A Lexicon-Enhanced Attention Network for Aspect-Level Sentiment Analysis

ZHIYING REN [1,3], **GUANGPING ZENG** [1,3], **LIU CHEN** [1,3], **QINGCHUAN ZHANG** [2], **CHUNGUANG ZHANG** [1,3], **AND DINGQI PAN** [1,3]

[1] University of Science and Technology Beijing, Beijing 100083, China
[2] National Engineering Laboratory for Agri-product Quality Traceability, Beijing Technology and Business University, Beijing 100048, China
[3] Beijing Key Laboratory of Knowledge Engineering for Materials Science, Beijing 100083, China

Corresponding author: Zhiying Ren (ryy0714@126.com) and Guangping Zeng (zenggpustb@163.com)

ABSTRACT Aspect-level sentiment classification is a fine-grained task in sentiment analysis. In recent years, researchers have realized the importance of the relationship between aspect term and sentence and many classification models based on deep learning network have been proposed. However, these end-to-end deep neural network models lack flexibility and do not consider the sentiment word information in existing methods. Therefore, we propose a lexicon-enhanced attention network (LEAN) based on bidirectional LSTM. LEAN not only can catch the sentiment words in a sentence but also concentrate on specific aspect information in a sentence. Moreover, leveraging lexicon information will enhance the model's flexibility and robustness. We experiment on the SemEval 2014 dataset and results find that our model achieves state-of-the-art performance on aspect-level sentiment classification.

INDEX TERMS Natural language processing, sentiment analysis, aspect-level, sentiment lexicon, attention mechanism.

I. INTRODUCTION

Sentiment analysis (also called opinion mining) has been one of the most active fields in Natural Language Processing (NLP) due to its important value to business and society. It is the field of study that tries to extract opinion (positive, neutral, negative). However, the meaning of the same word in different sentences may be opposite, or the polarity of the same entity in the same sentence is different in different aspects. Therefore, aspect level sentiment classification is proposed to identify opinions from a text about specific entities and their aspect.

One important challenge in aspect-level sentiment analysis is how to model the semantic relationship between aspect terms and sentences. Many methods have been proposed to solve the problem. The traditional NLP approach is a method based on sentimental knowledge, which relies on the existing sentiment lexicon or domain lexicon in subjective text and the combined evaluation unit with sentiment

The associate editor coordinating the review of this manuscript and approving it for publication was Shiping Wen [ID].

polarity. Some of them use the features to train a sentiment classifier, such as Support Vector Machines (SVM). A common observation in the sentiment analysis field is that many high-quality sentiment lexicons, such as Bing Liu's Opinion Lexicon [1], NRC Sentiment 140 lexicon [2] and Subjectivity Lexicon [3]. A sentiment lexicon contains many sentiment words and their sentiment polarities. As for the sentiment classification, the lexicon is a useful resource. Incorporating sentiment lexicons into neural sentiment classification methods has attracted increasing attention recently [4]–[6]. Since these sentiment words are an important part to convey the sentimental polarity of sentences, and the sentiment lexicons play an important role in sentence sentiment classification. However, text sentiment classification based on sentiment dictionary has shortcomings as follows, first of all, text sentiment classification has multiple accurate discrimination criterion. Secondly, human language is a rather complicated cultural product. Because a sentence is not a simple linear combination of words, it has a rather complicated nonlinearity. Finally, different combinations, different orders, and different numbers of words can bring different meanings

and emotions, which leads to text emotions is difficult to classification.

Deep network architecture is widely used in various fields [7], [8], such as face recognition [9], video processing [10], language processing [11], etc. As for the text classification, the main reason why deep learning techniques are superior to traditional learning techniques is that they automatically learn semantic representations from high dimensional original data without carefully designed feature engineering. In the literature, lots of network applied to target-dependent sentiment analysis, such as Target-Dependent LSTM(TD-LSTM) [12] and Target-Connection LSTM (TC-LSTM) [12]. However, the network cannot catch which words should be paid attention in a sentence. Fortunately, attention mechanisms can effectively solve the problem. For example, Wang *et al.* [13] proposes AE-LSTM network, AT-LSTM and ATAE-LSTM network to address the aspect-level sentiment analysis.

Recently, incorporating sentiment lexicons into neural sentiment classification methods is used in aspect-level sentiment analysis. Wu *et al.* [6]. proposes a novel sentiment lexicon enhanced attestation-based LSTM(SLEA-LSTM) model to improve the performance of sentence-level sentiment classification. Shin *et al.* [14] performs separate convolutions for word and lexicon embeddings. Lu *et al.* [15] proposes an interactive rule attention network (IRAN) which includes a grammar rule encoder and constructs an interaction attention network to learn attention information from context and target. However, previous papers rarely notice both affective and aspect words. Thus, how to use the attention network to capture sentiment word information and aspect information in sentences is the key to improving the performance of sentiment classification.

