

# Modelling Context with Graph Convolutional Networks for Aspect-based Sentiment Analysis

Maoyuan Zhang  
Hubei Provincial Key Laboratory of  
Artificial Intelligence and Smart  
Learning  
Central China Normal University  
Wuhan, PR China  
zhangmy@mail.ccn.edu.cn

Jieqiong Zhang  
School of Computer Central China  
Normal University  
Wuhan, PR China  
2093913310@qq.com

Lisha Liu  
National Language Resources  
Monitoring&Research Center for  
Network Media  
Central China Normal University  
Wuhan, PR China  
liulisha9819@foxmail.com

**Abstract**—Aspect-based sentiment analysis is a fine-grained natural language processing task that aims to predict a specific target's sentiment polarity in its context. Existing researches mainly focus on the exploration of the interaction between the sentiment polarity of aspects and contexts. Models based on the self-attention mechanism can fully explore the syntactic structure of sentences. In contrast, models based on a convolutional neural network have the ability to make aspects and the semantics of contextual words alignment. These methods all have some limitations; that is, they lack the ability to make full use of syntactic information and long-range word dependencies to carry out relevant syntactic constraints while associating the target's sentiment with the local context. And they are not able to handle affective ambivalence in text. In this paper, we propose a stacked ensemble method for predicting the sentiment polarity by combining a local context embedding and a global graph convolutional network. It uses a Graph Convolutional Network (GCN) to supplement local information to improve the accuracy of the aspect sentiment classifier with revealing multi-level sentiments. Experimental results on three commonly used datasets show that our approach outperforms the state-of-the-art models in the Semeval-2014 dataset.

**Keywords**—sentiment classification, graph convolutional networks, multi-head self-attention, syntactic information, ambivalence handling

## I. INTRODUCTION

Aspect-based sentiment analysis is fine-grained Natural Language Processing (NLP) subtask in sentiment analysis, which differs from traditional sentiment analysis. It aims to determine the sentiment polarity (i.e., negative, neutral, or positive) of the targeted aspect word in the same sentence or document. For example, given a review, “The food was definitely good, but when all was said and done, I just couldn't justify it for the price...”. For two target words, “food” and “price”, we need to infer the sentiment polarities of them are positive and negative, respectively.

Existing Aspect-based sentiment classification (ASC) studies are based on various neural network models, such as Long and Short Term Memory Network (LSTM)[3], Convolutional Neural Network (CNN)[7], Recursive Neural Network (RNN)[21], etc. These models represent a sentence as a sequence of words and sometimes add target words information to supplement the syntactic relations between words. With the introduction of the attention mechanism, the attention model to establish the semantic correlation between the target word and the context is becoming more and more popular. However, attention models have the possibility of erroneous learning of contextual words that are not relevant to a specific target. Multiple complex attention will increase the computational effort and reduce the efficiency[1]. The models

based on CNN and RNN are not enough to learn the hidden features of unrelated words. The self-attention mechanism is an improvement of the attention mechanism, which is better at capturing semantic correlation between contexts. The Multi-Head Self-Attention (MHSA) can mitigate the potential function loss of the target's remote context words. Therefore, recent studies use self-attention models to learn the internal semantic relevance of contextual features and show better performance than other models.

The proposal of Local Context Focus (LCF) design[9] makes us notice that the sentiment polarity of the target word has a significant correlation with the nearby contextual words and eliminates the negative effects of the contextual terms far from the target. The current LCF models only focus on the adequate information of the target's neighbouring context but ignore the word co-occurrence information in the global context and the dependency relation of different types of terms, which can supplement the local context embedding. Based on the local context focus mechanism, we propose the global graph convolution network[8] combined with local context embedding named Local Context with Graph Convolution Network (LC-GCN). LC-GCN makes full use of syntactic information, distinguishing between different types of dependencies and word co-occurrence relation and the ability of the model to capture aspect-specific contextual words improved. Experimental results on three commonly used datasets show that this model plays an essential role in enhancing ASC performance. The main contributions of this paper are as follows:

- We propose a new local context GCN model, which uses global context information to supplement local context information. The model can handle affective ambivalence to accurately predict the sentiment polarity of the targeted aspect based on the self-attention mechanism.
- When learning global context features, a multi-layer graph convolutional network is used to obtain global word co-occurrence information. Syntactic information and word dependencies are used to fully learn multi-level sentiments, finally the self-attention mechanism is combined to enhance the ability of the model to learn global context features.
- The relationship between global word co-occurrence and local context embedding is explored. The important role of long-range word dependencies relation in supplementing the sentiment information of the targeted aspect in the local context is proved.

