2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

RWTLA-Prompt: Leveraging Prompt Learning and Deep Networks for Sentiment Analysis

Yunxiang Nan [1][202230603096@mails.zstu.edu.cn], Hongyun Huang [2][422720005@qq.com] and Zuohua Ding [1][zuohuading@zstu.edu.cn]

[1] School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China
[2] Center of Library Big Data Processing and Analysis, Zhejiang Sci-Tech University, Hangzhou, 310018, China

Abstract. Social media texts often contain informal expressions such as emojis, which pose challenges for conventional sentiment analysis models. To address this, we propose RWTLA-Prompt, a novel sentiment analysis framework that integrates deep neural networks with prompt learning. Our backbone model combines RoBERTa WWM with an improved ETextCNN module and BiLSTM to capture both local semantic cues and global contextual information. Specifically, the ETextCNN module is designed to enhance the representation of emojis and their surrounding text through separate convolutional pathways. In addition, we introduce Emoji2vec embeddings to enrich emoji semantics, and utilize soft prompt templates generated using keyword extraction techniques to guide the model toward sentiment-relevant features. Experiments conducted on two benchmark datasets, SMP2020 EWECT and Weibo_senti_100k, show that our model consistently outperforms strong baselines in both accuracy and F1 score. Ablation studies further demonstrate the effectiveness of emoji sensitive convolutional modeling and prompt driven guidance.

Keywords: Sentiment Analysis, RoBERTa-WWM, ETextCNN, Prompt Tuning.

1 Introduction

Sentiment analysis (SA) is an important task in Natural Language Processing (NLP) that aims to identify and classify the sentiment expressed in text. It is widely applied in fields such as public opinion monitoring, brand analysis, and social media analysis [1]. With the proliferation of social media and instant messaging tools, the way emotions are expressed in text has become increasingly diverse, especially with the use of emoji, which significantly enhances emotional communication in text [2]. However, traditional sentiment analysis methods often fail to fully understand and leverage the emotional information conveyed by non-text symbols like emojis, making the accurate analysis of sentiment in social media text more complex [3].
Most existing sentiment analysis models primarily focus on processing pure text data and overlook the use of emojis, which are widely used in social media. Although recent

studies have started to explore the impact of emojis on sentiment analysis, there are still several challenges: on the one hand, emojis often carry emotional information, but their semantics are closely tied to the textual content, and traditional word embedding methods struggle to effectively capture their emotional features [4]; on the other hand, emotional expression in social media text is highly variable and implicit, and how to capture the relationship between text and emoji to improve sentiment classification remains a challenge [5].

In addition, pre-trained language models (PLMs) have become mainstream methods in sentiment analysis, as these models can effectively capture the semantic information of text through large-scale pretraining [6]. However, PLMs still face challenges when applied to downstream tasks, especially in sentiment analysis, where emotional information is diverse and complex. Despite the powerful contextual modeling capabilities of PLMs, they still struggle to capture the nuances of emotional features due to the variety of sentiment categories and expressions [7].

To address these issues, prompt learning methods have emerged as a solution [8]. Prompt learning works by adding prompt templates to the input, guiding the model to better adapt to specific downstream tasks. In tasks such as sentiment analysis, which involve diversity and complexity, prompt learning helps the model focus on key emotional features, thus improving its performance [9]. While PLMs provide strong pre-training capabilities, the introduction of prompt learning can further compensate for the limitations of models in sentiment analysis, especially when dealing with social media text that involves both emojis and multiple sentiment polarities.

To address these challenges, this paper proposes an innovative hybrid sentiment analysis model that combines RoBERTa-WWM, ETextCNN, BiLSTM, and multi-head attention mechanisms, and introduces Emoji2Vec embedding technology, which effectively captures the emotional information conveyed by emojis [10]. The multi-head attention mechanism plays a crucial role in the model, learning to focus on important information in different feature spaces in parallel, thus improving the model's ability to perceive sentiment-related keywords[11].In addition, this study incorporates keyword extraction and prompt learning techniques by leveraging the KeyBERT model to automatically extract semantically meaningful keywords from raw text, which are then used to construct soft prompt templates. This approach effectively enhances the performance of sentiment classification in terms of both accuracy and generalization.

Experimental results show that the proposed model performs significantly better than traditional sentiment analysis methods and existing pre-trained model-based methods when handling social media text rich in emojis.

Our contributions are as follows:

1.A sentiment analysis model integrating emoji semantics: This paper proposes a hybrid neural architecture that combines RoBERTa-WWM, ETextCNN, BiLSTM, and multi-head attention mechanisms. By incorporating Emoji2Vec embeddings, the model differentiates between emoji symbols and regular text, enhancing its ability to interpret emotional cues embedded in informal expressions.

2.ETextCNN module for enhanced emoji feature modeling: To capture the unique affective roles of emojis in text, the proposed ETextCNN module extracts both local and global contextual features based on emoji positions. By applying convolutional kernels of varying sizes, it effectively models fine-grained semantic associations and improves the robustness of sentiment detection in emoji-rich scenarios.

3.Keyword-guided soft prompt generation mechanism: To address the task misalignment between pre-training and downstream fine-tuning, a soft prompt learning strategy is introduced. The model automatically generates prompt templates guided by sentiment-relevant keywords extracted using KeyBERT, which leverages contextualized embeddings from BERT to identify semantically salient phrases. This method reduces reliance on manually crafted templates, enhances the alignment between input prompts and sentiment categories, and significantly improves both accuracy and generalization performance.

