

## RESEARCH ARTICLE

# IDCN: A Novel Interactive Dual Channel Network for Aspect Sentiment Triplet Extraction

NING LIU<sup>1</sup>, JIE HU, SHUNYU YAO, DAN LIU, (Member, IEEE), AND MINGCHUAN YANG

China Telecom Research Institute, Beijing 102209, China

Corresponding author: Ning Liu (liun14@chinatelecom.cn)

**ABSTRACT** Aspect sentiment triplet extraction (ASTE) is one of the important subtasks of aspect-based sentiment analysis, it aims at detecting the aspect terms, opinion terms, and the corresponding sentiment polarity, simultaneously. Most methods directly employ GCNs to capture the syntactic dependency information in ASTE. However, these methods may lead to error propagation. Besides, the GCN-based methods are weak at capturing sequence information and long-distance information. The general neural networks such as LSTM are good at capturing this kind of information. However, these general neural networks are weak at modeling syntactic dependency information. To alleviate the above problems, we propose a novel interactive dual channel network (IDCN) for ASTE. In IDCN, an interactive word pair generating (IWPG) module is designed to model the sequence information, long-distance dependency information, and correlation relations between word pairs, simultaneously. In the IWPG module, the dual channels can learn different representations. Based on these representations, the informative word-pair representations can be learned by the interaction mechanism of dual channels. Besides, we design the syntactic dependency fusion module to model the syntax dependency information by constructing word pair dependency relation tensors and pooling mechanism, which can naturally inject the syntactic dependency knowledge into the general neural networks and reduce error propagation. Abundant experiments have been performed on multiple datasets. The experimental results show that IDCN acquires state-of-the-art results and validates the effectiveness of IDCN.

**INDEX TERMS** Aspect sentiment triplet extraction, aspect-based sentiment analysis, natural language processing, deep learning, BERT.

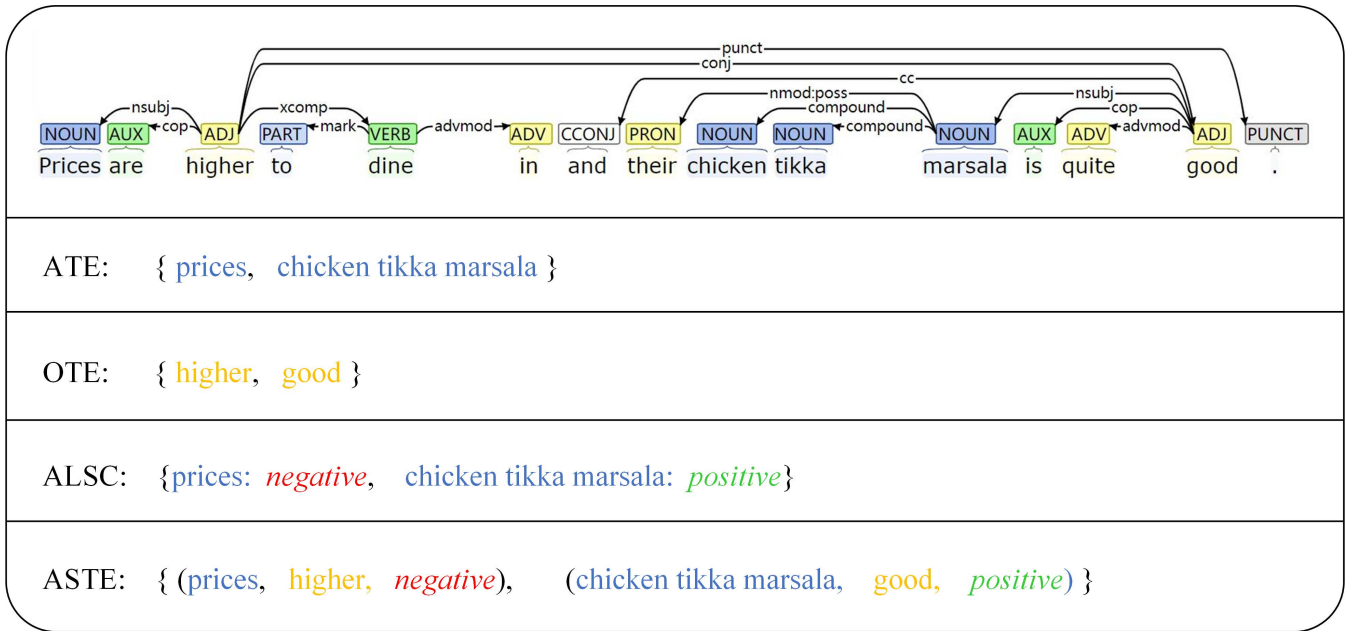
## I. INTRODUCTION

Aspect-based sentiment analysis (ABSA) has become a hot topic of research in natural language processing (NLP) [1], [2]. In the ABSA task, there are three basic tasks: aspect term extraction (ATE), opinion term extraction (OTE), and aspect level sentiment classification (ALSC) [3]. The ATE task aims at detecting aspect terms expressed in a specific sentence, the OTE task aims at detecting opinion terms expressed in a specific sentence, and the ALSC task aims at predicting the sentiment polarity toward a specific aspect term in a sentence [4]. Aspect sentiment triplet extraction (ASTE) is a combination of these three separate tasks. The ASTE task focuses on extracting aspect sentiment triplets that are expressed in a

sentence [5]. An aspect sentiment triplet consists of the aspect term, opinion term, and the associated sentiment polarity. The ATE, OTE, ALSC, and ASTE tasks are all illustrated in Fig. 1.

In the previous works of the ASTE task, there are three kinds of paradigms. The first type is the pipeline method. The pipeline methods extract aspect term, opinion term, and sentiment polarity of the aspect sentiment triplet independently [5]. The pipeline methods have achieved great success. However, the pipeline methods ignore the close relation between aspect term, opinion term, and sentiment polarity in the aspect sentiment triplet. And the error generated by the former approaches has a great influence on the performance of the subsequent approaches in pipeline methods. To alleviate the problems mentioned above. Some works treated the ASTE task as the machine reading comprehension task [3], [6]. The methods based on machine reading

The associate editor coordinating the review of this manuscript and approving it for publication was Agostino Forestiero<sup>1</sup>.



**FIGURE 1.** A sentence with its dependency relations is used to illustrate the ATE, OTE, ALSC, and ASTE tasks. The aspect terms and opinion terms are marked in blue and yellow, respectively. The positive sentiment polarity and the negative sentiment polarity are marked in red and green, respectively.

comprehension approaches are the second type in the ASTE task. The multiple subtasks are trained jointly in the machine reading comprehension approaches. The third type is the end-to-end method [7], [8], [9]. The end-to-end methods focus on extracting the aspect sentiment triplets by developing different kinds of tagging schemes based on pre-trained language models, such as BERT [10] and BART [11].

Graph convolutional networks (GCN) have obtained significant performance in many fields, such as the recommendation system, community detection, aspect-level sentiment classification, and so on. Recently, GCN has been also applied to the ASTE task which is constructed by syntactic parsing in a sentence [12]. Although these GCN-based methods have achieved excellent results, there still exist several challenges in the ASTE task.

These GCN-based methods are heavily dependent on the quality of the dependency tree of a sentence in ASTE. The errors from the syntactic parser which are out of the box can be propagated to GCN-based methods, and the model may learn the wrong and noise information for ASTE. Secondly, in ASTE, GCN directly operates on the dependency graph and it ignores the sequence information and long-distance dependency information in a sentence. Thirdly, the dependency relations can help the ASTE task to some extent. However, it is difficult to utilize general neural networks to model syntax dependency information. How to effectively integrate the syntax dependency information with general neural networks, such as Bi-LSTM, remains a challenge.

In this paper, we propose a novel interactive dual channel network (IDCN) to alleviate the aforementioned problems. Firstly, we design an interactive word pair generating (IWPG) module to model the sequence information and long-

distance information, as well as to capture the correlation relations between words in IDCN. In the IWPG module, two channels can learn different representations. Based on these representations, the informative word-pair representations can be obtained by the interaction mechanism of dual channels. Secondly, in order to model the sequence information and long-distance dependency information. The dual Bi-LSTM networks are employed in the IWPG module based on BERT. Thirdly, we develop a novel way to model the syntax dependency information and design the syntactic dependency fusion module in the IDCN model. We transform the dependency relations between words into the word pair dependency relation tensors. These word pair dependency relation tensors contain informative syntax dependency information about each word. The pooling mechanism is used to generate the syntactic dependency-aware word embeddings. And then these syntactic dependency-aware word embeddings are fed into the IWPG module. In short, the syntax dependency information can be learned by utilizing the word pair dependency relation tensors and the pooling mechanism in the IDCN model. The IDCN model does not operate on the dependency tree of a sentence. Hence, it can avoid error propagation from the dependency tree. The main contributions from this research are the following.

- We propose a novel IDCN model for the ASTE task. The IDCN model can simultaneously model the correlation relation between words and capture the sequence information, long-distance dependency information, and syntactic dependency information in a sentence.
- We propose a novel interactive word pair generating (IWPG) module in IDCN. IWPG designs dual channels to interactively learn the informative word-pair

representations and capture the semantic relations between word pairs.