## II. RELATED WORKS

In recent years, the research of ASC has focused more and more on deep neural network-based learning methods. However, traditional machine learning methods have gradually withdrawn from the stage of ASC tasks due to its tedious functions, low efficiency and poor generality. Here, we mainly introduce several types of deep learning models and review the development and evolution of deep neural networks in aspect level sentiment classification tasks. The attention mechanism aims to emphasize the target-specific context, encode sentences by using the semantic correlation between the context and the target, and generate a sentence representation containing attention weights. The attention mechanism is added to the deep neural network model, which greatly improves the performance of sentiment classification. Wang et al. proposed an attention-based LSTM[3] model, which can combine aspect embedding with word representation. However, it is not suitable for dealing with the interaction between context and aspects that will account for the loss of aspect information. Later, the Target-Dependent LSTM (TD-LSTM)[2] model was proposed, which could obtain the features of the target's left context and the right context, but it still lacked consideration for the target's sentiment information. Tay et al. and Ma et al. proposed an attention network based on the interactive association between targets and context[5]. Huang et al. proposed a model of aspects and sentence association learning and automatically focused on the essential parts of the sentence[6]. Fan et al. proposed a new multi-granularity attention network combining target and context information, which has better performance in capturing semantic interactions between targets and context[10].

In document-level sentiment classification, in order to integrate syntactic information, a graph-based model converts sentence dependency tree into a graph structure, uses GCN[8] or Graph Attention Network (GAT)[22] to aggregate contextual neighbourhood word information, and propagates the information to the target, or capture global word co-occurrence information for sentiment classification. Aspect-Specific Graph Convolutional Networks (ASGCN) propose a novel goal-specific GCN model, which uses syntactic dependency structures in sentences to solve the problem of remote multi-word dependency in aspect-based sentiment classification[16]. For the first time, a novel method to model Sentiment Dependencies with Graph Convolutional Networks (SDGCN)[18] introduces a bidirectional attention mechanism and position coding to model the representation of a specific target between the target and contextual words. The model can capture the sentiment dependencies between many aspects in a sentence. BIGCN[17] creatively proposes a two-layer interactive graph convolutional network to fully integrate the two conceptual levels of syntax and vocabulary.

A contextualized pre-trained language model has been used more and more to improve the performance of ASC. The BERT-PT model[13] proposed by Xu et al., which adopts the post-training method, can improve aspect sentiment analysis performance. The BERT-SPC model[12] proposed by Song et

al. that combines target and context for sentence encoding, and the AEN-BERT model[20] applies an attention mechanism to make semantic interaction between targets and contextual words on this basis. The LCF-BERT[9] and LCFS-BERT[15] models use syntactic relative distance to reduce the influence of target-unrelated words.

## III. METHODOLOGY

In this section, we mainly present the LC-GCN model proposed. Given a contextual sentence sequence  $S$  consists of  $n$  words including targeted aspects, where  $S = \{w_0, w_1, \dots, w_n\}$ . For  $m$  aspects contained in  $S$ ,  $a = \{a_0, a_1, \dots, a_m\}$  ( $m \geq 1$ ) constitutes the targeted aspect sequence, where  $A$  is a subsequence of  $S$ . Figure 1 is the local context graph convolutional network framework.

