Neurocomputing 438 (2021) 302–311

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Multiple perspective attention based on double BiLSTM for aspect and sentiment pair extract

Yujie Fu [a], Jian Liao [a], Yang Li [c], Suge Wang [a,b,*], Deyu Li [a,b], Xiaoli Li [d]

[a] School of Computer and Information Technology, Shanxi University, China
[b] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, China
[c] School of Finance, Shanxi University of Finance and Economics, China
[d] Institute for Infocomm Research, A*STAR, Singapore

ARTICLE INFO

Article history:
Received 13 September 2020
Revised 4 January 2021
Accepted 20 January 2021
Available online 26 January 2021
Communicated by Zidong Wang

Keywords:
Aspect-based sentiment analysis
Long short-term memory network
Multi-attention mechanism
Aspect category and sentiment pair extraction

ABSTRACT

Aspect category sentiment analysis has attracted increasing attention because of its outstanding performance in mining the fine-grained sentiment expression of users. In recent years, a new aspect category and sentiment pair extraction (ASPE) task has been proposed to simultaneously extract aspect categories and sentiment pairs. Most existing research works are designed in a two step pipeline, that is, they first perform aspect category detection, and subsequently conduct aspect category sentiment analysis. However, the pipeline method can clearly lead to error propagation from previous step. In this work, we propose a new framework for *multiple perspective attention based on double BiLSTM* with a novel joint strategy for ASPE to alleviate the accumulation of errors in the pipeline method. The experimental results on benchmark datasets SemEval and BDCI-2018 demonstrate the effectiveness of the proposed approach in terms of both accuracy and explainability for the ASPE task.

1. Introduction

Aspect-based sentiment analysis (ABSA) involves automatically identifying the sentiment polarity in the product aspect of a review text [17,18]. It has two subtasks: aspect term sentiment analysis (**ATSA**) and aspect category sentiment analysis (**ACSA**) [32]. Aspect term is a word or phrase in a text, and aspect category is an abstract concept that is predefined in a corpus. Aspect categories are typically coarser or higher level than aspect terms, and they do not necessarily occur as terms in a given text.

Typically, the ABSA task is formalized as a three-class sentiment classification problem, with class labels as positive, negative and neutral. The goal is to detect the sentiment for a given aspect category or aspect term. This framework was followed by most of the recent studies in this field [32,28,5,29,8,10,20,12,14,31,6,34]. Since multiple targets may be mentioned in a review text, it is necessary to consider the sentiment of different aspects of each target. Some other researchers have focused on targeted aspect-based sentiment analysis (TABSA) [21,22]. Below, we will elaborate the two subtasks, namely ACSA and ASPE in details and provide an example illustrate their concepts.

For ACSA, the data come from a wide range of sources and can be provided by major review websites. In addition, the annotation of aspect categories is easier to obtain than the annotation of aspect terms. For example, we can easily obtain the annotation of an aspect category from the review tag of a website. Therefore, the combination of aspect category detection and sentiment classification is highly suitable for real-world scenarios. However, in the ACSA task, the aspect category must be given before sentiment classification. This shortcoming limits the application of ACSA to practical scenarios. Therefore, Big Data & Computing Intelligence (BDCI)[1] proposed an aspect category and sentiment pair extraction (ASPE) contest for aspect category detection and sentiment analysis. Most works use the pipeline method to conduct aspect category detection first, followed by sentiment analysis for the corresponding aspect category. Aspect category detection can be regarded as a multilabel classification problem because a review may mention multiple aspect categories. Furthermore, we need to judge the sentiment polarity of each involved aspect category.

The ASPE task was proposed to determine which aspect categories are involved in a review and what is the polarity of the sentiment of each involved aspect categories.

* Corresponding author at: School of Computer and Information Technology, Shanxi University, China.
E-mail address: wsg@sxu.edu.cn (S. Wang).

[1] https://www.datafountain.cn/competitions/310.

https://doi.org/10.1016/j.neucom.2021.01.079
0925-2312/

Fig. 1 displays an example. A review has five aspect categories. In the example, the aspect term of the review is "food", and the corresponding sentiment polarity is "positive". The three aspect categories are "food, ambiance, and restaurant", and their corresponding sentiment polarities are "positive, positive, and positive" respectively. We can then see a certain correspondence between the term and the category, and different users express the term in different ways. The description of ambience and restaurant also appears in the review, but its corresponding aspect term is the default. We often identify "superb, homey, and intimate" as the sentiment descriptors of the aspects "food, ambiance, and restaurant", respectively.

In Fig. 1, we can see clearly the difference between ACSA task and ASPE task. The pipeline method detects the aspect categories that are mentioned in a review first. Then, the aspect categories are used for ACSA to distinguish different sentiment polarities. The goal of ACSA is to extract the corresponding sentiment of a given aspect category. In addition to a review as the input, ACSA needs to provide an annotated aspect category at first prior to sentiment analysis. By contrast, the output of the ASPE task is a pair of aspect category and sentiment. For the ASPE task in **Example 1**, our objective is to directly generate all pairs of aspect category and sentiment, including ("food-positive", "ambiance-positive", "restaurant-positive"), without providing the aspect category "food, ambiance, and restaurant". Clearly, we will focus on ASPE task as it can complete the two steps simultaneously without error propagation.

As mentioned in the Example 1, sentiment description information corresponds to the aspect mentioned in the sentence. In this paper, we propose a new framework called **m**ultiple **p**erspective **a**ttention based on **d**ouble **b**i-directional long short-term memory (MPADB). When a sentence mentions many aspect categories, these categories may not match other sentiment descriptions. To address this problem, we employ bi-directional long short-term memory (BiLSTM) to capture global semantic information [13] and syntactically dependent characteristics. In addition, we also model multiple BiLSTM to capture semantic and aspect category information, so as to alleviate the problem of mutual interference of sentiment information of multiple aspects. Existing methods apply single attention to aspect-related sentiment information, although single attention cannot be modeled effectively in the presence of multiple sentiments. Therefore, this study proposes to use the multi-attention mechanism to expand the representation space and focuses on information from multiple perspectives. It also aims to enhance the ability to capture sentiment information with corresponding aspect categories.

