Proceedings of CONF-CDS 2025 Symposium: Data Visualization Methods for Evaluation
DOI: 10.54254/2755-2721/2025.PO24450

Chinese Sentiment Analysis Based on CNN-BiLSTM-Attention Model

Haihao Zhou

School of Cyber Science and Engineering, Wuhan University, Wuhan, China
akatopiliy@outlook.com

Abstract. With the widespread proliferation of the Internet, individuals are increasingly inclined to express their opinions and comments on various matters online. This paper will discuss the feasibility of applying the Attention-Based CNN-BiLSTM model (ABCBM) to Chinese sentiment analysis. In order to demonstrate that the model can effectively adapt to the task, it is essential to select a high-quality dataset. This paper has chosen the WeiboSenti100k dataset, which contains 100,000 comments posted on Weibo. After splitting the dataset into train texts and test texts randomly, this study uses the model to capture the Semantic information inside the texts. The result of the experiment shows that this model can manage the Chinese sentiment analysis well, achieving 98% accuracy, 97% recall and 98% F1 score in classifying texts as positive and negative, which proves the capacity that ABCBM holds in clarifying the sentiment inside the Chinese language.

Keywords: CNN, BiLSTM, Attention, ABCBM, WeiBoSenti-100k

1. Introduction

With the advent of the information age, the venues for self-expression have gradually shifted from the physical realm to the digital space. An increasing number of individuals now prefer to articulate their attitudes and emotions towards various matters through online platforms. During this digital era, the modes of expression have become increasingly diverse, and correspondingly, the emotional inclinations of individuals towards various subjects have also expanded in complexity. With the continuous advancement of artificial intelligence, sentiment analysis of text has emerged as a prominent area of research within the AI domain. Specifically, the classification of textual sentiments has become one of the key metrics for evaluating the performance of a model. An increasing number of models have been applied to sentiment analysis, and many of them have achieved commendable results. However, due to the unique characteristics of Chinese characters, the distinctive semantic information embedded in Chinese texts differentiates Chinese sentiment analysis from that of other languages. This is primarily because Chinese characters possess inherent linguistic features such as polysemy, complex syntactic structures, and implicit emotional expressions. For instance, the same Chinese word may convey different emotional connotations depending on the context. Additionally, Chinese emotional expressions can be highly diverse, including direct statements, metaphors, and irony. These characteristics pose significant challenges for large language models when performing sentiment analysis on Chinese texts.

DOI: 10.54254/2755-2721/2025.PO24450

This paper introduces an Attention-Based CNN-BiLSTM model to manage the problem of Chinese sentiment analysis. The bidirectional long short-term memory (BiLSTM) network is capable of capturing the semantic relationships of text in both forward and backward directions. Combined with the convolutional neural network (CNN) and attention mechanisms, which further extract semantic features from the text. This integrated model is believed to be highly effective for sentiment prediction of Chinese text. To demonstrate the effectiveness of the proposed model, this study selected the Weibo-Senti-100k dataset, which contains 100,000 Chinese comments posted on Weibo, the same website as Twitter. The application of the model in this study is divided into three distinct phases: data loading and processing, model network construction, and training and evaluation. In the data processing phase of this study, the Pandas library was used to load the dataset. Subsequently, the Jieba library was employed for the tokenization of the text data, which is essential for handling the unique characteristics of the Chinese language. Finally, the LabelEncoder was applied to encode the labels, thereby transforming categorical data into a numerical format suitable for model training. During the model construction phase, the architecture is composed of three primary layers. Initially, this study employed Word2Vec to construct the word embedding layer, which generates word vectors. These word vectors are then fed into the BiLSTM layer to produce hidden states at each time step. The output from the BiLSTM layer is subsequently passed through the Attention layer to generate weights for each word. Finally, the CNN layer is utilized for feature extraction, and the extracted features are concatenated with attention weights and further used to generate the prediction results. The constructed model is subsequently subjected to rigorous training and evaluation processes to ascertain its performance, culminating in the derivation of the model's score. This paper ultimately aims to demonstrate the superior performance of the proposed model in the realm of Chinese sentiment analysis, thereby contributing to the advancement of research and understanding of BiLSTM models, attention mechanisms, and convolutional neural networks.

