Attention Feature Fusion Graph Convolutional Network for Target-Oriented Opinion Words Extraction

Shuaibo Li, Zhengpeng Li, Member, IANEG, Jiansheng Wu, Jiawei Miao, Yuhang Bai, Xinmiao Yu, Kejin Li

***Abstract*—Target-Oriented Opinion Words Extraction (TOWE) is a challenging sequence extraction task aimed at extracting opinion words corresponding to the opinion targets for a given sentence. Enhancing the performance of TOWE requires careful consideration of the semantic information within the sentence, particularly in relation to the opinion words and opinion targets. Although utilizing graph convolutional operations on the syntactic dependency tree allows for the utilization of syntactic dependency information, these operations do not effectively balance the degree of dependency on syntactic parsing results. This paper proposes an Attention Feature Fusion Graph Convolutional Network (AFFGCN) to address the issue. The proposed method enriches the feature representation of nodes through Graph Convolutional Networks (GCN) and captures the sequence features of the sentence using a Bi-LSTM. The Global Feature-aware Attention Module (GFA) guides the model to learn the global feature representation of the sentence to determine the absolute importance of a single word in the sentence. The Neighborhood-aware Attention in Feature Fusion Encoding Module (FFE) fully considers the syntactic structure of sentences to construct a high-quality syntactic perception representation. The experimental results demonstrate the effectiveness of our proposed method. The performance of AFFGCN is comparable to or even better than the state-of-the-art TOWE baseline models.**

***Index Terms*—Graph convolutional network, Target-oriented Opinion Words Extraction, Syntactic dependency tree, Attention feature fusion**

Manuscript received April 04, 2023; revised July 17, 2023. The research work was supported by Provincial Department Of Education Project of Liaoning (No. 2020LNZD05), and Science and Technology Innovation Project of University of Science and Technology Liaoning (LKDYC202109,LKDYC202219).

Shuaibo Li is a graduate student of University of Science and Technolog y Liaoning, Anshan, 114051, China. (e-mail: 1376111053@qq.com).

Zhengpeng Li is a doctoral student of University of Science and Technol ogy Liaoning, Anshan, 114051, China. (e-mail: 1156361257@qq.com).

Jiansheng Wu* is a professor of University of Science and Technology Liaoning, Anshan, 114051, China. (corresponding author, e-mail: ssewu@1 63.com).

Jiawei Miao is a doctoral student of University of Science and Technolo gy Liaoning, Anshan, 114051, China. (e-mail: 455605116@qq.com)

Yuhang Bai is a graduate student of University of Science and Technolo gy Liaoning, Anshan, 114051, China. (e-mail: 604760451@qq.com).

Xinmiao Yu is a graduate student of University of Science and Technolo gy Liaoning, Anshan, 114051, China. (e-mail: 2749936763@qq.com).

Kejin Li is a graduate student of University of Science and Technology Liaoning, Anshan, 114051, China. (e-mail: 1710107457@qq.com).

I. Introduction

Target-oriented Opinion Word Extraction (TOWE) is a specific subtask within the broader field of aspect-based sentiment analysis (ABSA). When analyzing review information, TOWE can identify fine-grained opinion information[1], positively impacting the development of various downstream tasks, such as sentiment classification and text mining. Therefore, TOWE has garnered significant attention in recent research. The goal of TOWE is to extract opinion words from a given sentence that align with the corresponding opinion targets. Opinion targets can also be referred to as attribute terms, and opinion words can also be referred to as sentiment terms. As shown in Figure 1, in the review text "The price is reasonable although the service is poor", "price" and "service" are two opinion targets. The objective of TOWE is to identify the sentiment term "reasonable" as the opinion word for "price", and the sentiment term "poor" as the opinion word for "service".

The TOWE task presents significant challenges due to the presence of multiple opinion targets and opinion words within a single comment. Additionally, opinion targets may have multiple representations. Early explorations used manually designed rules or templates[2] to accomplish this task. However, previous methods did not yield satisfactory results due to the complexity and non-standard nature of comment text data. Fan et al.[3] formally proposed the TOWE task, categorized it as a sequence labelling task. As shown in Figure 1, syntactic dependencies can serve as a shortcut for connecting target words and opinion words. Syntactic dependency information can help the model filter out irrelevant opinion words[4] for sentences containing multiple target-opinion pairs. Some works convert syntactic dependency trees into graph data structures, utilizing Graph Convolutional Networks to capture syntactic information[5, 6]. Many methods attempt to inject target information into sentence representations. Wu et al.[7] introduced position embeddings for target information, and Jiang et al.[5] introduced label embeddings for targets, both achieving good results.

However, the limitation of these techniques is that they do not fully exploit information about grammatical structure and semantics. In this paper, we propose a novel Attention Feature Fusion Graph Convolutional Network. During the training process, AFFGCN considers both local context information and global semantic information. Specifically, we use a processor to obtain the grammatical dependency tree of the sentence and construct graph data based on the dependency tree. Next, we utilize AFFGCN to encode the graph data and extract syntactic information features. Finally, we capture sequence information from the data by feeding the syntactic information into Bi-LSTM. We evaluated AFFGCN model on four benchmark datasets and conducted numerous experiments, demonstrating its superiority over baseline models. Remarkably, even when utilizing only global features, our model achieves comparable or superior performance compared to state-of-the-art models. Adding local context features further enhanced the performance of the model. The effectiveness of each model component has been proven through extensive experiments and studies.

The major contributions of this work are as follows:

1) We propose a method of attention feature fusion to capture both local context information and global feature information, enabling our model to learn both syntactic and semantic information.