For the ASPE task, we propose a joint strategy that integrates sentiment labels with category labels, and trains our model at the aspect level. On the basis of the original three-class classification problems, we add a label to identify the presence of such aspects. In this way, the task can be transformed into a multilabel

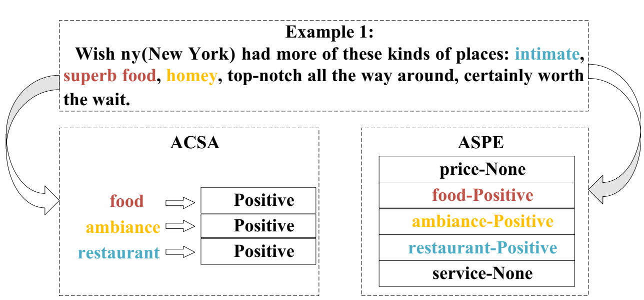

Fig. 1. An example showing the difference between the ACSA task and the ASPE task.

classification problem so that our model can detect the aspect category and distinguish the sentiment polarity while limiting the error accumulation problem caused by the pipeline method. These properties enable our model to make independent decisions for each aspect.

The main contributions of this work are summarized as follows:

1. We propose a new framework called **MPADB**. It avoids wrong matching between various aspects and sentiment descriptions through rich context expressions and the multi-perspective attention mechanism.
2. We propose the joint strategy **MPADB_Joint** to address the shortcomings of the pipeline method that depends on aspect category detection to avoid error accumulation. Moreover, we propose an orthogonal regularization constraint to solve the problem of overlapping attention weights caused by the joint model and enhance the explanatory ability of the model.
3. Our extensive results on SemEval and BDCI 2018 indicate that our approach is effective for the ASPE task.

The remainder of this paper is organized as follows. We first introduce Section 2 in the related work. Section 3 provides the details of our proposed MPADB. Section 4 discusses experiment and analysis. Finally, Section 5 concludes the paper with future directions.

2. Related work

In text sentiment analysis, ABSA is usually divided into ATSA and ACSA. In ABSA research, deep learning has become an important method.

2.1. Aspect term sentiment analysis

For ATSA, most researchers used the Long Short-Term Memory (LSTM) + Attention model as the basic framework and employed the attention or interaction of aspect term to obtain important sentiment information in sentences [28,5,8,10,20,12]. For instance, [29] first applied deep memory network to ATSA by using a computing layer shared by multiple layers of parameters, to make each context word learns the corresponding weight and employs this information for text representation. Other researchers introduced the location information and syntactic information of attribute items in comments to their models [8,10]. In addition, convolutional neural network (CNN) have also been utilized in ATSA tasks [14]. In order to tightly integrate the commonsense knowledge [21,2–4] proposed an extension of LSTM, termed Sentic LSTM, in TABSA tasks.

2.2. Aspect category sentiment analysis

For ACSA, aspect categories are often used as initialization vectors to participate in the attention mechanism. A typical ATAE model based on an LSTM network with an attention mechanism used category embedding to focus on category-specific sentiment information in sentences [31]. At present, the attention mechanism is often used for feature fusion [1] and widely used in ACSA tasks. [6] used aspect attention and sentiment attention to locate different information. To overcome the problem of inadequate semantic information in shallow network learning, [34] proposed a deep memory network with auxiliary memory, constructed two memory modules, and learned the category and sentiment features through their interaction. [32] proposed a model based on CNN and gate mechanism and used it to effectively control the transfer of sentiment information according to the given category information.

2.3. Aspect sentiment pair extraction

Most of the existing work focused on aspect term and sentiment pair extraction [15,16,9]. There is little work related to aspect category and sentiment pair extraction. By adopting a pipeline model, the ASPE task can be broken down into two subtasks, which are aspect category detection [26] and aspect category sentiment classification. Recently, [27] jointly modeled the detection of aspects and the classification of their polarity in an end-to-end trainable neural network. It conducted experiments with LSTM, CNN and word representations on the recent GermEval 2017 dataset. [33] proposed a joint learning framework with CNN, which used CNN to integrate the feature representation of character level and word level. However, these work lacks consideration of model structure, category imbalance, and interpretability of attention.

In practical application scenarios, we need to detect the aspect categories and the corresponding sentiments simultaneously. Therefore, studying the ASPE task is particularly challenging when aspect categories are not given.

3. Multiple perspective attention based on double BiLSTM

For the ASPE task, the different aspects may correspond to different sentiment descriptions. The real cases, involving many aspects and long text mutual interference or confusion of sentiment information between different aspects, are inevitable, where complex semantics need to be encoded. To this end, we design a network with two layers of BiLSTM. The first layer encodes semantic information to model the context dependency of a sentence. The second layer encodes the category information to fuse the aspects with the sentiment information in the context. The attention mechanism is used to locate the aspect information and to relate certain aspect categories to the expression information of the context. In addition, multiple perspective attention is used to expand the representation space and represent the attention information in multiple subspaces to enhance the attention information, thereby reducing the influence of the mutual interference of sentiment information. If a given review involves multiple aspect category sentiments, all aspects and the corresponding sentiments should be identified for the ASPE task. To this end, we design multiple classifiers to detect the aspect category and sentiment polarity. The **MPADB_Joint** structure is established on the basis of the BiLSTM network model with the multiple perspective attention and joint strategy (Fig. 2).

3.1. Framework of the proposed model

According to Fig. 2, the proposed model contains five layers: the input layer, the shared BiLSTM layer, the special BiLSTM layer, the multiple perspective attention layer, and the multilabel classifier layer. We describe each layer in detail as follows.

3.1.1. Input layer

Given a sentence $s = \{w_1, w_2, \ldots, w_n\}$ consisting of n words, we first map each word to a continuous vector space $E_w \in R^{v\times d}$ by using pretrained embedding, such as GloVe. In the expression of vector space, v is the size of the vocabulary set, and d is the embedded dimension. The set of aspect categories is denoted as $ASP = \{a_j\}_{j=1}^{m}$. We use the random initialization method to represent the aspect vectors. These vectors are then fine-tuned during the training stage.