### A. Context Input Representation

The input sequence of the model requires a series of processing at first, then we use BERT to encode sentence  $S$  and aspect  $A$ . The input format of global context  $G = [CLS]+S+{SEP}+A+{SEP}$ , where  $S$  is the contextual sentence and  $A$  is the aspect word, and  $L = [CLS]+S+{SEP}$  for local context input. MHSA in Transformer of BERT model is similar to multiple convolution cores in CNN, and can extract information from different perspectives. Therefore, when BERT encodes a token, it also makes use of the tokens in its context. We can better capture the semantic relationship between contextual sentences and aspects.

### B. Word Embeddings

We adopt two independent BERT embedding layers to model local context features and global context features, then feed local context and global context into context embedding to obtain local context vectors  $V^l$  and global context vectors  $V^g$ . Where Attention\_Mask can mask words that are not selected in global context. Local and global context features are obtained by using MHSA. Next, we adopt the context-feature dynamic masking (i.e. CDM)[9] technology to mask the context features that are less semantically related to the targeted aspect, so as to retain the feature focusing on the local context words. In order to capture the word co-occurrence information in the global context and the sentiment dependency relation of different types of aspects in the sentence. We add multi-layer convolution structure after normalizing the global context vector, so that the targeted aspect can interact with its context to obtain more accurate information.

### C. Graph Convolutional Networks Layer

GCN belongs to Graph Neural Network (GNN) and is a Graph Neural Network which adopts convolution operation. It can directly manipulate the multi-layer neural network of graphs and deduce the embedding vectors of nodes based on the neighboring nodes and their properties[8]. GCN can associate syntactically related words to the targeted aspect by the syntactic dependency tree of the sentence, and learn the long-range multi-word dependency and syntactic information by the GCN layer.

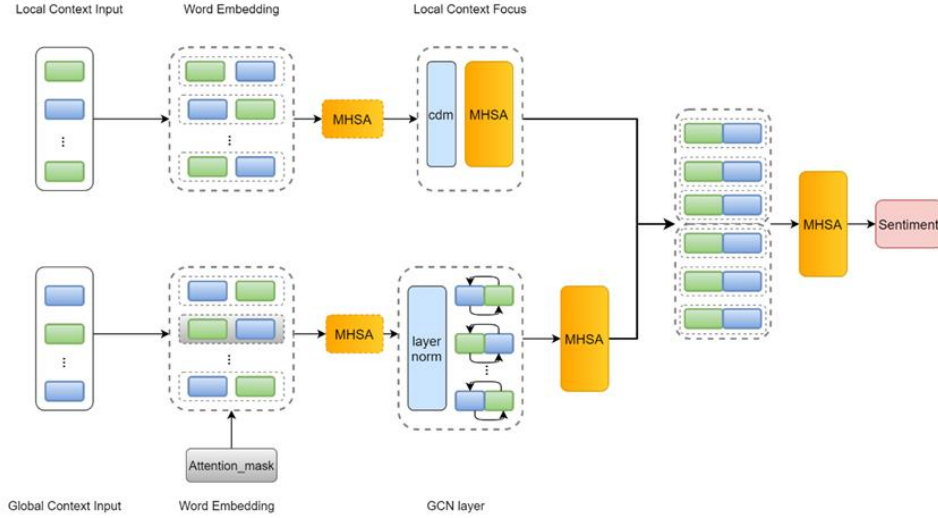


Fig. 1. The architecture of LC-GCN

Given a graph  $G=(V,E)$  with  $n$  nodes that represents a sentence containing  $n$  words, where  $V$  is a node set and  $E$  is an edge set. According to Kipf and Welling's self-looping idea, each node is connected to itself, and an adjacency matrix  $A \in \mathbb{R}^{n \times n}$  is generated by enumerating whether each vertex pair is adjacency. For  $L$ -layer GCN,  $l \in [1,2, \dots, L]$ ,  $h_i^l \in \mathbb{R}^d$  ( $d$  represents the embedding dimension) represents the  $l$ -th layer output of node  $i$ . So the update of node  $i$  after graph convolution is:

