



ARTICLE

SA-Model: Multi-Feature Fusion Poetic Sentiment Analysis Based on a Hybrid Word Vector Model

Lingli Zhang¹, Yadong Wu^{1,*}, Qikai Chu², Pan Li², Guijuan Wang^{3,4}, Weihang Zhang¹, Yu Qiu¹ and Yi Li¹

¹School of Computer Science and Engineering, Sichuan University of Science and Engineering, Zigong, 643000, China

²School of Automation and Information Engineering, Sichuan University of Science and Engineering, Zigong, 643000, China

³School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, 621000, China

⁴School of Information Engineering, Southwest University of Science and Technology, Mianyang, 621000, China

*Corresponding Author: Yadong Wu. Email: wuyadong028@163.com

Received: 19 October 2022 Accepted: 11 January 2023

ABSTRACT

Sentiment analysis in Chinese classical poetry has become a prominent topic in historical and cultural tracing, ancient literature research, etc. However, the existing research on sentiment analysis is relatively small. It does not effectively solve the problems such as the weak feature extraction ability of poetry text, which leads to the low performance of the model on sentiment analysis for Chinese classical poetry. In this research, we offer the SA-Model, a poetic sentiment analysis model. SA-Model firstly extracts text vector information and fuses it through Bidirectional encoder representation from transformers-Whole word masking-extension (BERT-wwm-ext) and Enhanced representation through knowledge integration (ERNIE) to enrich text vector information; Secondly, it incorporates numerous encoders to remove text features at multiple levels, thereby increasing text feature information, improving text semantics accuracy, and enhancing the model's learning and generalization capabilities; finally, multi-feature fusion poetry sentiment analysis model is constructed. The feasibility and accuracy of the model are validated through the ancient poetry sentiment corpus. Compared with other baseline models, the experimental findings indicate that SA-Model may increase the accuracy of text semantics and hence improve the capability of poetry sentiment analysis.

KEYWORDS

Sentiment analysis; Chinese classical poetry; natural language processing; BERT-wwm-ext; ERNIE; multi-feature fusion

1 Introduction

1.1 Background

The sentiment is a subjective response of living beings to external value relations and is an integral part of biological intelligence [1]. Chinese classical poetry is short text literature containing many semantic meanings and concise words, also known as the “Twitter” of ancient people's emotions [2]. As natural language processing technology takes off, the research on poetry has gradually received extensive attention. However, the existing research mainly focuses on using deep learning to generate



classic poetry [3], and the research on poetic sentiment analysis is less than others. The current research on poetic sentiment analysis mainly focuses on non-Chinese domains, such as Malay folklore [4] and English poetry [5].

1.2 Literature

Text sentiment analysis, also known as opinion mining, combines natural language processing, computer technique, and linguistic technique to identify, extract and analyze the sentiment embedded in a comment or text [6]. Along with the continuous development of the information times, short text information such as massive comments and microblogs are flooded around people, and short text sentiment analysis has been developed. Short text sentiment analysis may be separated into three components based on the research methodologies used: research-based on sentiment dictionary (SDST-sentiment analysis), research-based on traditional machine learning (TMLST-sentiment analysis), and research-based on deep learning (DLST-sentiment analysis) [7].

SDST-sentiment analysis refers to determining the emotional values of short texts by utilizing the sentiment dictionary. A sentiment dictionary is derived from words representing sentiment attributes in different texts, and the research in Chinese has also been well-developed in recent years, such as NTUSD [8]. The SDST-sentiment analysis is more accurate for the adaptive texts. However, the sentiment dictionary does not cover all sentiment attribute words in this language. In addition, there are few sentiment dictionaries in the minority language, and it is difficult to build them, so SDST-sentiment analysis is somewhat limited.

TMLST-sentiment analysis firstly maps text into multidimensional vectors through feature engineering; secondly, it employs classification models such as Support vector machines (SVM) for feature learning to perform text sentiment categorization [6]. Li et al. [9] applied traditional machine learning methods to short text classification and came up with a Multi-label maximum entropy model (MME) for sentiment classification. TMLST-sentiment analysis improves the accuracy of sentiment analysis by extracting sentiment features and combining them with different classifiers. However, these approaches frequently do not correspond with the text's contextual information, reducing their accuracy.

DLST-sentiment analysis converts text into a word vector matrix by word embedding, effectively weakening feature extraction's complexity [10]. Along with its continuous development, more neural network models are applied in sentiment analysis, such as Recurrent Neural Networks (RNN) [11]. Along with the constant evolution of pre-trained models, such as Bidirectional encoder representation from transformers (BERT) [12] and Embeddings from language models (ELMO) [13], can dynamically acquire word vectors, which can effectively relate the semantics between contexts, increase the model's generalization performance and effectively address the issue of "many interpretations of a word".

1.3 Contribution

Due to the tiny quantity of words in Chinese classical poetry, it has sparse characteristics, and sentiment features are not easy to obtain. In addition, words in Chinese classical poetry often have different meanings in different texts. Existing models are difficult to effectively solve all the problems. We propose a sentiment analysis model (SA-Model) to address existing poetic problems. First of all, our method leads the pre-trained model, which can effectively solve the problem of "different meanings in different texts in word vectors" and uses a dual-channel method to obtain text feature representations from multiple angles and enriches text vector information through fusion. Then, word vectors are feature extracted and fused to enrich the information of poetic text features and improve the accuracy of poetic sentiment analysis. Multi-feature fusion means integrating the different content

from different feature extraction methods to enhance the precision of text sentiment analysis. Multi-feature fusion can effectively and more comprehensively obtain text-related information and helps text vectors can more accurately display the semantic information of the text.

