

A Method for Classifying Ancient Poetry Emotions Based on Pre-trained Models and Attention Mechanisms

Guanyu Liu *

College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, 310058, China

* Corresponding Author Email: GuanyuLiu@zuua.zju.edu.cn

Abstract. Chinese classical poetry is a treasure of Chinese culture with rich and subtle emotional connotations. Studying the realization of automated emotional analysis of ancient poetry is of great significance for fields such as digital humanities and literary studies. This paper proposes an emotion classification model based on a pre-trained model and combined with the attention mechanism in response to the deficiencies in emotion classification accuracy and multi-label emotion classification in existing emotion classification models of ancient poetry and prose. This model is based on the RoBERTa-wwm-ext-large pre-trained model. An attention mechanism is introduced for the multi-label classification task to capture the emotional keywords in the verses, a weighted binary cross-entropy loss function was adopted to address the severe category imbalance problem in multi-label tasks. Experimental results show that the method achieves an accuracy of 90.80% for the single-label five-classification task on the THU fine-grained sentimental poetry corpus (FSPC) dataset, surpassing the CDBERT model's result of 51.40. On the Chinese Classical Poetry dataset (CCPD), the multi-label 12-class tasks Micro-F1 and Macro-F1 achieved 99.40% and 99.38% respectively, outperforming the CDBERT model by 89.95% and 74.90% respectively. This study validates the effectiveness of pre-trained models combined with attention mechanisms in sentiment analysis of classical poetry and provides new ideas for sentiment analysis research in classical literature.

Keywords: Sentiment Analysis, Classical Chinese poetry and prose sentiment analysis, Pre-trained models, Attention mechanism.

1. Introduction

Sentiment analysis, as one of the key research directions in the field of natural language processing, aims to identify and understand the subjective emotions, attitudes, and viewpoints contained in the text. In recent years, with the rapid development of deep learning techniques, especially driven by pre-trained language models such as BERT and RoBERTa, modern text sentiment analysis techniques have achieved fruitful results in multiple fields. However, researchers still face many challenges in applying these techniques to the particular field of Chinese classical poetry. Classical Chinese poetry holds a significant position in Chinese literature and is rich in emotional expressions. The deepening of research on the automatic classification of emotions in classical poetry not only holds significant theoretical value in the field of literary computing but also shows broad prospects in application scenarios such as the digitalization of cultural heritage, intelligent education systems, and cultural and creative industries. With the advancement of China's national cultural digitalization strategy, the digital protection and intelligent analysis of ancient poetry and prose have become one of the core topics in digital humanities research [1][2][3].

However, the expression of Chinese poetry is characterized by concise language, rich imagery, and implicit emotions, such as the common inversion and omission in grammatical structure, the use of various natural images as symbols of specific emotions ("willow" represents parting, "moon" conveys longing), and the indirect and implicit conveyance of emotions through expressions such as "expressing emotions through scenery" (Du Mu's "Autumn Evening" is a superficial description of the scenery, So how to achieve automatic emotion classification in Chinese poetry is one of the major challenges in the field of natural language processing. Early studies on sentiment analysis of ancient poetry mainly used a combination of dictionaries and rules. For example, Zhang et al. proposed a rule system for sentiment classification based on rhymes and imagery words, which

achieved an accuracy rate of 41.2% in Tang Dynasty poetry [4]. Zeng et al. compared the performance of SVM, naive Bayes, and maximum entropy models in sentiment analysis of ancient poetry. The best accuracy can reach 46.8%[5], but these methods, though highly interpretable, face flaws including difficulty in handling metaphors and allusions in ancient poetry and limited generalization ability due to reliance on artificial feature engineering; With the development of deep learning technology, various neural network architectures such as CNN, BiLSTM, and Transformer have been proposed and used to analyze the sentiment of Chinese verses. For example, Du used character-level CNN to capture local features of ancient poems and achieved an accuracy of 48.3% on FSPC [6] Although these methods have made further improvements in recognition accuracy, there are still problems such as underutilization of the transfer learning ability of pre-trained language models and insufficient modeling of temporal relationships between verses; Nowadays, large language models such as the GPT series and ChatGLM offer new ideas for sentiment analysis of ancient poetry. Recently, researchers have also discussed using multimodal recognition mechanisms to combine poetry, calligraphy and painting, and music to improve the accuracy of sentiment recognition in ancient poetry. Limited understanding of the special grammatical structure of ancient poetry, vulnerability to modern language bias, and high computational resource demands have restricted the further improvement of these methods in this field. In recent years, pre-trained language models have performed well in text classification tasks, but research on the particular genre of ancient poetry is still insufficient [7].

