang, 2020). The sentiment analysis of e-commerce comments requires handling long-term dependency relationships to understand emotional tendencies. Bi-LSTM can effectively capture the long-term dependency information of text and extract more complete semantic features. CRF can consider the dependency relationships between tags and determine the sentiment tag sequence through a global optimal strategy. Therefore, the Bi-LSTM+CRF combination can balance the requirements of feature extraction and sequence annotation. The commonly used sentiment analysis methods based on BERT currently have strong understanding ability for text, and can learn rich language knowledge and semantic information. However, the training and reasoning process incurs high computational costs and requires high hardware resources. The Bi-LSTM+CRF method, while ensuring good performance, has a relatively simple model structure and higher training and inference efficiency. In addition, traditional machine learning methods such as SVM, Naive Bayes, etc., although having relatively simple model structures and low computational complexity, have limitations in processing complex texts such as semantic ambiguity, uncertain emotional polarity, and ignoring contextual information. The Bi-LSTM+CRF fusion model can automatically learn text features and utilize the representation learning ability of deep learning to more accurately capture the semantic information and emotional tendencies of the text. The specific structure of the joint extraction model built on Bi-LSTM and CRF is Figure 4.

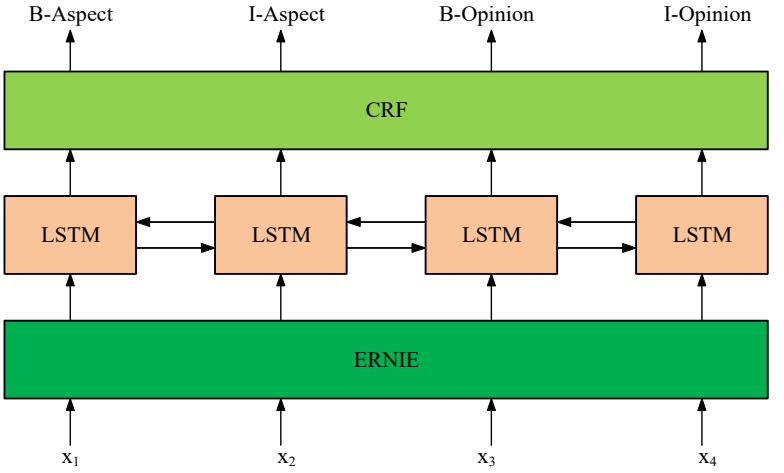


Figure 4 *Architecture of a joint extraction model incorporating Bi-LSTM and CRF*

In the extraction model with the combination of Bi-LSTM+CRF, the CRF part is somewhat different from the full CRF model. The model constitutes the score matrix of the text sequence by receiving the hidden layer states output from the neural network, so that the Viterbi algorithm can be utilized to plan the optimal annotation scheme for the classification of the comments. The calculation of the scores in the loss function of the CRF layer is displayed by Eq. (9).

$$s(C,l)=\sum_{j=0} T_{yj,yj+1}+\sum_{j=1} P_{j,yj} \tag{9}$$

In Eq. (9), T denotes the transfer score matrix of the dimension; $yj, yj+1$ denotes the sequence corresponding from the start label to the end label; P denotes the output matrix of the hidden layer state; $P_{j,yj}$ is the score value of the j-th comment word in the text sequence belonging to the yj category. The final state of the transfer score matrix is obtained from the joint model that has been trained and updated in several rounds, and it is a column of constraint rules generated by the structural layer of the CRF. The joint extraction model of the Bi-LSTM and the CRF represents the final labeled category of a comment in a probabilistic form. Therefore, a normalized probability calculation is performed on the basis of the scores computed in the CRF layer, which can be expressed in Eq. (10).

$$P(l/C)=\frac{e^{s(C,l)}}{\sum Le^{s(C,l)}} \tag{10}$$

In Eq. (10), e denotes the normalized processing features; e^s denotes the features corresponding to the s th feature. Since most of the comment word objects and viewpoints obtained by the joint extraction model are located in the same utterance, considering the problem of shorter and smaller texts, if an overly complex and cumbersome model is used to process them, it will not only increase the processing time, but also lead to a decrease in the effectiveness of the results. In order to improve the final performance of comment classification, this study further introduced the Enhanced Representation through kNowledge IntEgration (ERNIE) model. This model is a pre trained language model based on Transformer, which enhances semantic understanding ability by introducing knowledge masking strategy, and is particularly good at handling entity and relationship recognition in Chinese context. In this study, the ERNIE model was used for deep semantic encoding of comment texts, extracting entity and relationship features, and using these features as inputs to the CRF model to optimize the prediction of fine-grained sentiment label sequences. Specifically, ERNIE encodes input sequences through a multi-layer self attention mechanism to generate word vector representations rich in contextual information, thereby enhancing the modeling ability of semantic dependencies in sequence annotation tasks. Based on this, the study fuses the CRF model with the ERNIE model for processing the single-sentence comment text classification task as shown in Figure 5.

