International Journal of Information Technologies and Systems Approach
Volume 17 • Issue 1 • January-December 2024

A Fine-Grained Sentiment Analysis Method Using Transformer for Weibo Comment Text

Piao Xue
School of Marxism Studies, Xi'an Mingde Institute of Technology, China

Wei Bai
School of Mathematics and Computer Science, Ningxia Normal University, China

ABSTRACT

Many existing fine-grained sentiment analysis (FGSA) methods have problems such as easy loss of fine-grained information, difficulty in solving polysemy and imbalanced sample categories. Therefore, a Transformer based FGSA method for Weibo comment text is proposed. Firstly, the RoBERTa model with knowledge augmentation was used to dynamically encode the text so as to solving the polysemy issue. Then, BiLSTM is used to effectively capture bidirectional global semantic dependency features. Next, Transformer is used to fuse multi-dimensional features and adaptively strengthen key features to overcome the problem of fine-grained information loss. Finally, an improved Focal Loss function is utilized for training to solve the issue of imbalanced sample categories. As demonstrated by the experimental outcomes on the SMP2020-EWECT, NLPCC 2013 Task 2, NLPCC 2014 Task 1, and weibo_senti_100k datasets, the suggested method outperforms the alternatives for advanced comparison methods.

KEYWORDS

Fine-Grained Sentiment Analysis, Improved Focal Loss, RoBERTa, Semantic Feature Fusion, Transformer

INTRODUCTION

The digital age has democratized access to information, requiring the business and academic sectors to develop new methods to analyze and process this information. In addition, the tools and software for processing information are constantly being improved and updated (Gandhi et al., 2023; Li et al., 2022; Liu et al., 2022; Zeng et al., 2023). Structured data analysis alone is not enough to meet people's needs. Unstructured text data, like public opinion expressed through social media and surveys, is gaining importance as a valuable information source. For public stability, government departments should actively monitor public opinion, understand citizen sentiment, and keep them informed through transparent information flow (Fan, 2023). Public opinion control is not only a matter for relevant staff but also one that concerns the vital interests of all citizens (Djaballah et al., 2021; Li et al., 2021; Tamil et al., 2022; Zhang & Cui, 2023; Zhang et al., 2020).

It is common to use English texts as research objects in unstructured data analysis (Xiao et al., 2022; Yu et al., 2021; Zhang et al., 2022; Y. Zhang et al., 2023). However, when dealing with Chinese unstructured data, many problems remain, such as difficulty in sentence segmentation due to the absence of spaces between words, distinguishing words or phrases that may have different meanings in different contexts, and determining the impact of adverbs, conjunctions, and negations on sentence meaning (Hazarika et al., 2020; Kaur & Kautish, 2022; Li et al., 2020; Zhang et al., 2022).

DOI: 10.4018/IJITSA.345397

New media platforms like Weibo have become essential for capturing and shaping public opinion. These platforms provide a breeding ground for online discussions, where public sentiment surrounding current events can simmer, spread, and erupt (Liao et al., 2022; Lin et al., 2020; Liu, 2023; Poria et al., 2023). Netizens comment on events in the news through Weibo and express, disseminate, and interact with their emotions, thereby forming public opinion on these events (Fang et al., 2023; Yu et al., 2021; Zhu et al., 2022). However, information asymmetry also fuels the spread of online rumors, which can significantly impact public sentiment. Such effects typically last for an extended period, are difficult to control, and may bring about a certain amount of social instability. Therefore, it is necessary to conduct an effective sentiment analysis for Weibo comments (Jiang et al., 2023; Li et al., 2019; Lu et al., 2021). However, traditional sentiment analysis methods usually classify Weibo comment texts into three categories: positive, negative, or neutral. These coarse-grained sentiment analysis methods cannot capture the rich emotional information in Weibo texts. They can lead to confusion. Fine-grained sentiment analysis (FGSA) goes beyond simply classifying emotions as positive or negative. Depending on the requirements, it can identify more nuanced emotions, such as pleasure, anger, hatred, or sadness. This allows for a more fine-grained understanding of the emotional tendencies of Weibo texts (Huang et al., 2019).

Nonetheless, current FGSA methods are plagued by significant flaws, including the frequent loss of fine-grained data, the resolution of polysemy issues, and unbalanced sample categories. Therefore, this paper proposes a novel Transformer-based FGSA method for Weibo comment text. Its primary innovation lies in the following:

1) At the encoding layer, the RoBERTa model is employed to encode the text dynamically, and a representation dictionary is constructed to enhance knowledge, effectively solving the problem of polysemy. In the semantic representation layer, the bidirectional long short-term memory (BiLSTM) network is used to extract text semantic information in both forward and backward directions, better capturing bidirectional global semantic dependency features.
2) Transformer is used to fuse multi-dimensional features and adaptively strengthen key features, overcoming the problem of losing fine-grained information in traditional methods. By adopting a combination pooling operation that focuses on both global and local features, a richer emotional feature can be obtained.
3) During the model training process, an improved focal loss function is used to solve the class imbalance problem of sample labels.

RELATED WORK

The existing methods for text sentiment classification tasks can be divided into three categories. The first is based on manually constructed sentiment dictionaries for text sentiment classification. Although this approach is useful in classification, it can produce inaccurate classification results if some emotional words are not included, especially new words such as popular online phrases. The second category of methods is based on traditional machine learning. These methods focus more on the contextual semantics and context of the text. Although this type of method can overcome the limitations of sentiment dictionaries, machine learning algorithms struggle when dealing with extremely high volumes of data. The third category includes deep learning-based text sentiment classification methods, which achieve high model judgment accuracy via extensive dataset training. A deep learning-based approach can process massive amounts of data information, overcoming the shortcomings of machine learning in dealing with such vast quantities of data. Moreover, the neural network structure has excellent learning ability, and its application in text sentiment classification tasks can play a significant role.

Text Sentiment Analysis

Turney (2020) proposed an unsupervised sentiment analysis method based on text information classification using the contribution value of words in the vocabulary to determine the sentiment propensity of the text. However, this method and these rules are not closely compatible. Chen et al. (2020) applied the k-nearest neighbor algorithm to classify Lao language texts and proposed a corresponding sentiment analysis model for Lao language text information. This approach offers a valuable model for Lao language sentiment analysis, but further research is needed to explore its generalizability to other languages. Jiang et al. (2019) proposed an LSTM-CNN model based on the attention mechanism (AttM). The model used different convolutional layers to capture the dependency relationships between various features and added additional AttM layers to reduce the dimensionality. The utilization of these AttMs allowed the LSTM-CNN to extract deeper features. However, this method can only perform simple classification and is unsuitable for FGSA. Xu et al. (2020) proposed a CNN Text Word2vec model, which first trains text with word vectors. They investigated the effect of various semantic units on the model's precision and experimentally validated its superior performance. Nonetheless, this approach fails to efficiently integrate numerous distinct feature extraction techniques for feature fusion, thereby failing to fully exploit the model's optimal performance.

FGSA of Text

For the FGSA of text, Hao et al. (2022) proposed a joint model that effectively combines FGSA and text tasks. However, the model structure of this method is intricate and demands a substantial quantity of computational effort. Theodoropoulos & Alexandris (2022) designed a program for FGSA that classifies online comment text into "positive," "negative," or "neutral" categories. This program achieves FGSA to a certain extent through human–computer interaction. However, the FGSA results of this method only include three types, and the fine-grained size is insufficient. In response to the issue of unclear or often overlooked emotional trends in online comment texts, W. Z. Liao et al. (2021) proposed an FGSA model for texts utilizing syntactic rule matching and deep semantic networks. However, the convergence speed of this method is relatively slow. To address the challenge of limited and incomplete meaning expression during text interpretation, Yin et al. (2021) proposed a new text information interpretation model, DLER, based on dual learning as the basic framework. However, this method cannot achieve accurate FGSA of text information between different fields. Y. R. Zhang et al. (2023) proposed a cross-domain text information FGSA model, SKEP_Gram-CDNN, based on capsule networks to address the issue of significant feature differences between text information in multiple domains, which leads to insufficient or incorrect feature extraction of different information. However, this model may result in the loss of semantic information between words for complex sentences, affecting the final classification effect. Zhou et al. (2022) proposed a text FGSA model, BERT ftfl-SA, based on AttM and a support vector machine to address the issues of inaccurate classification and insufficient fine-grained granularity in some FGSA methods. However, this method struggles with universality in text sentiment classification in different languages.