The main contributions are as follows:

(1) An innovative dual-branch processing framework was designed for the coexisting single-label and multi-label classification tasks in the analysis of ancient poetry sentiment. In this framework, the two classifiers share a pre-trained base model, and at the same time derive models with different classification heads based on task characteristics, achieving precise adaptation and efficient processing of the two types of tasks in a single system.

(2) In view of the implicit, scattered and keyword-driven expression of emotions in ancient poetry, this study algorithm introduces an attention mechanism into the multi-label classification model, which automatically learns to capture the emotional keywords in the verses and generates weighted context vectors instead of simple sentence beginning vectors, significantly improving the ability to capture and recognize complex and multiple emotions in poetry.

(3) To address the common problem of category imbalance in datasets, this paper specifically designs a weighted binary cross-entropy loss function. By appropriately increasing the weights of categories with fewer supporting samples in the loss function, it effectively ensures that the model's recognition ability for all emotions is at a high level.

2. Research Tools and Methods

2.1. Model Architecture

The model structure proposed in this study is shown in Figure 1

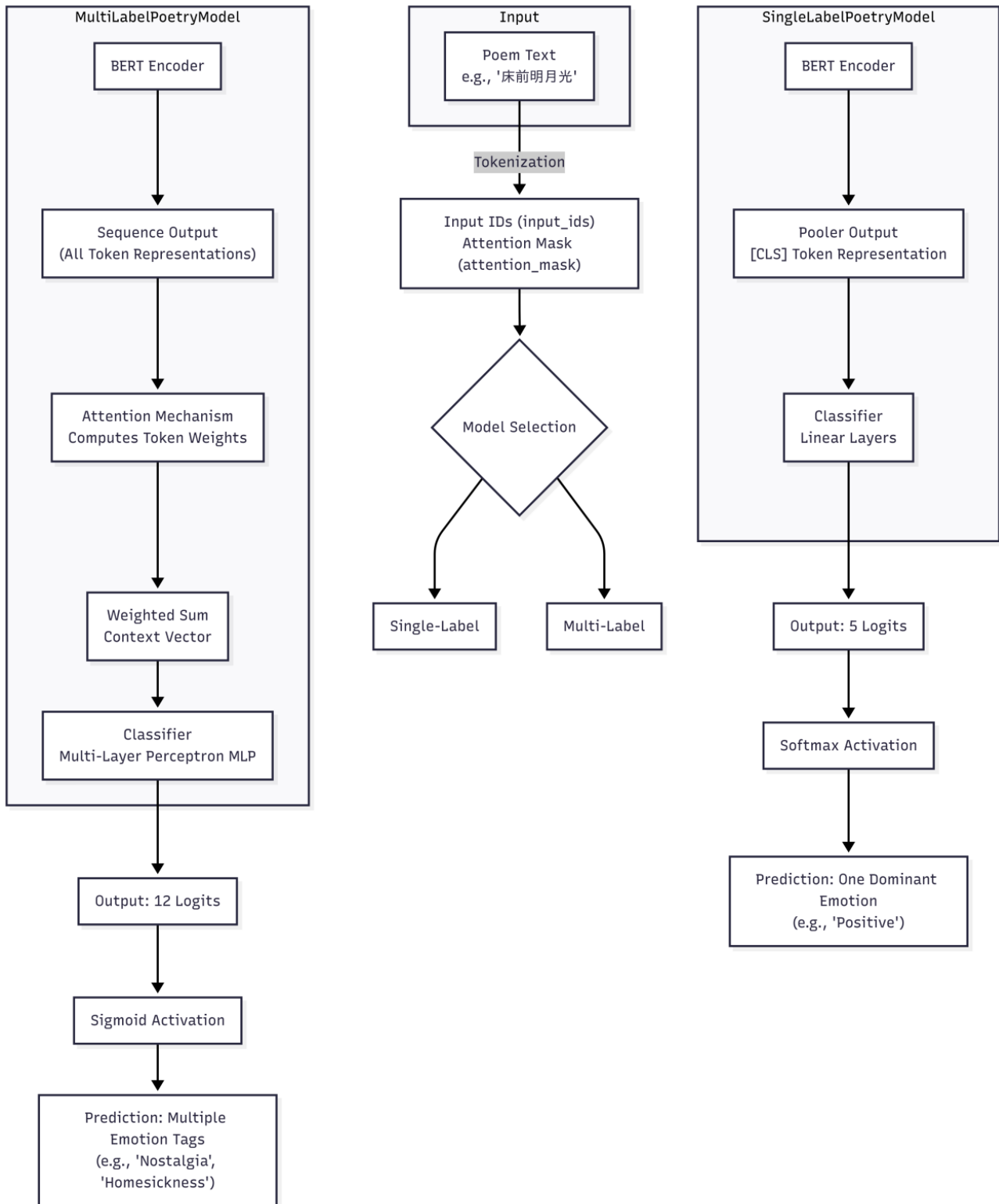


Fig. 1 Model Architecture diagram (Picture credit: Original)

This paper presents a model based on a pre-trained model and data processing, and selects the corresponding classifier according to the single-label or multi-label classification task. The core process is: first, preprocess and segment the input ancient poetry and prose information; Then, depending on whether the task type belongs to single-label classification or multi-label classification, choose to classify directly using the pooling output vector of the BERT encoder or introduce an additional attention mechanism to capture sentiment keywords, generate weighted context vectors before classification; Ultimately, the results are projected through the corresponding loss functions of the two classifiers to generate metric data for evaluation.

