

Dual Graph Convolutional Networks for Aspect-based Sentiment Analysis

Ruifan Li^{1*}, Hao Chen¹, Fangxiang Feng¹,
Zhanyu Ma¹, Xiaojie WANG¹, and Eduard Hovy²

¹ School of Artificial Intelligence, Beijing University of Posts and Telecommunications, China

² Language Technologies Institute, Carnegie Mellon University, USA

{rfli, ccchenhao997, fxfeng, mazhanyu, xjwang}@bupt.edu.cn
hovy@cmu.edu

Abstract

Aspect-based sentiment analysis is a fine-grained sentiment classification task. Recently, graph neural networks over dependency trees have been explored to explicitly model connections between aspects and opinion words. However, the improvement is limited due to the inaccuracy of the dependency parsing results and the informal expressions and complexity of online reviews. To overcome these challenges, in this paper, we propose a dual graph convolutional networks (DualGCN) model that considers the complementarity of syntax structures and semantic correlations simultaneously. Particularly, to alleviate dependency parsing errors, we design a SynGCN module with rich syntactic knowledge. To capture semantic correlations, we design a SemGCN module with self-attention mechanism. Furthermore, we propose orthogonal and differential regularizers to capture semantic correlations between words precisely by constraining attention scores in the SemGCN module. The orthogonal regularizer encourages the SemGCN to learn semantically correlated words with less overlap for each word. The differential regularizer encourages the SemGCN to learn semantic features that the SynGCN fails to capture. Experimental results on three public datasets show that our DualGCN model outperforms state-of-the-art methods and verify the effectiveness of our model.

1 Introduction

Sentiment analysis has become a popular topic in natural language processing (Liu, 2012; Li and Hovy, 2017). Aspect-based sentiment analysis (ABSA) talks an entity-level oriented fine-grained sentiment analysis task that aims to determine sentiment polarities of given aspects in a sentence. In

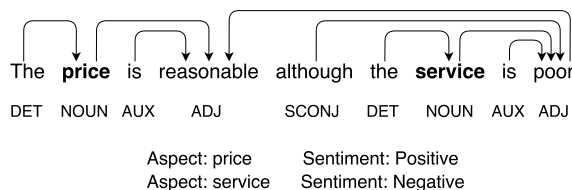


Figure 1: An example sentence with its dependency tree from the restaurant reviews. This sentence contains two aspects but with opposite sentiment polarities.

Figure 1, the comment is about a restaurant review. The sentiment polarity of the two aspects “price” and “service” are positive and negative, respectively. Thus, ABSA can precisely identify user’s attitudes towards a certain aspect, rather than simply assigning a sentiment polarity for a sentence.

The key point in solving the ABSA task is to model the dependency relationship between an aspect and its corresponding opinion expressions. Nevertheless, there probably exist multiple aspects and different opinion expressions in a sentence. To judge the sentiment of a particular aspect, previous studies (Wang et al., 2016; Tang et al., 2016a; Ma et al., 2017; Chen et al., 2017; Fan et al., 2018; Huang et al., 2018; Gu et al., 2018) have proposed various recurrent neural networks (RNNs) with attention mechanisms to generate aspect-specific sentence representations and have achieved appealing results. However, an inherent defect makes the attention mechanism vulnerable to noise in the sentence. Take Figure 1 as an example; for the aspect “service”, the opinion word “reasonable” may receive more attention than the opinion word “poor”. However, the “reasonable” refers to another aspect, i.e., “price”.