3.1.2. Shared BiLSTM layer

BiLSTM has been proven to be an effective way to fuse context information into word embedding [13]. The hidden layer state shared by the first BiLSTM is represented as a sequence of fusion context information. We abbreviate the computation of the forward LSTM as $\overrightarrow{LSTM}(w_i)$ and the backward LSTM as $\overleftarrow{LSTM}(w_i)$. We concatenate $\overrightarrow{LSTM}(w_i)$ and $\overleftarrow{LSTM}(w_i)$ as the output of BiLSTM at step i by using formula (1).

$$h1_i = \left[\overrightarrow{LSTM}(w_i); \overleftarrow{LSTM}(w_i)\right] \tag{1}$$

3.1.3. Special BiLSTM layer

In the ACSA task, the aspect category information has a strong guiding role. In the case in which the aspect category is not given in advance, we traverse the aspect category a_j in the aspect category set ASP. In this way, the information of different aspects is integrated into the shared BiLSTM representation $h1_i$. BiLSTM is

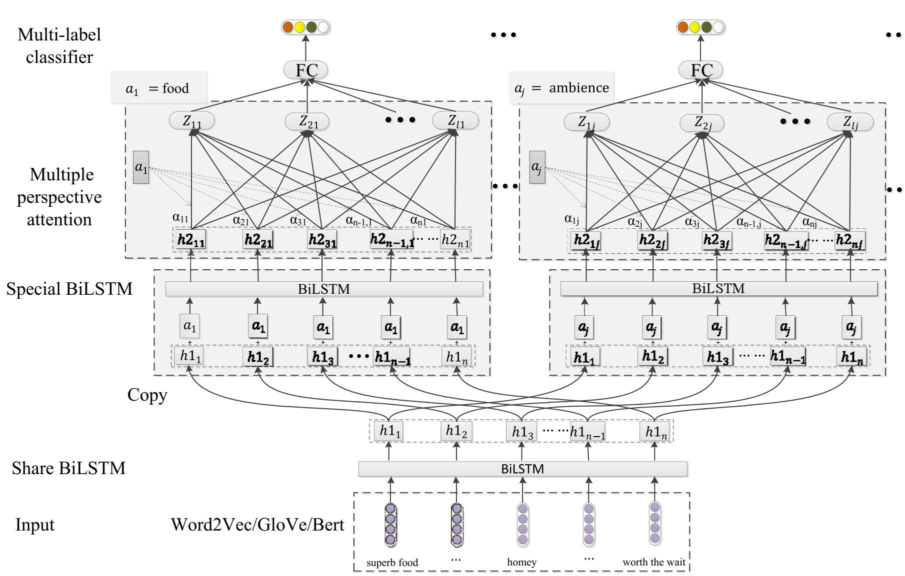

Fig. 2. Framework of MPADB_Joint.

used to model the internal association between the context and the aspect category to form the embedded representation of the fusion of the aspect category a_j and the context information via formula (2).

$$h2_{ij} = \left[\overrightarrow{LSTM}(h1_i; a_j); \overleftarrow{LSTM}(h1_i; a_j)\right] \tag{2}$$

3.1.4. Multiple perspective attention layer

In the ACSA task of a given aspect category, single attention is used to focus on the relevant information of the aspect category. For sentences that contain implicit multiple aspect categories, the information capturing ability of a single attention module cannot meet the requirements of the task. Hence, in each aspect category layer, we use the multiple perspective attention mechanism to make the model pay attention to different parts of the input during the training process, expand the representation space, and enhance the information that needs to be focused on.

For each aspect category level, attention input $h2_{ij}$ is the output of the special BiLSTM layer. Attention weight α_{ij} is obtained by calculating the correlation between aspect category a_j and state $h2_{ij}$. The weight is the probability that state $h2_{ij}$ is correctly noted when inferring the sentiment polarity of aspect category a_j. α_{ij} can be obtained by constructing a score function. Dot, general, and concatenation are three commonly used scoring functions [19]. However, a number of studies have shown that the linear offset of a vector can capture the relationship between two words [23]. Therefore, we also use minus [24] attention to capture the relationship between an aspect and a document. Four scoring functions are shown in formula (3).

$$score(h2_{ij}; a_j) = \begin{cases} h2_{ij}a_j, & \text{dot} \\ h2_{ij}Wa_j, & \text{general} \\ W\left[h2_{ij}a_j\right], & \text{concatenation} \\ V^T \tanh W(h2_{ij} - a_j) + b), & \text{minus} \end{cases} \tag{3}$$

where W is the parameter matrix, V is the parameter vector, and b is the bias.

For each aspect category, the following attention model is designed by using the score function of formula (3), as shown in formulas (4)–(5).

$$\alpha_{ij} = \frac{exp(score(h2_{ij}; a_j))}{\sum_{i=1}^{n} exp(score(h2_{ij}; a_j))} \tag{4}$$

$$z_j = \sum_{i=1}^{n} \alpha_{ij} h2_{ij} \tag{5}$$

For the attention representation of subspaces from different perspectives of a given aspect, formulas (4) and (5) are used several times in establishing multiple attention modules $\{z_{lj}\}_{l=1}^{t}$, corresponding to aspect category a_j. These attention modules can be executed in parallel. Inspired by the multiple attention in Transformer [30], $\{z_{lj}\}_{l=1}^{t}$ is concatenated and projected to construct the final representation of the sentence relative to aspect category a_j via formula (6).

$$Mult_j = \left[z_{1j}; z_{2j}; \ldots; z_{tj}\right] W_{j1} \tag{6}$$

where z_{lj} denotes the representation of the l-th attention module corresponding to a_j and W_{j1} is the parameter matrix.

3.1.5. Joint training strategy for ASPE task

For a whole sentence, the sentiment of each aspect category needs to be predicted accordingly. We design a training strategy that combines multiple classifiers to transform the ASPE task into a multilabel classification task.

For each aspect category as $a_j (j = 1, 2, \ldots, m)$, we use formula (6) to obtain the representation of the aspect-level attention layer and then input the result into the full connection layer for use by the sentiment classifier. To exclude aspect categories that are not involved in the sentence, we set the sentiment label to four types $\{-2, -1, 0, 1\}$, where $-2, -1, 0$, and 1 denotes none, negative, neutral, and positive, respectively. The sentiment polarity classification result of the aspect category is determined by the label with the highest probability obtained by formula (7).

$$y_j = softmax(Mult_j W_{j2} + b_j), \quad j = 1, 2, \ldots, m \tag{7}$$

By using formula (7), we can train multiple classifiers jointly and obtain the output result of multiple sentiment labels via formula (8).

$$y = (y_1, y_2, \ldots, y_m) \tag{8}$$

3.2. Orthogonal regularization

Multilabel joint training can be used for aspect detection and sentiment analysis simultaneously. In the process of network optimization, we find that the attentional weights of some samples would be excessive concentration of the nonexistent aspect in the review. That is, the distributed representation of attention with a different aspect category is learned to have an almost identical weight. Inspired by [11], to solve this problem, we propose the orthogonal attention mechanism and introduce it to model optimization.

