Sentiment Analysis with An Integrated Model of BERT and Bi-LSTM Based on Multi-Head Attention Mechanism

Yahui Wang, Xiaoqing Cheng, Xuelei Meng

***Abstract*—Sentiment analysis is one of the most important tasks in natural language processing. The goal of sentiment analysis is to classify the text-based sentiment tendencies by extracting text features. The quality of the classification can be assessed by the accuracy, precision and recall of the classification. So it is very important to design efficient tools or models to carry out the classification process. In this paper, we build a novel model, integrating the dealing method of Bidirectional Encoder Representation from Transformers (BERT), Bidirectional Long Short - Term Memory (Bi-LSTM) and the Multi-Head Attention (MHA), building an MHA based, integrating BERT and Bi-LSTM, Sentiment Analysis Model (MHA-BB-SAM). The model concludes two layers, the embedding layer and the Bi-LSTM layer. In the embedding layer, BERT is used to pre-process the text and train the required word vectors. In the Bi-LSTM layer, Bi-LSTM network is utilized to catch the order information of the context. Multi-Head Attention (MHA) is introduced to enhance the ability to capture the long-distance text features. The novel model is tested on two corpuses, SemEval-2020 corpus and College Learners' Spoken English Corpus (COLSEC). The computing results prove the high performance of the model. The approach presented in this paper can provide a reference for the sentiment analysis work.**

***Index Terms*—sentiment analysis, BERT, Bi-LSTM, MHA, MHA-BB-SAM**

I. Introduction

WITH development of technology and Internet, the information generating speed is getting higher and higher. On social media like Twitter and Facebook, people express their views in the form of text. On film review websites, film reviews are presented to the public in the form of text. The massive text contains a lot of valuable information that should be extracted and utilized fully. Sentiment analysis has been gaining importance in many applications such as recommendation systems, decision making support and prediction models. Sentiment analysis helps to understand and evaluate public opinion regarding social events, product services, and political trends, especially the feelings expressed through comments by users in social networks such as Twitter, Facebook, and Instagram.

Users in various platforms of short text contain strong emotion tendentiousness, reflecting the user's different attitudes. It has great value to collect massive amounts of text information and do the sentiment analysis. For instance, the government can understand the public's opinions on various social events and related policies by analyzing the different attitudes of citizens towards the same news event. By analyzing the user's evaluation of a function of the product, the manufacturer can improve the product in a targeted way. The potential value of sentiment analysis has attracted extensive attention from researchers in different fields such as data mining and natural language processing. Sentiment analysis for user-generated text has become an international research hotspot in relating fields.

In this paper, we propose a novel approach to solve the sentiment analysis problem. Our contributions are listed here.

A- We combine BERT, Bi-LSTM and Multi- attention mechanism, presenting a novel method to analyze the text.

B- We analyze the performance of the presented model, concluding that it has better performance than the Bi-LSTM model, which can not only capture the inner relationship in sentences from two directions but also obtain the local sequential information.

C- We tested the model on SemEval-2020 corpus and COLSEC and proved the effectiveness of it.

This paper is structured as follows: Section 2 presents the related work on this problem. And section 3 proposes the basic models and mechanisms for sentiment analysis, including BERT, Bi-LSTM and multi-attention mechanism. In Section 4, we build an integrated sentiment analysis model that consists of two layers-the embedding layer and the Bi-LSTM layer. In Section 5, a computation case is presented to test the model and we compare the computing results with the results obtained with other models. Section 6 draws conclusions.

Manuscript received July 4, 2022; revised December 24, 2022.

This research was funded by the Young Teachers program of Lanzhou Jiaotong University (Grant: 2019039), the Natural Science Foundation of Gansu Province (22JR5RA355), the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University (RCS2022K005), National Natural Science Foundation of China (Grant No. 71861022) and "Double-First Class" Major Research Programs, Educational Department of Gansu Province (No. GSSYLXM-04).

Yahui Wang is a lecturer of Foreign Languages School, Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China. (phone: 86-13919010651, e-mail: wangyh_lzjtu@163.com)

Xiaoqing Cheng is an associate professor of State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China. (e-mail: xqcheng@bjtu.edu.cn)

Xuelei Meng (corresponding author) is a professor of Traffic and Transportation School, Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China.(phone: 86-15117071015, e-mail: mengxl_lzjtu@163.com)

II. Related Work

Sentiment analysis, also known as opinion mining, is a research field that analyzes people's subjective feelings towards products, services, organizations, personal events, topics, attributes and other entities, such as opinions, emotional evaluation, views and attitudes. Text Sentiment Analysis (TSA) is one of the most crucial tasks in Natural Language Processing (NLP). It is a process to extract features and classify text with emotion. By comparing the current models with better performance, it is found that dividing sentences into word sequences and using the network model based on long - short memory or gated loop unit as the main tool for processing sequence data can extract the context relations of sentences effectively.