The rest of this paper is presented below. Section 2 described the structure of the proposed model. Section 3 described the analytical results of the model and provides a validation analysis of the model's predictive performance. Finally, Section 4 described the experimental conclusions and future work.

2 Sentiment Analysis Model of Chinese Classical Poetry with Multi-Features Fusion

2.1 Sentiment Analysis Model Construction with Multi-Feature Fusion

We offer a multi-feature fusion sentiment analysis model (SA-Model), and Fig. 1 illustrates the model design. SA-Model is constructed in three parts: word vectors training and fusion, feature information extraction and fusion, and output. We collated the information on the poetic emotions in the database and divided their emotional attributes into three sections: positive, neutral, and negative.

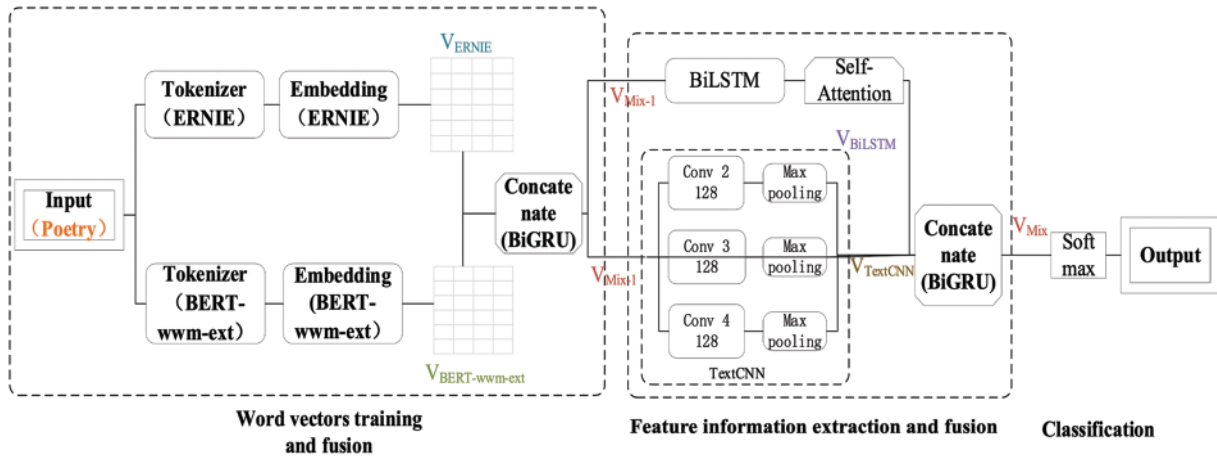


Figure 1: SA-Model architecture diagram

2.2 Word Vectors Training and Fusion

2.2.1 Word Vectors Extraction

Pre-trained language models are derived from neural network models, which can better solve the problem of contextual association by constructing dynamic word vector models, effectively improving the generalization ability of the models, and have been widely used in sentiment analysis models in recent years. Unlike some previous research programs, our method uses BERT-wwm-ext [14] and ERNIE [15] to encode word vectors, resulting in dynamic multi-level information.

SA-Model feeds the poetic sentiment corpus into the BERT-wwm-ext and ERNIE for training, respectively, and BERT-wwm-ext word vector matrix $B_{BERT-wwm-ext}$ ($V_{BERT-wwm-ext}$) and ERNIE word vector matrix E_{ERNIE} (V_{ERNIE}) are obtained as follows:

(1) Training Poetry Text Word Vectors (BERT-wwm-ext)

Bidirectional encoder representation from transformers-Whole word masking (BERT-wwm) [14] uses the Chinese word segmentation tool to split the text by words and imports whole-word masking strategies. BERT-wwm-ext [14] is a revised version of BERT-wwm. The improvement of this model

lies in the increase of training data and steps, which improves the accuracy of related tasks to a certain extent.

BERT-wwm-ext is composed of 12 transformer-blocks, the number of rows n equals the number of processed words in the text, and m equals the dimension of the processed word vector, which is 768. The word vector matrix $B_{BERT-wwm-ext}$ ($V_{BERT-wwm-ext}$) as shown in Eq. (1).

$$B_{BERT-wwm-ext} = \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1m} \\ B_{21} & B_{22} & \dots & B_{2m} \\ \dots & \dots & \dots & \dots \\ B_{n1} & B_{n2} & \dots & B_{nm} \end{bmatrix} \quad (1)$$

(2) Training Poetry Text Word Vectors (ERNIE)

ERNIE [15] has three pre-training tasks (word-aware pre-training tasks, structure-aware pre-training tasks and semantic-aware pre-training tasks), ERNIE can be the more comprehensive acquisition of syntax, semantics information, and so on. And it proposes sequential multi-task learning to solve catastrophic forgetting. However, ERNIE also has a drawback, ERNIE is no sufficient context information [16].

ERNIE consists of 12 transformer-blocks, the number of rows w is the number of processed words in the text, and m equals the dimension of the processed word vector is 768. The matrix E_{ERNIE} (V_{ERNIE}) can be obtained, as shown in Eq. (2).

$$E_{ERNIE} = \begin{bmatrix} E_{11} & E_{12} & \dots & E_{1m} \\ E_{21} & E_{22} & \dots & E_{2m} \\ \dots & \dots & \dots & \dots \\ E_{w1} & E_{w2} & \dots & E_{wm} \end{bmatrix} \quad (2)$$

2.2.2 Word Vectors Fusion

In this paper, based on the feature hierarchy fusion [17], the fusion of word vectors will be achieved by using concatenating [18]. We incorporate the word vectors obtained from BERT-wwm-ext and ERNIE.

