IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 8, 2019, accepted November 7, 2019, date of publication November 11, 2019, date of current version December 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2952888

Aspect Based Sentiment Analysis With Feature Enhanced Attention CNN-BiLSTM

WEI MENG[1], **YONGQING WEI**[2], **PEIYU LIU**[1], **ZHENFANG ZHU**[3], **AND HONGXIA YIN**[1]

[1]School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China
[2]Basic Education Department, Shandong Police College, Jinan 250014, China
[3]School of Information Science and Electrical Engineering, Shandong Jiaotong University, Jinan 250357, China

Corresponding author: Yongqing Wei (weiyongqing@sdpc.edu.cn)

This work was supported in part by the National Social Science Fund under Grant 19BYY076, in part by the Science Foundation of Ministry of Education of China under Grant 14YJC860042, and in part by the Shandong Provincial Social Science Planning Project under Grant 19BJCJ51, Grant 18CXWJ01, and Grant 18BJYJ04.

ABSTRACT Previous work has recognized the importance of using the attention mechanism to obtain the interaction between aspect words and contexts for sentiment analysis. However, for the most attention mechanisms, it is unrigorous to use the average vector of the aspect words to calculate the context attention. Besides, the feature extraction ability of the model is also essential for effective analysis, the combination of CNN and LSTM can enhance the feature extraction ability and semantic expression ability of the model, which is also a popular research trend. This paper introduces an aspect level neural network for sentiment analysis named Feature Enhanced Attention CNN-BiLSTM (FEA-NN). Our method is to extract a higher-level phrase representation sequence from the embedding layer by using CNN, which provides effective support for subsequent coding tasks. In order to improve the quality of context encoding and preserve semantic information, we use BiLSTM to capture both local features of phrases as well as global and temporal sentence semantics. Besides, we add an attention mechanism to model interaction relationships between aspect words and sentences to focus on those keywords of targets to learn more effective context representation. We evaluate the proposed model on three datasets: Restaurant, Laptop, and Twitter. Extensive experiments show that the effectivess of FEA-NN.

INDEX TERMS Aspect-based sentiment analysis, BiLSTM, CNN, attention mechanism.

I. INTRODUCTION

As a significant subtask in the field of natural language processing (NLP), sentiment analysis has attracted more and more researchers' attention in recent years [1]. It is also widely used in data mining, question and answer, recommendation systems, and information retrieval. The rise of this field has benefited from the rapid development of social media and e-commerce. Especially in the field of e-commerce, a massive number of reviews reflecting consumers' sentiments are published to different aspects of products and services, such as quality, price and so on. The purpose of Sentiment analysis is to capture opinions and emotions in user reviews that can be used to predict customs demands and to provide support in company decision-making. It is necessary to research to help users to concentrate on vital information and reduce the interference of spam [2]. According to the different granularity of processing texts, traditional sentiment analysis could be roughly divided into two levels: sentence-level and document-level. However, it is difficult for traditional sentiment analysis to find the sentiment for the aspects of the entity. For the product reviews, the product itself is usually the entity, and all things related to the product (e.g., price, quality, etc.) are aspects of it [3].

Aspect-based sentiment analysis is a fine-grained sentiment analysis task. Its purpose is to identify the sentiment polarity (e.g., positive, negative, and neutral) of one sentence towards an aspect word (also called target word). For example, the sentence of "the computer screen looks great, but the battery life is too short," the aspects of "screen" and "battery" are positive and negative, respectively. The complete ABSA (Aspect-based Sentiment Analysis) task includes two subtasks: aspect detection and sentiment classification. In this article, we use aspect identification to define aspect

The associate editor coordinating the review of this manuscript and approving it for publication was Wenge Rong.

embeddings, because the aspects we use are predefined. ABSA is a fundamental task in natural language processing and catches many researchers' attention [4]. Early researchs on the ABSA task typically use machine learning algorithms and build sentiment classifier in a supervised manner. Although machine learning has achieved much success in aspect-based sentiment analysis tasks, it requires large-scale text preprocessing and complex feature engineering. Representative approaches in the literature include feature-based Support Vector Machine (SVM) [5], [6].

Recently, the neural network has made a breakthrough in the field of text sentiment analysis with its superior performance. Bahdanau et al. [7] constructed a language model by using recurrent neural network and expressed word vectors in low-dimensional space, which could better measure the correlation between words and words. Convolutional Neural Networks model (CNN) has been successfully applied to text classification, which has multiple convolution kernels of different sizes to extract local features [8], [9]. Long Short-Term Memory (LSTM) solves the temporal relationship between words in sentences and the interactivity between words, which effectively improves the accuracy of sentiment analysis [10]–[12]. Grave et al. [13] proposed the use of recurrent convolutional neural networks for text classification, which uses a bidirectional cyclic structure to model text. Tai et al. [14] introduces tree structure on the basis of LSTM network to improve the semantic expression of sentences. Liu et al. [15] proposed an AC-BiLSTM model for text classification, and this model builds a network model by linking CNN and LSTM and enhanced the model feature learning ability to improve the model classification effect. Although these efforts have yielded good results, these models could not distinguish the different contributions of each word to the entire sentence, which limited their performance.

In order to distinguish the different contributions of different words in the sentence to the classification, the attention model is also widely introduced into the neural network model. Yin et al. [16] added attention mechanism based on the CNN model, which had better results than the model without attention mechanism. Researchers combined the LSTM network and attention mechanism to conduct sentiment classification based on specific aspects and achieved better results than the previous model [4], [17]. However, different from the general sentiment analysis task, if a sentence contains multiple aspects and each of which has related words to embellish it, and other target words would become noise, it is difficult to distinguish this sentiment polarity by different aspects. So, we need to consider the relevance among aspect words and context words when we designing models. At the same time, it proves the validity of the attention module in learning the relevance between aspect words with context words [18]–[20].