More recent efforts (Zhang et al., 2019; Sun et al., 2019b; Huang and Carley, 2019; Zhang and Qian, 2020; Chen et al., 2020; Liang et al., 2020; Wang et al., 2020; Tang et al., 2020) have been de-

*Corresponding author.

voted to graph convolutional networks (GCNs) and graph attention networks (GATs) over dependency trees, which explicitly exploit the syntactic structure of a sentence. Consider the dependency tree in Figure 1; the syntactic dependency can establish connections between the words in a sentence. For example, a dependency relation exists between the aspect “*price*” and the opinion word “*reasonable*”. However, two challenges arise when applying syntactic dependency knowledge to the ABSA task: 1) the inaccuracy of the dependency parsing results and 2) GCNs over dependency trees do not work well as expected on datasets that are not sensitive to syntactic dependency due to the informal expression and complexity of online reviews.

In this paper, we propose a novel architecture, the dual graph convolution network (DualGCN), as shown in Figure 2, to solve the aforementioned challenges. For the first challenge, we use the probability matrix of all dependency arcs from a dependency parser to build a syntax-based graph convolutional network (SynGCN). The idea behind this approach is that the probability matrix representing dependencies between words contains rich syntactic information compared with the final discrete output of a dependency parser. For the second, we construct a semantic correlation-based graph convolutional network (SemGCN) by utilizing a self-attention mechanism. The idea behind this approach is that the attention matrix shaped by self-attending, also viewed as an edge-weighted directed graph, can represent semantic correlations between words. Moreover, motivated by the work of DGEDT (Tang et al., 2020), we utilize a BiAffine module to bridge relevant information between the SynGCN and SemGCN modules.

Furthermore, we design two regularizers to enhance our DualGCN model. We observe that the semantically related terms of each word should not overlap. Therefore, we encourage the attention probability distributions over words to be orthogonal. To this end, we incorporate an orthogonal regularizer on the attention probability matrix for the SemGCN module. Moreover, the two representations learned from the SynGCN and SemGCN modules should contain significantly distinct information captured by the syntactic dependency and the semantic correlation. Therefore, we expect that the SemGCN module could learn semantic representations different from syntactic representations. Thus, we propose a differential regularizer between

the SynGCN and SemGCN modules.

Our contributions are highlighted as follows:

- We propose a DualGCN model for the ABSA task. Our DualGCN considers both the syntactic structure and the semantic correlation within a given sentence. Specifically, our DualGCN integrates the SynGCN and SemGCN networks through a mutual BiAffine module.
- We propose orthogonal and differential regularizers. The orthogonal regularizer encourages the SemGCN network to learn an orthogonal semantic attention matrix, whereas the differential regularizer encourages the SemGCN network to learn semantic features distinct from the syntactic ones built from the SynGCN network.
- We conduct extensive experiments on the SemEval 2014 and Twitter datasets. The experimental results demonstrate the effectiveness of our DualGCN model. Additionally, the source code and preprocessed datasets used in our work are provided on GitHub¹.

2 Related Work

Traditional sentiment analysis tasks are sentence-level or document-level oriented. In contrast, ABSA is an entity-level oriented and a more fine-grained task for sentiment analysis. Earlier methods (Titov and McDonald, 2008; Jiang et al., 2011; Kiritchenko et al., 2014; Vo and Zhang, 2015) are usually based on handcrafted features and fail to model the dependency between the given aspect and its context.

Recently, various attention-based neural networks have been proposed to implicitly model the semantic relation of an aspect and its context to capture the opinion expression component (Wang et al., 2016; Tang et al., 2016a,b; Ma et al., 2017; Chen et al., 2017; Fan et al., 2018; Huang et al., 2018; Gu et al., 2018; Li et al., 2018a; Tan et al., 2019). For instance, (Wang et al., 2016) proposed attention-based LSTMs for aspect-level sentiment classification. (Tang et al., 2016b) and (Chen et al., 2017) both introduced a hierarchical attention network to identify important sentiment information related to the given aspect. (Fan et al., 2018) exploited a multi-grained attention mechanism to capture the word-level interaction between aspects and their context. (Tan et al., 2019) designed a dual attention

¹<https://github.com/CCChenhao997/DualGCN-ABSA>

network to recognize conflicting opinions. In addition, the pre-trained language model BERT (Devlin et al., 2019) has achieved remarkable performance in many NLP tasks, including ABSA. (Sun et al., 2019a) transformed ABSA task into a sentence pair classification task by constructing an auxiliary sentence. (Xu et al., 2019) proposed a post-training approach on the BERT to enhance the performance of fine-tuning stage for the ABSA task.