2) We have conducted extensive experiments on benchmark datasets, and the results demonstrate that our model surpasses various baseline methods.

II. BACKGROUND AND RELATED WORK

A. Target-oriented Opinion Words Extraction

Target-oriented Opinion Words Extraction is a developing aspect of Aspect-Based Sentiment Analysis[8]. Early research tackled the extraction of target words and opinion words separately[9], including rule-based methods and topic modelling[10]. Yin et al.[11] learned distributed representations and dependency features of words in an unsupervised framework and used conditional random fields for target word extraction. Xu et al.[5] adopted a graph neural network with dual embedding for target word extraction. Some methods use variants of recurrent neural networks to model sentence representations[12, 13] and achieve promising performance.

Although these methods have achieved certain results, they are unable to extract the corresponding relationship between target words and opinion words. Wang et al.[14] explored the combined extraction of target words and opinion words. They designed a multi-layer attention network that Bidirectionally propagates target and opinion information through interactive learning of a pair of attentions, achieving significant effects. The TOWE task was initially suggested by Fan et al.[3], who also labeled the dataset and they designed a target fusion model for it. Wu et al.[7] established a latent opinion transfer network, achieving improved performance by transferring opinion information from expansive sentiment classification datasets to the resource-limited TOWE task. However, these methods cannot fully exploit dependency information[15]. Some studies propose to combine dependency trees with neural networks, allowing the model to learn word representations and preserve dependency information[16, 17]. Using Graph Convolutional Networks on syntactic dependency trees is a promising trend.

B. Graph Convolutional Networks

Graph Convolutional Networks are a type of neural networks that can effectively process graph data. GCN has achieved outstanding results in various, including chemistry, recommendation systems, and social networks. GCNs and their variations have been extensively used in ABSA tasks. Veyseh et al.[18] attempted to incorporate syntactic structure into their model. To learn syntactic dependencies, Zhou et al.[5] proposed a model built on graph neural networks and adversarial training. It is worth mentioning that giving equal weights to every node in the graph is unreasonable for real-world data[19]. Recent works have attempted to resolve this problem through the attention mechanism. Li et al.[6] designed a dual Graph Convolutional Network to learn the correlation between syntactic structure and semantics. Jiang et al.[5] proposed a distance-aware attention mechanism to improve R-GCNs and fully utilize syntactic information on the dependency graph. However, dependency parsing results can be inaccurate, and previous works overly rely on dependency information, lacking effective mechanisms to distinguish important features.

Inspired by previous work, we propose an innovative Attention Feature Fusion Graph Convolutional Network. Our method utilizes a global feature-aware attention mechanism to learn the global structure and feature information of the given graph, while the Neighborhood-aware Attention in Feature Fusion Encoding Module recognizes relationships between two nodes in the graph that are distant but have similar or adjacent features and structures. By fusing global features, local features, and relationship types between nodes, our method effectively incorporates syntactic dependency information into the node encoding process.

III. PROPOSED MODEL

A. Task Definition

Given a sentence S consisting of n words: $s = \{w_1, w_2, ..., w_i, ..., w_n\}$, there exists an opinion target $t = \{w_i, w_{i+1}, ..., w_{i+m}\}$ in the sentence, where $0 \le m \wedge i + m < n$, and the task is to extract the corresponding opinion word $a = \{w_i, w_{i+1}, ..., w_{i+q}\}$ for the opinion targets, where $0 \le q \wedge i + q < n$. The BIO tagging scheme[20] is used in this task. Each word in the sentence is labeled as $y_i \in \{B, I, O\}$ (B: Beginning, I: Inside, O: Others). Each opinion target and opinion word may consist of one or more words. As shown in the example in Figure 1, the sentence "The/O price/B is/O reasonable/O although/O the/O service/O is/O poor/O ./O" is provided, and the expected labeled result is "The/O price/O is/O reasonable/B although/O the/O service/O is/O poor/O ./O". In this case, the opinion target "price" corresponds to the opinion word "reasonable".

B. Overview

This section introduces the proposed model. Figure 2 shows the overall architecture of the model, comprising three major sections: the Global Feature-aware Attention module, the Feature Fusion Encoding module, and the Sequence Encoding module.

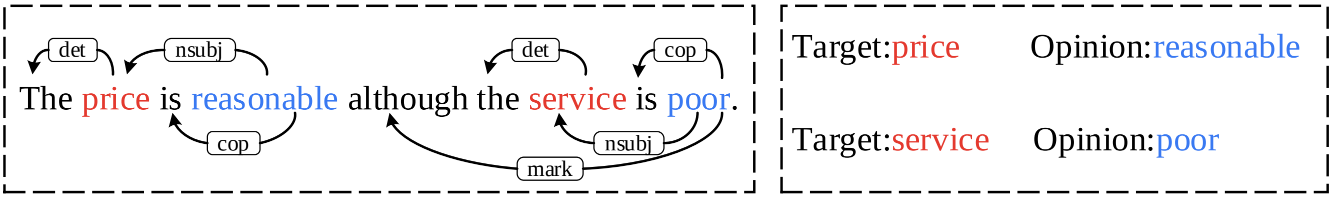

Fig.1 TOWE example with syntactic dependency tree. Opinion targets are marked in red, and opinion words are marked in blue.