Despite these advances in the attention mechanism, there are still some problems that have not been resolved. When an aspect word contains multiple words, most existing researches only compute the average attention vector of the aspect word, this oversimplified approach introduces noise into the attention module, reducing classification accuracy. For example, "bottle of milk," the average operation would introduce noise word "of." Besides, the instance in the aspect-target words has different effects on classification results. "Milk" is more important than "Bottle." If we can use accurate aspect-target information to calculate aspect and context interaction attention scores, the quality of the classification will high. Existing work does not solve this problem well. IAN uses an interactive approach to learn the attention relationship between sentence and aspect target, and it can extract features related to the aspect target through the attention mechanism. However, the pooling operation ignores the interaction between the word pairs between the sentence and the aspect target. Inspired by IAN, we believe that obtaining aspect and sentence interaction representations are useful. Therefore, we propose a Feature Enhanced Attention Mechanism (FEA) to extract interactive information from the target representation and sentence representation generated from CNN-BiLSTM. FEA not only generates common concern from aspect to sentence but also from sentence to aspect automatically. This is inspired by the observation that only a few words in a sentence contribute to one aspect of sentiment [21]. For example, there are two aspects "food" and "service" in the sentence "the food was delicious, but the service was supercilious." Based on our knowledge, we know that the negative word "supercilious" is more likely to describe "service" than the "food." Similarly, for an aspect phrase, we also need to concentrate on the most critical part.

Our work has the following contributions:

1: The high-quality word embedding model is an essential basis for sentiment analysis. We chose a new method, the improved word vector (IWV) [22], which improves the accuracy of pre-trained word embedding in sentiment analysis.

2: We propose a novel Feature Enhanced Attention model that could learn the interaction between the aspect and the sentence, distinguish the contribution of different aspect words to sentences, reduce the effect of interference words that are not related to the aspect word, and make full use of the keywords associated with the target to learn semantic representation.

3: To confirm the effectiveness of feature enhanced module, we propose the Feature Enhanced Attention CNN-BiLSTM (FEA-NN) to extract local information and long-distance dependent information. It can learn the interaction among context and aspect, and obtain more vital sentiment features through FEA.

We evaluate the proposed model on three datasets: Restaurant and Laptop reviews from SemEval 2014 [23] and Twitter dataset [24]. The experimental results verify the effectiveness of our model in solving the above problems.

The organizational structure of the article is as follows. Section II introduces the related work. The proposed approach is presented in detail in Section III. We evaluate the results of the experiment in section IV. Finally, Section V provides some conclusions and possible ways for future research.

II. RELATED WORK

Since the mission is presented, aspect-based sentiment analysis (ABSA) has attracted many researchers in the field of machine learning and natural language processing. Most of the research works are concentrated in two directions, one is the neural network, and the other is the attention mechanism.

A. ABSA BASED ON NEURAL NETWORK

ABSA aims at identifying sentiment polarity of a sentence expressed towards a target, which is one given aspect of a specific entity. Early works on the ABSA task relied on manual labeling lexicons and syntactic features engineering, then use the machine learning for classification, such as SVM [5], [6]. Although these methods have achieved better results, feature engineering requires a lot of manual labeling, which is time-consuming and laborious. Due to the success of deep learning models in the field of natural language processing (NLP), recent studies use deep neural networks to generate sentence embedding and then feed the vector representation of sentences to classifiers as low-dimensional feature vectors. Dong et al. [25] first introduced Recurrent Neural Networks (RNNs) into this field, they proposed an adaptive RNN, which can adaptively propagate the sentiment of context words to the aspect words. The results shown that the sentence representation of RNN was effective [26]. But they may fail due to common grammar analysis errors in practice [25], [27]. In contrast, in processing many (language) sequence learning tasks, the recurrent neural networks (RNNs) have been proven to be effective, most of the advanced solutions are based on RNNs [28], [27].

In deep learning, LSTM is mainly used for the processing of sequence data. In recent years, the application range of LSTM has expanded rapidly. In order to further improve the performance of LSTM to meet the requirements of various tasks, and to process variable-length sequence information, many researchers have proposed many methods to improve LSTM. Recently, LSTM and its variants have been applied to various tasks and achieved satisfactory results. Tang et al. [28] proposed a target-dependent long-short-term memory network (TD-LSTM) which learns representation directly from the left and right context of a given aspect by using two LSTM networks, respectively. Zhang et al. [27] used gated neural network structures to simulate the grammar and semantics of tweets and the interaction between context and aspect. IAN uses an interactive approach to learn the attention relationship between sentence and aspect target, and it can extract features related to the aspect target through the attention mechanism. The combination of LSTM and other network structures is also a popular research direction.

Researchers have combined LSTM with other network structures as an essential direction of exploration. Kolawole [29] used CNN-LSTM model to solve the challenging problems in sentiment analysis. This method is straightforward and can be well executed without excessive parameter optimization. Wang et al. [30] proposed a regional CNN-LSTM model consisting of two parts: regional CNN and LSTM, used to predict the value of text valence-arousal (VA) rating. The regional CNN used a single sentence as a region and divided the input sentence into several regions so that significant sentiment information in each region can be obtained and weighted according to their contribution to the VA prediction. Then, the LSTM is used to sequentially integrate such area information across regions to perform VA prediction. By combining the regional CNN and LSTM can fully consider the local (regional) information within the sentence and the long-distance dependency information across the sentence in the prediction process. Zhou et al. [31] proposed a C-LSTM model that combined the advantages of CNN and LSTM, where CNN to extract a sequence of higher-level phrase representations, and were fed into a long short-term memory neural network (LSTM) to obtain the sentence representation. C-LSTM was able to capture both local features of phrases as well as global and temporal sentence semantics. The above model demonstrates the validity of CNN combined with LSTM in extracting local information and long-distance dependent information. Besides, with the widespread use of attention mechanisms in the field of image processing, researchers have attempted to introduce it into the field of natural language processing.

