

Jointly Learning Aspect-Focused and Inter-Aspect Relations with Graph Convolutional Networks for Aspect Sentiment Analysis

Bin Liang^{1,2}, Rongdi Yin^{1,2}, Lin Gui^{3*}, Jiachen Du^{1,2}, Ruifeng Xu^{1,2,4*}

¹School of Computer Science and Technology,

Harbin Institute of Technology (Shenzhen), Shenzhen, China

²Joint Lab of HITSZ and China Merchants Securities, Shenzhen, China

³Department of Computer Science, University of Warwick, Coventry, U.K

⁴Peng Cheng Laboratory, Shenzhen, China

bin.liang@stu.hit.edu.cn, rongdiyin@gmail.com

lin.gui@warwick.ac.uk, Jacobvan199165@163.com

xurui Feng@hit.edu.cn

Abstract

In this paper, we explore a novel solution of constructing a heterogeneous graph for each instance by leveraging aspect-focused and inter-aspect contextual dependencies for the specific aspect. Based on it, we propose a novel graph-aware model with Interactive Graph Convolutional Networks (InterGCN) for aspect sentiment analysis. Specifically, an ordinary dependency graph is first constructed for each sentence over the dependency tree. Then we refine the graph by considering the syntactical dependencies between contextual words and aspect-specific words to derive the aspect-focused graph. Subsequently, the aspect-focused graph and the corresponding embedding matrix are fed into the aspect-focused GCN to capture the key aspect and contextual words. Besides, to interactively extract the dependencies between the aspect words and other aspects, an inter-aspect GCN is adopted to model the representations learned by aspect-focused GCN based on the inter-aspect graph. Hence, the model can be aware of the significant contextual and aspect words when interactively learning the sentiment features for a specific aspect. Experimental results on four benchmark datasets illustrate that our proposed model outperforms state-of-the-art methods and substantially boosts the performance in comparison with BERT.

1 Introduction

Aspect sentiment analysis is a fine-grained sentiment analysis task, which aims to identify the sentiment polarity (e.g. positive, negative, or neutral) towards a given aspect (term) in a sentence. For example, given the aspects: *food* and *service*, and a sentence of review: *great food but the service is dreadful*, the sentiment polarity of aspect *food* is positive, while for the aspect *service* is negative. That is, in the task of aspect sentiment analysis, we need to discriminate sentiment polarities according to different aspects. The main challenge is that some aspects may contain no explicit sentiment expression.

To vividly illustrate the challenge, we give examples shown in Figure 1, where the key contextual words and corresponding aspects are highlighted in the instances paired with their polarity labels. In Figure 1(a), there is a multiple words aspect (i.e. “*soup for the udon*”). We could readily resolve that the aspect word “*soup*” is the key aspect word of this aspect. Thus, syntactical dependencies between this aspect word and the contextual words need to be attended for predicting aspect-specific sentiment. In Figure 1(b), aspect “*toppings*” and “*place*” are mentioned simultaneously in the sentence. There is a sufficiently clear positive sentiment word (“*great*”) for aspect “*toppings*”, while for aspect “*place*”, which contains no sentiment expression, the sentiment polarity can also be identified thanks to the inter-aspect relations between aspect “*toppings*” and “*place*”. Hence, both aspect-focused and inter-aspect contextual relations should be considered for improving the performance of aspect sentiment analysis.

Recently, with the development of deep learning technique, many neural network-based methods achieve promising performance in aspect sentiment analysis (Wang et al., 2016a; Tang et al., 2016a;

* Corresponding Author

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>

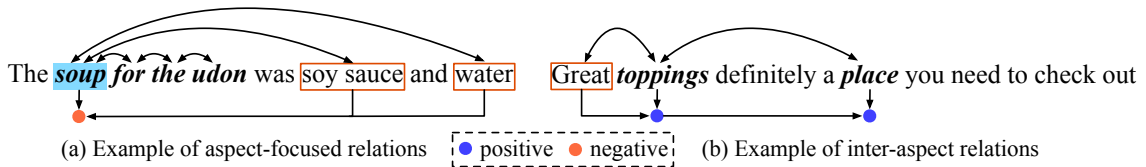


Figure 1: Examples of the contextual relations of different aspect words within an aspect and the sentiment relations of different aspects in a sentence.

Chen et al., 2017; Wang et al., 2018; Zheng et al., 2020). Subsequently, attention-based neural models are widely used in this task, which can enforce the model to focus on the given aspect (Wang et al., 2016b; Tang et al., 2016b; He et al., 2019). In most previous methods, however, they generally embed aspect information into the sentence representation to learn the pertinent sentiment features for the specific aspect, which leads to a lack of capturing the inter-aspect sentiment relations for a specific aspect. Analogously, most existing graph network-based model merely consider the syntactical dependencies between the specific aspect and the context (Zhang et al., 2019; Huang and Carley, 2019; Sun et al., 2019), which is insufficient to focus on which contextual dependencies along with aspect-specific words are essential for the specific aspect and also largely ignore the sentiment relations between different aspects in the sentence. Since intuitively, the role of distinct aspect word is different in deriving aspect expression. Besides, there are intricate sentiment relations among different aspects in many instances.