Pre-Trained Language Model

Regarding pre-training language models, Liu et al. (2020) proposed an aspect-based FGSA method, which pre-trains RoBERTa and combines long short-term memory (LSTM) and AttM to achieve a detailed classification of different types of emotions. Although this method can automatically determine whether the text content is positive, it cannot accurately determine the emotional orientation of the text. Joshy and Sundar (2022) compared and analyzed several FGSA methods based on data mining and deep learning and found the BERT model to be better than the others. However, no innovative proposal for a new model for FGSA has been made. Taking the war between Russia and Ukraine as the research object, Sirisha and Chandana (2022) proposed a new method for FGSA based on deep learning, combining the RoBERTa model, aspect-based sentiment analysis, and LSTM. However, this method cannot efficiently map text word vectors, making it difficult to improve the accuracy of contextual information. To address the inaccurate classification of emotional polarity in the FGSA process, W. X. Liao et al. (2021) proposed a new FGSA model based on a deep bidirectional transformer that combined the pre-trained RoBERTa model and aspect categories. However, this model may cause overfitting for particular texts due to too many parameters.

The above analysis of existing FGSA methods highlights that they are prone to defects in fine-grained information, multiple meanings of one word, and imbalanced sample categories. Therefore, this paper proposes a new FGSA method based on the pre-trained RoBERTa model for Weibo comment text. To solve different problems, we adopted corresponding strategies in the coding, semantic feature extraction, semantic feature fusion, and output layers. We enhanced the model architecture at the coding layer to prevent loss of fine-grained information, while the output layer incorporates techniques to address class imbalance issues. The problem of multiple meanings of a word is solved by introducing deeper semantic understanding. We introduced a semantic feature fusion layer to better integrate semantic features extracted from different levels. This integrated strategy aims to improve the FGSA performance of Weibo comment texts.

PROPOSED FGSA MODEL

The proposed FGSA model comprises four parts: a RoBERTa encoding layer, a semantic feature extraction layer, a semantic feature fusion layer, and an output layer. Figure 1 illustrates the structure of the proposed model.

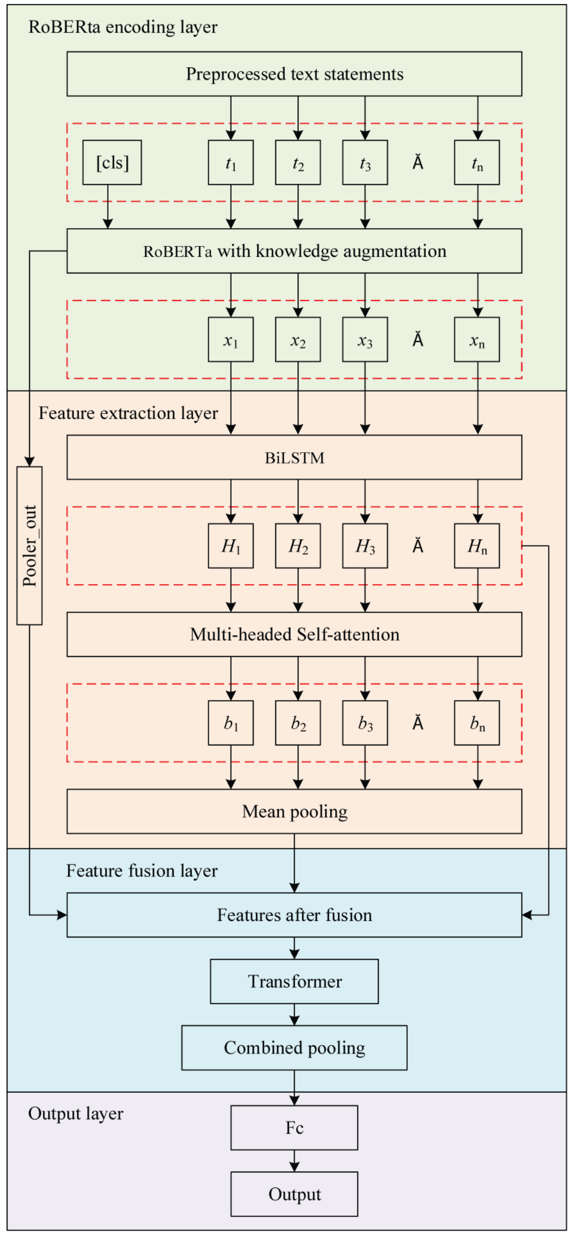
The function of each layer in the proposed model is as follows:

(1) RoBERTa encoding layer. This layer employs RoBERTa for dynamic text encoding and circumvents the insurmountable obstacle of polysemy faced by conventional encoding techniques. Consequently, the layer acquires the intricate grammatical and semantic characteristics of Weibo remark text.
(2) Semantic feature extraction layer. This layer adopts a combination of local and global features for extraction, obtaining rich semantic information from multiple dimensions. The BiLSTM network is utilized for global feature extraction. The network encodes text semantic information in both directions, capturing bidirectional global semantic dependencies. The capsule network model is used for local feature extraction. This network replaces the scalar output in traditional convolutional operations with vector output to better focus on the semantic information of text position.
(3) Semantic feature fusion layer. This layer adopts a multi-head self-AttM approach to fuse features from different dimensions. This approach allows the model to establish connections with multi-dimensional semantic information features, adaptively filter key features, and prevent information loss, a common problem of traditional fusion methods.

The proposed FGSA model conducts the emotional analysis process as follows:

(1) The preprocessed text is converted into a Token format that RoBERTa can accept.
(2) The RoBERTa encoding layer outputs the sentence embedding of the entire sentence in Pooler_Out and the embedding of each word.
(3) The embedding words are fed into the semantic feature extraction layer for local and global semantic extraction. The final output is obtained through the BiLSTM network, and the global input is obtained through multi-head self-AttM fusion. The local features of the statement are obtained through the capsule network.
(4) Multi-head self-AttM fusion and average pooling are performed on the obtained sentence features, global features of characters, and local features of characters to obtain fusion features.

Figure 1. Structure of the Proposed FGSA Model

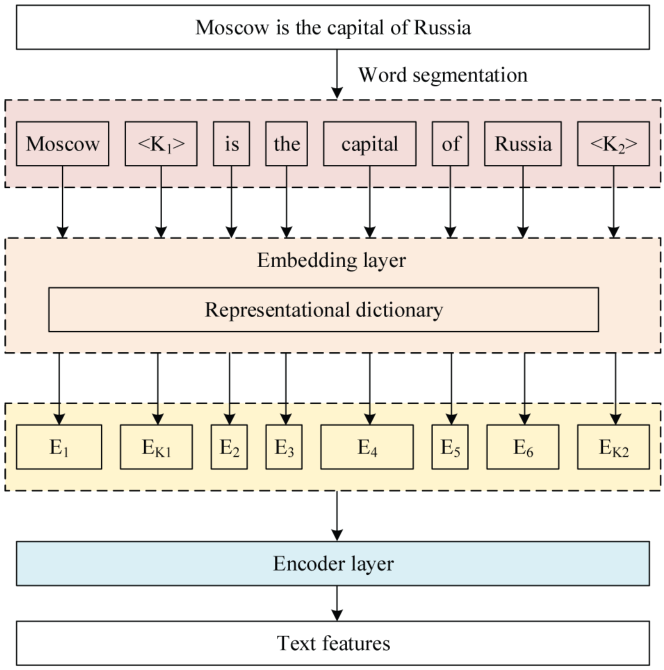


(5) The fused features pass through the Softmax output layer to obtain the final emotion class.

Encoding Layer Based on Knowledge Enhancement

The text features are extracted from Weibo comment texts using the RoBERTa model, while knowledge enhancement is accomplished by consulting representation dictionaries. The RoBERTa model comprises two primary components: the encoder layer and the embedding layer. To obtain text features, the embedding layer transforms words into numerical representations suitable for embedding

Figure 2. Structure of Text Feature Extraction Module



before passing them to the encoder layer for encoding. The process of knowledge enhancement occurs within the embedding layer. The RoBERTa model is described below.

