Classification of Microblog Users' Sentiments Based on BERT-BiLSTM-CBAM

Juan Chen, Ruyun Chen, Di Yu

School of Mathematics and Computer Science, Guangdong Ocean University, Zhanjiang, Guangdong, China

Abstract

This study proposes a method for more accurate classification of microblog users' sentiments based on the BERT-BiLSTM-CBAM hybrid model. First, the text information is pre-trained by the bidirectional encoder representation from transformers (BERT) model to get feature vectors. Then, the feature vectors are spliced and recombined using bidirectional long-short-term memory network (BiLSTM) and CBAM mechanism to obtain new feature vectors. Finally, these new feature vectors are input to the full connection layer and then processed by the softmax function to obtain the sentiment category of the text. The experiment conducted on the sample dataset demonstrates that the model proposed in this study yielded accurate and dependable result in classifying microblog texts in the sample dataset. The model based on BERT-BiLSTM-CBAM algorithm is more efficient than the traditional depth model in processing microblog contents.

Keywords: Sentiment analysis, BERT, BiLSTM, CBAM mechanism

I. Background

In this information age, the Internet generates huge amounts of data every day, and ordinary netizens are at a loss in the midst of this information tidal wave [1]. According to the 44th Statistical Report on Internet Development in China released by China Internet Network Information Center (CNNIC) in August 2019, the number of Internet users in China has reached 854 million, with an Internet penetration rate of 61.2%, and the total number of domain names is 48 million. In today's world, the data generated in the past five years account for 90% of the total [2]. In 2020, more than 50 billion Internet devices will generate tens of Zeta Byte of data, which means that the era of big data is coming [3]. Therefore, the analysis and utilization of massive data have become the focus of many research and engineering fields. In recent years, natural language processing has become a hot research topic, and sentiment analysis, an important part of natural language processing, has received a lot of attention. Sentiment analysis refers to the process of analyzing, processing, and extracting subjective texts with sentimental tint using natural language processing and text mining techniques [4]. Text sentiment analysis can be applied in various fields, and many examples are easily available. It is feasible to build the capability of understanding people's preference or degree of preference for something by training sentiment analysis model. One can build a model to predict stock prices by evaluating investors' degree of optimism based on their comments on online forums. By analyzing the public opinions on various issues using sentiment analysis model, government departments can prevent the publication and dissemination of some false information. Since Pang first combined machine learning with sentiment analysis in 2002, sentiment analysis methods have received a lot of attention [5]. In particular, the sentiment inclination analysis based on online comments has made great progress. Currently, the mainstream sentiment analysis methods can be roughly divided into three types. The first type is based on sentiment dictionary. The accuracy of this type of method depends heavily on sentiment dictionary which requires a lot of manpower and material resources to construct and maintain. The second type is based on machine learning. In 2018, Lee et al. proposed a sentiment analysis method based on dictionary and weak labeling information, which is a result of combining sentiment dictionary with machine learning [6]. The third type is based on deep learning. In 2018, Li proposed an algorithm based on the fusion of convolutional neural network and BiLSTM model [7]. This algorithm realizes the fusion of features by making

use of the complementarity between models, laying a good foundation for the text classification task. The sentiment analysis method proposed in this study combines neural network with CBAM mechanism and performs sentiment analysis based on the BERT-BiLSTM-CBAM hybrid model.