Another trend explicitly leverages syntactic knowledge. This type of knowledge helps to establish connections between the aspects and the other words in a sentence to learn syntax-aware feature representations of aspects. (Dong et al., 2014) proposed a recursive neural network to adaptively propagate the sentiment of words to the aspect along the dependency tree. (He et al., 2018) introduced an attention model that incorporated syntactic information to compute attention weights. (Phan and Ogunbona, 2020) utilized the syntactic relative distance to reduce the impact of irrelevant words.

Following this line, a few works extend the GCN and GAT models by means of a syntactical dependency tree and develop several outstanding models (Zhang et al., 2019; Sun et al., 2019b; Huang and Carley, 2019; Wang et al., 2020; Tang et al., 2020). These works explicitly exploit the syntactic structure information to learn node representations from adjacent nodes. Thus, the dependency tree shortens the distance between the aspects and opinion words of a sentence and alleviates the problem of long-range dependency.

Most recently, several works explore the idea of combining different types of graph for ABSA task. For instance, (Chen et al., 2020) combined a dependency graph and a latent graph to generate the aspect representation. (Zhang and Qian, 2020) observed the characteristics of word co-occurrence in linguistics and designed hierarchical syntactic and lexical graphs. (Liang et al., 2020) constructed aspect-focused and inter-aspect graphs to learn dependency feature of the key aspect words and sentiment relations between different aspects.

In this paper, we propose a GCN based method combining syntactic and semantic features. We use a dependency probability matrix with richer syntactic information and elaborately design orthogonal and differential regularizers to enhance the ability to precisely capture the semantic associations.

3 Graph Convolutional Network (GCN)

Motivated by conventional convolutional neural networks (CNNs) and graph embedding, a GCN is an efficient CNN variant that operates directly on graphs (Kipf and Welling, 2017). For graph structured data, a GCN can apply the convolution operation on directly connected nodes to encode local information. Through the message passing of multilayer GCNs, each node in a graph can learn more global information. Given a graph with n nodes, the graph can be represented as an adjacency matrix $A \in \mathbb{R}^{n \times n}$. Most previous work (Zhang et al., 2019; Sun et al., 2019b) extend GCN models by encoding dependency trees and incorporating dependency paths between words. They build the adjacency matrix A over the syntactical dependency tree of a sentence. Thus, an element A_{ij} in A indicates whether the i -th node is connected to the j -th node. Specifically, $A_{ij} = 1$ if the i -th node is connected to the j -th node, and $A_{ij} = 0$ otherwise. In addition, the adjacency matrix A , composed of 0 and 1, can be deemed as the final discrete output of a dependency parser. For the i -th node at the l -th layer, formally, its hidden state representation, denoted as h_i^l , is updated by the following equation:

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

where W^l is a weight matrix, b^l is a bias term, and σ is an activation function (e.g., ReLU).

4 Proposed DualGCN

Figure 2 provides an overview of DualGCN. In the ABSA task, a sentence-aspect pair (s, a) is given, where $a = \{a_1, a_2, \dots, a_m\}$ is an aspect. It is also a sub-sequence of the entire sentence $s = \{w_1, w_2, \dots, w_n\}$. Then, we utilize BiLSTM or BERT as sentence encoder to extract hidden contextual representations, respectively. For the BiLSTM encoder, we first obtain the word embeddings $x = \{x_1, x_2, \dots, x_n\}$ of the sentence s from an embedding lookup table $E \in \mathbb{R}^{|V| \times d_e}$, where $|V|$ is the size of vocabulary and d_e denotes the dimensionality of word embeddings. Next, the word embeddings of the sentence are fed into a BiLSTM to produce hidden state vectors $H = \{h_1, h_2, \dots, h_n\}$, where $h_i \in \mathbb{R}^{2d}$ is the hidden state vector at time t from the BiLSTM. The dimensionality of a hidden state vector d is output by a unidirectional LSTM.