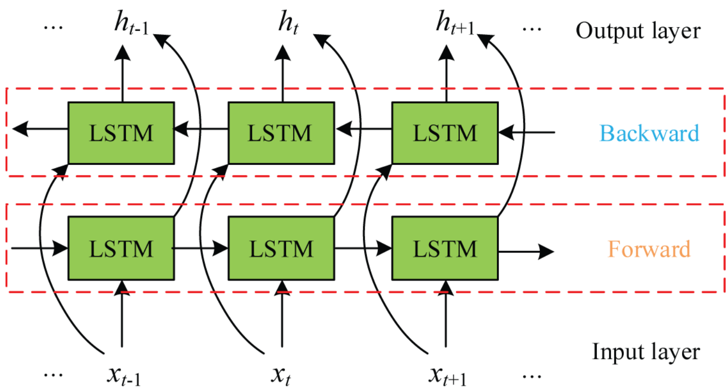
After the text in Figure 2 is preprocessed and irrelevant special characters like URLs are eliminated, word segmentation is executed. Word embedding is achieved by converting the segmented text via the embedding layer. The RoBERTa model utilizes the mask language task to acquire the word embedding representations of sentences.

The primary objective of this coding layer is to augment the model's semantic comprehension of knowledge phrases present in social media texts, such as named entities, English abbreviations, and neologisms. A lexicon of representations of knowledge phrases is created to enhance the model's comprehension of knowledge phrases. This lexicon is then used to replace knowledge phrases within sentences with their corresponding word embeddings, acting as background knowledge. The word embeddings are obtained through a masked language model, which is similar to a completion task, for predicting masked words and generating their word embeddings. This process uses the masked language model with whole-word masking. To generate representations of knowledge phrases, it is necessary to identify knowledge phrases in the corpus of the social media domain. This is achieved by using the named entity tool to pinpoint named entities and creating a custom lexicon or leveraging existing ones to identify neologisms and English shorthand. By augmenting the model's semantic comprehension of particular expressions in social media texts with preexisting knowledge and word embeddings, this strategy combines domain-specific corpora and lexicons.

Semantic Feature Extraction Layer

This layer refines the semantics of the word embeddings extracted by the RoBERTa layer. This model uses LSTM for semantic feature extraction. LSTM is a special form of recurrent neural network (RNN) characterized by the design of cell state and gating mechanisms. LSTM effectively solves the gradient vanishing problem in RNN training and can learn long-distance dependencies

Figure 3. Network Structure of BiLSTM in the Semantic Feature Extraction Layer



of text semantics. However, the LSTM model cannot encode text information from back to front. The BiLSTM model comprises two LSTM models oriented in opposite directions. In this instance, contextual text features are generated by encoding the contextual information of the text using the BiLSTM network:

$$\begin{cases} \vec{h}_t = LSTM\left(x_t, \vec{h}_{t-1}\right) \\ \overleftarrow{h}_t = LSTM\left(x_t, \overleftarrow{h}_{t-1}\right) \end{cases} \tag{1}$$

where x_t represents the current input time and $\vec{h}_{t-1}$ and $\overleftarrow{h}_{t-1}$ denote the forward and hidden states of the previous time, respectively. In BiLSTM, the hidden state H_t at a particular moment is connected by the hidden state $\vec{h}_t$ of the forward LSTM and the hidden state $\overleftarrow{h}_t$ of the reverse LSTM:

$$H_t = \left[\vec{h}_t, \overleftarrow{h}_t\right] \tag{2}$$

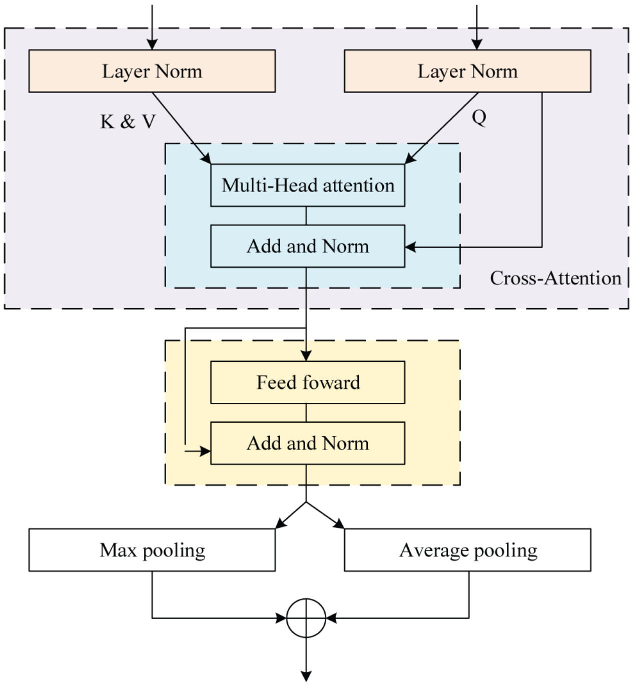
To avoid semantic loss, the output H_n of the last step of BiLSTM is taken out as the first part of the global semantic feature. Then, multi-head self-AttM is used to allocate the weight of each hidden state of BiLSTM, obtaining $B = [b_1, b_2, b_3, \ldots, b_n]$. Next, the global semantic feature representation M is obtained through an average pooling operation.

BiLSTM is a semantic feature extractor that efficiently captures semantic information in the input sequence. These semantic features are usually utilized in subsequent levels for higher-level contextual understanding and feature fusion in structures such as Transformer. Figure 3 illustrates the network structure of BiLSTM in the semantic feature extraction layer.

Semantic Feature Fusion Layer

Traditional approaches may use a simple feature splicing strategy to directly connect features of different dimensions. This approach may be too coarse when dealing with complex mapping relationships. Moreover, capturing non-linear relationships and long-distance dependencies between features is difficult. To overcome the limitations of the traditional approach, we implemented the Transformer encoder. Transformer is a model with a self-attention mechanism capable of modeling global dependencies in sequential data and is suitable for the fusion of multi-dimensional features. A Transformer encoder is used to fuse features of different dimensions. The self-attention mechanism

Figure 4. Multi-Dimensional Feature Fusion Based on Transformer



allows the Transformer to dynamically assess the importance of each feature in relation to every other feature. This enables the Transformer to capture complex mapping relationships between features more efficiently. Introducing the Transformer encoder allows the model to learn the mapping relationships between features of different dimensions more flexibly without the need to perform simple splicing of features beforehand. This helps the model gain a deeper understanding of feature relationships, improving the ability to extract fine-grained information and enhancing task performance. The subsequent architecture employs a Transformer encoder to combine characteristics of various dimensions.

In Figure 4, the low-level sentence embedding representation $P = pooler\_out$ is first obtained, and then local and global features are extracted for word embedding. The global features M and H_n are obtained through the global feature extraction layer. Multi-head self-AttM empowers the model to concentrate on crucial attributes while autonomously disregarding the encoded extraneous information. Using multi-head self-AttM to fuse the high-level semantic information matrices P, M, and H_n, information can be transmitted between the relatively isolated row vectors in the matrix, establishing connections, obtaining global dependencies, and improving the model's ability to understand Weibo text. Firstly, concatenate the obtained features of various granularity to synthesize the semantic vector Z:

$$Z = \left[P, M, H_n\right]^{\mathrm{T}} \tag{3}$$

The multi-dimensional embedding $ME \in R^{(n_t+n_i)\times d_t}$ is passed into the Transformer encoder. Firstly, ω_q, ω_k, and ω_v are used to map the multi-dimensional inputs into vectors Q, K, and V, respectively. Then, the AttM score between each token in the multi-dimensional feature sequence and other tokens is calculated using Q and K. Finally, the output is obtained by multiplying the AttM weight by V:

$$\begin{cases} Q^n = \omega_q \cdot F^{n-1} \\ K^n = \omega_k \cdot F^{n-1} \\ V^n = \omega_v \cdot F^{n-1} \end{cases} \tag{4}$$

$$F^n = softmax\left(\frac{Q^n (K^n)^T}{\sqrt{d_t}}\right) V^n \tag{5}$$

In the above equations, Q^n, K^n, and V^n represent the Q, K, and V matrices used by the nth Transformer encoder for self-AttM operations, respectively, F^n represents the output of the n^{th} Transformer encoder, $F^0 = ME$, $n \in [1, N]$, and N is a hyperparameter representing the number of encoders. The fused multi-dimensional feature sequence is the output $F^N = \left\{f_1, f_2, ..., f_{n_t} + f_{n_i}\right\}$ of the last layer Transformer encoder, $F^N \in R^{(n_t+n_i)\times d_t}$.