Sentiment analysis has caught the attention of researchers recently. Phan et al. (2019) proposed an approach based on integrating the sentiment towards a particular object from all tweets related to that object [1]. Ahmed et al. (2020) proposed a supervised neural weak model that aimed at learning a set of sentiment clusters embedding from the sentence global representation of the target domain [2]. Sreevidya et al. (2020) proposed a model for carrying out deep learning based multimodal sentiment analysis [3]. Peng et al. (2020) proposed an adversarial learning method for training sentiment word embedding, in which the discriminator was used to force the generator to produce high-quality word embedding by using semantic and sentiment information [4]. Tofighy and Fakhrahmad (2018) proposed a statistical and context-aware feature reduction algorithm that improved sentiment classification accuracy [5]. Mechulam et al. (2019) presented a model based on a context-graph which can be used for building domain specific sentiment lexicons (DL: Dynamic Lexicons) by propagating the valence of a few seed words [6]. Kumar and Parimala (2020) utilized sentiment analysis method to generate the plan for recommending smart phones [7]. Kulkarni et al. (2020) demonstrated sentiment analysis as a promising tool to quantify consumer responses towards branded viral video advertisements and thereupon, proposing a sentiment-based typology of viral ad sharers [8]. Yang et al. (2019) proposed a recurrent attention convolutional neural network (RACNN), which incorporated convolutional neural networks (CNNs)[9].

Lin et al. (2020) reformulated the classification task as a comparing problem, and proposed Comparison Enhanced Bi-LSTM with Multi-Head Attention (CE-B-MHA) to improve the performance of text sentiment analysis [10]. Khiabani et al. (2020) proposed an improved evidence-based aggregation method to do the sentiment analysis work [11]. Studiawan et al. (2020) proposed a sentiment analysis technique to automatically extract events of interest from log messages in the forensic timeline [12]. Asif et al. (2020) presented a novel approach for multilingual sentimental analysis, classifying the incorporated textual views into any of four categories, including high extreme, low extreme, moderate, and neutral, based on their level of extremism [13]. Ali et al. (2021) constructed a novel healthcare monitoring framework based on the cloud environment and a big data analytics engine to precisely store and analyze healthcare data, and to improve the classification accuracy. The computing results showed that the proposed model precisely handled heterogeneous data and improved the accuracy of health condition classification and drug side effect predictions [14].

There are some publications on Bi-LSTM application in language processing. Ali et al. (2021) proposed a social network‑based, real-time monitoring framework for traffic accident detection and condition analysis using ontology and latent Dirichlet allocation (OLDA) and bidirectional long short-term memory (Bi-LSTM), and they tested the framework using traffic-related data, comparing OLDA and Bi-LSTM with existing topic modeling methods and traditional classifiers respectively [15]. They proposed a novel fuzzy ontology-based semantic knowledge with Word2vec model to improve the task of transportation features extraction and text classification using the Bi-directional Long Short-Term Memory (Bi-LSTM) approach and proved that Bi-LSTM showed satisfactory improvement in both the extraction of features and the classification of the unstructured text of social media [16].

Bahad et al. (2019) used Bi-LTSM to detect the fake news [17]. Wang et al. (2020) proposed an improved emotion analysis model based on Bi-LSTM model to classify the further four-dimensional emotions of Pleasure, Anger, Sorrow and Joy [18]. Kim et al. (2019) developed a model to satisfy the requirements of Dialog System Technology Challenge 6 (DSTC6) Track 1: building an end-to-end dialog systems for goal-oriented applications, using Bi-LTSM [19]. Ye et al. (2019) proposed a web service classification method based on Wide & Bi-LSTM model [20]. Zhao et al. (2018) proposed an architecture of the Bi-LSTM neural network to enhance recognition of motion state accuracy [21]. Lin et al. (2019) proposed a neural-encoded mention-hypergraph model to use hypergraph to model overlapping or nested structure mentions based on Bi-LSTM [22]. Sreelakshmi et al. (2018) proposed a deep learning based framework using Bi-Directional Long Short-Term Memory (BLSTM) Networks for intent identification [23]. Jahangir et al. (2020) utilized Bi-LSTM to design a deep learning-based forecasting approach [24].

All of the publications gave us much enlightenment when we wrote the paper. We will propose a novel framework to realize the sentiment analysis work based on the references above in this paper.

III. Basics of Integrated Sentiment Analysis Model

We will combine two models and a mechanism to complete sentiment analysis work. So it is necessary to introduce the models and the mechanism firstly. The basic models are BERT, and Bi-LSTM and the mechanism is MHA mechanism.

BERT (Bidirectional Encoder Representation from Transformers) is a training model that was presented in 2018 (Devlin et al, 2018) [25], which is shown in Fig. 1. In the model, Trm is the transformer, E and T are the input sentence and the output sentence. It takes bi-directional transformers as the encoders and the MLM (Masked Language Model) and SLR (Sentence- Level Representation) are proposed. BERT is designed to pre-train deep bidirectional representations by jointly conditioning on both left and right context in all layers.