- We propose a novel way to utilize the syntax dependency information and design the syntactic dependency fusion module in IDCN, which can naturally inject the informative syntax dependency information into the general neural networks such as Bi-LSTM, and alleviate the error propagation.
- Extensive experiments are conducted on multiple benchmark datasets. The experimental results show that IDCN obtains state-of-the-art results. Further, these results also verify the effectiveness of IDCN.

## II. RELATED WORK

Sentiment Analysis (SA) can be divided into three granularities: document-level sentiment analysis, sentence-level sentiment analysis, and fine-grained aspect-based sentiment analysis (ABSA) [13]. In document-level sentiment analysis, it assumes that a whole document only contains a single sentiment polarity. In sentence-level sentiment analysis, it assumes that a whole sentence only contains a single sentiment polarity, while a document can contain multiple different sentiment polarities. In aspect-based sentiment analysis, it assumes that a whole sentence can contain several different sentiment polarities [14], [15]. In ABSA, there are multiple basic subtasks: aspect term extraction (ATE), opinion term extraction (OTE), and aspect-level sentiment classification (ALSC). Recently, the aspect sentiment triplet extraction (ASTE) task is developed as a new subtask of ABSA. In essence, the ASTE task is a combination of three basic subtasks and is completely different from these three basic subtasks.

The ASTE task aims at extracting aspect terms, opinion terms, and the corresponding sentiment polarity in a sentence, simultaneously. The ATE task only focuses on extracting the aspect terms in a sentence, the OTE task only focuses on extracting the opinion terms in a sentence, and the ALSC task only focuses on predicting the sentiment polarity toward an entity or a specific aspect of the entity in a sentence. The ASTE task is more complex and harder than the ATE task, the OTE task, and the ALSC task.

In the ATE task, Hu and Liu [16] were the first to study detecting the aspect terms in a sentence. They developed several association rules to extract aspect terms. To further improve the performance of the rule-based methods, Popescu and Etzioni [17] applied the pointwise mutual information (PMI) in ATE. Tubishat et al. [18] developed new rules and the combination of dependency-based rules and pattern-based rules for explicit aspect term extraction. And they proposed the improved whale optimization algorithm (IWOA) for the selection of different rules. Ozyurt and Akcayol [19] proposed a sentence segment LDA (SS-LDA) to alleviate the data sparsity problem and the lack of co-occurrence patterns in the short sentence.

Zhang et al. [20] proposed a CNN-based model with filters that are dynamically generated by the aspect informa-

tion. Venugopalan and Gupta [21] proposed an enhanced guided LDA model with BERT to extract aspect terms. The BERT model is employed as the semantic filter to enhance the ability to incorporate semantics. To save training time, Kumar et al. [22] proposed a novel hierarchical self-attention network (HSAN), and they fused two-attention mechanisms to improve the performance of the model. Chen and Qian [23] proposed a novel active domain adaptation method for ATE. They developed the syntactic bridge and the semantic bridge to transfer knowledge across different domains. Klein et al. [24] proposed another cross-domain aspect term extraction method by multi-task learning. They utilized the relational features to improve the performance of the model.

In general, the OTE task is often viewed as a co-extraction task with other subtasks of ABSA. Yu et al. [25] proposed a multi-task learning framework for the aspect and opinion terms co-extraction. They developed a global inference method to model the intra-relation and inter-relation between these two tasks. Zhao et al. [26] treated the co-extraction task as the aspect-opinion pair extraction (AOPE) problem. They developed a multi-task learning framework based on shared spans. The aspect and opinion terms are jointly recognized by the span representations. Gao et al. [27] viewed the AOPE task as a machine-reading comprehension task and proposed a question-driven span labeling model (QDSL). In QDSL, they explored the internal relation between aspect terms and opinion terms to identify the aspect-opinion pairs in a sentence.

Zhang et al. [28] proposed a two-stage neural network model for the quadruple extraction in ABSA. The elements in the quadruple extraction are the aspect term, aspect category, opinion term, and the associated sentiment polarity, respectively. Dai et al. [29] investigated reinforcement learning in OTE and proposed a Padding-Enhanced Reinforcement learning model (PER). In PER, a multiplex heterogeneous graph is proposed to model the sequential information and syntactic information. To address the problem of lacking labeled data, Wu et al. [30] proposed a hybrid unsupervised method for the ATE and OTE tasks. The GRU network is used to extract aspect and opinion terms with the pseudo-labeled data.

In the ALSC task, Tang et al. [31] proposed the target-dependent LSTM (TD-LSTM) and the target-connection LSTM (TC-LSTM). In TD-LSTM, the left and right parts of a sentence are fed into the model, separately. TD-LSTM is weak at capturing aspect-dependent sentence information. In order to further model the relation between aspect terms and the contexts, aspect embedding and word embedding are concatenated in TC-LSTM. Wang et al. [32] proposed an attention-based LSTM with aspect embedding (ATAE-LSTM) model for predicting the sentiment polarity towards an aspect. The ATAE-LSTM model utilized the attention mechanism and aspect embedding to obtain the aspect-dependent sentence representation. Xue and Li [33] proposed a gated convolutional network with aspect

embedding (GCAE). In GCAE, the novel gated tanh-reLU units can generate aspect-dependent sentiment features.

To tackle the aspect sentiment bias generated by the pre-trained model, Cao et al. [34] proposed a no-aspect differential sentiment (NADS) framework. And they utilized contrastive learning between the raw sentence and the sentence template to improve the robustness. Zhao and Yu [35] proposed a knowledge-enabled language representation model BERT for ALSC, the sentiment knowledge is learned by a sentiment knowledge graph in the language representation model. Li et al. [36] proposed a dual graph convolutional networks (DualGCN) model for ALSC. DualGCN can model the semantic and syntax information simultaneously. Liang et al. [37] utilized the SenticNet to inject the affective knowledge into the model. They developed a graph convolutional network based on SenticNet to model the contextual affective knowledge toward an aspect. To utilize the dependency types between words, Tian et al. [38] proposed a type-aware graph convolutional network (T-GCN). In T-GCN, the attention mechanism is employed to model different relations. Li et al. [39] investigated the generative model for ALSC and proposed a joint term-sentiment generator (JTSG) model. In JTSG, the encoder is used to encode the sentence information, and the decoder is used to predict the aspect term and its corresponding sentiment polarity. Yadav et al. [40] proposed a novel human-interpretable learning method for ALSC, they utilized the Tsetlin Machine (TM) to learn the particular sentiment toward the specific aspect.

Peng et al. [5] were the first to study the ASTE task. They proposed a two-stage framework to extract triplets from a sentence. In particular, the first stage of the model can output the aspect term, opinion term, and sentiment polarity. The second stage of the model can generate the triplet based on the outputs from the first stage. Mao et al. [3] viewed the ASTE task as a machine reading comprehension (MRC) problem. They developed two shared BERT-MRC models to solve two MRC problems which are constructed for the ASTE task. Meanwhile, Chen et al. [6] viewed the ASTE task as a multi-turn machine reading comprehension (MTMRC) problem. They designed three types of queries to model the relation between subtasks, and they proposed the bidirectional MRC (BMRC) to detect triplets in a sentence. Wu et al. [9] proposed a novel end-to-end framework that is based on the developed grid tagging scheme (GTS). In GTS, an effective inference strategy was designed to model the mutual indications between opinion factors. Zhang et al. [41] proposed a multi-task learning framework to jointly extract three elements of a triplet. Xu et al. [8] focused on the multi-word aspect and opinion terms and proposed a span-level approach. Besides, they developed a span pruning strategy for reducing the computation cost.

Yan et al. [42] investigated the ASTE task by the generation approach and proposed a unified generative framework. They utilized the pointer indexes and sentiment class indexes to extract the aspect term, opinion term, and the corresponding sentiment polarities. To model relations between words in

ASTE, Chen et al. [12] proposed an enhanced multi-channel graph convolutional network (EMC-GCN) model. They also employed four linguistic features to improve the performance of the EMC-GCN model. All these methods have achieved great success. However, these methods ignore modeling the syntactic dependency information through the general neural networks, such as Bi-LSTM. And the GCN-based models are weak at capturing the sequence information and long-distance dependency information. Besides, the dependency trees generated by the external dependency parsers can have great influence on the performance of the GCN-based models.

### III. METHODOLOGY

In this section, we describe detailedly the architecture of the proposed IDCN model.

#### A. TASK DEFINITION

Given a sentence  $S = \{w_1, w_2, \dots, w_n\}$ ,  $n$  is the total number of all words in a sentence, the goal of the ASTE task is to extract all triplets  $T = \{(a_1, o_1, s_1), \dots, (a_m, o_m, s_m)\}$  from a specific sentence  $S$ ,  $m$  is the total number of all triplets in a specific sentence  $S$ ,  $(a_i, o_i, s_i)$  denotes the aspect term, opinion term, and the corresponding sentiment polarity of the  $i$ -th triplet, separately. The sentiment polarity  $s$  of the aspect term  $a$  has three labels, they are *positive*, *negative*, and *neutral*. And the sentiment polarity belongs to one of three labels.