(1) Word Vector Input and Preprocessing

The Nth poetry corpus is obtained from the poetry corpus, and the poetry corpus of length L is received after the Tokenizer divides the poetry corpus (Note: the size of L is fixed, and the length of each poetry corpus is L after processing). Finally, we obtain the BERT-wwm-ext word vector $B_{BERT-wwm-ext}$ and ERNIE word vector E_{ERNIE} or the poetry corpus through the Embedding layer.

(2) Word Vectors Fusion

We will use concatenating and Bidirectional gating recurrent unit (BiGRU) [19] to materialize the fusion of text vectors, as shown in Eq. (3). This part implements bi-directional backward and forward state transfer based on BiGRU to improve the accuracy of the fused text vectors and finally get the fused vector V_{mix-1} .

$$V_{mix-1} = BiGRU(concat(B_{BERT-wwm-ext}, E_{ERNIE})) \quad (3)$$

2.3 Feature Information Extraction and Fusion

2.3.1 Feature Information Extraction

SA-Model will use two methods to achieve feature information extraction, they can extract features between contexts and key features of text at different levels, respectively, so our method can get multi-angle information, as follows:

- (1) Feature Information Extraction Based on Bi-Directional Long Short-Term Memory (BiLSTM) and Self-Attention Mechanism

BiLSTM [20] can extract features between contexts and obtain the vector- V_{BiLSTM} : the Long short-term memory (LSTM) cell is shown in Fig. 2. BiLSTM contains two states- the hidden state (HS_t) and the cell state (CS_t) (Note: B_i ($i = OG, IG, FG$) represents the offset, and t means time).

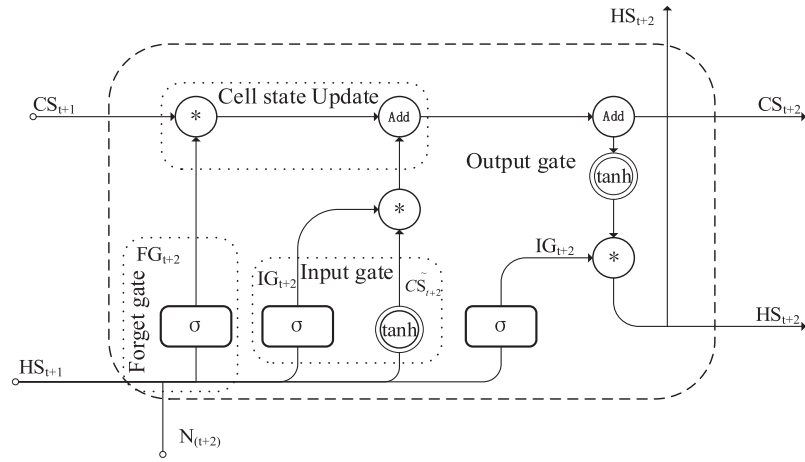


Figure 2: BiLSTM cell architecture

1) Partial Deletion of Historical Cell Status Information

The output gate controls it (FG) aims to remove some of the recorded cell state information, as shown in Eq. (4) (Note: FG_{t+2} is the output result of the forgetting gate)

$$FG_{t+2} = \sigma(W_{FG} \cdot \text{concat}(HS_{t+1}, N_{t+2}) + B_{FG}) \quad (4)$$

2) New Information Input for the Current CS_t

It is controlled by an input gate (IG) that aims to input part of the new information into the current cell state. The input gate consists of two activation functions, as shown in Eqs. (5) and (6).

$$IG_{t+2} = \sigma(W_{IG} \cdot \text{concat}(HS_{t+1}, N_{t+2}) + B_{IG}) \quad (5)$$

$$CS_{t+2}^{\sim} = \tanh(W_{IG} \cdot \text{concat}(HS_{t+1}, N_{t+2}) + B_{IG}) \quad (6)$$

3) Cell State Update (CSU)

The main task of CSU is to calculate/update the CS_t of the current cell, as shown in Eq. (7) (Note: \otimes represents the product).

$$CS_{t+2} = FG_{t+2} \otimes CS_{t+1} + IG_{t+2} \otimes \tilde{CS}_{t+2} \quad (7)$$

4) The Calculation for the Value of the Output Gate (OG_t) and Current HS_t

The primary role is to control the selective entry CS_{t+2} into HS_{t+2} , in other words, the update of the hidden state HS_{t+1} , owned by the output gate (OG), as shown in Eqs. (8) and (9) (Note: \otimes represents the product).

$$OG_{t+2} = \sigma (W_{OG} \cdot \text{concat} (HS_{t+1}, N_{t+2}) + B_{OG}) \quad (8)$$

$$HS_{t+2} = OG_{t+2} \otimes \tanh (CS_{t+2}) \quad (9)$$

The self-attention mechanism can be used for the augmented representation of features. In this paper, the correlation between the input vector sequences will be obtained by the attention mechanism, referred to as the HAN (Hierarchical Attention Networks) [21].

(2) Feature Information Extraction Based on Text Convolutional Neural Network (TextCNN)

TextCNN [22] can effectively extract key features of text at different levels. Based on Vmix-1, the text feature vector V_{TextCNN} will be obtained, the primary structure as shown in Fig. 3.