2. Related work

Previous researchers have employed a variety of methods for Chinese sentiment analysis [1]. This paper particularly focuses on deep learning techniques that related to this research. Long Short-Term Memory (LSTM) networks and their extension, Bidirectional Long Short-Term Memory (BiLSTM) networks, have been widely recognized for their effectiveness in capturing semantic information. These models selectively forget certain features to enhance semantic understanding. In the context of Chinese sentiment analysis, LSTM was applied in 2017 by Day and Lin [2], while BiLSTM was utilized early in 2016 by Wang et al. [3]. Apart from basic LSTM and BiLSTM networks, as early as 2015, researchers had combined LSTM with tree structures for Chinese sentiment analysis [4]. In addition to LSTM and BiLSTM, Convolutional Neural Networks (CNNs) have also been widely applied in Chinese sentiment analysis due to their remarkable performance in feature extraction [5,6,7]. With the advancement of research, it was soon realized that integrating multiple methods in an appropriate manner could yield better performance. Researchers quickly sensed the potential of combining BiLSTM, which excels at capturing contextual semantic information, with the then-popular multi-head attention mechanism. This integration leverages the strengths of both approaches: BiLSTM captures bidirectional contextual dependencies, while multi-head attention allows the model to focus on different aspects of the input sequence simultaneously, thereby enhancing the model's ability to understand complex semantic relationships [8]. Similarly, researchers have also explored the integration of CNN with LSTM or BiLSTM, achieving promising results [9,10]. This combination leverages the strengths of both architectures: CNN's ability to

DOI: 10.54254/2755-2721/2025.PO24450

capture local features and LSTM/BiLSTM's capacity to model temporal dependencies. Also, in 2022, Jia tried to combine the BERT model with the CNN and the attention mechanism. This attempt aims to combine the comprehension ability of the BERT model with the feature extract ability of the CNN and the attention mechanism. By combining these approaches, the models can effectively extract both spatial and temporal features from the text data, leading to improved performance in sentiment analysis tasks. The research methodology of this paper primarily draws inspiration from a study published by Gan Chenquan in 2021 [10], which integrates word attention mechanisms with CNN and BiLSTM networks. This model leverages the strengths of each component: CNN captures local features, BiLSTM models bidirectional temporal dependencies, and the attention mechanism assigns different weights to critical information, thereby enhancing the model's ability to understand and classify sentiments in Chinese text. However, their work focused on multi-type Chinese sentiment classification while this paper applying the similar but simpler model to Chinese sentiment binary classification.

3. Methods

This part will introduce the architecture of the model. The model can be introduced in five layers, which are Word Embedding Layer, BiLSTM Layer, Attention Layer, CNN layer and Output Layer. The preprocessed data is first transformed into word vectors through the Word Embedding Layer. It then passes through the BiLSTM Layer for the first semantic extraction, followed by further processing by the Attention Layer and the CNN Layer. The final result is processed through the Output Layer to get the final sentiment classification result.

3.1. Word embedding layer (including the preprocess of the datasets)

The dataset utilized in this paper consists of Chinese text and a label indicating the sentiment orientation. To employ the model for sentiment prediction, the initial step involves transforming the text into word vectors. After reading in the Chinese text data, word segmentation has always been a significant challenge in Chinese text processing. Here, we utilize the Jieba library for Chinese word segmentation. Jieba is a widely popular and highly efficient word segmentation method in the field of Chinese text processing. Following word segmentation, we employ the Word2Vec model for word vector generation and subsequently create a vocabulary table. Initially, we performed a train-test split on the WeiboSenti100k dataset, allocating 60% of the data to the training set. The remaining 40% was further divided equally, with half serving as the test set and the other half as the validation set. Subsequently, we utilized the training set to train a Word2Vec model, which generated word vectors for the vocabulary. Ultimately, each segment of textual information is transformed into a matrix of word vectors like vector A.

A[x_1,x_2...,]

3.2. BiLSTM layer

The word vectors are first processed in the Bidirectional Long Short-Term Memory (BiLSTM) layer. Compared to the classic RNN model, BiLSTM adds control gates such as the forget gate, enabling it to better capture textual information. Compared to the LSTM model, its bidirectional structure allows it to better capture the semantic relationships in both preceding and following contexts. The specific structure of the BiLSTM model is shown in Figure 1 (mentioning that the 'Very Good' in the figure refers to Chinese text in our actual training process). The BiLSTM is composed of

DOI: 10.54254/2755-2721/2025.PO24450

bidirectional LSTM networks, allowing it to capture semantic information from both preceding and following contexts due to its bidirectional structure. The LSTM network contains three gate mechanisms—namely, the forget gate, output gate, and input gate—that control the output of each LSTM cell, which ensures better outperform than the RNN networks.