$$h_i^l = \text{relu}(\sum_{j=1}^n A_{ij} W^l h_j^{l-1} + b^l) \quad (1)$$

Where  $W^l$  is the linear transformation weight vector, and  $b^l$  is the offset vector. They are both training parameters. Each layer of convolution of the GCN can obtain information about the immediate neighbors of each node. We apply multi-level graph convolution to the syntactic dependency tree of global context sentences and let them interact to refine the sentence representation:

$$h_i^l = h_i^l \oplus h_i^{l-1} \quad (2)$$

In this way, the influence of multi-level sentiments of related long-range words on the targeted aspect in local context can also be reinforced. The final representation is obtained with GELU activation function and initial global vector:

$$O^g = \text{Gelu}(H^L) + V^g \quad (3)$$

$$H^L = \{h_1^L, h_2^L, \dots, h_n^L\}, h_i^L \in \mathbb{R}^d$$

GELU is a high performance neural network activation function. It introduces the idea of random regularity into the activation. It is a probabilistic description of neuron input.[24]

In order to rebalance the feature distribution and learn the semantic associations within the global context, a new MHSA is deployed to learn global context features  $O^g$ .

#### D. Local Context Focus Layer

The LCF model adopts local context focus mechanism to extract local context features. In order to reduce the influence of remote irrelevant words that are far away from the targeted aspect, we use CDM to focus local context features.

The semantic relative distance (SRD)[15] between words can be calculated according to the position distance of nodes in the syntactic dependency of a sentence. For the targeted aspect consisting of multiple words, the SRD of contextual words and the targeted aspect is the average distance between the center position of the targeted aspect and the center position of context words. We select contextual words by setting the threshold  $\alpha$ .

CDM masks these less semantic features according the SRD between the feature word and the targeted aspect is greater than the given threshold, so as to retain more semantic context words. The mask vector  $V_i$  of the  $i$ -th contextual word in the local context  $V^l$  is:

$$V_i = \begin{cases} 0, & SRD_i > \alpha \\ 1, & SRD_i \leq \alpha \end{cases} \quad (4)$$

Mask matrices  $M = [V_1, V_2, \dots, V_n]$ , and the local context  $O^l = V^l \odot M$  is obtained. Similar to the global context feature, an MHSA is used to relearn the local context feature.

#### E. Output Layer

In order to fully integrate local context features and global context features, we connect  $O^l$  and  $O^g$  to feed into a feed-forward neural network. Because the exact same feed forward network is applied independently to the words at each position of an input sentence. Therefore, it is also called point-wise feed-forward neural network. Then, MHSA is used for adaptive fusion:

$$O^{out} = [O^g; O^l] \quad (5)$$

$$O_{ffn} = (W_1 O^{out} + b_1) W_2 + b_2 \quad (6)$$

$$O_{fuse} = MHS(A(O_{ffn})) \quad (7)$$

$W_1 \in R^{d_h \times 2d_h}$ ,  $b_1 \in R^{d_h}$  and  $W_2 \in R^{2d_h \times d_h}$ ,  $b_2 \in R^{d_h}$  are the weights and bias vector respectively,  $d_h$  is the hidden dimension; the order is linear transformation, then Relu nonlinear transformation, and linear transformation, then an MHS learns features and output  $O_{fuse}$  interactively. Finally, a Softmax layer is applied to predict the sentiment polarity:

$$X_{pool} = POOL(O_{fuse}) \quad (8)$$

$$\gamma = Softmax(X_{pool}) = \frac{exp(X_{pool})}{\sum_{k=1}^C X_{pool}} \quad (9)$$

where C is the number of sentiment classes,  $\gamma$  is the sentiment polarity predicted by LC-GCN model.

We use the cross-entropy loss function with  $L_2$  regularization as a loss function for the model:

$$\mathcal{E} = \sum_{i=1}^C \hat{y}_i \log y_i + \lambda \sum_{\theta \in \Theta} \theta^2 \quad (10)$$

Where  $\lambda$  is the  $L_2$  regularization parameter,  $\hat{y}_i$  is the predicted label corresponding to  $y_i$ , and  $\Theta$  is the parameter set of the LC-GCN model.