2 Related Work

2.1 Pre-trained Model

In In recent years, pretrained language models (PLMs) have made significant progress in various natural language processing (NLP) tasks, especially in sentiment analysis. BERT captures contextual information through bidirectional attention mechanisms, greatly improving text representation ability. However, BERT's application in Chinese texts faces challenges due to the lack of clear word boundaries in Chinese. To address this issue, the RoBERTa-WWM model was proposed, which masks entire words rather than sub-word units, improving the ability to process Chinese texts.Zhang et al. [12] proposed a method that combines RoBERTa-WWM with Bidirectional Long Short-Term Memory (BiLSTM). This approach dynamically integrates the vector weights generated by the 12 layers of transformers in RoBERTa-WWM and uses them as input to the BiLSTM network, fully utilizing the deep pretrained information of RoBERTa and the advantages of BiLSTM in sequence modeling. This combination not only enhances the model's contextual awareness but also achieves remarkable results in tasks like Chinese sentiment analysis.

2.2 Deep Learning Models in Sentiment Classification

Deep learning models, particularly Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have been widely applied in sentiment analysis tasks. Kim's [13] TextCNN model performs exceptionally well by using convolutional layers to extract local features from text, making it a fast and efficient model for sentiment classification. While RNNs can capture sequential relationships in text, they often suffer from issues like vanishing gradients and long training times. To overcome these issues, many variants of RNNs, such as Long Short-Term Memory (LSTM) [14], have been proposed. LSTMs can capture long-range dependencies in sequences, making them effective for sentiment classification tasks. However, LSTM models alone are insufficient for capturing local semantic information. Consequently, several hybrid models combining CNN and LSTM have been proposed to enhance sentiment analysis accuracy. For example, Rehman et al. [15] proposed a hybrid CNN-LSTM model that extracts features using CNN and processes them with LSTM for improved accuracy.

2.3 Application of Prompt Learning in Sentiment Analysis

As pretrained language models continue to scale up, researchers have focused on developing efficient methods to fully utilize these models, especially in low-resource settings. Prompt learning, which includes prompt tuning and context learning, has emerged as a promising solution for improving model performance with minimal data [16]. GPT-3 introduced the concept of in-context learning [17]. which inspired numerous follow-up studies In sentiment analysis, prompt learning works by embedding input texts into prompt templates, transforming classification tasks into fill-in-the-blank problems, thus leveraging pretrained models more effectively.

Several studies have explored how prompt learning can enhance sentiment classification. Gao et al. [18] proposed Pattern-Exploiting Training (PET), a semi-supervised training approach that transforms input samples into cloze forms using manually designed prompt templates, significantly improving performance in few-shot tasks. With the development of automatic prompt generation techniques, research has shifted towards reducing manual effort [19]. For instance, Deng et al. introduced a reinforcement learning-based method for generating discrete prompts, improving training efficiency while maintaining high performance [20].

3 RWTLA-Prompt Model

The RoBERTa-WWM-TextCNN-LSTM-Att-Prompt (RWTLA-Prompt) model proposed in this study integrates soft prompt learning with deep learning techniques to improve the robustness and accuracy of Chinese sentiment analysis. The model architecture consists of the following five key components,the model structure diagram is shown in Figure 1.

In the input layer,the original text is combined with prompt templates, and soft prompt templates are generated through keyword extraction. This process guides the model to focus on specific words and contexts relevant to sentiment analysis, enhancing its ability to capture emotional features. Soft prompt learning effectively improves the model's performance in low-resource tasks and optimizes the application of pre-trained language models.

In the word embedding layer, we use RoBERTa-WWM as the text encoder, leveraging its strong contextual modeling capabilities to address ambiguities in Chinese text. This is particularly useful for capturing semantic information in Chinese, where word boundaries are not always clear. Additionally, we introduce Emoji2Vec to represent emojis, integrating emoji embeddings with contextual word vectors. This enriches the emotional information dimension in the input text, enabling the model to better recognize emotional symbols.

In feature extraction layer,the model utilizes an improved convolution module, ETextCNN (Emoji-Improved TextCNN), along with BiLSTM (Bidirectional Long Short-Term Memory) for feature extraction. ETextCNN uses a dual-path convolution structure that retains the multi-scale convolution advantages of traditional TextCNN while adding a dedicated convolution path to focus on emojis and their contextual relationships. This enhances the model's ability to capture local emotional symbols and their interactions with the context. BiLSTM, on the other hand, is used to model long-range dependencies in the text, capturing the evolution of emotional information over time and improving the model's global semantic expression.

In the feature fusion layer,the model employs a multi-head attention mechanism to capture emotional keywords, syntactic structures, and emotional signals in parallel across different subspaces. This allows the model to process multi-level features of the text more effectively and strengthens its ability to identify emotional tendencies. The fully connected layer is responsible for fusing the different features and further enhancing the model's performance.

In the output layer,the model uses a softmax layer to perform sentiment classification and output the emotional category of the text. Based on the features extracted and fused in the previous layers, this layer enables precise sentiment analysis and prediction.