We explore the potential correlation between sentiment words and aspect information and sentence polarity in aspect-level sentiment classification. To be specific, our model consists of three components: 1) Lexicon-enhanced attention network (LEAN) model composes of two Bi-LSTM networks focusing on extracting the aspect features and sentence-level features respectively. 2) Obtaining the sentiment word information in the corresponding sentence and then converting the sentiment information into sentiment embedding. 3) Using a bidirectional attention mechanism to model aspect information and corresponding sentences. We evaluate our model on SemEval 2014 Datasets, and find that our model is more effective than other previous methods.

The main contributions of our work can be summarized as follows: 1) We attempt to explicitly investigate the effectiveness of the sentiment lexicon information for aspect-level sentiment analysis. 2) We propose a Lexicon-Enhanced Attention Network model based on Bi-LSTM, which leverage the high-quality sentiment lexicon to identify the emotional polarity of a sentence in a certain aspect. This has been shown to effectively improve sentiment analysis performance. 3) We apply a bidirectional attention mechanism,

which can enhance the mutual relation between the aspect term and its corresponding sentence.

The remainder of this paper is organized as follows. Section II discusses the related work. Section III gives details description of our model. Section IV shows comparative experimental results to justify the effectiveness of our model. Section V gives case studies to further elaborates model performance. Finally, it summarizes this work in Section VI.

II. RELATED WORK
A. ASPECT BASED SENTIMENT ANALYSIS

Aspect-level sentiment classification is often viewed as a classification problem in the literature. As the granularity continues to be divided, people try to discover the polarity of the sentence in a specific aspect. At first, people use emotional knowledge-based methods for aspect-level sentiment classification. It mainly relies on some existing sentiment dictionaries or domain dictionaries or combines evaluation units with emotional polarity in a subjective text to calculate the polarity of subjective text. Whitelaw *et al.* [16] uses a semi-automatic method combined the WordNet to construct a dictionary of adjective evaluation words and a dictionary of modifier words. Then, they calculate the attribute values of emotional words in a sentence to determine the polarity of the text. However, the results highly depend on the quality of features.

Machine learning is also a mainstream sentiment classification method. These methods mainly involve text representation and feature extraction, such as bag-of-words models and sentiment lexicon features, then training a sentiment classifier, such as support vector machine (SVM) and logistic regression (LR). Kiritchenko *et al.* [17] describes supervised machine-learning approaches are used to detect aspect terms and aspect categories as well as detect sentiment expressed towards aspect terms and aspect categories in customer reviews. However, traditional machine learning methods usually cannot model the contexts of many important features.

Therefore, a simple and effective approach to learn distributed representations is proposed by Mikolov *et al.* [18]. Also, the attention mechanism is widely applied to NLP fields. The neural networks advance sentiment analysis substantially [19]. Tang *et al.* [12] proposes a Target-Dependent Long Short-Term Memory model (TD-LSTM) and a target-connection long short-term memory (TC-LSTM). TD-LSTM split the whole context into two components, i.e., the left context with target and the right context with target adopt to a Long short-term memory (LSTM), respectively. The model concatenates the two target-specific representation as input for sentiment classification. The structure of the TC-LSTM is similar to the TD-LSTM and the only difference is TC-LSTM takes as input the concatenation of each word vector and target vector to incorporate the semantic relatedness of a target with its context words. To capture the sentence polarity of a particular aspect more accurately. Wang *et al.* [13] designs an attention-based LSTM with

Aspect Embedding (ATAE-LSTM), which append the word embedding and aspect embedding. Ma *et al.* [20] proposes the interactive attention networks (IAN) to interactively learn attentions in the contexts and the target can well represent a target and its collocative context. Tay *et al.* [21] proposes MNN-2 model which attempts to automatically learn the aspect-sentiment relationship between aspect and sentence sentiment.