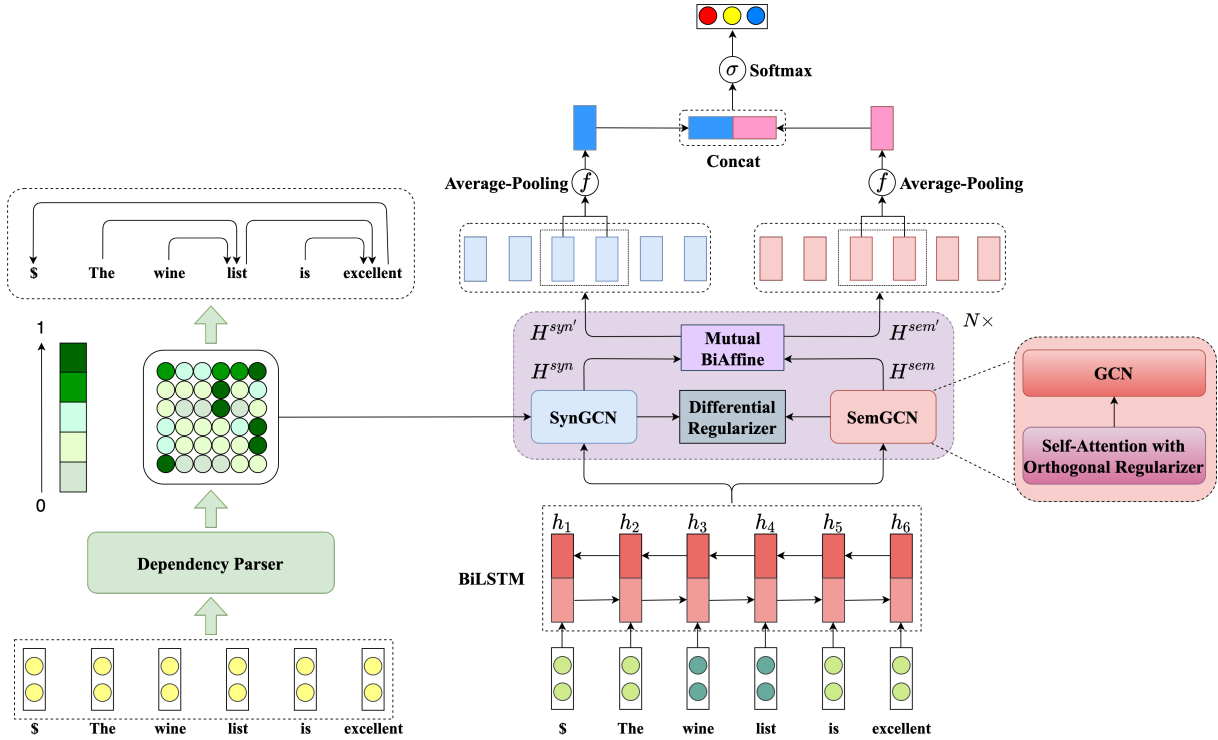


Figure 2: The overall architecture of DualGCN, which is composed primarily of SynGCN and SemGCN. SynGCN uses the probability matrix generated by the dependency parser, while SemGCN leverages the attention score matrix generated by the self-attention layer. The orthogonal and differential regularizers are designed to further improve the ability of capturing semantic correlations. Details of these components are described in the main text.