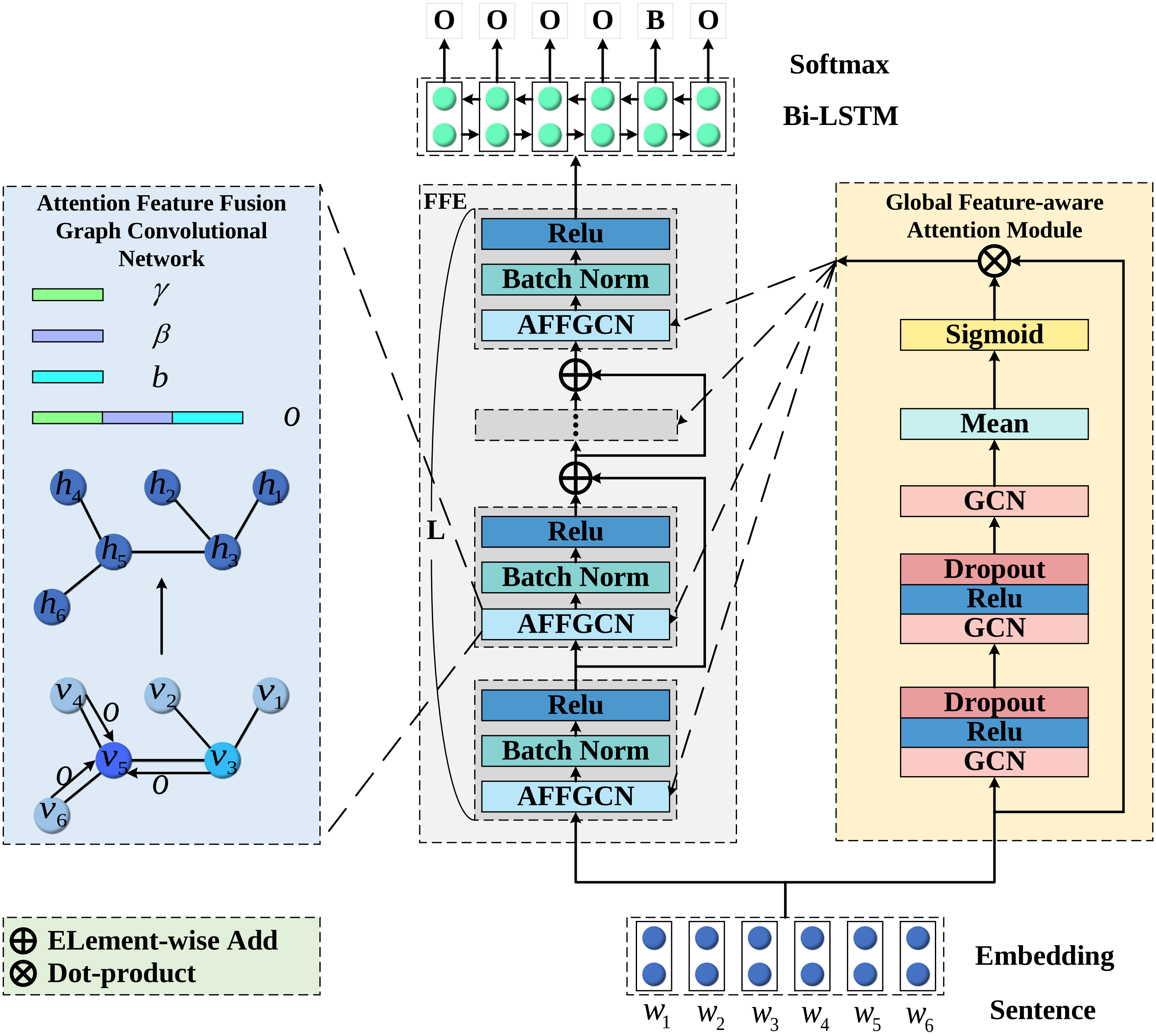

Fig.2 AFFGCN structure diagram. Firstly, the textual vectorized information is fed into the Global Feature-aware Attention Module and the Feature Fusion Encoding Module. The GFA module obtains the global feature-aware attention of the nodes, while the FFE module obtains the neighborhood perception attention. Then, the global attention, neighborhood attention, and the relationship between the nodes are concatenated to obtain the final attention representation, denoted as o. As an example, considering the node $v5$ as the central node, its neighbor nodes are $v3$, $v4$, and $v6$. The AFFGCN aggregates information from neighbor nodes using attention scores o, then updates the node $v5$ with the hidden information $h5$.

(1) The Global Feature-aware Attention module obtains the global information of syntactic dependency graph through GCN and then calculates the significance of each node in the graph.

(2) The Feature Fusion Encoding module is primarily used to learn the syntactic dependencies of sentences and enrich the feature representation of nodes. It comprises multiple layers of AFFGCN Network and normalization layers. The AFFGCN first calculates the similarity between the central node and its neighbor nodes, and then concatenates the local similarity, global association, and node relationship type information as attention scores. Finally, the new attention scores are used to aggregate the hidden information of the neighbor nodes around the central node. This strategy enables the model to consider both syntactic structure and semantic relevance, as well as capture contextual features.

(3) The Sequence Encoding module is composed of Bi-LSTM. Bi-LSTM is used to capture the sequential information of the syntactic representations extracted by AFFGCN.

In addition, to accurately extract opinion words for a given target, we employ target-aware representations to effectively encode target information in our model[5]. We define the target label as $P=\{p_1, p_2, \dots, p_i, \dots, p_n\}$, and the target-aware representation of the word w_i as follows,

$$e_i = [e_i^w, e_i^p], \tag{1}$$

Where e_i^w is the vectorized representation of a word w_i, e_i^p is the target label embedding for the word w_i, and [,] denotes concatenation. The target-aware representation of a particular sentence is shown as $E=[e_1; e_2; \dots e_i; \dots; e_n]$.