#### B. OVERVIEW OF IDCN

In order to capture the correlation relations between word pairs, the sequence information, and long-distance dependency information in a sentence, to alleviate the wrong and noise information that is introduced by the off-the-shelf dependency parser, and to model the syntax dependency information using the general neural networks, we proposed an end-to-end framework, named interactive dual channel network (IDCN). The overall framework of the proposed IDCN is illustrated in Fig. 2.

In IDCN, a novel interactive word pair generating (IWPG) module is designed to learn the informative word pair representation and model the correlation relations between word pairs. In the IWPG module, the feature learning layer is used to capture the sequence information and long-distance dependency information, as well as to model the syntactic dependency information from word pair dependency relation tensors. And the interaction layer is used to capture the comprehensive correlation relation between two words and output the informative word pair representation. The syntactic dependency fusion module is used to inject the syntactic dependency information into the model and generate the syntactic dependency-aware word embeddings from the constructed tensors. The embedding layer is employed to generate the contextualized word embeddings, it can provide general knowledge information for the model. In general, the pre-trained language model such as BERT is trained on the large-scale general data and can learn the inherent knowledge from the large-scale general data.

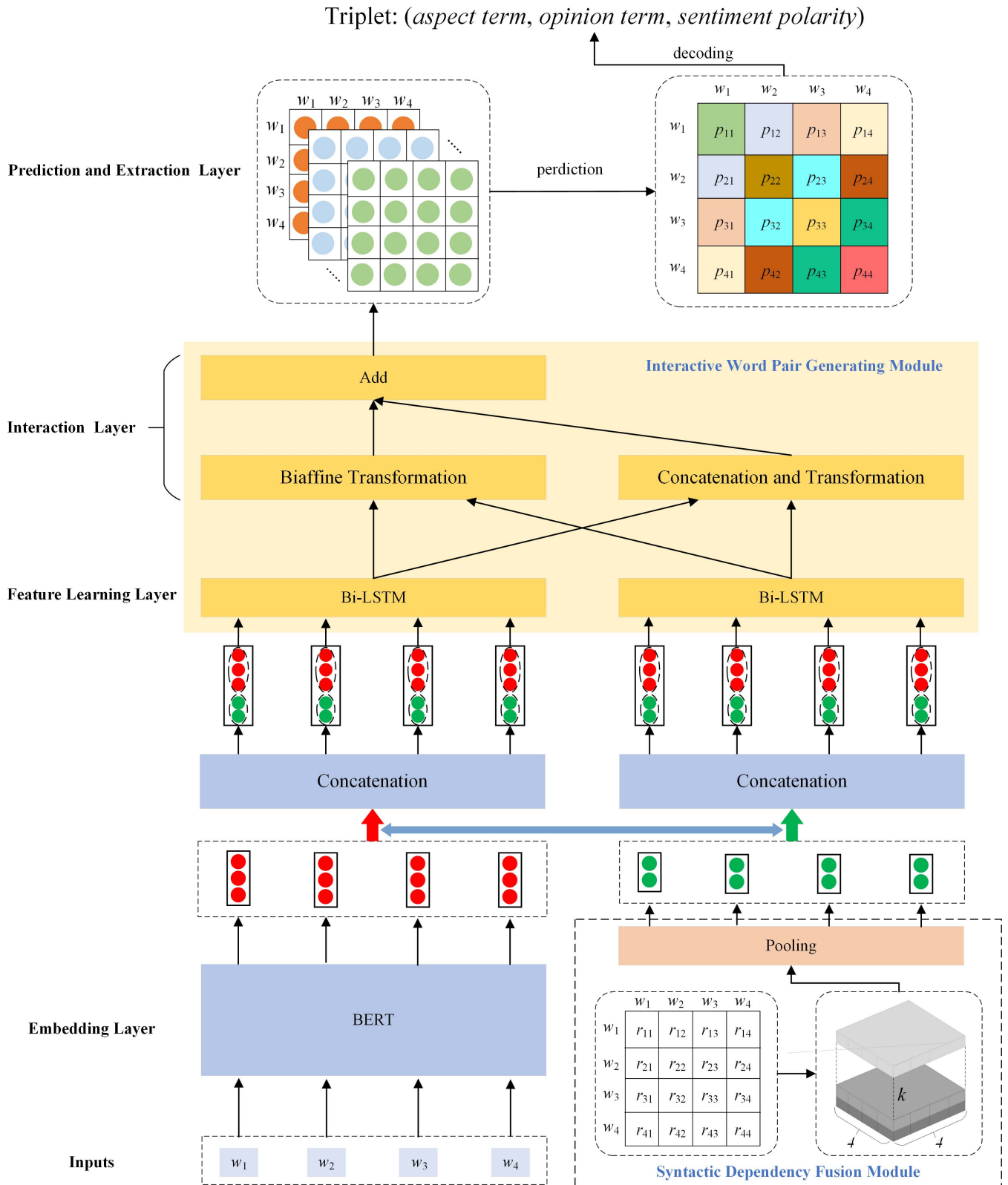


FIGURE 2. The overall framework of the proposed IDCN model.

### C. EMBEDDING LAYER

The embedding layer consists of BERT, the pre-trained BERT model can generate the contextual word embeddings and

provide a lot of general knowledge for the model. The contextual word embeddings can capture the more precise meaning of a word than the fixed word embedding according

to the dynamic context of a specific sentence. The general knowledge information can enhance the extraction ability of the model to some extent. The calculation formula is as following:

$$h_i = BERT(w_i) \quad (1)$$

where  $w_i$  is the  $i$ -th input of a sentence  $S$ ,  $h_i \in \mathbb{R}^{n \times d_{bert}}$  is the  $i$ -th output of BERT,  $n$  is the total length of a sentence  $S$ ,  $d_{bert}$  is the dimension of hidden contextual representations in BERT. The symbol  $BERT(\cdot)$  denotes all operations of the BERT model.

#### D. SYNTACTIC DEPENDENCY FUSION MODULE

The syntactic dependency fusion module is designed to inject the syntactic information into the model and generate syntactic dependency-aware word embeddings.

Firstly, we transform a specific sentence into a word pair dependency matrix. Each element of the word-pair dependency matrix denotes the syntactic dependency relation between two words, such as *nsubj* and *compound*. Then, the word pair dependency relation tensors can be obtained by the lookup table operation. The calculation processes can be obtained as following:

$$r_{ij}^D = lookup(r_{ij}) = \hat{r}_{ij} \cdot E \quad (2)$$

where  $\hat{r}_{ij} \in \mathbb{R}^k$  is the one-hot vector of the syntactic dependency relation  $r_{ij}$  between word  $i$  and word  $j$ ,  $k$  is the total number of syntactic dependency relations.  $E \in \mathbb{R}^{k \times d_r}$  is the embedding matrix and is the learning parameter of the model. Then the word pair dependency relation tensors  $R \in \mathbb{R}^{n \times n \times d_r}$  can be obtained,  $R_{i,j,:} = r_{ij}^D$ .

After we get the word pair dependency relation tensors  $R$ , we can obtain the syntactic dependency-aware word embeddings by the pooling mechanism such as average pooling and max-pooling, or the complicated networks such as attention-based methods. In other words, the syntax dependency information can be learned by utilizing the word pair dependency relation tensors and the pooling mechanism or complicated networks in IDCN. The proposed IDCN does not operate on the dependency tree of a sentence. Hence, it can avoid error propagation from the dependency tree.

Besides, we have performed experiments on the average pooling mechanism and the max-pooling mechanism. We found that the average pooling mechanism performs better than the max-pooling mechanism in the syntactic dependency fusion module. To further improve the performance of the model and simplify the model, we use the average pooling mechanism to obtain the syntactic dependency-aware word embeddings. The average pooling mechanism can be viewed as an aggregation of syntactic relations between the current word and its syntactically adjacent words.

The calculation processes can be obtained as following:

$$h_i^d = average\_pooling(R_{i,:,:}) \quad (3)$$

$$H^D = [h_1^d; h_2^d; \dots; h_n^d] \quad (4)$$

where  $R_{i,:,:} \in \mathbb{R}^{n \times d_r}$  denotes the  $i$ -th slice of  $R$ , the symbol  $average\_pooling(\cdot)$  denotes the average pooling operation along the first dimension,  $h_i^d \in \mathbb{R}^{d_r}$  is the syntactic dependency-aware word embedding of the word  $w_i$ ,  $H^D \in \mathbb{R}^{n \times d_r}$  denotes the embedding matrix of the specific sentence  $S$ , the symbol “;” denotes the concatenation of vectors along the first dimension. By the pooling mechanism, word pair dependency relation tensors can be easily converted to syntactic dependency-aware word embeddings. These syntactic dependency-aware word embeddings can be naturally utilized and modeled by the general neural network models such as Bi-LSTM. And the syntactic dependency information can be injected into the model by the syntactic dependency fusion module.

#### E. IWPG MODULE

The IWPG module consists of the feature learning layer and the interaction layer. In the following sections, we discussed them in detail.