B. ABSA BASED ON ATTENTION MODEL

Since Bahdanau et al. [7] have successfully applied attention mechanisms to machine translation tasks, many research efforts have used attention mechanisms to extract the relationship between targets and contexts. Recently, many attempts have been made. Tang et al. [32] proposed a deep memory network with multiple computational layers each of which is a neural attention model on external memory to extract the significance of each context word in deducing an aspect of sentiment polarity. Wang et al. [33] proposed two attention-based LSTM networks, namely AE-LSTM and ATAE-LSTM neural networks. The AE-LSTM neural network modeled the context through the LSTM neural network, and then combined the hiding state with the aspect embedding to generate the attention vector, and finally obtained the sentiment category of the aspect through the classifier. Based on AE-LSTM, the ATAE-LSTM network further enhanced the effects of aspect embedding. Yang et al. [34] proposed an attention-based BiLSTM to improve sentiment classification accuracy. Tay et al. [20] proposed an AF-LSTM based on the association between context words and targets. Thus, when given a target, it allows the model to focus on the correct context word adaptively. While the use of attention mechanisms improved the performance of ABSA tasks, all of this work only deals with targets by averaging pooling when calculating contextual attention scores. If the aspect involves multiple words, the performance of attention module will be reduced. Ma et al. [35] proposed a hierarchical attention model for ABSA analysis tasks, including target-level attention and

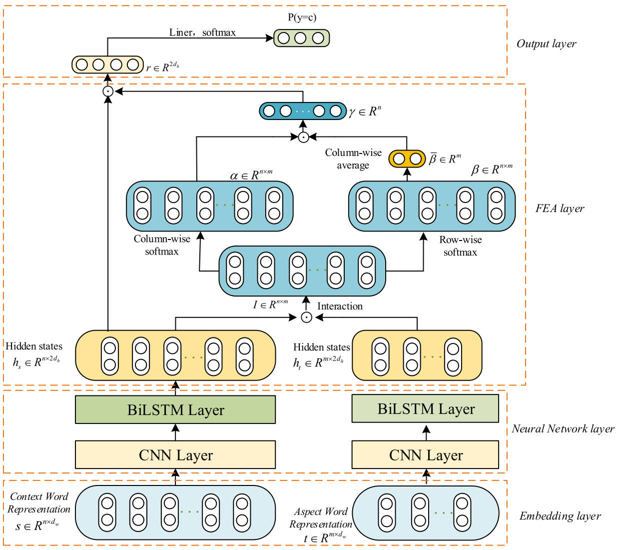

FIGURE 1. **The overall architecture of FEA-NN. The model (FEA-NN) we proposed mainly includes four modules: the first module is word embedding layer which convert input to word vectors. The second module is neural network layer, it can obtain the hidden representation from word embedding layer. The third module is Feature enhanced attention layer, which obtain more vital sentiment features. The last module is output layer, it is responsible for output polarity judgment results.**

sentence-level attention. Nevertheless, the attention at the aspect-level is a network of self-attention, with nothing but the hidden layer output itself as input. Without the guidance of context information, attention at the aspect-level is challenging to learn. Ma et al. [4] proposed an interactive attentional neural network that takes into account both the aspect and the contextual content, using two attention networks to detect important information between aspects and contextual content interactively, but it is too simple to calculate the interaction between aspect and context by means of the average vector, and the aspect information is not effectively utilized. Similar to IAN, we believe that contextual information will help to learn attention to the aspect-level.

III. METHOD

In this section, we first define the research task and introduce the FEA-CNN-BiLSTM, and then describe the model architecture we proposed, which is shown in Figure 1.

A. TASK DEFINITION

In this ABSA task, we suppose that the input sentence is $x = [x_1, x_{2,} \ldots, x_n]$, $t = [t_1, t_2, \ldots, t_m]$ is the given target-aspect words, y, $y_p \in$ *{Positive, Negative, Neutral}* denote real-sentiment and the predicted sentiment, respectively. The aspect-target may be a single word or including multiple words. Our goal is to predict the sentiment polarity of sentence x based on the aspect-target word t. For example, the sentiment polarity of the sentence "The food was great, but the service was dreadful." towards "food" is positive, while the polarity towards "service" is negative.

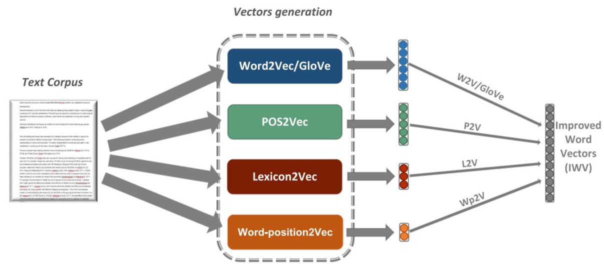

FIGURE 2. **The main architecture of the IWV [22].**

B. OVERVIEW

We propose a new attention mechanism model to solve the problems in the ABSA task. Figure 1 illustrates an overview of Feature Enhanced Attention CNN-BiLSTM (FEA-NN), which includes word embedding layer, CNN layer, BiLSTM layer, Feature Enhanced Attention layer, and sentiment classifier layer. The word embedding layer is used to encode the context and aspect into matrixes. We used CNN to extract local information, and then the BiLSTM was used to learn the long-dependent information and the hidden states of the sentence and the target. Given the hidden semantic representations of the text and the aspect-target generated by BiLSTM, we use the feature enhanced attention layer calculate the attention weights of context and target. Finally, the hidden state of the BiLSTM output was combined with the attention weight, and the softmax function was used to predict the probability of different sentiments.

C. WORD EMBEDDING

Choosing a suitable pre-trained word vector model has an important impact on improving the accuracy of sentiment analysis, this has been proven in the work of Syeda Rida-e-Fatima et al [36]. Existing context-based word embedding, such as Word2vec and GloVe, typically does not capture enough sentiment information, which may result in words with similar vector representations having opposite sentiment polarity (eg, good and bad), thereby reducing sentiment analysis performance [37]. In this paper, we chosed a new word embedding model, the Improved Word Vector (IWV) and all the baseline models in this paper used the IWV, which improved the accuracy of pre-training word embeddings in sentiment analysis. This method is based on the Part of Speech (POS) tagging technology, dictionary-based methods, word location algorithm, and Word2Vec/GloVe methods. In accordance with the method in Seyed's paper, we used Google's pre-trained word2vec and Stanford's pre-trained glove Vector to select the six dictionaries mentioned in the paper, introduced Part-of-speech (POS) tagging and Word-position2Vec. Then get the word vector model used in this paper. The main architecture of the IWV shown in Figure 2 [22].