In this paper, we explore a novel solution to construct heterogeneous graphs of sentences via enriching the contextual syntactical dependency representations of the key aspect words and leveraging the mutual sentiment relations between different aspects in the context. Based on it, an Interactive Graph Convolutional Networks (InterGCN) model is proposed to leverage the sentiment dependencies of the context. Here, the syntactical information from neighbors of each node is aggregated to derive the graph embeddings, so as to extract both aspect-focused and inter-aspect sentiment information for predicting aspect-specific sentiment polarity. The main contributions of our work can be summarized as follows:

- We explore a novel solution to construct the graph for each instance, in which both aspect-focused and inter-aspect syntactical dependencies are introduced.
- An Interactive Graph Convolutional Networks model is proposed to derive aspect-specific sentiment features by interactively extracting the sentiment relations within aspect words and across different aspects in the context.
- Experimental results on four benchmark datasets show that the proposed model achieves the state-of-the-art performance in aspect sentiment analysis.

2 Related Work

Some early works mostly use machine learning algorithms to capture the sentiment polarity based on rich features about content and syntactic structures in aspect sentiment analysis (Pang et al., 2008; Jiang et al., 2011; Kiritchenko et al., 2014). Recently, deep learning models have achieved promising performance in aspect sentiment analysis (Tang et al., 2016a; Wang et al., 2016b; Tang et al., 2016b; Chen et al., 2017; Ma et al., 2017; Xue and Li, 2018; Li et al., 2019; Liang et al., 2019). The majority of current approaches attempt to pay more attention to the specific aspect based on attention mechanism. (Wang et al., 2016b) exploited attention mechanism to capture the contextual representations via paying attention to the key parts of the sentence according to the given aspect. (Tang et al., 2016b) proposed an attention-based memory network to store contextual words and conducted multi-hop attention to derive the sentiment representation for the aspect. (Chen et al., 2017) utilized a weighted-memory mechanism to produce a tailor-made memory for different opinion aspects based on memory network. In addition, (Xue and Li, 2018) utilized a gated CNN to selectively model the sentiment features according to the given aspect. (Du et al., 2019) adopted a capsule network to construct vector-based feature representation and cluster features by an EM routing algorithm. (Majumder et al., 2018) considered the neighboring aspect-related information for the aspect-specific sentiment analysis with memory networks.

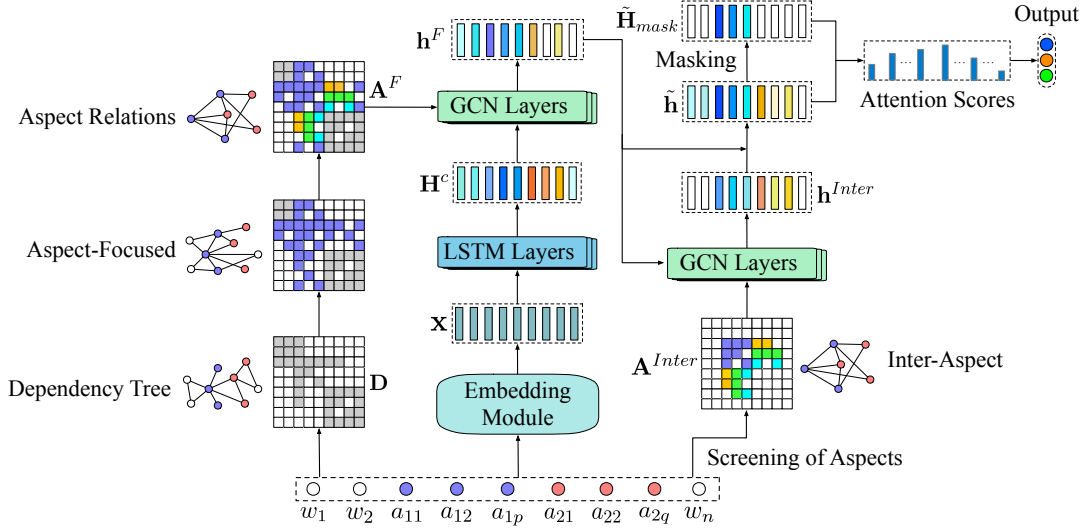


Figure 2: The architecture of the proposed Interactive Graph Convolutional Networks.

Graph convolutional network (GCN) has achieved promising performance in many NLP tasks (Kipf and Welling, 2017; Zhang et al., 2018; Huang et al., 2019; Yao et al., 2019). In aspect sentiment analysis, (Zhang et al., 2019) exploited GCN to capture syntactical information and word dependencies for the specific aspect over the dependency tree of a sentence. (Sun et al., 2019) proposed a GCN model over the dependency tree of the sentence to enhance the feature representations of aspects learned by a Bi-directional LSTM (Bi-LSTM). In addition, to develop the merit of BERT (Devlin et al., 2019), a GCN model based on selective attention was proposed to extract and aggregate the most important contextual features for the aspect representation (Hou et al., 2019). The above GCN-based models, however, neither considered the specific aspect when constructing the graph of the sentence nor extracted inter-aspect sentiment relations for the specific aspect. To this end, based on the merit of GCN in aspect sentiment analysis, we explore a novel solution of constructing syntactical dependency graph for a sentence according to the specific aspect and propose an Interactive Graph Convolutional Networks (InterGCN) model to extract both aspect-focused and inter-aspect sentiment features for the specific aspect.