This regularization term forces orthogonality between attention weight vectors of different aspects so that different aspect categories attend to different parts of the given sentence with *minimal overlap*. We apply this regularization to the multiple perspective attention layer. Let s be a sentence. Suppose s contains non-overlapping aspects category $a_1, a_2, \ldots, a_m, ,\alpha_j (j = 1, 2, \ldots, m)$ denotes the attention weight vectors of $a_j, \alpha_j = \left[\alpha_{1j}, \alpha_{2j}, \ldots, \alpha_{nj}\right]$, and $M \in \mathbb{R}^{m \times n}$ denotes a two-dimensional attention matrix to calculate the orthogonal regularization term R_o, as shown in formula (9).

$$R_o = ||M^T M - I||_2 \tag{9}$$

3.3. Model optimization

Given the imbalanced problem in the data, we can consider imposing the different costs caused by the sample misclassification of different aspect categories to balance the loss. Intuitively, the sentiment label -2 accounts for a large proportion in the data and is tagged as None, which is relatively insignificant in the discrimination of sentiment polarity of aspect categories. To obtain a high reward (penalty) when predicting the correct (wrong) polarity of a sentiment and to obtain a high weight for sentiment labels with few occurrences, we define the weight calculation by using formulas (10)–(12). Therefore, in the training process, the model focuses on optimizing the correctness of samples with sentiment labels.

$$\beta_{kj} = exp\frac{min\left(\{N_{kj}\}_{k=1}^{4}\right)}{N_{kj}} \tag{10}$$

$$\beta_{kj}^{*} = \frac{\beta_{kj}}{\sum_{k=1}^{4} \beta_{kj}} \tag{11}$$

where N_{kj} is the number of sentiment label k in the aspect category a_j. β_{kj} is the weight of sentiment label k in the aspect category a_j, and β^*_{kj} is the normalized representation of β_{kj}.

The objective function of model optimization training is to minimize the cross-entropy loss function, and it is given by formula (12).

$$L = -\sum_{j=1}^{m}\sum_{k=1}^{4}\beta^*_{kj}y_{kj}\log(\hat{y}_{kj}) \tag{12}$$

where y_{kj} is the ground truth sentiment label of aspect category a_j in given sentence, and $\hat{y}_{kj}$ is the probability of predicting sentiment label of aspect category a_j. L is the final objective function.

4. Experiments

We have conducted extensive experiments to compare our proposed techniques with existing state-of-the-art techniques.

4.1. Datasets

4.1.1. Restaurant 2014 (Res 2014)

The SemEval 2014 Restaurant data set is commonly used in ACSA task. A set of aspect categories includes "food", "price", "service", "ambiance" and "misc". The four sentiment polarities are positive, negative, neutral, and conflict. We remove the conflict label data.

4.1.2. Restaurant_Large (Res_Large)

[32] constructed a large dataset called "Restaurant-Large" by merging restaurant reviews for the period of 2014–2016. Data incompatibilities are fixed during merging.

4.1.3. CCF BDCI 2018

This dataset is a Chinese ASPE dataset and comprises users' evaluation of correlated content in an auto forum. One review may contain multiple aspect categories. The testing dataset is not released, and the test system is already inaccessible. Therefore, we use fivefold cross-validation on the training dataset to validate the performance of our proposed method.

We merge the samples with same review text and different aspect categories in the ACSA dataset. Then the merged data set is used for the ASPE task. The detailed statistics of the dataset are presented in Table 1.

4.2. Comparison methods

The research on ACSA is not as extensive as that on ATSA. For the baselines in this work, we use popular models that have exhibited excellent performances in ACSA. They are described as follows.

End-to-end LSTM/CNN [27] jointly models the detection of aspects and the classification of their polarities in an end-to-end trainable neural network, such as LSTM and CNN.

Char-CNN-CNN [33] uses CNN to integrate the feature representation of character level and word level.

TDLSTM [28] is a simple LSTM network for target-dependent sentiment classification.

AE-LSTM/AT-LSTM [31] is a variant of the ATAE model.

ATAE [31] is an attention-based LSTM for ACSA task. It fuses the aspect information at the input of LSTM and adds an attention mechanism on top of the LSTM layer.

HEAT [6] captures aspect information to help capture the sentiment information of the specific aspect of a sentence, so as to improve the accuracy of ACSA.

GCAE [32] uses the gated convolutional neural network, which adds aspect information in the process of convolution to conduct sentiment classification for different aspects.

MPADB is the model proposed for ACSA task in this work.

Bert [7] is a model in which we directly use the representation of "[CLS]" as a classification feature to fine-tune the BERT model for every aspect category classification.

MPADB_Bert is the initialization of MPADB replaced by BERT.

***_AD** (**A**spect **D**etection) denotes the aspect detection using * as core, uses multiple two-way classifier for multilabel classification.

***_pip** (**pip**eline model) denotes the pip method using * as core, which in turn performs *_AD for aspect detection and * for ACSA.

***_Joint** (**Joint** model) denotes using * as core and add our joint strategy for ASPE.

For the above models, The results with "♮" are retrieved from the original papers. In the comparison experiments, the models do not have corresponding experimental results for the datasets. Thus, we re-implement the three relatively advanced models of ATAE, HEAT, and GCAE. Each competitor is optimized independently.

4.3. Implementation details and evaluation metrics

The initialization of the input word embedding of the network adopts GloVe[2] [25] for English datasets and Word2Vec[3] [23] for Chinese datasets. The number of units in each hidden layer is set to 300. The dropout from the hidden layer to the output layer is 0.5. We adopt the Adam optimizer, the learning rate is 0.001, the batch size is 32. In this study, F1-measure[4] is adopted to test the effectiveness of "aspect category + sentiment label". We adopt the *accuracy* metric to evaluate the performance of aspect category sentiment classification. According to "aspect category + sentiment label", it is necessary to identify whether both the number and result are correct. More specifically, precision (P), recall (R) and F1 value (F) are defined using formula (13)–(15).

$$P = \frac{T_P}{T_P + F_P} \tag{13}$$

$$R = \frac{T_P}{T_P + F_N} \tag{14}$$

$$F = \frac{2 * P * R}{P + R} \tag{15}$$

T_P is the correct number of judgment where the judgment results of "aspect category + sentiment label" are completely correct.

F_P is the number of judgment errors, where models recognize additional "aspect category + sentiment label" beyond the actual quantity contained in the sample.