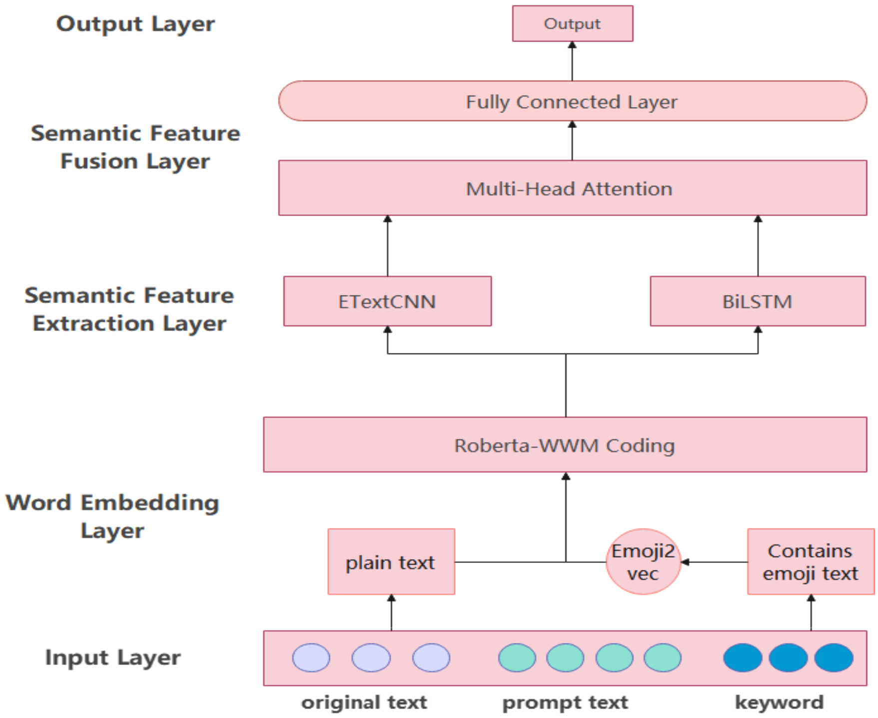

Fig. 1. The architecture diagram of our model.

3.1 Input Layer

The method of using KeyBERT for keyword extraction

To enhance the guidance ability of the prompt templates for textual sentiment features, this paper employs the KeyBERT model for keyword extraction from the raw text. KeyBERT is based on BERT embeddings and automatically identifies keywords by calculating semantic similarity [21], which allows it to better capture semantic relationships within the text. The process is as follows: first, BERT is used to extract document embeddings to obtain document-level representations; second, the document is tokenized, and BERT embeddings are extracted for each word or phrase; finally, cosine similarity is computed to identify the words or phrases most similar to the document, which are then selected as keywords. These keywords are subsequently used to construct soft prompt templates, further enhancing the sentiment feature guidance of the sentiment analysis model.,the specific workflow is shown in Figure 2.

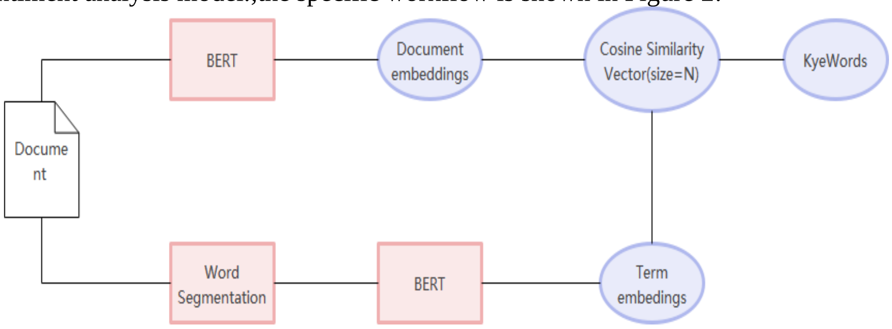

Fig. 2. The KeyBERT workflow figure.

We selected the top 3 keywords extracted using KeyBERT, and Table 1 presents a sample of the texts after keyword extraction.

Table 1. Text before and after partial keyword extraction.

Original text	Extracted keywords
Important festival! Wishing you happiness! For those celebrating, enjoy the holiday.	festival, happiness, celebrating
Just finished hotpot at this restaurant—it was way too spicy, I couldn't finish it, and the overall experience was poor.	spicy, couldn't, poor
I don't like this clothing brand. The quality is poor, the customer service was bad, and I left a negative review.	poor, bad, negative
Absolutely loved this tourist spot. The scenery was stunning and truly refreshing.	loved, stunning, refreshing

generation of prompt text

Another essential task in the input construction phase is the generation of prompt text. The key idea is to design prompt sentences containing a [MASK] token, aligning the sentiment classification objective with the original masked language modeling task of the pre-trained model. This alignment enhances the compatibility between downstream tasks and the pre-trained language model, thereby better leveraging its semantic understanding capabilities.

Based on a thorough investigation and the linguistic characteristics of the datasets used in this study, we adopt a postfix prompt structure, where the original input precedes the prompt. Accordingly, several prompt templates were designed, as illustrated in Table 2.

Table 2. Example of prompt template.

Prompt	MASK Position
Emotional inclination is [MASK]	6
The emotion behind this sentence is [MASK]	8
That's really amazing	4
The above content reflects the emotions of [MASK]	8
This is [MASK]	3

Soft prompt text construction

To fully leverage the capabilities of prompt learning, we concatenate the prompt template (Prompt), the original text (Input Sentence), and the keyword information extracted through dependency parsing (Keywords) to construct a comprehensive input. The concatenated input sequence is then fed into RoBERTa, and its form is represented as follows:

$$\mathrm{X}^{\text{input}} = [\mathrm{CLS}] + \mathrm{sen} + [\mathrm{SEP}] + \mathrm{X} + [\mathrm{SEP}] + \tau + [\mathrm{SEP}] \tag{1}$$

Where sen denotes the original input sentence, with length n. X represents the constructed prompt text, with length n.Generated from the keywords, with length m. Special tokens such as [CLS], and [SEP] are used to mark sentence boundaries and align the structure.

3.2 Word Embedding Layer

Roberta-WWM Embedding

BERT is a pre-trained model proposed by Google in 2018 that demonstrated exceptional capabilities in word vector extraction within English contexts. However, semantic understanding in Chinese significantly differs from that in English. Chinese sentences lack explicit word delimiters, with words often composed of multiple characters, complicating Chinese word segmentation. Directly applying BERT’s initial sentence model to Chinese can result in splitting a complete word into several independent characters. During BERT's pre-training, these characters are randomly replaced with "[MASK]", failing to effectively extract meaningful semantic information.