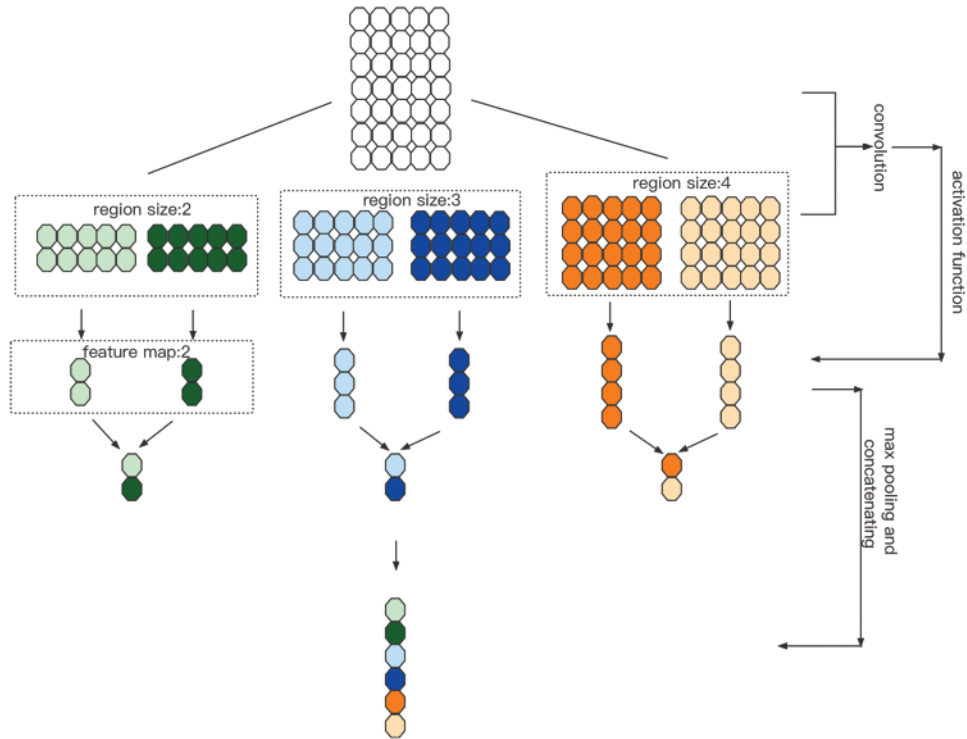


Figure 3: The primary structure of the TextCNN model

In this paper, filter size = (2, 3, 4), number of filter = 128, the width of each convolution kernel is the word vector dimension, and the vector is convolved according to the applicability of the dataset. The superposition of bias unit u , and finally, the activation of this data by the $ReLU$ function to obtain the feature vector V_i , $ReLU$ is the activation function; conv is the convolution operation, as shown in Eq. (10). The pooling layer adopts the max pooling strategy. The max pooling strategy example is shown in Eq. (11), $V_{i-\max}$ represents the result after using the max pooling strategy.

$$V_i = ReLU(conv(V_{mix-1}, n) + u) \quad (10)$$

$$V_{i-\max} = \max_{pooling}(V_i) \quad (11)$$

2.3.2 Feature Fusion

The most divergent information is obtained from the multiple original feature vectors involved in the fusion process, resulting in the learning of a better feature representation [16]. This section will use concatenating and BiGRU [19] to implement the fusion of feature vectors to obtain the vector- V_{mix} . The calculation method is shown in Eq. (12).

$$V_{mix} = BiGRU(concat(V_{BiLSTM}, V_{TextCNN})) \quad (12)$$

2.4 Output

The vector obtained after processing is passed through the softmax classifier to achieve the classification of sentiment. The loss function is set to categorical_crossentropy, X_j corresponds to the classification category, n means output size, and LN means loss, as shown in Eq. (13).

$$LN = - \sum_{j=1}^n x_j \cdot \log \hat{x}_j \quad (13)$$

3 Experiments and Results

3.1 Experimental Data

The primary source of data for this experiment is THUNLP-AIPoet [23]. The sentiment attributes of each poem are organized into three categories: positive, neutral, and negative.

3.2 Evaluation Standards

The evaluation metrics for the experiments in this paper consist of three components: Macro-P, Macro-R, and Macro-F1. Macro-F1 is a multi-categorization effect evaluation index based on Precision and Recall, which can be used to evaluate the model comprehensively. It takes values between 0 and 1, and the worst performance effect takes the value of 0 and vice versa. Macro-P and Macro-R are associated with Precision and Recall, respectively. The form of sentiment analysis results constructed according to the sentiment analysis dataset is shown in Table 1.

Table 1: Emotional analysis result form

Emotional classification situation	Emotion classification before the text belongs to the category of emotion	Emotion classification before the text does not belong to the category of emotion
The text belongs to this category	TP	TN
The text does not belong to this category	FP	FN

3.3 Experimental Parameters

The experimental environment of this paper is shown in [Table 2](#).

Table 2: Experiments environment table

Lab environment	Configuration
GPU	NVIDIA Quadro RTX 6000
Deep learning framework	Tensorflow 2.2.0
Operating system	Windows 10
Programming languages	Python 3.6
Software environment	Pycharm

Details parameters about Pre-trained language models are listed in [Table 3](#).

