Analysis of the Effectiveness of CNN-LSTM Models Incorporating Bert and Attention Mechanisms in Sentiment Analysis of Data Reviews

Lujuan Deng[1](✉), Tiantian Yin[1], Zuhe Li[1,2], and Qingxia Ge[1]

[1] School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, Henan, China
lujuandeng@163.com, {332007040532, 332007040514}@email.zzuli.edu.cn, zuheli@zzuli.edu.cn

[2] Henan Key Laboratory of Food Safety Data Intelligence, Zhengzhou University of Light Industry, Zhengzhou 450002, Henan, China

Abstract. This paper proposes a CNN-LSTM model based on Bert and attention mechanism, since current models cannot deal well with long-term dependencies in natural language. Firstly, the Bert-encoded text vector is fed into the CNN-LSTM model, and secondly, the output of the CNN-LSTM model is fed into the Attention-Based layer, which extracts the most relevant information from the input, and the important features are extracted by weighting the vector. The results show that compared with BiLSTM-ATT, Hierarchical Attention Network (HAN), Convolutional Neural Network (ABCNN), and Attention-Based models, the proposed model has significantly improved in accuracy, F1 score, and macro-averaged F1 metrics. The proposed model has significantly improved in accuracy, F1 score, and macro-average F1 metrics.

Keywords: Bert · convolutional neural networks · long- and short-term memory neural networks · attentional mechanisms

1 Introduction

In recent years, with the popularity of platforms such as social media and e-commerce, more and more users are posting reviews of goods or services on these platforms. These reviews contain rich emotional information, such as sentiment tendencies and emotional intensity, which are of great significance to both merchants and consumers [1]. Therefore, sentiment analysis techniques have important applications in analyzing these reviews. At present, deep learning techniques have been widely used in the field of sentiment analysis, with models such as Convolutional neural networks (CNN) and Long and short-term memory neural networks (LSTM) being the most widely used deep learning models [2]. However, these models do not deal well with long-term dependencies in natural language and therefore have limitations in sentiment analysis.

P. Qi and Z. Chen (Eds.): ICBDIE 2023, AISR 178, pp. 821–829, 2024.
https://doi.org/10.2991/978-94-6463-238-5_106

To address this problem, this study introduces Bert and attention mechanisms to enhance the performance of CNN-LSTM models to improve the accuracy and efficiency in data review sentiment analysis. Specifically, this study proposes a CNN-LSTM model based on the Bert and attention mechanisms, which retains the advantages of CNN and LSTM models while enhancing the representation learning ability and generalization ability of the model by introducing the Bert and attention mechanisms.

2 Related Studies

Sentiment analysis is an important research direction in natural language processing, which aims to identify and extract sentiment information from text [3]. Traditional sentiment analysis methods are mainly based on lexicons or machine learning models, such as plain Bayes, support vector machines, and random forests [4], however, these methods have limitations in dealing with complex text pre-processing, feature extraction, and model training [5].

In recent years, with the development of deep learning techniques, more and more scholars have started to explore the use of deep learning methods for sentiment analysis. Socher et al. [6] proposed a Recursive Neural Network (RNN) model for sentence-level sentiment analysis tasks, and achieved excellent performance on several datasets. The convolutional neural network model proposed by Kim [7] in 2014 is one of the classical approaches in the field of sentiment analysis. Chen [8] et al. proposed a method for sentiment analysis based on user and product concerns. The method first models the attention between users and products and then uses this attention information to weigh the user and product reviews. Later, Sun [9] et al. proposed a convolutional neural network-based sentiment analysis model that uses a convolutional neural network to extract features in the input comments and pass them to a multilayer perceptron for classification. Zeng et al. [10] proposed a new model, his paper implements the BERT combined with RCNN to judge the positive and negative directions of the text, and then uses BERT's next sentence prediction (NSP) to find out the topic-related sentences in the text. In addition, attention mechanisms [11] are another technique widely used in the field of natural language processing to focus attention on the most relevant parts of a text.

In summary, deep learning-based approaches have shown good results in the field of sentiment analysis, Therefore, A CNN-LSTM model incorporating Bert and attention mechanisms, proposed in this paper, aims to further improve the accuracy and efficiency of sentiment analysis.

3 Model Construction

We adopt a CNN-LSTM model with an attention mechanism to improve the accuracy of the model. In this process, we need to input the text vectors encoded by Bert into the CNN-LSTM model and preprocess the data into training, validation, and testing sets, respectively, for training, parameter tuning, and testing. The model structure diagram is shown in Fig. 1.

