IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 21, 2019, accepted November 9, 2019, date of publication November 20, 2019, date of current version December 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2954590

CE-HEAT: An Aspect-Level Sentiment Classification Approach With Collaborative Extraction Hierarchical Attention Network

YANG GAO, JIANXUN LIU, PEI LI, AND DONG ZHOU
Key Laboratory of Knowledge Processing and Networked Manufacturing, School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China

Corresponding authors: Jianxun Liu (ljx529@gmail.com) and Pei Li (8992077@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61872139, Grant 61572187, Grant 61876062, and Grant 61702181, in part by the Natural Science Foundation of Hunan Province under Grant 2018JJ3190, in part by the Scientific Research Fund of Hunan Provincial Education Department under Grant 18B199, and in part by the Fund of Hunan Provincial Key Research and Development Program under Grant 2018WK2022.

ABSTRACT Aspect-level sentiment classification is a fine-grained sentiment analysis task. It aims to predict the sentiment of a review in different aspects. Recently many works exploit the hierarchical attention network to capture the aspect-specific sentiment information. However, the prior work only attends to use the aspect terms to capture the aspect-specific sentiment information in the text. It may cause the mismatch of sentiment for the aspect-specific when the aspect words are extracted incorrectly. Since the number of aspect words is much more than that sentiment words, some uncommon aspect terms are more difficult to be extracted than uncommon sentiment terms. To solve this problem, we propose a collaborative extraction hierarchical attention network. It consists of two hierarchical attention units, i.e.. One unit extracts the sentiment features by the sentiment attention layer and uses the sentiment features to capture the specific aspect-related sentiment aspect information by the aspect attention layer. The other extracts the aspect features by the aspect attention layer and uses the aspect features to help capture the aspect-specific sentiment information by the sentiment attention layer. Moreover, we use the SemEval competition data set to verify our approach. The experiment shows that our approach achieves better performance than the methods which only use aspect features to extract sentiment feature for aspect-level sentiment classification.

INDEX TERMS Sentiment analysis, aspect-level sentiment classification, collaborative extraction hierarchical attention network.

I. INTRODUCTION

Sentiment classification of user-generated reviews is a key technique to comprehend individuals' attitudes on a product or service [1]. With the popularity of E-commerce platforms, the number of user-generated reviews has increased dramatically. Meanwhile, these resources have grown up to be an important basis for our investigation and analysis of products or services. For instance, we can collect users' reviews on a restaurant, analyze users' sentiment tendencies towards the restaurant and infer whether this food is successful and whether it is worth eating. A review usually contains an evaluation of multiple aspects of the restaurant, and users have different feelings about different aspects of the restaurant. Therefore, mining and analyzing users' sentiment tendencies towards various aspects of the restaurant can provide us a new chance to understand users' judgments about the restaurant.

The goal of aspect sentiment analysis is to identify the sentiment polarity (negative, neutral, or positive) of a aspect-specific expressed in a comment/review by a reviewer [2]. In the sentence ''The price is unaffordable but the swallow nest is so delicious'', the sentiment expression of the aspect-specific service is negative, and the sentiment expression of the aspect-specific food is positive. By extracting the sentiment preference for the product or service, sellers can understand customers more precisely and comprehensively. It may improve the quality of a product or a service. Using the relationship between sentiment information and aspect information is the key to improve the accuracy of

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

aspect-level sentiment classification. For the aspect-specific "price", the related sentiment feature is "unaffordable". For the aspect-specific "food", the related sentiment feature is "delicious". However, traditional sentiment classification methods are hard to extract correctly from multiple aspects of a given text to the related aspect-specific sentiment feature. In recent years, some people propose a model built on the attention mechanism to extract the related sentiment features according to the aspect-specific words and capture the sentiment information. It achieves quite good results [2], [3].

In previous studies of sentiment analysis, most of the efforts focus on how to extract the sentiment related to a given aspect, while ignoring the relationship between the aspect and sentiment in the given text. Most of the studies attend to use the aspect terms to capture the aspect-specific sentiment information in the text. However, the incorrect aspect feature extraction will result in a sentiment mismatch. This is because the number of aspect words is far more than sentiment words and some uncommon aspect terms are more difficult to be extracted. For example, in "The price is unaffordable but the swallow nest is so delicious", the uncommon words "swallow nest" cannot be extracted under the guidance of the aspect "food", which may result in incorrect polarity prediction.

Nevertheless, it is easier to use the sentiment terms to help capture the specific-sentiment aspect information in the text and make the correct prediction. In fact, for many descriptive words that express sentiment information, they can directly reflect the aspect information of descriptive words. For example, the aspect information of the sentiment word "unaffordable" is "price", and the aspect information of the sentiment word "delicious" is "food". Besides, in a given text, there is a positional relationship between sentiment and aspect information, and the closer their positions are, the more likely it is that sentiment information will be expressed as the sentiment in the given aspect.

We list some existing research for aspect sentiment classification in Table 1, where "Position" denotes the model considering the position information between aspect terms and sentiment terms, "Aspect→Sentiment" and "Sentiment→Aspect" means the model considering using aspect features to help extract sentiment features and using sentiment features to help extract aspect features respectively. As far as we know, we may be the first to use the sentiment feature to guide the extraction of the aspect feature.