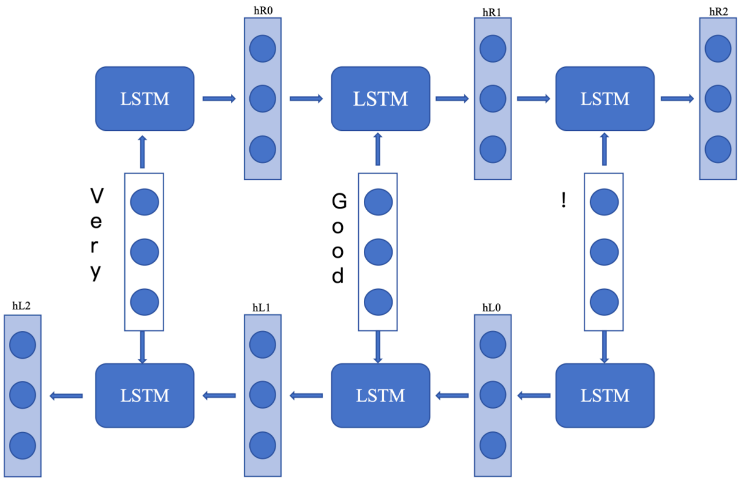

Figure 1. Structure of the BiLSTM model

To better understand how the LSTM networks works, To better understand how the LSTM network operates, let's denote c_t as the cell state at time step t . And h_t as the hidden state at the same time step. In every time steps, h_{t-1} will be dealt together with the input texts x_t . these signals will be processed sequentially through the forget gate, input gate, and output gate, shown in Figure 2. The forget gate(F_t) and input gate(I_t) jointly control whether the information from this time step is incorporated. The control signals are generated through the formula

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f), I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i).$$

For instance, when $F_t = 1$, $I_t = 0$ the information in this time step will be completely discarded. Specifically, the input gate does not directly control whether information enters the model, but rather it controls whether information is incorporated into the candidate memory cell. The candidate memory cell changes by the formula

$$\widetilde{\mathrm{C}}_{\mathrm{t}} = \tanh\left(\mathrm{X_t W_{xc}} + \mathrm{H_{t-1} W_{hc}} + \mathrm{b_c}\right)$$

The output of the LSTM is controlled by the output gate, which similarly generates control signals through a formula

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o)$$

DOI: 10.54254/2755-2721/2025.PO24450

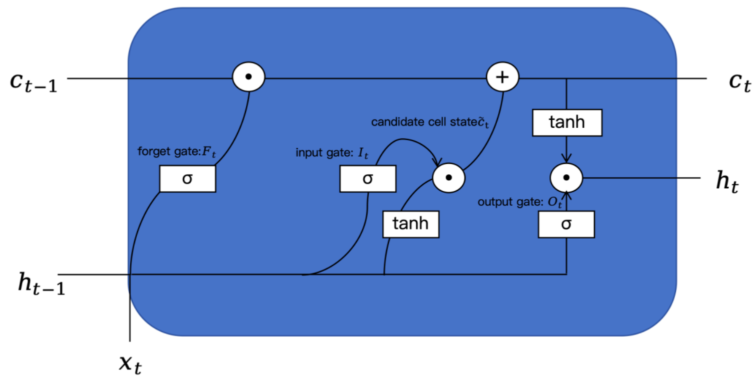

Figure 2. Detailed structure of LSTM model

The hidden states generated at each time step are eventually concatenated with their corresponding hidden states from the opposite direction, resulting in the final hidden state, shown in Figure 3. The final output is then passed to the Attention layer for further processing. Through this bidirectional processing approach, the BiLSTM model is able to better capture the contextual semantic relationships within the text.

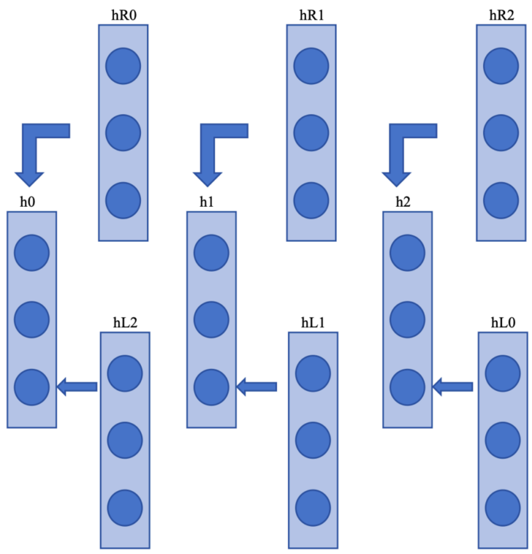

Figure 3. How the final hidden state is generated

3.3. Attention layer

The attention mechanism is a widely used method in natural language processing tasks and has been proven through extensive practice to be highly effective in extracting key features. In this paper, we introduce a learnable parameter v, which continuously evolves during the training process and is used to generate attention weights. The attention mechanism in this paper was shown in Figure 4.