B. LEXICON ENHANCED SENTIMENT ANALYSIS
In recent years, deep neural networks have been widely used in sentiment classification. Adding emotional lexicon information into neural networks can improve the performance of sentence-level sentiment classification. According to Mihalcea *et al.* [22], ''the meaning of a word has a certain relationship with its polarity, but the same meaning does not necessarily have the same polarity.'' Given this, the advantage of the dictionary-based method lies in the evaluation obtained. Therefore, lexicon-enhanced sentiment analysis is widely studied. Shin *et al.* [14] shows that using naïve concatenation, multichannel and separate convolution to integrate lexicon embeddings and an attention mechanism on Convolutional Neural Networks (CNN) can improve the accuracy, stability, and efficiency in sentiment classification. Lei *et al.* [5] proposes a novel sentiment lexicon enhance attention-based LSTM(SLEA-LSTM) model. Their method uses single-head and multi-head attention mechanisms with sentiment lexicon into the deep nature network. Wu *et al.* [6] set a word sentiment classification task to classify the sentiments of words in a sentence based on their hidden representations in the attention network. This method judges the sentiment polarity information and jointly with the natural sentiment classification. The above methods all use the attention network of sentiment lexicon for sentiment analysis. However, no attention has been paid to aspect information in a sentence. What's more, Bao *et al.* [4] describes an approach of leveraging numerical polarity features provided by existing lexicon resources in an aspect-based sentiment analysis environment with an attention LSTM. It is based on the AT-LSTM model and linearly transforms the emotional word embedding into the final regularization along with the sentence embedding.

As mentioned above, aspect information and sentimental word information can help attention to the closely related part of the context. Therefore, we build LEAN model which respectively sentiment lexicons and aspect words to compute the attention vector and learn the representation. In this way, LEAN can well acquire the appropriate final representation of aspect-level sentiment classification.

III. LEXICON-ENHANCED ATTENTION NETWORK
In this section, we describe the proposed model Lexicon-Enhanced Attention Network (LEAN) for aspect-level sentiment analysis and LEAN is shown in Figure 1.

A. LONG SHORT-TERM MEMORY (LSTM)
Long Short-term Memory network (LSTM) is first proposed by Hochreiter and Schmidhuber [23] which can overcome the gradient vanishing or exploding problems of Recurrent Neural Network (RNN) [24]. The main idea is to use three gates and cell memory state, which enable to keep the previous state and memorize the extracted features of the current data input. Given a sequence $S = [x_1, x_2, \ldots, x_l]$, where l is the length of input text and then give the word by word to LSTM. At time-step t, the memory c_t and the hidden state h_t are updated with the following equations:

$$\begin{bmatrix} i_t \\ f_t \\ o_t \\ \widehat{c_t} \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ tanh \end{bmatrix} W \cdot [h_{t-1}, x_t], \quad (1)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \widehat{c_t}, \quad (2)$$

$$h_t = o_t \odot \tanh(c_t), \quad (3)$$

where x_t is the input at the current time-step, i, f and o is the input gate activation, forget gate activation and output activation respectively. $\widehat{c_t}$ is the current cell state, σ denotes the logistic sigmoid function and $\odot$ denotes element-wise multiplication.

B. HIDDEN REPRESENTATION USING BI-LSTM MODEL
For the sequence modeling tasks, it is beneficial to have access to the past context as well as the future context. Schuster and Paliwal [25] proposed bidirectional LSTM model (Bi-LSTM) which can model the context dependency with forward LSTM and the backward LSTM. This model acquires the annotation of words by summing up information. The output of the ith word is shown in the following equation,

$$h_i = \left[\overrightarrow{h_i} \oplus \overleftarrow{h_i} \right]. \quad (4)$$

In this paper, we use Bi-LSTM to obtain the hidden representation of sentence, aspect and sentiment. Firstly, we obtain the representation of each word in aspect term and sentences, and formalize the notations in our work. We suppose that a sentence consists of M words $\left[w_1^c, w_2^c, w_3^c, \ldots, w_M^c \right]$ and an aspect term $\left[w_1^a, w_2^a, w_3^a, \ldots, w_M^a \right]$. We embed each word in a sentence to get context embedding $v^c \in \mathbb{R}^{d \times M}$ and aspect imbedding $v^a \in \mathbb{R}^{d_a}$ by looking up a word embedding matrix respectively, where d is size of hidden layers. Also, we get N sentiment words $\left[w_1^s, w_2^s, w_3^s, \ldots, w_N^s \right]$ and get the sentiment embedding $v^s = \left[v_1^s, v_2^s, v_3^s, \ldots, v_N^s \right]$. Then we input sentiment embedding v^s into the left Bi-LSTM to get the hidden representation H^s consisting of $\left[h_1^s, h_2^s, h_3^s, \ldots, h_N^s \right]$. For right Bi-LSTM structure, we take the concatenation of the aspect embedding v^a and sentence embedding v^c as the inputs, then, we obtain the final hidden representation $H^{ac}. \in \mathbb{R}^{d \times M}$ consist of $\left[h_1^{as}, h_2^{as}, h_3^{as}, \ldots, h_M^{as} \right]$ as the input.