Given a sentence $x = [x_1, x_2 \ldots x_n]$ with length n and an aspect $t = [t_1, t_2 \ldots t_m]$ with length m, we first map each word into a low-dimensional real-value vector, and each word is converted into a 300-dimensional word vector. Finally, the output of the embedding layer (n depends on the length of the longest input sentence) is a two-dimensional matrix of n*300, and each sentence forms a two-dimensional matrix $M = [w_1, w_2, \ldots, w_n]$ of $n*k$ $(k = 300)$. Where, $w_i = [w_{i1}, w_{i2}, \ldots, w_{ik}]$ is the word vector of the word x_i. Similarly, we can get the embedding matrix A of aspect words.

D. CNN

CNN is a feed-forward neural network. The basic structure of a convolutional neural network includes convolution layer, pooling layer, and fully connected layer.

1) CONVOLUTIONAL LAYER

The purpose of the convolutional layer is to extract the semantic features of the sentence. Each convolution kernel corresponds to extract a certain part of the feature. The number of convolution kernels was set to 150. Convolution operation for each sentence matrix M of the embedded layer output.

$$Z = f(MW + b) \tag{1}$$

where Z represents the extracted feature matrix after convolution operation; the weight matrix W and the bias vector b are the learning parameters of the network.

For the convenience of calculation, it is necessary to make a nonlinear mapping for the convolution result of each convolution kernel.

$$f = relu = max\,(0, x) \tag{2}$$

where relu function is one of the activation functions commonly used in neural network models.

In order to extract features more comprehensively, this paper uses convolution windows of size 2 and 3 to extract the binary and ternary features of the sentence.

2) K-Max POOLING LAYER

After the convolution operation, the extracted features were transferred to the pooling layer. The pooling layer further aggregate these features to simplify their expression. In this paper, we use K-Max pooling operation, which chooses Top-K maxima of each filter to represent the semantic information represented by the filter. The K-value expression is

$$K = [\frac{l - f_s + 1}{2}] \tag{3}$$

where l is the length of the sentence vector, f_s is the convolution window size.

After the pooling operation, the feature vector extracte by each convolution kernel is significantly reduce, and the semantic information of the core of the sentence was retained since the number of convolution kernels was set to 150, the sentence representation matrix generate after pooling is $W \in R^{k \times 150}$.

CNN convolution layer and pooling layer extract local features of short text sentences by convolution operation and pooling operation respectively and get generalized binary and ternary eigenvectors. After the fusion layer, the two kinds of eigenvectors are joined together as the input matrix of LSTM model.

E. BILSTM LAYER

After convolution operation, we use BiLSTM to extract the hidden semantics of words in the sentence and the targe, and it can also obtain long-dependent sequence information of sentences. The core of it is to use memory cells to remember long-term historical information and manage it with a door mechanism. The door structure does not provide information, but it is used to limit the amount of information. In fact, adding gate control mechanism is a multi-level feature selection method. The expressions of gates and memory cells in the gate mechanism are as follows:

$$f_i = \sigma(W_f[x_i, h_{i-1}] + b_f) \tag{4}$$
$$I_i = \sigma(W_I[x_i, h_{i-1}] + b_I) \tag{5}$$
$$\tilde{C}_i = tanh(W_C[x_i, h_{i-1}] + b_C) \tag{6}$$
$$C_i = f_i * C_{i-1} + I_i * \tilde{C}_i \tag{7}$$
$$o_i = \sigma(W_o[x_i, h_{i-1}] + b_o) \tag{8}$$
$$h_i = o_i * tanh(C_i) \tag{9}$$

where f_i, I_i and o_i are the forget gate, input gate, and output gate, respectively. W_f, W_I, W_o, W_c, b_f, b_I, b_c, and b_o are the weight matrix and bias for each gate. C_i is the cell state, and h_i is the hidden output. A single LSTM usually encodes sentence from just one direction. However, two LSTMs can be used as a bidirectional encoder, called bidirectional LSTM (BiLSTM). For a sentence $x = [x_1, x_{2 \ldots} x_n]$, bidirectional LSTM produces a sequence of hidden states $\overrightarrow{h_s} \in R^{n \times d_h}$ from CNN feature representation vectors, we generate another state $\overrightarrow{h_s} \in R^{n \times d_h}$ by another backward LSTM. In the BiLSTM network, the final output hidden states $h_s \in R^{n \times 2d_h}$ are generated by concatenating $\overrightarrow{h_s}$ and $\overleftarrow{h_s}$. We compute the hidden semantic states h_t for the aspect t in the same way.

$$\overrightarrow{h_s} = LSTM([v_1; v_2; \ldots; v_n]) \tag{10}$$
$$\overleftarrow{h_s} = LSTM([v_1; v_2; \ldots; v_n]) \tag{11}$$
$$h_s = [\overrightarrow{h_s}, \overleftarrow{h_s}] \tag{12}$$

Similarly, we can get the hidden state ht of aspect words.

F. FEA MECHANISM

For improving the accuracy of sentiment analysis, both the coattention model [2] and IAN [4] mentioned that it was significant to obtain the connection between aspects and context. Besides, the contribution of a sentence to the sentiment polarity of different aspects of a sentence should be different [32]. Inspired by this, we use a feature enhanced attention mechanism to track information that is most related to the sentiment polarity of a given aspect in the context.