A. BERT

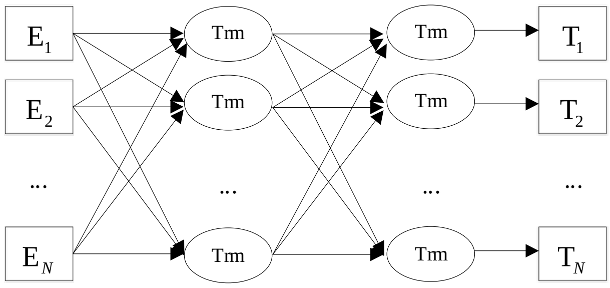

Fig. 1. Pre-training language model of BERT

15% of the words in the training set are marked with [MASK] due to MLM. The detailed rule is as follows. The word is marked with [MASK] with an 80% probability, replaced with another word with a 15% probability and set to be the original token with a 10% probability. The goal of SLR is to learn the relationship between the sentences. It replaces the sentences randomly and forecast the sentences. The key of BERT is to take transformer as the encoder, which is shown in Fig. 2.

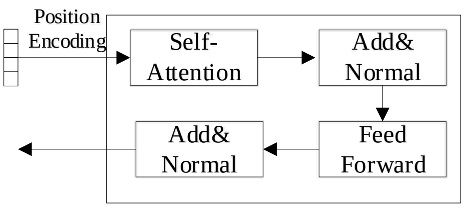

Fig. 2. Structure of encoding unit of transformer

There are four basic layers in the encoding unit of transformer, shown in Fig. 2. To avoid the problem that there is no position information in the data extracted from self-attention layer, a position encoding process is added before in-putting step, as shown in Fig. 2. The out-put of self-attention layer is transferred to the feed-forward neural network. Residual Network and layer normalization are added into the encoding unit of transformers to avoid the degeneration of deep learning. Add operation utilizes the structure of ResNet model to prevent the degeneration while realizing the multi-layer superposition. Layer normalization operation lowers the difficulty level by normalizing the vectors.

BERT model can fully utilize the information of both sides of a word to obtain better distribution representation.

B. Bi-LSTM

LSTM (Long Short- Term Memory) was proposed in

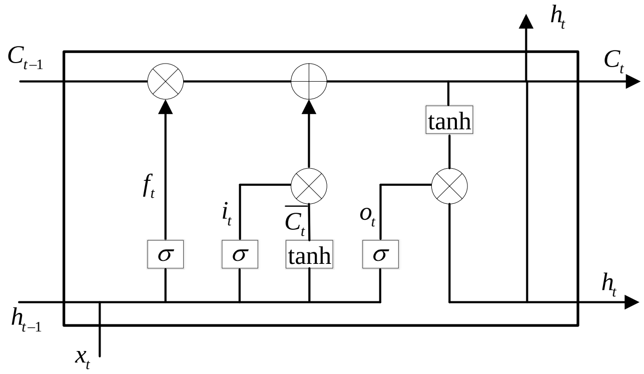

Fig. 3. Memory unit structure of LSTM

1997 [26], which is a special RNN (Recurrent Neural Network) model. LSTM solves the problems of gradient vanishing and gradient explosion.

LSTM can catch the long term dependence between the sentences and carry out sentiment analysis on the whole text. There are 3 control gates, forgotten gate, in-put gate and out-put gate. The structure for memory is shown in Fig. 3.

h_{t-1} and h_t stand for the out-put of the previous and current cells. x_t is the in-put of the current cell. σ is the activation function of Sigmoid. tanh is an activation function. f_t is the out-put of the forgotten gate that decides what to discard from the cell. The product of i_t and C_t is the out-put of the in-put gate, which decides how much information should be added into the cell. o_t is the out-put of the out-put gate that will be filtered based the cell status. The relationship of them can be described as the following equations.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\overline{C_t} = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = f_t \times C_{t-1} + i_t \times \overline{C_t} \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (6)$$

Bi-LSTM is proposed by Graves Abase on LSTM [27]. Bi-LSTM is composed of forward-LSTM and backward-LSTM and makes it catch the bi-directional semantic dependence.

C. MHA Mechanism

Attention mechanism was presented by Mnih et al. in 2014 to deal with image processing problem, which hired the RNN-Attention approach to identify and classify the images [28]. Attention mechanism simulates human attention to external information. The model will focus selectively on the related information in the in-put data. And self-attention mechanism makes it possible for the sentence to deal with itself without any additional information and extracts useful information. At the same time, self-attention mechanism can catch the characteristic relation between the words faraway.

Machine translation team of Google proposed Multi-Head Attention (MHA) in 2017 [29]. Order information is most important for NLP that stands for the logical structure of the manuscript. MHA mechanism can get the global structure of the manuscript, while it cannot catch the order information. Machine translation team of Google solved this problem with the position vector. They numbered each position and matched them to the word vectors. Thus it makes it possible to make the attention model to catch the order information of the text. In this paper, we use the MHA based Bi-LSTM to get the order information.

IV. An Integrated Sentiment Analysis Model

We build a novel sentiment analysis model, based on MHA, integrating BERT and Bi-LSTM, as shown in Fig. 4. The model consists of three layers, the embedding layer, the Bi-LSTM layer and MHA mechanism.

A. Embedding Layer

The text is usually turned to word vectors before it can be proceeded and an embedding layer should be inserted to encode the text. In this paper, we use BERT to pre-process the text and train the required word vectors. Sentences are divided into words and the low frequency words and stop words are deleted. The word vectors are loaded in the embedding layer and put into the model as the initialized parameters, as x_1 and x_t in Fig. 4.