Table 3: Details parameters (Pre-trained language models)

Model name	Parameter	Value
BERT-wwm-ext	Number of layer	12
	Hidden size	768
	Heads	12
	Hidden act	ReLU
ERNIE	Number of layer	12
	Hidden size	768
	Heads	12
	Hidden act	GELU

In this paper, BiLSTM and TextCNN are used for feature extraction, BiGRU is used for fusion, model's details parameters are listed in [Table 4](#).

Table 4: Details parameters (Pre-trained language models)

Model name	Parameter	Value
TextCNN	Filter size	(2,3,4)
	Number of filter	128
	Batch size	64
	Learning rate	5E-5
BiLSTM	Hidden size	256
	Batch size	64
BiGRU	Learning rate	5E-5
	Hidden size	256

3.4 Experimental Protocol and Analysis of Results

Five simulations are conducted for each model to ensure the efficacy of the outcomes, and the mean values are taken.

In this paper, four sets of comparative experiments are designed to verify the experimental effects of the poetic sentiment analysis model on the poetic dataset. The first set of experiments is used to verify the performance of the word vector model in this paper by enriching text vector information and improving the accuracy of text analysis; the second set of experiments demonstrates the superiority of the feature fusion model in the poetic sentiment analysis model by enriching text feature information based on the first set of experiments. The third set of experiments is used to exhibit the superiority of Macro-F1 in every epoch for the sentiment analysis model based on the second set of experiments. The last experimental design is to demonstrate the performance of the model in this paper through the visualization of the confusion matrix.

(1) Word Vectors Sentiment Analysis Experiments and Results

Based on the widely used word vector model and the approach of the ERNIE model, we design the experimental scheme. The models are as follows: BERT-wwm-ext [14], MacBert [14], RoBERTa [24], ERNIE [15], and the model which is from SA-Model.

- 1) BERT-wwm-ext [14]: This model is derived from the BERT-wwm model, which can effectively improve the accuracy of the related tasks through Whole Word Masking (WWM).
- 2) MacBert [14]: The model uses a correction for similar words for the Masking, effectively reducing the problem between pre-training and fine-tuning the model.
- 3) RoBERTa [24]: This model improves the pre-training method based on the BERT model by using more data for model training, dynamic Masking, and larger mini-batch to enhance the model's performance. In this paper, we use RoBERTa for Chinese.
- 4) ERNIE [15]: The method is applied to SA-Model. This method ERNIE model contains three masking strategies: Basic-level Masking, Entity-level Masking, and Phrase-level Masking, which integrate multi-level semantic information in training and improve the accuracy of word vectors.

- 5) BERT-wmm-ext_ERNIE_concat_BiGRU (concat(SA-Model)): the text vector extraction and fusion approach is derived from SA-Model.

Table 5 and Fig. 4 exhibit the experimental results on three evaluation indicators. Specifically, in the task of sentiment analysis, our method has a manifest improvement, with an improvement of 0.7% over the best model (ERNIE) in Macro-F1, 0.28% over the best model (ERNIE) in Macro-R, and so on. We believe that SA-Model is significantly ahead of other single pre-trained models in the sentiment analysis task by enriching the text vector information.

Table 5: Comparison of different models Macro-P, Macro-R, and Macro-F1 (No. 1)

Model name	Macro-P↑	Macro-R↑	Macro-F1↑
BERT-wmm-ext	58.72%	55.76%	54.49%
MacBERT	57.85%	55.53%	55.60%
RoBERTa	58.23%	54.16%	53.40%
ERNIE	66.52%	63.39%	62.88%
Concat (SA-model)	66.53%	63.67%	63.58%

(2) Experiments and Results of the Feature Fusion Model

We develop many experimental techniques to evaluate our model's efficacy. The models we apply are as follows: BERT-CMCNN [25], BERT+LSTM [26], BERT_ERNIE_BiGRU [16], BERT-DCA [27], and our model.

- 1) BERT-CMCNN [25]: This model consists of BERT and BiLSTM as well as the CNN model to improve sentiment polarity classification effectively.
- 2) BERT+LSTM [26]: This model consists of BERT and the LSTM model, which significantly enhances text feature extraction capability.
- 3) BERT_ERNIE_GRU [16]: This method uses a BERT model and an ERNIE model for text vector fusion using TextCNN to extract text features and direct concatenating fusion. Splicing and gating cycle unit (Gated Recurrent Unit, GRU) are used, respectively.
- 4) BERT-DCA [27]: This method uses a channel complex based on BERT, a two-channel attention mechanism, and BiGRU to extract semantic features.
- 5) SA-Model: The model is proposed by us.

The efficiency of different sentiment analysis models is compared and examined, and the text of the experimental findings is summarized in Table 6.

Table 6 and Fig. 5 exhibited the experimental results on three evaluation indicators, which demonstrates that our method is significantly improved compared with other pretrained models. Compared with other best models, our method increases by 0.95%, 1.18%, and 1.94% respectively on the corresponding assessment indicators (Macro-P, Macro-R, and Macro-F1). It shows that the performance of the model after word vector feature fusion is better than that of a single vector model. In addition, the performance of the model after feature fusion is also better than that of a single feature information extraction model. In summary, the experimental findings demonstrate that SA-Model can outperform other models in the sentiment analysis task.