1) CONTEXT-TARGET INTERACT ATTENTION

Combined with the hidden representation of the context and aspect words generated by BiLSTM, we calculate the attention weights of the sentences through an FEA module. Given the target representation $h_t \in R^{m\times 2d_h}$ and the sentence representation $h_s \in R^{n\times 2d_h}$, we first compute a pairwise interaction matrix:

$$I = h_s \cdot h_t^T \tag{13}$$

where the value of each vector represents the association between the sentence and the aspect-target word. Through column-wise softmax and row-wise softmax, we get the attention matrix of the target to the sentence α and the sentence to the aspect-target β:

$$\alpha_{ij} = \frac{exp(I_{ij})}{\sum_i exp(I_{ij})} \tag{14}$$

$$\beta_{ij} = \frac{exp(I_{ij})}{\sum_j exp(I_{ij})} \tag{15}$$

After column-wise averaging β, we get aspect-level attention $\overline{\beta} \in R^m$ that represents an important part of the aspect-target.

$$\overline{\beta}_j = \frac{1}{n}\sum_i \beta_{ij} \tag{16}$$

The final sentence-level attention matrix $\gamma \in R^n$ calculates each aspect-to-sentence attention by weighting. By explicitly considering the contribution of each word to the aspect, we can get the weight of each word in the sentence.

$$\gamma = \alpha \cdot \overline{\beta}^T \tag{17}$$

2) LOCATION-ENHANCED ATTENTION

Although our method works well on calculating aspect-level and sentence-level attention, there still exists one drawback, our method calculates the aspect-level attention weights with an average operation, which introduce interference information to reduce classification accuracy.

In order to solve this problem, we try to introduce position information in the attention calculation procedure with the inspiration that the closer to the aspect-target word is, the more significant the contribution to the prediction. Besides, adding position information to the sentence-level of attention can make the model more dependent on the context words around the aspect-target word to calculate the critical sentiment features of the aspect.

This method is inspired by the works of Liu et al. [38]. Firstly, we defined D_i to measure the relative distance between the current word and the aspect word. For instance, "the food was great, but the service was terrible." For the aspect "food," the relative distance of "great" to "food" was 2 and the distance of "terrible" was 8.

We also define E_i to measure the location significance of words relative to aspect:

$$E_i = 1 - \frac{|n - D_i|}{n} \tag{18}$$

where n is the length of the sentence.

E_i was used to initialize the context representation before incorporating the location information into the attention mechanism.

We first normalize E_i:

$$P_i = \frac{E_i}{\sum_{i=1}^{n} E_i} \tag{19}$$

Then we get the context representation through the weighting function:

$$h_s^{normal} = \sum_{i=1}^{n} p_i \cdot h_{si} \tag{20}$$

where h_{si} is the hidden representation of the context word x_i encoded by BiLSTM.

We add location information to sentence-level attention, using location information to modify attention weights to concentrate on words around the aspect:

$$R_i = E_i \cdot \gamma_i \tag{21}$$

After standardization, we can get the new sentence-level attention:

$$\gamma_i = \frac{Ri}{\sum_{i=1}^{n} Ri} \tag{22}$$

G. OUTPUT AND TRAINING

The final sentence representation is to use the sentence attention of FEA module to weigh the hidden semantic state of the sentence.

$$r = h_s^T \cdot \gamma \tag{23}$$

We represent this sentence as the final classification feature and feed it into a linear layer, projecting r into the space of the target class C.

$$x = W_l \cdot r + b_l \tag{24}$$

where W_l and b_l are weight matrix and bias. After the linear layer, we use the softmax layer to calculate the probability that the sentence S has a sentiment polarity of $c \in C$ for the aspect a:

$$P(y = c) = \frac{exp(x_c)}{\sum_{i\in C} exp(x_i)} \tag{25}$$

The ultimate predictive emotional polarity of one aspect of the goal is only the label with the highest probability. We minimize the cross-entropy loss by L_2 regularization to train our model

$$loss = -\sum_i \sum_{c\in C} I(y_i = c) \cdot log(P(y_i = c)) + \lambda||\theta||^2 \tag{26}$$

$I(\cdot)$ is the indicator function, λ is the L_2 regularization parameter, and θ is the weight matrix of a set of LSTM networks and linear layers. We further apply dropout to avoid overfitting, where we randomly remove some of the input from the BiLSTM unit.

Using the Adam [39] update rule, the mini-batch stochastic gradient descent is used to minimize the loss function of the weight matrix and the bias term in the model.

TABLE 1. Statistics of restaurant, laptop and twitter datasets.

Dataset	Positive		Negative		Natural	
	Train	Test	Train	Test	Train	Test
Restaurant	2164	728	898	196	637	196
Laptop	987	340	866	128	460	169
Tweet	1411	173	1411	173	2826	346

IV. EXPERIMENT

In this section, we describe the experiment setting and demonstrate the validity of the proposed model by setting up a series of comparative experiments.

A. DATASET

We experiment the proposed model on three public datasets, restaurant and laptop reviews from SemEval 2014 [2] have been widely used in previous studies, the twitter dataset collecte by Dong et al. [11]. The statistics of these datasets are presente in Table 1. For the sake of data balance, we removed some examples of ''conflict'' sentiment polarity, for example, ''Certainly not the best sushi in New York, however, it is always fresh'' [2]. The evaluation metrics are classification accuracy and F1 score.

B. EXPERIMENT SETTINGS

In the experiment, we first randomly select 20% of the training data as the verification set to adjust the hyperparameters. All weights are randomly initialize by a uniform distribution of U (-10^{-4}, 10^{-4}) with all bias terms set to zero. The L2 regularization rate is set to 10^{-4} and, the dropout is set to 0.2 [15]. The word embedding is initialized with a 300-dimensional pre-trained word vector and fixed during training. For words outside the vocabulary, we use random distribution U (-0.25, 0.25) to initialize randomly, as done in [9]. The dimension of the LSTM hidden state is set to 150. The initial learning rate of Adam optimizer is 0.01 [39]. If the training loss does not decrease after every three stages, our learning rate will be reduced by half. The batch size is set to 25.

C. MODEL COMPARISON

Our methods and previous ABSA methods are trained and evaluated on these four data sets, respectively. We use accuracy and macro-F1 to evaluate the model. To further validate the performance of the model, we compare it with several baseline approaches. In order to exclude the influence of word vectors on experiments, all the baseline models in this paper use the same word vector model–the pre-trained IWV.

SVM [5]:the classic SVM model using a series of manual features has the best results in SemEval-2014 Task 4.

CNN: based on the convolutional neural network model proposed by Kim [9].

LSTM [28]: uses single LSTM to model the sentence, and the last hidden state is used as the sentence representation for the final classification.