II. Related Techniques and the Model of Bert-Bilstm-Cbam

2.1 Related techniques

2.1.1 BERT model
In 2018, Google Inc. proposed a new BERT model [8]. Based on the transformer bidirectional language model, BERT model employs the masked language model (MLM) and next sentence prediction (NSP) method to pre-train the target, so it can obtain context-related bidirectional feature representation and is good at matching sentences or paragraphs. BERT is a cloze-like masking language model that combines the advantages of Word2Vec, ELMo, and GPT models. Its basic idea is to randomly mask some input marks according to a certain proportion and only predict the masked marks, that is, it only passes the final hidden vector corresponding to the masked marks to the normalization function of the output layer through the vocabulary instead of passing the whole input [9]. This cloze-like learning forces the model to rely more on context information to predict words, thus endowing the model with certain error correction ability. The structure of the BERT model is shown in Figure 1. In which E1, ..., EN represents the text inputs in the form of words, and the vectorized representations T1, ..., TN of the text are obtained after the text is processed by the bidirectional transformer encoder.
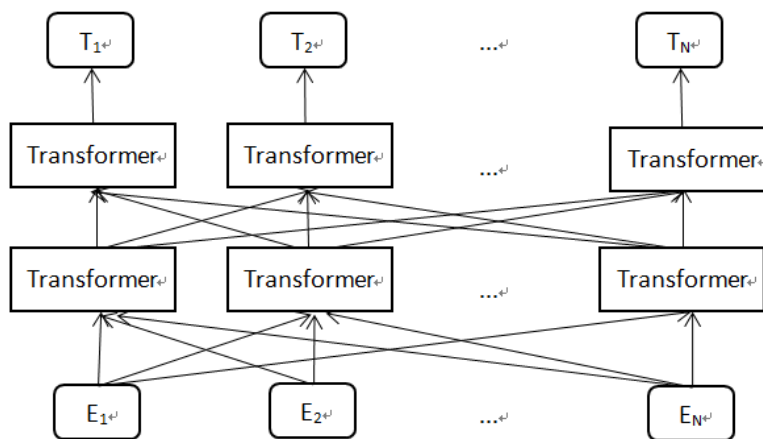

Fig 1: Structure of the BERT model

2.1.2 BiLSTM model
Recurrent neural network (RNN) is a neural network for processing sequential data [10]. Compared with the common neural networks, it has the advantage of processing data with sequential changes but suffers from the gradient disappearance problem (explosion problem). LSTM, a special RNN proposed by Jurgen Schmidhuber, solves the above problem by using a complex gate mechanism and can effectively detect long-distance dependence in sequences [11]. However, there is a problem in modeling sentences with LSTM, that is, it is impossible to encode information from back to front. The BiLSTM model proposed by Graves A [12] has stronger "memory ability" than uni-directional LSTM [13]. The network structure of BiLSTM is shown in Figure 2, where xi (i=0,1,2,3,4) represents the input content.
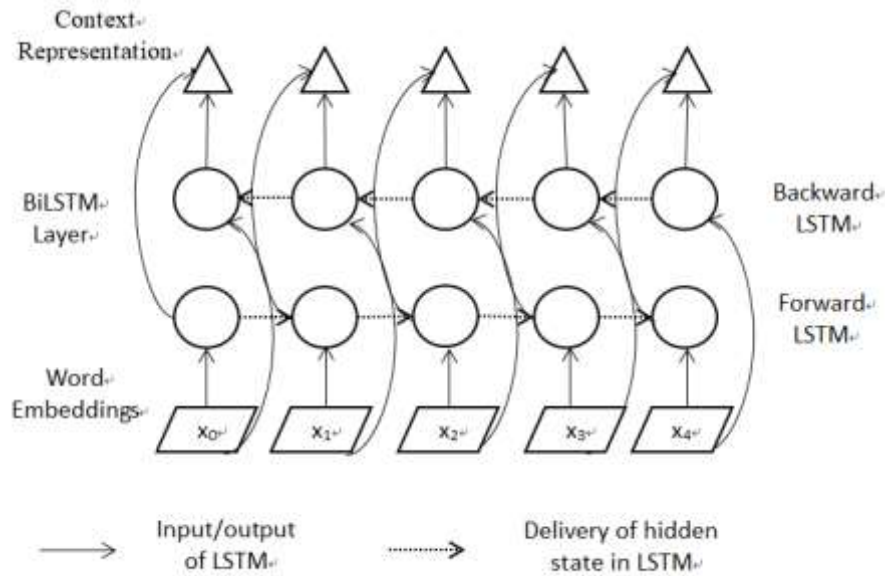
Fig 2: Structure of the BiLSTM network