For the BERT encoder, we construct a sentence-aspect pair “[CLS] sentence [SEP] aspect [SEP]” as input to obtain aspect-aware hidden representations of the sentence. Moreover, in order to match the wordpiece-based representations of BERT with the result of syntactic dependency based on word, we expand dependencies of a word into its all of subwords. Then, the hidden representations of sentence are input into the SynGCN and SemGCN modules, respectively. A BiAffine module is then adopted for effective information flow. Finally, we aggregate all the aspect nodes’ representations from the SynGCN and SemGCN modules via pooling and concatenation to form the final aspect representation. Next, we elaborate on the details of our proposed DualGCN model.

4.1 Syntax-based GCN (SynGCN)

The SynGCN module takes the syntactic encoding as input. To encode syntactic information, we utilize the probability matrix of all dependency arcs from a dependency parser. Compared to the final discrete output of a dependency parser, the dependency probability matrix could capture rich structural information by providing all latent syntactic

structures. Therefore, the dependency probability matrix is used to alleviate dependency parsing errors. Here, we use the state-of-the-art dependency parsing model LAL-Parser (Mrini et al., 2019).

With the syntactic encoding of an adjacency matrix $A^{\text{syn}} \in \mathbb{R}^{n \times n}$, the SynGCN module takes the hidden state vectors H from BiLSTM as initial node representations in the syntactic graph. The syntactic graph representation $H^{\text{syn}} = \{h_1^{\text{syn}}, h_2^{\text{syn}}, \dots, h_n^{\text{syn}}\}$ is then obtained from the SynGCN module using Eq. (1). Here, $h_i^{\text{syn}} \in \mathbb{R}^d$ is a hidden representation of the i^{th} node. Note that for aspect nodes, we use symbols $\{h_{a_1}^{\text{syn}}, h_{a_2}^{\text{syn}}, \dots, h_{a_m}^{\text{syn}}\}$ to denote their hidden representations.

4.2 Semantic-based GCN (SemGCN)

Instead of utilizing additional syntactic knowledge, as in SynGCN, SemGCN obtains an attention matrix as an adjacency matrix via a self-attention mechanism. On the one hand, self-attention can capture the semantically related terms of each word in a sentence, which is more flexible than the syntactic structure. On the other hand, SemGCN can adapt to online reviews that are not sensitive to syntactic information.

Self-Attention Self-attention (Vaswani et al., 2017) computes the attention score of each pair of elements in parallel. In our DualGCN, we compute the attention score matrix $A^{\text{sem}} \in \mathbb{R}^{n \times n}$ using a self-attention layer. We then take the attention score matrix A^{sem} as the adjacency matrix of our SemGCN module, which can be formulated as:

$$A^{\text{sem}} = \text{softmax} \left(\frac{QW^Q \times (KW^K)^T}{\sqrt{d}} \right) \quad (2)$$

where matrices Q and K are both equal to the graph representations of previous layer of our SemGCN module, while W^Q and W^K are learnable weight matrices. In addition, d is the dimensionality of the input node feature. Note that we use only one self-attention head to obtain an attention score matrix for a sentence. Similar to the SynGCN module, the SemGCN module obtains the graph representation H^{sem} . Additionally, we use the symbols $\{h_{a_1}^{\text{sem}}, h_{a_2}^{\text{sem}}, \dots, h_{a_m}^{\text{sem}}\}$ to denote the hidden representations of all aspect nodes.

BiAffine Module To effectively exchange relevant features between the SynGCN and SemGCN modules, we adopt a mutual BiAffine transformation as a bridge. We formulate the process as follows:

$$H^{\text{syn}'} = \text{softmax} \left(H^{\text{syn}} W_1 (H^{\text{sem}})^T \right) H^{\text{sem}} \quad (3)$$

$$H^{\text{sem}'} = \text{softmax} \left(H^{\text{sem}} W_2 (H^{\text{syn}})^T \right) H^{\text{syn}} \quad (4)$$

where W_1 and W_2 are trainable parameters.