Bas on the above ideas, we propose a Collaborative Extraction HiErarchical ATtention network (hereafter CE-HEAT) to solve this issue. It consists of two hierarchical attention units, i.e., the sentiment-guided aspect extraction unit and the aspect-guided sentiment extraction unit. Each unit consists of an aspect attention layer and a sentiment attention layer. The sentiment-guided aspect extraction unit extracts firstly the sentiment feature by the sentiment attention layer and then uses the sentiment feature to capture the aspect-related sentiment aspect information by aspect attention layer. The aspect-guided sentiment extraction unit extracts firstly the

TABLE 1. Comparison of existing research and our research on aspect-level sentiment classification.

Method	Position	Aspect→Sentiment	Sentiment→Aspect
TD-LSTM [4]	×	×	×
GRNN [5]	×	×	×
SCNN [6]	×	×	×
H-LSTM [7]	×	×	×
AT-LSTM [3]	×	√	×
IAN [8]	×	√	×
MAN [9]	√	√	×
HEAT-AS [2]	√	√	×
HEAT-SA	√	√	√
CE-HEAT	√	√	√

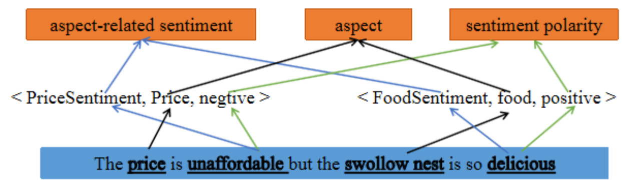


FIGURE 1. An example of aspect-level sentiment classification. The underline words are aspect terms. The bold words are aspect-related sentiment terms and sentiment words.

aspect features by aspect attention layer and then uses the aspect features to capture the aspect-specific sentiment information by the sentiment attention layer. The most important thing is to learn the potential relationship between aspect and sentiment.

The rest of our paper is structured as follows: Section 2 discusses related works. Section 3 presents the problem statement of the model. The detailed description of the model is given in section 4. Section 5 presents extensive experiments to justify the effectiveness of our proposals. Finally this work and its future direction are included in Section 6.

II. RELATED WORK

In early aspect level sentiment analysis methods [10], [11], the extraction of aspect feature and sentiment feature are captured separately and then combines the two features. Usually, the methods such as sequence labelling are used first, which extract sentiment information through traditional sentiment classification algorithms [12], [13]. The aspect-specific sentiment obtained by these methods is only the sentiment expression of the sentence. It is not the sentiment related to the target aspect, which results in a sentiment mismatch [3], [14], [15]. For instance, in Fig. 1, the given aspect "price" is extracted, but the sentiment extracted by the traditional sentiment classification algorithm may be "so delicious", which results in a sentiment mismatch.

In recent years, most models use RNN that achieve better performance on aspect level sentiment classification [4], [5], [7], [8]. For instance, Tang et al. [4] use two LSTM to encode the statements from the left and right sides of the aspect words, and then connect the outputs of the two networks as the sentiment related to the target aspect. The threshold neural network is applied to model the syntactic and semantic information of the statement and the context information of the aspect word [5]. Ma et al. [8] implement an interactive attention network to integrate attention from both target and context. These RNN-based models achieve better classification results because RNN has many advantages, such as the LSTM is better at extracting short-range dependencies among words in sentences. However, an approach only uses the RNN-based model can't extract the potential correlation between relatively distant sentiment words and aspect words in complex statements.

Subsequent studies indicate that the introduction of attention mechanisms can help solve this problem [9], [16]. A complex statement may contain multiple aspect words. Each word in the sentence may be associated with one or more aspects. The phrase in the sentence may also convey sentiment information about a specific aspect. By introducing the attention mechanism, the sentiment feature related to the given aspect can be extracted from a complex statement. For example, Wang et al. [3] proposes the AT-LSTM model based on the attention mechanism (referred to here as AT-LSTM), which extracts the context feature of each word in the text with a one-way LSTM, and then uses the attention network to extract the sentiment features of the text directly under the guidance of the given target aspect vector. However, this model only uses one attention layer to extract sentiment features. While the aspect attention layer extracts unrelated sentiment features, it will result in a sentiment mismatch. Another example, the HiErarchical ATtention network (hereafter HEAT-BiGRU-AS) proposed by Cheng et al. [2] uses a hierarchical attention model to solve the problem by extracting the aspect features and using the aspect features to extract the aspect-specific sentiment information. The experiment achieves better performance. However, due to the large number of aspect terms, the model is not able to identify those uncommon aspects terms well, such as ''chop suey'' and ''swallow nest''. Therefore, the incorrect aspect feature extraction will result in a sentiment mismatch.

Here we propose a CE-HEAT network to solve this problem by using the potential relationship between aspect and sentiment to help each other to extract their features.

III. PROBLEM STATEMENT

The task of aspect-level sentiment classification is to predict the sentiment polarity of the aspect-specific in the sentence. We adopt the most popular definition of aspect and aspect term which gave and used in [2], [6], [17]. In the following, we give some definitions first.

Definition 1 (Aspect-related Sentiment Category): An aspect-related sentiment category is a collection of emotions expressed in the same aspect.