B. Bi-LSTM Layer

In this paper, LTSM network is introduced to catch the order information of the context. LSTM network is a special recurrent neural network that not only has the external circulating system, but also has a structure of internal cellular LSTM, shown in Fig. 4. We use gate structure to realize the selective passage of information. The gate is constructed with a sigmoid function layer and a multiplication operation that can delete or add the information into a cell. In each cell of LSTM, there are 3 gates, forget gate, input gate and the output gate. They will perform different functions to control the cell status.

In Fig. 4, x_t is the input of the current unit, h_{t-1} is the input of the precious cell and h_t is the output of the current cell. C_{t-1} and C_t are the status of the precious cell and current cell.

The forget gate decides what information should be deleted from a cell. It reads h_{t-1} and x_t, puts out a weight which is between 0 and 1 to multiply it with the number of status C_{t-1}. If the value is 1, it means to keep thoroughly. If the value is 0, it means to delete thoroughly. The input gate is used to determine how to add new information into a cell. tanh layer is used to generate the candidate vectors and we determine the number to add into a cell by multiplying it with the result of Sigmoid layer. When the number is added into the cell, the status of the cell is change to C_t. The output number of the system is determined based on the status of C_t. Then we use tanh to deal with C_t and multiply it with the output of Sigmoid layer and get the input number h_t.

Then LSTM cell can be abstracted as a function that includes h_{t-1}, x_t (as the input) and h_t (as the output). The function is shown in Equation (7).

$$h_t = LSTM(x_t, h_{t-1}) \tag{7}$$

The Bi-LSTM network can extract the characteristics of the word sequence from the front and back more precisely and efficiently. We use Bi-LSTM to extract the local order information, combining the forward LSTM and backward LSTM.

The output of forward LSTM and backward LSTM are h_t^f and h_t^b respectively.

$$h_t^f = LSTM(x_t, h_{t-1}^f) \tag{8}$$

$$h_t^b = LSTM(x_t, h_{t-1}^b) \tag{9}$$

Then h_t^f and h_t^b are connected and a new vector- h_t is generated, which is the output of Bi-LTSM.

$$h_t = Concat(h_t^f, h_t^b) \tag{10}$$

C. MHA Mechanism

MHA is substantially a combination of several Self-Attention (SA) structures. In this paper, we use MHA to get the characteristic of the distance. The output of Bi-LTSM, from h1 to h_t will be combined into matrices, will be the input information of ***Q***, ***K*** and ***V*** of each Self- Attention unit, see Fig. 4. In Self – Attention, each vector should attend the Attention calculating process with the destination to learn the dependence relation in the sequence and catch the inner structure of the sequence.

As shown in Fig. 4, the Scaled Dot- Product Attention (SDPA) mechanism used in MHA-B is a variant of Self – Attention. It is reduced to $\sqrt{d_K}$ based on the calculation of point multiplication using ***Q*** and ***K***. The goal is to control the size of inner product. The first step is to calculate the point multiplication of ***Q*** and ***K***. The second step is to divide the inner product by a standard scale value. Then, the calculation result is turned to probability distribution with *Softmax* manipulation. Then it is multiplied by Matrix **V**. The SDPA manipulation is shown in Equation (11).

$$SDPA(Q,K,V) = softmax(\frac{QK^t}{\sqrt{d_K}}) \tag{11}$$

In MHA, ***Q***, ***K*** and ***V*** are conducted a linear transformation with different parameters W_i^Q, W_i^K, W_i^V, and input them into SDPA. The process is repeated for h times. And the process will make each result $head_i$ learn the characteristics in different representation space. It is shown as Equation (12).

$$head_i = SDPA(Q \cdot W_i^Q, K \cdot W_i^K, V \cdot W_i^V) \tag{12}$$

In this paper, ***Q***, ***K*** and ***V*** are conducted linear transformation with different W. Then $head_1$, $head_2$, …, and $head_h$ will be combined. Then the combined sequence is transformed linearly with parameters W. Then the result of MHA is obtained, as shown in Equation (13).

$$MHA(Q,K,V) = Concat(head_1, head_2, ..., head_h) \cdot W \tag{13}$$

To sum up, MHA merges h_1, h_2, … and h_t, (output of Bi-LSTM layer) into a matrix and calculates it with the dot product. And then it completes the linear transformation after forming several heads. The linear transformation of different parameters of ***Q, K*** and ***V*** is the essence of multi-head, with the destination to learn the information from different dimensions and describing subspace. MHA-B turns the input text into word vectors x_1 and x_t, using the Embedding layer.