C. BIDIRECTIONAL ATTENTION MECHANISM IN LEAN
As shown in figure 1, Bi-LSTM attention model in LEAN consists of two parts: including an aspect term to the
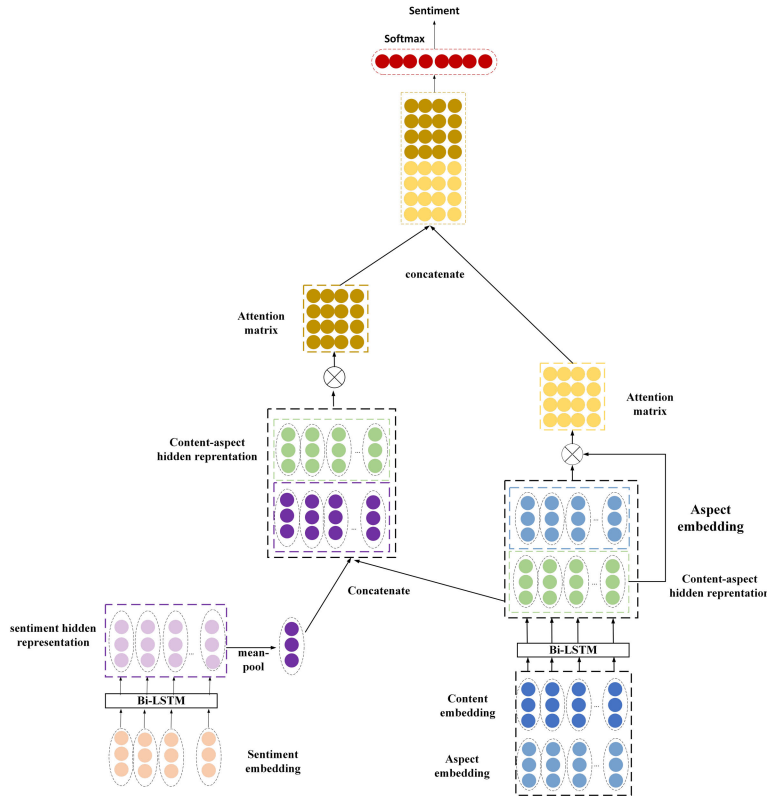
FIGURE 1. The architecture of the lexicon-Enhanced Attention Network for aspect-level sentiment analysis (LEAN). The right part is an aspect term to the aspect sentence attention model and the left part is Lexicon-enhanced attention to aspect sentence model.

aspect sentence attention part and the lexicon-enhanced to aspect sentence part. For the former part, we are able to obtain different parts of a sentence when different aspects are concerned [26]. For the later part, we leverage the sentiment lexicon information into an aspect-based sentences to focus on the affective words in the sentences.

1) AN ASPECT TERM TO THE ASPECT SENTENCE ATTENTION PART

Words in a sentence have different polarities in different aspects. Thus, we must concentrate on the effect of aspect term to the sentence polarity. In this part, we replicate ATAE-LSTM model as the baseline model. Furthermore, we get the final hidden contextual representation of the concatenation of the aspect embedding and sentence embedding by the right Bi-LSTM in Figure 1. According to Lei *et al.* [5], multi-head attention produces attention that allows jointly focus on the information from different representation subspaces at different positions. Thus, we use multi-head attention with concatenation function to calculate the weight representation of sentences with the given aspect using following equations:

$$r_{ac} = \sum_i^M \alpha_{i,j} h_i^{ac}, \qquad (5)$$

$$\alpha_{i,j} = \frac{exp\left(\sigma\left(\left[h_i^{ac}; v_j^a\right]\right)\right)}{\sum_i^M exp\left(\sigma\left(\left[h_i^{ac}; v_j^a\right]\right)\right)}, \qquad (6)$$

$$\sigma\left(\left[h_i^{ac}, v_j^a\right]\right) = u_{v1}^T tanh\left(W_h\left[h_i^{ac}; v_j^a\right]\right), \qquad (7)$$

where $\left[h_i^{ac}; v_j^a\right]$ denotes the concatenation of h_i^{ac} and v_j^a. W_h and u_{v1}^T are the parameters to learn. $\alpha_{i,j}$ is a vector consisting of attention weights and r_{ac} is the weighted representation of sentence with aspect.