Finally, we apply average pooling and concatenation operations on the aspect nodes of the SynGCN and SemGCN modules. Thus, we obtain the final feature representation for the ABSA task, i.e.,

$$h_a^{\text{syn}} = f(h_{a_1}^{\text{syn}}, h_{a_2}^{\text{syn}}, \dots, h_{a_m}^{\text{syn}}) \quad (5)$$

$$h_a^{\text{sem}} = f(h_{a_1}^{\text{sem}}, h_{a_2}^{\text{sem}}, \dots, h_{a_m}^{\text{sem}}) \quad (6)$$

$$r = [h_a^{\text{syn}}, h_a^{\text{sem}}] \quad (7)$$

where $f(\cdot)$ is an average pooling function applied over the aspect node representations. Then, the obtained representation r is fed into a linear layer, followed by a softmax function to produce a sentiment probability distribution p , i.e.,

$$p(a) = \text{softmax}(W_p r + b_p) \quad (8)$$

where W_p and b_p are the learnable weight and bias.

4.3 Regularizer

To improve the semantic representation, we propose two regularizers for the SemGCN module, i.e., orthogonal and differential regularizers.

Orthogonal Regularizer Intuitively, the related items of each word should be in different regions in a sentence, so the attention score distributions rarely overlap. Therefore, we expect a regularizer to encourage orthogonality among the attention score vectors of all words. Given an attention score matrix $A^{\text{sem}} \in \mathbb{R}^{n \times n}$, the orthogonal regularizer is formulated as follows:

$$R_O = \|A^{\text{sem}} A^{\text{sem}T} - I\|_F \quad (9)$$

where I is an identity matrix. The subscript F denotes the Frobenius norm. As a result, each nondiagonal element of $A^{\text{sem}} A^{\text{sem}T}$ is minimized to maintain the matrix A^{sem} orthogonal.

Differential Regularizer We expect that two types of feature representations learned from the SynGCN and SemGCN modules represent distinct information contained within the syntactic dependency trees and semantic correlations. Therefore, we adopt a differential regularizer between the two adjacency matrices of the SynGCN and SemGCN modules. Note that the regularizer is only restrictive to A^{sem} and is given as

$$R_D = \frac{1}{\|A^{\text{sem}} - A^{\text{syn}}\|_F}. \quad (10)$$

4.4 Loss Function

Our training goal is to minimize the following total objective function:

$$\ell_T = \ell_C + \lambda_1 R_O + \lambda_2 R_D + \lambda_3 \|\Theta\|_2 \quad (11)$$

where λ_1 , λ_2 and λ_3 are regularization coefficients and Θ represents all trainable model parameters. ℓ_C is a standard cross-entropy loss and is defined for the ABSA task as follows:

$$\ell_C = - \sum_{(s,a) \in \mathcal{D}} \sum_{c \in \mathcal{C}} \log p(a) \quad (12)$$

where \mathcal{D} contains all sentence-aspect pairs and \mathcal{C} is the collection of distinct sentiment polarities.

5 Experiments

5.1 Datasets

We conduct experiments on three public standard datasets. The Restaurant and Laptop datasets

Dataset	Division	# Positive	# Negative	# Neutral
Restaurant	Training	2164	807	637
	Testing	727	196	196
Laptop	Training	976	851	455
	Testing	337	128	167
Twitter	Training	1507	1528	3016
	Testing	172	169	336

Table 1: Statistics for the three experimental datasets.

are made public from the SemEval ABSA challenge (Pontiki et al., 2014). Following (Chen et al., 2017), we remove the instances using the “conflict” label. In addition, the Twitter dataset is a collection of tweets (Dong et al., 2014). All three datasets have three sentiment polarities: positive, negative and neutral. Each sentence in these datasets is annotated with marked aspects and their corresponding polarities. Statistics for the three datasets are shown in Table 1.