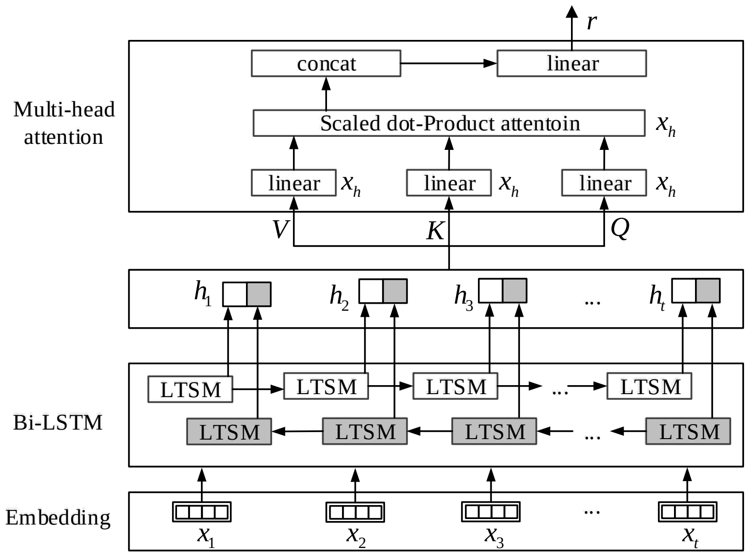

Fig. 4. MHA-BB-SAM model with embedding, Bi-LSTM and MHA layer

Bi-LSTM layer generates forward and backward word vectors and gets h_1 and h_t. Then a matrix is generated with h_1, h_2, … and h_t, and the matrix is taken as the input data of ***Q***, ***K*** and ***V*** of MHA. MHA-B and Bi-LTSM complement each other and improve the analysis ability of the model.

V. Data Experiment and Result Analysis

MHA-B model is trained with the selected data set. The model is evaluated with the classified result and compared the result with the baseline to acquire the sentiment analysis ability of the model. The data experiment is done based on the SemEval-2020 and the COLSEC. The model is tested at a Windows 10 platform of a personal computer with processor frequency of 2.5 GHz and memory size of 4 GB.

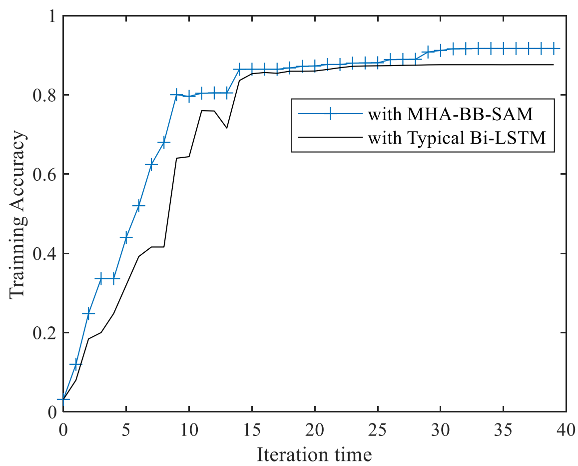

Fig. 5. Training accuracy with MHA-BB-SAM and typical Bi-LSTM

Firstly, we calculated training accuracy and verification accuracy of the MHA-BB-SAM and typical Bi-LSTM. It can be seen that the training accuracy and the verification accuracy rise with the increase of text classification processing number. Fig. 5 shows that the training accuracy tends to be stable when the calculation number reaches 30. It is about 0.917 with MHA-BB-SAM and 0.876 with typical Bi-LSTM. The verification accuracy turns to be stable when the calculation number reaches 35 as shown in Fig. 6. It is 0.840 with MHA-BB-SAM and 0.837 with typical Bi-LSTM.

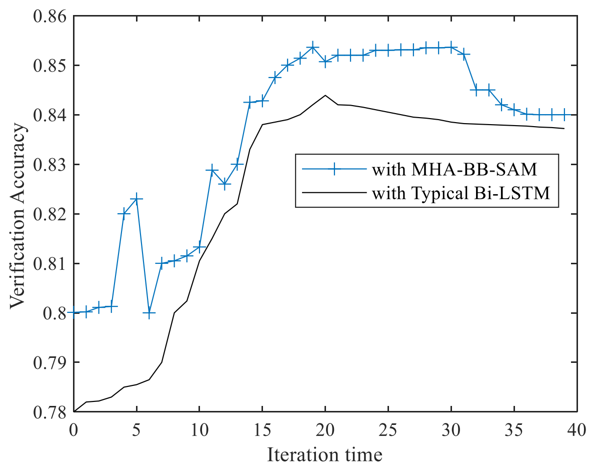

Fig. 6. Verification accuracy with MHA-BB-SAM and typical Bi-LSTM

It can be seen from Fig. 7 that when calculating training accuracy using MHA-BB-SAM and typical Bi-LSTM, the MHA-BB-SAM model consumes more time than the typical Bi-LSTM model. Taking completing 30 iterations for example, the total time spent using the MHA-BB-SAM model was 6940.5 seconds, compared to 6036.9 seconds using the typical Bi-LSTM. The MHA-BB-SAM model takes 13.02% more time than the typical Bi-LSTM model. Similarly, MHA-BB-SAM cost more time than typical Bi-LSTM model when carrying out the verification accuracy calculation, which took 14.15% more time than using typical Bi-LSTM. The reason is that each iteration of the MHA-BB-SAM model involves more computing processes and has more complexity. As a result, it costs more time.