2.1.3 The CBAM mechanism

In 2018, Woo et al. [14] put forward a new convolution module attention mechanism module in the paper CBAM: Convolutional Block Attention Module. The module improves the accuracy of the network through the feature weight calibration on both the spatial and channel dimensions, achieving better results than SENet, which only pays attention to the attention mechanism on the feature dimension. The working principle of CBAM is shown in Figure 3.
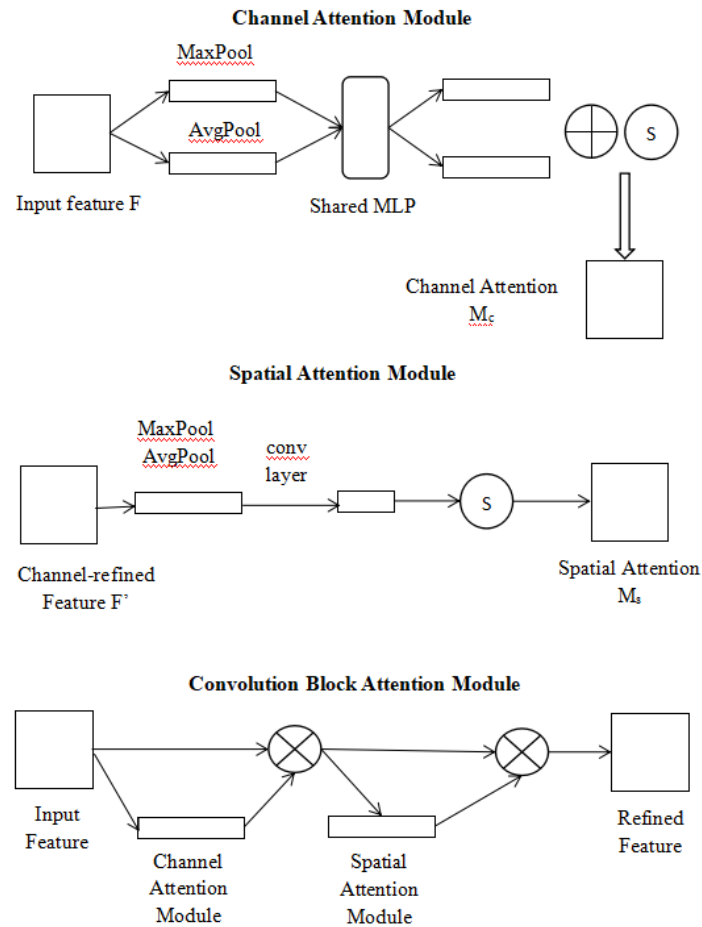
Channel Attention Module

Spatial Attention Module

Convolution Block Attention Module

Fig 3: Schematic diagram of CBAM

Figure 3 shows that the channel attention mechanism will use the maximum and average pooling algorithms at the same time. The data output by the two algorithms go through several MLP layers, and the outputs are applied to the two channels separately to obtain the channel attention results using the sigmoid function. First, the spatial attention mechanism reduces the dimensions of the channel itself. Then it obtains the results of maximum and mean pooling and splices them into a new feature. Finally, it uses a convolution layer for learning. The two mechanisms learn the importance of channel and space separately and can be easily embedded into other frameworks.

2.2 The model of BERT-BiLSTM-CBAM

This study combines the above three models and proposes a BERT-BiLSTM- CBAM hybrid text sentiment analysis model. The model is mainly composed of BERT, BiLSTM and CBAM layers. Its structure is shown in Figure 4.

The basic idea of the model is that the language model is first pre-trained by BERT, then word vector extraction is performed to obtain the word vector representation of each text, including context information. This ensures that no text information gets lost. In BERT, the language model is built by masked language model (LM); first, it uses a randomly selected word or phrase that is masked or replaced; then it allows the model to guess the masked word or phrase based on the understanding of the context. Therefore, only the masked part needs to be calculated when doing Loss. The procedure for the experiment is as follows. First, replace 80% of tokens in one sentence with [mask], 10% with wrong words, and 10% with correct words. Then, let the model predict and restore the masked or replaced parts. The calculated dimensions output by the hidden layer are Xhidden:[batch_size, seq_len, embedding_dim], and the