##### 1) FEATURE LEARNING LAYER

The feature learning layer is used to model the sequence information, long-distance dependency information, and syntactic dependency information. The feature learning layer consists of dual channels, and each channel can consist of any type of networks such as Bi-LSTM, CNN, and transformer. The Bi-LSTM is better than CNN and transformer at modeling the sequence information and long-distance information, hence we employ Bi-LSTM in each channel of the feature learning layer. Besides, we have performed experiments on dual Bi-LSTMs with shared parameters and dual Bi-LSTMs without shared parameters. We found that the dual Bi-LSTMs with shared parameters perform better than the dual Bi-LSTMs without shared parameters in the left and right channels of the feature learning layer. To further improve the performance of IDCN and reduce the parameters of IDCN, the dual Bi-LSTMs are set to share parameters in the left and right channels.

For capturing the sequence information, long-distance dependency information, and syntactic dependency information simultaneously, the contextual word embeddings generated by BERT and syntactic dependency-aware word embeddings generated by the syntactic dependency fusion module are concatenated as the inputs of the feature learning layer.

The calculation processes can be obtained as following:

$$h_i^l = BiLSTM_l([h_i; h_i^d]) \quad (5)$$

$$h_i^r = BiLSTM_r([h_i; h_i^d]) \quad (6)$$

where the symbol  $BiLSTM(\cdot)$  denotes all operations of two Bi-LSTMs in the left and right channels, respectively. The dual Bi-LSTMs share parameters in the left and right channels. The symbol “;” denotes the concatenation of vectors along the second dimension.  $h_i^l \in \mathbb{R}^{d_f}$  and  $h_i^r \in \mathbb{R}^{d_f}$  denote the  $i$ -th outputs of the left and right channels, respectively.

## 2) INTERACTION LAYER

The interaction layer is designed to capture the association relation between two words and generate the informative word pair representation. The association relation between two words in a sentence plays an important role in the ASTE task. For example, in the sentence of Fig. 1, the word “chicken” and word “tikka”, word “chicken” and word “marsala”, word “tikka” and word “marsala” belong to the same aspect term “chicken tikka marsala”. To effectively detect the aspect term “chicken tikka marsala”, the words “tikka” and “marsala” can provide effective information for the word “chicken” and vice versa. The opinion term “good” can provide the sentiment information for the aspect term “chicken tikka marsala”. In a word, the informative word pair representations are important for the ASTE task. The calculation processes of the interaction layer can be obtained as following:

$$h_{ij}^{lm} = h_i^l \cdot W_1 \cdot h_j^r + b_1 \quad (7)$$

$$h_{ij}^m = W_2 \cdot [h_i^l; h_j^r] \quad (8)$$

$$h_{ij}^a = h_{i,j}^{lm} + h_{i,j}^m \quad (9)$$

where  $h_{ij}^a \in \mathbb{R}^{d_a}$  is the word pair representation of word  $w_i$  and word  $w_j$  and reflects relations between word  $w_i$  and word  $w_j$ ,  $d_a$  is a hyper-parameter and denotes the dimension of the word pair representation,  $W_1 \in \mathbb{R}^{d_f \times d_a \times d_f}$ ,  $b_1 \in \mathbb{R}^{d_a}$ , and  $W_2 \in \mathbb{R}^{d_a \times (d_f + d_f)}$  are the learning parameters of the model for the left and right channels in the interaction layer, respectively. The symbol “;” denotes the concatenation of vectors.

The biaffine transformation is the left channel in the interaction layer, and the concatenation and transformation is the right channel in the interaction layer. The dual channels in the interaction layer can capture and model the more comprehensive association relation between two words. The addition of two channels can generate more informative word pair representations.

## F. PREDICTION AND EXTRACTION LAYER

After we get the word pair representation, the word pair representation  $h_{ij}^a$  is fed into the prediction and extraction layer, and the calculation formula of label distribution can be obtained as following:

$$p_{ij} = \text{softmax}(h_{ij}^a) \quad (10)$$

where the symbol  $\text{softmax}(\cdot)$  denotes the softmax function,  $p_{ij} \in \mathbb{R}^q$  is the label distribution of the word pair, and  $q$  is the total number of classes of output labels.

Following Chen et al. [12], we design ten kinds of word pair labels: *B-AT*, *I-AT*, *A-WP*, *B-OT*, *I-OT*, *O-WP*, *P-S*, *N-S*, *NEU-S*, and *None*. The label *B-AT* denotes any word in the word pair is the beginning of the aspect term. The label *I-AT* denotes any word in the word pair is the inside of the aspect term. The label *A-WP* denotes that the word pair forms the aspect term. The label *B-OT* denotes any word in the word pair is the beginning of the opinion term. The label *I-OT*

denotes any word in the word pair is the inside of the opinion term. The label *O-WP* denotes that the word pair forms the opinion term. The label *P-S*, *N-S*, and *NEU-S* denote that a word in the word pair is an aspect term and another word in the word pair is an opinion term. And they also denote that aspect-opinion term pair with positive, negative, and neutral sentiment polarity, respectively. The label *None* denotes that the relations between word pairs do not belong to the above relations.

By the argmax function on the label distribution, we can get the specific label of the word pair. Then the triplets can be obtained from the predicted results of a sentence. Firstly, we can obtain the candidate aspect terms and opinion terms from the predicted results. Specifically, the aspect terms and opinion terms can be identified by the *B-AT*, *I-AT*, *B-OT*, and *I-OT* labels along the diagonal of the predicted word pair table. The predicted word pair table can be seen in the upper right of Fig. 2. After we get all aspect terms and opinion terms, we utilize the lower triangular table to extract the triplets. We can choose an aspect term and an opinion term from the extracted aspect and opinion terms until all items are selected in the extracted aspect and opinion terms. If an aspect term and an opinion term have any sentiment relation, then they form the aspect-opinion pair. According to the sentiment relation of the aspect-opinion pair, we can obtain the most predicted sentiment relations. Finally, we can acquire a triplet (*aspect term*, *opinion term*, *sentiment polarity*).

## G. LOSS FUNCTION

Our goal is to minimize the loss function. The calculation processes can be obtained as following:

$$\mathcal{L} = \mathcal{L}_w + \alpha \mathcal{L}_d \quad (11)$$

$$\mathcal{L}_w = \sum_i^n \sum_j^n y_{ij} \log p_{ij} \quad (12)$$

$$\mathcal{L}_d = \sum_i^n \sum_j^n y_{ij} \log r_{ij}^D \quad (13)$$

where  $\mathcal{L}_w$  is the standard loss function for the aspect sentiment triplet extraction,  $\mathcal{L}_d$  is the constraint loss for word pair dependency relation tensors,  $\alpha$  is the hyper-parameter,  $y_{ij}$  is the true probability distribution of the word pairs.

## IV. EXPERIMENTS

### A. DATASET

We use two typical datasets for training and testing of the model in experiments, these two typical datasets are all from the SemEval 2014 [43], 2015 [44], and 2016 [45]. The first dataset is collected by Wu et al. [9]. The second dataset is collected by Xu et al. [7], and they corrected some errors based on the dataset proposed by Peng et al. [5]. Both datasets involve the restaurant and laptop domains. The statistics of both datasets are shown in Table 1.

**TABLE 1.** Statistics of both datasets. “D<sub>1</sub>” and “D<sub>2</sub>” respectively denote the first and the second dataset. “#Sen” and “#Tri” respectively denote the numbers of sentences and aspect sentiment triplets. “Res” and “Lap” respectively denote the restaurant and laptop datasets.

Dataset		14Res		14Lap		15Res		16Res	
		#Sen	#Tri	#Sen	#Tri	#Sen	#Tri	#Sen	#Tri
D <sub>1</sub>	Train	1259	2356	899	1452	603	1038	863	1421
	Dev	315	580	225	383	151	239	216	348
	Test	493	1008	332	547	325	493	328	525
D <sub>2</sub>	Train	1266	2338	906	1460	605	1013	857	1394
	Dev	310	577	219	346	148	249	210	339
	Test	492	994	328	543	322	485	326	514

## B. EVALUATION METRICS

We use precision, recall, and F1 as the evaluation metrics, which are often used to evaluate the performance of the model in the ASTE task. The calculation formulas of precision, recall, and F1 are given as following:

$$P = \frac{TP}{TP + FP} \quad (14)$$

$$R = \frac{TP}{TP + FN} \quad (15)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (16)$$

where  $P$  and  $R$  respectively denote the precision and recall.  $TP$  means the true positives,  $FP$  means the false positives, and  $FN$  means the false negatives.

## C. EXPERIMENT SETTINGS

We use Stanza [46] to obtain the syntactic dependency information in a sentence. The basic version of pre-trained BERT implemented by Transformers from Huggingface [47] is employed in the proposed IDCN. The AdamW optimizer is employed as the default optimizer. The learning rate for the BERT is set to  $2e-5$ , and the learning rate for other parameters in IDCN is set to  $1e-3$ . The dropout strategy is used to prevent model overfitting and the value is set to 0.5. The dimensionality of Bi-LSTM is 300. The dimensionality of syntactic dependency-aware word embeddings is the same as the number of word pair relations. The hyper-parameter of loss function is set to 0.01 in D<sub>2</sub> and 14Lap of D<sub>1</sub> datasets, and it is set to 0.1 in other datasets of D<sub>1</sub>. The epoch number is set to 100 and the batch size is 16.

