scientific reports

OPEN

A novel approach for multiclass sentiment analysis on Chinese social media with ERNIE-MCBMA

Youyang Sun[2], Ziyi Yu[1], Yuyu Sun[4], Yaqing Xu[3]✉ & Boming Song[1]✉

Weibo, one of the most widely used social media platforms in China, sees a vast number of users expressing their opinions and emotional tendencies. Conducting sentiment analysis on Weibo posts using natural language processing techniques is crucial for market research and public opinion observation, holding significant commercial and societal importance. However, Chinese expression is highly diverse, making sentiment polarity harder to discern. Traditional sentiment classification algorithms often struggle with insufficient semantic feature extraction and coarse-grained in sentiment classification for Chinese texts. To address these challenges, this paper proposes a Chinese sentiment multi-classification method based on ERNIE-MCBMA. The proposed model extracts the parallel local dependency features between words through the multi-channel CNN convolutional layer and then uses the collaborative architecture of bidirectional LSTM and multi-head attention mechanism to realize context-sensitive feature recalibration. Finally, the shallow syntactic features and deep semantic representation were fused through the cross-layer feature fusion layer to realize the complementary enhancement of more fine-grained semantic information. The SMP2020-EWECT public dataset is used to categorize texts into six classes: neutral, happy, angry, sad, fear, and surprise. Various comparative experiments were conducted on the dataset. The experimental results show that the ERNIE-MCBMA achieves an accuracy of 78.26% and an F1-score of 78.45% for the 6-class classification task, outperforming other baseline models.

Keywords Sentiment analysis, Cross-layer feature fusion, ERNIE, Multi-channel CNN, Multi-head attention

Social media has become increasingly important in modern society. Social media platforms generate vast amounts of content every day, including text, images, videos, and more, reflecting a wide range of personal and group viewpoints, emotions, and attitudes. It has also become a significant channel for business, politics, and cultural influence. Consequently, there is a growing demand for automated sentiment analysis methods[1].

Sentiment analysis, also known as opinion mining, aims to identify or study emotions, opinions, subjective information, or attitudes hidden within human communication[2]. This technology plays a crucial role in the field of natural language processing (NLP) and has seen significant development over the past few decades[3,4]. Sentiment analysis based on user-generated text can help governments, businesses, and individuals understand the emotions and opinions of users, providing support for decision-making in various fields[5,6]. Existing sentiment analysis methods can be classified into lexicon-based approaches, machine learning-based approaches, deep learning and pre-trained model-based approaches. After long-term development, these methods have been effectively applied and validated in multiple domains such as comment analysis[7], public opinion analysis[8], mental health analysis[9], recommendation systems[10], etc.

With the availability of abundant annotated datasets and the development of complex language models, the effectiveness of sentiment analysis has been predominantly demonstrated in English[11]. However, significant gaps remain in sentiment analysis for non-English languages, where annotated data is often insufficient or missing[12,13]. Unlike English texts which have relatively fixed grammar structures and direct expressions, Chinese texts often convey emotions implicitly, where emotional nuances may be embedded within the context. This demands models that can deeply understand the underlying meanings of the text. Additionally, Chinese includes numerous polysemous words, synonyms, and complex semantic roles and dependency relationships, necessitating models that can grasp a wide range of vocabulary usage and grammatical rules. Most sentiment

[1]School of Medical Information and Engineering, Xuzhou Medical University, Xuzhou 221004, Jiangsu, China. [2]School of Computer Science and Technology, Soochow University, Suzhou 215006, Jiangsu, China. [3]School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou 221116, Jiangsu, China. [4]School of Humanities, Jinling Institute of Technology, Nanjing 210038, Jiangsu, China. ✉email: yaqingxu@jsnu.edu.cn; boming.song@xzhmu.edu.cn

analysis methods only identify emotional tendencies at a macro level, such as positive, negative, or neutral[14–21]. However, the sentiment in Chinese texts is not as easy to quantify as in English, and more refined analysis methods are needed to assess the intensity of emotions.

Over the course of long-term development, Chinese sentiment analysis methods have achieved satisfactory performance in binary classification tasks[22,23]. However, their performance declines rapidly when faced with multi-class problems. To address the challenges of low accuracy and coarse-grained analysis in Chinese sentiment classification, this paper proposes ERNIE-MCBMA, a cross-layer feature fusion architecture specifically tailored for Chinses text sentiment analysis. Leveraging the pre-trained model ERNIE, the ERNIE-MCBMA is capable to capture intricate Chinese grammatical rules, lexical nuances, and deep semantic insights. On this basis, we propose a three-stage progressive feature fusion mechanism, which uses a multi-channel Convolutional Neural Network (CNN) to extract the local sequence features between words and then processes them through the BiLSTM-multi-attention layer to capture the comprehensive global sequence features. In the fusion stage, we employ one-dimensional Max pooling, feature concatenation, and Mish activation function to seamlessly integrate original, local, and global features of text sequences, and then forward them to fully connected layers for final decision or classification. This design constructs a complete cross-layer feature interaction path.

To validate our approach, we harness open-access datasets, meticulously preprocessing them to eliminate specific formatting strings (e.g., @xxx, < xxx>) and non-Chinese characters, facilitating smoother text processing and analysis downstream. Extensive experiments on these datasets demonstrate that ERNIE-MCBMA outperforms benchmark models across key metrics such as accuracy, highlighting its superiority. The objective of this study is to elevate the accuracy of sentiment recognition and classification to achieve a fine-grained sentiment analysis, thereby offering a novel and effective methodology for more nuanced Chinese text sentiment analysis.

This paper is structed as follows: “Related works” section, introduces the development of sentiment analysis and recent works related to Chinese sentiment analysis; “Methodology” section, proposes the novel architecture ERNIE-MCBMA; “Experiments and results” section, experiments and results analysis; “Discussion” section, discusses the advantages, limitations, and prospects of the proposed architecture; “Conclusion” section, summarizes the research in this work.

Related works

Sentiment analysis can be categorized into four main fields: lexicon-based methods, machine learning-based methods, deep learning-based methods, and pre-trained model-based methods.

Lexicon-based methods

The lexicon-based method has played a pivotal role in the field of sentiment analysis. These approaches rely on a meticulously constructed sentiment lexicon to classify texts by calculating sentiment weights. The foundation of this method lies in building a comprehensive and detailed sentiment dictionary that includes not only positive and negative keywords but also different levels of intensity for sentiment terms and their synonyms. Subasic et al.[24] focused on defining the lexicon, establishing the intensity levels of sentiment words, and specifying the relationship between vocabulary and sentiment categories, laid the groundwork for early sentiment analysis.

With the rapid development of the internet and the continuous emergence of new vocabularies, lexicon-based methods face the challenge of frequent updates to adapt to evolving language environments. Esuli et al.[25], developed SentiWordNet based on WordNet, using cosine similarity between words to identify synonyms, enriching the sentiment lexicon with more accurate and nuanced expressions. Additionally, Khoo et al.[26]’s WKWSCI lexicon demonstrated advantages in sentiment classification of non-comment texts such as news headlines by comparing it with five existing lexicons, emphasizing its effectiveness in scenarios with scarce training data. In the Chinese context, Lee et al.[27] ‘s “Chinese EmoBank” serves as a specialized sentiment resource for Mandarin, supporting multi-dimensional sentiment analysis. It captures more complex emotional states within texts, considering multiple dimensions such as types of emotions and their intensities. This approach provides more precise sentiment analysis results, which are crucial for understanding the underlying sentiments in Chinese texts.

The construction of sentiment lexicons laid the foundational groundwork for early sentiment analysis, with Hu et al.[28] proposing a straightforward method that counts positive and negative sentiment words for sentiment prediction. This approach, while effective in its simplicity, faced limitations when dealing with complex language expressions and context-dependent meanings. To address these challenges, Ding et al.[29] introduced a more holistic lexicon-based sentiment analysis technique that incorporated external evidence and linguistic conventions expressed in natural language. This allowed for better handling opinion words relevant to specific contexts, significantly improving the accuracy of sentiment predictions by considering the nuances of language use.

Building on this foundation, lexicon-based methods have been widely applied across various industries. Huang et al.[30] for instance, developed SentiCNN, a sentiment analysis model that integrates convolutional neural networks with dictionary information. By employing an attention mechanism to highlight emotionally significant vocabulary, SentiCNN enhances the performance of online comment sentiment classification. Similarly, Lin et al.[31] devised a lexicon-driven prompting approach tailored specifically for the financial domain, where labeled data is scarce. Their aim was to correct erroneous predictions related to financial terminologies, thus boosting overall performance and enabling more accurate analyses of financial documents.

The rapid evolution of the Internet has led to the emergence of a vast array of new vocabularies, making it challenging for traditional sentiment analysis methods that rely on static lexicons to keep pace, especially when dealing with internet slang and rapidly changing expressions. To address this issue, there is a growing trend towards utilizing machine learning technologies, which can enhance the accuracy and adaptability of sentiment analysis by automatically learning from data without needing specific rules for each case, thus better coping with the dynamic nature of language.

Machine learning-based methods

In the field of sentiment analysis, the application of machine learning mainly relies on supervised learning algorithms. Instead of relying on keywords representing sentiment categories or intensities, these algorithms automatically identify, and extract sentiment related information based on manually annotated text data, such as sentences or paragraphs. This approach allows the model to understand the complex sentiment behind the text and can handle texts that do not directly express sentiment but have an implicit sentiment tendency.