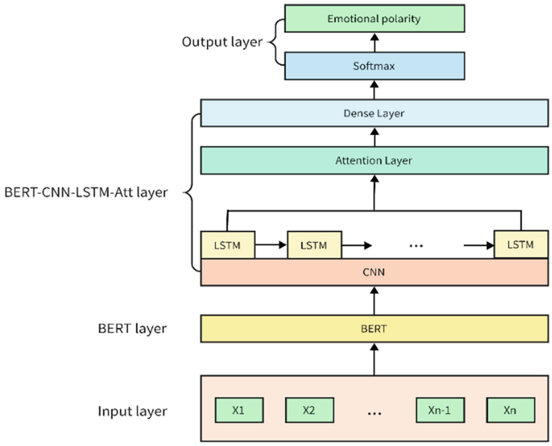

Fig. 1. Structure of BERT-CNN-LSTM-Att model

3.1 Input Layer

(1) Data pre-processing: the original text data is cleaned, divided into words, and deactivated to obtain a data format that can be processed by the model.
(2) Text embedding layer: The text sequence after word separation is mapped into a high-dimensional vector representation, where each word corresponds to a vector {X1, X2,…Xn−1, Xn}, which is used to capture the semantic information of each word. In the model in this paper, a Bert pre-training model is used for text embedding.

3.2 Embedded Layers

The embedding layer is used to map the input text data into a high-dimensional vector space and to capture the semantic information of the words [12]. Specifically, the input text data is first passed through the Bert pre-training model to obtain a vector representation of each word, and then these vectors are stitched together with the manually constructed word vectors to obtain a matrix of dimension (S, d), where S denotes the length of the input text and d denotes the dimension of each word vector. This matrix is used as the output of the embedding layer and as input to the subsequent model. The Bert model is shown in Fig. 2.

BERT architecture, where En is the nth token in the input sequence, Trm is the Transformer block, and Tn is the corresponding output embedding.

3.3 Convolutional Layers

The pre-processed text matrix S*d is used as the feature map of the input layer of the convolutional neural network, and the convolutional kernels are used to convolve the feature map and further extract local features. The convolutional layer of the convolutional neural network consists of convolutional kernels of different sizes, and the size of the convolutional kernels used in this paper is h*d. The height h of the convolutional

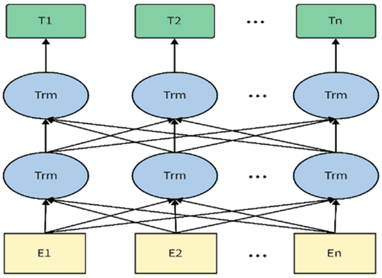

Fig. 2. BERT architecture

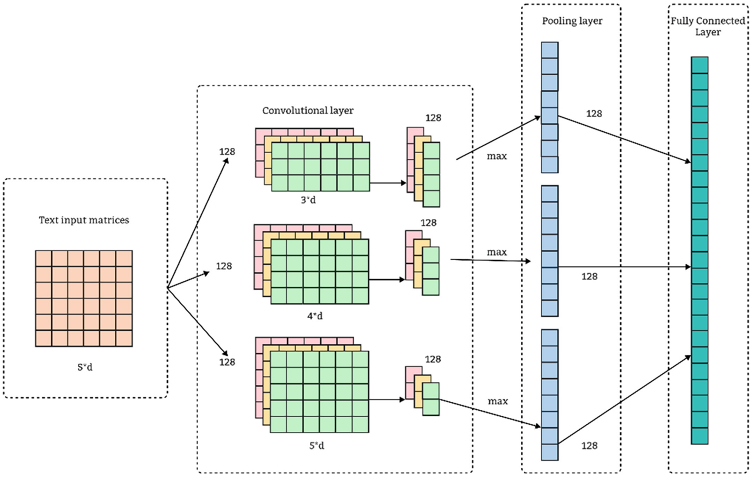

Fig. 3. CNN structure diagram

kernels is 3, 4, and 5 respectively, and there are 128 convolutional kernels of each size. The pooling layer, also known as the downsampling layer, is mainly used to reduce the dimensionality of the feature map, but does not change the number of feature maps, and is used to select the optimal feature values. The structure of a CNN is shown in Fig. 3.

3.4 LSTM Layer

The LSTM layer is a special kind of [13] RNN, which is mainly used to process sequential data with temporal order, such as text data. The LSTM layer consists of three gating units, an input gate, forget gate, and output gate, and a unified state. The application formula and detailed flow of the LSTM layer are as follows:

$$f_s = \sigma(W_f \cdot [h_{s-1}, x_s] + b_f) \tag{1}$$

$$i_s = \sigma(W_f \cdot [h_{s-1}, x_s] + b_i) \tag{2}$$

$$\tilde{C} = \tanh(W_c \cdot [h_{s-1}, x_s] + b_C) \tag{3}$$

$$C_s = f_s \cdot C_{s-1} + i_s \cdot \tilde{C} \tag{4}$$

$$o_s = (W_o[h_{s-1}, x_s] + b_C) \tag{5}$$

$$h_s = o_s \cdot \tan(C_s) \tag{6}$$

The input gate controls how much information can flow into the LSTM cell state, as in Eq. (4). The forgetting gate controls how much information needs to be discarded, as Eq. (3). The update of the cell state is determined by Eqs. (5) and (6), and the output gate controls the output of the cell state as Eq. (7). Where i_s is the output of the input gate, σ denotes the sigmoid function, C_s and is the unit state at the current moment.