Fig. 8 shows the deviation of ACC (Accuracy) values during each computation iteration on SemEval-2020. It is easy to find that when calculating with MHA-BB-SAM, the deviation value fluctuated between 0 and 0.1741, with an average value of 0.1096. When calculating with the Typical Bi-LSTM, the deviation value fluctuated between 0 and

0.1929, with an average value of 0.1284. Since the standard value of the ACC is 1, its deviation values were all positive. It is obvious that the MHA-BB-SAM model outperforms the Typical Bi-LSTM model for the computational deviation control.

TABLE I
COMPUTING RESULTS WITH SEMEVAL-2020 CORPUS AND THE COLSEC

Corpuses	Model	ACC	AUC	Precision	Recall
SemEval -2020	MHA-B B-SAM	0.8904	0.9517	0.8993	0.8525
	Typical Bi-LSTM	0.8716	0.9288	0.9090	0.8409
COLSE C	MHA-B B-SAM	0.8528	0.8913	0.9028	0.8354
	Typical Bi-LSTM	0.7984	0.8892	0.9133	0.9927

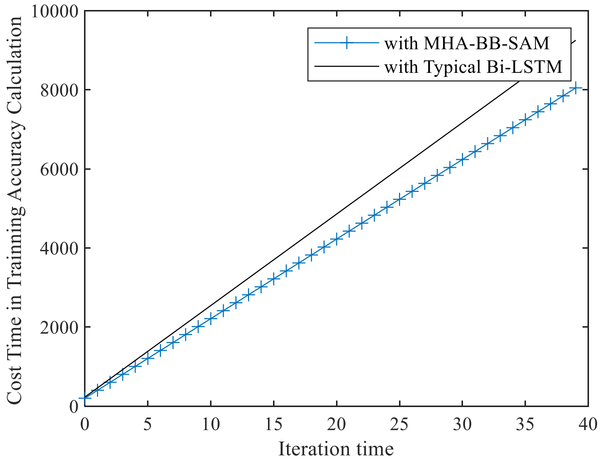

Fig. 7. Cost time in training accuracy calculation

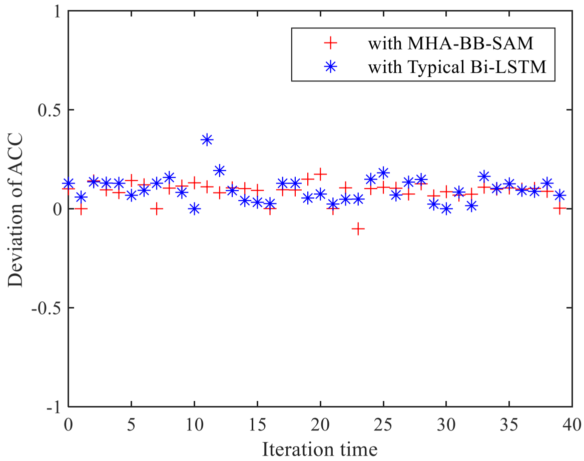

Fig. 8. Deviation of ACC

Similarly, it is concluded from Fig. 9, Fig. 10 and Fig. 11 that the MHA-BB-SAM models all showed better deviation control ability than the Typical Bi-LSTM models in the AUC (Area Under Curve), Precision, and Recall calculations.

The typical Bi-LSTM is taken as the baseline to analyze the performance of MHA-BB-SAM. Typical Bi-LSTM is often used to extract the characteristics of the text, which is appropriate to be used to compare the performance. As we can see in TABLE I, ACC value obtained with MHA-BB-SAM on SemEval-2020 corpus is 0.8904, which is 2.16% higher than that with typical Bi-LSTM. The value of AUC obtained with MHA-B is 0.9517, which is 2.47% higher than that with Bi-LSTM. ACC and AUC values obtained with the two models on COLSEC show the same characteristics. Then we can conclude that MHA-B model has the performance on the ACC and AUC than the typical Bi-LSTM model.