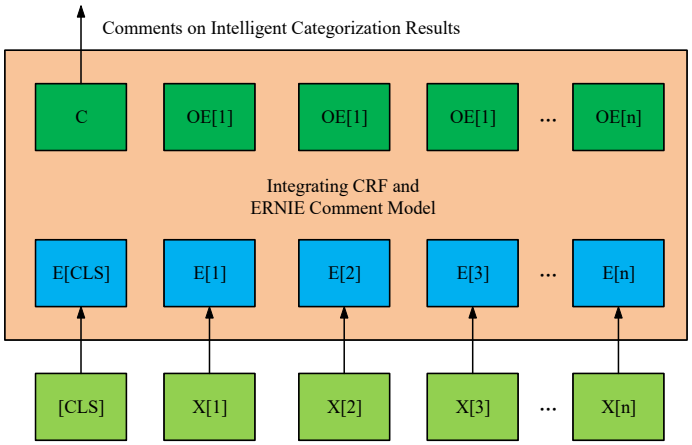


Figure 5 *Process flowchart of comment text classification task integrating CRF model and ERNIE model*

The workflow of the two functional modules of comment statement classification and viewpoint sentiment analysis: Firstly, the comment viewpoints and product attributes extracted by the entity recognition module are formatted; CLS tags are added at the beginning of each sentence, and then the processed sequence information is inputted into the hybrid model, which learns the semantic relationship between the words inside the comments by masking and filling the comment sentences, and finally the output of the model containing the whole comment is inputted into the FCL for linear classification. The model learns the semantic relationship between the words inside the comment by masking and filling operations, and finally inputs the vector containing the whole comment output from the model to the fully connected layer to do the linear classification process, and the result obtained through the above steps is the result of intelligent classification of e-commerce platform comments using FGSA of sentiment viewpoints.

4. Performance Analysis of Review Intelligent Classification Application Based on FGSA Models

To verify the performance of intelligent classification of reviews on e-commerce platforms based on FGSA model, this study collected public comments on mobile products from mainstream e-commerce platforms such as JD.com and Taobao between 2020 and 2021 using web crawling technology. The comment content was controlled to around 30 words, and a total of 38572 comments were collected, mainly for training and testing the model. After data collection, the original comment data was cleaned and preprocessed, including removing HTML tags, filtering missing values and duplicate data, to ensure data quality and consistency. At the same time, label each comment with three types of emotional tags: positive, neutral, and negative. When labeling, refer to the comment rating, emotional words, and other information to comprehensively determine the emotional tendency. In addition, the publication date and time of comments were also counted to analyze the temporal distribution characteristics of user comments.

4.1 Performance analysis of comment intelligent classification models

The dataset used in this study is sourced from publicly available e-commerce platform comment data and does not involve any privacy or confidential information. This study sets the embedding dimension to 128, the number of hidden layer units in Bi-LSTM to 64, the initial learning rate to 0.001, and uses the Adam optimizer for training. The batch size is set to 32, and the training period is set to 50 epochs. During the model training process, the Bi-LSTM+CRF model is first initialized based on the set hyperparameters. Then, the preprocessed data is input into the model in batches for training. The difference between the predicted and true values is calculated using a loss function, and the model parameters are updated using backpropagation algorithm. Subsequently, the model performance is evaluated on the validation set, and hyperparameters and training strategies are adjusted based on the validation results. Finally, after training is completed, save the optimal model for subsequent testing and application. To verify the contributions of each component in the proposed model, ablation experiments were conducted in this study. Compare individual Bi-LSTM, Bi-LSTM+SAM, Bi-LSTM+CRF, and the complete model. Using accuracy, F1 score, and loss value as evaluation indicators. To ensure the reliability of the results, each experiment was repeated 10 times, and the results of each run were recorded and the mean and standard deviation were calculated. In addition, statistical significance analysis was conducted on the performance differences between the complete model and other variants through paired t-tests. The results of the ablation experiment are shown in Table 1.

