Aspect-Based Attention LSTM for Aspect-Level Sentiment Analysis

Jingang Ma
School of Intelligence and Information Engineering
Shandong University of Traditional Chinese Medicine
Jinan, China
ma_jingang@126.com

Xiaohong Cai*
School of Intelligence and Information Engineering
Shandong University of Traditional Chinese Medicine
Jinan, China
caixhsdutcm@163.com
*Corresponding author

Dejian Wei
School of Intelligence and Information Engineering
Shandong University of Traditional Chinese Medicine
Jinan, China
289608678@qq.com

Hui Cao
School of Intelligence and Information Engineering
Shandong University of Traditional Chinese Medicine
Jinan, China
caohui63@163.com

Jing Liu
School of Intelligence and Information Engineering
Shandong University of Traditional Chinese Medicine
Jinan, China
Liuj_jn@163.com

Xuqiang Zhuang
School of Information Science and Engineering
Shandong Normal University
Jinan, China
zhuangxq@sdnu.edu.cn

2021 3rd World Symposium on Artificial Intelligence (WSAI) | 978-1-6654-0451-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/WSAI51899.2021.9486323

***Abstract*—With the rapid development of the mobile Internet, how to efficiently extract information from the massive data in cyberspace and use it has become a key concern of all walks of life. Sentiment analysis aims to dig out the subjective emotional information contained in large-scale texts. It has become a hot research topic in the field of natural language processing and has important application value in many fields. This paper focuses on the aspect-level sentiment analysis problem, which is an important sub-problem in fine-grained sentiment analysis. The purpose of aspect-level sentiment classification is to dig out the sentiment polarity of users' opinions on a specific aspect in the text. This paper proposes a sentiment classification method combining BiLSTM and aspect attention module. BiLSTM has fewer parameters, faster model training, and can effectively extract deep-level information of text; combining the attention mechanism with aspect information can fully extract specific aspects of information. The experimental results show that the F1 score that solves the problem of aspect term extraction and aspect sentiment classification on two data sets at the same time, compared with the existing sentiment analysis method, obtains a better classification effect.**

***Keywords*—aspect-level sentiment analysis, Bi-LSTM, attention mechanism**

I. INTRODUCTION

Emotion classification is one of the important directions of natural language processing research. Its main purpose is to dig out the subjective emotional information expressed by people for a thing from massive text data. One of the most critical aspects of sentiment analysis is to extract emotional features from natural language, and then use algorithms to analyze them based on the extracted features. According to the target size of sentiment analysis, sentiment analysis tasks can be divided into "coarse-grained sentiment analysis" and "fine-grained sentiment analysis" according to the granularity. Sentiment analysis on user's daily comment data belongs to the category of fine-grained sentiment analysis. For example, for a restaurant review such as "The food is delicious, but the environment is poor." In which the user has a positive attitude towards "food" but a negative attitude towards "environment", and coarse-grained sentiment analysis cannot be used. Comprehensively and accurately analyze the emotion in the sentence. The problem of fine-grained sentiment analysis has gradually become a key concern in the field of sentiment analysis.

In recent years, deep learning has gradually become the mainstream method for solving natural language processing tasks due to its excellent feature extraction capabilities, and deep learning has been widely used in aspect-level sentiment classification tasks [1]-[10]. The method based on deep learning does not need to manually design features, and usually uses an end-to-end method for prediction. You only need to input the text vector into the deep learning model to automatically learn features through the model, which greatly reduces labor costs. Dong et al. [11] proposed an adaptive recurrent neural network. Given a sentence and its dependency tree, the root node of the dependency tree is the target word in the sentence, and try to transform the dependency tree for different target words, so that the aspect words are related to each other. Adjacent words are combined recursively. Nguyen et al. [12] proposed a phrase recurrent neural network model based on the adaptive recurrent neural network, which uses the syntactic information of the dependency tree and the component tree to generate aspect representations. Wang et al. [13] proposed a recursive neural conditional random domain model, which combines the recurrent neural network constructed based on the dependency tree and the conditional random domain to jointly extract aspects and express their opinions on aspects. The performance of the method based on