#### IV. EXPERIMENTS

##### A. Datasets and Experimental Settings

We evaluate and compare the proposed ASC models on three benchmark datasets: Restaurant, Laptop and Twitter, to prove the effectiveness of LC-GCN. Each sample sentence in the datasets provides masked aspects with their sentiment polarities which are categorized into neutral (*Neu*), positive (*Pos*) and negative (*Neg*). TABLE I. describes the details of three datasets, where the Restaurant and Laptop datasets are taken from SemEval-2014 Task 4 challenge [11], and a common ACL 14 twitter social dataset are introduced by Dong et al.

TABLE I. NUMBER OF SENTENCES BY POLARITY IN THREE DATASETS

Dataset	Train			Test		
	Pos	Neg	Neu	Pos	Neg	Neu
Restaurant	2164	807	637	728	196	196
Laptop	994	870	464	341	128	169
Twitter	1561	1560	3127	173	173	346

We adopt accuracy and Macro-F1 score to evaluate the performance of all models. In order to obtain the best performance of the LC-GCN model, we set the learning rate at  $2 \times 10^{-5}$ , dropout rates at 0.1, and  $L_2$  regularization at  $1 \times 10^{-5}$ . The hidden dimension and embedding dimension are set to 768. For other GCN models, 300-dimensional pre-trained Glove vectors[23] are used to initialize word embedding. So their hidden dimension is set to 300, the learning rate is set to 0.001, the  $L_2$  regularization is same with the above, and the number of GCN layers is set to 2. Moreover, all models utilize the Adam optimizer[14].

##### B. Models for Comparison

In order to comprehensively evaluate the performance of our model, we compare the LC-GCN model with multiple

baseline models including state-of-the-art GCN-based models and various BERT-based models on three data sets, as listed below:

**TD-LSTM**(Tang et al.,2016)[2] constructs the connection between aspects and its context, and then select the relevant parts of the context to infer the sentiment polarity of the targeted aspect in the sentence. The basic idea of this model is to model the left context and the right context with the targeted aspect so that the context in both directions can be used as a feature representation for sentiment classification.

**ATAE-LSTM**(Wang et al.,2016)[4] implements an attention mechanism based LSTM on aspect-based sentiment classification. This mechanism allows the model to focus on different parts of a sentence with different aspects. Meanwhile, ATAE-LSTM proposes two different ways to incorporate aspect information. One is to concatenate the aspect vector with the hidden level representation of the sentence to calculate the weight of attention. Another directly splices aspect vector with input word vector.

**IAN**(Ma et al.,2017)[5] models aspects and context separately and use attention mechanism to connect them. When modeling aspects, the attention mechanism is added with the context as the Query vector. On the contrary, when modeling the context, the attention mechanism is added with aspects as the Query vector. The resulting aspects and context representation merge their interaction information.

**AOA**(Huang et al.,2018)[6] learns aspects and sentence representations together and automatically focus on the important parts of the sentence. AOA can capture the interaction between aspects and contextual sentences with multiple attention layers.

**MGAN**(Fan et al.,2018)[10] uses different grained attention to capture word-level interaction between aspects and sentence.

**ASGCN**(Zhang et al.,2019)[16] builds a GCN over the dependency tree of a sentence to exploit syntactical information and word dependencies.

**SDGCN**(Zhao et al.,2019)[18] firstly introduces bidirectional attention mechanism with position encoding to model aspect-specific representation between each aspect and its context words, then employs GCN over the attention mechanism to capture the sentiment dependencies between different aspects in one sentence.

**BIGCN**(Zhang et al.,2020)[17] employs a global lexical graph to encode the corpus level word co-occurrence information, and builds a concept hierarchy on both the syntactic and lexical graphs for differentiating various types of dependency relations or lexical word pairs. Finally, a bi-level interactive graph convolution network designed to fully exploit these two graphs.

**BERT-PT**(Xu et al.,2018)[13] proposes a post-training approach on the BERT pre-trained model to enhance the quality of word representations to the end-task, the method also can be adapted to ASC task.