3.5 Attention Layer

The Attention Layer is a mechanism commonly used in deep learning to help models better focus on the important parts of the input sequence [14]. In this paper's model, the Attention Layer is applied to the output of a convolutional-long and short-term memory network to extract the most relevant information from the input. The process is as follows.

$$\begin{aligned} u_i &= \tanh(W_w h_i + b_w) \\ a_i &= \mathit{soft}\max(u_i) \\ H_a &= \sum a_i h_i \end{aligned} \tag{7}$$

3.6 Output Layers

The output layer is mainly used to predict the classification result of the comment sentiment. The final output vector is normalized by the softmax activation function to obtain the probability distribution of each class. The formula is Eq. (10).

$$y = \mathit{soft}\max(Wh_{fc} + b) \tag{8}$$

4 Experiments

4.1 Data Sets

The dataset used in this paper is The Stanford Sentiment Treebank (SST), which is a widely used dataset for sentiment analysis. This dataset contains 11,855 training samples, 1,016 validation samples, and 2,210 test samples and is widely used for the evaluation of sentiment analysis models. Figure 4 shows an example from the SST dataset, which is a parse tree structure with all nodes labeled.

4.2 Experimental Parameter Settings

The specific parameters of the model and their descriptions are shown in Table 1.

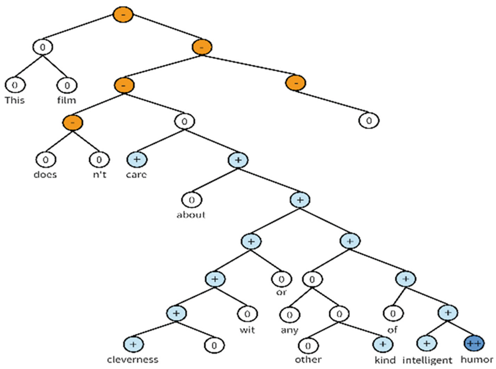

Fig. 4. Example SST dataset

Table 1. Model parameter settings

Name of experimental parameter	Parameter values
Embedding size	300
hidden size	150
attention size	150
Window Size	3, 4, 5
Epochs	10
Dropout_rate	0.5
Optimizer	Adam
Batch Size	32
learning rate	0.001

4.3 Evaluation Indicators

Accuracy, F1 Score, and Macro F1 are used as evaluation metrics in this paper. Accuracy provides a clear judgment of our model's performance, F1 Score is the summed average of accuracy and recall, which takes into account both accuracy and recall of the classification model, and Macro F1 simply averages the F1 scores for each category and can provide an overview of the overall performance evaluation. Here are the calculation equations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$\Pr ecision(P) = \frac{TP}{TP + FP} \tag{10}$$

$$\text{Re}call(R) = \frac{TP}{TP + FN} \tag{11}$$

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{12}$$

where TP indicates the number of sentiment predictions that are positive and correct, TN indicates the number of sentiment predictions that are negative and correct, FP indicates the number of negative category errors predicted as positive, and FN indicates the number of positive category errors predicted as negative.

4.4 Experimental Results and Analysis

In this paper, the proposed model was compared with five other commonly used sentiment analysis models in the experiment. This is shown in Table 2. These comparison models include CNN, LSTM, and models that include attention mechanisms such as BiLSTM-ATT, Hierarchical Attention Network (HAN), and Attention-Based Convolutional Neural Network (ABCNN). The experimental results show that the model proposed in this paper achieves the best performance so far on the SST-2 dataset, and its accuracy on the test set reaches 90.4%, which is about 1 percentage point higher than the highest accuracy of all other models. Also, the model in this paper performed well in both weighted F1 and macro-averaged F1, with 89.5% and 88.2% respectively, which were 0.5% and 0.7% higher than the highest values of the other models, respectively.

Specifically, in comparison with the CNN model, this paper's model improved by 0.9%, 1.3%, and 1.5% in accuracy, F1, and macro-average F1 metrics, respectively. In comparison with the LSTM model, the model improved by 1.4%, 0.4%, and 0.9% in accuracy, F1, and macro-average F1 metrics, respectively. In comparison with the BiLSTM-ATT model, the model improved by 0.7%, 0.8%, and 0.8% in the accuracy, F1, and macro-average F1 metrics respectively. In the comparison with the HAN model, the model in this paper improved by 0.7%, 0.8%, and 0.9% in accuracy, F1, and macro-average F1 metrics, respectively. In comparison with the ABCNN model, the model improved by 1.1%, 0.4%, and 0.4% in accuracy, F1, and macro-average F1 metrics, respectively. These results indicate that the model in this paper has a good performance advantage on the SST-2 dataset.