In multi-dimensional embedding, each token performs self-AttM calculations with other tokens. This allows the model to simultaneously capture both the internal effects of single-dimensional features and the mapping relationships between multi-dimensional features. By dynamically adjusting the weights assigned by AttM to each dimension, the model can more effectively allocate attention and learn meaningful representations. Due to the advantages of pooling operations, such as noise suppression, reduced information redundancy, simplified model computation, and overfitting prevention, we selected three pooling operations: combination pooling, maximum pooling, and average pooling. Combination pooling yields a more comprehensive feature layer. Maximum pooling captures local features at each instant, and average pooling directs the model's attention toward global features. The outcomes of the maximum and average pooling operations are combined into a single output for the AttM module. The calculation process is outlined below:

$$\begin{cases} F_{k_{\max}}^{L\to A} = maxpooling\left(F_k^{L\to A}\right) \\ F_{k_{avg}}^{L\to A} = averagepooling\left(F_k^{L\to A}\right) \\ F_k^{L\to A} = Concat\left(F_{k_{\max}}^{L\to A}, F_{k_{avg}}^{L\to A}\right) \end{cases} \tag{6}$$

The fused feature vectors cannot be directly used for sentiment classification. Therefore, a simple AttM is used to obtain the final feature representation for classification as follows:

$$a_i' = ReLU\left(f_i \omega_1 + b_1\right) \omega_2 + b_2 \tag{7}$$

$$a_i = exp\left(a_i' / \sum_{k=1}^{n_t+n_i} a_k'\right) \tag{8}$$

$$Z = \sum_{i=1}^{n_t+n_i} a_i f_i \tag{9}$$

where $ReLU$ denotes the activation function, ω_1 and ω_2 denote the weight matrices, and b_1 and b_2 denote the bias matrices.

Sentiment Classification

The above feature vector is input into the fully connected layer, and the Softmax classifier is used to complete the classification. Following traversal by the classification layer, the likelihood of emotion category output is computed as follows:

$$Z' = ReLU\left(Z\omega_3 + b_3\right) \tag{10}$$

$$P = \frac{e^{Z_i}}{\sum_{k=1}^{n} e^{Z_k}} \tag{11}$$

In the above equations, ω_3 and b_3 represent the weight matrix and bias matrix of the linear transformation, respectively. Z_i is the vector of the multidimensional representation of the i^{th} sample used for classification after passing through the fully connected layer, and P represents a vector composed of probabilities that the i^{th} sentence belongs to each emotion category.

Model Training

Traditional approaches for Weibo sentiment fine-grained classification often use mean square error or standard cross entropy as loss functions. However, these methods suffer from two limitations. First, the learning rate of the model is very slow at the beginning of training. Second, standard cross-entropy can be skewed by imbalanced classes, where the loss function is easily dominated by a redundant category, affecting the model's effectiveness. Therefore, we adopted an improved focal loss function to solve the problem of imbalanced sample categories in fine-grained classification of Weibo statements. During training, the initial term of the function was the original focal loss function (Hao et al., 2022). This function decreases the impact of samples readily classified by the model. This allows the model to prioritize learning from more challenging samples, ultimately improving its overall performance. To avoid problems such as vanishing gradients and exploding gradients during the training process, we introduced orthogonal constraints after training the function and added a regularization term to the second term of the function:

$$\begin{aligned} FL = & -\gamma(c_i)\left[1 - q(c_i)\right]^{\beta}\log\left[q(c_i)\right] \\ & + \eta\sum_{i}\left|\delta_i^{\,T}\delta_i - I\right|^2 \end{aligned} \tag{12}$$

In the formula, the term $\gamma(c_i)$ represents the weight of the c_i class sample, and $q(c_i)$ represents the probability distribution of the predicted class c_i. The term η represents the penalty coefficient, I represents the identity matrix, and δ_i represents the i^{th} type weight matrix.

EXPERIMENT AND ANALYSIS

Datasets

We used four different datasets for the FGSA of Weibo statements.

(1) SMP2020-EWECT Dataset (Poria et al., 2023). This dataset is divided into the "Epidemic" and "General" datasets. It contains 16,000 pieces of data with six emotional labels: Happy, Angry, Sad, Fear, Surprised, and Neutral.

(2) NLPCC2013 Task 2 Weibo Sentence Emotion Dataset (Zhu et al., 2022). This dataset uses eight emotional labels: None, Sadness, Like, Anger, Happiness, Disgust, Fear, and Surprise. It contains a total of 14,000 pieces of data.
(3) NLPCC2014 Task 1 Emotion Dataset (Fang et al., 2023). This dataset uses the same eight emotional labels as the NLPCC2013 Task 2 Weibo Sentence Emotion Dataset. It contains a total of 20,000 pieces of data.
(4) weibo_senti_100k Dataset (Jiang et al., 2019). This dataset contains a total of 100,000 Weibo texts. It provides a balanced split of 50,000 positive and 50,000 negative emotional texts.

We divided the four datasets following a 6:2:2 split. In this scheme, 60% of the combined data was allocated to the training set, while each remaining 20% was allocated to the validation and testing sets. The statistical information about the four datasets is presented in Table 1.

Experimental Setup

The parameter settings of the model in the experiment are presented in Table 2.

The specific experimental environment information is provided in Table 3.

To evaluate the experimental results, we employed accuracy (ACC) and the F1 value as evaluation indicators for the experiment (Hao et al., 2022).

Comparison With State-of-the-Art Methods

To verify the effectiveness of the proposed Transformer-based Weibo comment text FGSA method, we compared it with several models, including the LSTM-CNN model (Jiang et al., 2019), the CNN-Text-Word2vec model (Xu et al., 2020), the DLER model (Yin et al., 2021), the SKEP_Gram-CDNN model (Y. R. Zhang et al., 2023), and the BERT-ftfl-SA model (Zhou et al., 2022). We evaluated these models on four datasets: SMP2020-EWECT, NLPCC 2013 Task 2, NLPCC 2014 Task 1, and weibo_ senti_100k.

Figure 5 and Table 4 show the results of various models on the SMP2020-EWECT dataset recorded under identical experimental conditions.

Table 4 shows that the proposed model achieves significant improvements in ACC and F1 score on the SMP2020-EWECT dataset compared to the state-of-the-art models. Compared to the highest-performing SKEP_Gram-CDNN model, the proposed model achieves an improvement of 0.58% in ACC and 0.56% in F1 score. The proposed model also achieves improvements of 1.41% in ACC and 1.35% in F1 score over the LSTM-CNN model, 1.49% in ACC and 1.43% in F1 score over the DLER model, 2.15% in ACC and 2.07% in F1 score over the CNN-Text-Word2vec model, and 2.89% in ACC and 2.79% in F1 score over the BERT-ftfl-SA model.

Figure 6 and Table 5 illustrate the results of various models utilizing the NLPCC 2013 Task 2 dataset. All results were recorded under identical experimental conditions.

Table 5 shows that the proposed model achieves considerable improvements on the NLPCC 2013 Task 2 dataset. Compared to the best-performing SKEP_Gram-CDNN model, the proposed model achieves an increase of 0.48% in ACC and 0.38% in F1 score. Similarly, the proposed model achieves improvements of 1.18% in ACC and 0.92% in F1 score over the LSTM-CNN model, 1.25% in ACC and 0.97% in F1 score over the DLER model, 1.8% in ACC and 1.4% in F1 score over the CNN-Text-Word2vec model; and 2.42% in ACC and 1.89% in F1 score over the BERT-ftfl-SA model.

Figure 7 and Table 6 illustrate the results generated by the various models utilizing the NLPCC 2014 Task 1 dataset under identical experimental conditions.

Table 6 compares the performance of the proposed model with other models on the NLPCC 2014 Task 1 dataset. Compared to the top-performing SKEP_Gram-CDNN model, the proposed model achieves improvements of 0.52% in ACC and 0.42% in F1 score. Similar improvements are observed against LSTM-CNN (1.25% ACC, 1.02% F1), DLER (1.33% ACC, 1.08% F1), and BERT-ftfl-SA (1.91% ACC, 1.57% F1).

Table 1. Statistical Information for the Dataset

		Training Set	Verification Set	Test Set
SMP2020-EWECT	Happy	1737	579	579
	Angry	1420	473	473
	Sad	1736	579	579
	Fear	1790	597	597
	Surprise	1543	514	514
	Neutral	1373	458	458
NLPCC 2013 Task 2	None	1083	361	361
	Sadness	1047	349	349
	Like	1099	366	366
	Anger	1304	435	435
	Happiness	892	297	297
	Disgust	1093	364	364
	Fear	1004	335	335
	Surprise	877	292	292
NLPCC 2014 Task 1	None	1559	520	520
	Sadness	1532	511	511
	Like	1435	478	478
	Anger	1729	576	576
	Happiness	1469	490	490
	Disgust	1592	531	531
	Fear	1435	478	478
	Surprise	1249	416	833
weibo_senti_100k	None	3797	1266	1266
	Sadness	3686	1229	1229
	Like	4019	1340	1340
	Anger	3614	1205	1205
	Happiness	3669	1223	1223
	Disgust	3771	1257	1257
	Fear	3764	1255	1255
	Surprise	3680	1227	1227

The results generated by various models under identical experimental conditions utilizing Weibo_Senti_100k are displayed in Figure 8 and Table 7.