978-1-6654-0451-8/21/$31.00 ©2021 IEEE

syntactic analysis depends on the accuracy of the syntactic parser. Because the text on the Internet usually has a lot of noise, the analysis accuracy of the syntactic parser is greatly restricted, which also causes the performance of the syntactic parsing algorithm. influences. Vo et al. [14] proposed a model based on the structure of the sentence itself, which can automatically extract features and obtain the same excellent performance as the syntax-based parsing method. The model divides the text into the left context and the right context according to the position of a given aspect, uses a series of neural network functions to extract features, and also uses an emotional dictionary in the process of extraction. This method uses a pooling function to extract features from word vector sequences, and cannot display and capture information related to a given aspect and its context, and cannot clarify the aspect-level semantics of the text. In order to overcome the defect that the pooling function cannot clearly model the level semantics of the text, Zhang et al. [15] proposed a sentence-level neural network model, using a threshold loop neural network to model the semantic information and syntactic information inside the text. In addition, the interaction information between the context and the aspect is modeled. In this way, the expression of the aspect and the context can be obtained more effectively than directly using the pooling function. Tang et al. [16] proposed a target link long and short memory network. The model spliced each word in the sentence face to face with each word in the sentence, and used a two-way LSTM to process the spliced word vector sequence, and combine the aspects and contexts. Semantic association is used for modeling to achieve the purpose of promoting model performance.

Most of the related work focuses on the research of the BiLSTM neural network combined with the attention model, and it has a good performance on sentiment classification tasks. The method in this paper is based on previous research results, using a simple structure, less calculation, and a small storage space required by the Bi LSTM neural network and a hybrid network model of the aspect attention module for aspect sentiment analysis. This method enables different aspects of the sentence to be input into the model at the same time and processed at the same time. For this reason, a multi-faceted attention module is adopted. Each aspect attention module has independent aspect information and attention parameters, can extract feature information of a specific aspect, and can well identify the emotional polarity of different aspects.

II. ASPECT-BASED ATTENTION BILSTM

A. LSTM

LSTM have long-term memory function by introducing the cell state with long-term memory, which can be regarded as a kind of complex cyclic neural network. LSTM introduces three gating units: input gate, forget gate, and output gate, the structure of which is as follows:

$$
\begin{aligned}
f_t &= \sigma(W_f \bullet [h_{t-1}, x_t] + b_f) \\
i_t &= \sigma(W_i \bullet [h_{t-1}, x_t] + b_i) \\
\tilde{C}_t &= \tanh(W_c \bullet [h_{t-1}, x_t] + b_c) \\
C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\
o_t &= \sigma(W_o \bullet [h_{t-1}, x_t] + b_o) \\
h_t &= o_t * \tanh(C_t)
\end{aligned}
\tag{1}
$$

where t represents the information processing state of the current network; b and c represent the bias vectors of the parameters; x_t represents the input of state t; h_t represents the hidden layer output of state t; o_t represents the output layer input of state t; c_t is hidden Layer status.

B. BiLSTM

Bidirectional Long Short-Term Memory is composed of LSTMs in two directions, forward LSTM and backward LSTM, and can capture order-dependent information in both forward and backward directions. LSTM can model sequence data and can well capture long-distance dependencies. However, an LSTM cannot model dependent information from back to front. Therefore, the Bi LSTM layer uses two-direction LSTM to model the word embedding sequence output by the embedding layer, capturing the forward and backward dependency information, and obtaining the contextual abstract representation of the word. The words in the sentence do not exist alone, they all exist in the context. Therefore, the word-independent word embedding feature cannot express its meaning in the entire text. The words in the text have a front-to-back order, and the words arranged in order collectively express attitudes and opinions. BiLSTM can capture the order-dependent features of the context words in the two directions around the aspect term, and obtain a contextualized feature representation.

For the word vector of the word sequence of a sentence S and the given aspect word sequence word vector, BiLSTM is used to extract the bidirectional semantic information of the text sequence. The hidden state output of BiLSTM is as follows:

$$
\begin{aligned}
h_i^f &= \overrightarrow{LSTM}(h_{i-1}^f, e_{w_i}) \\
h_i^b &= \overrightarrow{LSTM}(h_{i-1}^b, e_{w_b}) \\
h_i &= h_i^f \oplus h_i^b
\end{aligned}
\tag{2}
$$

C. Aspect-Based Attention BiLSTM

The Aspect-Based Attention BiLSTM (ABABiLSTM) model is a codec structure, as shown in Figure 1. The input layer and BiLSTM layer are the encoder part, and the decoder part is composed of multiple Aspect-Based Attention modules.