Take the restaurant review as an example. In the sentence ''food is delicious and tasty, but the service is slow and rough'', ''delicious'' and ''tasty'' are attributes of the foodSentiment category. ''Slow'' and ''rough'' are attributes of the serviceSentiment category. FoodSentiment and serviceSentiment are all regarded as a kinds of aspect-related sentiment categories.

Definition 2 (Aspect-related Sentiment Category Term): An aspect-related sentiment category term, also called explicit saspect-related sentiment expression, it is a word or phrase that occurs in the text indicating an aspect-related sentiment category.

An aspect-related sentiment category can express as different aspect-related sentiment category terms in different texts. For example, the foodSentiment category can be mentioned in texts with ''delicious'', ''tasty'' and ''salty''. However, for the serviceSentiment category, it can be mentioned in texts with ''fast'' or ''friendly''.

IV. COLLABORATION EXTRACTION HIERARCHICAL NETWORK

In this section, we present the CE-HEAT network. Firstly, we give an overview of the CE-HEAT network. Then we show the details of each module in the CE-HEAT network. Finally, we will demonstrate the application of this approach on aspect-level sentiment classification.

A. AN OVERVIEW OF THE CE-HEAT NETWORK

Fig. 2 is an illustration of the CE-HEAT network that includes three modules. The three modules are as follows:

(1) **Input Module**. It turns sentence, aspect, and aspect-related sentiment category into distributing embedding vectors. It captures the features of each word in the sentence.

(2) **Collaborative Extraction Module**. It consists of two hierarchical attention units and each unit consists of an aspect attention layer and a sentiment attention layer. The sentiment-guided aspect extraction unit extracts firstly the sentiment features by the sentiment attention layer and then uses the sentiment features to capture the aspect-related sentiment aspect information by aspect attention layer. The aspect-guided sentiment extraction unit extracts firstly the aspect terms by aspect attention layer and then uses the aspect features to capture the aspect-specific sentiment information by the sentiment attention layer.

(3) **Sentiment Classification Module**. It checks the consistency of the sentiment polarity predicted by the two units of the collaboration extraction module and gives the final sentiment polarity prediction result.

B. INPUT MODULE

The input of the CE-HEAT network includes a sentence, aspect-related sentiment category, and a given aspect.

1) ASPECT AND ASPECT-RELATED SENTIMENT CATEGORY EMBEDDING

Using a matrix $\Phi_a \in R^{d_a \times |A|}$ to represent all aspects, where Φ_a is the collection of aspects φ_{a_i}, the embedding vector φ_{a_i} represents aspect a_i, and d_a is the dimension of the embedding vector φ_{a_i}. Matrix $\Phi_s \in R^{d_s \times |S|}$ represents all aspect-related sentiment categories, where Φ_S is the collection of aspect-related sentiment categories φ_{s_i}, the embedding vector φ_{s_i}