The RoBERTa-wwm model builds upon BERT to better capture the deep bidirectional relationships in Chinese. BERT’s core architecture utilizes 12 layers of bidirectional transformer encoders and employs a multi-head self-attention mechanism (MHA) to reduce the distance between any two words in a sentence to 1. BERT's pre-training objectives consist of the Masked Language Model (MLM) and Next Sentence Prediction (NSP). In MLM, 15% of the tokens in the input sentence are randomly selected, with 80% of these replaced by "[MASK]", 10% replaced by random words, and the remaining 10% left unchanged, thereby improving the model's feature representation and generalization.The structural diagram is shown in Figure 3.

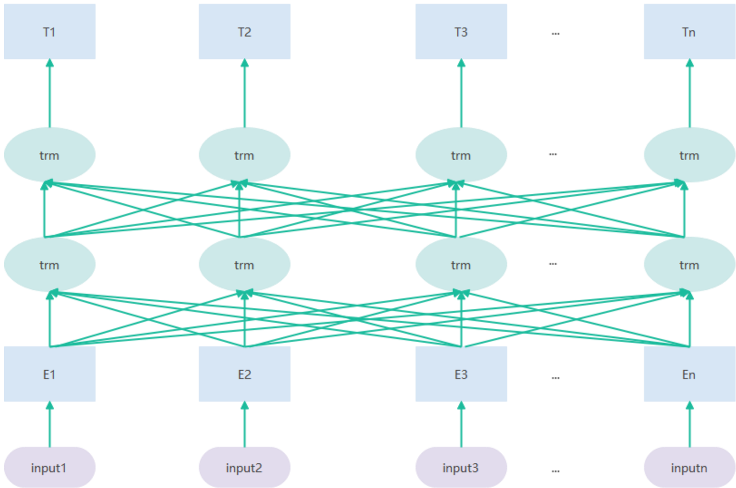

Figure 3. The structure of the RoBERTa-wwm model.

Emoji2vec Embedding

With the widespread use of emoji in social media, sentiment analysis models need to effectively understand these non-textual symbols. To address this, this paper introduces Emoji2vec technology, which embeds emojis into the same semantic space as traditional text to enhance the model's ability to represent emojis. Similar to word vectors, Emoji2vec trains on the Unicode descriptive text of emojis (e.g., "😂" corresponds to "face with tears of joy") to convert them into semantically consistent embedding vectors. These vectors are then used as additional features, combined with text vectors, to further improve the model's accuracy in sentiment analysis for text containing emojis.The core idea of this method is to train a mapping function f that assigns a vector representation to each emoji. This vector representation is optimized by minimizing the difference between the emoji's semantic vector *w(e)* and its generated vector $f(e)$.In other words, the goal is to ensure that the generated emoji vector $f(e)$ is as close as possible to its semantic description vector *w(e)*.Finally, the obtained emoji vectors, denoted as X_{emoji}, are input into the model along with traditional text vectors for further processing. The complete representation of the embedding layer is as follows:

$$\mathrm{X} = \mathrm{X_{Text}} \oplus \mathrm{X_{emoji}} \tag{2}$$

3.3 Semantic Feature Extration Layer

ETextCNN Model

In sentiment analysis tasks, capturing local semantic features of text is crucial. To effectively extract these local features, the traditional TextCNN model performs feature extraction of local information by parallel operations of multi-scale convolutional kernels on the input text. However, when dealing with texts that contain special symbols such as emojis, the conventional TextCNN convolution operations are not specifically optimized for the emotional expression conveyed by these symbols.

To address this issue, this paper proposes an enhanced emoji feature extraction module, EmojiImproved-TextCNN (ETextCNN). The ETextCNN module combines the multi-scale feature extraction capabilities of traditional TextCNN while enhancing the modeling ability of emojis and their surrounding context through improved convolutional pathways. ETextCNN first receives the contextual features output by the traditional TextCNN and further processes the parts containing emojis through a dedicated convolutional path. This path utilizes small-scale convolutional kernels to allow the model to focus on the local regions containing emojis, thereby improving the modeling of the emotional relationship between emojis and their neighboring context. In this path, smaller convolutional kernels (1×256) and a smaller stride (stride of 1) are used to finely capture short-range information near the emoji. This operation aims to retain the close connection between the emoji and its context, avoiding the loss of information.
Additionally, ETextCNN uses another path with larger convolutional kernels (3×256) and a longer stride (stride of 2) to perform convolution on the entire text, further extracting sentence-level semantic features. The structure of the model is shown in Figure 4. This path not only captures global semantic information but also enhances the understanding of long-distance semantic relationships, which is particularly important for tasks such as sentiment analysis, where determining the overall emotional tendency of the entire sentence is crucial.

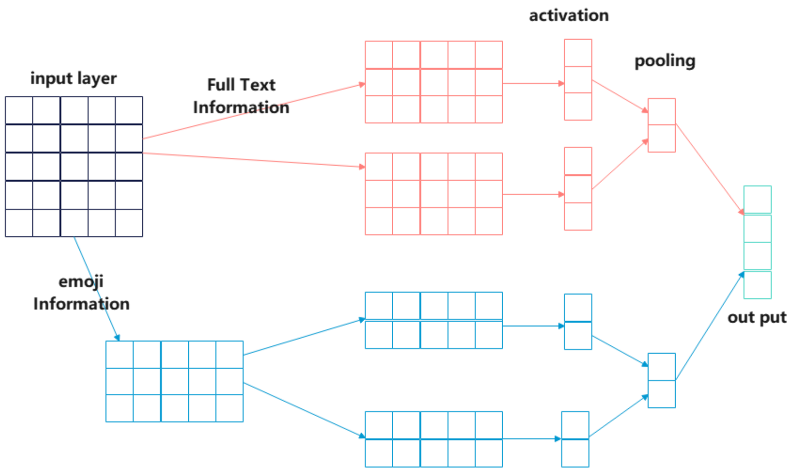

Figure 4. The structure of the ETextCNN.