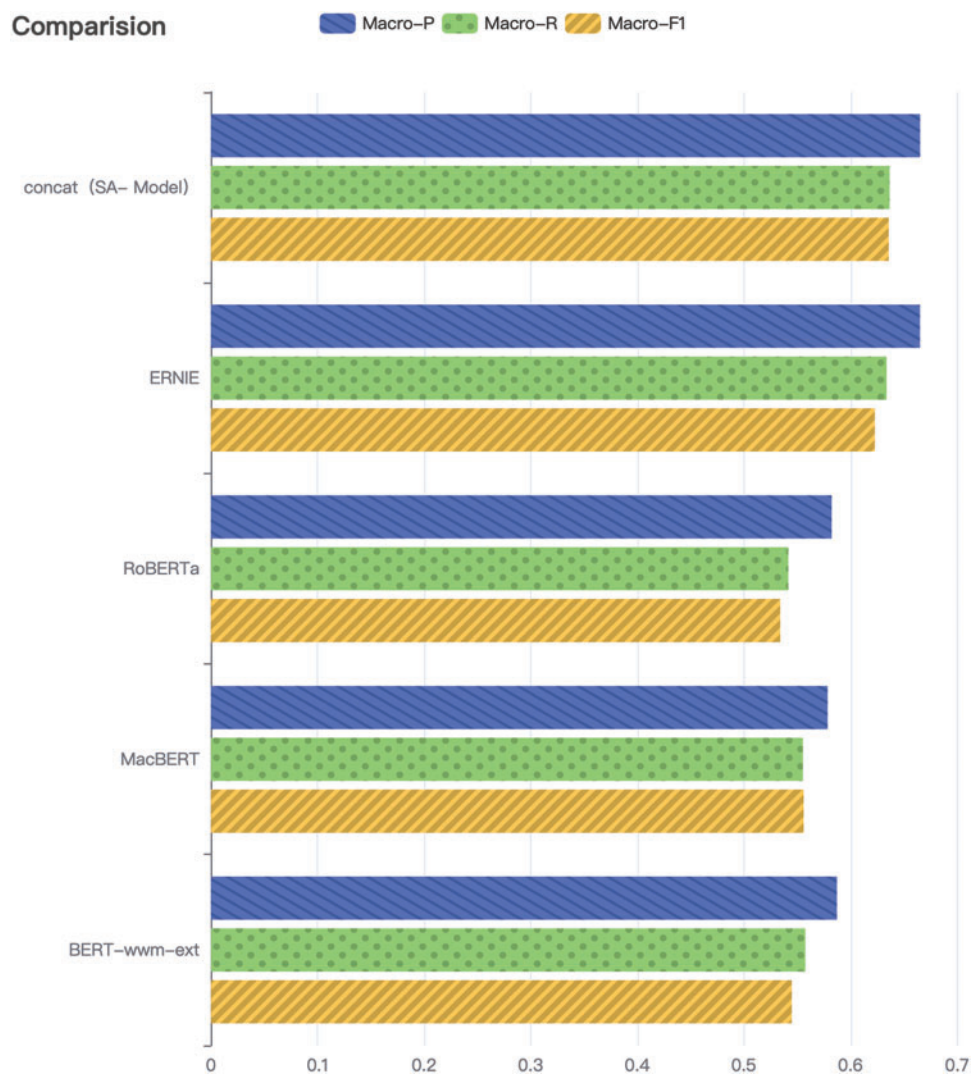


Figure 4: Comparative view of the results of Macro-P, Macro-R, and Macro-F1 in this experiment (No. 1)

Table 6: Comparison of different models Macro-P, Macro-R, and Macro-F1 (No. 2)

Model name	Macro-P↑	Macro-R↑	Macro-F1↑
Bert-CMCNN	60.03%	55.08%	55.06%
BERT+LSTM	59.18%	54.34%	53.62%
BERT_ERNIE_GRU	66.38%	63.71%	62.43%
BERT-DCA	58.47%	54.88%	54.17%
SA-model	67.36%	64.89%	64.38%