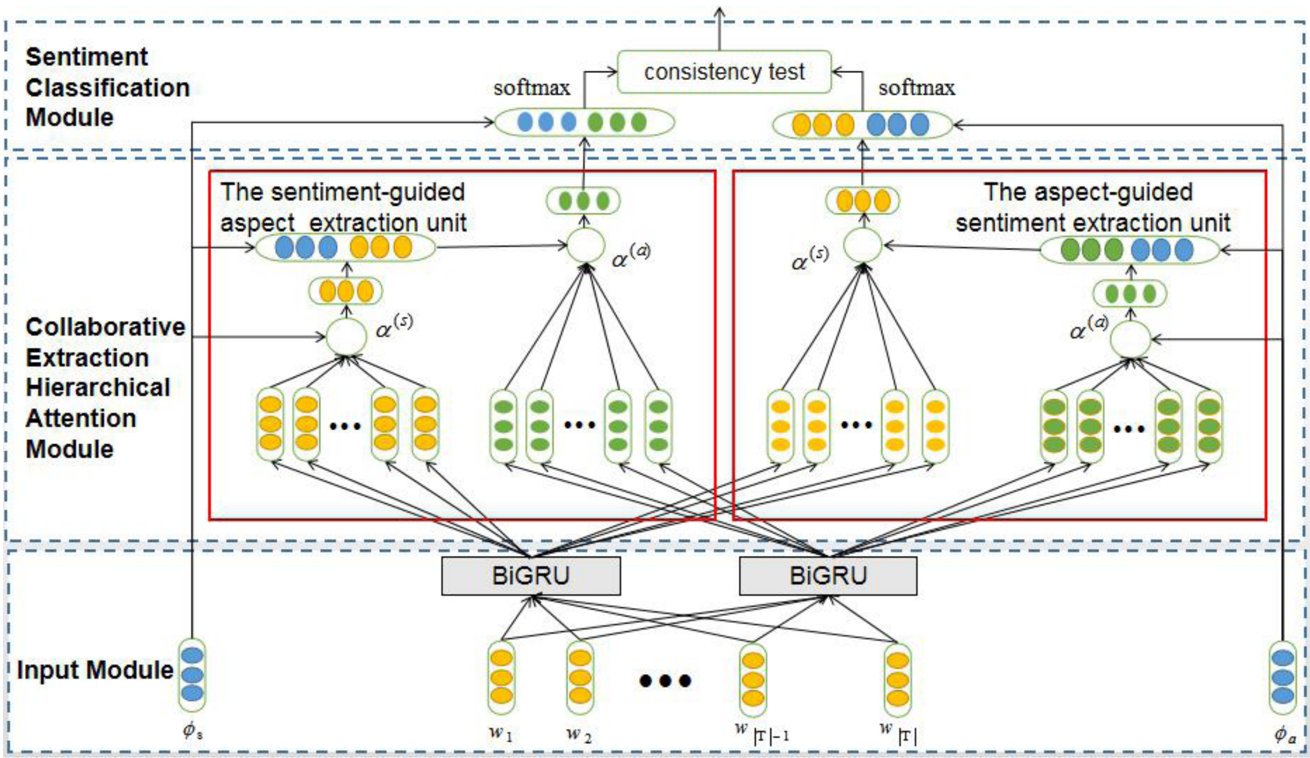


FIGURE 2. An illustration of aspect-level sentiment classification based on Collaborative Extraction HiErarchical ATtention network.

TABLE 2. Meaning of important parameters.

$\|A\|$	The number of aspects in the reviews
$\|S\|$	The number of sentiment categories in the reviews
$\|V\|$	The numbers of the word in the data set
$e_{w_i} \in i^{d_w}$	The word vector related to the word in the text
$\overset{\leftrightarrow}{H}^{(a)}$	The output of BiGRU
$\overset{\leftrightarrow}{h}^{(a)}$	The aspect features of the word w_i in the text
$\overset{\leftrightarrow}{H}^{(s)}$	The output of BiGRU
$\overset{\leftrightarrow}{h}^{(s)}$	The sentiment features of the word w_i in the text
$g_i^{(a)}$	The weight of the ith word in expressing the aspect information in the text
$g_i^{(s)}$	The weight of the ith word in expressing the sentiment information in the text
$v_{(a)}$	The aspect features of the input text
$v_{(s)}$	The sentiment features of the input text

represents aspect-related sentiment category S_i, and d_s is the dimension of embedding vector φ_{s_i}.

2) TEXT EMBEDDING

For input text $T = \{w_1, w_2, \ldots, w_n\}$, the module transforms firstly the text into a matrix of distributed word vectors representations, and then uses a recurrent neural network to extract the contextual feature of each word in the text. The word vectors embedding matrix is $E \in i^{d_x \times |V|}$. It is applied to map each word into a distributed word vectors, where d_w is the word vector dimension. The input text can be noted as $X = [e_1, e_2, \ldots, e_{|T|}]$. In the experiment, we pre-trained word vectors with a large amount of unsupervised data in the same field and then the pre-trained word vectors are used as the initial parameters of the model for sentiment classification training.

3) RECURRENT NEURAL NETWORK SELECTION

After word vectors representation, the module uses a recurrent neural network to extract the contextual features of each word in the sentence for subsequent operations. Both LSTM and GRU have great performance on extracting the contextual features of each word in the sentence. Nevertheless, GRU is better at computing speed and resource occupancy than LSTM. Therefore, we utilize GRU as the recurrent neural network for subsequent tasks. Besides, the contextual features of a word are related to the words that before and after in the sentence, we use BiGRU to extract the context features of each word in the sentence.

C. COLLABORATION EXTRACTION MODULE

The collaboration extraction module consists of two hierarchical attention units and each unit contains an aspect attention layer and a sentiment attention layer. We use the potential relationship between aspect features and sentiment features to help attention layer extract features.

1) LOCATION MASK LAYER

Here is a positional relationship between aspect term and aspect-specific sentiment term. The closer their positions are, the more likely it is that the sentiment information will be expressed as the sentiment in the given aspect. We introduce a location mask layer to calculate the sentiment and aspect attention weights in the hierarchical attention unit more accurately. It represents the location relationship between the aspect-related sentiment category term and aspect term. We use a position matrix $M \in R^{|T\times|T|}$ to represent the position relationship between a word and others in the text. The position relationship between two words is as

$$M_{ij} = 1 - \frac{|i-j|}{|T|}, \tag{1}$$

where $i, j \in \{1, 2, \ldots, |T\}$ and the value M_{ij} is the distance of i, j. The position mask is calculated as

$$m = M\alpha, \tag{2}$$

where α is the vector attention weight.

2) THE SENTIMENT-GUIDED ASPECT EXTRACTION UNIT

It extracts firstly the sentiment features by the sentiment attention layer and then uses the sentiment features to capture the aspect-related sentiment aspect information by aspect attention layer.

Firstly, the sentiment attention layer extracts the aspect-related sentiment features in the text under the guidance of a given aspect-related sentiment category. In most cases, the sentiment features in the text are some unique sentiment words that represent the aspect-related sentiment category, such as "friendly" for serviceSentiment category, "delicious" for foodSentiment category, and so on. The sentiment attention layer takes the aspect-related sentiment category vector φ_s and $\overset{\leftrightarrow(s)}{H} = \left[\overset{\leftrightarrow(s)}{h}_{(1)}, \overset{\leftrightarrow(s)}{h}_{(2)}, \ldots, \overset{\leftrightarrow(s)}{h}_{(T)}\right] \in R^{2\,d\times|T|}$ as the input of the sentiment attention layer. $g_i^{(s)}$ is defined as

$$g_i^{(s)} = \left(u^{(s)}\right)^T \tanh\left(W^{(s)}\left[\varphi_s;\ \overset{\leftrightarrow(s)}{h}_i\right] + b^s\right), \tag{3}$$

where $W^{(s)} \in R^{d\times 3\,d}$ is the weight matrix parameter, $b^{(s)} \in R^d$ is the bias vector parameter, $u^{(s)} \in R^d$ is the weight vector parameter and $\left(u^{(s)}\right)^T$ is the transpose of $u^{(S)}$.