Through this dual-path convolutional structure, ETextCNN not only inherits the global feature extraction ability of TextCNN under multi-scale convolutions but also enhances its sensitivity to local emotional symbols, such as emojis, and their surrounding context. The resulting feature vector exhibits significant expressive power in both global and local semantic levels.
The innovation of the ETextCNN module lies in its combination of the traditional TextCNN's convolutional feature extraction capabilities with a specialized convolutional path designed to strengthen the emotional features of emojis. This fusion improves the model's ability to understand and model rich emotional symbols in sentiment analysis tasks, especially when special symbols like emojis are prese

BiLSTM Model

Bidirectional Long Short-Term Memory networks (BiLSTM) differ from traditional recurrent neural networks. LSTM effectively addresses the long-term dependency problem through its carefully designed memory cell architecture. BiLSTM consists of two LSTM units with opposite directions, which integrate the outputs of both forward and backward LSTMs to simultaneously capture forward and backward semantic information in text sequences. This bidirectional processing mechanism gives BiLSTM significant advantages in extracting bidirectional semantic features from text. Specifically, when a sequence of word vectors is input to the model, BiLSTM performs two parallel computational processes: one processes the text in its natural order (from front to back), while the other processes it in reverse order (from back to front). The outputs from both directions are then fused to obtain a feature representation that contains complete contextual information. The structure of the model is shown in Figure 5.

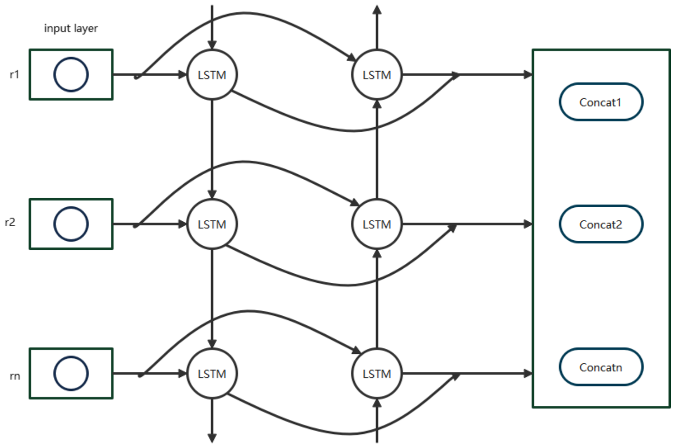

Figure 5. The structure of the BiLSTM diagram.

The BiLSTM layer enhances textual sentiment feature extraction through its bidirectional semantic modeling mechanism. The word vector sequence $X=\{x_1,x_2,...,x_n\}$ is simultaneously processed by both forward and backward LSTM networks. The forward LSTM generates hidden state sequences $\{\vec{h}1,\vec{h}2,...,\vec{h}t\}\in\mathbb{R}^d$ word by word to capture left-to-right contextual information, while the backward LSTM produces $\{\overleftarrow{h}1,\overleftarrow{h}2,...,\overleftarrow{h}t\}\in\mathbb{R}^d$ to capture reverse semantic relationships, where d represents the hidden layer dimension of the LSTM unit. This bidirectional processing enables the model to concurrently consider both preceding and subsequent semantic relationships in the text, more comprehensively capturing long-range dependency features and effectively improving the recognition capability of sentiment keywords. Finally, the hidden state sequences from both directions are concatenated at corresponding positions to form a complete bidirectional feature representation.

$$\overrightarrow{h_t} = \overrightarrow{\mathrm{LSTM}}(x_t) \quad (3)$$

$$\overleftarrow{\mathrm{h}}_{\mathrm{t}} = \overleftarrow{\mathrm{LSTM}}(\mathrm{x}_{\mathrm{t}}) \tag{4}$$

$$\mathrm{h} = \left[\overrightarrow{h_{\mathrm{t}}} \oplus \overleftarrow{\mathrm{h}}_{\mathrm{t}}\right] \tag{5}$$

3.4 semantic feature fusion layer

Multi-head Attention Layer

The multi-head attention mechanism breaks down the attention process into multiple "heads," each learning an independent attention distribution. This allows the model to focus on different parts of the sequence simultaneously, capturing local features and filtering key information, thus improving sentiment analysis performance. Specifically, the attention heads focus on different sentiment feature spaces, including: the ETextCNN module consists of two convolutional paths. One path extracts local sentiment keywords, identifying emotionally charged words or phrases in the text, such as "like" and "hate," especially in n-gram structures. The other path focuses on processing sentiment symbols like emojis and their context, enhancing the model's attention to emotional expressions through emojis. It captures the semantic interactions between emojis and nearby words while preserving global semantic information, improving sentiment analysis accuracy.Other attention heads focus on processing long-range dependencies extracted by BiLSTM, capturing long-distance semantic relationships in the context. This is crucial for sentiment analysis, especially when sentiment information spans longer text spans.

Specifically, the model's input is passed into the multi-head attention mechanism through the Query (Q), Key (K), and Value (V) matrices. Each head calculates its attention score, normalizes it using the Softmax function, and then multiplies the weights by the Value matrix to obtain the weighted output. The outputs of all heads are finally concatenated and passed through.

$$\mathrm{Attention}(\mathrm{Q},\mathrm{K},\mathrm{V}) = \mathrm{softmax}(\frac{\mathrm{QK^T}}{\sqrt{\mathrm{d_k}}})\mathrm{V} \tag{6}$$

the weighted outputs from each attention head are computed, they are combined to form the overall output of the multi-head attention mechanism. The combined output is then passed through a linear transformation to obtain the final result of the multi-head attention mechanism, as shown in the following formula:

$$\mathrm{head_i} = \mathrm{Attention}(\mathrm{QW_i^Q}, \mathrm{KW_i^Q}, \mathrm{VW_i^Q}) \tag{7}$$

$$\mathrm{M} = \mathrm{Concat}(h\mathrm{ead_i}, \ldots., h\mathrm{ead}_h)\mathrm{W^o} \tag{8}$$

In the multi-head attention mechanism, i denotes the index of the i-th attention head. The matrices W_i^Q 、 W_i^Q 、 , W_i^Q represent the learnable parameter matrices for queries, keys, and values corresponding to the i-th attention head, respectively. Through matrix multiplication operations QW_i^Q 、 KW_i^Q 、 VW_i^Q the model generates independent query, key, and value representations for each attention head. The Concat() function is responsible for concatenating the outputs of multiple attention heads, while W^o is the linear transformation matrix that maps the concatenated matrix to the final output space, where the superscript *O* denotes the Output transformation. This design enables the model to process multiple representation subspaces in parallel, thereby more comprehensively capturing the semantic information of the input sequence.