DOI: 10.54254/2755-2721/2025.PO24450

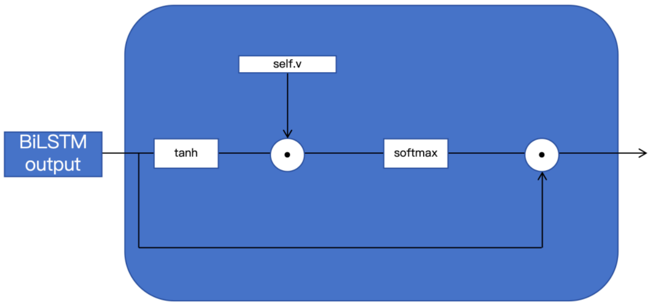

Figure 4. Attention mechanism

The output from BiLSTM layer is handled by a linear layer, transforming its dimension from [batch_size ,seq_len, lstm_hidden_size] to [batch_size, seq_len, attention_size] for further processes. Subsequently, the output is pointwise multiplied by the learnable parameter matrix v , yielding the attention scores for each time step of the BiLSTM layer. The scores are then handled by a softmax layer, promising the scores fall within the range of 0 to 1. Finally, the scores are multiplied by the corresponding BiLSTM outputs and summed to obtain a context vector that incorporates the semantic weight relationships.

3.4. CNN layer

Convolutional Neural Network (CNN) is a classic and highly effective type of network in natural language processing. It captures important textual semantic information by performing convolutional compression on the semantic features. In this paper, we apply conv1d to construct the CNN layer. The input to the CNN layer, which comes from the output of the BiLSTM layer, is first processed by a 1D convolutional layer and then passed through a max-pooling layer to extract its features.

3.5. Output layer

The results from CNN layer and Attention layer are concatenated, and passed through a linear layer, generating the final result that is ready for classification. After concatenating the outputs from the CNN layer and the Attention layer, the model is able to not only reflect the features extracted by both layers but also enhance the weights of sentiment-related words. This allows the model to better predict the semantic information of the text. The outlook of the model is shown in Figure 5 (mentioning that the 'Text' in the figure refers to Chinese text in our actual training process).

DOI: 10.54254/2755-2721/2025.PO24450

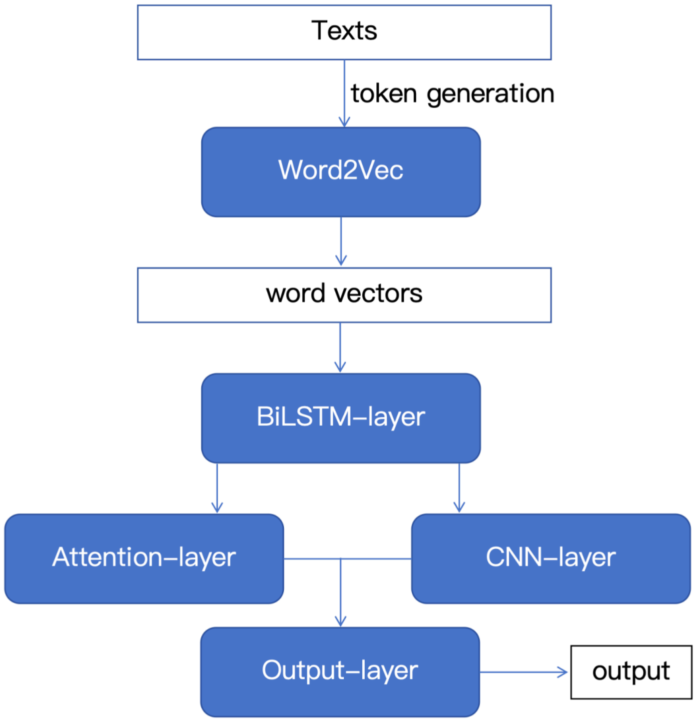

Figure 5. Outlook of the model

4. Experiments

This section will describe the methods used to evaluate our model's performance, providing a comprehensive demonstration of our model's capabilities.