## D. BASELINES

In the experiments, we compare the proposed IDCN with the variants of IDCN and the state-of-the-art baselines.

### 1) VARIANTS OF IDCN

The variants of the proposed IDCN are used to validate the effectiveness of the designed modules or layers in IDCN. The IDCN-v1, -v2, and -v3 are designed to verify the effectiveness of the IWPG module. The IDCN-v4 and -v5 are used to verify the effectiveness of the syntactic dependency fusion module.

#### IDCN-v1

1. : This is the first variant of IDCN. It removes the IWPG module in IDCN. The concatenation of the con-

textualized word embeddings generated by BERT and the dependency-aware word tensors generated by the syntactic dependency fusion module is directly fed into the prediction and extraction layer. The concatenation tensors are directly used to predict the labels of word pairs.

#### IDCN-v2

2. : This is the second variant of IDCN. It removes the interaction layer in the IWPG module. The outputs of the dual Bi-LSTMs are concatenated with the dependency-aware word tensors, and then the dual concatenation tensors are added as inputs to the prediction and extraction layer.

#### IDCN-v3

3. : This is the third variant of IDCN. It replaces the dual Bi-LSTMs with the dual transformers. The outputs of the dual transformers are fed into the interaction layer. The dual transformers also share the same parameters.

#### IDCN-v4

4. : This is the fourth variant of IDCN. It replaces the average pooling with the max pooling mechanism in the syntactic dependency fusion module. The outputs of the max pooling operation are viewed as the outputs of the syntactic dependency fusion module.

#### IDCN-v5

5. : This is the fifth variant of IDCN. It discards the syntactic dependency fusion module in the IDCN. In other words, the inputs of the IWPG module in IDCN-v5 are only the contextualized word embeddings generated by BERT, not involving the syntactic dependency information.

### 2) STATE-OF-THE-ART METHODS

The state-of-the-art methods are used to compare with the proposed IDCN model. In general, there are three kinds of methods in the ASTE task, which are pipeline-based methods, end-to-end methods, and machine reading comprehension-based methods. The pipeline-based methods include CMLA+, RINANTE+, Li-unified-R, Peng-two-stage, Peng+IOG, and IMN+IOG. The end-to-end methods include GTS-CNN, GTS-BiLSTM, GTS-BERT, OTE-MTL, JET-BERT, S<sup>3</sup>E<sup>2</sup>, and EMC-GCN. The machine reading comprehension-based methods include BMRC.

1. **CMLA+**: This model employs the attention mechanism to model dependency relations between words.



- In the first stage, the model jointly extracts the spans of aspect terms, opinion terms, and sentiment polarity. In the second stage, the MLP classifier is used to determine whether the candidate triplet can be viewed as the predicted results or not [5].
2. **RINANTE+**: This model employs LSTM-CRF to jointly extract aspect terms, opinion terms, and the corresponding sentiment. In the first stage, they utilized rules as weak supervision to model dependency relations between words. In the second stage, it also employs the MLP to output all the valid triplets as CMLA+ [5].
  3. **Li-unified-R**: This model designs a customized multi-layer LSTM to respectively identify aspect terms with the corresponding sentiment polarity and the opinion terms. In the second stage, it also employs the MLP to output all the valid triplets as CMLA+ [5].
  4. **Peng-two-stage**: This model extracts the aspect terms with the corresponding sentiment and opinion spans, simultaneously. The GCN is used to model the dependency relations between words for enhancing the performance of the model. In the second stage, it also employs the MLP to output all the valid triplets as CMLA+ [5].
  5. **Peng+IOG**: This model combines Peng-two-stage and the IOG model [48] for the ASTE task [9].
  6. **IMN+IOG**: This model combines the IMG model [49] and the IOG model [48] for the ASTE task [9].
  7. **GTS-CNN**: This model views the ASTE task as a unified tagging task, and designs a novel grid tagging scheme for the ASTE task in an end-to-end fashion instead of a pipeline fashion. Specifically, GTS-CNN employs the CNN model to extract the triplets [9].
  8. **GTS-BiLSTM**: This model is also based on the grid tagging scheme presented in GTS-CNN. GTS-BiLSTM employs Bi-LSTM to extract the triplets [9].
  9. **GTS-BERT**: This model is also based on the grid tagging scheme presented in GTS-CNN. GTS-BERT employs BERT to extract the triplets [9].
  10. **OTE-MTL**: This model is a multi-task learning framework for the opinion triplet extraction task, which jointly detects aspects, opinions, and sentiment dependencies. It employs Bi-LSTM for sentence encoding [41].
  11. **JET-BERT**: This model is based on a position-aware tagging scheme for jointly extracting the aspect sentiment triplets, which can specify the structural information for a triplet. It develops factorized features to effectively capture the interactions between elements in a triplet [7].
  12. **S<sup>3</sup>E<sup>2</sup>**: This model is a semantic and syntactic enhanced aspect sentiment triplet extraction. Specifically, it designs a graph representation for integrating the syntactic dependency information and semantic information. Besides, the LSTM is used to capture the contextual semantics [50].
  13. **EMC-GCN**: This model is the current state-of-the-art framework for ASTE. It employs GCN to model the word relations between words. Besides, the part-of-speech information, syntactic dependency information, tree-based position information, and relative position information are used to enhance the model performance [12].
  14. **BMRC**: This model is a multi-turn machine reading comprehension framework. It develops three types of queries for capturing the correlation relation between different subtasks. Two directions in the model can benefit each other and generate the more precise triplets [6].

## E. OVERALL RESULTS

In this section, we report all experimental results and discuss the reasons why the proposed IDCN model can obtain better results than the baselines.

### 1) RESULTS OF VARIANTS

The ablation experiments are used to verify the effectiveness of the important modules or layers of the proposed IDCN. The experiment results of all variants of IDCN on dataset  $D_1$  are shown in Table 2. And the experiment results of all variants of IDCN on the dataset  $D_2$  are shown in Table 3.

We can observe from the experimental results of Table 2 and Table 3 that the proposed IDCN acquires the best precision, recall, and F1 values on all datasets.

The difference between IDCN-v1 and IDCN is that the IDCN-v1 discards the overall IWPG module. The IDCN-v1 model lacks the ability to model sequence information and long-distance information, as well as to capture the correlation relation between word pairs simultaneously. The above information and relation are crucial for the ASTE task. Hence, the absence of the IWPG module leads to the worst performance of the IDCN-v1 model on all datasets. The experimental results of IDCN-v1 verify the effectiveness of the designed IWPG module in IDCN.

There are two layers in the IWPG module, which are the interaction layer and the feature learning layer. The IDCN-v2 and IDCN-v3 are used to verify the effectiveness of them. The difference between IDCN-v2 and IDCN is that IDCN-v2 does not consider the interaction mechanism of dual channels and does not capture the correlation relation between word pairs. Compared with IDCN, the performance degradation of IDCN-v2 validate the importance of the interaction layer. However, only modeling the correlation relation between word pairs cannot guarantee the performance of the model. This is demonstrated by the experimental results of IDCN-v3 on all datasets. The IDCN-v3 model ignores modeling the sequence information. Besides, the experimental results of IDCN-v3 also show that Bi-LSTM is more appreciated than transformers in the feature learning layer of the proposed IDCN model.

The difference between IDCN-v4 and IDCN is that IDCN-v4 uses the max pooling mechanism to obtain the

**TABLE 2.** The experimental results of variants on the dataset  $D_1$ . The bold indicates the best results.

Method	14Res			14Lap			15Res			16Res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
IDCN-v1	72.27	57.03	63.75	69.54	41.47	51.95	67.79	49.09	56.94	74.94	57.74	65.23
IDCN-v2	78.62	53.51	63.68	67.94	39.27	49.77	68.44	47.06	55.77	75.77	56.21	64.54
IDCN-v3	71.66	70.08	70.86	53.3	55.96	54.61	62.28	56.59	59.30	65.53	69.79	67.59
IDCN-v4	70.40	71.89	71.14	63.43	51.56	56.88	63.27	56.42	59.65	64.93	71.51	68.06
IDCN-v5	72.38	70.78	71.57	61.28	54.31	57.90	59.35	59.23	59.29	68.22	67.30	67.76
IDCN	72.34	70.88	<b>71.60</b>	60.43	56.33	<b>58.31</b>	62.34	57.40	<b>59.77</b>	68.65	68.26	<b>68.46</b>

**TABLE 3.** The experimental results of variants on the dataset  $D_2$ . The bold indicates the best results.

Method	14Res			14Lap			15Res			16Res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
IDCN-v1	78.07	55.04	64.56	67.65	42.51	52.21	67.14	49.28	56.84	70.71	60.23	65.05
IDCN-v2	76.83	55.65	64.54	68.00	40.85	51.04	69.59	49.07	57.56	69.12	58.48	63.36
IDCN-v3	68.38	70.40	69.37	62.64	53.60	57.77	61.85	59.18	60.49	64.39	68.03	66.16
IDCN-v4	73.49	67.96	70.61	58.94	53.60	56.15	62.45	58.97	60.66	65.85	68.03	66.92
IDCN-v5	69.83	75.59	72.59	60.92	56.19	58.46	62.49	60.21	61.33	68.54	68.81	68.68
IDCN	72.07	73.25	<b>72.65</b>	65.97	52.68	<b>58.58</b>	62.26	61.24	<b>61.75</b>	66.61	72.32	<b>69.35</b>

**TABLE 4.** The experimental results of state-of-the-art methods on the dataset  $D_1$ . The bold indicates the best results.