3.5 Output Layer

The output layer primarily utilizes the Softmax function on the final fused feature to obtain the ultimate sentiment result. The equation is as follows.

$$\mathrm{P} = \mathrm{softmax}(\mathrm{WfF} + \mathrm{b_f}) \qquad (9)$$

The feature extraction result *F* is transformed into a probability distribution *P* through the Softmax function, where each element represents the predicted probability of the corresponding category. This process is achieved via a linear transformation using the learnable weight matrix *Wf* and bias term *bf*, ultimately selecting the category with the highest probability value as the predicted output.

4 Experiment

This section presents experiments aimed at thoroughly evaluating the performance of the proposed model.

4.1 Datasets

To validate the effectiveness of our algorithm, we conducted experiments on two Chinese datasets: SMP2020-EWECT and WeiboSenti-100k.
SMP2020-EWECT： This dataset is composed of virus-related and normal content. We used the commonly used subset, which includes over 27,000 Weibo texts. These texts are labeled into six emotion categories: happy, angry, sad, scared, surprised, and neutral.
WeiboSenti-100k： This dataset consists of over 110,000 comments scraped from popular topics on Weibo. The total corpus includes approximately 50,000 positive sentiment texts and 50,000 negative sentiment texts.

4.2 Experimental Setup

The experiments were conducted using the PyTorch deep learning framework. An NVIDIA GeForce RTX 4090 GPU was used, and the system ran on Ubuntu 22.04. The model parameters and their descriptions are provided in Table 3.

4.3 Results Analysis

Evaluation method

Table 3. Experimental Parameters.

Name of experimental parameter	Parameter Values
Max Length Of Sentences	120
Learning Rate	0.00001
Batch Size	64
Epochs	30
Dropout_rate	0.5
Optimizer	Adam

Consistent with evaluation metrics used in other sentiment analysis studies, we use accuracy and macro F_1 score to compare our model with baseline methods.

Template experiment

Based on the RWTLA-Prompt model, a comparative analysis of the impact of different templates on model performance enables a comprehensive evaluation of the effectiveness of prompt learning in sentiment analysis tasks. The specific results are shown in Table 4.

Table 4. Comparison of Experimental Results with Different Templates.

Template	ACC(%)		F_1(%)	
	SMP2020-EWECT	Weibo_senti_100k	SMP2020-EWECT	Weibo_senti_100k
Emotional inclination is [MASK]	**81.73**	**98.52**	**78.53**	**98.44**
The emotion behind this sentence is [MASK]	81.29	98.43	78.15	98.38
That's really amazing	80.87	98.31	77.43	98.26
The above content reflects the emotions of [MASK]	81.43	98.42	78.29	98.34
This is [MASK]	81.12	98.36	77.84	98.28

Through comparative experiments conducted on multiple Chinese sentiment analysis datasets, it was observed that the prompt “The emotional inclination is [MASK]” achieved the highest accuracy and F_1 scores, consistently outperforming other templates across two datasets. This indicates that the semantic structure of this prompt aligns well with the sentiment classification task, effectively guiding the pre-trained language model to identify and predict the emotional orientation within the text.

It is important to highlight that all prompt templates in this study were constructed by integrating sentiment-related keywords extracted via dependency parsing. By concatenating these keywords with the original input text and then appending the prompt template, the final input sequence not only preserved the overall semantic context but also emphasized the core emotional elements of each sentence. This fusion strategy played a positive role across all evaluated prompts, contributing to the overall improvement in sentiment classification performance.

Analysis of experimental results

Table 5. Here Is The Performance Comparison Of Different Methods On Datasets

Model	*ACC*(%)		F_1(%)	
	SMP2020-EWECT	Weibo_senti_100k	SMP2020-EWECT	Weibo_senti_100k
TextCNN	71.84	94.15	63.23	94.11
BiLSTM+Attention	73.63	95.37	67.79	95.36
RoBERTa-WWM	76.52	95.46	72.17	95.41
RoBERTa-WWM+TextCNN	77.68	96.36	73.85	96.27
RoBERTa-WWM+BiLSTM	78.14	96.48	75.36	96.38
RWTLA-Prompt	**81.73**	**98.52**	**78.53**	**98.44**

The experimental results indicate that the proposed method demonstrates significant performance improvements across two datasets (see Table 5).