4.1. Evaluation criteria

To clearly explain the calculation of our evaluation scores, this study defines the number of correctly predicted positive and negative instances as TP and TN , respectively, while the number of incorrectly predicted instances are denoted as FP and FN .

Accuracy(Acc) is the proportion of correctly predicted instances out of all samples,

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

Recall(R)demonstrates the proportion of correctly classified positive samples in the total number of positive samples,

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$$

Precision (P) is the proportion of correctly classified positive samples in the total number of samples predicted to be positive, however, this paper only uses the Precision for the calculation and explanation of the F1 score, without computing the Precision separately.

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$$

The F1 is the harmonized average of precision and recall,

$$\mathrm{F1} = \mathrm{P} \times \mathrm{R}$$

These metrics are widely recognized evaluation criteria in the field of sentiment analysis and can more specifically demonstrate the effectiveness of the model presented in this paper.

DOI: 10.54254/2755-2721/2025.PO24450

4.2. Datasets

This paper utilizes the WeiboSenti-100k Chinese sentiment binary classification dataset, which contains almost 120000 comments collected from Weibo. These comments encompass a broad range of expressions found on the Chinese internet, with each text entry accompanied by a corresponding sentiment label: 1 for positive sentiment and 0 for negative sentiment. Weibo is one of the largest social network sites on the Chinese internet, where comments essentially cover a wide range of emotional expression methods used on the Chinese internet. This dataset is suitable for training models and effectively demonstrates the feasibility of the model in the field of Chinese sentiment analysis.

Before feeding the dataset to the model for processing, it is necessary to preprocess the data. In this paper, we use the Jieba segmentation library to tokenize the sentences in the dataset, after which the tokenized data is passed to the model for further processing.

4.3. Training details

Adam is applied to train the network, with the learning rate set to 0.00001. Other factors are set to default. The vocab size is set to 10000, which can cover most of the words that are important for sentiment analysis. This section introduces the hyperparameters used in our model, which are carefully chosen to optimize performance and ensure efficient training. The dimensionality of the embedding layer is 128, which maps each word into a 128-dimensional vector space for semantic representation. The hidden dimension of the LSTM layer is set to 64, controlling the capacity of the LSTM to capture sequential dependencies. The kernel size of the 1D convolutional layer is 3, defining the width of the convolutional filters used to extract local features. The number of output channels in the CNN layer is 64, determining the number of feature maps generated by the convolutional operation. The dimensionality of the attention mechanism is 64, enabling the model to focus on relevant parts of the input sequence. The batch size is set to 32, balancing computational efficiency and model convergence during training. The model is trained for 30 epochs to allow sufficient learning over the entire dataset. To enhance the performance of the model, this paper introduces a warm-up step. We designed a linear learning rate warm-up scheduler, where the learning rate is linearly increased over a set number of steps and then linearly decreased back to 0. This warm-up phase enhances the model's performance and improves its fitting capability. The main loss function of training is the cross-entropy loss.

4.4. Results

This study experimented with various batch sizes and learning rates, and ultimately achieved the best performance with a batch size of 32 and a learning rate of 0.00001. An accuracy of 0.9824, a recall value of 0.9676, and an F1 score of 0.9821 clearly demonstrate the model's outstanding capability in Chinese sentiment analysis. The model achieved a good fitting result and a good score within the first twenty epochs. The training process was carefully monitored, and the detailed results are illustrated in Figure 6. This figure provides insights into the model's learning trajectory, including the evolution of the loss function and the convergence behavior during training. From this figure, we can see that the model achieves excellent results. This not only demonstrates the high efficiency of training with this dataset but also confirms the model's strong performance in the field of Chinese sentiment analysis.