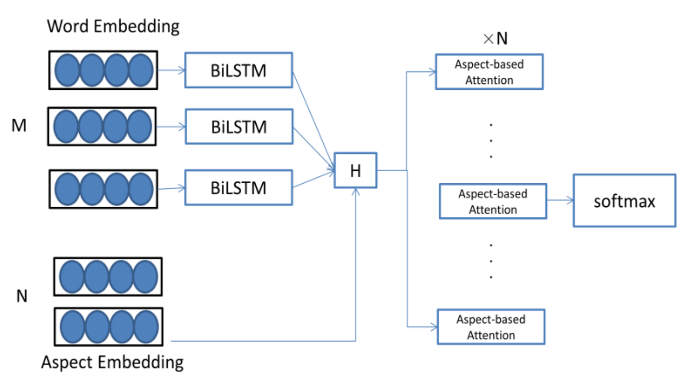

Fig. 1. ABABiLSTM model structure

The encoder part extracts the semantic features of the sentence, including forward and backward directions. Through the BiLSTM network, obtain the hidden layer state ht at the t-th time step. For a sentence of length n, the hidden layer output matrix is $H=[h_1,h_2,\cdots,h_n]$, and the overall vector of the sentence is as follows:

$$v_s=\frac{1}{n}\sum_{t=1}^{n}h_t \tag{3}$$

The decoder part is a multi-aspect attention module, corresponding to different aspects of the data set. For sentences containing multiple aspects, the output H obtained by encoding is sent to the corresponding aspect attention module. The aspect vector $v_{a,i}$ is the aspect information in each aspect attention module. The attention modules of various aspects perform operations independently to extract hidden features of specific aspects. The aspect attention module consists of an attention part, a probability part and a reconstruction part. The structure is shown in Figure 2.

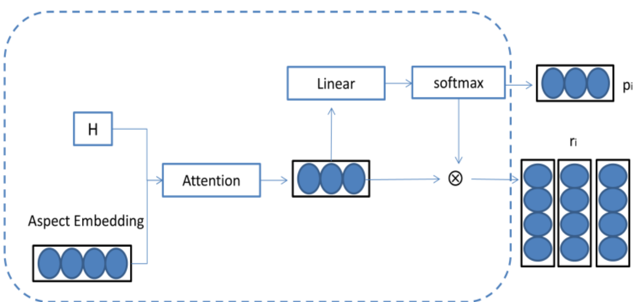

Fig. 2. Aspect-based attention module.

The attention part splices the aspect vector $v_{a,i}$ with each hidden layer state in the input H, and then performs the attention operation. The vector of specific aspect text obtained by weighted average is represented by $v_{c,i}$, and the calculation formula is as follows:

$$\begin{aligned} t_{t,i}&=\tanh(w_{a,i}[h_t,v_{a,i}])+b_{a,i}\\ a_{t,i}&=\frac{\exp(u_{t,i})}{\sum_{j=1}^{n}\exp(u_{j,i})}\\ v_{c,i}&=\sum_{t=1}^{n}a_{t,i}h_t \end{aligned} \tag{4}$$

where i represents the i-th aspect attention module; $w_{a,i}$ represents the weight matrix of attention; $b_{a,i}$ represents the bias of attention. The vector $v_{c,i}$ of the text of a specific aspect passes through the fully connected layer, and the probability of the input sample in each category is calculated by the softmax function and output. Calculated as follows:

$$p_i=softmax(w_{p,i}v_{c,i}+b_{p,i}) \tag{5}$$

where $w_{p,i}\in U^{C\times d}$ is the weight matrix of the fully connected layer; $w_{p,i}$ i the bias of the fully connected layer; c is the number of classification.

The reconstruction part reconstructs the vector $v_{c,i}$ related to the specific aspect of the text, and obtains the sentence overall vector v_s the new vector of sentiment classification in the specific aspect. By reconstructing $v_{c,i}$ by classification probability, multiple reconstructed texts can be obtained. The vector of reconstructed text is, and the calculation formula is as follows:

The reconstruction part reconstructs the vector $v_{c,i}$ related to the specific aspect of the text, and obtains the sentence overall vector v_s the new vector of sentiment classification in the specific aspect. By reconstructing $v_{c,i}$ by classification probability, multiple reconstructed texts can be obtained. The vector r_i of the reconstructed text is $r_{i,j}$, and the calculation formula is as follows:

$$r_{i,j}=p_{i,j}v_{c,i}, 1\le \mathrm{j}\le c \tag{6}$$

where $r_{i,j}\in U^d$,c is the number of classification.