TD-LSTM: TD-LSTM [28] uses BiLSTM to encode the sequential structure of the sentence and represent the given target using a vector that averages the hidden output of the target instance.

ATAE-LSTM: an attention-based LSTM network proposed by Wang et al. [33]. Used the pre-trained word2vec word vector as input to train the model, combined aspect information and attention mechanisms. The model focuses on specific aspects during training and effectively identifies sentiment polarity.

BILSTM-ATT-G [19]: It models left and right contexts using two attention-based LSTMs and introduces gates to measure the importance of left context, right context, and the entire sentence for the prediction.

RAM [18]: RAM is a multilayer architecture where each layer consists of attention-based aggregation of word features and a GRU cell to learn the sentence representation.

IAN [4]: an LSTM-based network which uses attention module to obtain significant information to generate representations of aspect and context separately by the interactive way.

AF-LSTM [20]: learn to participate according to the relationship between sentences and aspects which allows our model to concentrate on the correct words given an aspect term adaptively.

TNET [40]: use CNN as a feature extractor for classification problems and rely on context-preserving and location-correlation mechanisms to maintain the advantages of LSTM.

IAN-NN: inspired by the TNET model, we added a CNN layer in IAN to extract significant features.

FEA-NN*: the model proposed in this paper without adding location attention.

D. MAIN RESULT

It can be seen from Table 2 that our model FEA-NN maintains the best performance on all datasets, which verifies the effectiveness of our entire model. Meantime, majority method has the poor performance, and the main reason is that the model cannot obtain target information accurately, especially those models that use the average pooling method, which ignored the target information, such as TD-LSTM, ATAE-LSTM, AF-LSTM, BILSTM-ATT-G, IAN and RAM. TD-LSTM improves performance by approximately 2% compared to LSTM, it is improved from the basis of LSTM, uses a forward LSTM and a backward LSTM to model the left context and right context with an aspect, but it may be incapable of capturing the interactions between aspects and contexts, resulting in poor performance. ATAE-LSTM improve performance by approximately 2$\sim$3% compared with TD-LSTM because it uses attention mechanisms to model the relationship between aspect and context. AF-LSTM adopts circular convolution and circular correlation to model the similarity between aspect and words, the performance of the model has been significantly improved. BILSTM-ATT-G improved performance by approximately 2% compared with AF-LSTM, close to IAN, because it uses two attention-based LSTMs, and gates were introduced to measure the

TABLE 2. Comparison with other models.

Dataset	Restaurant		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
CNN	62.33	57.17	57.14	53.43	63.20	64.12
LSTM	71.62	68.54	63.25	59.89	59.31	55.52
SVM	72.23	67.65	64.34	61.43	62.23	61.72
TD-LSTM	73.56	65.90	66.90	61.73	-	-
ATAE-LSTM	76.82	72.56	67.93	60.81	-	-
AF-LSTM	77.13	72.85	72.32	68.21	66.60	60.82
BILSTM-ATT-G	77.82	72.33	72.61	67.42	67.52	62.47
IAN	78.33	73.12	73.64	67.31	-	-
RAM	78.95	74.21	74.10	68.42	67.94	63.02
TNET	80.79	74.32	75.21	69.73	68.63	64.31
IAN-NN	81.99	75.21	76.21	70.13	70.71	65.85
FEA-NN*	82.21	76.05	76.43	70.62	71.10	67.31
FEA-NN	**83.21**	**77.57**	**78.55**	**72.65**	**73.31**	**68.67**

importance of the left and right contexts and the entire sentence. IAN has better results because it models aspect representations by using attention modules while modeling context, which makes better use of aspect information than AF-LSTM. The performance of RAM is 2% higher than that of IAN, close to TNET, because it uses multiple attention mechanisms to capture distant sentiment features, which is more robust to interference from uncorrelated information. The performance of TNET is 2% higher than that of IAN, instead of using attention mechanism, TNET uses CNN to extract significant features from BiLSTM's context representation and adds a component between the two layers to maintain the original context information, and this effectively avoids the natural defects of the previous attention mechanism.

Among our proposed models, FEA-NN obtains the best classification accuracies on three datasets. The IAN-NN combine the advantages of CNN extraction features in TNET and the interactive attention mechanism in IAN, and achieved better results, this proved the effectiveness of introducing the CNN layer, but the average pooling method in IAN limited the further improvement of the model. FEA-NN adopts FEA without adding location attention mechanism. FEA-NN uses FEA without adding location attention mechanism, which also achieves better results, but the performance improvement is limited compared with IAN-NN because of the lack of location attention mechanism, and the ability to distinguish the importance of context words related to aspect needs to be improved. As expected, FEA-NN can better predict the sentiment polarity of a given aspect in a sentence by considering the attention mechanism of location information between aspect words and context words. The above results show that adding attention mechanism on the basis of considering the interaction between aspect and context words can better distinguish the importance of different words to aspect and help to improve the accuracy of sentiment prediction.

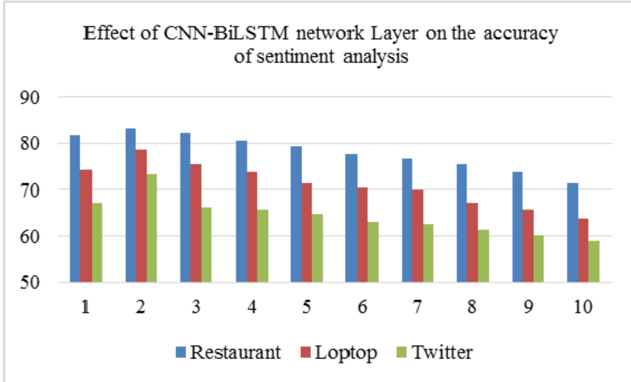

FIGURE 3. Effect of CNN-BiLSTM network Layer on the accuracy of sentiment analysis.

TABLE 3. Comparison of results using different word embedding model.