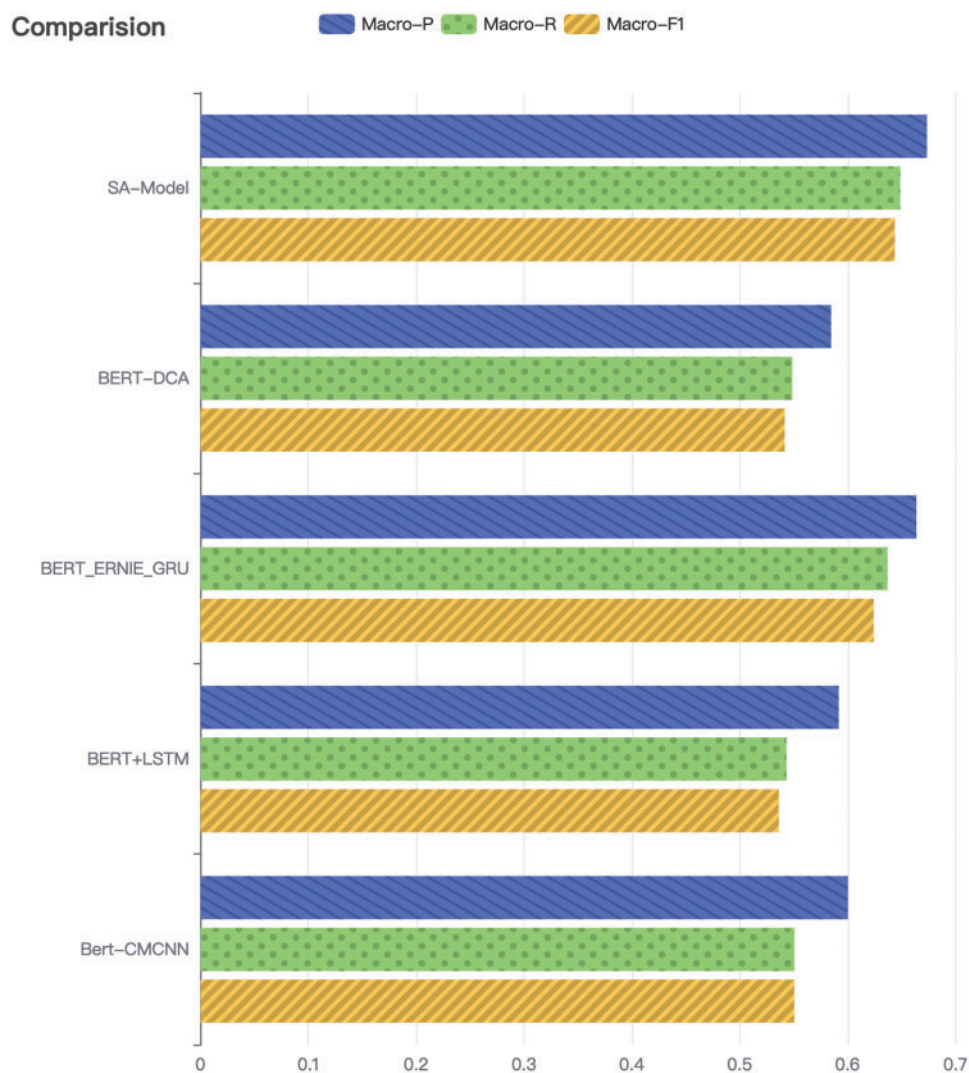


Figure 5: Comparative view of the results of Macro-P, Macro-R, and Macro-F1 in this experiment (No. 2)

(3) Line Chart of Comparation

Line chart will be created to visualize the change over the number of Macro-F1 in every epoch for the sentiment analysis model. From Fig. 6, using the first 15 epochs as an example, based on the previous set of experimental information, the performance of our method shows better than other sentiment analysis models, which shows that with multi-feature fusion, the performance of at each stage has a certain advantage, although the Macro-F1 value in a small number of epochs is lower than that of experiment 3).

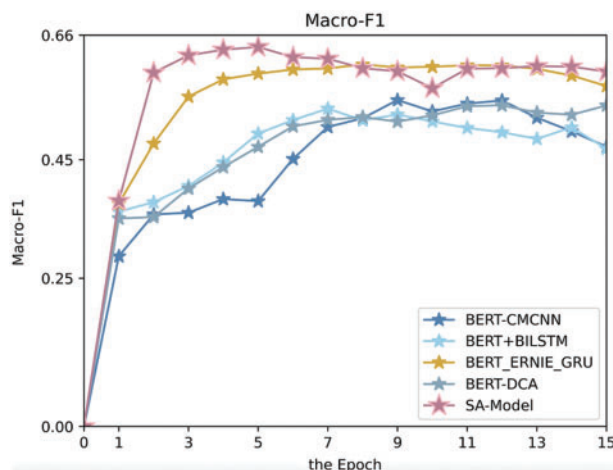


Figure 6: Macro-F1 comparisons for each model across the first 15 epochs

(4) Confusion Matrix of SA-Model

The efficiency of SA-Model is proven by showing the confusion matrix, as shown in Fig. 7. The output contains three probabilities: the probabilities of positive, neutral and negative, which correspond to the number 0, 1, and 2. The model can distinguish positive and negative emotions more accurately. Although it is not as apparent as negative and positive emotions in processing neutral emotions, it can effectively determine the sentiment of the text data of poems.

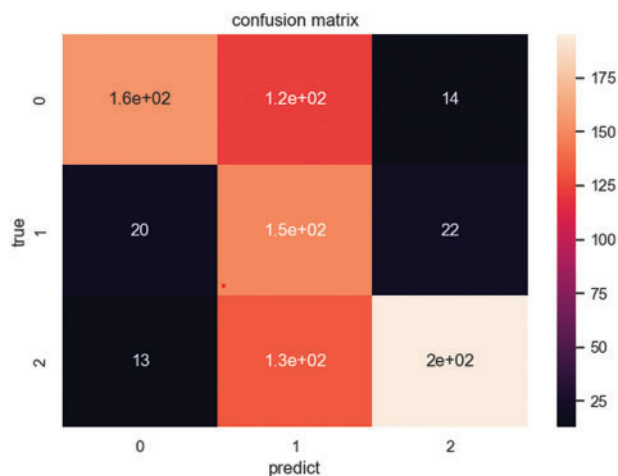


Figure 7: Visualization of the confusion matrix

4 Conclusion

4.1 Research Content and Results

Aiming at the existing problems, this paper proposes a poetic analysis model called SA-Model. We use BERT-wwm-ext and ERNIE to dynamically extract word vectors, which can successfully solve the problem of “polysemy” in poetry texts and obtain word vector information that is more consistent with poetry texts. In addition, they are also used to receive information from multiple angles, enrich the

content of text vector information, and obtain text differences; the multi-level extraction of text feature information based on BiLSTM, self-attention mechanism, and TextCNN improves the accuracy of text sentiment analysis. Therefore, SA-Model effectively solves the problem of sparse features and improves the accuracy of text sentiment analysis.