2.2. Data Preprocessing and Classifier Selection

In ancient poetry sentiment classification tasks, the sentiment labels of the verses provided by the dataset range from unique to multiple, and the content formats of different texts vary greatly. Therefore, parsing and cleaning the data in advance is of great significance for improving recognition accuracy.

The FSPC dataset and CCPD dataset used in this study are examples. The FSPC dataset is mainly in json format, with lines separated by the "|" symbol (e.g., "窗前明月光|疑是地上霜"), and it needs to be preprocessed to convert it into a complete text sequence (e.g., "窗前明月光 疑是地上霜") And convert the values corresponding to the sentiment labels into indexes of the appropriate size range to fit the PyTorch loss function; The CCPD dataset file is mainly in csv format, with multiple sentiment labels for each verse, which need to be processed and converted into a 12-dimensional Multi-hot Encoding vector.

At the same time, to address the problem that the imbalance of different types of data in the dataset affects the generalization ability of the model, the model enhances the data obtained after cleaning, such as randomly swapping the order of adjacent lines, randomly deleting a line of poetry, or randomly selecting a word from a line of poetry for replacement, forcing the model to reason in context during training. These operations invisibly expand the size of the dataset and enhance the generalization ability of the model [8].

chinese-roberta-wwm-ext-large is preferred as the feature encoder in this paper. The model, which was pre-trained on a large Chinese corpus using full-word masking techniques, outperforms earlier models in terms of language representation ability, especially in the holistic understanding of Chinese vocabulary. Given an input sequence, the encoder converts it into a series of token ids and attention masks through the Tokenizer. Let the input sequence be, where n is the sequence length, the encoder BERT maps it to a sequence of hidden states rich in context information:

$$H = \text{BERT}(X) = (h_{[\text{CLS}]}, h_{[1]}, h_{[2]}, \dots, h_{[n]}, h_{[\text{SEP}]}) \quad (1)$$

Where, d_h is the dimension of the hidden layer, $H \in \mathbb{R}^{n \times d_h}$.