Table 1 *Results of ablation experiment*

Model	Accuracy/%	F1/%	Loss
Bi-LSTM	85.42±0.41	84.26±0.44	0.106±0.003
Bi-LSTM+SAM	87.96±0.36	86.34±0.39	0.084±0.002
Bi-LSTM+CRF	89.27±0.30	88.17±0.35	0.093±0.002
Complete model	92.15±0.23	91.26±0.30	0.078±0.002

From Table 1, it can be seen that the average accuracy and F1 score of the complete model are 92.15% and 91.26%, respectively, which are significantly higher than the Bi-LSTM, Bi-LSTM+SAM, and Bi-LSTM+CRF models ($p<0.05$). Meanwhile, the loss value of the complete model is 0.078, which is statistically significant ($p<0.05$) compared to the other three models, indicating that it has the best fitting effect on the training data. The results indicate that CRF, SAM, and Bi-LSTM components significantly contribute to the improvement of model performance. To test the performance of the intelligent classification model of comments based on sentiment analysis, the study compares the LSTM, CRF with Bi-LSTM+CRF (the model constructed by the study) using LSTM and CRF in the constructed dataset, and the comparison metrics are the detection rate, the error rate, the PR curves, and the F1 value. Conduct multiple independent running experiments. Each experiment was randomly divided into a 70% training set, a 15% validation set, and a 15% test set in the same way to ensure the reliability of the results. The results represent the average performance of multiple runs, aimed at evaluating the stability and generalization ability of the model under different data partitions. The data detection accuracy and error rate of the three models in intelligent categorization of comments

are shown in Figure 6.

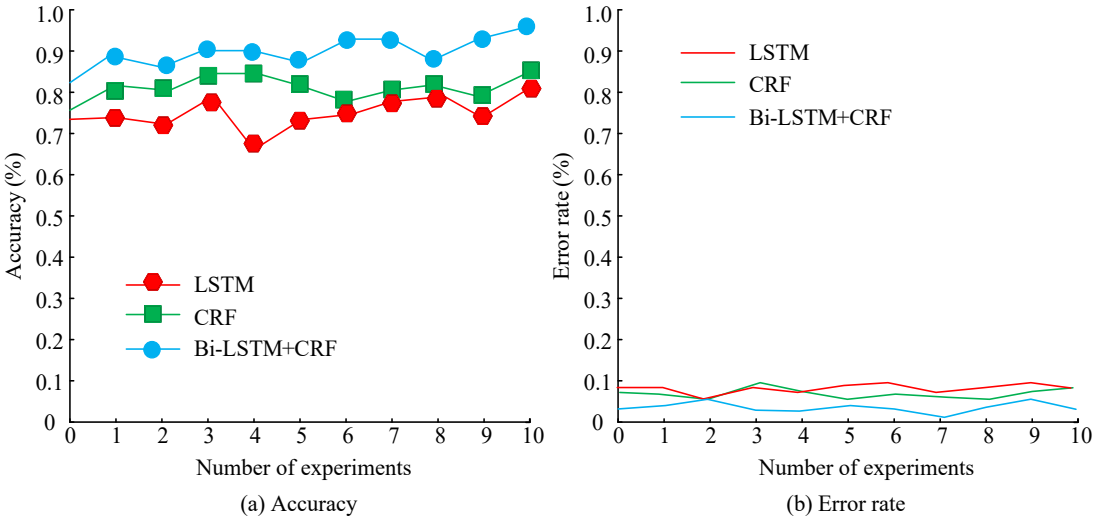


Figure 6 *Comparison of detection accuracy and error rate of three models*

From Fig. 6(a), the average detection accuracy of Bi-LSTM+CRF, CRF and LSTM on comment categorization data is 92.15%, 83.53% and 76.42%, respectively. From Fig. 6(b), the Bi-LSTM+CRF has the lowest error rate in the process of classifying the comment data, with an average value of 5.31%, while the average error rates of CRF and LSTM are 8.75% and 10.06%, respectively. And the Bi-LSTM+CRF model has a lower error rate mean value of 3.44% and 4.75% than CRF and LSTM moses. This indicates that Bi-LSTM+CRF model has the highest detection accuracy and lowest error rate in the dataset. The result of PR curves and F1 values of the three models after training in the training set is Figure 7.