F_N is the number of missed judgments, which means that the number of identified "aspect category + sentiment label" is less than the actual number contained in the sample.

Accuracy is defined as:

$$accuracy = \frac{T}{N} \tag{16}$$

where T is the number of correctly predicted samples, N is the total number of samples.

[2] https://nlp.stanford.edu/projects/glove/.
[3] https://github.com/Embedding/Chinese-Word-Vectors.
[4] https://www.datafountain.cn/competitions/310/datasets.

Table 1
Proportion of experimental datasets.

DataSets	Aspect Category	Len	Positive		Negative		Neutral		Total	
			Train	Test	Train	Test	Train	Test	Train	Test
Res 2014	5	69	2179	657	839	222	500	94	3518	973
Res_Large	8	69	2710	1505	1189	680	757	241	4656	2426
BDCI	10	128	2048		2036		8488		12572	

4.4. Experimental results and analysis

4.4.1. Comparison experiment

To verify the effectiveness of the proposed method, we conduct the following comparative experiments on three data sets, namely, Res 2014, Res_Large, and BDCI datasets respectively.

4.4.1.1. Experiments in aspect category detection. Our objective is to predict the aspect category of a review. Most existing research focused on ACSA, and only a few studies have solely focused on the detection of aspect category detection alone. We modify the last layer of our framework to make it suitable for the aspect category. ATAE, HEAT, and GCAE are modified accordingly under our model framework (we use multiple classifiers to detect whether or not an aspect category exists and convert the results into a two-way classification problem). The results are shown in Table 2.

Note that GCAE and HEAT apply aspect information only to gated units or attention mechanisms. ATAE makes the model slightly better than GCAE and HEAT by using aspect information twice. Unlike that in the aforementioned models, we use the aspect category information several times in the special BiLSTM layer and multiple perspective attention layer; the shared BiLSTM layer then captures the context information. In this way, our model can easily capture the corresponding aspect information. Hence, our model is better than other models consistently and thus is effective for the ACSA task.

Table 2
F1 scores of aspect category detection.

models	Res 2014	Res_Large	BDCI
ATAE_AD	87.42	74.03	88.29
HEAT_AD	87.37	74.35	88.25
GCAE_AD	87.31	74.66	88.13
MPADB_AD	**88.42**	**75.37**	**89.08**

The best scores are in bold.

4.4.1.2. Experiments in aspect category sentiment analysis. To verify whether or not the proposed method is effective in predicting the sentiment polarity of a given aspect category, we choose *accuracy* of the evaluation measure and compare it with existing models. As ATAE, HEAT, and GCAE lack corresponding experimental results in BDCI data, we reimplemented them for comparison. The comparative experimental results are shown in Table 3.

As shown in Table 3, the experimental results of the proposed method are better than those of most existing models for the three datasets, especially in Res_Large and BDCI. However, in Res 2014, the results of the proposed method are slightly lower than those of the HEAT method. The main reason is that multi-head attention has many parameters while the sample of the Res 2014 dataset is small, thereby potentially leading to insufficient training.

Table 3
Accuracy of aspect category sentiment analysis.

models	Res 2014	Res_Large	BDCI
TDLSTM	82.6$^{\natural}$	–	–
AE-LSTM	82.5$^{\natural}$	–	–
AT-LSTM	83.1$^{\natural}$	–	–
ATAE	84.0$^{\natural}$	83.91$^{\natural}$	73.52
HEAT	**85.1**$^{\natural}$	85.12	73.56
GCAE	84.6	85.92$^{\natural}$	73.52
MPADB	85.0	**86.48**	**73.81**

The best scores are in bold.

4.4.1.3. Experiments in aspect category and sentiment pair extraction. The experimental objective is to predict the aspect category of a review and the corresponding sentiment polarity where we employ F1 to evaluate the performance. As the existing ACSA methods are not applied to experiments on aspect category detection and sentiment classification, we reimplement them in this work. The result of Char-CNN-CNN on BDCI is retrieved from the original paper, where their dataset is a subset of the dataset in this paper. The results are listed here for reference. The experiment results are reported in Table 4. The best results of each method (pipeline and joint) are highlighted in bold.

Table 4 indicates that the proposed model obtains the best experimental results among all the joint models. In addition, the results of the proposed framework are better than the pipeline results for the Res_Large and BDCI datasets. Hence, our model can limit the error accumulation caused by pipeline methods. We find that the Res 2014 dataset only has five aspects, which are relatively easy to classify. Tables 2 and 3 indicate that the model has high scores in the aspect detection task and the ACSA task in Res 2014. Therefore, in the Res 2014 dataset, the pipeline method is superior to our joint model. The pipeline model and the joint model use the same framework, and only the output part of the classifier is different. Therefore, error accumulation can be reduced only when the aspect detection and sentiment analysis are performed well. However, these pipeline models do not perform well for all datasets. With the increase in the number of categories and the data volume, our model becomes fully trained and thus achieves good results for the Res_Large and BDCI datasets. The experimental results demonstrate the effectiveness of the proposed framework.

Table 4
F1 scores of aspect category and sentiment pair extraction.

models	Res 2014	Res_Large	BDCI
ATAE_pip	79.44	66.69	69.81
HEAT_pip	79.60	66.74	69.55
GCAE_pip	78.82	67.16	**69.92**
MPADB_pip(Our)	**80.41**	**67.82**	69.10
End-to-end LSTM	75.82†	65.35†	68.47†
End-to-end CNN	77.58†	64.54†	70.54†
Char-CNN-CNN	–	–	64.42$^{\natural}$
ATAE_Joint	78.54†	67.83‡	71.97†
HEAT_Joint	78.82‡	67.15†	72.32‡
GCAE_Joint	78.86†	68.24‡	71.36†
MPADB_Joint(Our)	**79.99**	**68.83**	**73.43**
Bert_Joint	83.88	75.61	73.63
MPADB_Bert_Joint(Our)	**85.01**	**76.64**	**73.80**

We have carried out P-test upon the 5-folded experiment performances. MPADB_Joint is respectively paired with each baselines. The marker † refers $p < 0.05$, and the marker ‡ refers $p < 0.1$.

Table 5
The results of P, R, F.

models	P	R	F
End-to-end LSTM	60.45	78.94	68.47
End-to-end CNN	60.52	84.52	70.54
ATAE	63.11	83.74	71.97
HEAT	63.97	83.18	72.32
GCAE	63.31	81.77	71.36
MPADB	**64.63**	**85.01**	**73.43**

The best scores are in bold.