3 Proposed Approach

As demonstrated in Figure 2, the architecture of the proposed InterGCN model mainly contains two components: 1) aspect-focused graph convolutional networks, which aims to extract the aspect-specific sentiment features based on our novel syntactical dependency graph of the sentence, and 2) inter-aspect graph convolutional networks, which is designed to derive the sentiment relations between different aspects. The feature representations captured from these two components are interactively combined to produce the sentiment features for the specific aspect. We first assume that there is a sentence with n words and two aspects, i.e. $s = \{w_1, w_2, \dots, a_{11}, a_{12}, \dots, a_{1p}, \dots, a_{21}, a_{22}, \dots, a_{2q}, \dots, w_n\}$, where w_i represents the i -th contextual word and a_{ij} represents the j -th word of aspect i . Each instance contains a sentence and one or more aspects corresponding to different sentiment polarities (*Positive*, *Negative*, or *Neutral*), and each aspect may consist of single or multiple words. The aim of aspect sentiment analysis is to predict the sentiment polarity over a given aspect in a sentence.

3.1 Embedding Module

In our InterGCN model, each word embedding is a distributed representation of a word in the sentence, which is retrieved from the embedding lookup table $\mathbf{V} \in \mathbb{R}^{m \times |N|}$ according to the word index, where $|N|$ is the vocabulary size. And thus, for a sentence with n words, we can get the corresponding embedding matrix $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, where $\mathbf{x}_i \in \mathbb{R}^m$ is the word embedding of w_i , m is the dimension of word vectors. We exploit the pre-trained word embeddings GloVe (Pennington et al., 2014) and BERT (Devlin et al., 2019) to initialize word vectors, and fine-tune them during the training process.

3.2 Producing Ordinary Graphs over Dependency Tree

Inspired by the previous GCN-based works (Zhang et al., 2019; Sun et al., 2019), we first produce an ordinary dependency graph for each input sentence over the dependency tree¹:

$$D_{i,j} = \begin{cases} 1 & \text{if } i = j \text{ or } w_i, w_j \text{ in the dependency tree of the sentence} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

After that, an adjacency matrix $\mathbf{D} \in \mathbb{R}^{n \times n}$ is derived via the dependency tree of the input sentence.

3.3 Refining Graphs for Specific Aspect

To highlight the specific aspect from the contextual words and capture aspect-focused enhanced dependency graph for the sentence, we refine the graph via computing a relative position weight for each element of the adjacency matrix according to the specific aspect:

$$W_{i,j}^F = \begin{cases} 1 & \text{if } w_i \in \{a_i^s\} \text{ and } w_j \in \{a_i^s\} \\ 1/(|j - p^s| + 1) & \text{if } w_i \in \{a_i^s\} \\ 1/(|i - p^s| + 1) & \text{if } w_j \in \{a_i^s\} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $|\cdot|$ is an absolute value function, p^s is the beginning position of the specific aspect, $\{a_i^s\}$ is the word set of the specific aspect. And thus, we can capture the relative dependencies between words of the specific aspect and other contextual words.

To augment the syntactical dependencies of contextual words and produce the relations between aspect and contextual words, here, we integrate the aspect-focused weights and the ordinary dependency graph to derive an aspect-focused syntactical dependency adjacency matrix:

$$G_{i,j} = \begin{cases} D_{i,j} + D_{i,j} * W_{i,j}^F & \text{if } D_{i,j} = 1 \\ W_{i,j}^F & \text{otherwise} \end{cases} \quad (3)$$

Intuitively, as mentioned above, some aspects might not signal distinct sentiment expression in the context. That is, the aspect-focused syntactical dependencies derived by the aspect might be insufficient for identifying the accurate sentiment relations, since the sentiment dependency derivation of those aspects should be provided with the help of other aspects. Thus to leverage the connections of multiple aspects in the sentence, we further refine the aspect-focused graph via incorporating relative graphs from other aspects into the aspect-focused adjacency matrix:

$$\mathbf{A}^F = \mathbf{G} + \frac{1}{l} \sum_{a \in \{a_i^o\}} (\alpha * \mathbf{G}_a) \quad (4)$$

$$\alpha = 1/(|p^o - p^s| + 1) \quad (5)$$

where $\{a_i^o\}$ is the word set of length l of other aspects, and the p^o for each $a \in \{a_i^o\}$ denotes the beginning position of the other aspect. The procedure of generating the adjacency matrix for each sentence via focusing on the specific aspect is depicted in Algorithm 1. Here, to enrich the information of dependencies for the input sentence, we construct the adjacency matrix in un-directional, i.e. $A_{i,j}^F = A_{j,i}^F$.

3.4 Constructing Inter-Aspect Graphs

According to Figure 1, sentiment polarities of some aspects require to be predicted through the sentiment relations between others in the sentence. Hence, we screen the aspects from the sentence and construct

¹In this work, we use spaCy toolkit for producing the dependency tree of the input sentence: <https://spacy.io/>.