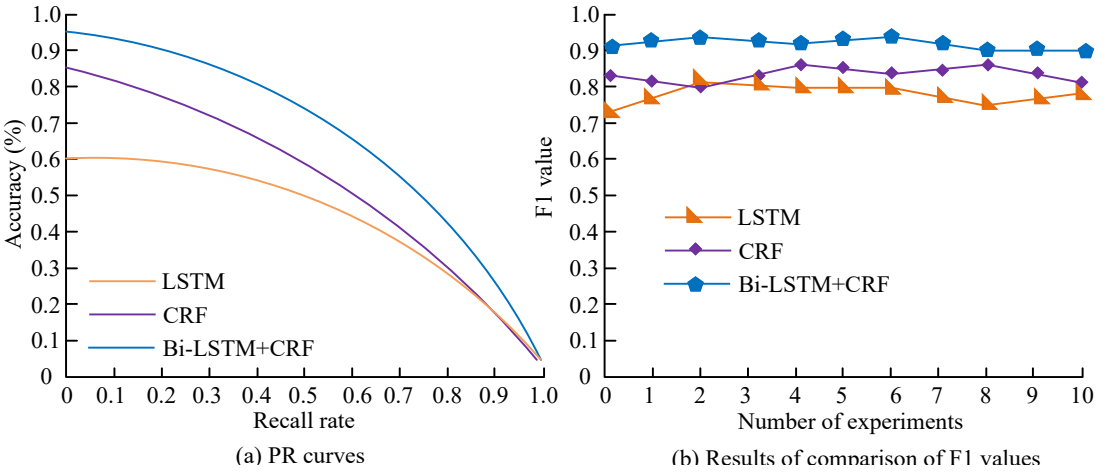


Figure 7 *Comparison results of PR curves and F1 values for the three models*

From Figure 7(a), the PR curve area of the Bi-LSTM+CRF is 0.89, and the PR curve areas of the CRF and LSTM are 0.81 and 0.69. From Fig. 7(b), the F1 mean of the Bi-LSTM+CRF is the highest, which is 0.91; and the F1 mean of the CRF and LSTM are 0.82 and 0.79. The Bi-LSTM+CRF significantly outperforms the CRF and LSTM models in terms of precision and recall dimensions. To further verify the Bi-LSTM+CRF's performance, the research pair utilizes the loss value to reflect the comparative performance of the three models, and the comparison results are shown in Figure 8.

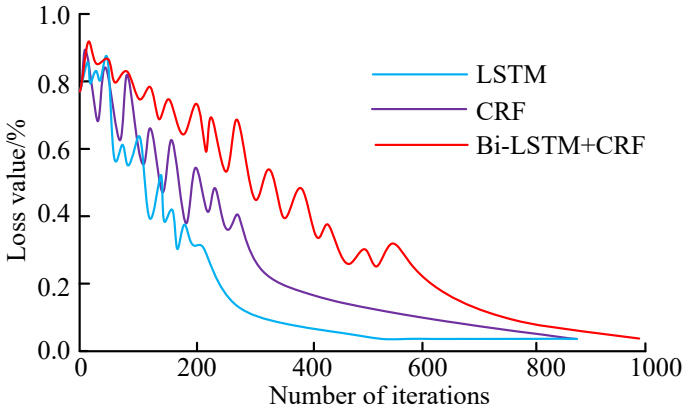


Figure 8 *Comparison results of loss values for the three models*

From Figure 8, the loss values of the three models gradually decrease with the increase of the number of iterations, in which the loss value of the Bi-LSTM+CRF stabilizes at 400 iterations, while the loss values of the CRF and LSTM models stabilize at 753 and 816 iterations, respectively. The size of the iteration loss value can reflect the degree of the model's fit to the training data. The smaller the loss value, the better the model fits the training data, and the model can better capture the complex patterns and relationships in the data. To verify the classification ability of the models in different review data, the same three models are also utilized for comparison, and the comparison results are exhibited in Figure 9.

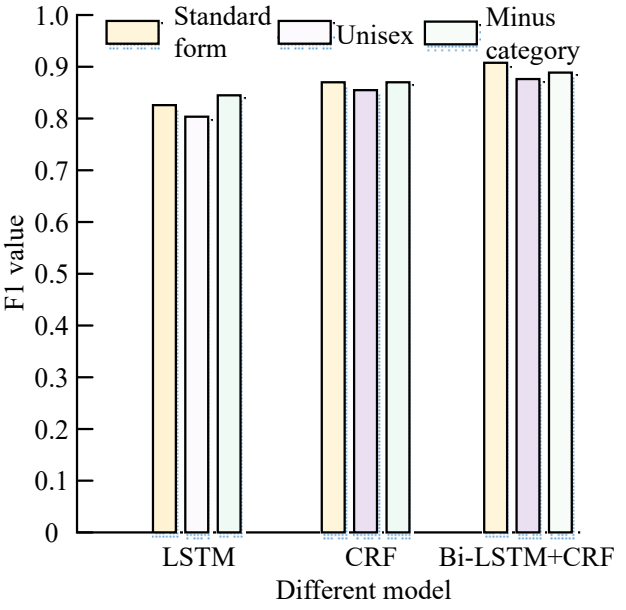


Figure 9 Comparative results of the classification ability of different models in different comment data categories

In Figure 9, the F1 value of the Bi-LSTM+CRF model is higher than the corresponding F1 value of the CRF and LSTM models in both positive and negative classification samples. This verifies that the Bi-LSTM+CRF model also has a more excellent classification ability for different comment data, although the data of the neutral category of comments is less and the definition of neutral is not as clear as that of the positive and negative classification samples, the results are still within the acceptable range. To further validate the superiority of the proposed Bi-LSTM+CRF model, this study compared it with various baseline and advanced models. Including Convolutional Neural Networks (CNN), SVM, BERT, RoBERTa ABSA, and the Knowledge Aware Dependency Graph Network (KADGN) proposed by Wu et al. (2023). Accuracy, recall, and semantic relevance score (SRS) are used as evaluation indicators. The performance comparison results of each model are shown in Table 2.