2) LEXICON-ENHANCED ATTENTION TO ASPECT SENTENCE

Sentiment words are important clues of sentiment and are very informative for inferring the sentiments of a sentence. Therefore, we get the hidden contextual representation of the inputs by the left Bi-LSTM in Figure 1 and the hidden representation of sentiment words. Then we use the aspect term and sentence information to calculate the attention weight with sentiment features:

$$r_{cs} = \sum_i^M \gamma_i h_i^{ac}, \qquad (8)$$

$$\gamma_i = \frac{exp\left(\sigma\left(\left[h_i^{ac}; z_s\right]\right)\right)}{\sum_i^M exp\left(\sigma\left(\left[h_i^{ac}; z_s\right]\right)\right)}, \qquad (9)$$

$$\sigma\left(\left[h_i^{ac}, z_s\right]\right) = u_{s1}^T tanh\left(W_h\left[h_i^{ac}; z_s\right]\right), \qquad (10)$$

$$z_s = \sum_i^N h_i^s / N, \qquad (11)$$

where γ_i stands for the attention weights from aspect term and sentence with the sentiment words, which would focus on the sentiment words in a sentence in a different aspect. z_s is the output from the mean-pool layer, which donated an entire representation of sentiment words.

Later, we concatenate the two weighted representation from above two parts and get the sequence representation x^* is obtained by using a non-linear layer:

$$x^* = tanh\left(W_{acs}\left[r_{ac}; r_{cs}\right] + b_{acs}\right), \qquad (12)$$

where W_{acs} and b_{acs} are the weight matrix and bias respectively.

We feed x into a linear layer, which length is equals with the number of class labels. Finally, a *softmax* layer followed to judge the sentiment polarities as positive, negative, neutral:

$$y = softmax\left(W_s x^* + b_s\right), \qquad (13)$$

where W_s and b_s are the parameters for the *softmax* layer.

D. MODEL TRAINING
The model can be trained in an end-to-end way in order to optimize the whole parameters and minimize the loss function as much as possible. In our work, we let y is the target distribution for sentence, $\widehat{y}$ is the predicted sentence distribution and use cross-entropy as the loss function:

$$loss = -\sum_i \sum_j y_i^j \log \widehat{y}_i^j + \lambda \|\theta\|^2, \qquad (14)$$

where i is an index of sentence and j is an index of class. λ is L_2 regularization factor. θ is the parameter set.

IV. EXPERIMENTS
In this section, we present our experiment settings and conduct experiments on the task of aspect-level sentiment classification.

A. EXPERIMENTS SETTING
In our experiments, all the word vectors are initialized by Pennington *et al.* [27] (Pre-trained word vectors of Glove can be obtained from http://nlp.stanford.edu/projects/glove/). The dimension of sentence, aspect term and sentiment words are set to 300, and the number of hidden units is set to 200. We use Keras to implement our nature work. Furthermore, a momentum of 0.9, L_2 regularization weight of 0.001 and the learning rate of 0.01 for AdaGrad.

B. DATASET
To evaluate our proposed model, we conduct experiments on SemEval 2014 Task 4 (The introduction of this dataset can be seen at: http://alt.qcri.org/semeval2014/task4/). It consists of reviews in Restaurant and Laptop datasets. Each review determines to whether the polarity of each aspect term is positive, negative or neutral. For example, there is an aspect term staff is negative in the sentence "But the staff was so horrible to us". The statistics are presented in Table 1.

TABLE 1. Statistic of SemEval 2014 datasets on restaurant and laptop.

Dataset	positive		Negative		Neural	
	Train	Test	Train	Test	Train	Test
Restaurant	2012	288	692	99	580	83
Laptop	896	331	776	287	450	167

C. SENTIMENT LEXICON
In this paper, we build our lexicon by merging 2 existing lexicons: Opinion Lexicon [1] and Subjectivity Lexicon [3]. The lexicons include sentiment information and the frequency of positive and positive polarity. Specifically, the fused lexicon contains 4,723 positive words such as "favorable", "attractive", "moderate" etc. And 4,781 negative words such as "unpleasant", "worst", "unusable" etc.

D. EVALUATION METRICS
To evaluate the performance of the model, we have used the classification *Accuracy* to measure the overall sentiment classification performance. The *Accuracy* can be calculated by formulation (15) and Figure 2 is the loss function graph.

$$Accuracy = \frac{T}{N}, \qquad (15)$$

where T it the number of samples correctly predicted and N is the total number of the test dataset.
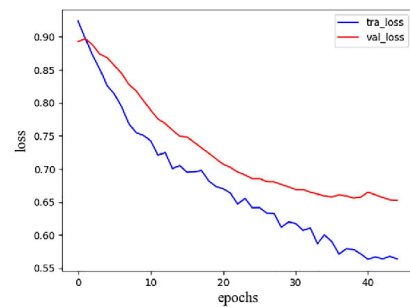

FIGURE 2. The loss function graph after smoothing.

E. BASELINES
To evaluate the performance of LEAN, we compare our model with several baseline approaches. The baselines are introduced as follows.

LSTM: The paper takes LSTM network to model the sentence with opinions. Then it uses the last hidden vector as the sentence representation, and inputs it to the softmax layer to predict the polarity of each sentence. But this method does not capture the features in some aspect term of the sentence [13].