Table 6
F1 scores of different scoring functions.

models	Res 2014	Res_Large	BDCI
BiLSTM-Dot-Matt	78.80	66.51	72.58
BiLSTM-General-Matt	78.51	67.60	72.56
BiLSTM-Concat-Matt	78.45	67.16	72.90
BiLSTM-Minus-Matt	**78.86**	**67.98**	**73.22**

The best scores are in bold.

After switching to BERT representations, we show that MPADB_Bert_Joint achieves enhanced performance. Although the original BERT_Joint model already provides strong prediction power, MPADB_Bert_Joint consistently improves over Bert_Joint after fine-tuning. Hence, our model can effectively utilize these semantic representations. Moreover, capturing aspect category information explicitly is useful for the BERT-based model.

4.4.2. Result analysis of BDCI

We performed a 5-fold cross validation on the BDCI. Then merge the results of the five test sets to obtain the prediction result of a complete dataset. Then, we compute three valuation metrics Precision (P), Recall (R), F-Measure (F), which are used to analyze the performance of the different models again.

As can be seen from Table 5, the results once again demonstrate our proposed MPADB outperform existing 5 models consistently in terms of all the evaluation metrics, namely precision, recall and F score. In particular, our model is significantly higher than other comparison methods in terms of P and R. We have a high recall due to our repeated use of relevant aspect information. And multiple perspective attention allows sentiment of each aspect to be accurately judged.

4.4.3. Effectiveness of different attention scoring functions

According to the attention score functions in the hierarchical attentional mechanism in Section 3, four different BiLSTM-X-

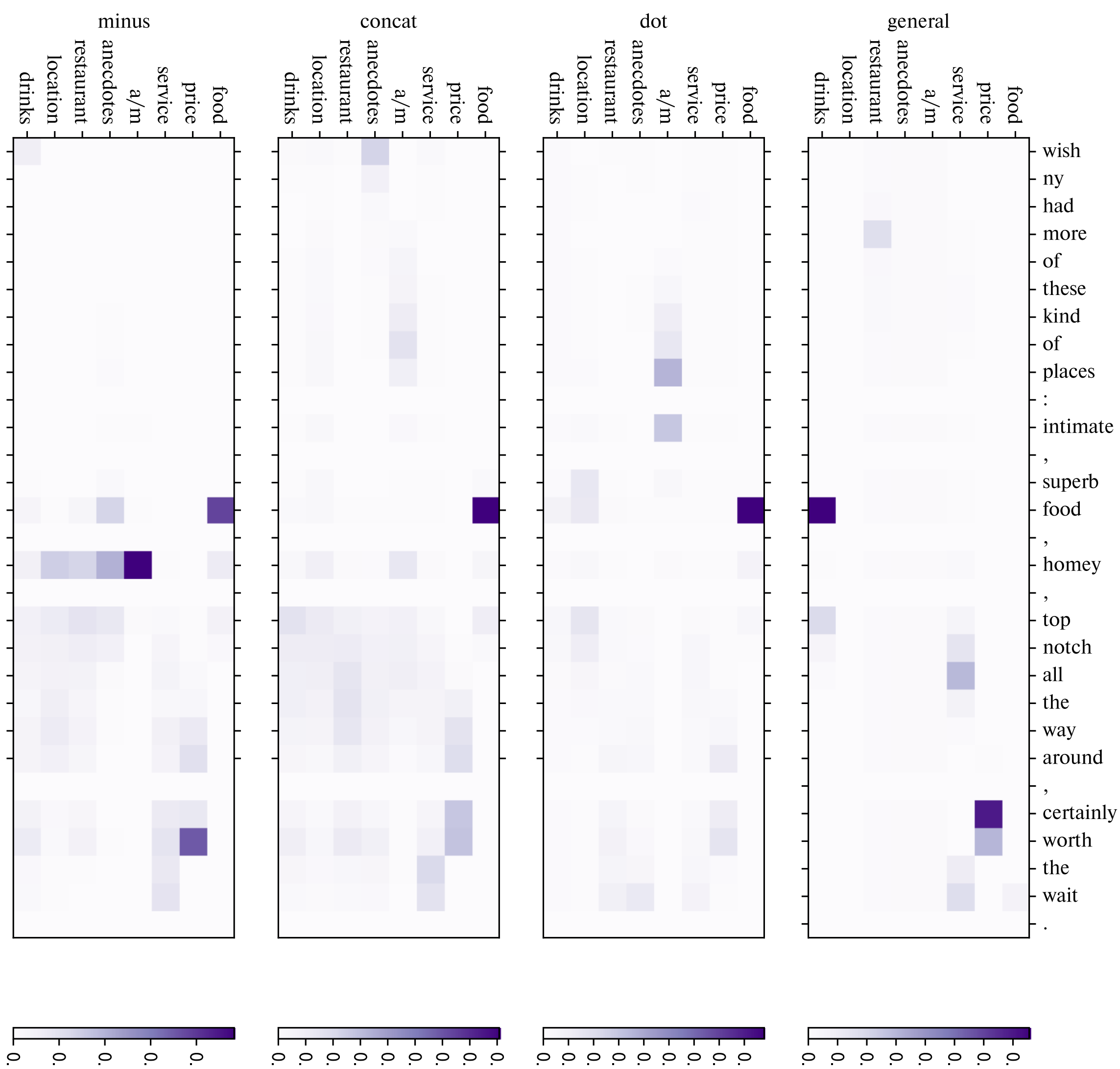

Fig. 3. Visualization of attention weight for Example 1.

Matt models are obtained by using formulas (3)–(5). X denotes Dot, General, Concat, and Minus. To verify the performance of the four score functions in ASPE, we conduct a comparative experiment of three datasets. The experimental results are shown in Table 6.

From Table 6, we can see that the Minus attention has the best F1 score among all score functions. To further illustrate the advantages and disadvantages of the four functions of attention, we show in Fig. 3 the visualization of the attention weight.

As shown in Fig. 3, the aspect categories involved in Example 1 are "food", "ambiance", and "restaurant". Only Minus attention makes the correct predictions for the aspect category and sentiment label. Therefore, Minus-attention can pay attention to aspect-related words in the text. Moreover, it has a strong ability to capture context information, thereby improving the recognition effectiveness of the aspect category and sentiment pair.

4.4.4. Influence of number of attention modules on multiple attention

Clearly, the number of attention modules affects model performance. To analyze the influence of the number of attention modules on model performance, we adopt the minus attention mechanism to evaluate our framework with multiple attention models. The results are shown in Table 7, where BiLSTM_Minus_Matt×N denotes N attention models.