5.2 Implementation Details

The LAL-Parser (Mrini et al., 2019), which is used for dependency parsing, provides an off-the-shelf parser². For all the experiments, we use pretrained 300-dimensional Glove³ vectors (Pennington et al., 2014) to initialize the word embeddings. The dimensionality of the position (i.e., the relative position of each word in a sentence with respect to the aspect) embeddings and part-of-speech (POS) embeddings is set to 30. Thus, we concatenate the word, POS and position embeddings and then input them into a BiLSTM model, whose hidden size is set to 50. To alleviate overfitting, we apply dropout at a rate of 0.7 to the input word embeddings of the BiLSTM. The dropout rate of the SynGCN and SemGCN modules is set to 0.1, and the number of SynGCN and SemGCN layers is set to 2. All the model weights are initialized from a uniform distribution. We use the Adam optimizer with a learning rate of 0.002. The DualGCN model is trained in 50 epochs with a batch size of 16. The regularization coefficients, λ_1 and λ_2 are set to (0.2, 0.3), (0.2, 0.2) and (0.3, 0.2) for the three datasets, respectively, and λ_3 is set to 10^{-4} . For DualGCN+BERT, we use the bert-base-uncased⁴ English version. See our code for more details about BERT’s experiments. Additionally, following (Marcheggiani and Titov, 2017), we add a self-loop for each node in

²<https://github.com/KhalilMrini/LAL-Parser>

³<https://nlp.stanford.edu/projects/glove/>

⁴<https://github.com/huggingface/transformers>

the SynGCN and SemGCN modules.

5.3 Baseline Methods

We compare DualGCN with state-of-the-art baselines. The models are briefly described as follows.

1) **ATAE-LSTM** (Wang et al., 2016) utilizes aspect embedding and the attention mechanism in aspect-level sentiment classification.

2) **IAN** (Ma et al., 2017) employs two LSTMs and an interactive attention mechanism to generate representations for the aspect and sentence.

3) **RAM** (Chen et al., 2017) uses multiple attention and memory networks to learn the sentence representation.

4) **MGAN** (Fan et al., 2018) designs a multigrained attention mechanism to capture word-level interactions between the aspect and context.

5) **TNet** (Li et al., 2018b) transforms BiLSTM embeddings into target-specific embeddings and uses CNN to extract final embeddings for classification.

6) **ASGCN** (Zhang et al., 2019) first proposed using GCN to learn the aspect-specific representations for aspect-based sentiment classification.

7) **CDT** (Sun et al., 2019b) utilizes a GCN over a dependency tree to learn aspect representations with syntactic information.

8) **BiGCN** (Zhang and Qian, 2020) uses hierarchical graph structure to integrate word co-occurrence information and dependency type information.

9) **kumaGCN** (Chen et al., 2020) employs a latent graph structure to complement syntactic features.

10) **InterGCN** (Liang et al., 2020) utilizes a GCN over a dependency tree to learn aspect representations with syntactic information.

11) **R-GAT** (Wang et al., 2020) proposes a aspect-oriented dependency tree structure and then encodes new dependency trees with a relational GAT.

12) **DGEDT** (Tang et al., 2020) proposes a dependency graph enhanced dual-transformer network by jointly considering flat representations and graph-based representations.

13) **BERT** (Devlin et al., 2019) is the vanilla BERT model by feeding the sentence-aspect pair and using the representation of [CLS] for predictions.

14) **R-GAT+BERT** (Wang et al., 2020) is the R-GAT model that uses a pre-trained BERT to replace BiLSTM as an encoder.

15) **DGEDT+BERT** (Tang et al., 2020) is the DGEDT model that uses a pre-trained BERT to replace BiLSTM as an encoder.