III. Experiments

A. Dataset

This paper uses the SemEval2014 data set and Twitter data set for comparative experiments. The emotional polarity of the data samples is divided into positive, negative and neutral. Among them, the SemEval2014 dataset is a dataset for semantic evaluation competition tasks, including user comments in two areas: laptop and restaurant. Through comparative experiments, it is verified that the ABABiLSTM proposed in this article has achieved good sentiment classification performance on different domain data sets.

In order to better select hyperparameters of our model, this paper randomly divides 20% of the data from the training set as the validation set. This paper uses 300-dimensional Glove vector to initialize the word embedding, and sets the dimension of the BiLSTM hidden vector to 300 dimensions. In order to prevent over-fitting, this article adds a dropout layer to the LSTM layer and the output layer, and the loss probability is set to 0.5. In addition, this article also uses the L1 regular and L2 regular, and the parameters are set to 1 and 0.1 respectively. This article uses the root mean square backpropagation algorithm to update the parameters during training, and the initial value of the learning rate is set to 0.001. In the training process, this paper uses the early stopping mechanism, and stops training when the performance of the verification set no longer improves. This article trains the model, takes the best model on the validation set, and then evaluates it on the test set.

B. Baseline

Experiment with the ABABiLSTM model and 7 baselines SemEval2014 data set and Twitter data set.

1) SVM. The feature-based SVM classification method is a common method of traditional machine learning.

2) CNN. Based on the convolutional neural network model, it is the most basic convolutional neural network.

3) ATT-CNN. Based on the convolutional neural network based on the attention mechanism.

4) LSTM. The basic LSTM network uses the last hidden state as a sentence representation and inputs it into the final classifier.

5) TDLSTM. Based on the two LSTM networks proposed by Tang et al. [17], they act on the text before and after the specific aspect respectively, and then use the stitching of the last hidden state of the two LSTM networks to predict the emotional polarity.

6) AT-LSTM. Based on the LSTM network proposed by Wang et al. [18], the text context is modeled, and then the hidden state and the specific aspect are jointly embedded in the generation of the supervised attention vector, and then the generated attention vector generates the final specific aspect emotional polarity.

7) IAN. Based on Ma et al. [19] proposed using two LSTM networks to model sentences and specific aspects respectively, and interactively generate two parts of attention vector for emotion classification.

C. Analysis of Experimental Results

It can be seen from Table I that ABABiLSTM achieved the best results in all groups. By analyzing many examples in these groups, it is found that when there are multiple targets in a sentence, the adaptive target word representation can enable the attention mechanism to better capture the relevant contextual opinions of a given target. At the same time, using BiLSTM to extract the local features of the target word context also improves the classification performance. Combining the advantages of attention extraction features with the advantages of BiLSTM to extract the context has achieved better results.

TABLE I. COMPARISON OF SENTIMENT ANALYSIS ON SEMEVAL2014 DATASET AND TWITTER DATASET

Model	Accuracy (%)		
	Laptop	***Restaurant***	***Twitter***
SVM	60.23	69.16	72.34
CNN	60.45	68.94	71.59
AT-CNN	62.74	71.28	72.50
LSTM	66.41	73.28	74.47
TDLSTM	68.17	74.67	75.91
AT-LSTM	69.96	76.53	79.15
IAN	71.22	79.12	83.18
ABABiLSTM	75.59	82.15	86.87

It can be seen from the comparison of experimental results that adding aspect information or using the attention mechanism can significantly improve the classification effect. Reflected in the performance of the evaluation index, it is to improve the accuracy rate. However, because the function of the attention layer to highlight information is carried out by constant weighting calculation, the time cost is increased. The time cost of the ABABiLSTM model is relatively high. Because the calculation of the BLSTM neural network is relatively complicated, the calculation time is increased, and the attention layer also increases the weighted calculation time while highlighting the key information.

In the model that combines the attention mechanism with the neural network, the accuracy rate is significantly improved compared with the model that does not use attention, which proves the effectiveness of the attention mechanism in certain aspects of sentiment analysis tasks. Due to the limitations of CNN in NLP, the results of ATT-CNN are not ideal compared to the model that combines attention with LSTM. Both AT-LSTM and IAN are methods that combine the attention mechanism with LSTM, and both have improved accuracy. IAN models the text and specific aspects separately, and generates attention interactively. It not only learns the relatively important information in the text for a specific aspect, but also learns the more important information in the specific aspect for the text, which once again proves the effectiveness of the attention mechanism. Sex. ABABiLSTM supervises the important information in the context of a specific aspect through the attention mechanism, and embeds the specific aspect with the text jointly, which generates a more reasonable representation for the sentiment classification of the specific aspect.