Table 2 *Performance comparison results of various models*

Model	Accuracy/%	Recall/%	SRS
CNN	84.62	83.71	0.80
SVM	77.75	76.42	0.74
BERT	88.97	87.56	0.85
RoBERTa-ABSA	90.31	89.74	0.86
KADGN	91.22	90.53	0.88
Bi-LSTM+CRF	92.15	91.27	0.90

From Table 2, it can be seen that the performance of the CNN and SVM baseline models is relatively poor, indicating their difficulty in capturing the deep semantic information and contextual dependencies of text. The performance of the KADGN model is second only to the Bi-LSTM+CRF model. The proposed Bi-LSTM+CRF model achieved the highest values in

accuracy, recall, and SRS, with values of 92.15%, 91.27%, and 0.90, respectively, indicating its good applicability and effectiveness in sentiment analysis of comments on e-commerce platforms. This is mainly due to the fusion of the automatic feature learning ability of deep learning and the structured prediction advantage of probability graph models in the Bi LSTM+CRF model, which has significant advantages in solving semantic ambiguity, emotional polarity uncertainty, and lack of contextual information problems.

4.2 Analysis of the effect of the application of the intelligent classification model of comments

To verify the effectiveness of the application of the model algorithm, the study utilizes the training set to train the algorithm and obtain the prediction model. Then the accuracy of the obtained prediction model is verified using the test set, and the verification results are shown in Figure 10.

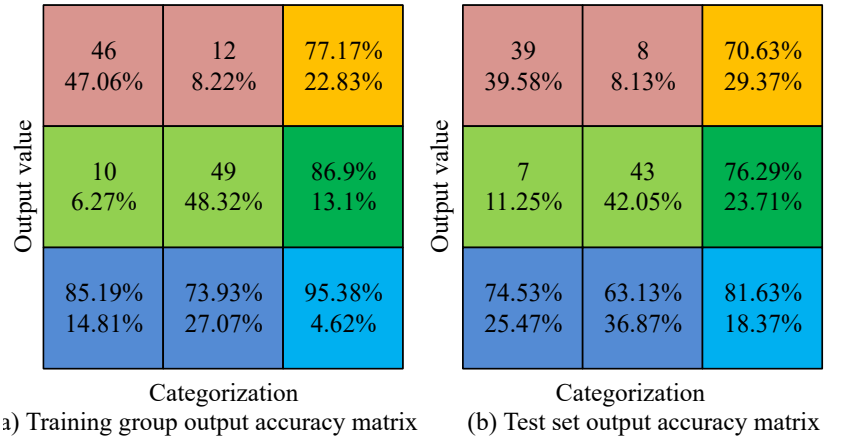


Figure 10 *Comparison of the results in the training and test sets*

From Figure 10(a), in the training set, the Bi-LSTM+CRF model has a recognition rate of 95.38% for the comment data and is able to predict 95 samples. As Figure 10(b), in the test set, the recognition rate of Bi-LSTM+CRF on review data is 81.63%, and it is able to predict 82 samples, which indicates that the Bi-LSTM+CRF has a good effect on the recognition of reviews on e-commerce platforms. To better verify the model's ability to intelligently classify the reviews, the study applies the information of cell phone reviews in an online shopping platform to the model and obtains the results of classifying the reviews of the customers after purchasing a cell phone, and the results are shown in Figure 11.