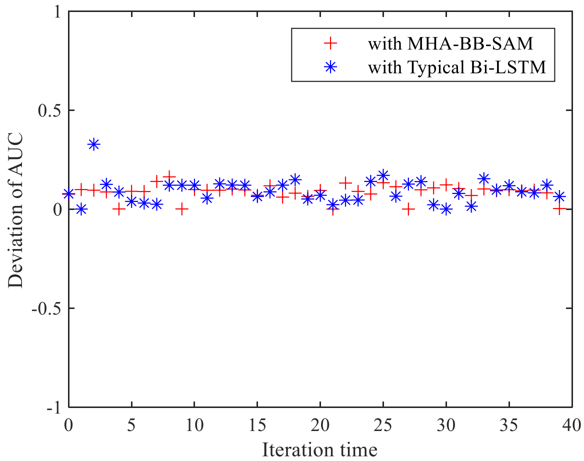

Fig. 9. Deviation of AUC

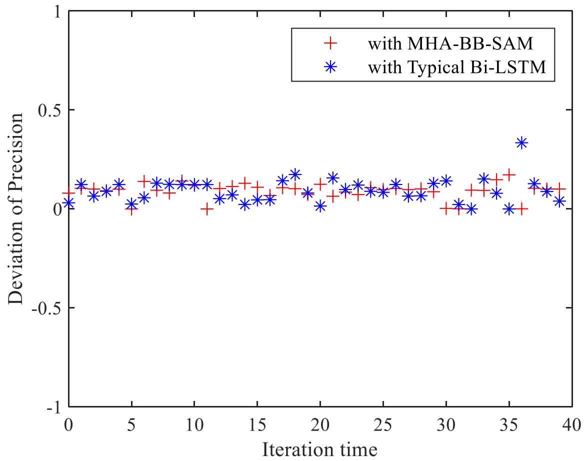

Fig. 10. Deviation of precision

On the other hand, the precision obtained with MHA-BB-SAM on SemEval-2020 corpus is 0.8993 and it is 0.9090 with typical Bi-LSTM. And the recall ratio value is 0.8525 with MHA-BB-SAM and it is 0.8409 with Bi-LSTM. Although the precision and recall values obtained with MHA-BB-SAM are not better than that obtained with typical Bi-LSTM, they still maintain a high level. The same conclusion can be drawn when we did the experiment on COLSEC.

Typical Bi-LSTM is taken as the baselines in this paper, which is a common model to extract the characteristics of texts. The essence of attention mechanism is actually weighted summation that has the stronger processing ability. This baseline is suitable to compare the performance of MHA-BB-SAM model, which is enhanced by the multi-head attention mechanism.

MHA mechanism can more efficiently catch the text characteristics of long text, compared with typical Bi-LSTM. MHA mechanism can learn important information from different dimensions and representing subspace, compared with the typical Bi-LSTM with attention. MHA-BB-SAM model has more powerful classifying ability than typical Bi-LSTM.

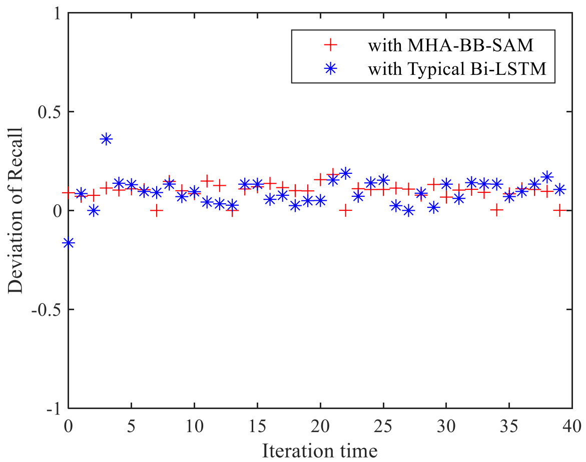

Fig. 11. Deviation of recall

VI. Conclusion

We proposed a novel sentiment analysis model-MHA-BB-SAM model, which combines Bi-LSTM and multi-head attention mechanism. The model is tested on SemEval-2020 corpus and COLSEC. The presented model has better performance than the Bi-LSTM model through comparing the computing results.

In this paper, Bi-LSTM is hired to catch the order information of the sequence, supplementing the multi-head attention mechanism. The advantage of MHA-BB-SAM is that it can capture the inner relationship in sentences from two directions with Bi-LSTM and obtain the local sequential information. Additionally, it can catch the characteristics of long distance with multi-head attention. It integrates the advantages of the ability of capturing the local sequential information of Bi-LSTM and the ability of catching the global information. As a novel sentiment analysis tool, it has high performance in sentiment analysis. The method we proposed in this paper can extract the public sentiment from the text, presented on the Internet, like MicroBlog, Twitter and Facebook. It is a method to monitor the public opinions and the sentiment.

There is still need for further research to extend and refine the analytical techniques described in this paper. Computational efficiency is one area of concern, especially in bi-directional relationship information catching. And the model should be tested on more corpuses to analyze the performance. The application of the model in sentiment analysis should also be paid attention in the future research work.

References

[1] Phan, H.T.; Nguyen, N.T.; Tran, V.C.;Hwang, D.; Szczerbicki, E.;Trawiń ski, B. A sentiment analysis method of objects by integrating sentiments from tweets. Journal of Intelligent & Fuzzy Systems. 2019, 37(6), 7251-7263.

[2] Ahmed, M.; Chen, Q.; Li, Z. Constructing domain-dependent sentiment dictionary for sentiment analysis. Neural Computing and Applications 2020, 10.1007/s00521-020-04824-8.

[3] Sreevidya, P.; Murthy, O. V. R.; Veni, S. Sentiment analysis by deep learning approaches. Telkomnika, 2020, 18(2), 752-760.

[4] Peng, B.; Wang,J.; Zhang, X. Adversarial learning of sentiment word representations for sentiment analysis. Information Sciences 2020, 541: 426-441.

[5] Tofighy, S.; Fakhrahmad, S.M. A proposed scheme for sentiment analysis. Kybernetes, 2018, 47(5), 957-984.

[6] Mechulam, N.; Salvia, D.; Ros, A.; Etcheverry, M. Building Dynamic lexicons for sentiment analysis. Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial, 2019, 22(64), 1-13.

[7] Kumar, G.; Parimala, N. An integration of sentiment analysis and MCDM approach for smartphone recommendation. International Journal of Information Technology & Decision Making. 2020, 19(4), 1037-1063.

[8] Kulkarni, K.K.; Kalro, A.D.; Sharma, D.; Sharma, P. A typology of viral ad sharers using sentiment analysis. Journal of Retailing and Consumer Services. 2020, 53, doi:10.1016/j.jretconser.2019.01.008.