The first step, comparable to earlier work, involves converting the syntactic dependency tree into graph-structured data to utilize graph neural networks in the TOWE task. A directed graph is defined as $G=\{V,E,R\}$, where V represents the set of nodes, E represents the set of edges connecting nodes, and R represents the set of node connection classes. We obtain a sparse representation of the adjacency matrix A based on G, where $A \in \mathbb{R}^{2\times n}$ and n represent the size of E.

C. Global Feature-Aware Attention Module

An effective encoding module should h be able to capture the semantic information of the sentence, and we introduce a global attentive mechanism. To calculate the global feature representation $g \in \mathbb{R}^d$ of the graph, a simple approach is to take the average of node embeddings,

$$g = \tanh\left(\frac{1}{|V|} W_2 \sum_{v_i \in V} \tilde{h}_i\right), \tag{2}$$

$$\tilde{h}_i^{l+1} = f_1\left(\frac{1}{d_i} A_{ij}\left(W_1^l \tilde{h}_j^l + b^l\right)\right), \tag{3}$$

Where $\tilde{h}_j^l$ refers to the hidden information generated by the l-th layer GCN for the node v_j, while $\tilde{h}_i^{l+1}$ refers to the hidden information obtained by aggregating neighbor information for the node v_i. f_1 represents a non-linear activation function (such as $RELU(\cdot)$), d_i is the degree of the node v_i, and W_1^l, W_2, and b^l are trainable parameters.

To enable a node v_i to accurately learn a global feature representation, we compute the dot product between this node and the global feature g. In this way, we calculate an attention score for each node. Intuitively, nodes with features that are similar to the global feature should receive more attention.

$$\gamma_{ig} = f_2\left(\tilde{h}_i \otimes g\right), \tag{4}$$

in the equation, the sigmoid function f_2 is used, and the attention scorer γ_{ig} determines which nodes are more important, guiding the model to learn specific associations.

D. Feature Fusion Encoding Module

L layers are set in the Feature Fusion Encoding module, with each layer is composed of the Attention Feature Fusion Graph Convolution Network and Batch Normalization. Among them, AFFGCN is an improvement based on R-GCNs.

In the FFE, we have designed a Neighborhood-aware Attention to guide the model to capture local feature information of the graph. We assume that adjacent nodes have similar features, and the attention coefficients between nodes depend on the similarity between them. Specifically, we obtain the $query^{h_i}$ and key^{h_j} through the projection matrix of the central node feature h_i and the neighbor node feature h_j. The local attention weight of node v_j to node v_i is calculated as follows,

$$\beta_{ij} = soft\max(h_i \odot h_j), \tag{5}$$

finally, we concatenate the relational information, local attention scores, and global attention scores. The node v_i updates its hidden information via AFFGCN,

$$h_i^{l+1} = \sum_{r\in R} \sum_{j \in N_i} o_{ij} W_0^l h_j^l, \tag{6}$$

$$o_{ij} = f_1\left(\left[b_r, \gamma_{jg}, \beta_{ij}\right] W_3^l\right), \tag{7}$$

in this formulation, o_{ij} denotes the final weight of neighbor node v_j with respect to center node v_i, W_3^l is a trainable parameters, and b_r is the vector representation of relation r. The function f_1 is an activation function. This part integrates node relation features, global features, and local context features. AFFGCN fully considers features at different scales when updating node information.

Furthermore, we employ a multi-head attention mechanism to obtain contextual information, which improves the performance and stability of our model. AFFGCN updates node information using independent attention mechanisms. The output of each layer of multi-head attention is as follows,

$$\bar{h}_i = W_d\left[h_i^1, h_i^2, ..., h_i^k, \dots, h_i^K\right], \tag{8}$$

among them, $W_d \in \mathbb{R}^{Kd_K}$ and d_K are the dimensions of each attention head. h_i^k is the hidden information of node v_i in the k-th attention head, where $1 \le k < K$.

GCN updates the hidden information of the central node by aggregating hidden information from neighbor nodes. A single GCN only considers the first-order neighborhood information. However, for complex structures such as long sentences, a simple GCN may lose the dependency relationship between the opinion targets and the opinion words. Therefore, we use multiple layers of consecutive AFFGCNs to capture these relationships. As the number of layers of GCN in the model increases, the multi-hop neighborhood features aggregate towards the edge nodes, which may result in the problem of over-smoothing. To address this issue, we introduce residual connections,

$$\bar{h}_i^{l+1} = \bar{h}_i^l + \bar{h}_i'^l, \tag{9}$$

Where $\bar{h}_i^l$ is the hidden information of node v_i at the input of the l-th layer, and $\bar{h}_i'^l$ is the hidden information of node v_i after being updated by the l-th layer AFFGCN.

E. Sequence Encoding Module

The TOWE task is essentially a sequence labeling task. However, graph neural networks are cannot directly capture sequence information, which may result in unsatisfactory model performance. We used the Bi-LSTM network to extract sequence information from the syntactic information obtained by AFFGCN,

$$\hat{h}_i = \left[\vec{h}_i, \overleftarrow{h}_i\right], \tag{10}$$

$$\overrightarrow{h}_i = \overrightarrow{LSTM}\left(\overrightarrow{s}_{i-1}, \left[\overrightarrow{h}_{i-1}, \bar{h}_i^l\right]; c_i\right), \quad (11)$$

$$\overleftarrow{h}_i = \overleftarrow{LSTM}(\overleftarrow{s}_{i+1}, \left[\overleftarrow{h}_{i+1}, \bar{h}_i^l\right]; c_i), \quad (12)$$

here, $\overrightarrow{h}_{i-1} \in \mathbb{R}^{d_h}$ and $\overrightarrow{s}_{i-1} \in \mathbb{R}^{d_h}$ refer to the hidden information and memory information of the previous time step, respectively. $\hat{h}_i \in \mathbb{R}^{2d_h}$ is the concatenation of the forward and backward hidden information, and c_i represents all relevant parameters of the Bi-LSTM.