**BERT-SPC**(Devlin et al.,2019)[12] proposes a pre-trained BERT model for sentence-pair classification which constructs the input sequence combining the context and aspects as sentence-pair.

TABLE II. COMPARISON RESULTS OF OUR MODEL IN TERMS OF ACCURACY(%) AND F1 SCORES WITH THE STATE-OF-THE-ART METHODS.

(We use “-” denotes the unreported experimental results, “\*” means the results of models are obtained by our implementations according to the methodology published in the paper. Others are retrieved from the previous papers.)

	Model	Restaurant		Laptop		Twitter	
		Acc	F1	Acc	F1	Acc	F1
<b>Baselines</b>	TD-LSTM	75.63	-	68.13	-	70.8	69
	ATAE-LSTM	77.32	66.57	69.14	63.18	69.65	67.40
	IAN	79.26	70.09	72.05	67.38	72.50	70.81
	AOA	79.97	70.42	72.62	67.52	72.30	70.20
	MGAN	81.25	71.94	75.39	72.47	72.54	70.81
<b>GCN models</b>	ASGCN	80.86	72.19	75.55	71.05	72.15	70.40
	SDGCN	83.57	76.47	81.35	78.34	-	-
	BIGCN	81.97	73.48	74.59	71.84	74.16	73.35
<b>BERT models</b>	BERT-PT	84.95	76.96	78.07	75.08	-	-
	BERT-SPC	83.57*	75.10*	77.59*	72.14*	74.57*	73.07*
	AEN-BERT	83.12	73.76	79.93	76.31	74.71	73.13
	LCF-BERT	84.55*	77.42*	77.74*	72.72*	75.00*	73.54*
	LCFS-BERT	85.54*	78.19*	79.94*	75.21*	73.99*	72.79*
	<b>LC-GCN</b>	<b>87.23</b>	<b>81.32</b>	<b>80.09</b>	<b>76.78</b>	<b>75.58</b>	<b>73.26</b>

**AEN-BERT** (Song et al., 2019)[20] proposes an Attentional Encoder Net-work which eschews recurrence and employs attention based encoders for the modeling between context and the targeted aspect.

**LCF-BERT** (Zeng et al., 2019)[9] proposed a Local Context Focus (LCF) mechanism for ASC based on MHSA which utilizes the CDM and Context Features Dynamic Weighted (CDW) layers to pay more attention to the local context words.

**LCFS-BERT** (Phan et al., 2020)[15] constructs an end-to-end aspect-based sentiment analysis method based on LCF-BERT, which can make full use of grammatical information and use self-attention mechanism to fully mine syntactic structure.

### C. Results analysis

TABLE II. demonstrates that our proposed LC-GCN model has the relatively better performance than other models in three datasets, especially on the Restaurant dataset. All BERT models achieve impressive better performance than GCN models. In the Restaurant dataset and Laptop dataset, BERT models improve the performance by 2-5%. It show us the BERT-shared layer powerful ability to extract and learn context features. But the results on the Twitter dataset are not significantly improved. Results of all models are worse in the Twitter dataset than in the Restaurant dataset and Laptop dataset. Because the Twitter dataset is a social dataset with lots of misspelled words and unknown symbols, and sentences from there are less grammatical. The experimental results indicate that the LC-GCN obtains considerable performance on three datasets, compared with GCN models and BERT models. Because the local context focus mechanism plays an important role in local feature extraction

and the Graph Convolution Network layer is better at capturing long-range word dependencies which can improve the ability of our model to extract global context features. Besides, it also demonstrates the GCN has a complementary effect on MHSA.