The experimental results demonstrate that our proposed RWTLA-Prompt model exhibits comprehensive advantages when compared against TextCNN, BiLSTM+Attention, RoBERTa-WWM, and its hybrid variants RoBERTa-WWM+TextCNN and RoBERTa-WWM+BiLSTM. Among the baseline neural network models, TextCNN effectively captures local textual features, while BiLSTM+Attention enhances comprehension of lengthy texts through sequence modeling and attention mechanisms, though these conventional models still show limitations in deep semantic representation. The introduction of the RoBERTa-WWM pretrained model significantly improves text representation quality through its dynamic word embeddings and extensive pretrained knowledge, confirming the fundamental value of pretrained language models. The enhanced hybrid architecture RoBERTa-WWM+TextCNN demonstrates notable performance gains by combining global contextual understanding with local feature extraction, whereas RoBERTa-WWM+BiLSTM further improves performance in complex contextual scenarios through reinforced sequence modeling capabilities - these results collectively validate the effectiveness of multi-feature fusion strategies. Our final RWTLA-Prompt model innovatively incorporates prompt learning mechanisms, preserving the semantic depth of pretrained models while strengthening the capture of crucial emotional cues through feature interaction mechanisms, achieving optimal performance across all test scenarios with remarkable adaptability and robustness. The progressive performance improvement from basic neural networks to pretrained models and ultimately to RWTLA-Prompt conclusively demonstrates the efficacy of our proposed multi-component fusion architecture and prompt learning strategy, offering a novel solution for textual sentiment analysis tasks.

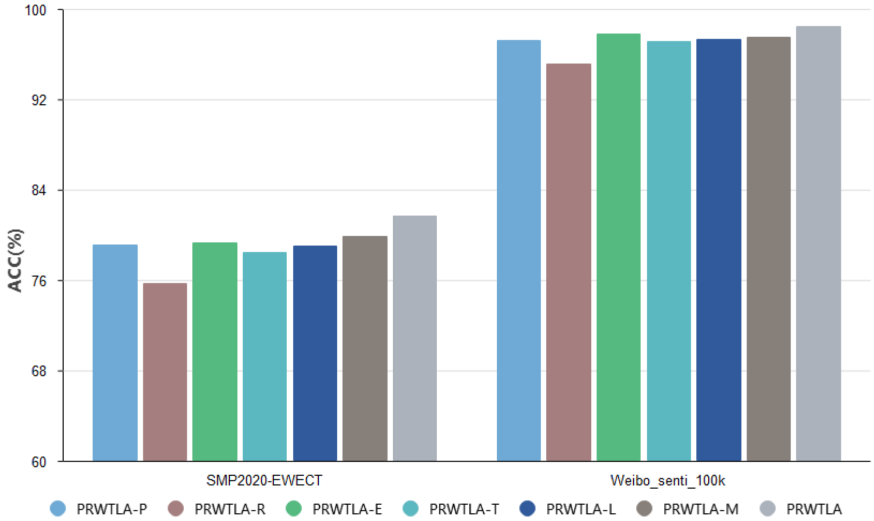

Fig. 6. The Accuracy of ablation experiment.

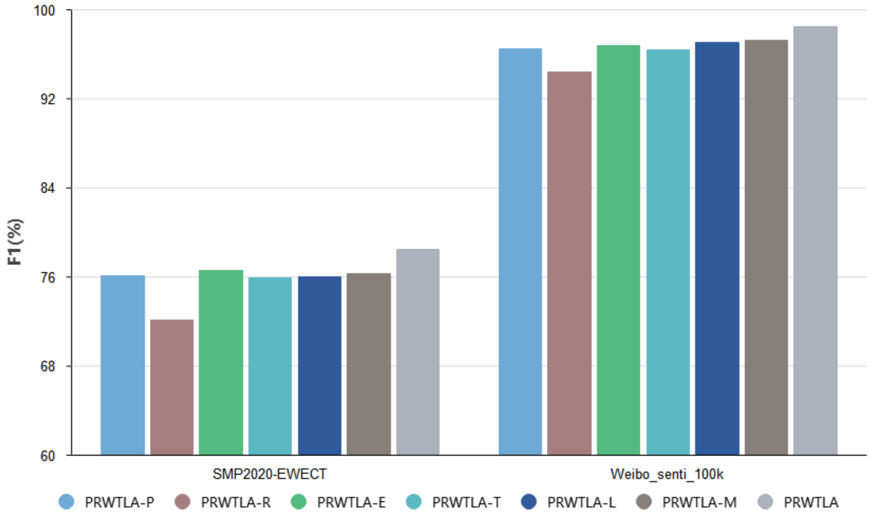

Fig. 7. The F_1 of ablation experiment.

4.4 Ablation study

PRWTLA-P: Removes the prompt learning component
PRWTLA-R: Removes the RoBERTa-WWM pretrained model

PRWTLA-E: Removes the Emoji2vec module for emoji handling
PRWTLA-T: Removes the ETextCNN module
PRWTLA-L: Removes the BiLSTM component
PRWTLA-A: Removes the multi-head attention mechanism
By conducting these ablation experiments, we aim to quantitatively assess the contribution of each component to the overall performance of the PRWTLA model and provide insights into the effectiveness of each individual component in improving sentiment analysis results. The result are shown in Figure 6 and Figure 7.
The PRWTLA model achieved the best performance on multiple datasets, with accuracy rates of 81.73% and 98.52%, and F1 scores of 78.53% and 98.44% on the SMP2020-EWECT and Weibo_senti_100k datasets, respectively. These results demonstrate that PRWTLA significantly improves sentiment analysis performance by integrating multiple modules.
In the ablation study, removing the prompt learning module significantly reduced model performance, indicating that prompt learning is crucial for enhancing the model's understanding of input text. Removing RoBERTa-WWM caused a sharp decline in performance across all datasets, highlighting its essential role in handling complex semantics and long-range dependencies. Removing Emoji2vec weakened sentiment feature extraction, particularly in recognizing emoji and sentiment-related features. Removing ETextCNN, a convolutional module designed to enhance emoji feature modeling, negatively impacted short-text sentiment analysis. ETextCNN effectively captures fine-grained relationships between emojis and their surrounding context, improving the representation of emoji sentiment expressions. Removing BiLSTM led to a performance drop in long-text analysis, as BiLSTM is critical for modeling long-distance dependencies. Removing the multi-head attention mechanism weakened the model's ability to capture global features, affecting the processing of long-range dependencies.
In summary, the ablation study results indicate that each component of the PRWTLA model plays a critical role in sentiment analysis. Removing any key module results in a noticeable performance decline, validating the effectiveness and complementarity of these components. Furthermore, the superior performance of the complete PRWTLA model confirms that integrating multiple deep learning structures can significantly enhance sentiment analysis, especially in multi-category and long-text scenarios.