Algorithm 1: The procedure of constructing the aspect-focused adjacency matrix for a sentence

Input: a sequence of words $s = \{w_1, w_2, \dots, w_n\}$; a set of aspect-specific words $a^s = \{a_1^s, a_2^s, \dots, a_p^s\}$; a set of other aspects words $a^o = \{a_1^o, a_2^o, \dots, a_q^o\}$; the dependency tree of the sentence $dependency(s)$

```
1 for  $i = 1 \rightarrow n; j = 1 \rightarrow n$  do
2    $\triangleright$  Producing the ordinary graph over dependency tree
3   if  $dependency(w_i, w_j) \in dependency(s)$  or  $i = j$  then
4      $D_{i,j} \leftarrow 1$ 
5   else
6      $D_{i,j} \leftarrow 0$ 
7    $\triangleright$  Refining the adjacency matrix of the graph for the specific aspect
8   if  $w_i \in a^s$  and  $w_j \in a^s$  then
9      $W_{i,j}^F \leftarrow 1$ 
10  else if  $w_i \in a^s$  then
11     $W_{i,j}^F \leftarrow 1/(|j - p^s| + 1)$ 
12  else if  $w_j \in a^s$  then
13     $W_{i,j}^F \leftarrow 1/(|i - p^s| + 1)$ 
14  else
15     $W_{i,j}^F \leftarrow 0$ 
16  if  $D_{i,j} = 1$  then
17     $G_{i,j} \leftarrow D_{i,j} + D_{i,j} * W_{i,j}^F$ 
18  else
19     $G_{i,j} \leftarrow W_{i,j}^F$ 
20  $\triangleright$  Leveraging the relations between multiple aspects corresponding to the specific aspect
21 for  $a \in a^o$  do
22    $\mathbf{A}^F \leftarrow \mathbf{G} + \frac{1}{q} \sum_{a \in a^o} ([1/(|p^o - p^s| + 1)] * \mathbf{G}_a)$ 
23 for  $i = 1 \rightarrow n; j = 1 \rightarrow n$  do
24    $A_{i,j}^F \leftarrow A_{j,i}^F$ 
```

an inter-aspect adjacency matrix for these aspects to derive the contextual sentiment dependencies of these aspects:

$$A_{i,j}^{Inter} = \begin{cases} 1 + 1/(|j - p^s| + 1) & \text{if } w_i \in \{a_i^s\} \text{ and } w_j \in \{a_i^o\} \\ 1 + 1/(|i - p^s| + 1) & \text{if } w_j \in \{a_i^s\} \text{ and } w_i \in \{a_i^o\} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Analogously, to capture the interactive dependencies between multiple aspects in the sentence, we also construct the inter-aspect graph of the sentence in un-directional: $A_{i,j}^{Inter} = A_{j,i}^{Inter}$.

3.5 Interactive Graph Convolutional Network

In InterGCN, aspect-focused GCN takes each aspect-focused graph and corresponding word embedding matrix as input, and the inter-aspect GCN receives the inter-aspect graph and hidden representations learned by aspect-focused GCN layers to produce interactive sentiment features for the specific aspect. Each node in the l -th GCN layer is updated according to the hidden representations of its neighborhoods:

$$\mathbf{h}_i^l = \text{ReLU}(\tilde{\mathbf{A}}_i^F \mathbf{g}_i^{l-1} \mathbf{W}^l + \mathbf{b}^l) \quad (7)$$

where $\mathbf{g}_i^{l-1} = \mathcal{F}(\mathbf{h}_i^{l-1})$ is the hidden representation evolved from the preceding GCN layer. $\mathcal{F}(\cdot)$ is a position-aware transformation function, which is utilized in a previous GCN-based work (Zhang et al., 2019). $\tilde{\mathbf{A}}^F$ is a normalized symmetric of an aspect-focused adjacency matrix:

$$\tilde{\mathbf{A}}_i^F = \mathbf{A}_i^F / (\mathbf{D}_i^F + 1) \quad (8)$$

where $\mathbf{D}_i^F = \sum_{j=1}^n A_{i,j}^F$ is the degree of \mathbf{A}_i^F . The original nodes of the aspect-focused GCN layers are derived from the hidden representations of Bi-LSTM layers, which takes word embeddings as input:

$$\mathbf{H}^c = \{\mathbf{h}_1^c, \mathbf{h}_2^c, \dots, \mathbf{h}_n^c\} = \text{Bi-LSTM}(\mathbf{x}) \quad (9)$$

Interactively, original nodes of the inter-aspect GCN layers are generated by the aspect-focused GCN layers. After that, we can successively capture the final representations of the aspect-focused and inter-aspect GCN layers, i.e. \mathbf{h}^F and \mathbf{h}^{Inter} . And thus, we combine these two final representations to extract the interactive relations between aspect-focused features and inter-aspect features:

$$\tilde{\mathbf{h}} = \mathbf{h}^F + \gamma \mathbf{h}^{Inter} \quad (10)$$

where γ is the coefficient of inter-aspect features. To highlight the significant features of aspect words, we exploit aspect-specific masking to mask the non-aspect representations: $\tilde{\mathbf{H}}_{mask} = \{\mathbf{0}, \dots, \tilde{\mathbf{h}}_\tau, \dots, \tilde{\mathbf{h}}_{\tau+k-1}, \dots, \mathbf{0}\}$, where $\tilde{\mathbf{h}}_t$ is the representation of the t -th word learned by InterGCN, τ is the beginning index of the specific aspect, and k is the length of the aspect. Then inspired by (Zhang et al., 2019), we adopt a retrieval-based attention mechanism to capture significant sentiment features from the context representations for the specific aspect:

$$\beta_t = \sum_{i=1}^n \mathbf{h}_t^c \top \tilde{\mathbf{h}}_i = \sum_{i=\tau}^{\tau+k-1} \mathbf{h}_t^c \top \tilde{\mathbf{h}}_i \quad (11)$$

$$\alpha_t = \frac{\exp(\beta_t)}{\sum_{i=1}^n \exp(\beta_i)} \quad (12)$$

Hence, the final representation of the input with respect to the specific aspect is formulated as:

$$\mathbf{y} = \text{softmax}(\mathbf{W}_o \mathbf{r} + \mathbf{b}_o) \quad (13)$$

$$\mathbf{r} = \sum_{i=1}^n \alpha_i \mathbf{h}_i^c \quad (14)$$

where $\text{softmax}(\cdot)$ is the softmax function to obtain the output distribution of the classifier.