Chikersal et al.[32] developed a Support Vector Machine (SVM) text classification scheme integrated with semantic rules, which demonstrated good performance in text sentiment analysis. Govindarajan et al.[33] integrated a naive Bayes algorithm with a genetic algorithm to devise an innovative text classification approach, which demonstrated a marked enhancement in classification accuracy in movie reviews analysis. By optimizing the feature selection process, this method effectively reduces the influence of noise features and enhances the learning effect of key features. This contribution complements the work of Chikersal et al. as it shows how model performance can be further improved by improving feature engineering. Tarmer et al.[34] evaluated three methods of naive Bayes, SVM, and decision trees, for sentiment analysis of movie reviews, with SVM achieving the highest accuracy. This suggests that although different algorithms have their own advantages, certain algorithms may be more suitable to deal with complex feature dimension problems in specific tasks such as sentiment analysis. This also indirectly supports the importance of Chikersal et al. for context understanding and Govindarajan et al. for feature selection optimization. Li et al.[35] proposed a two-stage hybrid model for sentiment classification of Chinese micro-blogs. An enhanced sentiment lexicon is used and SVM, and KNN, with an improved feature selection technique is employed to boost feature discriminability. Considering the unique characteristics of Chinese microblogs, such as their brevity, colloquialism and cultural background, this approach represents a deepening and development of previous work. It demonstrates how to adjust models and techniques in accordance with specific language and social cultural contexts to better meet the demands of practical applications.

While machine learning-based methods have significantly reduced the dependence on sentiment dictionaries and demonstrated superior performance within specific domains, their application in cross-domain sentiment analysis is still constrained by limited generalization capabilities. In contrast, deep learning offers substantial advantages, such as the automatic extraction of comprehensive features, outstanding generalization ability, and superior performance on large-scale datasets. These strengths have made deep learning an increasingly important focus in the field of sentiment analysis.

Deep learning-based methods

The difficulty in capturing long-distance dependencies within texts and the challenge of better extracting global and local features from text sequences have long been constraints hindering the development of deep learning-based sentiment analysis. Although traditional recurrent neural networks (RNNs) and their variants, LSTM and GRU, can partially address the problem of long-distance dependencies, they still encounter the issue of information attenuation as the sentence length increases. In order to solve this problem, Alayba et al.[36] proposed a model combining LSTM and CNN, this approach not only solves the problem of local information loss, but also is able to identify long-term dependencies in text sequences. In addition, Sun et al.[37] constructed a deep learning model based on CNN, Attention, and BiLSTM to fully extract deep hidden features from short texts and used residual networks to solve the problem of gradient vanishing as the number of network layers increases.

The TextCNN model was first proposed by Kim[38]. The innovation lies in its ability to convert text into word vectors and then use convolutional neural networks for text classification. This approach is particularly suitable for capturing local features in text. However, in order to capture both global and local features, Li et al.[39] proposed a deep learning sentiment classification model called RSCOEWR, which integrates CNN, BiLSTM, BIGRU and attention mechanism, effectively solves high-dimensional problems and enhances contextual attention, and shows excellent generalization ability on public Chinese sentiment datasets.

Abyaad et al.[40] used Word2vec to represent text and used LSTM and gated recurrent unit (GRU) networks as the backbone of the model to classify news articles from titles and short texts. Lyu et al.[41] proposed a text-driven method that uses an additional user-product cross-context module to model the association between users and products, and achieved an accuracy of 59.0%, 72.1% and 72.6% on IMDb, Yelp2013 and Yelp-2014 English benchmark test sets, respectively. Kanwal et al.[42] proposed a model that combines Stacked Auto-encoder and LSTM. The former is used to extract relevant information features, and the latter performs sentiment analysis based on the extracted features, achieving an accuracy of 87% on the IMDB dataset.

Data quality issues or limited data volumes can severely impact model performance, causing unstable and unreliable results. The data annotation process is time-consuming, labor-intensive, and requires professional knowledge to ensure consistency and accuracy. Low-quality data can lead models to learn incorrect patterns, while insufficient data restricts their ability to capture complex features, both weakening the model's generalization and overall performance.

To address these challenges, researchers have begun to employ pre-trained models. These models, pre-trained on large-scale datasets, not only alleviate data scarcity issues but also significantly enhance performance on specific tasks. Leveraging transfer learning, pre-trained models can quickly adapt to new tasks with limited data, achieving better results. The pre-training strategy accelerates model convergence and improves adaptability to different domains and tasks, providing an effective solution.

Pre-trained model-based methods

Pre-trained models can capture profound semantic information. With additional fine-tuning and optimization, their performance can be substantially enhanced, which has led to their widespread adoption. Vaswani et al.[43] introduced a novel neural network model, named Transformer. Unlike traditional RNNs or CNNs, Transformer models can more accurately identify interrelationships in input sequences by introducing attention mechanisms and multi head attention techniques. One important application of Transformer-based architectures is BERT, a pretrained language model developed by Devlin et al.[44]. BERT has demonstrated excellent performance in a variety of NLP tasks, so more and more researchers choose to use this model for text processing and sentiment analysis. For example, Jia et al.[45]proposed a sentiment classification framework for Chinese microblogs that utilizes character-level and sentence-level vectors for sentiment mining, achieving significant improvements over baseline methods in two classification tasks. Deng et al.[46] developed a Chinese word-level BERT and a transformer fusion framework for sentiment intensity prediction, outperforming other methods on the Chinese EmoBank and demonstrating adaptability for various downstream tasks. Peng et al.[47] proposed that while deep learning (DL) models have some weaknesses, such as long training time and difficult convergence, the biologically inspired learning (BL) model has the advantages of simple structure, incremental modeling, and short training time. They explore the combination of BERT and BL, named BBL, for the task of text emotion classification. Additionally, the performance of the proposed BBL is verified through experiments conducted on two public datasets. Ding et al.[48] proposed a model combining BERT with a custom attention mechanism that adjusts encoder weights to distinguish between similar yet distinct emotions, improving classification accuracy. This model outperformed traditional ones in precision, recall, and F1-score on two six-emotion datasets.

Sun et al.[49] proposed the ERNIE model, which improves BERT's Mask strategy by enhancing language representation through knowledge masking strategies, including entity level and phrase level masking, to improve the accuracy and richness of semantic representation. They confirmed that the model has achieved good results in five Chinese natural language processing tasks. Huang et al.[50] proposed the ERNIE-BiLSTM-Att (EBLA) model for three-class sentiment classification, which uses ERNIE's dynamic word vectors, BiLSTM, and an attention mechanism to achieve high precision, recall, and F1-scores on JD.com's Chinese product reviews, outperforming other deep learning models in e-commerce sentiment analysis. Yang et al.[51] proposed an advanced sentiment classification model that combines ERNIE with a modified DPCNN, achieving a Macro-F1 score of 87.03% on the EmoInt dataset and outperforming six baseline models.

Currently, there is a scarcity of research on Chinese sentiment classification beyond three classes. In terms of research methods, there are few methods that explore the feature fusion between deep and shallow layers of the model, and there are problems of insufficient semantic feature capture. To better tackle the task of Chinese multi-level sentiment analysis, this paper employs the ERNIE model for text vector representation and proposes the -MCBMA architecture to capture more nuanced semantic feature information. Experimental results demonstrate that the proposed -MCBMA architecture, when used with ERNIE, BERT, and Word2vec as word embedding layers, achieves higher accuracy and better classification performance, with the optimal results observed when paired with ERNIE.

Methodology

This paper proposed a Multi-channel CNN-BiLSTM-Multi-Att (MCBMA) architecture based on ERNIE, which specifically includes a word embedding layer, multi-channel convolutional layer, BiLSTM-Multi-Attention layer, cross-layer feature fusion layer, and fully connected layer, to achieve cross layer feature fusion for text classification. The framework of the model is shown in Fig. 1. The upcoming content will describe the role of each component within the architecture and explain how to integrate these components.

The text data is first fed into the ERNIE pre-trained model to obtain a representation of the word vectors of the text sequence. Then, the text sequence is converted into a two-dimensional matrix by dimension transformation, so that the 2D convolution kernel can simultaneously capture the parallel local dependencies and deep semantic associations between words. Then, Bidirectional Long Short-Term Memory (BiLSTM) layers were stacked directly after the 2D convolution output to avoid the dimensionality reduction loss of local features caused by the pooling layer. Then, the context-sensitive feature recalibration was realized through the collaborative architecture of bidirectional LSTM and multi-head attention mechanism. Finally, the output of the embedding layer, the output of the convolutional layer, and the output of the collaborative architecture with LSTM and multi-head attention mechanism were fused and connected with the output of the pooling layer from the ERNIE pre-trained language model. Through the above multi-level design, we construct a complete cross-layer feature interaction path, which integrates shallow syntactic features and deep semantic representation to realize the complementary enhancement of more fine-grained semantic information.

Word embedding

In the proposed architecture, the ERNIE pre-trained model serves as the word embedding layer. The process of constructing word embedding vectors for ERNIE includes the following three parts:

(1) Token Embedding: Map the characters in the input text to corresponding vector representations based on the pre-trained word embedding table.
(2) Position Embedding: The absolute positional encoding is employed in the ERNIE. A fixed positional encoding vector is appended to each word before the embedding layer to represent its position in the sentence.
(3) Sentence Embedding: To convert the semantic information within a complete sentence into a corresponding sentence vector, Dialogue Embedding method and Semantic Embedding method are employed. Dialogue Embedding, which is utilized for text modeling, incorporates both the global and local contextual information of the text, enabling a deeper understanding of the textual context. Meanwhile, Semantic Em-