As shown in Table 7, our model with two or three attention models are satisfactory in terms of F1, but the attention number is inconsistent among different datasets. Multiple-times attention can capture sufficient sentiment information from different subspaces using more attention models. Meanwhile, F1 does not monotonically increase with the attention number. We observe that 4-times is not as good as 2-times, because the more complex the models are, the more difficult they are to train and the less generalizable they are. Therefore, in the subsequent experiments, we adopt the optimal attention number for different datasets. Similar results type have also appeared in [5].

4.4.5. Effects of orthogonal attention mechanism

This regularization term forces orthogonality between attention weight vectors of different aspects. For the aspect categories that do not exist in sentences, the orthogonal attention mechanism attends to the different parts of the sentence with minimal overlap. To verify the effectiveness of the regularized attention mechanism in the BDCI dataset, we show the experimental results in Table 8.

We select the attention weight of an attention block in a sample for visual analysis, as shown in Fig. 4.

Table 7
F1 scores of attention module number.

models	Res 2014	Res_Large	BDCI
BiLSTM_Minus_Matt×1	78.86	67.98	73.22
BiLSTM_Minus_Matt×2	79.21	**68.88**	**73.43**
BiLSTM_Minus_Matt×3	**79.99**	68.58	73.41
BiLSTM_Minus_Matt×4	79.14	68.31	73.36

The best scores are in bold.

Table 8
F1 scores of orthogonal attention mechanism.

Models	No regularization	Regularization	
BiLSTM_Minus_Matt×1	73.22	73.26	↑
BiLSTM_Minus_Matt×2	73.43	73.36	↓
BiLSTM_Minus_Matt×3	73.41	73.51	↑
BiLSTM_Minus_Matt×4	73.36	73.42	↑

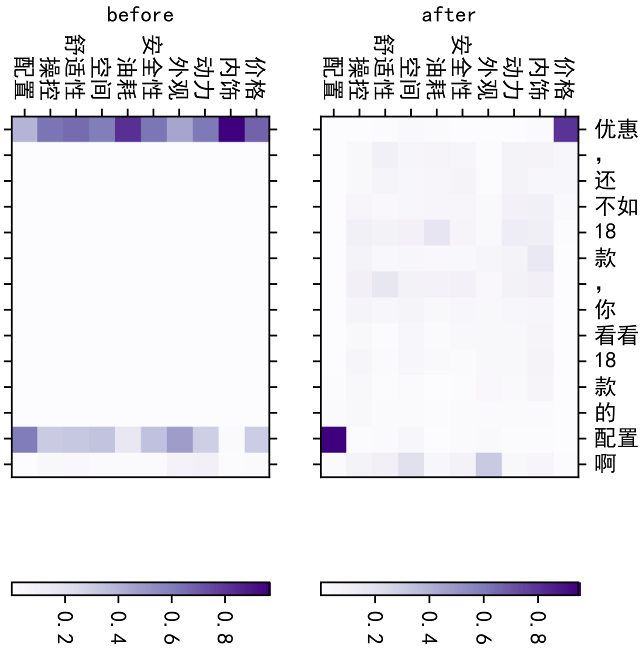

Fig. 4. Visualization of orthogonal for Example 2.

Example 2: 优惠好少，还不如 18 款，你看看 18 款的配置啊。(Discount is low, is not as good as 18, you see the configurations of 18.).

From Table 8 and Fig. 4, we can see that the F1 score of the regularized attention mechanism is superior to that without the regularized attention mechanism for BiLSTM_Minus_Matt×1, BiLSTM_Minus_Matt×3, and BiLSTM_Minus_Matt×4 on the BDCI dataset and is slightly lower than that for BiLSTM_Minus_Matt×2. We also obtain a more interpretable result. As shown in Fig. 4 on the left, the first row shows several overlaps prior to the application of the regularized attention mechanism. The visualization results on the right show that regularization can largely solve the problem of overlapping attention after using the regularized attention mechanism.

5. Conclusion

This study proposes **MPADB**, which uses two layers of BiLSTM to capture global semantic information and aspect category information. At the same time, it uses the multiple attention mechanism to pay attention to aspect category and sentiment information for a couple of times, thereby alleviating the problem of mutual interference between sentiment information of multiple aspect categories. In addition, we use a multilabel joint training model to predict aspect categories and corresponding sentiments simultaneously in ASPE task. The approach can reduce the accumulation of errors caused by pipeline methods. Finally, the MPADB model with multiple attention is used to achieve good results in ACSA and ASPE tasks given the Res_Large and BDCI datasets regardless of whether the aspect category is given. Clearly, the computational complexity of the model is naturally increased due to the use of multiple perspective attention and orthogonal regularization to improve our models' accuracy and interpretability. How to further reduce time complexity while maintain the accuracy and interpretability will be an important direction for our future research. In addition, we will consider using a multitask model to predict aspects and sentiments. Finally, we will also explore a new attention mechanism for effectively capture the interaction between aspect categories and contexts.

CRediT authorship contribution statement

Yujie Fu: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Jian Liao:** Writing - review & editing, Funding acquisition. **Yang Li:** Visualization, Investigation, Formal analysis. **Suge Wang:** Writing - review & editing, Supervision, Resources, Project administration, Funding acquisition. **Deyu Li:** Project administration, Funding acquisition. **Xiaoli Li:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank all anonymous reviewers for their valuable comments and suggestions which have significantly improved the quality and presentation of this paper. The works described in this paper are supported by the National Natural Science Foundation of China (61632011, 62076158, 62072294, 61906112), the Key Research and Development Program of Shanxi Province (201803D421024), Natural Science Foundation of Shanxi Province, China (201901D211174), Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (2019L0008).