3.6 Model Training

The objective to train the classifiers is defined as minimizing the cross-entropy loss between predicted and ground-truth distribution:

$$\mathcal{L} = -\sum_{i=1}^S \sum_{j=1}^C \hat{y}_i^j \cdot \log(y_i^j) + \lambda \|\Theta\|_2 \quad (15)$$

Where S is the number of training samples, C is the number of classes. \hat{y} is the ground-truth distribution of sentiment. λ is the weight of the L_2 regularization term. Θ denotes all trainable parameters.

Dataset	Positive		Neutral		Negative	
	Train	Test	Train	Test	Train	Test
REST14	2164	728	637	196	807	196
LAP14	994	341	464	169	870	128
REST15	1178	439	50	35	382	328
REST16	1620	597	88	38	709	190

Table 1: Statistics of the experimental datasets.

4 Experiments

4.1 Dataset and Experiment Setting

We conduct experiments on four benchmark datasets from SemEval 2014 (Pontiki et al., 2014) (Restaurant14, Laptop14), SemEval 2015 (Pontiki et al., 2015) (Restaurant15), and SemEval 2016 (Pontiki et al., 2016) (Restaurant16). Each sample consists of the review sentences, aspects (single or multiple words), and the sentiment polarity towards to the aspects. The statistics of the datasets are shown in Table 1.

In our experiments, we use GloVe vectors (Pennington et al., 2014) to initialize each word into 300-dimensional word embedding for all non-BERT models. The dimensionality of hidden vector representations is set to 300. The number of GCN layers is set to 2, which is the optimal depth in pilot studies. The coefficient γ is set to 0.2, and the coefficient λ of L_2 regularization item is set to 10^{-5} . Adam is utilized as the optimizer with a learning rate of 10^{-3} to train the model, and the mini-batch size is 16. We randomly initialize all the \mathbf{W} and \mathbf{b} with uniform distribution².

4.2 Comparison Models

We compared the proposed model (**InterGCN**) with the following methods:

SVM (Kiritchenko et al., 2014) trains a SVM classifier based on conventional feature extraction methods.

TD-LSTM (Tang et al., 2016a) models bidirectional contextual features for a given aspect with LSTMs.

ATAE-LSTM (Wang et al., 2016b) explores aspect-specific attention mechanism based on LSTM.

MemNet (Tang et al., 2016b) exploits word and position attention to focus on specific aspect by a multi-hop memory network.

IAN (Ma et al., 2017) learns the interactive relationships for aspect and context representations by an interactive attention network.

RAM (Chen et al., 2017) proposes a recurrent attention memory network for aspect sentiment analysis.

GCAE (Xue and Li, 2018) explores a gated CNN to control the flow of features for a given aspect.

MGAN (Fan et al., 2018) exploits fine-grained and coarse-grained attention mechanisms to capture the word-level interaction between aspect and context.

AOA (Huang et al., 2018) utilizes an attention-over-attention model to learn the interaction between aspect words and contextual words.

TNet-LF (Li et al., 2018) exploits a target-specific transformation component to better integrate target information into the word representations.

IARM (Majumder et al., 2018) extracts the influence of the neighboring aspects related information for the aspect sentiment analysis.

TransCap (Chen and Qian, 2019) utilizes a transfer capsule network model for aspect sentiment analysis.

IACapsNet (Du et al., 2019) adopts a capsule network to model vector-based feature representation and cluster features by an EM routing algorithm for a specific aspect.

AEN (Song et al., 2019) explores an attention encoder network to extract contextual sentiment interactions for the aspect.

TD-GAT (Huang and Carley, 2019) exploits a target-dependent graph attention network over dependency tree for aspect sentiment analysis.

ASGCN-DT and **ASGCN-DG** (Zhang et al., 2019) extracts syntactical information and word dependencies with directional and un-directional graph respectively over the dependency tree based on GCN.

BERT (Devlin et al., 2019) is the vanilla BERT model, which adopts “[CLS] sentence [SEP] aspect [SEP]” as input³.

AEN+BERT (Song et al., 2019) is the AEN model based on pre-trained BERT.

SA-GCN+BERT (Hou et al., 2019) is a GCN-based model with dependency tree and BERT, which employs the selective attention find important words to derive the representations of aspects.

InterGCN is our complete proposed model.

AFGCN is identical to our proposed InterGCN model, but only utilize the aspect-focused component.

InterGCN+BERT integrates our proposed InterGCN model with pre-trained BERT.

AFGCN+BERT integrates our proposed AFGCN model with pre-trained BERT.

4.3 Main Results

Table 2 shows the comparison results on four benchmark datasets, which demonstrate that the proposed InterGCN consistently outperforms all comparison models. This verifies the effectiveness of our proposed InterGCN in aspect sentiment analysis. We can note that the proposed InterGCN+BERT achieves

²The source code of this work is released at <https://github.com/BinLiang-NLP/InterGCN-ABSA>

³Here, we use the pre-trained uncased BERT-base model for all BERT-based models.