The experiment results indicate that: the SA-Model extracts text features at various levels gathers text information from numerous viewpoints, and efficiently increases poetry sentiment analysis accuracy.

4.2 Future Directions

In future research, we may consider combining poetic tone and other pre-trained models with sentiment analysis of poetic texts. Additionally, we may investigate merging visual analysis approaches with sentiment analysis techniques to increase research efficiency and validate model results.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4), 344–350. <https://doi.org/10.1511/2001.4.344>
2. Wu, B., Ji, J., Meng, L., Shi, C., Zhao, H. D. et al. (2016). Transfer learning based sentiment analysis for poetry of the Tang dynasty and Song dynasty. *Acta Electronica Sinica*, 44(11), 2780.
3. Tang, Y., Wang, X., Qi, P., Sun, Y. (2020). A neural network-based sentiment analysis scheme for Tang poetry. *2020 International Wireless Communications and Mobile Computing (IWCMC)*, pp. 1783–1788. Jbail, Lebanon.
4. Saad, M. M., Jamil, N., Hamzah, R. (2018). Evaluation of support vector machine and decision tree for emotion recognition of Malay folklores. *Bulletin of Electrical Engineering and Informatics*, 7(3), 479–486. <https://doi.org/10.11591/eei.v7i3.1279>
5. Sreeja, P. S., Mahalakshmi, G. S. (2016). Comparison of probabilistic corpus based method and vector space model for emotion recognition from poems. *Asian Journal of Information Technology*, 15(5), 908–915.
6. Zhang, X. Y., Bai, Y. (2022). Chinese online comments sentiment analysis based on weighted char-word mixture word representation. *Application Research of Computers*, 39(1), 31–36.
7. Feng, Y., Cheng, Y. (2021). Short text sentiment analysis based on multi-channel CNN with multi-head attention mechanism. *IEEE Access*, 9, 19854–19863. <https://doi.org/10.1109/ACCESS.2021.3054521>
8. Ku, L. W., Chen, H. H. (2007). Mining opinions from the Web: Beyond relevance retrieval. *Journal of the American Society for Information Science and Technology*, 58(12), 1838–1850. <https://doi.org/10.1002/asi.20630>
9. Li, J., Rao, Y., Jin, F., Chen, H., Xiang, X. (2016). Multi-label maximum entropy model for social emotion classification over short text. *Neurocomputing*, 210(9), 247–256. <https://doi.org/10.1016/j.neucom.2016.03.088>
10. Jiang, S. Y., Guo, L. D., Wang, L. X., Fu, S. H. (2018). Survey on opinion target extraction. *Acta Automatica Sinica*, 44(7), 1165–1182.

11. Cheng, J., Li, P., Ding, Z., Zhang, S., Wang, H. (2016). Sentiment classification of Chinese microblogging texts with global RNN. *2016 IEEE First International Conference on Data Science in Cyberspace (DSC)*, pp. 653–657. Changsha, China.
12. Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
13. Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C. et al. (2018). Deep contextualized word representations. arXiv preprint arXiv:1802.05365.
14. Cui, Y., Che, W., Liu, T., Qin, B., Yang, Z. (2021). Pre-training with whole word masking for Chinese BERT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 3504–3514. <https://doi.org/10.1109/TASLP.2021.3124365>
15. Sun, Y., Wang, S., Li, Y., Feng, S., Tian, H. et al. (2020). Ernie 2.0: A continual pre-training framework for language understanding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(5), 8968–8975. <https://doi.org/10.1609/aaai.v34i05.6428>
16. Jie, C., Jing, M., Li, X. F. (2021). Short-text classification method with text features from pre-trained models. *Data Analysis and Knowledge Discovery*, 5(9), 21–30.
17. Zhang, X. C., Yu, L. F., Sang, R. T., Zhang, Y. H. (2018). A study of the short text classification with CNN and LDA. *Software Engineering*, 21(6), 17–21.
18. Nie, W., Chen, Y., Ma, J. (2019). A text vector representation model merging multi-granularity information. *Data Analysis and Knowledge Discovery*, 3(9), 45–52.
19. Cui, X., Long, H., Xion, X., Shao, Y., Du, Q. (2020). Chinese text sentiment classification based on parallel bi-directional gated recurrent unit and self-attention. *Journal of Beijing University of Chemical Technology*, 47(2), 115.
20. Jin, C., Li, W. H., Ji, C., Jin, X. Z., Guo, Y. P. (2018). Chinese word segmentation based on bidirectional LSTM neural network model. *Journal of Chinese Information Processing*, 32(2), 29–37.
21. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A. et al. (2016). Hierarchical attention networks for document classification. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1480–1489. San Diego, California, USA.
22. Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.
23. Chen, H., Yi, X., Sun, M., Li, W., Yang, C. et al. (2019). Sentiment-controllable Chinese poetry generation. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*, pp. 4925–4931. Macao, China.
24. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M. et al. (2019). Roberta: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.
25. Hu, R. Y., Liu, J. H., Bu, G. N., Zhang, D. Y., Luo, Y. X. (2021). Research on sentiment analysis of multi-level semantic collaboration model fused with BERT. *Computer Engineering and Application*, 57(13), 176–184.
26. Yang, J., Yang, W. J. (2021). Text comment sentiment analysis based on BERT model. *Journal of Tianjin University of Technology*, 37(2), 12–16.
27. Xie, R. Z., Li, Y. (2020). Text sentiment classification model based on BERT and dual channel attention. *Journal of Tianjin University of Technology*, 35(4), 642–652.