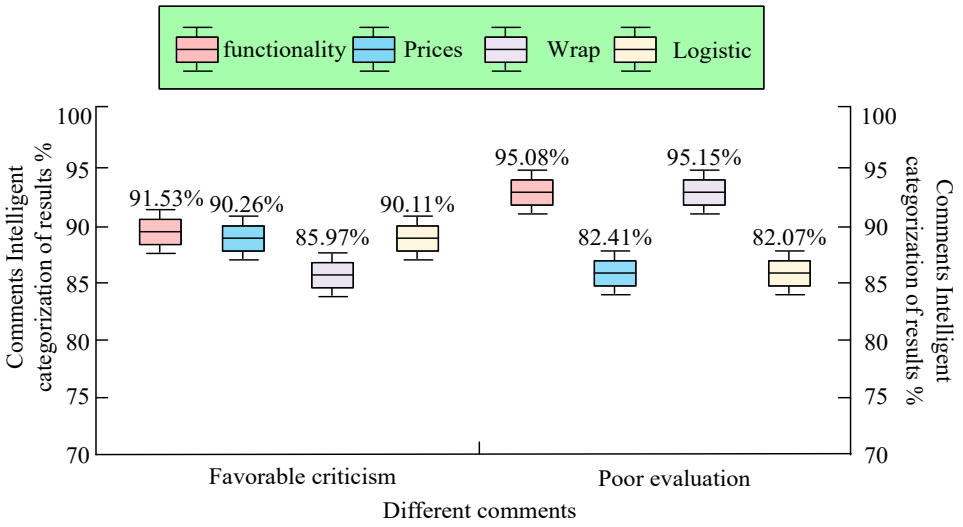


Figure 11 *Results of the model's intelligent categorization of comments*

As can be seen in Figure 11, the main information in the reviews is summarized and classified by the classification of the Bi-LSTM+CRF model, which mainly consists of the phone's features, price, packaging and logistics. The function of the cell phone is most described in the positive reviews, followed by price, logistics and packaging; while the packaging is most described in the negative reviews, followed by function, price and logistics. This verifies that the Bi-LSTM+CRF can effectively and intelligently categorize the information in reviews. To further verify the performance, the study similarly categorizes the content of reviews with positive, negative and neutral concepts, and reflects the practical application effect of the model through the final statistical results, which are shown in Figure 12.

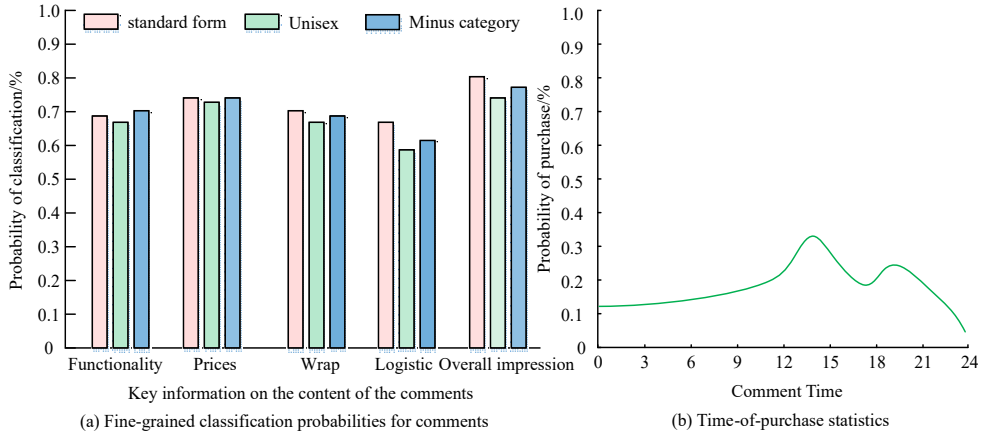


Figure 12 *Review content key messages and purchase probability analysis*

From Figure 12 (a) and (b), by analyzing the emotional information in the content of the comments, customers are most interested in the overall feeling of the cell phone goods, followed by price, function, packaging and logistics. Meanwhile, by analyzing the time of posting comments, customers have the highest probability of purchasing between 13:00 and 14:00. This

indirectly verifies that merchants can classify customer comments according to the Bi-LSTM+CRF to obtain the shopping habits of the buying group, which can be used as the basis for customizing related promotional activities.

5. Theoretical Implications

The experimental results of this study provide valuable insights into the core theories of sequence modeling and fine-grained sentiment analysis in natural language processing. Firstly, the synergistic effect of Bi LSTM and CRF validates the effectiveness of the collaborative theory of "local features and global constraints". The context aware features generated by Bi LSTM provide high-quality input for the global sequence optimization of CRF, while CRF ensures the structural rationality of the output by modeling label transition probabilities. The combination of the two provides the optimal path for solving semantic ambiguity and ensuring label consistency. Secondly, the significant effect of SAM supports the dynamic feature weighting theory, indicating that focusing on key emotional words in complex long texts can significantly improve semantic capture accuracy. In addition, the successful integration of ERNIE and CRF deepens the application theory of pre training fine-tuning paradigm in domain adaptation, confirming the feasibility of transferring general semantic knowledge to vertical domains through specific sequence models. These findings have empirically advanced our understanding of core theoretical issues such as feature learning, attention mechanisms, and knowledge transfer in sequence annotation tasks, providing valuable theoretical references for related research.