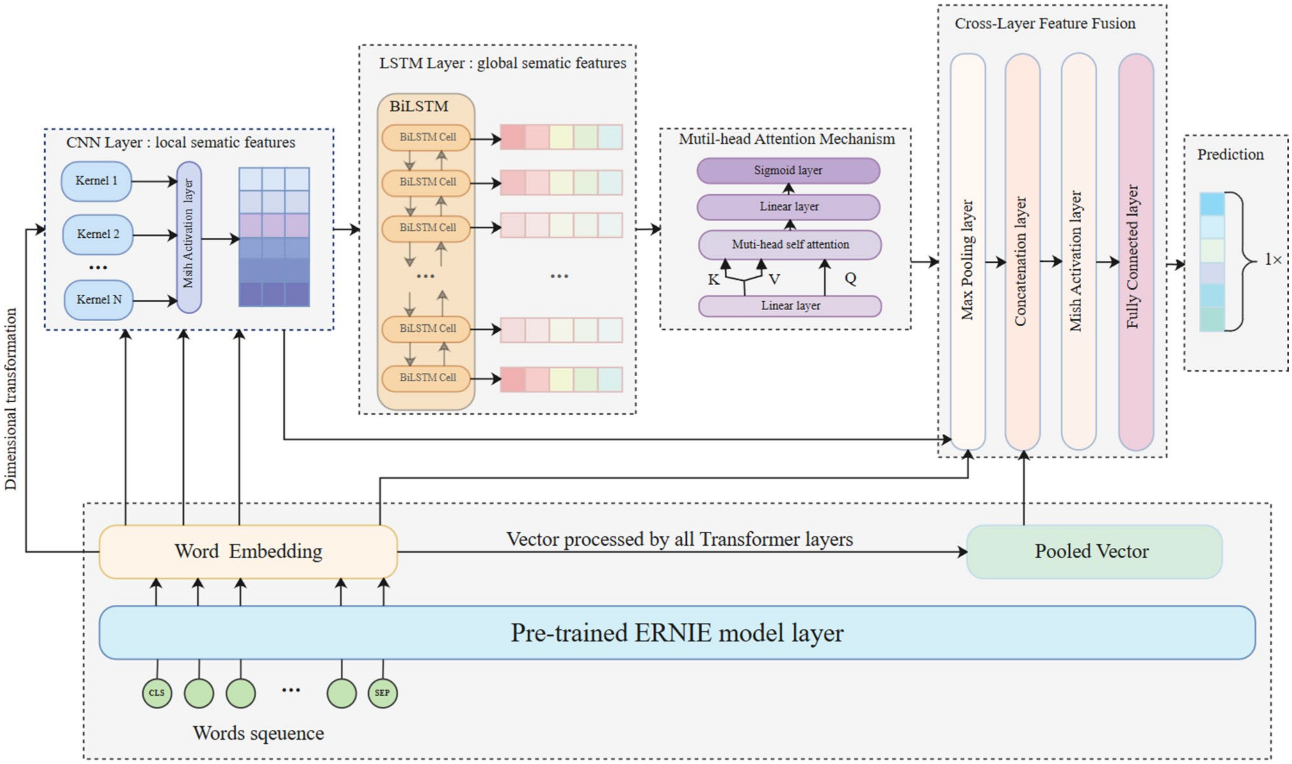

Fig. 1. The architecture of the ERNIE-MCBMA.

bedding integrates concepts such as entities from a knowledge graph into the model, thereby enriching the model's semantic comprehension of the input text.

Then, the word embedding vectors can be represented as:

$$E = E_{x_i} + P_i + S_{st_i} \tag{1}$$

where x is the text sequence input, x_i is the *i*-th vocabulary in the text sequence, E_{x_i} is the word embedding vector of x_i, P_i is the positional embedding vector of x_i, st_i is the sentence type to which the vocabulary x_i belongs, and S_{st_i} is the sentence type embedding vector.

Convolutional layers

The main purpose of using multi-channel CNN convolutional layers is to extract sequence text features within one step. Through the transformation of dimensions, the original one-dimensional sequence text representation is converted to two-dimensional space, so that the 2D convolution kernel can synchronously capture the parallel local dependencies between words in the orthogonal dimension. Its convolution process, as shown in Fig. 2.

For single channel convolution:

$$c_i = f\left(w \cdot x_{i:i+j-1} + b\right) \tag{2}$$

where w is the weight matrix of the corresponding convolution kernel, $x_{i:i+j-1}$ represents the local vector composed of the *i*-th element to the $(i + j - 1)$-th element in the single channel sequence after dimensional transformation, $f\,(\bullet)$ is the activation function, b is the bias, c_i represents the *i*-th eigenvalue after convolution.

An eigenvector C^k can be obtained for the sequence of length *n*:

$$C^k = (c_1, c_2, \ldots, c_{n-j+1}), \quad k = 1{,}2, \ldots, K \tag{3}$$

Finally, the output feature matrix of convolution is formed by all single channel:

$$C = \left\{C^1, C^2, \ldots, C^K\right\} \tag{4}$$

By using activation functions in the convolutional layer, the network can transform linear features into nonlinear features, providing a richer data representation. We used the Mish function as the activation function for CNN to solve the problem of gradient vanishing in the ReLu function under negative input. The Mish can maintain non-linear characteristics even for negative values, alleviating gradient vanishing. Meanwhile, it has smoothness, which can introduce more feature information and promote smooth convergence of the network.

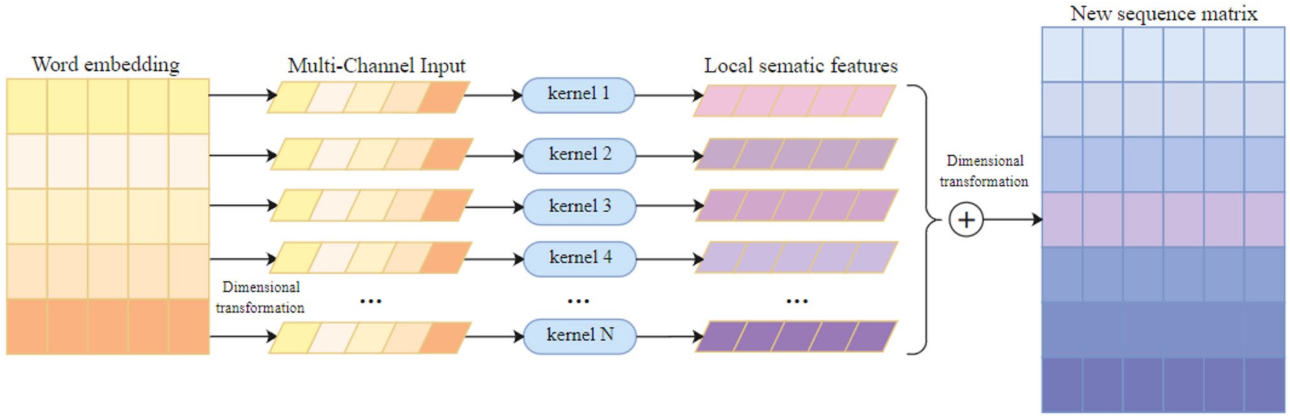

Fig. 2. Structure of the multi-channel CNN.

Feature extraction layers by BiLSTM

The BiLSTM, an extension of LSTM, captures long-term sequence dependencies bidirectionally, providing comprehensive context for feature extraction and stronger non-linear expression[52].

At each moment, the BiLSTM outputs a vector that contains information from both the forward and backward LSTMs. Sequence data is fed into the BiLSTM, where the forward LSTM generates the hidden sequence $\overrightarrow{h_t}$ and the reverse LSTM generates $\overleftarrow{h_t}$. These sequences are then concatenated to produce the final hidden output h_t at time *t*, which is as follows:

$$\underset{h_t}{\rightarrow} = \underset{w_x}{\rightarrow} x_t + \underset{w_h}{\rightarrow} \underset{h_{t-1}}{\rightarrow} + \underset{b_h}{\rightarrow} \tag{5}$$

$$\underset{h_t}{\leftarrow} = \underset{w_t}{\leftarrow} x_t + \underset{w_h}{\leftarrow} \underset{h_{t+1}}{\leftarrow} + \underset{b_h}{\leftarrow} \tag{6}$$

$$h_t = \underset{w_h}{\rightarrow} \underset{h_t}{\rightarrow} + \underset{w_h}{\leftarrow} \underset{h_t}{\leftarrow} + b_h \tag{7}$$

where w is the hidden weights, b represents the bias vector, and x_t is the input time series vector.

Multi-head attention layer

The multi-head attention mechanism boosts model flexibility by enabling parallel self-attention across multiple subspaces, allowing each head to compute attention weights independently and focus on different sequence parts for richer information capture. As shown in Fig. 3, the final output is obtained by concatenating and summing the outputs from all the heads.

The multi-head attention mechanism splits the multi-channel CNN output into multiple heads, each processing the sequence vector to generate Q, K, and V matrices via independent weight multiplication. After computing attention weights with SoftMax, the results are multiplied by V and concatenated across all heads, followed by a linear transformation to yield the final output. The entire process can be represented by the following equations:

$$Attention(Q, K, V) = SoftMax(QK^T/\sqrt{d_k})V \tag{8}$$

$$head_i = Attention\left(Q_i, K_i, V_i\right) \tag{9}$$

$$Multihead = Concat\left(head_1, head_2, \ldots, head_h\right) \tag{10}$$

Cross-layer feature fusion

Cross-Layer Feature Fusion is a technique in deep learning used to integrate feature information from different depth levels. Features from different layers contain varied information: shallow layers typically capture local and detailed information, while deeper layers capture more abstract semantic information. Fusing these features aids the model in comprehensively understanding the input data, thereby enhancing its performance on complex tasks. The concatenation method joins multiple feature sets along the channel dimension, creating a combined feature set with multiple feature channels. In this work, the cross-layer feature fusion technique employs one-dimensional max pooling combined with feature channel concatenation and introduces the Mish activation function. The cross-layer feature fusion process is illustrated in Fig. 4.

The word embedding vectors, the features after convolution, and the features processed by BiLSTM-Multi-Attention are subject to one-dimensional max pooling along a fixed channel dimension. This aims to retain important features while reducing data complexity, computation load, and parameter count. To further enhance the semantic richness of the model, the Pooled layer from the ERNIE model is also used.