The findings presented in this paper indicate that the proposed model achieves significant improvements on the weibo_senti_100k dataset. Compared to the highest-performing SKEP_Gram-CDNN model, the proposed model demonstrates an improvement of 0.69% for both ACC and F1 scores. Similar improvements are observed against LSTM-CNN (1.68% in both metrics), DLER (1.78% in both metrics), CNN-Text-Word2vec (2.57% in both metrics), and BERT-ftfl-SA (3.73% in ACC and 3.46% in F1).

Table 2. Model Parameters

	Parameters	Value
RoBERTa	Learning rate	0.00003
	Interlayer learning coefficient	0.95
	Weight attenuation	0.00001
	Word vector dimension	768
	Maximum sentence length	175
Multi-headed Attention	Number of heads	4
Fc	Dropout	0.4
Transformer	Number of Transformer layers	4
Training	Optimizer	Adam
	Batch size	32
	Epoch	25

Table 3. Experimental Environment

Parameters	Configuration
OS	Linux
CPU	Intel(R) Xeon(R) Gold 5118
Memory	16G
Programming language	Python 3.8
Programming environment	PyTorch 1.12.1

The above bar charts and graphs clearly indicate that the proposed model achieves the maximum ACC and F1 scores for all four datasets. The results are 82.69% and 79.65% for the SMP2020-EWECT

Figure 5. Results on the SMP2020-EWECT Dataset

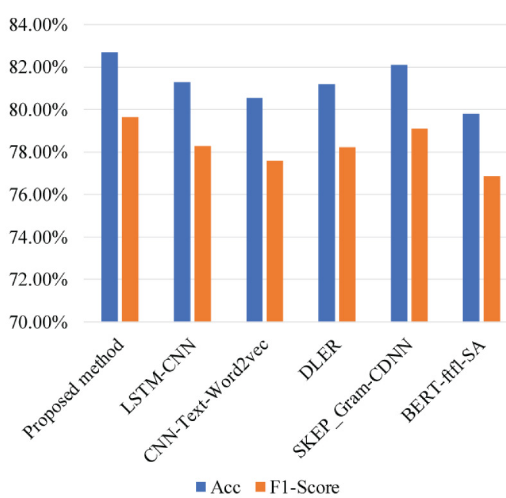


Table 4. Results on the SMP2020-EWECT Dataset

Model	Indicator	
	Accuracy	F1 Score
Proposed method	82.69%	79.65%
LSTM-CNN	81.28%	78.30%
CNN-Text-Word2vec	80.54%	77.58%
DLER	81.20%	78.22%
SKEP_Gram-CDNN	82.11%	79.09%
BERT-ftfl-SA	79.80%	76.86%

Table 5. Results on the NLPCC 2013 Task 2 Dataset

Model	Indicator	
	Accuracy	F1 Score
Proposed method	69.28%	53.87%
LSTM-CNN	68.10%	52.95%
CNN-Text-Word2vec	67.48%	52.47%
DLER	68.03%	52.90%
SKEP_Gram-CDNN	68.80%	53.49%
BERT-ftfl-SA	66.86%	51.98%

dataset, 69.28% and 53.87% for the NLPCC 2013 Task 2 dataset, 73.64% and 60.22% for the NLPCC 2014 Task 1 dataset, 98.87% and 98.75% for the weibo_senti_100k dataset, respectively.

Figure 6. Results on the NLPCC 2013 Task 2 Dataset

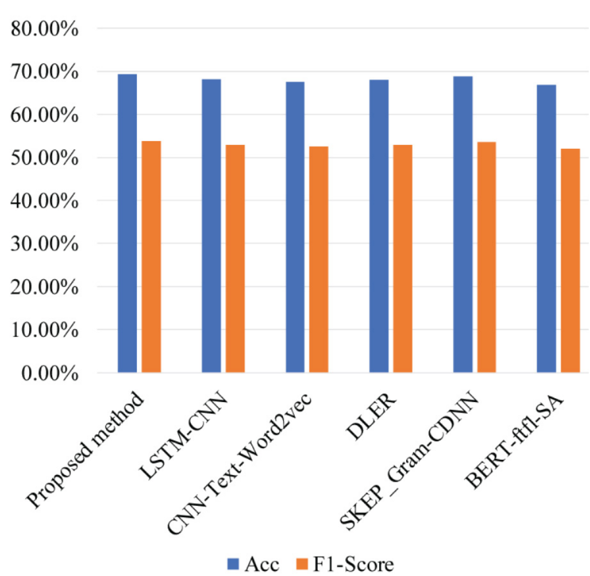


Figure 7. Results on the NLPCC 2014 Task 1 Dataset

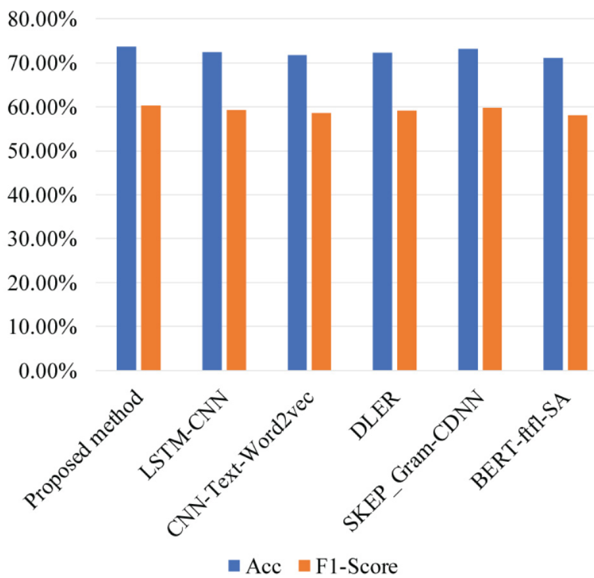


Table 6. Results on the NLPCC 2014 Task 1 Dataset

Model	Indicator	
	Accuracy	F1-Score
Proposed method	73.64%	60.22%
LSTM-CNN	72.39%	59.20%
CNN-Text-Word2vec	71.73%	58.65%
DLER	72.31%	59.14%
SKEP_Gram-CDNN	73.12%	59.80%
BERT-ftfl-SA	71.06%	58.11%

In contrast to the top-performing SKEP_Gram-CDNN model, the model presented in this paper utilizes a BiLSTM network to extract both local and global semantic information from the text. While the SKEP_Gram-CDNN model may result in the loss of semantic information between words for complex utterances, which impacts the final classification effect, the proposed model extracts both local and global semantic information from the text through bidirectional processing.

For sentiment analysis, LSTM-CNN only executes simple classification with poor granularity. In contrast, the proposed model acquires the fused features using multi-head self-attention fusion and average pooling. This improves the model's ability to comprehend the input sequences and extract crucial contextual information. It also enhances the precision of text sentiment classification and the level of detail in sentiment analysis.

The DLER model encounters limitations in accurately generating feature-grained similarity across domains (FG-SA) and exploring the features of distinct information or incorrect features (due to the vast difference in feature sets) among text information from multiple domains. In contrast, this paper employs the RoBERTa model to dynamically encode the text and augment understanding through the construction of a representation dictionary.