2.3. Single-label Classification Model

For data with a single label, a relatively standard processing structure can be adopted [9]. $h_{[\text{CLS}]} \in \mathbb{R}^{n \times d_h}$ The model first uses the pooled output from the previous BERT encoder as the semantic representation of the entire sentence, a vector that was specifically optimized during pre-training to more clearly represent the overall information of the sequence. Then the vector passes through a classifier head, which is a relatively small feedforward neural network. The classifier head contains Dropout layers, LayerNorm layers, and nonlinear activation functions, which can help reduce overfitting and accelerate training. Ultimately, the classifier maps high-dimensional vectors representing verses to logits (unnormalized scores) for each category:

$$z = W \cdot \text{ReLU}(\text{LayerNorm}(W_1 \cdot h_{[\text{CLS}]} + b_1)) + b \quad (2)$$

Here, W_1, b_1, W, b are learnable parameter matrices and bias terms. The final predicted probabilities are obtained through the Softmax function:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \quad (3)$$

2.4. Multi-label Classification Models and Attention Mechanisms

For multi-classification tasks such as dealing with CCPD datasets, simply using the representations adopted in single-label classification models may not be sufficient to fully capture the relatively scattered emotional keyword information in the verses. For this purpose, this study introduces the attention mechanism. First the model uses the entire hidden state sequence H of the BERT encoder instead of just using it, $h_{[\text{CLS}]}$ and then an attention network f_{attn} assigns an importance weight score to each token in the sequence:

$$e_i = w_a^T \cdot \tanh(W_a \cdot h_i + b_a) \quad (4)$$

$$a_i = \frac{\exp(e_i)}{\sum_{j=1}^n \exp(e_j)} \quad (5)$$

Here, $W_a \in \mathbb{R}^{d_a \times d_h}$, $b_a \in \mathbb{R}^{d_a}$, $w_a \in \mathbb{R}^{d_a}$ is the learnable parameter, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ is the activation function. It is the correlation probability distribution, where $\sum_{i=1}^n \alpha_i = 1$. Then weighted summation is performed on these vectors, resulting in a Context Vector c :

$$c = \sum_{i=1}^n \alpha_i h_i \quad (6)$$

The context vector replaces the simple sequence vector average, adjusts the corresponding weights according to the distribution of the emotional keywords in the verse through adaptive learning of the model, and finally obtains the logits of each label in the multi-label through a deep perceptron located at a deeper level: $z = \mathbb{R}^n$

2.5. Loss Function Design

For calculating the loss of the single-label classification task, this study uses a cross-entropy loss function with category weights. The study first calculate the weight w_c of each category c , which is inversely proportional to the number of samples in that category N_c , to balance the category distribution:

$$w_c = \frac{N_{total}}{c \cdot N_c} \quad (7)$$

The expression of the loss function is:

$$L_{CE} = - \sum_{c=1}^n w_c \cdot y_c \log(\hat{y}_c) \quad (8)$$

Among them is the code of the real label y_c

In completing the multi-label classification task, this study uses weighted binary cross-entropy loss to address the imbalance in the number of different table label samples in the training task dataset. This study calculates the sample weights for each label k in the dataset: pos_weight_k

$$pos_weight_k = \frac{\text{number of negative samples for } k}{\text{number of positive samples for } k} \quad (9)$$

This weight automatically raises the contribution of labels with fewer positive samples to the loss function when calculated, and the weighted sum of the BCE losses for each label is the final loss:

$$L_{WBCE} = - \frac{1}{n} \sum_{k=1}^n [pos_weight_k \cdot y_k \log(\sigma(z_k)) + (1 - y_k) \log(1 - \sigma(z_k))] \quad (10)$$

Where σ is the Sigmoid function, $y_k \in 0,1$ is the true value of label k .

2.6. Loss Function Design

The following takes Du Mu's 《泊秦淮》 as an example to show the processing of this poem in a multi-label classifier, as shown in Figure 2 below.