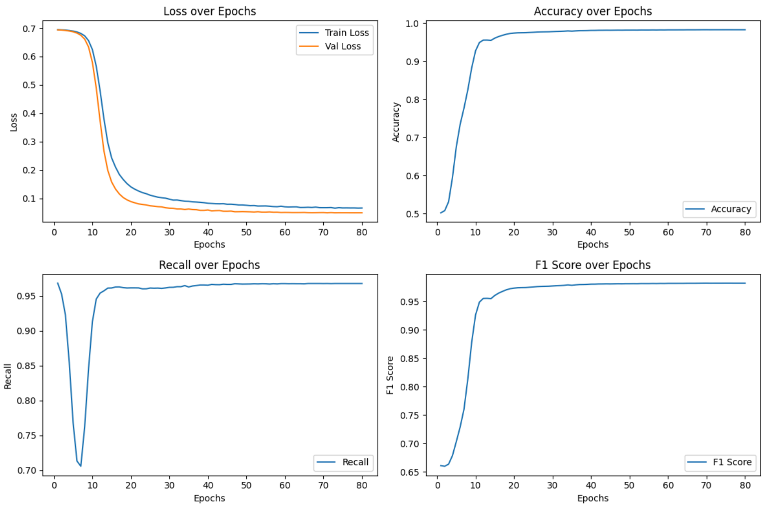

Figure 6. Detailed training results

5. Conclusion

In this paper, we introduce a model based on the attention mechanism, BiLSTM, and CNN, and through designed experiments, successfully demonstrate its superior performance in Chinese sentiment analysis. By leveraging the powerful feature extraction capabilities of the attention mechanism and CNN, along with the BiLSTM model's exceptional ability to capture semantic information, our model achieves an accuracy (ACC) of 98%, a recall rate of 97%, and an F1 score of 97.5% on Chinese sentiment analysis tasks. These results fully validate the rationality and effectiveness of the proposed model architecture, and contribute to the advancement and application of BiLSTM, CNN, and attention mechanisms in the field. However, due to the complexity of Chinese text and the multiple semantic meanings contained within Chinese characters and words, this model struggles to achieve perfect prediction accuracy. This indicates that there is still room for improvement in the model.

Looking ahead, there are several promising directions for future work. Firstly, exploring more advanced techniques for handling the polysemy and complexity of Chinese characters could further enhance model performance. Secondly, incorporating external knowledge sources such as knowledge graphs or lexicons might provide additional context to improve prediction accuracy. Additionally, experimenting with different architectures or hybrid models that combine the strengths of various deep learning methods could lead to breakthroughs in capturing nuanced semantic information. Lastly, refining preprocessing steps, such as better word segmentation and feature engineering, could also contribute to more accurate predictions. By addressing these areas, the author aims to build even more robust and effective models for Chinese sentiment analysis.

References

[1] Wang, Z., Huang, D., Cui, J., Zhang, X., Ho, S. B., & Cambria, E. (2025). A review of Chinese sentiment analysis: subjects, methods, and trends. Artificial Intelligence Review, 58(3), 75.
[2] Day, M. Y., & Lin, Y. D. (2017, August). Deep learning for sentiment analysis on google play consumer review. In 2017 IEEE international conference on information reuse and integration (IRI) (pp. 382-388). IEEE.
[3] Wang, Y., Feng, S., Wang, D., Zhang, Y., & Yu, G. (2016). Context-aware chinese microblog sentiment classification with bidirectional LSTM. In Web Technologies and Applications: 18th Asia-Pacific Web Conference,

APWeb 2016, Suzhou, China, September 23-25, 2016. Proceedings, Part I (pp. 594-606). Springer International Publishing.

[4] Tai, K. S., Socher, R., & Manning, C. D. (2015). Improved semantic representations from tree-structured long short-term memory networks. arXiv preprint arXiv: 1503.00075.

[5] Yanmei, L., & Yuda, C. (2015, December). Research on Chinese micro-blog sentiment analysis based on deep learning. In 2015 8th international symposium on computational intelligence and design (ISCID) (Vol. 1, pp. 358-361). IEEE.

[6] Li, Q., Jin, Z., Wang, C., & Zeng, D. D. (2016). Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems. Knowledge-based systems, 107, 289-300.

[7] Xu, F., Zhang, X., Xin, Z., & Yang, A. (2019). Investigation on the Chinese text sentiment analysis based on convolutional neural networks in deep learning. Computers, Materials & Continua, 58(3).

[8] Long, F., Zhou, K., & Ou, W. (2019). Sentiment analysis of text based on bidirectional LSTM with multi-head attention. Ieee Access, 7, 141960-141969.

[9] Zhang, W., Li, L., Zhu, Y., Yu, P., & Wen, J. (2022). CNN-LSTM neural network model for fine-grained negative emotion computing in emergencies. Alexandria Engineering Journal, 61(9), 6755-6767.

[10] Gan, C., Feng, Q., & Zhang, Z. (2021). Scalable multi-channel dilated CNN–BiLSTM model with attention mechanism for Chinese textual sentiment analysis. Future Generation Computer Systems, 118, 297-309.