Figure 8. Results on the weibo_senti_100k Dataset

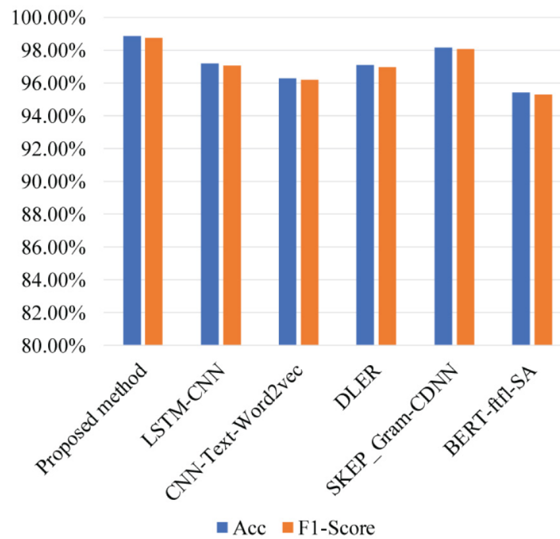


Table 7. Results on the weibo_senti_100k Dataset

Model	Indicator	
	Accuracy	F1-Score
Proposed method	98.87%	98.75%
LSTM-CNN	97.19%	97.07%
CNN-Text-Word2vec	96.30%	96.18%
DLER	97.09%	96.97%
SKEP_Gram-CDNN	98.18%	98.06%
BERT-ftfl-SA	95.41%	95.29%

The CNN-Text-Word2vec model struggles to combine multiple distinct feature extraction methods efficiently, thereby impeding the model's performance. The proposed model solves this issue by utilizing the Transformer to merge the multidimensional features, adaptively improve the key features, and circumvent the CNN-Text-Word2vec method's propensity to lose textual information.

In contrast to the BERT-ftfl-SA model, the proposed approach employs the enhanced focal loss function to address class imbalance issues and the multi-head attention layer to produce the final feature representation for classification.

In summary, the proposed method utilizes the RoBERTa model within the embedding layer to extract superficial text features, substantially augmenting the text's semantic features. The model utilizes BiLSTM to extract semantic features, enabling it to capture bidirectional semantic dependencies more effectively. Furthermore, implementing a multi-layer Transformer encoder significantly improves the accuracy of FGSA by facilitating multi-dimensional feature fusion and precisely capturing the mapping relationship between multi-dimensional features.

Table 8. Model Ablation Experimental Results

Model	Indicator		Test time for a single sample (s)
	Accuracy	F1-Score	
Model 1	84.43%	84.33%	0.019
Model 2	56.55%	56.49%	0.016
Model 3	49.73%	49.67%	0.016
Model 4	55.27%	55.20%	0.014
Model 5	88.79%	88.68%	0.020
Model 6	78.90%	78.80%	0.017
Model 7	90.27%	90.16%	0.021
Model 8	89.48%	89.37%	0.020
Proposed model	98.87%	98.75%	0.022

Ablation Experiment

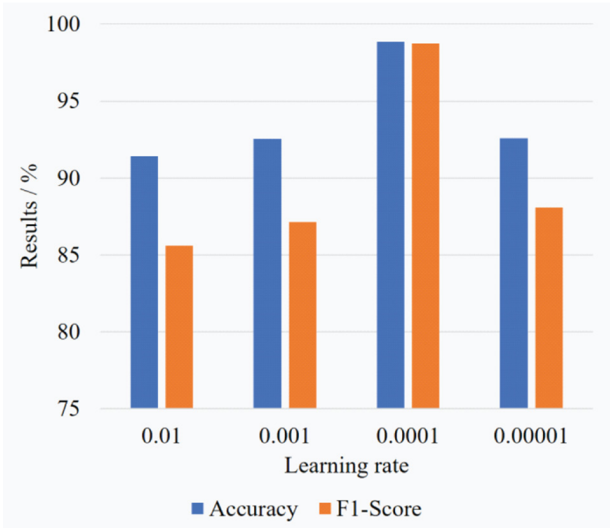
To enhance the validation of the efficacy of individual modules within the model, we devised a model ablation experiment to juxtapose the effects of distinct modules on the model's overall performance. The design consists of the following:

Model 1, w/o knowledge enhancement: RoBERTa is used for dynamic encoding directly in the word encoding layer without introducing knowledge enhancement.
Model 2, w/o RoBERTa: The traditional One Hot encoding is directly used instead of using RoBERTa for dynamic encoding in the word encoding layer.
Model 3, w/o BiLSTM: The semantic feature representation of BiLSTM is removed in the proposed model, and the features obtained from the word embedding layer are directly input into the Transformer.
Model 4, w/o Transformer: The feature fusion of the Transformer is removed from the proposed model, and pooling, full connection, and classification operations are directly performed on the features obtained from the semantic feature extraction layer.
Model 5, w/o Avg pooling: The Transformer only uses the maximum pooling operation in the feature fusion layer.
Model 6, w/o Max pooling: The Transformer only uses the average pooling operation in the feature fusion layer.
Model 7, Ls1: The improved focal loss function is replaced with the mean square error (MSE) loss function in the proposed model.
Model 8, Ls2: The improved focal loss function is replaced with the standard cross-entropy (CE) loss function.

We carried out the experiment utilizing the SMP2020-EWECT dataset, and the obtained results are listed in Table 8.

For the SMP2020-EWECT datasets, Table 8 indicates that the ACC and F1 score of the model's sentiment analysis will decrease if any module is removed. The substantial reduction in evaluation indicators for Model 2, Model 3, and Model 4 across all modules suggests that the BiLSTM, RoBERTa, and Transformer modules substantially enhance the results. The absence of any one of these three modules will affect the model's performance. Furthermore, alternative models cannot attain optimal performance, suggesting that enhancing the loss function and integrating pooling contribute to the

Figure 9. Influence of Learning Rate on the Proposed Model



overall enhancement of the results. The ablation experiment unequivocally demonstrates that each module of the proposed model is essential for optimal performance.

In practical applications, the model is usually trained offline and then deployed in the application system. The test experiments on a single sample show that the proposed model has a short test time and good practicality. When the Transformer module is removed, the performance of the proposed FGSA model decreases most significantly, which indicates that the semantic features are beneficial in enhancing the overall performance of the model. Moreover, the proposed FGSA most significantly reduces the test time when the Transformer module is removed. The reason is that the Transformer module contains a multi-head self-attention mechanism, which increases the computational overhead to a certain extent. However, despite the improved loss function, the testing time of the proposed FGSA does not increase significantly and remains acceptable for real-world applications.

Hyperparameter Analysis

Learning Rate Training

To evaluate the impact of different learning rates, we conducted an experiment on the weibo_senti_100k dataset using the proposed model. The results are illustrated in Figure 9.

Figure 9 illustrates that both the ACC and F1 score steadily improve as the learning rate decreases from 0.01 to 0.0001. Nevertheless, with the transition of the learning rate from 0.001 to 0.00001, a discernible decline in the values of the evaluation indicators emerged. The model's optimal performance is achieved with a learning rate of 0.001.

Dropout Analysis

The impact of Dropout on the model's performance was also investigated. Utilizing the "weibo_senti_100k" dataset, an experimental evaluation was conducted to assess the efficacy of the proposed model across various dropout conditions. The findings are presented in Figure 10.

Figure 10 shows that the ACC and F1 score both increase steadily from 0.2 to 0.4 as the dropout rate increases. However, a further increase in the dropout rate from 0.4 to 0.6 leads to a decline in the aforementioned evaluation indicators. The optimal performance of the model is observed when the dropout rate is configured to 0.4.

Figure 10. Influence of Dropout on the Proposed Model

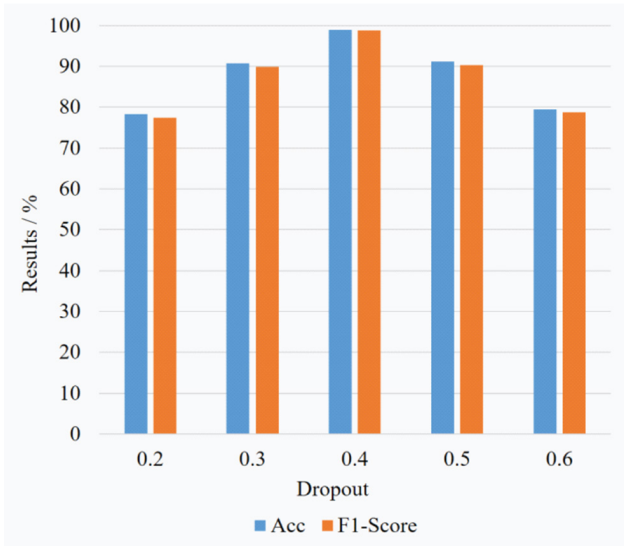
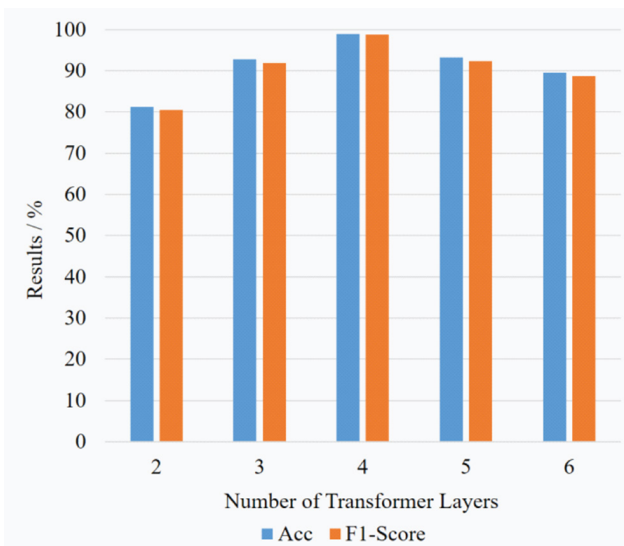


Figure 11. Influence of Transformer Layers



Transformer Layer Analysis

The configuration of Transformer layers may influence the training of a model. Therefore, we experimentally assessed the performance of the proposed model across various Transformer layers using the weibo_senti_100k dataset. The findings are illustrated in Figure 11.