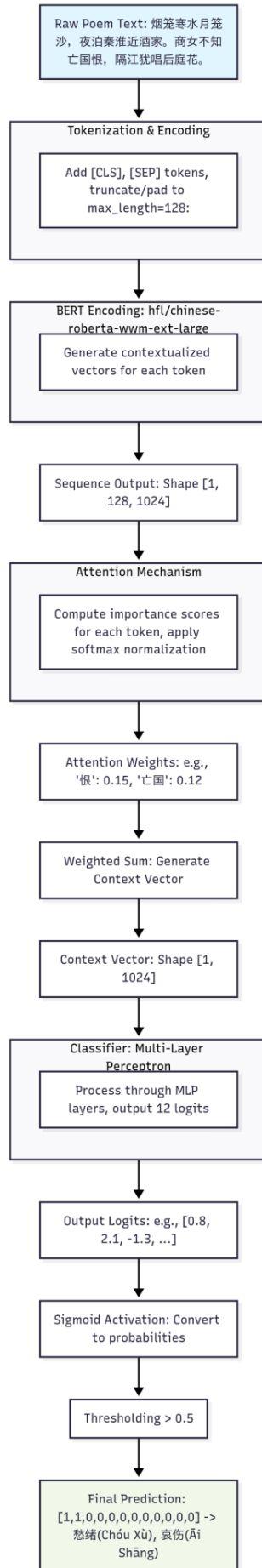


Fig. 2 Processing of the ancient poem 《泊秦淮》 in the model (Picture credit: Original)

the original input is "烟笼寒水月笼沙，夜泊秦淮近酒家。商女不知亡国恨，隔江犹唱后庭花。". The corresponding label is [1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]. First, the Tokenizer splits the sentence into subword units and adds special tokens [CLS] and [SEP]. Then convert these tokens into input_ids in the vocabulary [101, 2345, 1123, 3456,...] and generate attention_mask[1, 1, 1,...]. Then the BERT model takes input_ids and attention_mask and computes a batch of 128 1024-dimensional tokens through a multilayer Transformer; The attention layer then calculates weights for each token in the batch, and the model learns to assign appropriate weights to different emotional keywords such as "恨" (0.15), "亡国" (0.12), "愁" (0.10), "哀" (0.09),... "沙" (0.001); After processing through the linear layer and activation function, the output was mapped to the same number of output nodes as the number of labels, and the probability of each label was calculated independently and compared with the threshold to obtain the prediction results

[1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], namely "melancholy" and "sadness", consistent with the real labels.

3. Experimental Process and Result Analysis

3.1. Experimental Dataset

THU Fine-grained Sentimental Poetry Corpus (FSPC) is a dataset [9] with five classifications designed for a single-label classification task, containing 5,000 ancient poems with different emotional connotations, each poem consisting of four sentences. Each sentence was assigned an emotional label from one of five categories: "negative", "implied negative", "neutral", "implied positive" and "positive"; The Chinese Classical Poetry Dataset (CCPD) contains 52,479 ancient poems, each labeled with multiple possible emotional tags, totaling 12 fine-grained emotions: melancholy, grief, fear, surprise, tears, disappointment, longing, loneliness, joy, resentment, anger, homesickness [10].

3.2. Comparative Experiments and Analysis of Results

To evaluate the effectiveness of the study method, models pre-trained with CDBERT were selected as the comparison subjects. The FSPC dataset was divided into a training set of 8514, a validation set of 1065, and a test set of 1065 in an 8:1:1 ratio, with accuracy as the evaluation metric. The CCPD dataset was divided into a training set of 36,735 pieces, a validation set of 10,496 pieces, and a test set of 5,248 pieces, with Micro-F1 and Macro-F1 as the test criteria. The experimental results are shown in Table 1 below:

Table 1. Comparison of Experimental Results

Model	Dataset	Accuracy	Micro-F1	Macro-F1
CDBERT	FSPC	0.5140	-	-
Study Model	FSPC	0.9080	-	-
CDBERT	CCPD	-	0.8995	0.7490
Study Model	CCPD	-	0.9940	0.9938

Based on the analysis of the above experimental results, on the FSPC dataset, the accuracy rate of the model proposed in this paper reaches 90.80%, which is a relatively significant improvement of 39.4% compared to the 51.40% of CDBERT. This indicates that, compared to some commonly used standard models such as BERT, the powerful pre-trained model RoBERTa-wwm-ext-large and the well-tuned training strategy, such as data augmentation and assigning appropriate weights to sentiment keywords, can enable the model to make significant progress in the single-label sentiment classification task of ancient poetry. On the CCPD dataset, the model in this paper scored over 99% in both Micro-F1 and Macro-F1, outperforming the CDBERT model. This result fully demonstrates

the effectiveness of the strategies introduced in this paper for multi-label tasks, such as the attention mechanism and weighted loss function.