Under the guidance of the given aspect-related sentiment category vector φ_s, the sentiment attention layer uses the attention mechanism to calculate the weight of each word in expressing the aspect-related sentiment category information. The total weight of the sentiment information of each word in the sentence is 1 as shown in Formula (4),

$$\sum_{i=1}^{|T|} \alpha_i^{(s)} = \sum_{i=1}^{|T|} \frac{\exp\left(g_i^{(s)}\right)}{\sum_{j=1}^{|T|} \exp\left(g_j^{(s)}\right)} = 1, \tag{4}$$

where *exp* is the exponential function and $\alpha_i^{(S)}$ represents the weight of the word W_i which contains the target sentiment information.

The sentiment feature extracted from the sentiment attention layer are the weighted sum of the sentiment features $\alpha^{(S)}$ of all words of the input text, $v^{(s)}$ is defined as

$$v^{(s)} = \sum_{i=1}^{|T|} \alpha_i^{(s)} \overset{\leftrightarrow(s)}{h}_i. \tag{5}$$

Aspect attention layer extracts aspect features under the guidance of the position mask layer and sentiment information. It takes sentiment features $v^{(s)}$ shown in Formula (5), aspect-related sentiment category vector φ_s and $\overset{\leftrightarrow(a)}{H} = [\overset{\leftrightarrow(a)}{h}_{(1)}, \overset{\leftrightarrow(a)}{h}_{(2)}, \ldots, \overset{\leftrightarrow(a)}{h}_{(T)}] \in R^{2\,d\times|T|}$ as the input of the aspect attention layer. Under the guidance of aspect-related sentiment category vector φ_s and sentiment features $v^{(s)}$, the aspect attention layer uses the attention mechanism to calculate the weight of each word in expressing the aspect information, $g_i^{(a)}$ is defined as

$$g_i^{(a)} = \left(u^{(a)}\right)^T \tanh\left(W^{(a)}\left[\varphi_s;\ \overset{\leftrightarrow(a)}{h}_i\right] + b^a\right), \tag{6}$$

while calculating aspect features, we can take the positional relationship between aspect-related sentiment category term, aspect term into account and introduce the position mask layer. The weight of each word on the aspect is defined as

$$\alpha_i^{(a)} = \frac{\exp\left(m_i g_i^{(a)}\right)}{\sum_{j=1}^{|T|} \exp\left(m_i g_i^{(a)}\right)}. \tag{7}$$

The aspect features $v^{(a)}$ extracted by the aspect attention layer are the weighted sum on the aspect features $\alpha^{(a)}$ of all words in the input text, $v^{(a)}$ is defined as

$$v^{(a)} = \sum_{i=1}^{|T|} \alpha_i^{(a)} \overset{\leftrightarrow(a)}{h}_i. \tag{8}$$

3) THE ASPECT-GUIDED SENTIMENT EXTRACTION UNIT

It extracts firstly the aspect features by the aspect attention layer and then uses the aspect features to capture the aspect-specific sentiment information by the sentiment attention layer.

Firstly, the aspect attention layer extracts the relevant aspect features in the text under the guidance of a given aspect. In most cases, the aspect features in the text are represented by some unique aspects words. "pizza" is a part of the food aspect, and "waitress" is a part of the service aspect. The aspect attention layer takes the aspect-specific vector $\alpha_i^{(a)}$ and $\overset{\leftrightarrow(a)}{H} = [\overset{\leftrightarrow(a)}{h}_{(1)}, \overset{\leftrightarrow(a)}{h}_{(2)}, \ldots, \overset{\leftrightarrow(a)}{h}_{(T)}] \in R^{2\,d\times|T|}$ as the input of the aspect attention layer. $g_i^{(a)}$ is defined as

$$g_i^{(a)} = \left(u^{(a)}\right)^T \tanh\left(W^{(a)}\left[\varphi_a;\ \overset{\leftrightarrow(a)}{h}_i\right] + b^a\right), \tag{9}$$

where $W^{(a)} \in R^{d\times 3\,d}$ is the weight matrix parameter, $b^{(a)} \in R^d$ is the bias vector parameter, $u^{(a)} \in R^d$ is the weight vector parameter and $\left(u^{(a)}\right)^T$ is the transpose of $u^{(a)}$.