The Pooled layer in the ERNIE model is used to capture the global semantic information of the entire input sequence. A [CLS] token is typically added at the beginning of the sequence, and the final hidden state serves

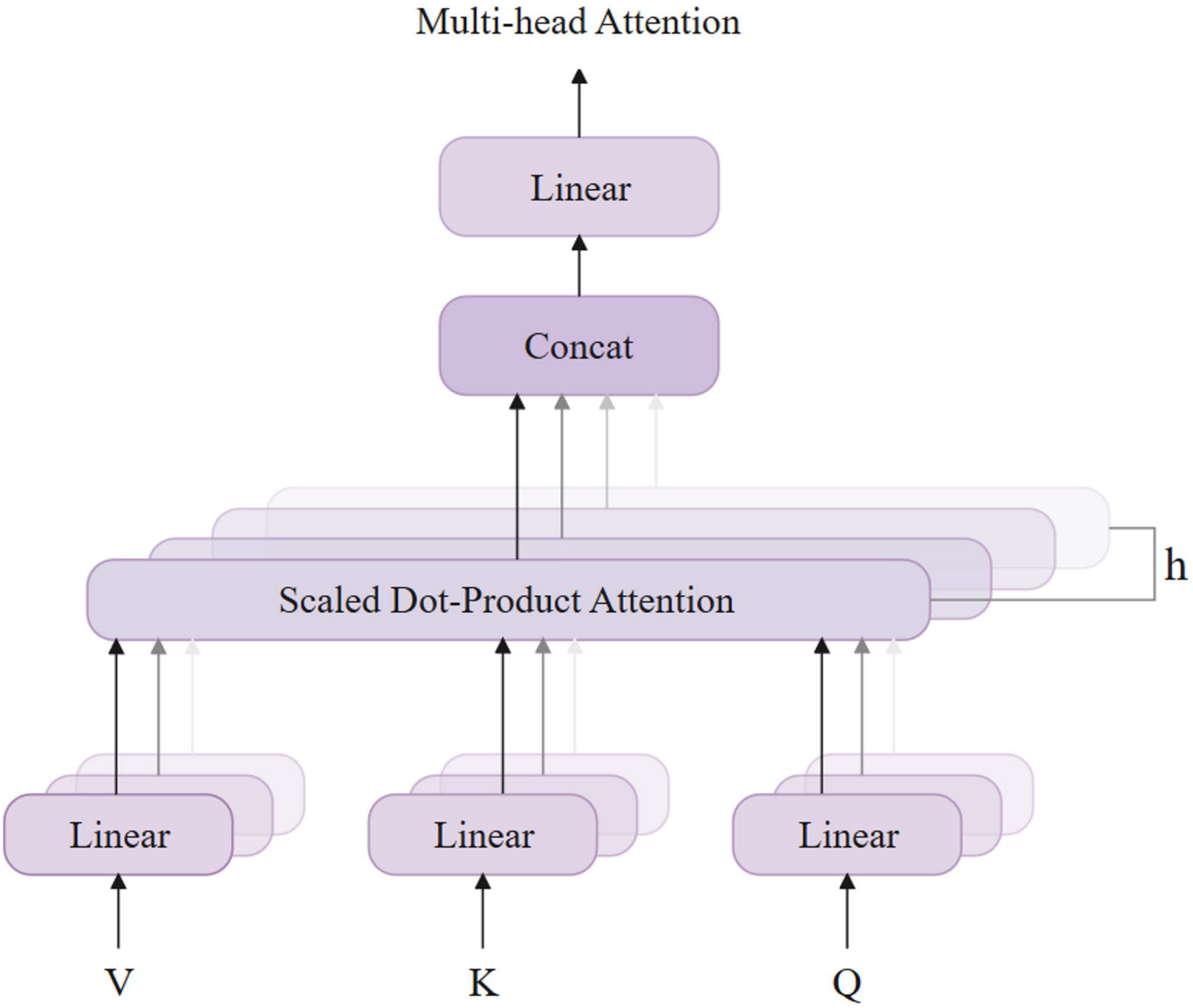

Fig. 3. The illustration of multi-head attention.

as the Pooled layer's output. Leveraging the bidirectional Transformer architecture, the contextual information of each word from the entire sequence can be captured, as illustrated in Fig. 5. While its output can be used for classification tasks, it is often combined with other layer outputs for richer feature representation.

Concatenate the reduced-dimensional sequence features with the outputs of the Pooled layer from the ERNIE model and utilize the Mish activation function to enhance nonlinear expression capabilities. As a result, rich semantic representations have been formed, and the model's language expression ability has been enhanced. The process can be represented as:

$$E_{Cross-Layer} = Mish\left(Concat\left(E_{Pooled}, MaxPool\left(E, E_{CNN}, E_{Multi-Att}\right)\right)\right) \tag{11}$$

where E represents the word embedding vector, E_{CNN} is the convolved feature vector of the sequence, and $E_{Multi-Att}$ is feature vector that has undergone BiLSTM-Multi-Attention of the sequence, $E_{Cross-Layer}$ is the feature vectors after cross-layer feature fusion, and E_{Pooled} represents the feature vectors from the pooled layer of the pre-trained model ERNIE.

ERNIE-MCBMA algorithm

The ERNIE-MCBMA algorithm obtains the corresponding sentiment classification labels from the input text sequence data. The detailed algorithm process is shown in Table 1.

Experiments and results

Dataset and pre-processing

The Chinese text dataset used in this work is the publicly released dataset SMP2020-EWECT (Evaluation of Weibo Emotion Classification Technology). The dataset includes two types of data: one is a general Weibo

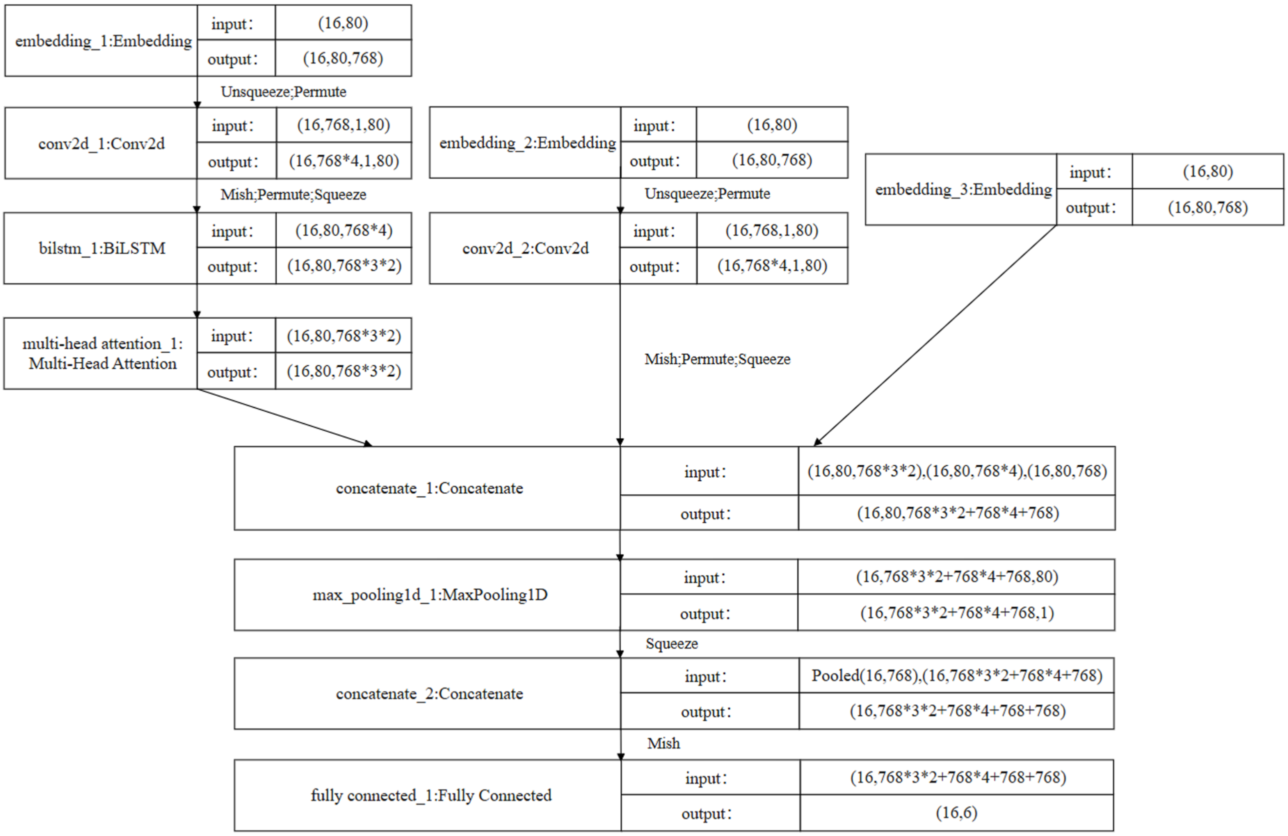

Fig. 4. The process of cross-layer feature fusion.

dataset, which consists of data randomly collected from Weibo encompassing various topics; the other is a Weibo dataset collected during the COVID-19, where all the data is related to the pandemic. Since this study aims to explore the emotions of social media users across different topics, the general Weibo dataset is used.

The general Weibo dataset is divided into the following six categories: neutral (20.70%), happy (19.37%), angry (30.05%), sad (17.97%), fear (4.39%), surprise (7.51%), where percentage represents the proportion in the total sample size. The training dataset in the general dataset includes 27,768 Weibo posts, the validation set contains 2,000 Weibo posts, and the test dataset contains 5,000 Weibo posts. Examples of posts for each emotion category are shown in Table 2.

Data preprocessing: Special characters and punctuation marks in the original text are removed, retaining complete Chinese text, and emotion labels are encoded. After processing, it was found that 25,324 out of the total dataset have text lengths less than 80 characters, accounting for approximately 92%. Therefore, a text truncation length of 80 characters is chosen for this experiment.

Experiment setup

In this work, the experimental platform consists of Python 3.8 running on Ubuntu 20.04 and utilizes the PyTorch framework for programming. The detailed environment configurations are shown in Table 3.

Parameter settings

In this work, the word embedding vector dimension is 768, batch size is set to 16, the number of convolutional kernels is 256, hidden nodes are 768, learning rate is set to 5e-5, and Epoch is 3. The model parameters are shown in Table 4. Besides, the CNN convolutional kernel size is 3 × 3 with padding size of 1.