### D. Ablation Study

To further study the improvement that LC-GCN brings to the ASC tasks, we conducted ablation experiments on LC-GCN. The results are listed in TABLE III. We examined the performance without GCN, without CDM and without FIL respectively. It can be seen that removal of the GCN layer leads to the performance degradation of the model, which proves that the GCN layer plays an important role in improving the capability of aspect sentiment classification of the LC-GCN model. And the performance of the model without CDM is worse than the performance of the model without GCN, which could be concluded that CDM contributes more in LC-GCN than GCN. Moreover, when the model lacks an adaptive fusion layer to interact the local context features and the global context features, the results are better than the other two ablated models(such as LC-GCN w/o GCN and LC-GCN w/o CDM) but worse than our proposed LC-GCN model. However, the performance improvement of LC-GCN is not significant in twitter dataset. After ablation study, we found that results of LC-GCN w/o CDM in twitter dataset are not well as expected. We suspected the effect of global context information on the targeted aspect in local context is not very positive in twitter dataset that is not so sensitive to long-range word dependencies.

TABLE III. ABLATION STUDY RESULT (%). “w/o” MEANS “WITHOUT”. “GCN” DENOTES THE GRAPH CONVOLUTIONAL NETWORK LAYER, “CDM” DENOTES THE LOCAL CONTEXT FOCUS LAYER, “FIL” DENOTES LOCAL CONTEXT FEATURES AND GLOBAL CONTEXT FEATURES INTERACTIVE LEARNING LAYER WITH MHSA AND FFN (FEED FORWARD NETWORK).

Model	Restaurant		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
LC-GCN	<b>87.23</b>	<b>81.32</b>	<b>80.09</b>	<b>76.78</b>	<b>75.58</b>	<b>73.26</b>
w/o GCN	85.36	77.57	78.21	74.05	74.13	72.97
w/o CDM	84.64	74.72	78.68	73.20	73.55	72.38
w/o FIL	86.34	79.96	79.62	76.78	75.14	73.08

TABLE IV. SEVERAL CASES OF ASC TASK FROM ASGCN, LCFS-BERT AND LC-GCN ON TEST DATASET, ALONG WITH THEIR PREDICTION P(ASGCN), P(LCFS-BERT) AND P(LC-GCN), RESPECTIVELY.

Sentence	Aspect	P(ASGCN)	P(LCFS-BERT)	P(LC-GCN)	Polarity
A beautiful atmosphere, perfect for drinks and/or appetizers.	drinks	positive	positive	neutral	0
	appetizers	positive	positive	neutral	0
From the speed to the multi touch gestures this operating system beats Windows easily.	Windows	negative	positive	negative	-1
I can understand the prices if it served better food, like some Chinese restaurants in midtown/uptown area.	prices	positive	positive	negative	-1
	food	positive	negative	negative	-1

### E. Case Study

We provided a case study with several testing examples to understand our LC-GCN model more intuitively. TABLE IV. shows the prediction of these examples in ASGCN, LCFS-BERT and LC-GCN, respectively, along with their ground truth labels. “0”, “1”, “-1” means neutral, positive and negative, which represents three sentiment polarities.

The first and third sentences both have two aspects within sentence, which makes it difficult for the model to predict the sentiment polarity of aspects accurately. It can be seen that ASGCN and LCFS-BERT almost make wrong predictions in these two sentences. Because they focus too much on local context features but ignore the sentiment tendency of the global context. Considering this association of global and local context information, our LC-GCN can give correct prediction for each aspect. In addition, other models always are weak in predicting neutral sentiment, but LC-GCN can leverage global context information to counteract the negative influence of local context, make correct “neutral” prediction. It proves that multi-level sentiments information extracted by GCN can handle affective ambivalence. The second sentence clearly depends on its syntactic structure and word dependency relation to make a correct prediction. So our LC-GCN correctly handles all the three examples, which indicates that local context embedding with GCN effectively integrates semantic feature of global context into sentiment

dependency information of local context for the targeted aspect.