Dataset	Restaurant	Laptop	Twitter
FEA-NN+fasttext	70.99	65.41	60.41
FEA-NN+glove	73.14	68.93	63.14
FEA-NN+word2gm	74.51	69.38	64.71
FEA-NN+prob-fasttext	76.23	71.34	66.23
FEA-NN+word2Vec	80.13	76.93	69.24
FEA-NN+IWV	83.21	78.55	73.31

E. EFFECT OF CNN-BILSTM NETWORK LAYER ON THE ACCURACY

Choosing a suitable CNN-BiLSTM network layer is important for improving the accuracy of the model. We determined the number of layers through experiments. The more the number of layers, the more accurate the surface is represented, but the more difficult it is to obtain the global optimal value. We experiment with 1 to 10, and selecte 2. Figure 3 shows the details.

F. COMPARISON OF RESULTS USING DIFFERENT WORD EMBEDDING MODEL

For sentiment analysis tasks, it is crucial to choose a high-quality word embedding model. We combine the neural network model proposes in this paper with different word vector models, to ensure the selection of the most suitable vectors for this model so that our model can achieve the best performance. The experimental results are shown in Table 3.

Word2vec: is an open-source tool proposed by Google in 2013 to represent words as real-valued vectors. It is a typical representative of the method of predicting word vector by context information.

Glove: Using statistical methods to model the frequency of occurrence of words and their context words.

Fasttext: When fasttext does word embedding, it thinks that words consist of English letters and words with similar letter structures should have standard features. This method uses Word2vec for reference and adds information of letters in words to assist.

word2gm: Drawing on Word2vec, it is believed that words may have different semantics in different contexts (multiple meanings), one word corresponding to a vector is not enough

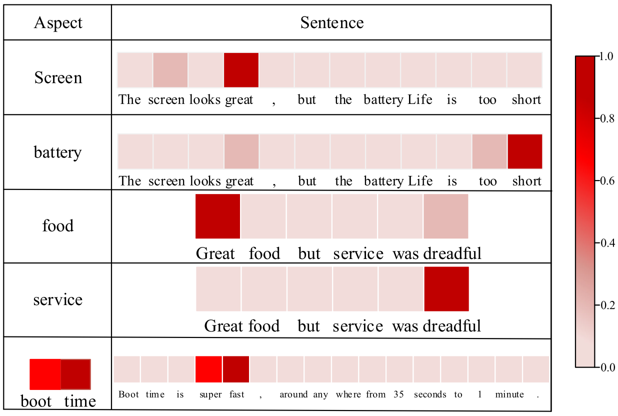

FIGURE 4. **Examples of final attention weights for sentences.**

to reflect such information, consider multiple words embedding, and learn each sub-vector of words with Gaussian mixture model.

Prob-fasttext: Fasttext considers the letter information, but does not consider the word polysemy, it mixes the idea of word2gm and fasttext, each word is represented by two embedding results, one of which is embedding of letters, one is embeeding of its own.

It can be seen from the experimental results that the FEA-NN using the IWV word vector model achieves the best sentiment classification accuracy.

G. CASE STUDY

In Figure 4, we list five examples of test sets that analyze the essential aspects of emotional polarity words, and we see the last sentence note vector in Figure 4. Color depth indicates the importance of a word in a sentence. The darker the color, the more critical it is. In the first two examples, there are two aspects "screen" and "battery." In the sentence "The screen looks great, but the battery life is too short." We can observe that when there are two aspects in a sentence, our model automatically points to the correct sentiment for the word representing each aspect, as in the third and fourth examples. In the last example, the aspect is the phrase "boot time." Boot time is super fast, around anywhere from 35 seconds to 1 minute. The color depth denotes the importance degree of the weight in attention vector γ.

V. CONCLUSION

In this paper, a novel neural network structure for target ABSA task is proposed. The model uses CNN to provide BiLSTM with higher quality text representation and uses FEA attention module to encode the target and complete sentences. Target-level attention learns to focus on the salient emotional part of the target expression and generates a more accurate representation of the target. Sentence-level attention searches for the relevance between the target and the aspect in the whole sentence, location-enhanced attention can be used to distinguish better the importance related to sentiment prediction base on the distance between context words and aspect words. We evaluate the model on three datasets. The experimental results verified the effectiveness of the proposed neural network and shown that the model achieves the most advanced performance.

REFERENCES

[1] L. Zhang and B. Liu, "Sentiment analysis and opinion mining," *Synth. Lectures Hum. Lang. Technol.*, vol. 30, no. 1, p. 167, May 2016.
[2] C. Yang, H. Zhang, B. Jiang, and K. Li, "Aspect-based sentiment analysis with alternating coattention networks," *Inf. Process. Manage.*, vol. 56, no. 3, pp. 463–478, 2019.
[3] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 813–830, Mar. 2016.
[4] 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. (IJCAI)*, 2017, pp. 4068–4074.
[5] S. Kiritchenko, X. Zhu, C. Cherry, and S. Mohammad, "NRC-Canada-2014: Detecting aspects and sentiment in customer reviews," in *Proc. Int. Workshop Semantic Eval.*, 2014, pp. 437–442.
[6] J. Wagner, P. Arora, S. Cortes, U. Barman, D. Bogdanova, J. Foster, and L. Tounsi, "DCU: Aspect-based polarity classification for SemEval task 4," in *Proc. 8th Int. Workshop Semantic Eval. (SEMEVAL)*, 2014, pp. 223–229.
[7] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *Comput. Sci.*, Sep. 2014.
[8] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," 2014, *arXiv:1404.2188*. [Online]. Available: https://arxiv.org/abs/1404.2188
[9] Y. Kim, "Convolutional neural networks for sentence classifcation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1746–1751.
[10] Q. Qian, M. Huang, J. Lei, and X. Zhu, "Linguistically regularized LSTM for sentiment classification," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2017, pp. 1679–1689.
[11] S. Ruder, P. Ghaffari, and J. G. Breslin, "A hierarchical model of reviews for aspect-based sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2016, pp. 999–1005.
[12] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu, "Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling," in *Proc. 26th Int. Conf. Comput. Linguistics (COLING)*, 2016, pp. 3485–3495.
[13] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.
[14] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *Comput. Sci.*, vol. 5, no. 1, p. 36, 2015.
[15] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, Apr. 2019.
[16] Z. Lin, M. Feng, C. N. D. Santos, M. Yu, B. Xiang, B. Zhou, and Y. Bengio, "A structured self-attentive sentence embedding," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–15.
[17] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol. (NAACL)*, 2016, pp. 1480–1489.
[18] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2017, pp. 452–461.
[19] J. Liu and Y. Zhang, "Attention modeling for targeted sentiment," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics (EACL)*, 2017, pp. 572–577.
[20] 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. (AAAI)*, 2018, pp. 5956–5963.
[21] Y. Cui, Z. Chen, S. Wei, S. Wang, T. Liu, and G. Hu, "Attention-over-attention neural networks for reading comprehension," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, 2017, pp. 593–602.