6. Conclusion

Aiming at the problem of intelligent classification of customer reviews in e-commerce platforms, the study utilizes the FGSA model of Bi-LSTM+CRF to classify the reviews based on sentiment analysis and achieves good results. The study firstly utilizes a Bi-LSTM to extract features from the review text; then the output sequence of the Bi-LSTM is used as the input, and the CRF is used to infer the sentiment labels to realize the intelligent classification of the reviews; finally, the dataset is used to validate the performance of the model. The results show that in the test set, the recognition rate of the Bi-LSTM+CRF model on the review data is 81.63%, and it can predict 82 samples, which indicates that the Bi-LSTM+CRF model has a good effect on the recognition of reviews on e-commerce platforms; meanwhile, in classifying the reviews of purchasing cell phones, it can accurately refine the four key information. This indicates that the model has strong adaptability and generalization ability and can effectively classify the review data. Although the combination of Bi-LSTM and CRF is not the first proposed in this study, unlike many early applications, this study focuses on FGSA of comments on e-commerce platforms, aiming to simultaneously identify evaluation aspects and emotional polarity, which is more in line with the actual decision-making needs of e-commerce platforms. And this study also integrated SAM into the Bi-LSTM network structure to solve the semantic ambiguity problem in e-commerce reviews, and combined the deep semantic understanding ability of ERNIE model with the sequence annotation advantage of CRF to further optimize the prediction of fine-grained sentiment label sequences. In addition, the proposed model was systematically validated on a large-scale and real e-commerce review dataset, demonstrating its effectiveness and superiority in handling FGSA tasks for e-commerce reviews. However, there are still shortcomings in the study, which only utilizes Bi-LSTM to fuse with the CRF model, and other deep learning models and algorithms can be considered in the next step to further

improve the accuracy and effectiveness of the sentiment analysis model.

Funding: No funding supports.

Conflict of Interest: The author has declared that no conflict interests exist.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

Akande, O. N., Lawrence, M. O., Ogedebe, P. (2023). Application of bidirectional LSTM deep learning technique for sentiment analysis of COVID-19 tweets: post-COVID vaccination era. *Journal of Electrical Systems and Information Technology*, 10(1): 1-17. https://doi.org/10.1186/s43067-023-00118-w

Annadurai, S., Arock, M., Vadivel, A. (2023). Real and fake emotion detection using enhanced boosted support vector machine algorithm. *Multimedia Tools and Applications*, 82(1), 1333-1353. https://doi.org/10.1007/s11042-022-13210-6

Bouras, D., Amroune, M., Bendjenna, H., Bendib, I. (2022). Improving fine-grained opinion mining approach with a deep constituency tree-long short-term memory network and word embedding. *Recent Advances in Computer Science and Communications* (*Formerly: Recent Patents on Computer Science*), 15(4): 571-582. https://doi.org/10.2174/2666255813999200922142212

Chen, X., Xie, H., Qin, S. J., Chai, Y., Tao, X., Wang, F. L. (2024). Cognitive-inspired deep learning models for aspect-based sentiment analysis: A retrospective overview and bibliometric analysis. *Cognitive Computation*, 16(6), 3518-3556. https://doi.org/10.1007/s12559-024-10331-y

D'Aniello, G., Gaeta, M., La Rocca, I. (2022). KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis. *Artificial Intelligence Review*, 55(7): 5543-5574. https://doi.org/10.1007/s10462-021-10134-9

Fu, Y., Liao, J., Li, Y., Wang, S., Li, X. (2021). Multiple perspective attention based on double BiLSTM for aspect and sentiment pair extract. *Neurocomputing*, 438(1): 302-311. https://doi.org/10.1016/j.neucom.2021.01.079

He, Z., Dumdumaya, C. E., Machica, I. K. D. (2023). Text sentiment analysis based on Multi-Layer Bi-Directional LSTM with a trapezoidal structure. *Intelligent Automation & Soft Computing*, 37(1): 639-654. https://doi.org/10.32604/iasc.2023.035352

Hou, L., Pan, X. (2022). Consumers with specialised and diverse experience produce more helpful reviews. *Online Information Review*, 46(4): 645-659. https://doi.org/10.1108/OIR-06-2020-0244

Huang, B., Guo, R., Zhu, Y., Fang, Z., Liu, J., Wang, Y., Fujita, H., Shi, Z. (2022). Aspect-level sentiment analysis with aspect-specific context position information. *Knowledge-Based Systems*, 243(7): 1-11. https://doi.org/10.1016/j.knosys.2022.108473

Huang, C., Jiang, W., Wu, J. (2020). Personalized review recommendation based on users' aspect sentiment. *ACM Transactions on Internet Technology* (*TOIT*), 20(4): 1-26. https://doi.org/10.1145/3414841

Huang, Y., Liu, H., Li, W., Wang, Z., Hu, X., Wang, W. (2020). Lifestyles in Amazon: Evidence from

online reviews enhanced recommender system. *International Journal of Market Research*, 62(6): 689-706. http://dx.doi.org/10.1177/1470785319844146

Liu, H., Ma, X., Zhang, L. (2023). Aspect-based sentiment analysis model integrating match-LSTM network and grammatical distance. *Journal of Computer Applications*, 43(1): 45-50. https://doi.org/10.11772/j.issn.1001-9081.2021111874

Liu, P., Zhang, L., Gulla, J. A. (2021). Multilingual review-aware deep recommender system via aspect-based sentiment analysis. *ACM Transactions on Information Systems* (*TOIS*), 39(2): 1-33. https://doi.org/10.1145/3432049