Evaluation metrics

In this study, we use Acc (Accuracy), Precision (Pre), Recall and F1-score (F1) to evaluate the sentiment classification results of the proposed method. By using TP to represent samples where the actual sentiment is positive and the prediction is also positive, FP to denote samples where the actual sentiment is negative but predicted as positive, TN to represent samples where the actual sentiment is negative and the prediction is also negative, and FN to denote samples where the actual sentiment is positive but predicted as negative, the four evaluation indicators can be defined as follows:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{12}$$

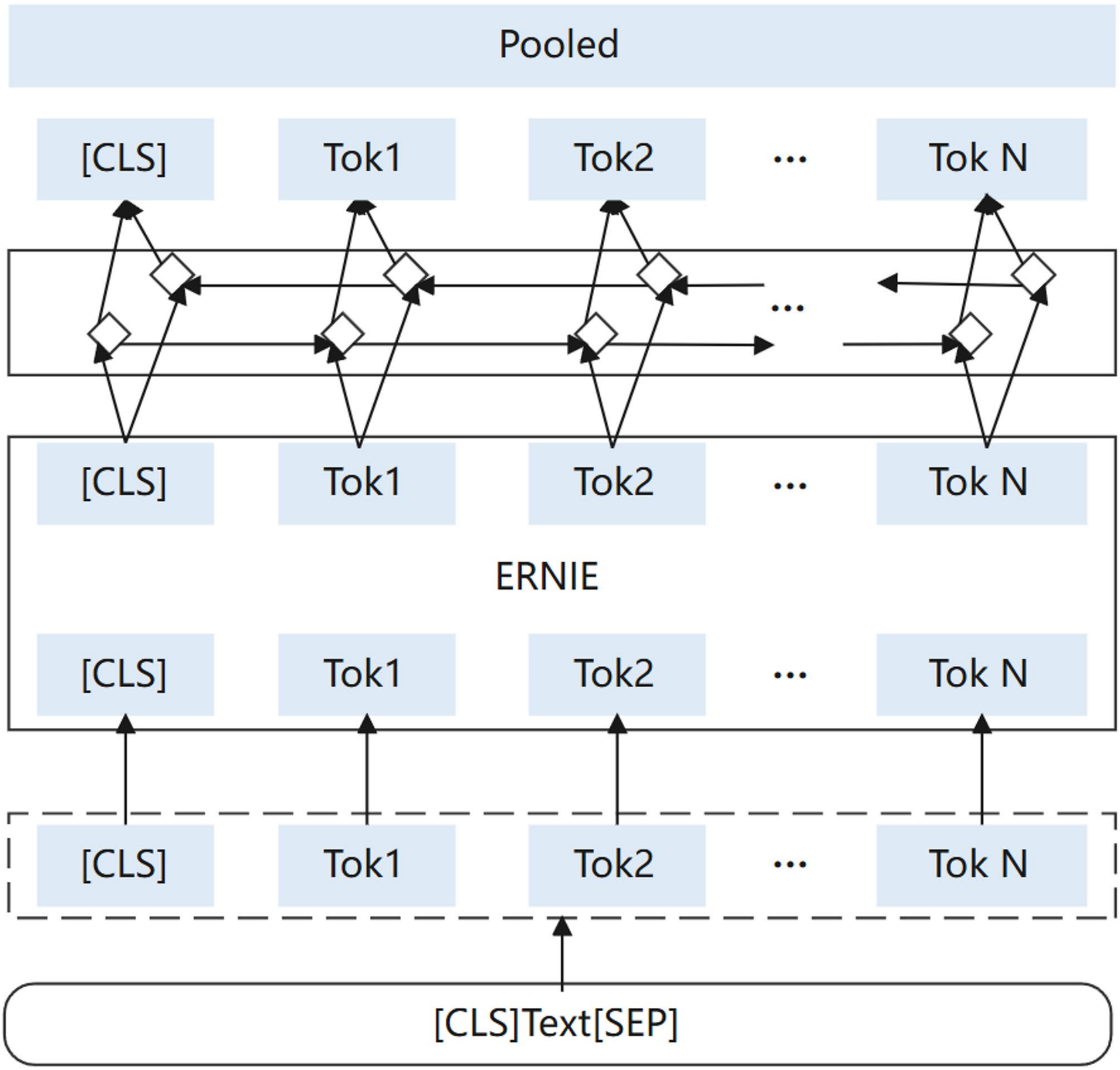

Fig. 5. The structure of the Pooled layer in ERNIE.

Inputs: $X = (x_1, x_2, x_3, \ldots, x_n)$, where X is the input text sequence
Outputs: sentiment classification labels

```
encoder_out, text-cls = ERNIE(context, attention-Mask = mask)
_, pooled = ERNIE(context, attention-Mask = mask)
out = Reshape(encoder_out, [batch_size, 1, seq_len, hidden_size])
out = Permute(out, [0, 3, 1, 2])
cnn_out = Mish(Conv2d(out, kernel_size = 3, padding = 1))
cnn_out = Permute(cnn_out, [0, 3, 1, 2]).squeeze(2)
out, _ = BiLSTM(cnn_out)
Att_out = Mutil-Attention(out, out, out, mask = None)
combined_out = Concatenate(encoder_out, cnn_out, Att_out, axis = 2)
combined_out = Permute(combined_out, [0, 2, 1])
combined_out = MaxPool1d(combined_out, kernel_size = seq_len).squeeze(1)
combined_out = Mish(Concatenate(pooled, combined_out, axis = 1))
final_out = FullyConnected(combined_out)
return final_out
```

Table 1. ERNIE-MCBMA algorithm.

Sentiment	Examples
Neutral	Original: 我们都已经开始自由, 我要做个压抑时放纵的坏人. Translation: We have all started to be free; I want to be a bad person who indulges when feeling oppressed.
Happy	Original: 明明可以靠颜值, 而我却要靠实力. Translation: I could clearly rely on my good looks, but I insist on relying on my abilities.
Angry	Original: 麻将桌上就那男的从一开始就不停的讲. Translation: At the mahjong table, that man has been talking nonstop from the very beginning.
Sad	Original: 太快了。爱情来的太快也走的太快. Translation: It's too fast. The love comes and goes too fast.
Fear	Original: 我想知道这天下究竟要发生什么…简直风云突变. Translation: I want to know what's going to happen in this world… it's like a sudden change in the situation.
Surprise	Original: 中午吃着饭一抬头, 发现小果发芽了! Translation: While eating lunch, I looked up and found that the small fruit had sprouted!

Table 2. Examples of Weibo posts for each emotion category.

Environment	Configuration
Operation system	Ubuntu20.04
CPU	16 vCPU Intel(R) Xeon(R) Gold 6430
GPU Memory	24GB
GPU	NVIDIA GeForce RTX 4090
Framwork	Pytorch2.2.0

Table 3. Environment configurations.

Parameters	Values
dimension of the word embedding vector	768
batchsize	16
number of convolutional kernels	256
hidden_size	768
epoch	3
learning_rate	5e-5
dropout	0.1

Table 4. Parameter settings.

$$Pre = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$F_1 = \frac{2 \times \ Recall \times \ Rre}{Pre + Recall} \tag{15}$$

Due to the presence of class imbalance in the SMP2020-EWECT dataset, the F1-score should be calculated using weighted averaging:

$$F_1 = \frac{2 \times \sum_i w_i Pre_i \times \sum_i w_i Recall_i}{\sum_i w_i Pre_i + \sum_i w_i Recall_i} \tag{16}$$

where w_i represents the proportion of total samples that belong to class i.

Experimental results and analysis

In this study, neural networks, pre-trained models, and their ensemble models are compared on the SMP2020-EWECT dataset. In order to ensure the comparability of experimental results, the selected data sets of the benchmark model are processed in the same way. The benchmark models are roughly divided into three categories:

(1) Neural networks and ensemble models, which include TextCNN, TextRNN, TextRCNN, TextRNN_Att, FastText, and the Transformer. The word embedding vectors of these models are generated by Word2Vec embeddings, except for FastText, which does not require pre-trained word vectors. Through repeated experiments

and meticulous tuning of model parameters, we ultimately identified the parameter settings that enabled these models to achieve optimal performance. The parameters of these models are shown in Table 5.

Additionally, in the CNN network, the convolutional kernel sizes are (2, 3, 4), and the ReLU activation function is used. In the RNN network, the number of LSTM layers is 3, with 128 hidden nodes per LSTM layer. In the RCNN network, there is 1 LSTM layer with 256 hidden nodes.

(2) Pre-trained models: BERT and ERNIE;

(3) Ensembled pre-trained models: Bert-TextCNN, Bert-BiLSTM, Bert-TextRCNN, Ernie-TextCNN, Ernie-BiLSTM, Ernie-TextRCNN, BBL[47] and BCAM[48].

The proposed architecture in this study is ERNIE-MCBMA, which has been compared with the benchmark models. We generate pre-trained word vectors by loading pre-trained models. The parameter settings are shown in Table 4. In addition, to further validate the effectiveness of the proposed -MCBMA architecture, we introduced it into the BERT model for comparative experiments. Due to the differences in the word embedding vectors generated by ERNIE and BERT, we determined through repeated tuning that the BERT-MCBMA model would use the Gelu function as its activation function. All other architectures and processes were kept the same as those in the ERNIE-MCBMA model to ensure consistency and comparability of the experiments. Figure 6 shows the prediction accuracy of the ERNIE-MCBMA model across different classes. The values on the diagonal (0.8, 0.78, 0.84, 0.72, 0.66, 0.74) indicate the probability of each class being correctly predicted. Figure 7 shows the training loss consistently decreased and the training accuracy steadily increased, indicating effective learning and improvement of the model. The model demonstrated good convergence properties, with the loss stabilizing around 0.3 and the accuracy reaching up to 85%, showing promising performance. Table 6 presents the results of various benchmark models and our proposed model on the dataset.