[9] Yang, H.D.; Zhang, J.; Li, S.H.; Luo, T.J. Bi-direction hierarchical LSTM with spatial-temporal attention for action recognition. Journal of Intelligent & Fuzzy Systems. 2019, 36(1), 775-786.

[10] Lin, J.C.W.; Shao, Y.N.; Zhou, Y.J.; Pirouz, M.; Chen, H.C. A Bi-LSTM mention hypergraph model with encoding schema for mention extraction. Engineering Applications of Artificial Intelligence. 2019, 85, 175-181.

[11] Khiabani, P.J.; Basiri, M.E.; Rastegari, H. An improved evidence-based aggregation method for sentiment analysis. Journal of Information Science. 2020, 46(3), 340-360.

[12] Studiawan, H.; Sohel, F.; Payne, C. Sentiment Analysis in a forensic timeline with deep learning. IEEE Access. 2020, 8, 60664-60675.

[13] Asif, M.; Ishtiaq, A.; Ahmad, H.; Aljuaid, H.; Shah,J. Sentiment analysis of extremism in social media from textual information. Telematics & Informatics 2020, 48: 101345. doi:10.1016/j.tele.2020.101345

[14] Ali, F.; El-Sappagh, S., Riazul Islam S.M. and et al. An intelligent healthcare monitoring framework using wearable sensors and social networking data. Future Generation Computer Systems, 2021, 114: 23-43.

[15] Ali, F., Ali, A., Imran, M., and et al. Traffic accident detection and condition analysis based on social networking data. Accident Analysis & Prevention, 2021, 151: 105973, doi: 10.1016/j.aap.2021.105973

[16] Ali, F., El-Sappagh, S., Kwak, D. Fuzzy Ontology and LSTM-Based Text Mining: A Transportation Network Monitoring System for Assisting Travel. Sensors, 2019, 19(2): 234. doi: 10.3390/s19020234.

[17] Bahad, P.; Saxena, P.;Kamal, R. Fake news detection using bi-directional LSTM-Recurrent neural network. Procedia Computer Science. 2019, 165, 74-82.

[18] Wang, S.K.; Chen, Y.H.; Ming, H.J.; Huang, H.; Mi, L.X.;Shi, Z.Y. Improved danmaku emotion analysis and its application based on Bi-LSTM model. IEEE Access. 2020, 8, 114123-114134.

[19] Kim, B.; Chung, K.; Lee, J.; Seo, J.; Koo, M.W. A Bi-LSTM memory network for end-to-end goal-oriented dialog learning. Computer Speech and Language, 2019, 53: 217-230.

[20] Ye, H.F.; Cao, B.Q.; Peng, Z.L.; Chen, T.; Wen, Y.P.; Liu, J.X. Web services classification based on wide & Bi-LSTM model. IEEE Access. 2019, 7, 43697-43706.

[21] Zhao, H.; Hou, C.N.; Tiwari, S.; Trivedi, M.; Kohle, M.L. Recognition of motion state by smartphone sensors using Bi-LSTM neural network. Journal of Intelligent & Fuzzy Systems. 2018, 35(2), 1733-1742.

[22] Lin, Y.; Li, J.P.; Yang, L.; Xu, K.; Lin, H.F. Sentiment Analysis With Comparison Enhanced Deep Neural Network. IEEE Access. 2020, 8, 78378-78384.

[23] Sreelakshmi, K.; Rafeeque, P.C.; Sreetha, S.; Gayathri,E.S. Deep bi-directional LSTM network for query intent detection. Procedia Computer Science. 2018, 143, 939-946.

[24] Jahangir, H.; Tayarani, H.; Gougheri, S.S.; Golkar, M.A.; Ahmadian, A.; Elkamel, A. Deep learning-based forecasting approach in smart grids with micro-clustering and bi-directional LSTM network. IEEE Transactions on Industrial Electronics. 2020, doi: 10.1109/TIE.2020.3009604.

[25] Devlin, J.; Chang, M. W.; Lee, K.; et al. BERT: Pre-training of deep bidirectional transformers for language understanding. 2018. arXiv: 1810.04805 [cs.CL]. 11. Oct. 2018.

[26] Hochreiter, S.; Schmidhuber, J. Long short- term memory. Neural Computation. 1997, 9(8), 1735-1780.

[27] Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. J. Neural Networks. 2005, 18(5-6), 602-610.
[28] Mnih, V.; Heess, N.; Graves, A.; et al. Recurrent models of visual attention. Advances in neural information processing systems. 2014. arXiv:1406.6247v1 [cs.CL]. 24. Jun. 2014.
[29] Vaswani, A.; Shazeer, N.; Parmar, N.; et al. Attention is all you need. Advances in Neural Information Processing systmes. Long Beach, USA. 2017, 5998-6008.

Y. Wang received the MA degree in Foreign Linguistics and Applied Linguistics in School of Foreign Languages, Lanzhou Jiaotong University in 2012. She is currently a lecturer in Lanzhou Jiaotong University. Her research interests are in the areas of scientific and technical translation, translation theory and practice.