[22] S. M. Rezaeinia, R. Rahmani, A. Ghodsi, and H. Veisi, "Sentiment analysis based on improved pre-trained word embeddings," *Expert Syst. Appl.*, vol. 117, pp. 139–147, Mar. 2019.

[23] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval.*, 2014, pp. 27–35.

[24] X. Li, L. Bing, P. Li, and W. Lam, "A unified model for opinion target extraction and target sentiment prediction," in *Proc. AAAI*, 2019, pp. 1–8.

[25] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "Adaptive recursive neural network for target-dependent Twitter sentiment classification," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, Baltimore, MD, USA, 2014, pp. 49–54.

[26] T. H. Nguyen and K. Shirai, "PhraseRNN: Phrase recursive neural network for aspect-based sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Lisbon, Portugal, 2015, pp. 2509–2514.

[27] M. Zhang, Y. Zhang, and D. Tang, "Gated neural networks for targeted sentiment analysis," in *Proc. 30th AAAI Conf. Artif. Intell.*, Phoenix, AZ, USA, 2016, pp. 3087–3093.

[28] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective LSTMs for target-dependent sentiment classification," in *Proc. 26th Int. Conf. Comput. Linguistics (COLING)*, Osaka, Japan, 2016, pp. 3298–3307.

[29] S. K. Sønderby, C. K. Sønderby, H. Nielsen, and O. Winther, "Convolutional LSTM networks for subcellular localization of proteins," in *Proc. 2nd Int. Conf. Algorithms Comput. Biol.* Mexico City, Mexico: Springer-Verlag, 2015, pp. 68–80.

[30] J. Wang, L.-C. Yu, K. R. Lai, and X. Zhang, "Dimensional sentiment analysis using a regional CNN-LSTM model," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, Berlin, Germany, 2016, pp. 225–230.

[31] C. Zhou, C. Sun, Z. Liu, and F. C. M. Lau, "A C-LSTM neural network for text classification," *Comput. Sci.*, vol. 1, no. 4, pp. 39–44, 2015.

[32] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2016, pp. 214–224.

[33] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proc. EMNLP*, 2016, pp. 606–615.

[34] M. Yang, W. Tu, J. Wang, F. Xu, and X. Chen, "Attention based LSTM for target dependent sentiment classification," in *Proc. AAAI*, 2017, pp. 5013–5014.

[35] Y. Ma, H. Peng, and E. Cambria, "Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM," in *Proc. 32nd AAAI Conf. Artif. Intell. (AAAI)*, 2018, pp. 5876–5883.

[36] S. Rida-E-Fatima, A. Javed, A. Banjar, A. Irtaza, H. Dawood, H. Dawood, and A. Alamri, "A multi-layer dual attention deep learning model with refined word embeddings for aspect-based sentiment analysis," *IEEE Access*, vol. 7, pp. 114795–114807, 2019.

[37] L.-C. Yu, J. Wang, K. R. Lai, and X. Zhang, "Refining word embeddings using intensity scores for sentiment analysis," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 3, pp. 671–681, Mar. 2018.

[38] Q. Liu, H. Zhang, Y. Zeng, Z. Huang, and Z. Wu, "Content attention model for aspect based sentiment analysis," in *Proc. World Wide Web Conf.*, 2018, pp. 1023–1032.

[39] A. Kosiorek, "Attention mechanism in neural networks," *Robot Ind.*, vol. 6, pp. 12–17 2017.

[40] X. Li, L. Bing, W. Lam, and B. Shi, "Transformation networks for target-oriented sentiment classification," 2018, *arXiv:1805.01086*. [Online]. Available: https://arxiv.org/abs/1805.01086

WEI MENG was born in 1995. He is currently pursuing the master's degree with the College of Computer Science and Engineering, Shandong Normal University, China. His research interests include natural language processing, data mining, and sentiment analysis.

YONGQING WEI received the bachelor's degree from Shandong Normal University, in 1983. She is currently a Second-level Professor and a Master's Supervisor, with the Basic Education Department, Shandong Police College, China. She is the middle-aged and young expert with outstanding contribution in Shandong province, the outstanding policeman in Shandong Province, the top-notch scientist in Shandong Public Security, and the advanced individual in Shandong Public Security Science and Technology Work. Her research interests include network information security, information retrieval, natural language processing, and artificial intelligence.

PEIYU LIU received the master's degree from East China Normal University, in 1986. He is currently a second-level Professor and a Doctoral Supervisor with the School of Information Science and Engineering, Shandong Normal University, China. He is the National Outstanding Science and Technology Worker, a middle-aged and young expert with outstanding contribution in Shandong province, and a famous teacher of Shandong province. His research interests include network information security, information retrieval, natural language processing, and artificial intelligence.

ZHENFANG ZHU received the Ph.D. degree from Shandong Normal University, in 2012. He was a Postdoctoral Fellow with Shandong University, from 2012 to 2015. He is currently an Associate Professor and a Master's Supervisor with the School of Information Science and Electrical Engineering, Shandong Jiaotong University, China. His research interests include network information security, natural language processing, and applied linguistics.

HONGXIA YIN was born in 1993. She is currently pursuing the master's degree with the College of Computer Science and Engineering, Shandong Normal University, China. Her research interests include natural language processing, data mining, and textual sentiment analysis.