IV. SUMMARY AND OUTLOOK

With the rapid development of the mobile Internet, how to efficiently extract information from the massive data in cyberspace and use it has become a key concern of all walks of life. Sentiment analysis aims to dig out the subjective emotional information contained in large-scale texts. It has become a hot research topic in the field of natural language processing and has important application value in many fields. This paper focuses on the aspect-level sentiment analysis problem, which is an important sub-problem in fine-grained sentiment analysis. The purpose of aspect-level sentiment classification is to dig out the sentiment polarity of users' opinions on a specific aspect in the text. This paper proposes a sentiment classification method combining BiLSTM and aspect attention module. BiLSTM has fewer parameters, faster model training, and can effectively extract deep-level information of text; combining the attention mechanism with aspect information can fully extract specific aspects of information.，It is verified in the datasets, and its performance is better than many advanced emotion analysis models based on attention mechanism.

ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of China (NSFC) (No.81973981); Key Project of Research and Development in Shandong Province (No.2019RKB14090); Project of Traditional Chinese Medicine and Technology Development Plan Program in Shandong province (No.2019-0018). Shandong Postgraduate Education Quality Improvement Plan（SDYKC19147）.

REFERENCES

[1] Schouten, Kim, and Flavius Frasincar. "Survey on aspect-level sentiment analysis." IEEE Transactions on Knowledge and Data Engineering 28.3 (2015): 813-830.

[2] Chen, Peng, et al. "Recurrent attention network on memory for aspect sentiment analysis." Proceedings of the 2017 conference on empirical methods in natural language processing. 2017.

[3] Xianghua, Fu, et al. "Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon." Knowledge-Based Systems 37 (2013): 186-195.

[4] Do, Hai Ha, et al. "Deep learning for aspect-based sentiment analysis: a comparative review." Expert Systems with Applications 118 (2019): 272-299.

[5] Phan, Minh Hieu, and Philip O. Ogunbona. "Modelling context and syntactical features for aspect-based sentiment analysis." Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.

[6] Thet, Tun Thura, Jin-Cheon Na, and Christopher SG Khoo. "Aspect-based sentiment analysis of movie reviews on discussion boards." Journal of information science 36.6 (2010): 823-848.

[7] García-Pablos, Aitor, Montse Cuadros, and German Rigau. "W2VLDA: almost unsupervised system for aspect based sentiment analysis." Expert Systems with Applications 91 (2018): 127-137.

[8] Hu, Mengting, et al. "Can: Constrained attention networks for multi-aspect sentiment analysis." arXiv preprint arXiv:1812.10735 (2018).

[9] Peng, Haiyun, et al. "Knowing what, how and why: A near complete solution for aspect-based sentiment analysis." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. No. 05. 2020.

[10] Ruder, Sebastian, Parsa Ghaffari, and John G. Breslin. "A hierarchical model of reviews for aspect-based sentiment analysis." arXiv preprint arXiv:1609.02745 (2016).

[11] Dong L, Wei F, Tan C, et al. “Adaptive recursive neural network for target-dependent twitter sentiment classification”, Proceedings of the 52nd annual meeting of the association for computational linguistics. 2014: 49-54.

[12] Nguyen T H, Shirai K, “Phrasernn: Phrase recursive neural network for aspect-based sentiment analysis”, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015: 2509-2514.

[13] Wang W, Pan S J, Dahlmeier D, et al, “Recursive neural conditional random fields for aspect-based sentiment analysis”, arXiv preprint arXiv:1603.06679, 2016.

[14] Vo D T, Zhang Y, “Target-dependent twitter sentiment classification with rich automatic features”, Twenty-fourth international joint conference on artificial intelligence. 2015.

[15] Zhang, Meishan, Yue Zhang, and Duy-Tin Vo. "Gated neural networks for targeted sentiment analysis." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 30. No. 1. 2016.

[16] Tang, Duyu, et al. "Effective LSTMs for target-dependent sentiment classification." arXiv preprint arXiv:1512.01100 (2015).

[17] Tang, Duyu, et al. "Effective LSTMs for target-dependent sentiment classification." arXiv preprint arXiv:1512.01100 (2015).

[18] Wang, Yequan, et al. "Attention-based LSTM for aspect-level sentiment classification." Proceedings of the 2016 conference on empirical methods in natural language processing. 2016.

[19] Ma, Dehong, et al. "Interactive attention networks for aspect-level sentiment classification." arXiv preprint arXiv:1709.00893 (2017).