Model	REST14		LAP14		REST15		REST16	
	Acc. (%)	F1 (%)	Acc. (%)	F1 (%)	Acc. (%)	F1 (%)	Acc. (%)	F1 (%)
SVM [‡] (Kiritchenko et al., 2014)	80.16	-	70.49	-	-	-	-	-
TD-LSTM (Tang et al., 2016a)	78.00	66.73	71.83	68.43	76.39	58.70	82.16	54.21
ATAE-LSTM (Wang et al., 2016b)	78.60	67.02	68.88	63.93	78.48	60.53	83.77	61.71
MemNet [‡] (Tang et al., 2016b)	79.61	69.64	70.64	65.17	77.31	58.28	85.44	65.99
IAN [‡] (Ma et al., 2017)	79.26	70.09	72.05	67.38	78.54	52.65	84.74	55.21
RAM (Chen et al., 2017)	80.23	70.80	74.49	71.35	79.98	60.57	83.88	62.14
GCAE (Xue and Li, 2018)	75.74	62.45	71.98	68.71	77.56	56.03	83.70	62.69
MGAN (Fan et al., 2018)	81.25	71.94	75.39	72.47	79.36	57.26	87.06	62.29
AOA [‡] (Huang et al., 2018)	79.97	70.42	72.62	67.52	78.17	57.02	87.50	66.21
TNet-LF [‡] (Li et al., 2018)	80.42	71.03	74.61	70.14	78.47	59.47	89.07	70.43
IARM (Majumder et al., 2018)	80.00	-	73.80	-	-	-	-	-
TransCap (Chen and Qian, 2019)	79.29	70.85	73.87	70.10	-	-	-	-
IACapsNet (Du et al., 2019)	81.79	73.40	76.80	73.29	-	-	-	-
AEN (Song et al., 2019)	80.98	72.14	73.51	69.04	-	-	-	-
TD-GAT (Huang and Carley, 2019)	81.10	-	73.70	-	-	-	-	-
ASGCN-DT [‡] (Zhang et al., 2019)	80.86	72.19	74.14	69.24	79.34	60.78	88.69	66.64
ASGCN-DG [‡] (Zhang et al., 2019)	80.77	72.02	75.55	71.05	79.89	61.89	88.99	67.48
AFGCN (ours)	81.79	73.42	76.96	73.29	81.55	65.08	89.12	70.60
InterGCN (ours)	82.23	74.01	77.86	74.32	81.76	65.67	89.77	73.05
BERT (Devlin et al., 2019)	84.11	76.68	77.59	73.28	83.48	66.18	90.10	74.16
AEN+BERT (Song et al., 2019)	83.12	73.76	79.93	76.31	-	-	-	-
SA-GCN+BERT (Hou et al., 2019)	85.80	79.70	81.70	78.80	-	-	-	-
AFGCN+BERT (ours)	86.57	80.36	82.53	79.10	85.02	70.89	90.76	76.88
InterGCN+BERT (ours)	87.12	81.02	82.87	79.32	85.42	71.05	91.27	78.32

Table 2: Main experimental results. Acc. represents accuracy, F1 represents Macro-F1 score. Best scores are in bold. The results with ‡ are retrieved from (Zhang et al., 2019).

Model	REST14		LAP14		REST15		REST16	
	Acc. (%)	F1 (%)	Acc. (%)	F1 (%)	Acc. (%)	F1 (%)	Acc. (%)	F1 (%)
InterGCN w/o dependency tree	82.03	73.82	76.70	72.63	81.56	65.12	89.43	72.67
InterGCN w/o aspect-focused	81.32	72.57	75.89	71.45	80.06	62.73	89.05	68.36
InterGCN w/o aspect relations	81.80	73.61	76.38	72.21	81.34	64.57	89.21	71.89
InterGCN w/o inter-aspect (AFGCN)	81.79	73.42	76.96	73.29	81.55	65.08	89.12	70.60
InterGCN	82.23	74.01	77.86	74.32	81.76	65.67	89.77	73.05

Table 3: Experimental results of the ablation study.

significant improvement on all datasets compared to the other BERT-based models, which indicates that InterGCN can be easily integrated with pre-trained BERT and improve the performance of aspect sentiment analysis. It is also noteworthy that both AFGCN and InterGCN perform significantly better than the previous GCN-based models (for both non-BERT and BERT), which fundamentally verifies the effectiveness of the novel solution of constructing graphs exploited in this work. Compared with AFGCN, which only utilizes aspect-focused GCN layers, the complete InterGCN achieves better performance on all datasets. This denotes that exploiting the interaction of the features extracted from aspect-focused and inter-aspect GCN layers can further improve the performance of aspect sentiment analysis.