5 Conclusions

This paper proposes an innovative sentiment analysis model that combines RoBERTa-WWM, ETextCNN, BiLSTM, and multi-head attention mechanisms, with a particular focus on optimizing texts containing emoji symbols. By incorporating Emoji2Vec embeddings, the model effectively distinguishes between emoji symbols and regular text, enhancing its ability to interpret emotional cues embedded in informal expressions, thereby improving sentiment analysis performance in such contexts.
Additionally, the proposed ETextCNN module utilizes multi-scale convolution operations to capture both local and global contextual features around emojis. This fine-grained modeling of the semantic associations between emojis and their surrounding

words greatly enhances the model's sentiment recognition capability, particularly in emoji-rich texts. The design enables the model to better handle emotional symbols and their interactions with other parts of the text, ultimately improving sentiment analysis accuracy.

Furthermore, a keyword-guided soft prompt learning mechanism is introduced. By using KeyBERT to extract sentiment-related keywords, the model automatically generates prompt templates. This method addresses the task misalignment between pre-training and downstream fine-tuning, reduces reliance on manually crafted templates, and significantly improves both accuracy and generalization performance.

Experimental results demonstrate that the proposed model significantly enhances sentiment analysis accuracy and F1 scores on datasets like SMP2020-EWECT and Weibo_senti_100k, validating the effectiveness and advantages of the methods proposed in this study.

Overall, this work provides new insights for sentiment analysis, especially when dealing with emoji-rich and informal texts. Future research could explore applying this model in low-resource scenarios or further optimizing model performance by incorporating additional emotional symbols and contextual features.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), pp. 79–86. Association for Computational Linguistics (2002)
2. Shiha, M., Ayvaz, S.: The effects of emoji in sentiment analysis. International Journal of Computer and Electrical Engineering (IJCEE) 9(1), 360–369 (2017)
3. Zhang, L., Wang, S., Liu, B.: Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8(4), e1253 (2018)
4. Eisner, B., Rocktäschel, T., Augenstein, I., et al.: emoji2vec: Learning emoji representations from their description. In: Proceedings of the 4th Workshop on Natural Language Processing for Social Media (SocialNLP 2016), pp. 48-54. Association for Computational Linguistics (2016)
5. Xuanyuan, M., Xiao, L., Duan, M.: Sentiment classification algorithm based on multi-modal social media text information. IEEE Access 9, 33410-33418 (2021)
6. Devlin, J., Chang, M.W., Lee, K., et al.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019), vol. 1, pp. 4171-4186. Association for Computational Linguistics (2019)
7. Sun, C., Qiu, X., Xu, Y., et al.: How to fine-tune BERT for text classification? In: Tang, J., Kan, M.-Y., Zhao, D. (eds.) CCL 2019: China National Conference on Chinese Computational Linguistics. LNCS, vol. 11856, pp. 194-206. Springer, Cham (2019)

8. Schick, T., Schütze, H.: Exploiting Cloze-Questions for Few-Shot Text Classification and Natural Language Inference. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021): Main Volume, pp. 255-269. Association for Computational Linguistics (2021)
9. Zhou, P., Shi, W., Tian, J., et al.: Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016), vol. 2, pp. 207-212. Association for Computational Linguistics (2016)
10. Liu, Y., Ott, M., Goyal, N., et al.: RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692 (2019)
11. Vaswani, A., Shazeer, N., Parmar, N., et al.: Attention Is All You Need. In: Advances in Neural Information Processing Systems 30 (NIPS 2017), pp. 5998-6008. Curran Associates, Inc. (2017)
12. Zhang, Y.Q., Wang, Y., Li, B.: Identifying Named Entities of Chinese Electronic Medical Records Based on RoBERTa-wwm Dynamic Fusion Model. Data Analysis and Knowledge Discovery 6(2/3), 242-250 (2022)
13. Kim, Y.: Convolutional Neural Networks for Sentence Classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), pp. 1746-1751. Association for Computational Linguistics, Stroudsburg (2014)
14. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. Neural Computation 9(8), 1735-1780 (1997)
15. Rehman, A.U., Malik, A.K., Raza, B., et al.: A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis. Multimedia Tools and Applications 78, 26597-26613 (2019)
16. Brown, T., Mann, B., Ryder, N., et al.: Language Models Are Few-Shot Learners. In: Larochelle, H., Ranzato, M., Hadsell, R., et al. (eds.) Advances in Neural Information Processing Systems 33 (NeurIPS 2020), pp. 1877-1901. Curran Associates, Inc. (2020)
17. Reynolds, L., McDonell, K.: Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21), pp. 1-7. Association for Computing Machinery, New York (2021)
18. Gao, T., Fisch, A., Chen, D.: Making Pre-trained Language Models Better Few-Shot Learners. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021), pp. 3816-3830. Association for Computational Linguistics (2021)
19. Schick, T., Schütze, H.: Exploiting Cloze-Questions for Few-Shot Text Classification and Natural Language Inference. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021): Main Volume, pp. 255-269. Association for Computational Linguistics (2021)
20. Deng, M., Wang, J., Hsieh, C.-P., et al.: RLPrompt: Optimizing Discrete Text Prompts with Reinforcement Learning. arXiv preprint arXiv:2205.12548 (2022)
21. Issa, B., Jasser, M.B., Chua, H.N., et al.: A Comparative Study on Embedding Models for Keyword Extraction Using KeyBERT Method. In: Proceedings of the 13th IEEE International Conference on System Engineering and Technology (ICSET 2023), pp. 40-45. IEEE (2023)