Lv, Y., Wei, F., Cao, L., Peng, S., Wang, C. (2021). Aspect-level sentiment analysis using context and aspect memory network. *Neurocomputing*, 428(11): 195-205. https://doi.org/10.1016/j.neucom.2020.11.049

Mannepalli, K., Singh, S. P., Kolli, C. S., Raj, S., Bojja, G. R., Rajakumar, B. R., Binu, D. (2023). Popularity prediction model with context, time and user sentiment information: An optimization assisted deep learning technique. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 31(02): 283-302. https://doi.org/10.1142/S0218488523500150

Mewada, A., Dewang, R. K. (2023). SA-ASBA: a hybrid model for aspect-based sentiment analysis using synthetic attention in pre-trained language BERT model with extreme gradient boosting. *The Journal of Supercomputing*, 79(5), 5516-5551. https://doi.org/10.1109/10.1007/s11227-022-04881-x

Nimrah, S., Saifullah, S. (2022). Context-free word importance scores for attacking neural networks. *Journal of Computational and Cognitive Engineering*, 1(4): 187-192. https://doi.org/10.47852/bonviewJCCE2202406

Ortega, J. L. (2022). Classification and analysis of Pubpeer comments: How a web journal club is used. *Journal of the Association for Information Science and Technology*, 73(5): 655-670. https://doi.org/10.1002/asi.24568

Ridderhof, J., Tsiotras, P. (2022). Chance-constrained covariance steering in a Gaussian random field via successive convex programming. *Journal of Guidance, Control, and Dynamics*, 45(4), 599-610. https://doi.org/10.2514/1.G005941

Tang, F., Fu, L., Yao, B. (2019). Aspect based fine-grained sentiment analysis for online reviews. *Information Sciences*, 488(2): 190-204. https://doi.org/10.1016/j.ins.2019.02.064

Wang, C., Yun, M. H. (2020). Cross-cultural difference in product preference in consumer review-based text mining methods: A case study on smart band. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 64(1): 1383-1387. https://doi.org/10.1177/1071181320641330

Wang, Y. (2020). Fine-grained opinion mining on Chinese car reviews with conditional random field. *Journal of Shanghai Jiaotong University* (*Science*), 25(9): 325-332. https://doi.org/10.1007/s12204-020-2184-1

Wu, H., Huang, C., Deng, S. (2023). Improving aspect-based sentiment analysis with Knowledge-aware Dependency Graph Network. *Information Fusion*, 92: 289-299. https://doi.org/10.1016/j.inffus.2022.12.004

Xing, B., Tsang, I. W. (2022). Understand me, if you refer to aspect knowledge: Knowledge-aware gated recurrent memory network. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(5), 1092-1102. https://doi.org/10.1109/TETCI.2022.3156989

Zeng, J., Liu, T., Jia, W., Zhou, J. (2021). Fine-grained question-answer sentiment classification with

hierarchical graph attention network. *Neurocomputing*, 457(5): 214-224. https://doi.org/10.1016/j.neucom.2021.06.040

Zhang, J., Zhang, A., Liu, D., Bian, Y. (2021). Customer preferences extraction for air purifiers based on fine-grained sentiment analysis of online reviews. *Knowledge-Based Systems*, 228(9): 1-15. https://doi.org/10.1016/j.knosys.2021.107259

Zhang, W., Li, X., Deng, Y., Bing, L., Lam, W. (2022). A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 35(11), 11019-11038. https://doi.org/10.1109/TKDE.2022.3230975

Zhang, X., Chen, Y., He, L. (2023). Information block multi-head subspace based long short-term memory networks for sentiment analysis. *Applied Intelligence*, 53(10): 12179-12197. https://doi.org/10.1007/s10489-022-03998-z

Zhao, P., Mao, C., Yu, Z. (2020). Semi-supervised aspect-based sentiment analysis for case-related microblog reviews using case knowledge graph embedding. *International Journal of Asian Language Processing*, 30(3): 12-19. https://doi.org/10.1142/S2717554520500125

Zhou, T., Tao, A., Sun, L., Qu, B., Wang, Y., Huang, H. (2023). Behavior recognition based on the improved density clustering and context-guided Bi-LSTM model. *Multimedia Tools and Applications*, 82(29), 45471-45488. https://doi.org/10.1007/s11042-023-15501-y

Hongliang Cheng
Linzhou College of Architectural Technology, Linzhou 456500, China
E-mail address: lzjzchl@163.com
Major area(s): Architectural Technology & Science

Qingheng Sun
Zhengzhou Urban Construction Vocational College, Zhengzhou 451263, China
North China University of Water Resources and Electric Power, Zhengzhou 450045, China
E-mail address: 15290866637@139.com
Major area(s): Water Resources and Hydropower Engineering

(Received April 2025; accepted December 2025)