weight of a mapping layer is initiated as Wvocab: [embedding_dim, vocab_size]. The mapping from the hidden dimension to the number of word vectors is established using Wvocab. This process only requires the matrix multiplication (dot product) of Xhidden and Wvocab; the sum of vocab_size is 1. The prediction result of the model is obtained using the word with the highest probability in vocab_size, and the prediction result can be used to do Loss with the prepared Label and transfer the gradient back.
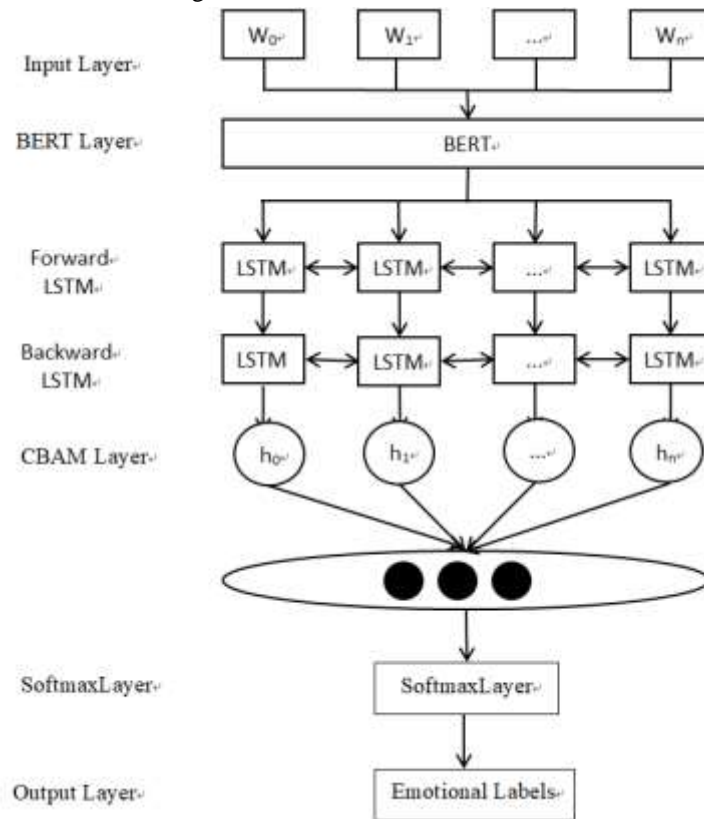

Fig 4: The BERT-BiLSTM-CBAM model

The total of n vectors are obtained by the BERT model, then the obtained vectors are spliced by BiLSTM model, where hi (i=0,1,...,n) represents the feature vectors output to the front and back hidden layers at time i. The feature vector h of the final text can be obtained by splicing the feature vectors in two directions at the last time. The resulting h is put under global max pooling and global average pooling calculations separately based on the width and height and then passes through the MLP layer. The features output by the MLP layer undergo an elementwise addition operation and then sigmoid activation, resulting in the channel attention featuremap. Perform elementwise multiplication operation on the channel attention featuremap and input featuremap to obtain the input features required by the spatial attention module, as shown in Equation (1)

$$M_c(h) = \sigma(MLP(AvgPool(h))) + MLP(MaxPool(h))$$
$$= \sigma(W_1(W_0(h^c_{avg})) + W_1(W_0(h^c_{max}))), W_0, W_1 \in R^{C/r \times C} \tag{1}$$

where seigema represents the sigmoid operation, and r represents the reduction rate.

The featuremap output by the channel attention module is taken as the input featuremap of this module. First, perform global max pooling and global average pooling based on the channel, and then do concat operation based on the channel. Next, perform a convolution operation to reduce the dimensions to 1 channel. Then generate spatial attention feature through sigmoid. Finally, make this feature to multiply with the input feature of the module to obtain the final features, as shown in Equation (2)

$$M_s(h) = \sigma(f^{7\times7}([AvgPool(h); MaxPool(h)]))$$
$$= \sigma(f^{7\times7}([h^s_{avg}; h^s_{max}]))$$

(2)

where seigema represents sigmoid operation, and 7*7 indicates the size of convolution kernel; the 7*7 convolution kernel is better than the 3*3 convolution kernel. The obtained feature vectors are input to the full connection layer, and the emotion classification is performed by softmax function.