F. Module Training

The numerical representation of each classification label $\{B, I, O\}$ for each opinion word was set to $\{1, 2, 0\}$. We fed the sequence information into a linear layer and used the softmax function to decode it,

$$\hat{Y}_i = soft\max(W_{fc}\hat{h}_i + b_{fc}), \quad (13)$$

Where W_{fc} and b_{fc} are trainable parameters of the linear layer, and $\hat{Y}_i$ represents the predicted label distribution.

The loss function is specified as follows, and we use cross-entropy as the training criterion,

$$\mathcal{L} = -\sum_{w_i \in s}\sum_{k=0}^{2} y_i^k \log(\hat{y}_i^k; \Theta), \quad (14)$$

in the equation, n represents the number of training samples, k is the number of classes, y_i^k is the true distribution of the labels, and $\hat{y}_i^k$ is the output distribution of the model, where Θ represents the trainable parameters. During the model training process, the loss $\mathcal{L}$ is minimized through back-propagation.

IV. EXPERIMENTS

A. Datasets and Metrics

Following previous works, our model was evaluated on four popular datasets. 14res and 14lap are from SemEval challenge 2014 Task 4[21], 15res is from SemEval challenge 2015 Task 12[22], and 16res is from SemEval challenge 2016 Task 5[23]. The suffix "res" and "lap" indicate that they come from reviews of restaurants and laptops, respectively. One of the most widely used benchmarks for many ABSA subtasks is the original SemEval challenge datasets. The annotated opinion targets are available in the original datasets, but the corresponding opinion words and relations are not annotated. To perform the TOWE task, Fan et al.[3] annotated the corresponding opinion words for the opinion targets in the datasets and ignored instances where there were no explicit opinion words. Table 1 displays the statistical information for these datasets.

Like most classification tasks, we use Precision, Recall, and F1 as assessment measures. Only when the anticipated location of the opinion term completely accords with the actual situation is a forecast deemed to be accurate.

B. Settings

In our experiments, we initialized word embedding vectors[24] using 300-dimensional GloVe vectors[25] for all non-BERT-based models. For the pre-trained AFFGCN, we used the last hidden state of pre-trained BERT[26] as the representation for words and fine-tuned it jointly, with the dimension of the word embedding set to 768. We set the depth of the GCN layer to 10 and the number of channels to 128 in the model architecture, while the hidden dimension of Bi-LSTM was set to 128. We used spaCy as the dependency parser[27].

TABLE I
THE STATISTICS OF THE DATASETS.
Note that a sentence may contain multiple opinion targets.

Dataset	Split	Sentences	Targets
14res	Training	1627	2643
	Testing	500	865
14lap	Training	1158	1634
	Testing	343	482
15res	Training	754	1076
	Testing	325	436
16res	Training	1079	1512
	Testing	329	457

C. Comparison models

We contrast our model with the subsequent methods:

1) **Distance-rule**[5]: This is a distance-based approach that uses the distance information between opinion targets and opinion words, which is an important feature of text data. For each word in the sentence, this method generates a part-of-speech tag and selects the closest adjective to the given target as the opinion word.

2) **Dependency-rule**[5]: This method uses rule templates to extract opinion pairs. It mines the dependency relations between opinion targets and opinion words in the training data through syntax dependency trees and records the POS tags of the opinion targets and opinion words, as well as their dependency relations, as rule templates.

3) **LSTM/Bi-LSTM**: Fan et al.[3] built an LSTM/Bi-LSTM network based on the work by Liu et al.[28]. This is a method used for sentence-level opinion words extraction, which passes word embedding vectors into an LSTM or Bi-LSTM and performs three-class classification on each hidden state.

4) **Pipeline**[3]**:** This method combines the LSTM and distance-rule approaches. It uses a Bi-LSTM model to extract all opinion words from a sentence and selects the opinion word closest to the target as the result.

5) **TC-BiLSTM**[29]: This method takes into account the target information. In TC-LSTM, the target vector and target word embedding are concatenated as the input at each location. The target vector is created by aggregating the target word embeddings.

6) **IOG**[3]: IOG models the sentence from six different directions. This strategy effectively encodes target information into both the left and right contexts while integrates global contextual information.

7) **LOTN[7]:** This is a transfer learning method that uses a Bi-LSTM network with position embeddings, a simple yet effective strategy. To alleviate the relative scarcity of training data for the TOWE task, LOTN first trains a model on a large-scale sentiment classification dataset. It then utilizes this pre-trained model to learn potential opinion words in the data before combining the hidden states of the two modules to complete the TOWE task.