The experimental results show that the model proposed in this paper performs well in terms of accuracy, score F1, and macro F1. Compared with the best benchmark model, the model proposed in this paper improved by about 0.7 percentage points in macro F1 value. This indicates that the model proposed in this paper has better performance in the sentiment analysis task.

Table 2. Comparison of experimental results

Model	F1 Score	Macro F1	Accuracy
TextCNN	88.2%	86.7%	89.5%
LSTM	89.1%	87.3%	89.0%
BiLSTM-ATT	88.7%	87.4%	89.7%
Hierarchical Attention Network(HAN)	88.7%	87.4%	89.7%
Attention-Based Convolutional Neural Network(ABCNN)	89.1%	87.8%	89.3%
Bert-CNN-LSTM-Attention	**89.5%**	**88.2%**	**90.4%**

In addition, further analysis is carried out in this paper. By visualizing the attentional weight matrix, the paper finds that the attentional mechanism does focus better on key information and improves the model's focus on sentiment words. The importance of each part of the model is also analyzed and it is found that the attention mechanism and the LSTM layer play a more significant role in the model proposed in this paper. These analyses provide useful insights for further improving and optimizing the model proposed in this paper.

5 Conclusion

In summary, this paper presents a CNN-LSTM model incorporating Bert and attention mechanisms for data review sentiment analysis. The experimental results show that the model has better performance than the traditional CNN-LSTM model and the model using only Bert while achieving the best performance on the SST dataset. Through the analysis of the experimental results, we found that the model achieves significant improvements in different levels of sentiment analysis tasks, especially for complex negative sentiment analysis tasks. However, we are also aware that there are still some limitations of the model, for example, it may not perform well for texts with high semantic complexity. Future research can explore more effective model structures and richer datasets to improve the performance of sentiment analysis tasks.

Founding. This paper was supported by the National Natural Science Foundation of China under Grant 61702462, the Henan Provincial Science and Technology Research Project under Grant 222102210010 and 222102210064.

References

1. Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: A survey[J]. Ain Shams engineering journal, 2014, 5(4): 1093–1113.
2. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: A survey[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, 8(4): e1253.
3. Yue W, Li L. Sentiment analysis using word2vec-cnn-bilstm classification[C]//2020 seventh international conference on social networks analysis, management and security (SNAMS). IEEE, 2020: 1–5.
4. Liu H. Sentiment analysis of citations using word2vec[J]. arXiv preprint arXiv:1704.00177, 2017.
5. Wang J, Yu L C, Lai K R, et al. Dimensional sentiment analysis using a regional CNN-LSTM model[C]//Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers). 2016: 225–230.
6. Socher R, Huval B, Bath B, et al. Convolutional-recursive deep learning for 3d object classification[J]. Advances in neural information processing systems, 2012, 25.
7. Kim Y. Convolutional Neural Networks for Sentence Classification.
8. Zhou J, Huang J X, Chen Q, et al. Deep learning for aspect-level sentiment classification: survey, vision, and challenges[J]. IEEE access, 2019, 7: 78454–78483.

9. Huang L, Sun X, Li S, et al. Syntax-aware graph attention network for aspect-level sentiment classification[C]//Proceedings of the 28th international conference on computational linguistics. 2020: 799–810.
10. H. -T. Tseng, Y. -Z. Zheng and C. -C. Hsieh, "Sentiment Analysis using BERT, LSTM, and Cognitive Dictionary," 2022 IEEE International Conference on Consumer Electronics - Taiwan, Taipei, Taiwan, 2022, pp. 163–164, https://doi.org/10.1109/ICCE-Taiwan55306.2022.9868974.
11. Bhuvaneshwari P, Rao A N, Robinson Y H, et al. Sentiment analysis for user reviews using Bi-LSTM self-attention based CNN model[J]. Multimedia Tools and Applications, 2022, 81(9): 12405–12419.
12. Jia K. Sentiment classification of microblog: A framework based on BERT and CNN with attention mechanism[J]. Computers and Electrical Engineering, 2022, 101: 108032.
13. Huang F, Li X, Yuan C, et al. Attention-emotion-enhanced convolutional LSTM for sentiment analysis[J]. IEEE transactions on neural networks and learning systems, 2021, 33(9): 4332–4345.
14. Rani S, Bashir A K, Alhudhaif A, et al. An efficient CNN-LSTM model for sentiment detection in# BlackLivesMatter[J]. Expert Systems with Applications, 2022, 193: 116256.