Method	14Res			14Lap			15Res			16Res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Peng+IOG	58.89	60.41	59.64	48.62	45.52	47.02	51.70	46.04	48.71	59.25	58.09	58.67
IMN+IOG	59.57	63.88	61.65	49.21	46.23	47.68	55.24	52.33	53.75	-	-	-
GTS-CNN	70.79	61.71	65.94	55.93	47.52	51.38	60.09	53.57	56.64	62.63	66.98	64.73
GTS-BiLSTM	67.28	61.91	64.49	59.42	45.13	51.30	63.26	50.71	56.29	66.07	65.05	65.56
S <sup>3</sup> E <sup>2</sup>	69.08	64.55	66.74	59.43	46.23	52.01	61.06	56.44	58.66	71.08	63.13	66.87
GTS-BERT	70.92	69.49	70.20	57.52	51.92	54.58	59.29	58.07	58.67	68.58	66.60	67.58
BMRC	-	-	70.01	-	-	57.83	-	-	58.74	-	-	67.49
EMC-GCN	74.00	72.29	<b>73.13</b>	59.80	54.86	57.23	56.57	62.88	59.56	66.41	66.16	66.28
IDCN	72.34	70.88	71.60	60.43	56.33	<b>58.31</b>	62.34	57.40	<b>59.77</b>	68.65	68.26	<b>68.46</b>

syntactic dependency-aware word embeddings in the syntactic dependency fusion module. The experimental results of IDCN-v4 validate the effectiveness of the average pooling mechanism in IDCN.

The syntactic dependency information can improve the performance of the model for the ASTE task. The syntactic dependency relation can provide more informative information for the model to extract the aspect sentiment triplet. The difference between IDCN-v5 and IDCN is that IDCN-v5 discards the syntactic dependency fusion module and ignores the syntactic dependency information. The performance degradation on all datasets verifies the effectiveness of the syntactic dependency fusion module. Besides, the experimental results also demonstrate that syntactic dependency relation can enhance the expression ability of the model to some extent.

## 2) RESULTS OF STATE-OF-THE-ART METHODS

The state-of-the-art methods are used to compare with the proposed IDCN and validate the superior performance of the proposed IDCN. We reported the experimental results of EMC-GCN by running their original codes, and the experimental results of the other state-of-the-art methods are reported from their original papers or Chen et al. [12]. All experimental results of state-of-the-art methods on datasets  $D_1$  and  $D_2$  are reported in Table 4 and Table 5, respec-

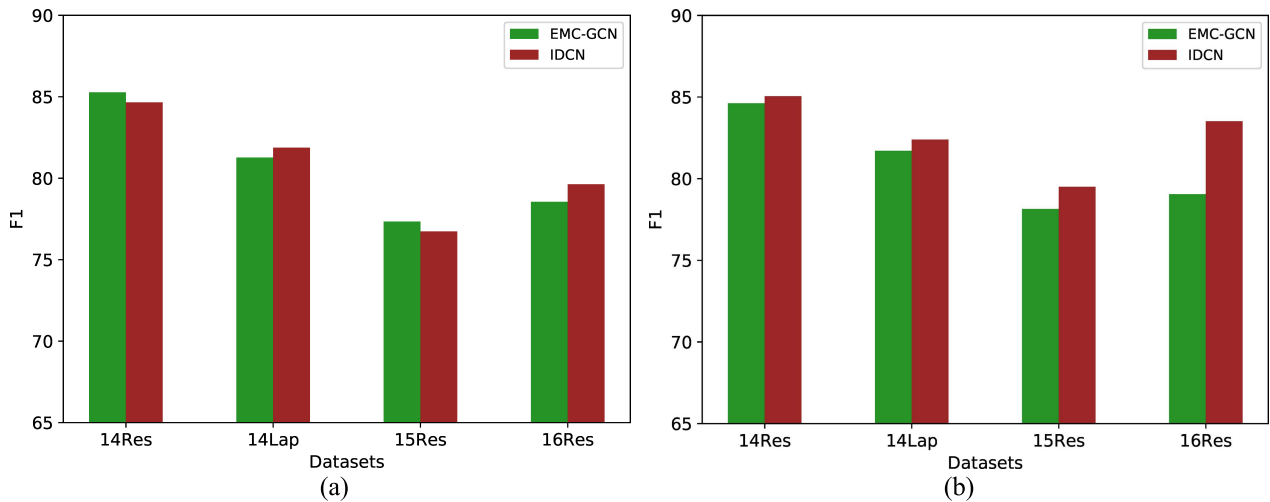
tively. In Table 4 and Table 5, the symbol ‘-’ denotes that the experimental results are not reported in their original paper.

As shown in Table 4, the proposed IDCN outperforms state-of-the-art models on most of the datasets, it surpasses state-of-the-art models by 0.21%-0.88% F1 values. As seen shown in Table 5, the proposed IDCN achieves the best performance on all datasets, it surpasses state-of-the-art models by 0.56%-1.77% F1 values.

The first reason that the proposed IDCN performs better than other state-of-the-art methods is that the IWPG module in the proposed IDCN can model the sequence information, long-distance information, and the association relation between word pairs, simultaneously. Other state-of-the-art methods lack the ability to capture multiple types of information and relations. The sequence information, long-distance information, and association relation between word pairs are important for the ASTE task. In the feature learning layer of the IWPG module, it can capture the sequence and long-distance information by employing the Bi-LSTMs. The dual channels in the IWPG module can learn different informative representations. By the interaction of dual channels, the interaction layer can capture the association relation between word pairs and generate informative word-pair representations.

**TABLE 5.** The experimental results of state-of-the-art methods on the dataset  $D_2$ . The bold indicates the best results.

Method	14Res			14Lap			15Res			16Res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
CMLA+	39.18	47.13	42.79	30.09	36.92	33.16	34.56	39.84	37.01	41.34	42.10	41.72
RINANTE+	31.42	39.38	34.95	21.71	18.66	20.07	29.88	30.06	29.97	25.68	22.30	23.87
Li-unified-R	41.04	67.35	51.00	40.56	44.28	42.34	44.72	51.39	47.82	37.33	54.51	44.31
Peng-two-stage	43.24	63.66	51.46	37.38	50.38	42.87	48.07	57.51	52.32	46.96	64.24	54.21
OTE-MTL	62.00	55.97	58.71	49.53	39.22	43.42	56.37	40.94	47.13	62.88	52.10	56.96
JET-BERT	70.56	55.94	62.40	55.39	47.33	51.04	64.45	51.96	57.53	70.42	58.37	63.83
GTS-BERT	68.09	69.54	68.81	59.40	51.94	55.42	59.28	57.93	58.60	68.32	66.86	67.58
BMRC	75.61	61.77	67.99	70.55	48.98	57.82	68.51	53.40	60.02	71.20	61.08	65.75
EMC-GCN	69.50	74.87	72.09	59.49	56.19	57.80	57.91	64.12	60.86	60.33	71.15	65.30
IDCN	72.07	73.25	<b>72.65</b>	65.97	52.68	<b>58.58</b>	62.26	61.24	<b>61.75</b>	66.61	72.32	<b>69.35</b>

**FIGURE 3.** The F1 values of different methods on the ATE task, (a) reports the experimental results on the dataset  $D_1$ , (b) reports the experimental results on the dataset  $D_2$ .

The second reason is that the syntactic dependency fusion module can capture the syntactic dependency information between word pairs and inject the informative syntactic dependency information into the general neural network such as Bi-LSTM. In general, the syntactic dependency information is often modeled by GCNs, which are weak at capturing sequence information and long-distance information. In general neural networks, the syntactic dependency information is difficult to combine with the model. In the syntactic dependency fusion module, we construct the word pair dependency relation tensors from the syntactic dependency information between word pairs and then utilize the pooling mechanism to obtain the syntactic dependency-aware word embeddings. In this way, the syntactic dependency fusion module can naturally inject the informative syntactic dependency information into the general network works. Besides, the syntactic dependency information can also be captured and learned in the syntactic dependency fusion module.

The third reason is that the dual architecture of the proposed IDCN model can effectively learn different representations and integrate these different representations to generate the informative word-pair dependent representations. In other words, the dual architecture can provide more useful and

comprehensive information for extracting the aspect sentiment triplets.