Under the guidance of the given aspect vector φ_a, the aspect attention layer uses the attention mechanism to calculate the weight of each word in expressing the aspect information. The total weight of aspect information of each word in the sentence is also 1, $\alpha_i^{(a)}$ is defined as

$$\alpha_i^{(a)} = \frac{\exp\left(m_i g_i^{(a)}\right)}{\sum_{j=1}^{|T|} \exp\left(m_i g_i^{(a)}\right)}. \tag{10}$$

The aspect features $v^{(a)}$ extracted from the aspect attention layer are the weighted sum of the aspect features $\alpha^{(a)}$ of all words of the input text, $v^{(a)}$ is defined as

$$v^{(a)} = \sum_{i=1}^{|T} \alpha_i^{(a)} \overset{\leftrightarrow(a)}{h}_i . \tag{11}$$

Sentiment attention layer extracts sentiment features under the guidance of the position mask layer and aspect weight information. It takes aspect features $v^{(a)}$ shown in Formula (11), the given aspect vector φ_a and $\overset{\leftrightarrow(s)}{H} = [\overset{\leftrightarrow(s)}{h}_{(1)}, \overset{\leftrightarrow(s)}{h}_{(2)}, \ldots, \overset{\leftrightarrow(s)}{h}_{(T)}] \in R^{2\,d\times|T|}$ as the input of the sentiment attention layer. Under the guidance of the given aspect vector φ_a and aspect features $v^{(a)}$, the sentiment attention layer uses the attention mechanism to calculate the weight of each word in expressing the sentiment information,

$$g_i^{(s)} = \left(u^{(s)}\right)^T \tanh\left(W^{(s)}\left[\varphi_a; v^{(a)}; \overset{\leftrightarrow s}{h}_i\right] + b^s\right), \tag{12}$$

while calculating sentiment features, we take the positional relationship between the aspect term and the sentiment term into account, and introduce the position mask layer. The weight of sentiment attention is defined as

$$\alpha_i^{(s)} = \frac{\exp\left(m_i g_i^{(s)}\right)}{\sum_{j=1}^{|T|} \exp\left(m_i g_i^{(s)}\right)}. \tag{13}$$

The sentiment features $v^{(s)}$ extracted by the sentiment attention layer are the weighted sum on the sentiment features $\alpha^{(s)}$ of all words in the input text. $v^{(s)}$ is defined as

$$v^{(s)} = \sum_{i=1}^{|T} \alpha_i^{(s)} \overset{\leftrightarrow(s)}{h}_i . \tag{14}$$

D. SENTIMENT CLASSIFICATION MODULE

The sentiment classification module includes two subtasks. The first is to predict the sentiment polarity of the two units in the collaboration extraction module. The second is to check the consistency of the sentiment polarity predicted by the two units of the collaboration extraction module and obtain the final sentiment prediction result.

1) PREDICTION OF SENTIMENT POLARITY

In the expression of sentences, different collocation between sentiment words and aspect words may result in different sentiment polarity. For example, "hot" pairs with food the sentiment polarity is positive, but "hot" pairs with the circumstance the sentiment polarity is negative. Therefore, in the aspect-guided aspect extraction unit, we take the target aspect-related sentiment category vector φ_s, $v^{(s)}$ in Formula (14) and the output of aspect attention layer as the input of the sentiment classification module,

TABLE 3. Accuracy on aspect-level sentiment classification. Bin indicates binary prediction. Thr stands for 3-class prediction. Best scores are in bold.

Model	Rest14		Rest15		Rest16		Lapt15	
	Bin.	Thr.	Bin.	Thr.	Bin.	Thr.	Bin.	Thr.
AT-LSTM	89.6	83.1	81.0	77.2	87.6	83.0	86.3	82.1
ATAE-LSTM	89.9	84.0	80.9	77.4	87.2	82.7	85.8	82.3
HEAT-BiGRU-AS	91.3	86.2	83.0	80.1	90.8	87.1	87.9	84.9
HEAT-BiGRU-SA	91.9	85.8	83.6	80.5	91.4	87.3	88.4	85.2
CE-HEAT-BiGRU	**92.4**	**86.3**	**84.0**	**80.8**	**92.0**	**88.4**	**88.9**	**85.6**

$$y_1 = softmax\left(w_{y_1}\left[\varphi_s; v^{(s)}\right] + b^{y_1}\right), \tag{15}$$

$$p_1^i = softmax(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^{d_c} \exp(x_j)}, \tag{16}$$

where p_1^i is the probability of the ith classification result.

In the aspect-guided aspect extraction unit, we take the target aspect vector φ_a, $v^{(a)}$ in Formula (8) and the output of sentiment attention layer as the input of the sentiment classification module,

$$y_2 = softmax\left(w_{y_2}\left[\varphi_a; v^{(a)}\right] + b^{y_2}\right), \tag{17}$$

$$p_2^i = softmax(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^{d_c} \exp(x_j)}. \tag{18}$$

2) THE FINAL PREDICTION RESULT OBTAINS BY THE CONSISTENCY TEST

Here we list some examples of the aspect consistency with the given aspect and the aspect-related sentiment category of the input module. ServiceSentiment category is for the sentiment expression of "service" and foodSentiment is for the sentiment expression of "food". It means they are consistent in aspect category. Therefore, we only need to conduct a consistency test of the sentiment polarity prediction of the two hierarchical attention units. If the two results are the same, we can take it as the final result. However, if the two results are different, we should select the higher probability in p_1^i and p_2^i as the final result.