TD-LSTM: TD-LSTM firstly uses the target string as the last unit of the sentence. Then taking a sentence with the target string into left LSTM and right LSTM. Afterwards, they concatenate the last hidden vectors of left LSTM and right LSTM and feed them to a softmax layer to predict the polarity of sentence [12].

TC-LSTM: TC-LSTM is structurally similar to TD-LSTM. However, in TC-LSTM the input of each word is concatenation of word embedding and target embedding. Thus, it can make much better use of the connection between target each context word [12].

AT-LSTM: AT-LSTM is first to propose attention-based LSTM network for aspect-level sentiment analysis. It devises a model that set the aspect embeddings which concatenate the aspect embedding with the word embedding as the hidden contextual representation and finally calculates the attention weight. Then using the weight to judge the sentiment polarity [13].

ATAE-LSTM: ATAE-LSTM extends AE-LSTM by attaching aspect embedding content to each word embedding to enhance the importance of aspect information. Then, the output of hidden representations concatenates with the aspect embedding to compute the attention weights for predicting the sentiment polarity [13].

IAN: IAN uses two attention networks to model the target and context interactively. The model can pay close attention to the important parts of the target and context. Finally, its representation for predicting the polarity of sentence [20].

MNN-2: MNN-2 uses CNN and LSTM to capture textual information at the word level and character level. Then use the self-attention mechanism and interactive-attention mechanism to focus on the key emotional cues for the target. Finally, the sequence of hidden representation is inputted into the CRF layer to predict the sentence polarity [21].

F. MODEL COMPARISONS

Table 2 shows the sentiment classification performance of our model and baseline models on the "restaurant and laptop" dataset. We can observe that our proposed LEAN model achieves the best performance among all methods. It can be seen that LSTM model gets the worst performance of all nature network baseline methods. Because it ignores the importance of aspect information and treats aspect information equally with the sentence information. TD-LSTM and TC-LSTM all process the left and right contexts with target and focus the target representations, which are obtained from

TABLE 2. Comparison with baselines. The accuracy on Three-class prediction about Restaurant and Laptop dataset.

Datasets	Restaurant	Laptop
LSTM	0.743	0.665
TD-LSTM	0.756	0.681
TC-LSTM	0.760	0.682
AT-LSTM	0.762	0.689
ATAE-LSTM	0.772	0.687
IAN	0.786	0.721
MNN	0.791	0.729
LEAN	**0.791**	**0.737**

word vectors into the input of LSTM cell unit. Therefore, the performance of TD-LSTM and TC-LSTM exceed about 1% and 2% then LSTM on the Restaurant dataset and exceed about 1.5% and 1.6% then LSTM on the Laptop dataset. Furthermore, both AT-LSTM and ATAE-LSTM perform better than TC-LSTM model. Because they not only consider the importance of aspect information in a sentence but also utilize the attention mechanism. Especially, ATAE-LSTM appends the aspect embedding to each word embedding and treats them as input. So, it gets better performance than AT-LSTM. Moreover, IAN knows the importance of the aspect term and context, thus, it utilizes the attention mechanism associated with a target to get important information from the context and context representation for sentiment classification. Also, it makes use of interactive information from context to supervise the modeling of the target. Finally, IAN predicts the sentiment polarity by using the two representations. The MNN-2 model does not use a manual functional design or any external language resources but is committed to solving the end-to-end ABSA one-stop solution by using a multi-task neural learning framework. The model uses CNN and Bi-LSTM to capture the sentence information in the sentence sequence, respectively. Then, use self-attention mechanism and interactive attention mechanism to pay attention to the relationship between aspect information and sentence polarity. This method not only focuses on using Bi-LSTM which can capture the current word and preceding words and successive words information but also uses the attention mechanism to pay attention to sentence information. Compared with the MNN-2 model, our method has no obvious advantage in the restaurant dataset, but on the Laptop dataset, the accuracy rate is improved by 0.8%.

According to the idea of the baseline models, LEAN considers both the aspect information and sentiment word information in the sentence. Unlike ATAE-LSTM, LEAN uses a bidirectional LSTM network to capture information. Meanwhile, we use existing sentiment dictionaries that classify sentiment words. We use emotional words in the model to make sentences more focused on words with emotion, which is what our emotional perception attention model does. Inspired by the IAN and MNN-2 models learning the attention between the aspect words and their corresponding sentences in an interactive way, we use a bidirectional attention mechanism to pay attention to the influence of the sentiment words in the sentence on the polarity and capture the aspect information.