In order to further validate the generalization ability of the proposed IDCN, we have performed additional experiments on the aspect term extraction (ATE) task. The compared state-of-the-art method is EMC-GCN. The experimental results on all datasets are reported in Fig. 3. In Fig. 3, (a) and (b) report the F1 values of EMC-GCN and IDCN on the datasets  $D_1$  and  $D_2$ , respectively. As seen in Fig. 3, the proposed IDCN acquires the best F1 values on most datasets. The experimental results demonstrate that the proposed IDCN is not only good at the ASTE task, but also good at the ATE task. Besides, the experimental results also verify the effectiveness of the proposed IDCN.

### 3) EFFECT OF DIMENSION OF FEATURE LEARNING LAYER

We have performed additional experiments to investigate the effect of different dimensions of the feature learning layer of IDCN. Specifically, we employ Bi-LSTMs in the feature learning layer. The performances of different dimensions of feature learning layer are illustrated in Fig. 4. The specific dimensions are 100, 200, 300, and 400, separately. In Fig. 4, (a) reports the F1 values on 14Res and 14Lap of the dataset  $D_1$ , and (b) reports the F1 values on 14Res and 14Lap of the

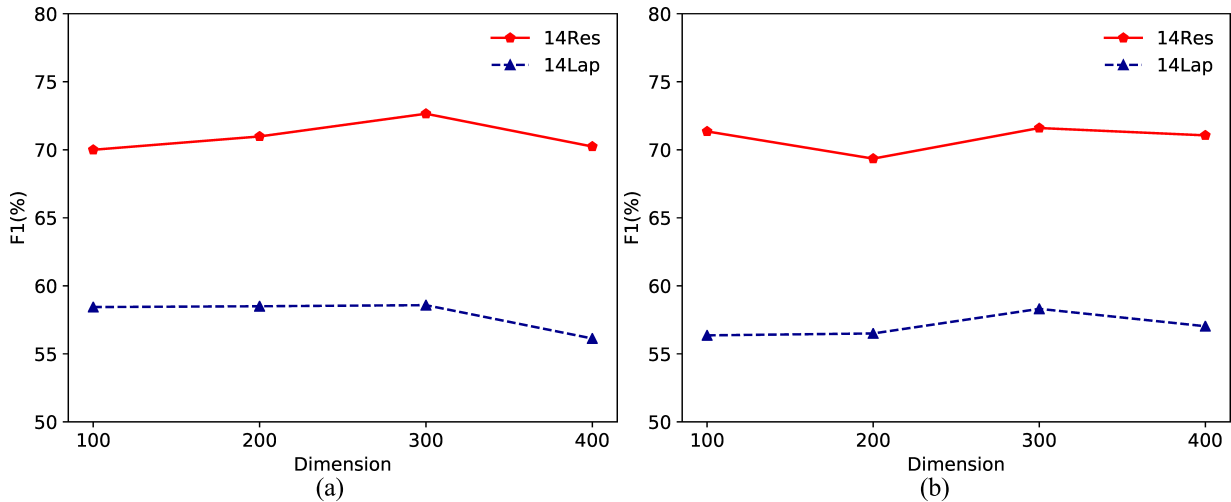


FIGURE 4. The F1 values of different dimensions, (a) reports the experimental results on 14Res and 14Lap of the dataset D<sub>1</sub>, (b) reports the experimental results on 14Res and 14Lap of the dataset D<sub>2</sub>.

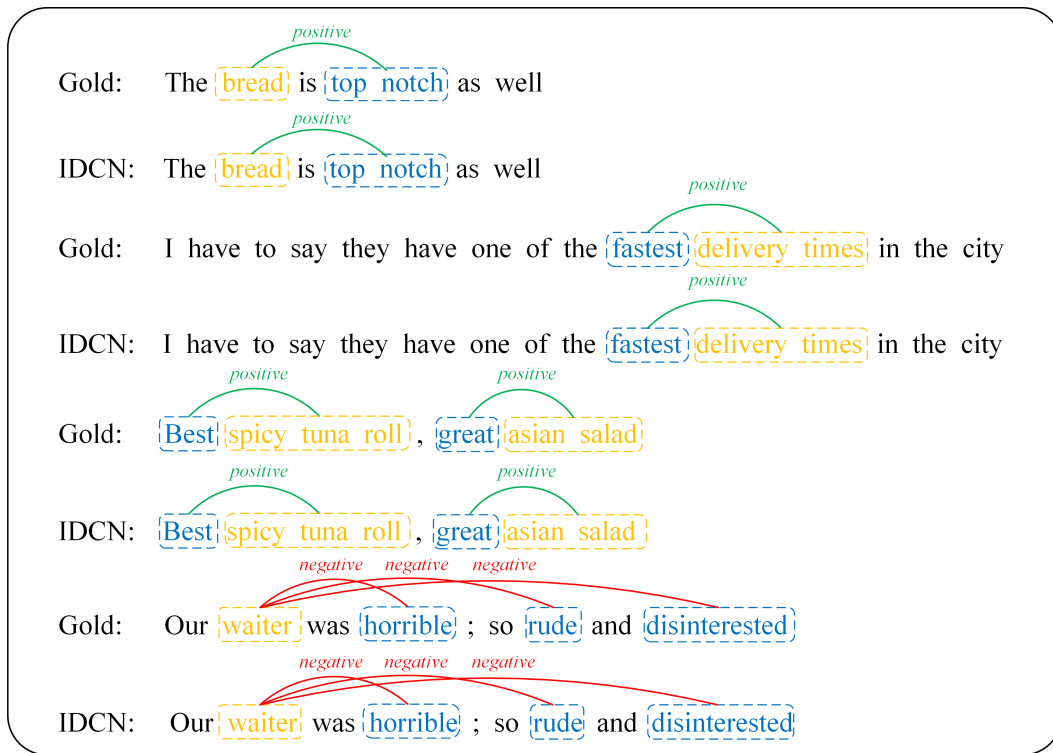


FIGURE 5. The predicted results of IDCN for some sentences.

dataset D<sub>2</sub>. We can observe that the IDCN model acquires the best F1 values when the dimension is 300 on all datasets. When the dimension is less than 300, the performance of the model increases with the increase of dimensions. When the dimension is greater than 300, the performance of the model decreases with the increase of dimensions.

#### 4) CASE STUDY

We report some predicted results of IDCN in Fig. 5. In Fig. 5, the aspect terms are marked in orange, the opinion terms are marked in blue, and the green solid line denotes that the aspect term and the opinion term belong to the same

triplet and the corresponding sentiment polarity is positive, the red solid line denotes that the sentiment polarity of a triplet is negative, the symbol “Gold” denotes the gold true triplets in a sentence, and “IDCN” denotes the predicted results of the proposed IDCN. As seen in Fig. 5, the IDCN model can extract multiple complicated triplets in a sentence. The predicted results further verify the effectiveness of the proposed IDCN.

#### V. CONCLUSION

In this paper, we propose a novel interactive dual channel network (IDCN) model for the ASTE task. In the

IDCN model, the IWPG module is designed to capture the sequence information, long-distance dependency information, and association relations between word pairs, simultaneously. The dual channels can effectively learn different representations. Based on these representations, the informative word pair representation can be obtained by the designed interaction mechanism of dual channels. In order to naturally inject the syntactic dependency information into the general neural networks and alleviate error propagation, the syntactic dependency fusion module is developed. It can model the syntactic relations and provide the syntactic dependency information for the IWPG module by constructing the word pair dependency relation tensors and average pooling. Multiple datasets are used to verify the effectiveness and generalization ability of the proposed IDCN model. The experimental results show that the proposed IDCN model acquires state-of-the-art results. In the next step, we plan to apply the proposed IDCN model to other related tasks. Besides, incorporating multi-hop reasoning with the proposed IDCN model can further increase the interpretability of the prediction results.