E. LOSS FUNCTION

The objective function (loss function) is the cross-entropy loss, it defined as

$$L^{(C)} = -\sum_i \sum_j y_i^j \hat{y}_i^j + \lambda \left\|\theta^2\right\|, \tag{19}$$

where y is the distribution of the sentence, $\hat{y}$ is sentiment prediction distribution of the sentence, i is the index of the sentence, j is the index of class, λ is the regularization term, and θ is the parameter set.

TABLE 4. Datas sets with number of sentences(S) and <aspect, sentiment> tuples(T).

Dataset	Train		Test		Total	
	#S	#T	#S	#T	#S	#T
Rest14	3401	3713	800	1025	4201	4738
Rest15	1315	1476	685	730	2000	2206
Rest16	2000	2107	676	703	2676	2810
Lapt15	1739	1901	761	912	2500	2813

V. EXPERIMENTS

A. DATASETS

We use the SemEval competition dataset to evaluate the CE-HEAT network at the sentence level of aspect-level sentiment classification [3]. It includes task 4 of the SemEval 2014 competition, task 12 of SemEval 2015 competition and task 5 of the SemEval 2015 competition. The data set contains user-generated reviews of restaurants and laptops. Each data set contains a target aspect and the aspect-specific sentiment polarity. However, it does not have a aspect-related sentiment category. So we manually mark the aspect-related sentiment category and map it to each aspect-related sentiment category terms. We present the details of the data set in Table 4, where ''Rest'' means the restaurant domain and ''Lapt'' means the laptop domain.

B. EXPERIMENT SETTING

We train word embeddings on two domain-specific corpora. Following [2], [3], we use the Yelp Challenge and the Amazon electronic data set for restaurant domain and laptop domain respectively. Yelp Challenge data set contains 4.1M reviews on restaurants. Amazon electronic data set contains more than 1.5M reviews. We use the Amazon electronic data set for laptop domain. For both corpora, we train 100-dimension word embeddings with the word2vec tool.

We apply the CE-HEAT network in aspect-level sentiment classification. Some parameters initialize by a uniform distribution U(−0.01, 0.01). The dimension of word vectors, aspect, sentiment category embeddings and the size of the hidden layer are 32. We train all models with a batch size of 30 examples, a momentum of 0.9, L2-regularization weight of 0.001 and the initial learning rate are 0.01. In order to reduce the influence of initialization parameters, we train each model three times under different initialization parameters and took the average of the test accuracy as the final result of the model.

C. MODEL FOR COMPARISON

1) AT-LSTM (LSTM MODEL BASED ON ATTENTION NETWORK) [3]

The model extracts the context features of each word in the text with a one-way LSTM, and then uses the attention network to extract the sentiment features of the text directly under the guidance of the given target aspect vector.

2) HEAT-BiGRU-AS (BiGRU BASED ON HIERARCHICAL ATTENTION NETWORK) [2]

The model uses aspect information in the review extracted by the aspect attention layer under the guidance of the target aspect to capture aspect-specific sentiment features.

3) HEAT-BiGRU-SA (BiGRU bASED ON HIERARCHICAL ATTENTION NETWORK)

We propose the model that uses the aspect-related sentiment information in the review extracted by the sentiment attention layer under the guidance of the target sentiment category. Then the aspect attention layer uses it to help extract the aspect features.

4) CE-HEAT-BiGRU (BiGRU BASED ON COLLABORATIVE EXTRACTION HIERARCHICAL ATTENTION NETWORK)

We propose the model which consists of two hierarchical attention units, i.e., it uses the relationship between aspect and sentiment to guide each other to extract their features.

D. EVALUATION AND ANALYSIS

The results of Table 3 indicate the excellent performance of the CE-HEAT network on aspect-level sentiment classification. We observe the following results: 1) the CE-HEAT network achieves better result than the HEAT model; 2) it can learn the potential relationship between the aspect and sentiment more effectively; 3) it can solve the problem of sentiment mismatch caused by the extraction error of the aspect feature extracted by the aspect attention layer. Besides, we will list some advantages of our model and analyze the limitations of the model.

1) CASE STUDIES

In Figure. 3, we list some correct examples of sentiment polarity prediction of the CE-HEAT network. On the left of the Figure. 3 is the input sentence and the shading color depth of words indicates the weight of the sentiment attention layer. On the right of the Figure. 3 is the tuple lists the target aspect-related sentiment category, aspect and the aspect-specific sentiment polarity. Examples (1)-(6) are examples from restaurant reviews. Example (1) and Example (2) are the same sentences. It includes two categories, priceSentiment and foodSentiment. Giving the priceSentiment category and the price, the model can accurately predict the sentiment polarity of the price. Giving the foodSentiment category and the food, the model can accurately predict the sentiment polarity of the food. Example (3) and Example (4) are two sentences that do not have aspect words and aspect-related sentiment category words. Because of their high frequency of occurrence in the data set, the model can accurately predict the sentiment polarity of such sentences.

2) ERROR ANALYSIS

The CE-HEAT network will make incorrect sentiment polarity prediction for sentences that require reasoning.