G. ANALYSIS OF LEAN MODEL

In this section, we will design a series of models to prove the validity of our LEAN model. Firstly, we designed a LEAN model without sentiment lexicon and its structure is similar to the ATAE-LSTM model. The only difference from the ATAE-LSTM is that we use attention mechanism based on Bi- LSTM network instead of LSTM network to obtain hidden re- presentations. We test the hidden representation of the concatenate of aspect embedding and sentence embedding as

Aspect term	Sentiment words	Sentence	polarity
food and service	highly disappointed	I was highly disappointed by their service and food	negative
pizza	good and huge	The pizza was pretty good and huge	positive

FIGURE 3. Case study: the visualized attention weights for sentiment words and aspect term in a sentence by LEAN. The color depth indicates the importance degree of the weight.

the input of Bi-LSTM without using the sentiment lexicon to judge the sentiment polarity of the sentence. The analysis of LEAN model is shown in Table 3, we can conclude that LEAN has a better performance comparing the other model.

TABLE 3. Analysis of LEAN model.

Dataset	Restaurant	Laptop
ATAE-LSTM	0.772	0.687
LEAN (without sentiment lexicon)	0.780	0.691
LEAN	0.791	0.737

Because Bi-LSTM structure has a big advantage over LSTM. It is obvious that ATAE-Bi-LSTM has better performance than ATAE-LSTM. As for the AEAN, we can see that the sentiment lexicon can improve the accuracy of the sentiment classification. In our model, adding the attention mechanism based on sentiment dictionary can pay more attention to the emotional words in the sentence, and with using the attention mechanism based on aspect words, which effectively improves the accuracy of sentiment classification. From Table 3, it can be seen that LEAN cores 1.1% and 4.6% higher on the Restaurant and Laptop dataset than LEAN without sentiment lexicon respectively. Now we can conclude that our design is effective when addressing sentiment-level classification.

V. CASE STUDY

To have an intuitive understanding of our model, we visualize the attention weights on the aspect term and sentiment words of a sentence in Figure 3. The color depth indicates the importance degree of the weight. In the first sentence, "*I was highly disappointed by their service and food.*", the evaluation of both service and things is negative. Our model can discriminate the different aspect term and identifies the negative sentiment word "*highly disappointed*" in the sentence as same time, which is enable to help the judge the sentiment polarity. In second sentence, "*The pizza was pretty good and huge.*". Both the aspect term, "*pizza*", and the sentiment words influent the sentiment of the sentence. We can find that LEAN not only recognizes the aspect words, but also

catches the emotional words "*pretty good*" and "*huge*" that have a great influence on the polarity of sentiment. The lexicon-Enhanced Attention Network that integrating aspect words and affective words, greatly improving classification efficiency.

VI. CONCLUSION

In this paper, we propose a Lexicon-Enhanced Attention Net-work (LEAN) based on Bi-LSTM for aspect-level sentiment analysis. The main idea of LEAN is to leverage the exits sentiment lexicon to catch the sentiment information in a sentence to calculate the attention weights. Also, LEAN adopts the bidirectional attention mechanism, which can catch the aspect information of a sentence and find the relationships with the sentiment words. LEAN experiment on SemEval 2014 Datasets and obtain superior performance over the baseline models.

REFERENCES

[1] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. Knowl. Discovery Data Mining*, 2004, pp. 168–177.

[2] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment analysis of short informal texts," *J. Artif. Intell. Res.*, vol. 50, pp. 723–762, Aug. 2014.

[3] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proc. Conf. Hum. Lang. Technol. Empirical Methods Natural Lang. Process. (HLT)*, 2005, pp. 347–354.

[4] L. Bao, P. Lambert, and T. Badia, "Attention and lexicon regularized LSTM for aspect-based sentiment analysis," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics, Student Res. Workshop*, 2019, pp. 253–259.

[5] Z. Lei, Y. Yang, and M. Yang, "Sentiment lexicon enhanced attention-based LSTM for sentiment classification," in *Proc. Nat. Conf. Artif. Intell.*, 2018, pp. 8105–8106.

[6] C. Wu, F. Wu, J. Liu, Y. Huang, and X. Xie, "Sentiment lexicon enhanced neural sentiment classification," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 1091–1100.

[7] Y. Cao, Y. Cao, S. Wen, T. Huang, and Z. Zeng, "Passivity analysis of delayed reaction-diffusion memristor-based neural networks," *Neural Netw.*, vol. 109, pp. 159–167, Jan. 2019.

[8] Y. Cao, S. Wang, Z. Guo, T. Huang, and S. Wen, "Synchronization of memristive neural networks with leakage delay and parameters mismatch via event-triggered control," *Neural Netw.*, vol. 119, pp. 178–189, Nov. 2019.