TABLE II
MAIN EXPERIMENTAL RESULTS (%).
The best results without using the pre-trained model are shown in italics. The best result in the case of using the pre-trained model is shown in bold.

model	14lap			14res			16res			15res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Distance-rule	50.13	33.86	40.42	50.13	43.59	49.92	61.90	44.57	51.83	54.12	39.96	45.97
Dependency-rule	45.09	31.57	37.14	45.09	52.72	58.04	76.03	56.19	64.62	65.49	48.88	55.98
LSTM	55.71	57.53	56.52	55.71	65.47	58.34	62.46	68.72	65.33	57.27	60.69	58.93
Bi-LSTM	64.52	61.45	62.71	64.52	61.73	59.95	68.68	70.51	69.57	60.46	63.65	62.00
Pipeline	72.58	56.97	63.83	72.58	62.33	69.18	81.46	67.81	74.01	74.75	60.65	66.97
TC-BiLSTM	62.45	60.14	61.21	62.45	67.67	67.61	73.46	72.88	73.10	66.06	60.16	62.94
IOG	73.24	69.63	71.35	73.24	77.38	80.02	82.25	78.51	81.69	76.06	70.71	73.25
LOTN	77.08	67.62	72.02	77.08	80.52	82.21	86.57	80.89	83.62	76.61	70.29	73.29
TS-GCN	72.37	73.89	73.12	72.37	83.30	83.34	85.15	83.04	84.08	81.08	75.65	78.28
ARGCN	79.45	71.60	75.32	79.45	82.72	84.65	86.16	84.19	85.16	76.57	76.88	76.72
PER	80.68	70.72	75.38	80.68	80.39	83.30	90.00	84.00	*86.90*	81.50	75.05	78.14
AFFGCN	79.87	75.13	*77.43*	79.87	83.01	*85.05*	87.53	84.04	85.74	80.40	79.71	*80.05*
ARGCN(pre)	75.83	76.90	76.36	75.83	83.59	85.42	88.49	84.96	86.69	78.81	77.69	78.24
AFFGCN(pre)	80.60	77.57	**79.06**	80.60	85.67	**85.98**	88.88	85.56	**87.19**	82.34	79.43	**80.86**

TABLE III
EXPERIMENTAL RESULTS (%) FOR DIFFERENT MODEL DESIGNS.

model	14res			14lap			15res			16res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
AFFGCN	87.20	83.01	85.05	79.87	75.13	77.43	80.40	79.71	80.05	87.53	84.04	85.74
R-GCN	85.23	81.67	83.40	76.24	70.26	73.12	74.08	72.74	73.35	86.25	80.76	83.39
W/o global	85.68	83.03	84.33	79.19	75.21	77.15	79.60	78.42	79.01	86.01	84.08	85.03
W/o local	86.43	82.72	84.53	79.09	75.38	77.19	78.75	79.92	79.33	86.74	84.00	85.35
W/o pos	85.37	82.50	83.91	78.03	73.68	75.79	78.80	78.98	78.89	85.90	83.77	84.82

8) **TS-GCN**[4]**:** This method learns the dependencies between opinion words and targets using a multi-scale syntactic representation strategy. The input to this method consists of word embeddings and position embeddings, and Bi-LSTM is used to generate the initial representation of nodes in the graph. The memory unit aggregates historical information and local features from the GCN layer to update the hidden state.

9) **ARGCN**[5]: This paper proposes a target-aware representation and extends R-GCNs with a distance-aware attention mechanism.

10) **PER**[1]: This method applies reinforcement learning to extract opinion words from multiple heterogeneous graphs. Additionally, it constructs a padding module to enrich node information.

D. Results

Table 2 presents the experimental results of our model and the baseline models on four datasets. It can be observed that ARGCN and PER achieve similar scores, which demonstrates the effectiveness of introducing reinforcement learning into the TOWE task. In comparison to PER, AFFGCN has achieved significant improvements on the 14res, 15res, and 14lap datasets. However, on dataset 16res, our method is weaker than PER, which may be due to the advantage of PER in capturing long-range node information on this dataset. Our model, AFFGCN, performs similarly to ARGCN on 14res and 16res, while significantly outperforming ARGCN on 14lap and 15res. These results verify the effectiveness of incorporating both global and local features in the TOWE task.

Additionally, when using the pre-trained language model BERT, ARGCN achieved even more advanced performance. Correspondingly, our model also exhibited an average improvement of 0.96 points. BERT possesses robust word embedding capabilities and the ability to capture syntactic dependencies, which is beneficial for sentiment analysis tasks.

E. Ablation study

We conducted ablation experiments to evaluate the impact of different components of AFFGCN on the results. We report the performance of the following variants: W/o global, which does not use global feature information and relies solely on local features and node relationship features; W/o local, which does not use local contextual features and relies on node importance in the graph and node relationships to aggregate information from neighbor nodes; W/o pos, which does not incorporate part-of-speech information. The AFFGCN model is based on R-GCNs.

The performance comparison presented in Table 3 demonstrates the superiority of our model over R-GCNs, with significant F1 score improvements of 1.65, 4.31, 6.7, and 2.35, respectively. Overall, excluding any component from AFFGCN results in a decrease in average performance, highlighting the significant contribution of our approach to overall enhancement. Specifically, the variant without local features achieves comparable results to previous methods, while incorporating local features further enhances the performance of the model. These findings indicate that our approach effectively captures both semantic information and syntactic structures. Moreover, the observed decline in performance when part-of-speech information is removed aligns with previous research conclusions, emphasizing the importance of target-related information for TOWE performance. These results affirm the ability of our proposed method to learn target-specific representations.

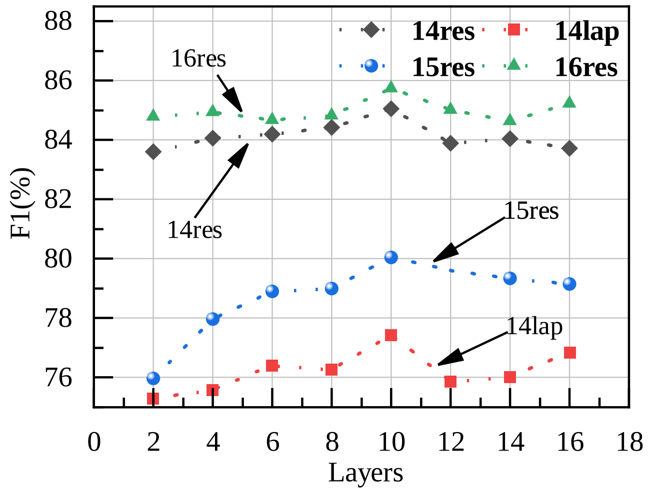

Fig.3 Results of F1 score for different numbers of layers.