Table 6 presents the baseline results achieved by neural network-based models alongside ours results. It is evident that the -MCBMA architecture demonstrates significant performance advantages. Especially for the ERNIE-MCBMA model, the ACC has improved by 0.62, demonstrating a more significant performance enhancement. The performance of BERT-based and ERNIE-based models is significantly higher than that of neural network-based models, which confirms the significant improvement in performance of word embedding techniques based on pre-trained models such as BERT and ERNIE compared to Word2Vec.

Discussion

Comparison between BiLSTM and BiGRU

To verify the superiority of the BiLSTM layer used in our -MCBMA architecture compared to the BiGRU layer, a control experiment was conducted using the two layers, separately. Other parameters, such as dropout, learning rate, and epoch, were kept completely consistent. The comparison included the time required for every 200 iterations, the total training time, and the model's ACC and F1-score. Experiments were conducted separately on ERNIE and BERT, and the time consumption results are shown in Fig. 8; Tables 7 and 8.

From Fig. 7; Tables 7 and 8, whether based on the ERNIE model or the BERT model, using the BiLSTM layer significantly outperforms using the BiGRU layer in key metrics. In the application of the ERNIE model, the accuracy using the BiLSTM layer is 1.46% higher than that of the BiGRU layer, and the F1 score has increased by 1.62%. The -MCBMA architecture combines the temporal modeling capabilities of the BiLSTM layer with the self-attention mechanism and cross-layer feature fusion design, achieving deep extraction and effective integration of language features. This not only improves the model's training efficiency but also significantly enhances its performance in complex natural language processing tasks. Consequently, the superiority of the -MCBMA architecture has been thoroughly validated and demonstrated in both the ERNIE and BERT mainstream pre-trained language models.

Ablation study

To better understand the contributions of the various components to effectiveness, we conducted ablation experiments on our proposed model.

ERNIE The pooled layer data in the ERNIE model is directly used and input into the fully connected layer for sentiment analysis.

Multi-channel CNN We eliminate the BiLSTM-Multi-Attention layer and the cross-layer feature fusion. Instead, we directly input the data generated by ERNIE, after extracting local semantic information through multi-channel convolution, into the fully connected layer for classification.

Parameters	Values
Dimension of word embedding vector	300
Batchsize	128
Number of convolutional kernels	256
Epoch	20
Learning_rate	1e-3
Dropout	0.5

Table 5. Parameter settings of the benchmark models.

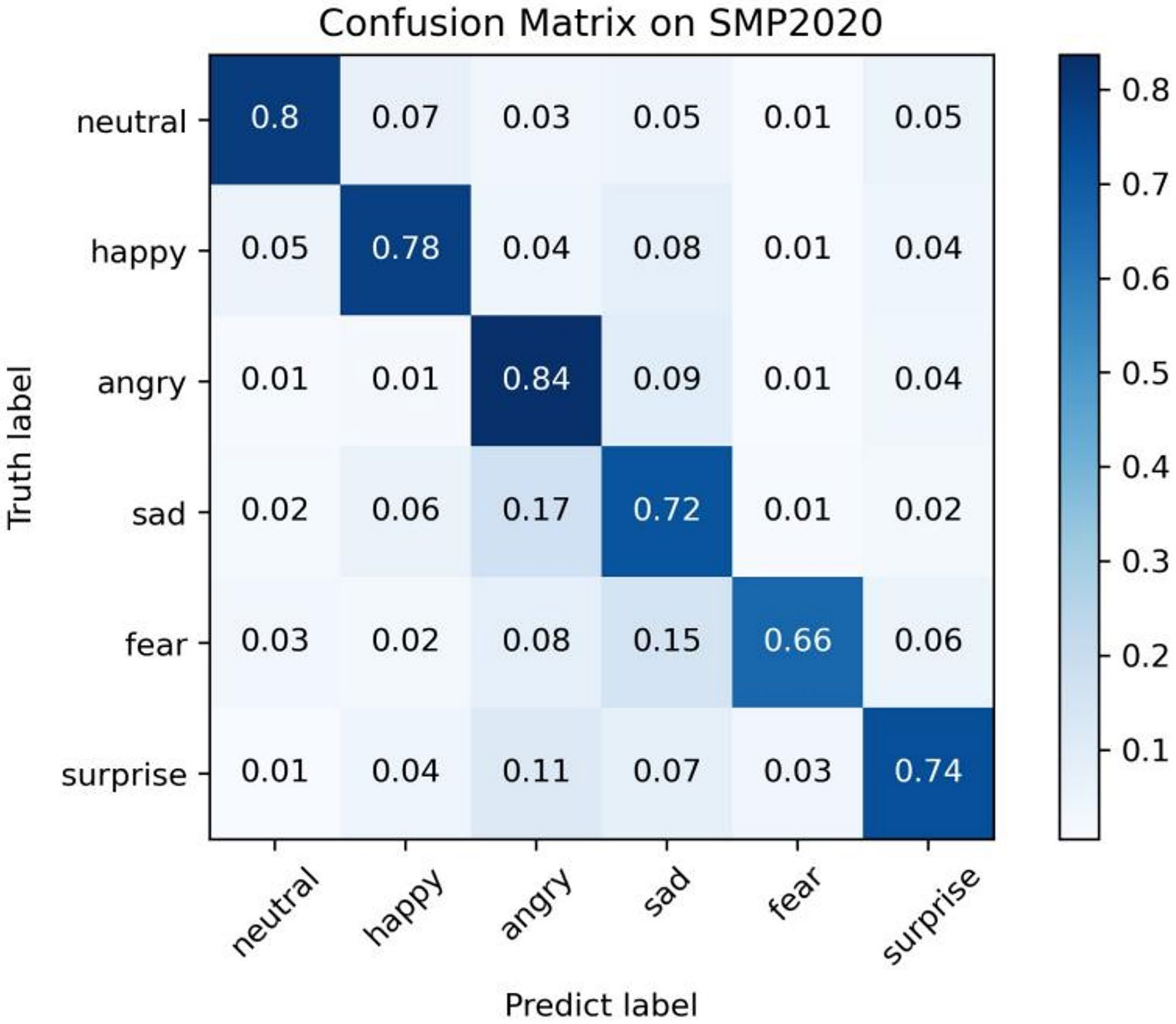

Fig. 6. Confusion matrix of our proposed model.

BiLSTM-Multi-Attention We exclude the multi-channel CNN layer and the cross-layer feature fusion, inputting the word embedding vectors generated by ERNIE directly into the BiLSTM-Multi-Attention layer. The extracted global semantic information is then fed into the fully connected layer for classification.

The results of the ablation experiment are shown in Table 9.

From Table 9, multi-channel CNN has an advantage in extracting local features of word vector through convolution, while BiLSTM-Multi-Attention performs well in extracting details of global features. The Pooled layer of the ERNIE not only captures global features more accurately through pooling but also decreases computational complexity by reducing the data dimensionality, resulting in slightly better performance. By combining local and global features, ERNIE-MCBMA achieves improvements of 0.62%, 0.78%, and 1.08%, respectively, compared to the other three scenarios in accuracy. This highlights the advantage of the ERNIE-MCBMA model in cross-layer feature fusion.

Additionally, to eliminate the advantage brought by the superior performance of the ERNIE model itself, we also explored the scenario of replacing ERNIE with Word2Vec. A comparison is conducted with a series of benchmark neural network models based on Word2Vec embeddings. The activation function for all comparison models was ReLU to ensure consistency and comparability in the experimental settings. Besides, the cross-layer feature fusion layer was not integrated with the pooled layer, while other architectural components remained unchanged. As shown in Table 10, even with Word2Vec as the word embedding layer, the -MCBMA architecture exhibited outstanding performance, achieving up to a 4.54% improvement in accuracy compared to other benchmark models.