4.4 Ablation Study

We conduct an ablation study to further analyze the impact of different components of InterGCN. The results are shown in Table 3. We can observe that removal of “dependency tree” degrades the performance slightly, which indicates that the dependency tree based graph construction can improve the quality of dependency representations but it is not an essential part of InterGCN. We can also notice that model without “aspect-focused” performs most unsatisfactorily on all datasets, verifying that incorporating aspect-focused information into the model is the most important improvement for aspect sentiment analysis. In addition, both removal of “aspect relations” and “inter-aspect” lead to performance drops evidently, which further indicates that extracting sentiment relations between different aspects can largely

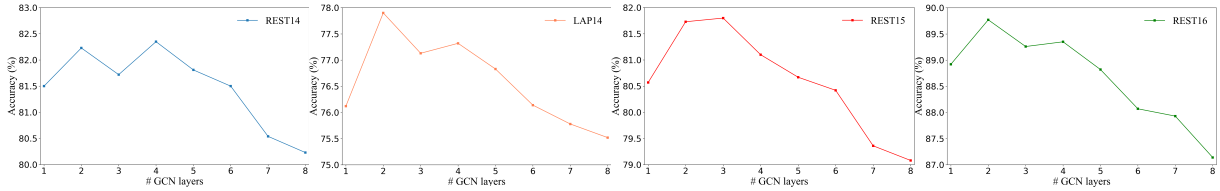


Figure 3: The impact of the number of GCN layers. Accuracy based on different numbers of GCN layers is reported.

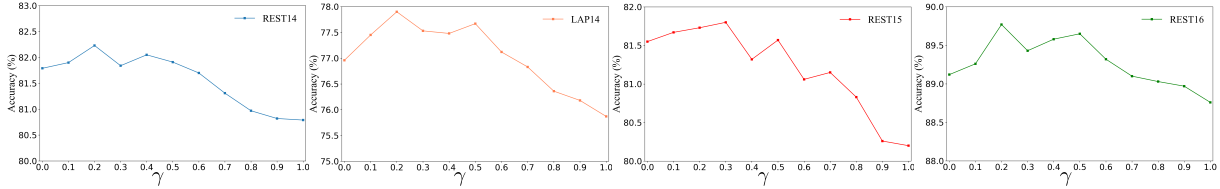


Figure 4: The impact of the coefficient of Inter-Aspect. Accuracy based on different values of γ is reported.

improve the performance of aspect sentiment analysis.

4.5 Impact of GCN Layers

We investigate the impact of the layer number on the performance of our proposed InterGCN. We vary the layer number from 1 to 8 and report the results in Figure 3. Overall, 2-layer GCN achieves better performance on all datasets, and thus we set the number of GCN layers as 2 in our experiments. Comparatively, 1-layer InterGCN performs unsatisfactorily, which indicates that 1-layer GCN is insufficient to derive precise syntactical dependencies of the context towards the specific aspect. Additionally, the performance of InterGCN fluctuates with the increasing number of GCN layers and essentially tends to decline when the model depth is greater than 4. This implies that roughly increasing the depth of GCN is vulnerable to slash the learning ability of the model due to the sharp increase of model parameters.

4.6 Impact of the Coefficient of Inter-Aspect

To further analyze the effect of extracting inter-aspect relations in InterGCN, we conduct experiments based on different values of γ and demonstrate the results in Figure 4. We can observe that as the value of γ increases from 0 to 0.2 the performance improves steadily, which implies that appropriately incorporating the interactive features extracted from inter-aspect GCN layers can assist the aspect-focused component to learn precise aspect-specific sentiment features and improve the performance of aspect sentiment analysis. However, the curve quite fluctuates when the value of γ is greater than 0.2 and considerably tends to decline when γ is greater than 0.5. This indicates that excessively consider extracting inter-aspect relations of the sentence may hamper the learning of aspect-specific sentiment features.

4.7 Analysis of Multiple Aspects

To further analyze the improvement of the performance brought by multiple aspects sentences with the proposed InterGCN, we separate the training instances into different groups according to the number of aspects in the sentences and report the training accuracy for different data groups in comparison with a precious GCN-based model (ASGCN-DG) on REST14 and LAPI14 dataset⁴. The comparison results are demonstrated in Figure 5. We can observe that the fitting results of the proposed InterGCN are superior in comparison with ASGCN-DG on all numbers of aspects. In addition, InterGCN achieves remarkable fitting results on all vary of aspect numbers, however, the performance of fitting is insufficient with ASGCN-DG when the number of aspects is either particular small or great. This implies that InterGCN can capture significant sentiment features for a specific aspect with aspect-focused GCN when the number of aspects is small in the sentence. Concurrently, InterGCN can commendably discriminate the sentiment features of multiple aspects sentences with the help of inter-aspect relations.

⁴Following (Zhang et al., 2019), we ignore the instances whose aspect numbers are more than 7 since the count of these instances is too small for any meaningful comparison.

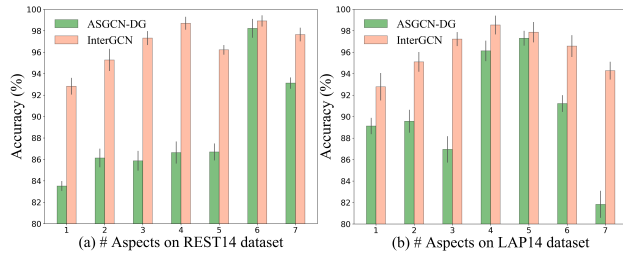


Figure 5: Comparison results of different aspect numbers.

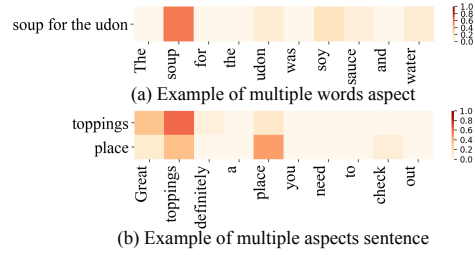


Figure 6: The attention visualizations of typical samples.