F. The effect of the number of layers

One important parameter in AFFGCN is the number of GCN layers, which directly affects the performance of the model. To further explore the effect of the number of GCN layers on the end performance, we conduct experiments on four datasets while maintaining the same values for other hyperparameters.

As shown in Figure 3, the model achieved the best performance when the number of GCN layers was 10. From the trend in the figure, our model benefited from an increase in the number of layers. However, when the number of layers exceeded 10, the performance decreased. One possible reason for this is that excessive aggregation of multi-hop neighborhood information led to the model being overly smooth. This trend was particularly noticeable in the 14lap dataset, indicating a correlation between the depth of GCN layers and data characteristics. Previous studies generally recommend 2 to 4 GCN layers as optimal. In contrast to these works, we utilize residual connections to avoid overfitting.

G. Case study

Table 4 presents the case study results of our models, AFFGCN and ARGCN, on the 14lap dataset. Table 4 presents three cases, where each sentence contains 1, 2, and 3 opinion targets, respectively. It can be observed that, overall, both models accurately identify opinion words corresponding to different opinion targets. This accuracy can be attributed to the incorporation of opinion targets information in both models. In the third sentence, both models identified the corresponding opinion words for opinion targets "screen" and "keyboard". However, for the opinion target "features", although ARGCN identified the corresponding opinion word "expected", it also identified "more" as another opinion word. This may be because the ARGCN model benefits from the semantic and syntactic correlations between words, but is affected in sentences with a large number of targets. In contrast, our model focuses on local contextual information and has sufficient awareness of syntactic structure. Therefore, our model can accurately identify opinion words in comments with multiple opinion targets.

TABLE IV
CASE STUDY EXAMPLES
Examples with opinion targets in bold and corresponding opinion words underlined.

SETENCE	ARGCN	AFFGCN
Luckily, for all of us contemplating the decision, the Mac Mini is **priced** just right.	right	right
Its **size** is deal and the weight is acceptable.	deal	deal
Its size is deal and the **weight** is acceptable.	acceptable	acceptable
It has all the expected **features** and more +plus a wide screen and more than roomy keyboard.	expected, more	expected
It has all the expected features and more +plus a wide **screen** and more than roomy keyboard.	wide	wide
It has all the expected features and more +plus a wide screen and more than roomy **keyboard**.	roomy	roomy

V. CONCLUSION

In this work, we extended R-GCNs by incorporating global feature-aware attention and neighborhood-aware attention. We designed a neural network model to accomplish the TOWE task. By encoding the syntactic information of sentences using graph neural networks, our method effectively learns the syntactic structure and semantic information of the comment text. The effectiveness of the proposed method, which enhances the performance of the TOWE task by combining global and local features, is supported by experimental findings. In fact, the AFFGCN model can be viewed as a type of graph attention network. In graph neural networks, designing a well-defined attention mechanism for encoding syntactic information is beneficial for the TOWE task.

REFERENCES

[1] Y. Dai, P. Wang, and X. Zhu, "Reasoning over multiplex heterogeneous graph for Target-oriented Opinion Words Extraction," *Knowledge-Based Systems*, vol. 236, 2022.

[2] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," *Computational Linguistics*, vol. 37, no. 1, pp. 9-27, 2011.

[3] Z. Fan, Z. Wu, X.-Y. Dai, S. Huang, and J. Chen, "Target-oriented opinion words extraction with target-fused neural sequence labeling," in *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2019, June 2, 2019 - June 7, 2019*, Minneapolis, MN, United states, 2019, vol. 1, pp. 2509-2518: Association for Computational Linguistics (ACL).

[4] J. Zhang *et al.*, "Integrate syntax information for target-oriented opinion words extraction with target-specific graph convolutional network," *Neurocomputing*, vol. 440, pp. 321-335, 2021.

[5] J. Jiang, A. Wang, and A. Aizawa, "Attention-based relational graph convolutional network for target-oriented opinion words extraction," in *16th Conference of the European Chapter of the Associationfor Computational Linguistics, EACL 2021, April 19, 2021 - April 23,*

2021, Virtual, Online, 2021, pp. 1986-1997: Association for Computational Linguistics (ACL).

[6] R. Li, H. Chen, F. Feng, Z. Ma, X. Wang, and E. Hovy, "Dual graph convolutional networks for aspect-based sentiment analysis," in *Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL-IJCNLP 2021, August 1, 2021 - August 6, 2021*, Virtual, Online, 2021, pp. 6319-6329: Association for Computational Linguistics (ACL).

[7] Z. Wu, F. Zhao, X.-Y. Dai, S. Huang, and J. Chen, "Latent opinions transfer network for target-oriented opinion words extraction," in *34th AAAI Conference on Artificial Intelligence, AAAI 2020, February 7, 2020 - February 12, 2020*, New York, NY, United states, 2020, pp. 9298-9305: AAAI press.

[8] Z. Li, J. Wu, J. Miao, X. Yu, and S. Li, "Multi-model Fusion Attention Network for News Text Classification," *International Journal for Engineering Modelling*, vol. 35, no. 2, pp. 1-15, 2022.

[9] R. Alfred, L. B. Jing, J. H. Obit, Y. Lim, H. Haviluddin, and R. Alfred, "Concept Trending of Social Media Data Using Apriori Algorithm," *IAENG International Journal of Computer Science,* vol. 50, no. 1, pp. 287-296, 2023.