### V. CONCLUSION AND FUTURE WORKS

This paper proposed a novel network to supplement BERT with GCN for aspect-based sentiment classification. We combine a local context embedding and a global graph convolutional network using MHSA mechanism for adaptive fusion, which can handle affective ambivalence in text. The experimental results indicated that GCN has a significantly effect in extracting multi-level sentiments in global context by integrating syntactical information and long-range word dependencies. Meanwhile, the important correlation between local feature extraction and co-occurrence information in global context has been proved. The LC-GCN model outperforms state-of-the-art performance on three ASC datasets. In the future, we can incorporate domain knowledge with GCN to develop more syntax-based contextualized embedding, and we will apply LC-GCN techniques on other NLP tasks, such as text classification and semantic disambiguation. Moreover, we can explore in great depth the information of the multi-level graph convolution to generate finer characterizations of sentiments as well as emotions involved, or to obtain multi-level fine-scaled sentiments as well as different types of emotions.

### REFERENCES

- [1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.
- [2] Tang, D.; Qin, B.; Feng, X.; Liu, T. Effective LSTMs for Target-Dependent Sentiment Classification. In *Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Osaka, Japan, 11–17 December 2016; pp. 3298–3307.
- [3] Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation, Inc.: Montréal, QC, Canada, 2014; pp. 3104–3112.
- [4] Wang, Y.; Huang, M.; Zhao, L. Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, TX, USA, 1–5 November 2016; pp. 606–615.
- [5] Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. In *Proceedings of IJCAI*, pages 4068–4074.
- [6] Binxuan Huang, Yanglan Ou and Kathleen M. Carley. 2018. Aspect Level Sentiment Classification with Attention-over-Attention Neural Networks. *Arxiv preprint arXiv:1804.06536*.
- [7] Wei Xue and Tao Li. 2018. Aspect based sentiment analysis with gated convolutional networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2514–2523.
- [8] Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*.
- [9] Biqing Zeng, Heng Yang, Ruyang Xu, Wu Zhou, and Xuli Han. 2019. Lcf: A local context focus mechanism for aspect-based sentiment classification. *Applied Sciences*, 9(16):3389.
- [10] Feifan Fan, Yansong Feng, and Dongyan Zhao. 2018. Multi-grained attention network for aspect-level sentiment classification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3433–3442.
- [11] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect-based sentiment analysis. In *Proceedings of [15] the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35.
- [12] Song, Y.; Wang, J.; Jiang, T.; Liu, Z.; Rao, Y. Attentional Encoder Network for Targeted Sentiment Classification. *arXiv 2019*, arXiv:1902.09314.
- [13] Xu, H.; Liu, B.; Shu, L.; Philip, S.Y. BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, MN, USA, 2–7 June 2019; pp. 2324–2335.
- [14] Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv 2014*, arXiv:1412.6980.
- [15] Phan, Minh & Ogunbona, Philip. (2020). Modelling Context and Syntactical Features for Aspect-based Sentiment Analysis. 3211-3220. 10.18653/v1/2020.acl-main.293.
- [16] Chen Zhang, Qiuchi Li, and Dawei Song. 2019. Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *EMNLP-IJCNLP*, pages 4567–4577.
- [17] Zhang, Mi & Qian, Tiejun. (2020). Convolution over Hierarchical Syntactic and Lexical Graphs for Aspect Level Sentiment Analysis. 3540-3549. 10.18653/v1/2020.emnlp-main.286.
- [18] Zhao, Pinlong & Hou, Linlin & Wu, Ou. (2019). Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification. *Knowledge-Based Systems*. 193. 105443. 10.1016/j.knosys.2019.105443.
- [19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- [20] Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao. 2019. Targeted sentiment classification with attentional encoder network. In *International Conference on Artificial Neural Networks*, pages 93–103. Springer.
- [21] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- [22] Chen, Xiao & Sun, Changlong & Wang, Jingjing & Li, Shoushan & Si, Luo & Zhang, Min & Zhou, Guodong. (2020). Aspect Sentiment Classification with Document-level Sentiment Preference Modeling. 3667-3677. 10.18653/v1/2020.acl-main.338.
- [23] Pennington, J.; Socher, R.; Manning, C. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
- [24] D. Hendrycks and K. Gimpel, “Bridging nonlinearities and stochastic regularizers with Gaussian error linear units,” *ArXiv Prepr. ArXiv 160608415*, 2016, doi:https://arxiv.org/abs/1606.08415v2.