III. Experimental Results and Analysis

An experiment based on a sample dataset was carried out. The sample dataset contains 10,000 pieces of training data and 1,000 pieces of test data. To verify the effectiveness of the model, two other models, BERT and BERT+BiLSTM, were also trained and tested for comparison. The performance of each model was assessed using the calculated accuracy and fi-score values. The experimental results of all the models are shown in Table 1.

Table 1 Experimental results of different models

S/N	Model	Accuracy	f1-score
1	BERT	69.0%	0.81657
2	BERT+BiLSTM+CBAM	89.4%	0.9203
3	BERT+BiLSTM+CBAM	92.2%	0.9437

A comparison of Models 1 and 2 reveals that adding the BiLSMT model can significantly improve the prediction accuracy and f1-score value because BiLSTM can better capture the bidirectional semantic dependence. It can be seen from the comparison of Models 2 and 3 that the accuracy and f1-score values of BERT+BiLSTM+CBAM have been improved because the addition of the CBAM mechanism enables the model to obtain more internal features of the text. The comparison of the three models demonstrates that the BERT-BiLSTM-CBAM model delivers better performance than BERT+BiLSTM when processing microblog content dataset, improving the f1-score value by 2.34%.

IV. Conclusions

The proposed microblog content sentiment analysis model based on BERT-BiLSTM-CBAM algorithm is more efficient than the traditional depth model in processing microblog contents. The experimental results show that the processing ability of BERT-BiLSTM improves with the help of the CBAM mechanism. This is because we have added the CBAM mechanism to the model to remedy one weakness of traditional models, that is, low sensitivity to sentimental words in the process of comment text processing. The addition of the CBAM mechanism enhances the ability of the model to capture sentimental words in the text, thereby improving the accuracy and f1-score values of the model.

References

[1] S. Khusro, Z. Ali, I. Ullah, "Recommender Systems: Issues, Challenges, and Research Opportunities," Information Science and Applications (ICISA) vol. 376, pp. 1179-1189, 2016.

[2] China Internet Network Information Center, China Internet Network Information Center (CNNIC). vol. 9, pp. 30-31, 2019.

[3] H. Karau, A. Konwinski, P. Wendell, et al., "Learning Spark: Lightning-Fast Big Data Analytics," O'Reilly Media,inc. pp. 258, 2015.

[4] D. Xu, Z. Tian, R. Lai, et al., "Deep Learning Based Emotion Analysis of Microblog Texts," Information Fusion. pp. 64, 2020.

[5] B. Pang, L. Lee, "Opinion mining and sentiment analysis," Comput Linguist. vol. 35, pp. 311-312, 2009.

[6]   G. Lee, J. Jeong, S. Seo, et al., "Sentiment classification with word localization based on weakly supervised learning with a convolutional neural network," Knowledge-Based Systems. vol. 152, pp. 70-82, 2018.

[7]   J. Yang, X. Huang, H. Wu, et al., "EEG-based emotion classification based on Bidirectional Long Short-Term Memory Network," Procedia Computer Science. vol 174, pp. 491-504, 2020.

[8]   T. Mikolov, K. Chen, Corrado G S, et al., "Efficient Estimation of Word Representations in Vector Space," International conference on learning representations. vol. 1, pp. 1-12, 2013.

[9]   Y. Arase, J. Tsujii, "Transfer fine-tuning of BERT with phrasal paraphrases," Computer speech & Language. 66: 101164, 2021.

[10]   W. Zaremba, I. Sutskever, O. Vinyals, "Recurrent Neural Network Regularization," Eprint Arxiv. vol. 1, pp. 1-15, 2014.

[11]   S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory. Neural Computation," vol 9, pp. 1735-1780, 1997.

[12]   A. Graves, J. Schmidhuber, "Special Issue: Framewise phoneme classification with bidirectional LSTM and other neural network architectures," Neural Networks. vol. 18, pp. 602-610, 2005.

[13]   L. Gu, W. Zhang, Y. Wang, et al., "Named Entity Recognition in Judicial Field Based on BERT-BiLSTM-CRF Model," 2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI). vol. 10, pp. 170-174, 2020.

[14]   S. Woo, J. Park, J. Lee, et al., "CBAM: Convolutional Block Attention Module," European Conference on Conputer Vision. vol. 11211, pp. 3-19, 2018.