Figure 11 illustrates that both the ACC and F1 score increase as the number of Transformer layers increases from two to four. Nevertheless, as the number of Transformer layers further increases from four to six, the evaluation indicators' values decrease. Thus, the model demonstrates its highest level of efficacy with four Transformer layers.

Table 9. Case Analysis Results of the Proposed FGSA Model on the NLPCC2014 Task1 Dataset

Sample	Text	Emotional label	FGSA w/o RoBERTa	FGSA w/o BiLSTM	FGSA w/o Transformer	FGSA
(a)	It’s a major bug.	None	√	√	√	√
(b)	I feel tired.	Sadness	×	√	√	√
(c)	I like apples.	Like	√	×	×	√
(d)	The washing machine I bought broke down in one day.	Anger	√	×	√	√
(e)	Happy New Year!	Happiness	√	√	√	√
(f)	I haven’t done my homework yet. It's 12 o ‘clock.	Disgust	×	√	√	√
(g)	Haunted houses have always scared me.	Fear	×	×	×	×
(h)	What Asia has done is amazing.	Surprise	×	×	√	√

Case Analysis

To evaluate more intuitively the influence of each critical component on the performance of the proposed FGSA model, we randomly selected eight samples from the NLPCC2014 Task 1 dataset. We conducted classification experiments on FGSA w/o RoBERTa, FGSA w/o Transformer, FGSA w/o BiLSTM, and FGSA. The classification results are shown in Table 4. The symbols “√” and “×” indicate whether the model correctly predicts the emotional polarity.

Table 9 shows that FGSA w/o RoBERTa and FGSA w/o BiLSTM correctly predicted four of the eight selected samples. However, FGSA w/o Transformer correctly predicted six, indicating that the Transformer encoder module is crucial to improving the final prediction performance. The FGSA model correctly predicted seven with an accuracy of 87.5%, demonstrating that the RoBERTa, BiLSTM, and Transformer modules all contribute to the analysis of sentiment features. In particular, the fusion of these three modules can effectively enhance the performance of sentiment analysis. Among them, the prediction for (g) is incorrect for all four models. The reason is that the meaning of the text in the examples is unclear and even requires some domain knowledge and general knowledge accumulation, which results in the models being unable to understand the emotions expressed in the text well. This also indicates that the proposed FGSA model needs further improvement.

The proposed Transformer-based method for FGSA of Weibo comment text is highly significant in practical applications. The proposed method provides an effective tool for social media analysis, market research, and customer feedback by solving the problems of word polysemy, information loss, and sample imbalance. In social media analysis, accurate sentiment analysis helps to monitor the trend of social opinion and user emotions in real time. In market research, precise sentiment analysis helps quickly obtain users’ feelings about products or services and improves competitive advantages. For customer feedback, sentiment analysis helps enterprises promptly identify and respond to potential crises and optimize customer experience. This methodology equips businesses to streamline operations, strengthen brand reputation, and deliver experiences that delight users, ultimately driving business growth.

CONCLUSION

This paper has proposed a Transformer-based Weibo comment text FGSA method to address the issues of fine-grained information loss, polysemy, and imbalanced sample categories in the existing FGSA methods. We conducted multiple experiments to verify the performance of the proposed method. The obtained results validate that introducing the RoBERTa model in the embedding layer can effectively enhance the semantic features of the text and solve the problem of polysemy. Utilizing BiLSTM to extract text semantic information from both directions better captures bidirectional global semantic dependency features. Using a Transformer to fuse multi-dimensional features can adaptively strengthen key features, overcoming the problem of fine-grained information loss. The improved focal loss function can effectively solve the issue of imbalanced categories in sample labels, improving the accuracy of sentiment analysis.

However, the proposed method also has limitations in terms of scalability and practicality. In future research, we will apply the proposed FGSA model to user sentiment analysis on larger and more complex social media platforms and other platforms, such as e-commerce and online education, to improve its scalability. As new deep-learning network models emerge, the proposed FGSA model can be further enhanced. The integration of techniques like graph convolutional networks, advanced pre-trained language models, and next-generation Transformers holds the potential to improve the model's accuracy in sentiment analysis. Furthermore, in terms of model training efficiency, the number of parameters will be reduced by methods such as model optimization to improve the model's practicality.

Conflicts of Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Funding Statement

This research supported by Scientific Research Fund of Xi'an Mingde Institute of Technology [No.2022XY07L05].

Process Dates

Received: This manuscript was initially received for consideration for the journal on 02/21/2024, revisions were received for the manuscript following the double-anonymized peer review on 03/26/2024, the manuscript was formally accepted on 04/13/2024, and the manuscript was finalized for publication on 04/24/2024.

Corresponding Author

Correspondence should be addressed to Piao Xue; hlqldyt@163.com

REFERENCES

Chen, Z., Zhou, L. J., Li, X. D., Zhang, J. N., & Huo, W. J. (2020). The Lao text classification method based on KNN. *Procedia Computer Science*, *166*, 523–528. 10.1016/j.procs.2020.02.053

Djaballah, K. A., Boukhalfa, K., Boussaid, O., & Ramdane, Y. (2021). An improved sentiment analysis approach to detect radical content on Twitter. [IJITWE]. *International Journal of Information Technology and Web Engineering*, *16*(4), 52–73. 10.4018/IJITWE.2021100103

Fan, X. (2023). Artificial intelligence technology-based semantic sentiment analysis on network public opinion texts. *International Journal of Information Technologies and Systems Approach*, *16*(2), 1–14. 10.4018/IJITSA.318447

Fang, M., Peng, S., Liang, Y., Hung, C. C., & Liu, S. (2023). Control, a multimodal fusion model with multi-level attention mechanism for depression detection. *Biomedical Signal Processing and Control*, *82*, 104561. 10.1016/j.bspc.2022.104561

Gandhi, A., Adhvaryu, K., Poria, S., Cambria, E., & Hussain, A. (2023). Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. *Information Fusion*, *91*, 424–444. 10.1016/j.inffus.2022.09.025

Hao, F., Li, C. L., Ji, D., & Li, F. (2022). Mutual disentanglement learning for joint fine-grained sentiment classification and controllable text generation. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 1555–1565. Wuhan University.

Hazarika, D., Zimmermann, R., & Poria, S. (2020). MISA: Modality-invariant and -specific representations for multimodal sentiment analysis. In *Proceedings of the ACM International Conference on Multimedia (ACM MM)*, 1122–1131. ACM. 10.1145/3394171.3413678

Huang, F., Zhang, X., Zhao, Z., Xu, J., & Li, Z. (2019). Image–text sentiment analysis via deep multimodal attentive fusion. *Knowledge-Based Systems*, *167*, 26–37. 10.1016/j.knosys.2019.01.019

Jiang, D., Hao, L., Wei, R., & Tu, G. (2023). CSAT-FTCN: A fuzzy-oriented model with contextual self-attention network for multimodal emotion recognition. *Cognitive Computation*, *15*(3), 1–10. 10.1007/s12559-023-10119-6

Jiang, M., Zhang, W., Zhang, M., Wu, J., & Wen, T. (2019). An LSTM-CNN attention approach for aspect-level sentiment classification. *Journal of Computational Methods in Sciences and Engineering*, *19*(4), 859–868. 10.3233/JCM-190022

Joshy, A., & Sundar, S. (2022). Analyzing the performance of sentiment analysis using BERT, DistilBERT, and RoBERTa. In *Proceedings of the IEEE International Power and Renewable Energy Conference (IPRECON)*, 16–21. IEEE. 10.1109/IPRECON55716.2022.10059542