References

[1] M.E. Basiri, S. Nemati, M. Abdar, E. Cambria, U.R. Acharya, ABCDM: an attention-based bidirectional CNN-RNN deep model for sentiment analysis, Fut. Gen. Comput. Syst. 115 (2021) 279–294.
[2] E. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2016) 102–107.
[3] E. Cambria, A. Hussain, C. Havasi, C. Eckl, Sentic computing: exploitation of common sense for the development of emotion-sensitive systems, in: Proceedings of the Second International Conference on Development of Multimodal Interfaces: Active Listening and Synchrony, 2009, pp. 148–156.
[4] E. Cambria, Y. Li, F.Z. Xing, S. Poria, K. Kwok, Senticnet 6: ensemble application of symbolic and subsymbolic AI for sentiment analysis, in: Proceedings of the 29th ACM International Conference on Information and Knowledge Management, 2020, pp. 105–114.
[5] P. Chen, Z. Sun, L. Bing, W. Yang, Recurrent attention network on memory for aspect sentiment analysis, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 452–461.
[6] J. Cheng, S. Zhao, J. Zhang, I. King, X. Zhang, H. Wang, Aspect-level sentiment classification with HEAT (hierarchical attention) network, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017, pp. 97–106.
[7] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, 2018. arXiv preprint arXiv:1810.04805..
[8] S. Gu, L. Zhang, Y. Hou, Y. Song, A position-aware bidirectional attention network for aspect-level sentiment analysis, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 774–784.
[9] P. Haiyun, X. Lu, B. Lidong, H. Fei, L. Wei, S. Luo, Knowing what, how and why: a near complete solution for aspect-based sentiment analysis, in: Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, 2020, pp. 8600–8607.
[10] R. He, W.S. Lee, H.T. Ng, D. Dahlmeier, Effective attention modeling for aspect-level sentiment classification, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 1121–1131.
[11] M. Hu, S. Zhao, L. Zhang, K. Cai, Z. Su, R. Cheng, X. Shen, CAN: constrained attention networks for multi-aspect sentiment analysis, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 4601–4610.
[12] B. Huang, Y. Ou, K.M. Carley, Aspect level sentiment classification with attention-over-attention neural networks, in: R. Thomson, C. Dancy, A. Hyder, H. Bisgin (Eds.), Social, Cultural, and Behavioral Modeling, 2018, pp. 197–206..
[13] S. Lai, L. Xu, K. Liu, J. Zhao, Recurrent convolutional neural networks for text classification, in: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015, pp. 2267–2273.
[14] X. Li, L. Bing, W. Lam, B. Shi, Transformation networks for target-oriented sentiment classification, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018, pp. 946–956.
[15] X. Li, L. Bing, P. Li, W. Lam, A unified model for opinion target extraction and target sentiment prediction, in: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, 2019, pp. 6714–6721.
[16] X. Li, L. Bing, W. Zhang, W. Lam, Exploiting BERT for end-to-end aspect-based sentiment analysis, in: Proceedings of the 5th Workshop on Noisy User-generated Text, 2019, pp. 34–41..
[17] B. Liu, Opinion Mining and Sentiment Analysis, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 459–526.
[18] B. Liu, L. Zhang, A Survey of Opinion Mining and Sentiment Analysis, Springer, US, Boston, MA, 2012, pp. 415–463.
[19] T. Luong, H. Pham, C.D. Manning, Effective approaches to attention-based neural machine translation, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 1412–1421.
[20] D. Ma, S. Li, X. Zhang, H. Wang, Interactive attention networks for aspect-level sentiment classification, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017, pp. 4068–4074.
[21] Y. Ma, H. Peng, E. Cambria, Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, 2018, pp. 5876–5883.
[22] Y. Ma, H. Peng, T. Khan, E. Cambria, A. Hussain, Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis, Cogn. Comput. (2018) 639–650.
[23] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Proceedings of the 26th International Conference on Neural Information Processing Systems, 2013, pp. 3111–3119..
[24] L. Mou, R. Men, G. Li, Y. Xu, L. Zhang, R. Yan, Z. Jin, Natural language inference by tree-based convolution and heuristic matching, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016, pp. 130–136.
[25] J. Pennington, R. Socher, C. Manning, GloVe: global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1532–1543.
[26] S. Poria, I. Chaturvedi, E. Cambria, F. Bisio, Sentic LDA: improving on LDA with semantic similarity for aspect-based sentiment analysis, in: Proceedings of the 2016 International Joint Conference on Neural Networks, 2016, pp. 4465–4473.
[27] M. Schmitt, S. Steinheber, K. Schreiber, B. Roth, Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 1109–1114.
[28] D. Tang, B. Qin, X. Feng, T. Liu, Effective LSTMs for target-dependent sentiment classification, in: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016, pp. 3298–3307.
[29] D. Tang, B. Qin, T. Liu, Aspect level sentiment classification with deep memory network, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 214–224.
[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, U. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 6000–6010..
[31] Y. Wang, M. Huang, X. Zhu, L. Zhao, Attention-based LSTM for aspect-level sentiment classification, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 606–615.
[32] W. Xue, T. Li, Aspect based sentiment analysis with gated convolutional networks, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018, pp. 2514–2523.
[33] Z. Zeng, J. Ma, M. Chen, X. Li, Joint learning for aspect category detection and sentiment analysis in chinese reviews, Inf. Retrieval (2019) 108–120.
[34] P. Zhu, T. Qian, Enhanced aspect level sentiment classification with auxiliary memory, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 1077–1087.

Yujie Fu is currently pursuing the Ph.D. degree with the School of Computer and Information Technology, Shanxi University, China. His research interests include sentiment analysis, stance detection and argument mining.

Jian Liao, Ph.D., associate professor in the School of Computer and Information Technology, Shanxi University, China. His research interests include implicit sentiment analysis and representation learning.

Yang Li, Ph.D., lecturer in the School of Finance, Shanxi University of Finance and Economics, China. Her research interests include natural language processing and text sentiment analysis.

Suge Wang, received the Ph.D. degree from Shanghai University. She is currently a Professor with the Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, School of Computer and Information Technology, Shanxi University. She has published more than 50 papers in international journals. Her research interests include natural language processing, text sentiment analysis, and machine learning.

Deyu Li received the Ph.D. degree from Xi'an Jiaotong University. He is currently a Professor with the Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, School of Computer and Information Technology, Shanxi University. He has published more than 60 articles in international journals. His research interests include artificial intelligence, granular computing, data mining, and machine learning.

Xiaoli Li (Senior Member, IEEE) is currently a Principal Scientist at the Institute for Infocomm Research, A*STAR, Singapore. He also holds Adjunct Professor positions at Nanyang Technological University. His research interests include AI, data mining, machine learning, and bioinformatics. He has been serving as area chairs, senior PC Members in leading AI and data mining related conferences (including KDD, ICDM, SDM, PKDD/ECML, WWW, IJCAI, AAAI, ACL and CIKM). Xiaoli has published more than 200 high quality papers with more than 10,000 citations and won six best paper awards. He has led over ten research projects in collaboration with industry partners across a range of sectors.