3.3. Analysis of Experimental Results and Conclusions for Sentiment Classification with Different Labels

This study also recorded the recognition performance of the model on each emotion label during the experiment, as shown in Tables 2 and 3.

Table 2. Classification results of labels on FSPC

category	precision	recall	F1-score	sample number
Negative	0.93	0.87	0.90	87
Hidden negative	0.89	0.92	0.90	293
Neutral	0.88	0.89	0.88	266
Hidden frontal	0.93	0.90	0.91	313
Front	0.97	0.97	0.97	106

Table 3. Comparison of Experimental Results

category	F1-score	category	F1-score	category	F1-score	category	F1-score
Melancholy	0.9962	Surprise	0.9964	Longing	0.9949	Resentment	0.9931
Sorrow	0.9850	Tears	0.9915	Loneliness	0.9890	Anger	1.000
Fear	0.9973	Disappointment	0.9938	Joy	0.9925	Homesickness	0.9962

It is not difficult to see from the above experimental results that the model proposed in this study performed well in the classification of different sentiment labels on both datasets, especially from the test set selected from the FSPC dataset, which can be analyzed that even though there are considerable differences in the number of samples supporting different sentiment labels. The model's performance in recognizing different labels, both in terms of accuracy and recall, remained at a high level, which is consistent with the conclusions in Table 1 that the Micro-F1 model of this study has high scores and the Macro-F1 model has almost equal scores with the Micro-F1, proving that the model proposed in this study can effectively solve the problem of unbalanced training sample categories. It can effectively identify the more subtle emotions conveyed in ancient poetry and prose.

4. Conclusion

This study presents a model based on a large-scale pre-trained model, combined with an attention mechanism, for the task of classical poetry sentiment classification. Experimental results show that this model can effectively identify various emotions in ancient Chinese poetry and prose, and its performance is superior to that of models based on the commonly used pre-trained model CDBERT. It provides a beneficial exploration for sentiment analysis of classical Chinese poetry in the field of natural language processing and has positive significance for digital humanities research.

The results of this study can be further improved in the following aspects: Consider integrating the prosodic features of poetry into the feature analysis module when analyzing the verses; Introduce an attention mechanism to identify emotional keywords more effectively and combine an external knowledge base to enhance the model's ability to understand the cultural context contained in the verses.

References

- [1] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019: 4171-4186.
- [2] Zhang W, Wang H, Song M, et al. A method of constructing a fine-grained sentiment lexicon for the humanities computing of classical Chinese poetry. *Neural Computing and Applications*, 2023, 35 (3): 2325-2346.
- [3] Yi X, Sun M, Li R, et al. Chinese poetry generation with a working memory model. *arXiv preprint arXiv:1809.04306*, 2018.
- [4] Zhang J, Duan Y, Gu X. Research on emotion analysis of Chinese literati painting images based on deep learning. *Frontiers in Psychology*, 2021, 12: 723325.
- [5] Zeng Z, Li Y. Multi-modal Chinese text emotion metaphor computation based on mutual information and information entropy. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2024, 23 (8): 1-18.
- [6] Su C, Liu S, Luo C. Misc: A multimodal approach for sentiment classification of classical Chinese poetry. *International Conference on Intelligent Computing*. Singapore: Springer Nature Singapore, 2023.
- [7] Du X, Pei H, Zhang H. Picturized and recited with dialects: A multimodal Chinese representation framework for sentiment analysis of classical Chinese poetry. *arXiv preprint arXiv:2505.13210*, 2025.
- [8] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [9] Liu T, Wang F, Chen M. Rethinking tabular data understanding with large language models. *arXiv preprint arXiv:2312.16702*, 2023.
- [10] Wang C, Liu X, Awadallah A H. Cost-effective hyperparameter optimization for large language model generation inference. *International Conference on Automated Machine Learning*, PMLR, 2023: 21/1-17.