[9] M. Dong, S. Wen, Z. Zeng, Z. Yan, and T. Huang, "Sparse fully convolutional network for face labeling," *Neurocomputing*, vol. 331, pp. 465–472, Feb. 2019.

[10] S. Wen, W. Liu, Y. Yang, T. Huang, and Z. Zeng, "Generating realistic videos from keyframes with concatenated GANs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2337–2348, Aug. 2019.

[11] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: https://arxiv.org/abs/1406.1078

[12] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective LSTMs for target-dependent sentiment classification," 2015, *arXiv:1512.01100*. [Online]. Available: https://arxiv.org/abs/1512.01100

[13] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 606–615.

[14] B. Shin, T. Lee, and J. D. Choi, "Lexicon integrated CNN models with attention for sentiment analysis," 2016, *arXiv:1610.06272*. [Online]. Available: https://arxiv.org/abs/1610.06272

[15] Q. Lu, Z. Zhu, D. Zhang, W. Wu, and Q. Guo, "Interactive rule attention network for aspect-level sentiment analysis," *IEEE Access*, vol. 8, pp. 52505–52516, 2020, doi: 10.1109/ACCESS.2020.2981139.

[16] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," in *Proc. 14th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2005, pp. 625–631.

[17] S. Kiritchenko, X. Zhu, C. Cherry, and S. Mohammad, "NRC-Canada-2014: Detecting aspects and sentiment in customer reviews," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, 2014, pp. 437–442.

[18] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," 2013, *arXiv:1310.4546*. [Online]. Available: http://arxiv.org/abs/1310.4546

[19] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," 2014, *arXiv:1406.6247*. [Online]. Available: https://arxiv.org/abs/1406.6247

[20] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 4068–4074.

[21] Y. Tay, L. A. Tuan, and S. C. Hui, "Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 1–8.

[22] R. Mihalcea, C. Banea, and J. Wiebe, "Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis," in *Proc. Meeting Assoc. Comput. Linguistics*, 2007, pp. 976–983.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[24] J. Cui, Q. Gao, and D. Li, "Improved long short-term memory network based short term load forecasting," in *Proc. Chin. Automat. Congr.*, 2019, pp. 4428–4433.

[25] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[26] X. Yao, "Attention-based BiLSTM neural networks for sentiment classification of short texts," in *Proc. Inf. Sci. Cloud Comput. PoS (ISCC)*, Feb. 2018, pp. 110–117.

[27] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.

GUANGPING ZENG received the Ph.D. degree from the University of Science and Technology Beijing, China. He worked as a Postdoctoral Researcher with the Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley for two years. He is currently a Full Professor and the Ph.D. Supervisor of computer science and technology with the University of Science and Technology Beijing, China, and the Deputy Director of the Beijing Key Laboratory of Materials Science Knowledge Engineering. His interested fields include distributed and migrating computing, linux operating systems and embedded systems, intelligent robotics and softman technology, smart systems and soft computing as well as natural language processing, and data mining.
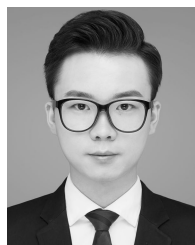
LIU CHEN received the B.E. degree in information security from the University of Science and Technology Beijing, China, where he is currently pursuing the Ph.D. degree in computer science and technology. His researches focus on applications in natural language processing, knowledge representation learning, recommendation systems with graph neural networks, and probabilistic graph models.

QINGCHUAN ZHANG received the B.E. and M.S. degrees with the Department of Computer Science and Technology, Hebei University of Science and Technology, and the Ph.D. degree in computer science and technology from the University of Science and Technology Beijing, China. He has been a Postdoc with the University of Science and Technology Beijing, for three years. He is currently an Associate Professor with the School of Computer and Information Engineering, Beijing Technology and Business University, China. His research interests include semantic computing, natural language processing, big data, and artificial intelligence.

CHUNGUANG ZHANG is currently pursuing the Ph.D. degree in computer science and technology with the University of Science and Technology Beijing, China. His research fields include artificial intelligence, intelligent networks and communications, data mining, and the Internet of Things.

ZHIYING REN received the B.E. degree in software engineering from Shanxi University. She is currently pursuing the M.S. degree in computer science and technology with the University of Science and Technology Beijing, China. She is currently a Researcher with the Beijing Key Laboratory of Knowledge Engineering for Materials Science. Her current interested fields include natural language processing and semantic computing.

DINGQI PAN is currently pursuing the Ph.D. degree with the Department of Computer Science and Technology, School of Computer and Communication Engineering, University of Science and Technology Beijing, China. His research fields include DIS systems and flight simulation systems, AI systems and intelligent networks, communications, and data mining.

• • •