(1) Price is cheap， food is dreadful.	<PriceSentiment,price,positive>
(2) Price is cheap， food is dreadful.	<FoodSentiment,food,negtive>
(3) I would go back again	<GeneralSentiment,general,positive>
(4) I recommend this to you.	<GeneralSentiment,general,positive>
(5) Let us go for the swollow nest,it's delicious.	<foodSentiment,food,positive>
(6) I will not recommend the chop suey,even it's yummy.	<foodSentiment,food,positive>

FIGURE 3. Examples of collaborative extraction HiErarchical ATtention network. The colored word and the color depth illustrate the weight of word in sentiment attention layer.

An example is "if the price is affordable, it will be a great meal". It expresses negative sentiment about the meal, our approach can easily extract the sentiment information "great" but can not capture the word "if". It is the implicit expression of sentiment. For example, in "Save your money and your time and go somewhere else", it expresses negative feelings towards the restaurant as a whole, but there are no obvious negative words, it is difficult to make a correct prediction.

VI. CONCLUSION AND FUTURE WORK

In this paper, the CE-HEAT network achieves a better performance on aspect-level sentiment classification. it can learn the potential relationship between the aspect and sentiment more effectively. It can solve the problems that the sentiment mismatch caused by the incorrect aspect feature extraction.

The CE-HEAT model achieves better performance on aspect-level sentiment classification. However, there are still some logical relations we can not extract effectively and it can cause incorrect sentiment polarity prediction. In the future, we will consider how to use auxiliary information such as sentence structure, logical relation and semantic features to improve the accuracy of aspect-level sentiment classification.

REFERENCES

[1] M. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews," *J. Comput. Sci.*, vol. 27, pp. 386–393, Jul. 2017.

[2] J. Cheng, S. Zhao, J. Zhang, I. King, X. Zhang, and H. Wang, "Aspect-level sentiment classification with HEAT (HiErarchical ATtention) network," in *Proc. ACM Conf. Inf. Knowl. Manage.*, 2017, pp. 97–106.

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

[4] 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

[5] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1422–1432.

[6] H. Wu, Y. Gu, S. Sun, and X. Gu, "Aspect-based opinion summarization with convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 3157–3163.

[7] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, "Sentic LSTM: A hybrid network for targeted aspect-based sentiment analysis," *Cogn. Comput.*, vol. 10, no. 4, pp. 639–650, 2018.

[8] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," 2017, *arXiv:1709.00893*. [Online]. Available: https://arxiv.org/abs/1709.00893

[9] H. Han, X. Li, S. Zhi, and H. Wang, "Multi-attention network for aspect sentiment analysis," in *Proc. 8th Int. Conf. Softw. Comput. Appl.*, 2019, pp. 22–26.

[10] D. Bollegala, T. Mu, and J. Y. Goulermas, "Cross-domain sentiment classification using sentiment sensitive embeddings," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 2, pp. 398–410, Feb. 2016.

[11] D. Rao and D. Ravichandran, "Semi-supervised polarity lexicon induction," in *Proc. 12th Conf. Eur. Chapter Assoc. Comput. Linguistics*, 2009, pp. 675–682.

[12] J. Khairnar and M. Kinikar, "Machine learning algorithms for opinion mining and sentiment classification," *Int. J. Sci. Res. Publications*, vol. 3, no. 6, pp. 1–6, Jun. 2013.

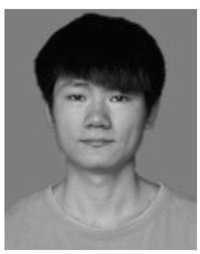
[13] D. Zhang, H. Xu, Z. Su, and Y. Xu, "Chinese comments sentiment classification based on word2vec and SVM^{perf}," *Expert Syst. Appl.*, vol. 42, no. 4, pp. 1857–1863, 2015.

[14] J. Cheng, X. Zhang, P. Li, S. Zhang, Z. Ding, and H. Wang, "Exploring sentiment parsing of microblogging texts for opinion polling on Chinese public figures," *Appl. Intell.*, vol. 45, no. 2, pp. 429–442, 2016.

[15] J. Yu and J. Jiang, "Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 236–246.

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[17] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.



YANG GAO is currently pursuing the master's degree with the School of Computer Science and Engineering, Hunan University of Science and Technology. His research interest includes natural language processing.

JIANXUN LIU received the Ph.D. degree in computer science from Shanghai Jiao Tong University, China, in 2003. He is currently a Professor and also the Dean of the School of Computer Science and Engineering, and the Associate Director of the Key Laboratory of Knowledge Processing and Networked Manufacturing, Hunan University of Science and Technology. His research interests include workflow management, service computing, and cloud computing. He has published more than 100 articles in well-known conferences and journals, such as the IEEE TPDS, IEEE TSC, IEEE TNSM, IR, IM, FGCS, JCSS, ICWS, and SCC. He has served as a program committee member or program chair for the major conferences in his area, such as ICWS, SCC, SCA, CGC, APBPM, and SKG, as well as a Reviewer for TKDE, TSC, and TPDS.

PEI LI received the B.S. and Ph.D. degrees in computer science from the University of Science and Technology of China, Hefei, China, in 2004 and 2009, respectively. He is currently an Associate Professor with the School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan, China. His research interests include social network modeling and information diffusion analysis.

DONG ZHOU is currently a Professor and the Vice Dean of the School of Computer Science and Engineering, Hunan University of Science and Technology, China. His research interests include information retrieval, and natural language processing.