Kaur, R., & Kautish, S. (2022). Multimodal sentiment analysis: A survey and comparison. In *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 1846–1870). Information Resources Management Association. 10.4018/978-1-6684-6303-1.ch098

Li, M., Bi, X., Wang, L., Han, X., Wang, L., & Zhou, W. (2022). Text similarity measurement method and application of online medical community based on density peak clustering. [JOEUC]. *Journal of Organizational and End User Computing*, *34*(2), 1–25. 10.4018/JOEUC.302893

Li, R., Wu, Z., Jia, J., Bu, Y., Zhao, S., & Meng, H. (2019). Towards discriminative representation learning for speech emotion recognition. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 5060–5066. IJCAI. 10.24963/ijcai.2019/703

Li, W., Qi, F., Tang, M., & Yu, Z. (2020). Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing*, *387*(C), 63–77. 10.1016/j.neucom.2020.01.006

Li, Z., Sun, Q., Guo, Q., Wu, H., Deng, L., Zhang, Q., Zhang, J., Zhang, H., & Chen, Y. (2021). Visual sentiment analysis based on image caption and adjective–noun–pair description. *Soft Computing*, ●●●, 1–13. 10.1007/s00500-021-06530-6

Liao, W., Zeng, B., Liu, J., Wei, P., & Fang, J. (2022). Image-text interaction graph neural network for image-text sentiment analysis. *Applied Intelligence*, *52*(10), 11184–11198. 10.1007/s10489-021-02936-9

Liao, W. X., Zeng, B., Yin, X. W., & Wei, P. (2021). An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. *Applied Intelligence*, *5*(6), 3522–3533. 10.1007/s10489-020-01964-1

Liao, W. Z., Ma, Y. H., Cao, Y. L., Ye, G., & Zuo, D. (2021). Two-stage fine-grained text-level sentiment analysis based on syntactic rule matching and deep semantic. *IEICE Transactions on Information and Systems*, *104*(8), 1274–1280. 10.1587/transinf.2020BDP0018

Lin, C., Zhao, S., Meng, L., & Chua, T. S. (2020). Multi-source domain adaptation for visual sentiment classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(3), 2661–2668. 10.1609/aaai.v34i03.5651

Liu, J. (2023). Sentiment classification of social network text based on AT-BiLSTM model in a big data environment. *International Journal of Information Technologies and Systems Approach*, *16*(2), 1–15. 10.4018/IJITSA.324808

Liu, L. L., Yang, Y., & Hu, J. (2020). Simple but effective: A model for aspect-based sentiment analysis. In *Proceedings of the 15th Symposium of Intelligent Systems and Knowledge Engineering (ISKE) held jointly with the 14th International FLINS Conference (FLINS)*, 52–59. 10.1142/9789811223334_0007

Liu, S., Li, K., & Li, Z. (2022). A robustly optimized BMRC for aspect sentiment triplet extraction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 272–278. 10.18653/v1/2022.naacl-main.20

Lu, Q., Zhu, Z., Zhang, G., Kang, S., & Liu, P. (2021). Aspect-gated graph convolutional networks for aspect-based sentiment analysis. *Applied Intelligence*, *51*(7), 4408–4419. 10.1007/s10489-020-02095-3

Poria, S., Hazarika, D., Majumder, N., & Mihalcea, R. (2023). Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research. *IEEE Transactions on Affective Computing*, *14*(1), 108–132. 10.1109/TAFFC.2020.3038167

Sirisha, U., & Chandana, B. S. (2022). Aspect based sentiment and emotion analysis with ROBERTa, LSTM. *International Journal of Advanced Computer Science and Applications*, *13*(11), 766–774. 10.14569/IJACSA.2022.0131189

Tamil, S. P., Balasubramaniam, K., Vidhya, S., Jayapandian, N., Ramya, K., Poongodi, M., Hamdi, M., & Tunze, G. B. (2022). Social network user profiling with multilayer semantic modeling using ego network. [IJITWE]. *International Journal of Information Technology and Web Engineering*, *17*(1), 1–14. 10.4018/IJITWE.304049

Theodoropoulos, P., & Alexandris, C. (2022). Fine-grained sentiment analysis of multi-domain online reviews. In Kurosu, M. (Ed.), Lecture Notes in Computer Science: *Vol. 13303*. *Human–Computer Interaction. Technological Innovation. HCII 2022* (pp. 264–278). Springer. 10.1007/978-3-031-05409-9_20

Turney, P. D. (2020). Thumbs up or thumbs down: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 35–40. ACL.

Xiao, Q., Li, S., Zhang, X., Zhang, F., Yue, Q., & Wan, S. (2022). Deconstructing online hospitality review systems: User quality experience toward design features. [JOEUC]. *Journal of Organizational and End User Computing*, *34*(2), 1–17. 10.4018/JOEUC.292523

Xu, D., Tian, Z., Lai, R., Kong, X., Tan, Z., & Shi, W. (2020). Deep learning based emotion analysis of microblog texts. *Information Fusion*, *64*, 1–11. 10.1016/j.inffus.2020.06.002

Yin, Z. Y., Kou, Y., Wang, G., Shen, D., & Nie, T. (2021). Explainable recommendation via neural rating regression and fine-grained sentiment perception. In Xing, C., Fu, X., Zhang, Y., & Borjigin, C. (Eds.), *Web Information Systems and Applications. WISA* (pp. 580–591). Lecture Notes in Computer Science. Springer. 10.1007/978-3-030-87571-8_50

Yu, W., Xu, H., Yuan, Z., & Wu, J. (2021). Learning modality-specific representations with self-supervised multi-task learning for multimodal sentiment analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(12), 10790–10797. 10.1609/aaai.v35i12.17289

Zeng, Y., Li, Z., Tang, Z., Chen, Z., & Ma, H. (2023). Heterogeneous graph convolution based on in-domain self-supervision for multimodal sentiment analysis. *Expert Systems with Applications*, *213*(C), 119240. 10.1016/j.eswa.2022.119240

Zhang, J., Liu, X., Chen, M., Ye, Q., & Wang, Z. (2022). Image sentiment classification via multi-level sentiment region correlation analysis. *Neurocomputing*, *469*, 221–233. 10.1016/j.neucom.2021.10.062

Zhang, X., & Cui, Y. (2023). Analysis of application and creation skills of story-based MV micro video and big multimedia data in music communication. [IJITWE]. *International Journal of Information Technology and Web Engineering*, *18*(1), 1–17. 10.4018/IJITWE.325213

Zhang, Y., Song, D., Li, X., Zhang, P., Wang, P., Rong, L., Yu, G., & Wang, B. (2020). A quantum-like multimodal network framework for modeling interaction dynamics in multiparty conversational sentiment analysis. *Information Fusion*, *62*, 14–31. 10.1016/j.inffus.2020.04.003

Zhang, Y., Wang, J., Liu, Y., Rong, L., Zheng, Q., Song, D., Tiwari, P., & Qin, J. (2023). A multitask learning model for multimodal sarcasm, sentiment and emotion recognition in conversations. *Information Fusion*, *93*, 282–301. 10.1016/j.inffus.2023.01.005

Zhang, Y., & Yang, Q. (2022). A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, *34*(12), 5586–5609. 10.1109/TKDE.2021.3070203

Zhang, Y. R., Zhu, C. X., & Xie, Y. X. (2023). Fine-grained sentiment analysis of cross-domain Chinese e-commerce texts based on SKEP_Gram-CDNN. *IEEE Access : Practical Innovations, Open Solutions*, *11*(5), 74058–74070. 10.1109/ACCESS.2023.3296447

Zhou, F. G., Zhang, J., & Song, Y. N. (2022). Chinese fine-grained sentiment classification based on pre-trained language model and attention mechanism. In *Proceedings of the* 6th *International Conference on Smart Computing and Communication (SmartCom)*, 37–47. 10.1007/978-3-030-97774-0_4

Zhu, T., Li, L., Yang, J., Zhao, S., Liu, H., & Qian, J. (2022). Multimodal sentiment analysis with image text interaction network. *IEEE Transactions on Multimedia*, *22*(8), 2087–2099.

Piao Xue, Lecturer. Master's degree, Graduated from Qinghai Normal University in 2019. Worked in Xi'an Mingde Institute of Technology. His research interests include mental health education and psychological counseling.

Wei Bai received his M.S. degree in Education from Ningxia University in 2009, China. He is currently worked in Ningxia Normal University, associate professor. His research interest in the area of Computer Science and Technology. He has published several research papers in scholarly journals in the above research areas.