[10] Z. Li, J. Wu, J. Miao, and X. Yu, "Improve the Response Diversity of Multi-turn Dialogue System by Combining Knowledge," *IAENG International Journal of Computer Science,* vol. 49, no. 3, pp. 701-709, 2022.

[11] Y. Yin, F. Wei, L. Dong, K. Xu, M. Zhang, and M. Zhou, "Unsupervised word and dependency path embeddings for aspect term extraction," in *25th International Joint Conference on Artificial Intelligence, IJCAI 2016, July 9, 2016 - July 15, 2016*, New York, NY, United states, 2016, vol. 2016-January, pp. 2979-2985: International Joint Conferences on Artificial Intelligence.

[12] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, September 9, 2017 - September 11, 2017*, Copenhagen, Denmark, 2017, pp. 452-461: Association for Computational Linguistics (ACL).

[13] O. Irsoy and C. Cardie, "Opinion mining with deep recurrent neural networks," in *2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25, 2014 - October 29, 2014*, Doha, Qatar, 2014, pp. 720-728: Association for Computational Linguistics (ACL).

[14] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, "Coupled multi-layer attentions for co-extraction of aspect and opinion terms," in *31st AAAI Conference on Artificial Intelligence, AAAI 2017, February 4, 2017 - February 10, 2017*, San Francisco, CA, United states, 2017, pp. 3316-3322: AAAI press.

[15] J.-R. Chang, L.-S. Chen, and L.-W. Lin, "A Novel Cluster based Over-sampling Approach for Classifying Imbalanced Sentiment Data," *IAENG International Journal of Computer Science,* vol. 48, no. 4, pp. 1118-1128, 2021.

[16] L. Mou, H. Peng, G. Li, Y. Xu, L. Zhang, and Z. Jin, "Discriminative neural sentence modeling by tree-based convolution," in *Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, September 17, 2015 - September 21, 2015*, Lisbon, Portugal, 2015, pp. 2315-2325: Association for Computational Linguistics (ACL).

[17] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-Term memory networks," in *53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL-IJCNLP 2015, July 26, 2015 - July 31, 2015*, Beijing, China, 2015, vol. 1, pp. 1556-1566: Association for Computational Linguistics (ACL).

[18] A. P. B. Veyseh, N. Nouri, F. Dernoncourt, D. Dou, and T. H. Nguyen, "Introducing syntactic structures into target opinion word extraction with deep learning," in *2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, November 16, 2020 - November 20, 2020*, Virtual, Online, 2020, pp. 8947-8956: Association for Computational Linguistics (ACL).

[19] G. Sun, Z. Wang, Z. Ding, and J. Zhao, "An Ensemble Classification Algorithm for Short Text Data Stream with Concept Drifts," *IAENG International Journal of Computer Science,* vol. 48, no. 4, pp. 1056-1061, 2021.

[20] L. A. Ramshaw and M. P. Marcus, "Text Chunking Using Transformation-Based Learning," in *Natural Language Processing Using Very Large Corpora*, S. Armstrong, K. Church, P. Isabelle, S. Manzi, E. Tzoukermann, and D. Yarowsky, Eds. Dordrecht: Springer Netherlands, 1999, pp. 157-176.

[21] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," in *8th International Workshop on Semantic Evaluation, SemEval 2014, August 23, 2014 - August 24, 2014*, Dublin, Ireland, 2014, pp. 27-35: Association for Computational Linguistics (ACL).

[22] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "SemEval-2015 Task 12: Aspect Based Sentiment Analysis," in *9th International Workshop on Semantic Evaluation, SemEval 2015, June 4, 2015 - June 5, 2015*, Denver, CO, United states, 2015, pp. 486-495: Association for Computational Linguistics (ACL).

[23] M. Pontiki *et al.*, "SemEval-2016 task 5: Aspect based sentiment analysis," in *10th International Workshop on Semantic Evaluation, SemEval 2016, June 16, 2016 - June 17, 2016*, San Diego, CA, United states, 2016, pp. 19-30: Association for Computational Linguistics (ACL).

[24] Venkatesh, S. U. Hegde, S. B. Basapur, and Y. Nagaraju, "DistilBERT-CNN-LSTM Model with GloVe for Sentiment Analysis on Football Specific Tweets," *IAENG International Journal of Computer Science,* vol. 49, no. 2, pp. 420-432, 2022.

[25] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25, 2014 - October 29, 2014*, Doha, Qatar, 2014, pp. 1532-1543: Association for Computational Linguistics (ACL).

[26] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2019, June 2, 2019 - June 7, 2019*, Minneapolis, MN, United states, 2019, vol. 1, pp. 4171-4186: Association for Computational Linguistics (ACL).

[27] M. Honnibal and M. Johnson, "An improved non-monotonic transition system for dependency parsing," in *Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, September 17, 2015 - September 21, 2015*, Lisbon, Portugal, 2015, pp. 1373-1378: Association for Computational Linguistics (ACL).

[28] P. Liu, S. Joty, and H. Meng, "Fine-grained opinion mining with recurrent neural networks and word embeddings," in *Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, September 17, 2015 - September 21, 2015*, Lisbon, Portugal, 2015, pp. 1433-1443: Association for Computational Linguistics (ACL).

[29] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective LSTMs for target-dependent sentiment classification," in *26th International Conference on Computational Linguistics, COLING 2016, December 11, 2016 - December 16, 2016*, Osaka, Japan, 2016, pp. 3298-3307: Association for Computational Linguistics, ACL Anthology.