4.8 Visualization

To qualitatively demonstrate how the proposed InterGCN improves the performance of aspect sentiment analysis, we visualize the attention weights by showing two typical examples, which are cited in Figure 1. The results of attention weight visualization are demonstrated in Figure 6. According to Figure 6(a), the proposed InterGCN can pay more attention to the key aspect word for extracting aspect-specific sentiment. In addition, the multiple-aspects example shown in Figure 6(b) denotes that InterGCN can connect the related aspect when dealing with an aspect without intuitive sentiment expression. This verifies that interactively learning the aspect-focused and inter-aspect sentiment relations can derive more precise aspect-specific sentiment features and improve the performance of aspect sentiment analysis.

5 Conclusion

In this paper, we explore a novel solution of constructing aspect-focused and inter-aspect dependency graphs for aspect sentiment analysis. Based on it, an Interactive Graph Convolutional Networks (InterGCN) model is proposed to extract the aspect-specific sentiment features from the aspect-focused and inter-aspect perspective. To this end, the proposed InterGCN model can pay significant attention to the key aspect words when dealing with multiple words aspect, and connect the valuable sentiment features of related aspects when considering an aspect without distinct sentiment expression. Experimental results on four benchmark datasets show that the proposed InterGCN can outperform state-of-the-art methods including remarkable GCN-based models and BERT-based models.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (No. 61632011 and No. 61876053), the Guangdong Province Covid-19 Pandemic Control Research Funding (No. 2020KZDZX1224), the Shenzhen Foundational Research Funding (No. JCYJ20180507183527919 and No. JCYJ20180507183608379), and the EU-H2020 (grant No. 794196).

References

- Zhuang Chen and Tiejun Qian. 2019. Transfer capsule network for aspect level sentiment classification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 547–556.
- Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. Recurrent attention network on memory for aspect sentiment analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 452–461.
- 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.
- Chunning Du, Haifeng Sun, Jingyu Wang, Qi Qi, Jianxin Liao, Tong Xu, and Ming Liu. 2019. Capsule network with interactive attention for aspect-level sentiment classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5488–5497.

- 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.
- Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2019. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 504–515, July.
- Xiaochen Hou, Jing Huang, Guangtao Wang, Kevin Huang, Xiaodong He, and Bowen Zhou. 2019. Selective attention based graph convolutional networks for aspect-level sentiment classification. *arXiv preprint arXiv:1910.10857*.
- Binxuan Huang and Kathleen Carley. 2019. Syntax-aware aspect level sentiment classification with graph attention networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5469–5477.
- Binxuan Huang, Yanglan Ou, and Kathleen M Carley. 2018. Aspect level sentiment classification with attention-over-attention neural networks. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, pages 197–206.
- Lianzhe Huang, Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2019. Text level graph neural network for text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3444–3450.
- Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 2011. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 151–160.
- Thomas N Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *ICLR*.
- Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. 2014. NRC-canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 437–442.
- Xin Li, Lidong Bing, Wai Lam, and Bei Shi. 2018. Transformation networks for target-oriented sentiment classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 946–956.
- Zheng Li, Xin Li, Ying Wei, Lidong Bing, Yu Zhang, and Qiang Yang. 2019. Transferable end-to-end aspect-based sentiment analysis with selective adversarial learning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4582–4592, November.
- Bin Liang, Jiachen Du, Ruifeng Xu, Binyang Li, and Hejiao Huang. 2019. Context-aware embedding for targeted aspect-based sentiment analysis. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4678–4683.
- Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 4068–4074.
- Navonil Majumder, Soujanya Poria, Alexander Gelbukh, Md. Shad Akhtar, Erik Cambria, and Asif Ekbal. 2018. IARM: Inter-aspect relation modeling with memory networks in aspect-based sentiment analysis. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3402–3411.
- Bo Pang, Lillian Lee, et al. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.
- 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 the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35.

- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495.
- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphee De Clercq, Veronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, N ria Bel, Salud Mar a Jim nez-Zafra, and G l sen Eryiđit. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30.
- Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao. 2019. Attentional encoder network for targeted sentiment classification. *arXiv preprint arXiv:1902.09314*.
- Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. 2019. Aspect-level sentiment analysis via convolution over dependency tree. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5679–5688.
- Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. 2016a. Effective LSTMs for target-dependent sentiment classification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3298–3307.
- Duyu Tang, Bing Qin, and Ting Liu. 2016b. Aspect level sentiment classification with deep memory network. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 214–224.
- Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016a. Recursive neural conditional random fields for aspect-based sentiment analysis. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 616–626.
- Yequan Wang, Minlie Huang, xiaoyan zhu, and Li Zhao. 2016b. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615.
- Shuai Wang, Sahisnu Mazumder, Bing Liu, Mianwei Zhou, and Yi Chang. 2018. Target-sensitive memory networks for aspect sentiment classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 957–967.
- 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*, pages 2514–2523.
- Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. Graph convolutional networks for text classification. In *AAAI*, pages 7370–7377.
- Yuhao Zhang, Peng Qi, and Christopher D. Manning. 2018. Graph convolution over pruned dependency trees improves relation extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2205–2215.
- Chen Zhang, Qiuchi Li, and Dawei Song. 2019. Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4567–4577.
- Yaowei Zheng, Richong Zhang, Samuel Mensah, and Yongyi Mao. 2020. Replicate, walk, and stop on syntax: An effective neural network model for aspect-level sentiment classification. In *AAAI*, pages 9685–9692.