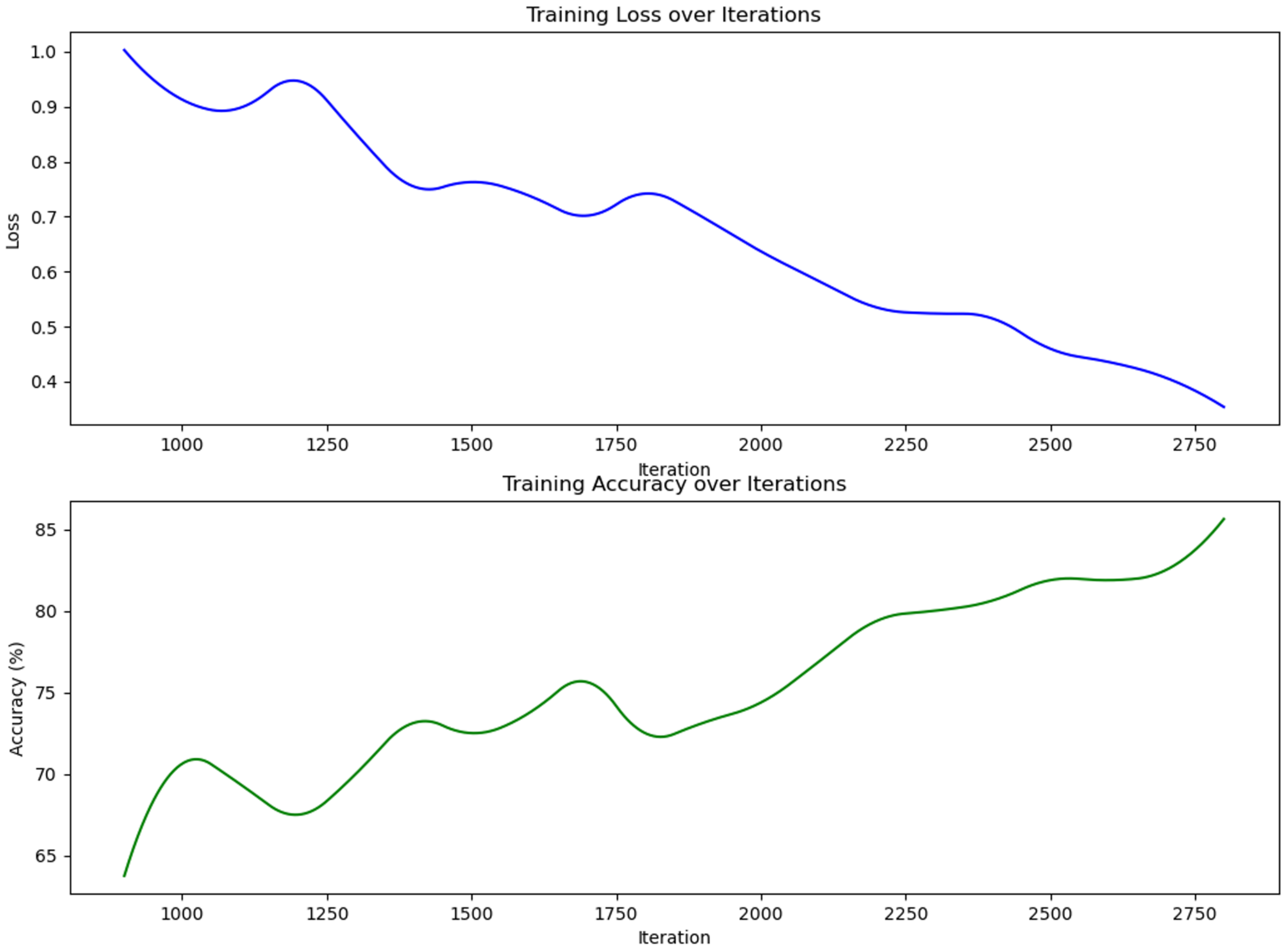

Fig. 7. Training dynamics: loss and accuracy trends.

Models	Acc (%)	Pre(%)	Recall(%)	F1 (%)
TextCNN	70.00	70.59	70.00	70.15
TextRNN	66.00	65.75	66.00	65.60
TextRCNN	69.26	69.09	69.26	68.77
TextRNN_Att	68.58	69.88	68.58	68.76
FastText	67.52	67.60	67.52	67.45
Transformer	66.78	67.80	66.78	67.16
BERT	74.02	75.4	74.02	74.46
BERT-TextCNN[53]	74.44	74.78	74.43	74.47
BERT-BiLSTM[54]	74.10	74.26	74.10	74.20
BERT-TextRCNN[55]	74.00	74.82	74.00	74.04
BERT-MCBMA (comparison)	**74.66**$_{+0.22}$	**75.42**	**74.66**	**74.77**$_{+0.30}$
ERNIE	77.64	78.49	77.64	77.91
ERNIE-TextCNN[56]	77.02	76.79	77.01	76.85
ERNIE-BiLSTM[57]	76.98	77.46	76.98	77.19
ERNIE-TextRCNN[58]	76.04	76.23	76.04	75.82
BBL(2024)[47]	76.51	-	-	72.54
BCAM(2024)[48]	77.40	74.90	74.0	74.3
ERNIE-MCBMA (ours)	**78.26**$_{+0.62}$	**78.95**	**78.26**	**78.45**$_{+0.54}$

Table 6. Sentiment classification results on SMP2020-EWECT.

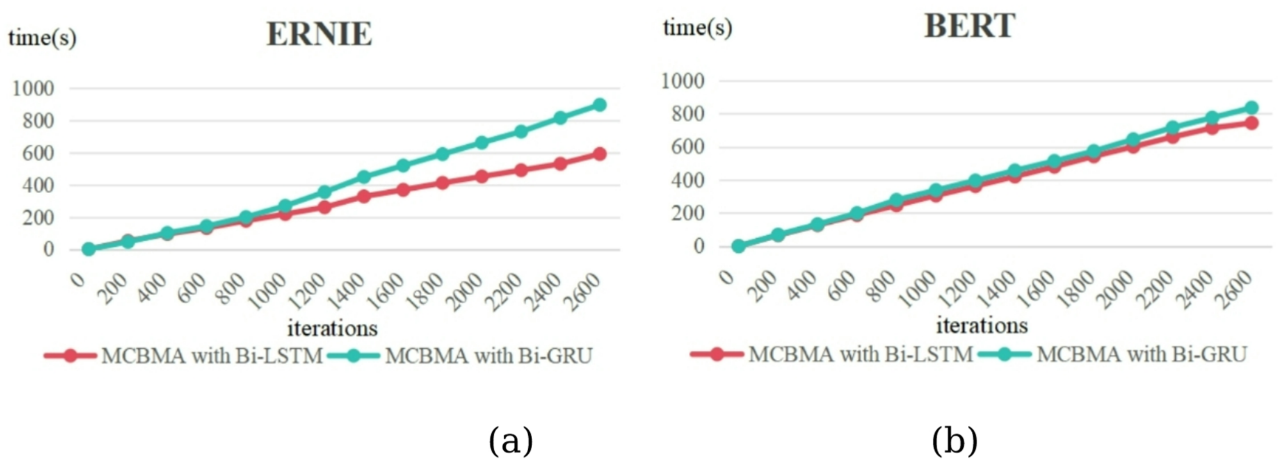

Fig. 8. Time consumption results. (**a**) The time consumption of ERNIE-based models; (**b**) the time consumption of BERT-based models.

Layers	Duration (s)	Acc (%)	F1 (%)
with BiLSTM	**669**	**78.26**	**78.45**
with BiGRU	1033	76.80	76.83

Table 7. Time consumption results for ERNIE-MCBMA.

Layers	Duration (s)	Acc (%)	F1 (%)
with BiLSTM	**773**	**74.66**	**74.77**
with BiGRU	866	74.6	74.23

Table 8. Time consumption results for BERT-MCBMA.

Models	Acc	Pre	Recall	F1
ERNIE	77.64	78.49	77.64	77.91
ERNIE with Multi-channel CNN	77.48	77.33	77.48	77.35
ERNIE with BiLSTM-Multi-Attention	77.18	78.34	77.18	77.50
ERNIE-MCBMA	**78.26**	**78.95**	**78.26**	**78.45**

Table 9. Ablation experiment results.

Models	Acc (%)	F1 (%)
TextCNN	70.00	**70.15**
TextRNN	66.00	65.60
TextRCNN	69.26	68.77
TextRNN_Att	68.58	68.76
FastText	67.52	67.45
Transformer	66.78	67.16
Word2Vec-MCBMA	**70.54**	69.74

Table 10. Results of Word2Vec-based models on SMP2020-EWECT dataset.

Conclusion

We propose a multi-channel CNN BiLSTM Multi Attention model based on ERNIE, which has significant differences from traditional text sentiment analysis models in terms of word embedding vector convolution and cross-layer feature fusion. In the strategy of word embedding vector convolution, different from the use of one-dimensional convolution in traditional text processing, we adopted 2D convolution techniques to capture the parallel local dependencies between words and the deep semantic association features.

Thereafter, we propose a three-stage progressive feature fusion mechanism:

(1) The primary feature enhancement layer: the bidirectional LSTM layer is directly stacked after the 2D convolution output to avoid the dimensionality reduction loss of local features caused by the pooling layer.
(2) Attention refining layer: Through the collaborative architecture of bidirectional LSTM and multi-head attention mechanism, context-sensitive feature recalibration is realized.
(3) Cross-layer feature fusion layer: fusion of shallow grammatical features and deep semantic representations to achieve fine-grained semantic information complementarity and enhancement.

Through this approach, we effectively integrate superficial detail information with profound semantic information, substantially enhancing the model's capacity for semantic comprehension. Ultimately, our model outperformed both the benchmark models and the other mainstream models in terms of accuracy and F1-score, substantiating its outstanding performance in sentiment analysis of texts.

However, there are some limitations and deficiencies in this study. First, to ensure comparability with the baseline models, we do not address the problem of data imbalance in the original dataset. Secondly, we have not yet established a multidimensional model to analyze and solve the diversity of emotional expression in social media. In the future, this research will focus on processing the original data set and exploring sentiment analysis models with multidimensional inputs.

This study provides a new perspective for more fine-grained Chinese text sentiment multi classification problems and can serve as a valuable reference.

Data availability

The data used in this article is from the SMP2020-EWECT public dataset (https://smp2020ewect.github.io).