## REFERENCES

- [1] H. H. Dohaiha, P. W. C. Prasad, A. Maag, and A. Alsadoon, "Deep learning for aspect-based sentiment analysis: A comparative review," *Expert Syst. Appl.*, vol. 118, pp. 272–299, Mar. 2019.
- [2] G. D'Aniello, M. Gaeta, and I. La Rocca, "KnowMIS-ABSA: An overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis," *Artif. Intell. Rev.*, vol. 55, pp. 5543–5574, Jan. 2022.
- [3] Y. Mao, Y. Shen, C. Yu, and L. Cai, "A joint training dual-MRC framework for aspect based sentiment analysis," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 13543–13551.
- [4] Y. Xiao and G. Zhou, "Syntactic edge-enhanced graph convolutional networks for aspect-level sentiment classification with interactive attention," *IEEE Access*, vol. 8, pp. 157068–157080, 2020.
- [5] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si, "Knowing what, how and why: A near complete solution for aspect-based sentiment analysis," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 8600–8607.
- [6] S. Chen, Y. Wang, J. Liu, and Y. Wang, "Bidirectional machine reading comprehension for aspect sentiment triplet extraction," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 12666–12674.
- [7] L. Xu, H. Li, W. Lu, and L. Bing, "Position-aware tagging for aspect sentiment triplet extraction," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 2339–2349.
- [8] L. Xu, Y. K. Chia, and L. Bing, "Learning span-level interactions for aspect sentiment triplet extraction," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics 11th Int. Joint Conf. Natural Lang. Process.*, vol. 1, 2021, pp. 4755–4766.
- [9] Z. Wu, C. Ying, F. Zhao, Z. Fan, X. Dai, and R. Xia, "Grid tagging scheme for aspect-oriented fine-grained opinion extraction," in *Proc. Findings Assoc. Comput. Linguistics (EMNLP)*, 2020, pp. 2576–2585.
- [10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1, 2019, pp. 4171–4186.
- [11] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 7871–7880.
- [12] H. Chen, Z. Zhai, F. Feng, R. Li, and X. Wang, "Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2022, pp. 2974–2985.
- [13] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 8, no. 4, 2018, Art. no. e1253.
- [14] N. Liu and B. Shen, "Aspect-based sentiment analysis with gated alternate neural network," *Knowl.-Based Syst.*, vol. 188, Jan. 2020, Art. no. 105010.
- [15] N. Liu, B. Shen, Z. Zhang, Z. Zhang, and K. Mi, "Attention-based sentiment reasoner for aspect-based sentiment analysis," *Hum.-Centric Comput. Inf. Sci.*, vol. 9, no. 1, pp. 1–17, Dec. 2019.
- [16] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 168–177.
- [17] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Natural Language Processing and Text Mining*. London, U.K.: Springer, 2007, pp. 9–28.
- [18] M. Tubishat, N. Idris, and M. Abushariah, "Explicit aspects extraction in sentiment analysis using optimal rules combination," *Future Gener. Comput. Syst.*, vol. 114, pp. 448–480, Jan. 2021.
- [19] B. Ozyurt and M. A. Akcayol, "A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA," *Expert Syst. Appl.*, vol. 168, Apr. 2021, Art. no. 114231.
- [20] Z. Zhang, Y. Rao, H. Lai, J. Wang, and J. Yin, "TADC: A topic-aware dynamic convolutional neural network for aspect extraction," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 25, 2021, doi: 10.1109/TNNLS.2021.3119026.
- [21] M. Venugopalan and D. Gupta, "An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis," *Knowl.-Based Syst.*, vol. 246, Jun. 2022, Art. no. 108668.
- [22] A. Kumar, A. S. Veerubhotla, V. T. Narapareddy, V. Aruru, L. B. M. Neti, and A. Malapati, "Aspect term extraction for opinion mining using a hierarchical self-attention network," *Neurocomputing*, vol. 465, pp. 195–204, Nov. 2021.
- [23] Z. Chen and T. Qian, "Bridge-based active domain adaptation for aspect term extraction," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics 11th Int. Joint Conf. Natural Lang. Process.*, vol. 1, 2021, pp. 317–327.
- [24] A. Klein, O. Pereg, D. Korat, V. Lal, M. Wasserblat, and I. Dagan, "Opinion-based relational pivoting for cross-domain aspect term extraction," in *Proc. 12th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal.*, 2022, pp. 104–112.
- [25] J. Yu, J. Jiang, and R. Xia, "Global inference for aspect and opinion terms co-extraction based on multi-task neural networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 1, pp. 168–177, Jan. 2019.
- [26] H. Zhao, L. Huang, R. Zhang, Q. Lu, and H. Xue, "SpanMI: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3239–3248.
- [27] L. Gao, Y. Wang, T. Liu, J. Wang, L. Zhang, and J. Liao, "Question-driven span labeling model for aspect–opinion pair extraction," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 12875–12883.
- [28] H. Zhang, Z. Chen, B. Chen, B. Hu, M. Li, C. Yang, and B. Jiang, "Complete quadruple extraction using a two-stage neural model for aspect-based sentiment analysis," *Neurocomputing*, vol. 492, pp. 452–463, Jul. 2022.
- [29] Y. Dai, P. Wang, and X. Zhu, "Reasoning over multiplex heterogeneous graph for target-oriented opinion words extraction," *Knowl.-Based Syst.*, vol. 236, Jan. 2022, Art. no. 107723.
- [30] C. Wu, F. Wu, S. Wu, Z. Yuan, and Y. Huang, "A hybrid unsupervised method for aspect term and opinion target extraction," *Knowl.-Based Syst.*, vol. 148, pp. 66–73, May 2018.
- [31] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective LSTMs for target-dependent sentiment classification," in *Proc. 26th Int. Conf. Comput. Linguistics (COLING)*, 2016, pp. 3298–3307.
- [32] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 606–615.
- [33] W. Xue and T. Li, "Aspect based sentiment analysis with gated convolutional networks," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2018, pp. 2514–2523.
- [34] J. Cao, R. Liu, H. Peng, L. Jiang, and X. Bai, "Aspect is not you need: No-aspect differential sentiment framework for aspect-based sentiment analysis," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2022, pp. 1599–1609.
- [35] A. Zhao and Y. Yu, "Knowledge-enabled BERT for aspect-based sentiment analysis," *Knowl.-Based Syst.*, vol. 227, Sep. 2021, Art. no. 107220.
- [36] R. Li, H. Chen, F. Feng, Z. Ma, X. Wang, and E. Hovy, "Dual graph convolutional networks for aspect-based sentiment analysis," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics 11th Int. Joint Conf. Natural Lang. Process.*, vol. 1, 2021, pp. 6319–6329.

- [37] B. Liang, H. Su, L. Gui, E. Cambria, and R. Xu, "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," *Knowl.-Based Syst.*, vol. 235, Jan. 2022, Art. no. 107643.
- [38] Y. Tian, G. Chen, and Y. Song, "Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2021, pp. 2910–2922.
- [39] Z. Li, L. Li, A. Zhou, and H. Lu, "JTSG: A joint term-sentiment generator for aspect-based sentiment analysis," *Neurocomputing*, vol. 459, pp. 1–9, Oct. 2021.
- [40] R. K. Yadav, L. Jiao, O.-C. Granmo, and M. Goodwin, "Human-level interpretable learning for aspect-based sentiment analysis," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 14203–14212.
- [41] C. Zhang, Q. Li, D. Song, and B. Wang, "A multi-task learning framework for opinion triplet extraction," in *Proc. Findings Assoc. Comput. Linguistics: EMNLP*, 2020, pp. 819–828.
- [42] H. Yan, J. Dai, T. Ji, X. Qiu, and Z. Zhang, "A unified generative framework for aspect-based sentiment analysis," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics 11th Int. Joint Conf. Natural Lang. Process.*, vol. 1, 2021, pp. 2416–2429.
- [43] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, 2014, pp. 27–35.
- [44] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "SemEval-2015 task 12: Aspect based sentiment analysis," in *Proc. 9th Int. Workshop Semantic Eval. (SemEval)*, 2015, pp. 486–495.
- [45] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, and O. De Clercq, "Semeval-2016 task 5: Aspect based sentiment analysis," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 19–30.
- [46] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, and C. D. Manning, "Stanza: A Python natural language processing toolkit for many human languages," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics, Syst. Demonstrations*, 2020, pp. 101–108.
- [47] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, and M. Funtowicz, "Transformers: State-of-the-art natural language processing," in *Proc. Conf. Empirical Methods Natural Lang. Process., Syst. Demonstrations*, 2020, pp. 38–45.
- [48] Z. Fan, Z. Wu, X.-Y. Dai, S. Huang, and J. Chen, "Target-oriented opinion words extraction with target-fused neural sequence labeling," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1, 2019, pp. 2509–2518.
- [49] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "An interactive multi-task learning network for end-to-end aspect-based sentiment analysis," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 504–515.
- [50] Z. Chen, H. Huang, B. Liu, X. Shi, and H. Jin, "Semantic and syntactic enhanced aspect sentiment triplet extraction," in *Proc. Findings Assoc. Comput. Linguistics (ACL-IJCNLP)*, 2021, pp. 1474–1483.



**JIE HU** received the master's degree in software engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2017. She is currently the Deputy Director of the Research Center for Big Data and Cognitive Computing, China Telecom Research Institute. Her research interests include information retrieval, multi-hop reasoning, and natural language processing.



**SHUNYU YAO** received the master's degree in computer technology from the Beijing University of Posts and Telecommunications, Beijing, China, in 2022. He is currently working as an Algorithm Engineer at the China Telecom Research Institute. His research interests include event extraction and natural language processing.



**DAN LIU** (Member, IEEE) received the M.D. degree in communication and information systems from Beijing Jiaotong University, China, in 2019. She is currently working as an Engineer at China Telecom Research. Her research interests include edge computing, edge intelligence, and digital twin.



**NING LIU** received the Ph.D. degree from Beijing Jiaotong University, Beijing, China, in 2022. He is currently working as an Algorithm Engineer at the China Telecom Research Institute. His research interests include sentiment analysis, deep learning, and NLP techniques.



**MINGCHUAN YANG** received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, in 2002. He is currently the Director of the Institute of Big Data and Artificial Intelligence and the China Telecom Research Institute. His research interests include big data and artificial intelligence.

...