Received: 17 July 2024; Accepted: 22 May 2025
Published online: 28 May 2025

References

1. Sahoo, C., Wankhade, M. & Singh, B. K. Sentiment analysis using deep learning techniques: A comprehensive review. *Int. J. Multimedia Inform. Retr.* **12**(2) (2023).
2. Birjali, M., Kasri, M. & Beni-hssane, A. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowl. Based Syst.* 226 (2021).
3. Yadav, A. & Vishwakarma, D. K. Sentiment analysis using deep learning architectures: a review. *Artif. Intell. Rev.* **53**(6), 4335–4385 (2020).
4. Gandhi, A. et al. Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. *Inform. Fusion*. **91**, 424–444 (2023).
5. Schuller, B., Mousa, A. E. & Vryniotis, V. Sentiment analysis and opinion mining: on optimal parameters and performances. *Wiley Interdisciplinary Reviews-Data Min. Knowl. Discovery*. **5**(5), 255–263 (2015).
6. Sánchez-rada, J. F. & Iglesias, C. A. Social context in sentiment analysis: formal definition, overview of current trends and framework for comparison. *Inform. Fusion*. **52**, 344–356 (2019).
7. Nistor, S. C. et al. Building a Twitter sentiment analysis system with recurrent neural networks. *Sensors* **21**(7) (2021).
8. Aljabri, M. et al. Sentiment analysis of Arabic tweets regarding distance learning in Saudi Arabia during the COVID-19 pandemic. *Sensors* **21**(16) (2021).
9. Yu, H. et al. LUX: smart mirror with sentiment analysis for mental comfort. *Sensors* **21**(9) (2021).
10. Dang, C. N., Moreno-García, M. N. & De La Prieta, F. An approach to integrating sentiment analysis into recommender systems. *Sensors* **21**(16) (2021).
11. Sarker, I. H. Machine learning: algorithms, Real-World applications and research directions. *SN Comput. Sci.* **2**(3), 160 (2021).
12. Das, R. & Singh, T. D. Multimodal sentiment analysis: A survey of methods, trends, and challenges. *ACM Comput. Surveys* **55**(13S), (2023).
13. Mercha, E. & Benbrahim, H. Machine learning and deep learning for sentiment analysis across Languages: A survey. *Neurocomputing* **531**, 195–216 (2023).
14. Ombabi, A. H., Ouarda, W. & Alimi, A. M. *Improving Arabic Sentiment Analysis across context-aware Attention Deep Model Based on Natural Language Processing* (Language Resources and Evaluation, 2024).
15. Zhu, W. L. et al. PHNN: A prompt and hybrid neural Network-Based model for Aspect-Based sentiment classification. *Electronics* **12**(19) (2023).
16. Wang, Z. X. & Lin, Z. P. Optimal feature selection for Learning-Based algorithms for sentiment classification. *Cogn. Comput.* **12**(1), 238–248 (2020).
17. Lango, M. Tackling the problem of class imbalance in multi-class sentiment classification: an experimental study. *Found. Comput. Decis. Sci.* **44**(2), 151–178 (2019).
18. Alzamzami, F., Hoda, M. & El Saddik, A. Light gradient boosting machine for general sentiment classification on short texts: a comparative evaluation. *IEEE Access*. **8**, 101840–101858 (2020).
19. Hota, S. & Pathak, S. KNN classifier based approach for multi-class sentiment analysis of Twitter data. *Int. J. Eng. Technol.* **7**(3), 1372–1375 (2018).
20. Liu, Y., Bi, J-W. & Fan, Z-P. Multi-class sentiment classification: the experimental comparisons of feature selection and machine learning algorithms. *Expert Syst. Appl.* **80**, 323–339 (2017).
21. Syaekhoni, M. A., Seo, S. H. & Kwon, Y. S. Development of deep learning models for multi-class sentiment analysis. *J. Inform. Technol. Serv.* **16**(4), 149–160 (2017).
22. Li, M. Y. et al. A PERT-BiLSTM-Att model for online public opinion text sentiment analysis. *Intell. Autom. Soft Comput.* **37**(2), 2387–2406 (2023).

23. Peiliang, Z. et al. Dynamic word vector based sentiment analysis for microblog short text. *PRML 2022, 3rd International Conference on Pattern Recognition and Machine Learning* (IEEE, 2022).
24. Subasic, P. & Huettner, A. Affect analysis of text using fuzzy semantic typing. *IEEE Trans. Fuzzy Syst.* **9**(4), 483–496 (2001).
25. Esuli, A., Sebastiani, F. & Sentiwordnet A publicly available lexical resource for opinion mining. *LREC Proceedings of the 5th international conference on language resources and evaluation* (European Language Resources Association (ELRA), 2006).
26. Khoo, C. S. & Johnkhan, S. B. Lexicon-based sentiment analysis: comparative evaluation of six sentiment lexicons. *J. Inform. Sci.* **44**(4), 491–511 (2018).
27. Lee, L-H., Li, J-H. & Yu, L-C. Chinese EmoBank: Building valence-arousal resources for dimensional sentiment analysis. *Trans. Asian Low-Resource Lang. Inform. Process.* **21**(4), 1–18 (2022).
28. Hu, M. & Liu, B. Mining and summarizing customer reviews. *KDD. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (Association for Computing Machinery, 2004).
29. Ding, X., Liu, B. & Yu, P. S. A holistic lexicon-based approach to opinion mining. *WSDM. Proceedings of the 2008 international conference on web search and data mining* (Association for Computing Machinery, 2008).
30. Huang, M. et al. Lexicon-based sentiment convolutional neural networks for online review analysis. *IEEE Trans. Affect. Comput.* **13**(3), 1337–1348 (2020).
31. Lin, W. & Liao, L-C. Lexicon-based prompt for financial dimensional sentiment analysis. *Expert Syst. Appl.* **244**, 122936 (2024).
32. Chikersal, P., Poria, S. & Cambria, E. SeNTU: sentiment analysis of tweets by combining a rule-based classifier with supervised learning. *SemEval. Proceedings of the 9th international workshop on semantic evaluation* (Association for Computational Linguistics, 2015).
33. Govindarajan, M. Sentiment analysis of movie reviews using hybrid method of Naive Bayes and genetic algorithm. *Int. J. Adv. Comput. Res.* **3**(4), 139 (2013).
34. Tarımer, İ., Çoban, A. & Kocaman, A. E. Sentiment analysis on IMDB movie comments and Twitter data by machine learning and vector space techniques. *ArXiv Preprint arXiv*: 190311983 (2019).
35. Li, F. F. et al. Chinese micro-blog sentiment classification through a novel hybrid learning model. *J. Cent. South. Univ.* **24**(10), 2322–2330 (2017).
36. Alayba, A. M. et al. A combined CNN and LSTM model for Arabic sentiment analysis. Machine Learning and Knowledge Extraction: Second IFIP TC 5, TC 8/WG 84, 89, TC 12/WG 129 International Cross-Domain Conference, CD-MAKE 2018, Hamburg, Germany, August 27–30, Proceedings 2, 2018 (Springer, 2018).
37. Sun, X. & Huo, X. Word-level and Pinyin-level based Chinese short text classification. *IEEE Access.* **10**, 125552–125563 (2022).
38. Kim, Y. *Convolutional Neural Networks for Sentence Classification* (Eprint Arxiv, 2014).
39. Li, J. & Sun, G. Y. R. S. C. O. E. W. R. Radical-Based sentiment classification of online education website reviews. *Comput. J.* **66**(12), 3000–3014 (2023).
40. Abyaad, R., Kabir, M. R. & Hasan, S. A novel approach to categorize news articles from headlines and short text. *IEEE Region 10 Symposium (TENSYMP)* (IEEE, 2020).
41. Lyu, C. et al. *Exploiting Rich Textual User-Product Context for Improving Personalized Sentiment Analysis* (Association for Computational Linguistics (ACL), 2023).
42. Kanwal, I. et al. Sentiment analysis using hybrid model of stacked Auto-Encoder-Based feature extraction and long short term Memory-Based classification approach. *IEEE Access.* **11**, 124181–124197 (2023).
43. Vaswani, A. et al. Attention is all you need. *Adv. Neural. Inf. Process. Syst.* 30 (2017).
44. Devlin, J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:181004805 (2018).
45. Jia, K. L. Sentiment classification of microblog: A framework based on BERT and CNN with attention mechanism. *Comput. Electr. Eng.* 101 (2022).
46. Deng, Y. C. et al. Toward transformer fusions for Chinese sentiment intensity prediction in Valence-Arousal dimensions. *Ieee Access.* **11**, 109974–109982 (2023).
47. Peng, S. & Cao, L. BERT and broad learning for textual emotion classification. *Textual Emotion Classification Using Deep Broad Learning* 31–46 (Springer, 2024).
48. Ding, L. et al. Enhancing sentiment analysis for Chinese texts using a BERT-based model with a custom attention mechanism. *International Conference on Web Information Systems and Applications* (Springer, 2024).
49. Sun, Y. et al. Ernie: enhanced representation through knowledge integration. *ArXiv Preprint arXiv*: 190409223 (2019).
50. Huang, W. D., Lin, M. & Wang, Y. Sentiment analysis of Chinese E-Commerce product reviews using ERNIE word embedding and attention mechanism. *Appl. Sciences-Basel* **12**(14) (2022).
51. Yang, Y., Dong, X. & Qiang, Y. A deep learning sentiment analysis method based on ERNIE and modified DPCNN. *Proceedings of the 2023 12th International Conference on Computing and Pattern Recognition* (2023).
52. Zhang, S. et al. Bidirectional long short-term memory networks for relation classification. *PACLIC. Proceedings of the 29th Pacific Asia conference on language, information and computation* (2015).
53. Zhang, J., Li, L. & Yu, B. Short Text Classification of Invoices Based on BERT-TextCNN. *ICAICT 2023, International Conference on Artificial Intelligence and Communication Technology* (Springer, 2023).
54. Cai, R. et al. Sentiment analysis about investors and consumers in energy market based on BERT-BiLSTM. *IEEE Access.* **8**, 171408–171415 (2020).
55. Lu, J. et al. Text classification for distribution substation inspection based on BERT-TextRCNN model. *Front. Energy Res.* **12**, 1411654 (2024).
56. Wang, M. & Xu, J. & Networks, N. Research on Chinese short text classification based on ERNIE-TEXTCNN model. *AANN 2023, Third International Conference on Advanced Algorithms and Neural Networks* (SPIE, 2023).
57. Hsieh, Y-H. & Zeng, X-P. Sentiment analysis: an ERNIE-BiLSTM approach to bullet screen comments. *Sensors* **22**(14), 5223 (2022).
58. Wang, Q. & Li, X. Chinese News Title Classification Model Based on ERNIE-TextRCNN. MLNLP. *Proceedings of the 2022 5th International Conference on Machine Learning and Natural Language Processing* (Association for Computing Machinery, 2022).

Acknowledgements

This study was supported by the Basic Science (Natural Science) of Higher Education Institutions in Jiangsu Province (Grant Number: 21KJB510026), by the Scientific Research Foundation for Excellent Talents of Xuzhou Medical University (Grant Number: D2020048), by the Science Foundation of Jiangsu Normal University (Grant Number: 20XSRX016), by Jiangsu Province College Students' Innovation and Entrepreneurship Training Program (Grant Number: 202410313032Z).

Author contributions

B.S. and Y.X. formulated the overall research strategy and guided the work. Y.S. primarily engages in model building and algorithm implementation. Z.Y. is responsible for data collection and data processing. Y.S. prepared

all the figures and tables. All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.X. or B.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.