5.4 Comparison Results

To evaluate the ABSA models, we use the accuracy and macro-averaged F1-score as the main evaluation metrics. The main experimental results are reported in Table 2. Our DualGCN model consistently outperforms all attention-based and syntax-based methods on the Restaurant, Laptop and Twitter datasets. These results demonstrates that our DualGCN effectively integrates syntactic knowledge and semantic information. In addition, the DualGCN accurately fits datasets that contain formal, informal or complicated reviews. Compared to attention-based methods such as ATAE-LSTM, IAN and RAM, our DualGCN model utilizes syntactic knowledge to establish dependencies between words, so it can avoid noises introduced by the attention mechanism. Moreover, the syntax-based methods, such as ASGCN, CDT, R-GAT and so on, achieve better performance than attention-based methods, but they ignore the semantic correlation between words. However, when considering informal or complicated sentences, using only syntactic knowledge results in poor performance. In Table 2, on the other side, the results from the last group shows that the basic BERT outperforms most of the models based on static word embedding. Moreover, based on BERT, our DualGCN+BERT achieves better performance.

5.5 Ablation Study

To further investigate the role of modules in the DualGCN model, we conduct extensive ablation studies. The results are reported in Table 2. The SynGCN-head model uses the discrete outputs of a dependency parser to construct the adjacency matrix of the GCNs. In contrast, SynGCN leverages the probability matrix generated in a dependency parser as the adjacency matrix. The SynGCN model outperforms the SynGCN-head on the Restaurant and Laptop datasets, which demonstrates that rich syntactic knowledge can alleviate dependency parsing errors. The SemGCN model utilizes a self-attention layer to construct the adjacency matrix of the semantic graph. This SemGCN model outperforms the SynGCN on the Twitter dataset because the reviews from Twitter, compared to those from Restaurant and Laptop datasets, are largely informal and insensitive to syntactic information. DualGCN w/o BiAffine means that we remove the BiAffine module so that the SynGCN and SemGCN modules cannot interact with each

other. Therefore, the performance degrades substantially on the Restaurant and Laptop datasets. DualGCN w/o R_O & R_D indicates that we remove both the orthogonal and differential regularizers. Similarly, DualGCN w/o R_O or R_D denotes that we remove only one of the regularizers. The experimental results show that our two regularizers encourage the DualGCN to capture semantic correlations precisely. Overall, our DualGCN with all modules achieves the best performance.

5.6 Case Study

Table 4 shows a few sample cases analyzed using different models. The notations P, N and O represent positive, negative and neutral sentiment, respectively. We highlight the aspect words in red and in blue. For the aspect “*food*” in the first sample, the attention-based methods, i.e., ATAE-LSTM and IAN, are prone to attend to the noisy word “*dreadful*”. Although the syntactic dependency can establish direct connections between an aspect and some words, no association exists between the aspect and the opinion words for complicated sentences. Take the second sample as an example; the aspect “*apple os*” is far from the opinion word “*happy*” in terms of syntactic distance. Thus, the SynGCN model fails. Additionally, in the third sample, feature representations of the key words “*did not*” are not captured by the SynGCN model. In contrast, the SemGCN model can attend to the semantic correlation between words. The last two samples demonstrate that our DualGCN, which fully considers the complementarity of syntactic knowledge and semantic information, can address complicated and informal sentences with the help of the orthogonal and differential regularizers.

5.7 Attention Visualization

To investigate the effectiveness of the two regularizers in capturing the semantic correlations between words, we visualized the attention score matrix of the DualGCN w/o R_O & R_D and the intact DualGCN. Consider the sample sentence, i.e., “*Web browsing is very quick with Safari browser.*” with “*Safari browser*” as an aspect. As shown in Figure 3 (a), the attention score matrix is dense, and the related terms of each word overlap in the DualGCN w/o R_O & R_D model. This result is attributed to the lack of semantic constraints in the self-attention layers. The overlap of semantic correlations will lead to redundancy and noise during information propagation. The seventh and eighth rows of the

Models	Restaurant		Laptop		Twitter	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
ATAE-LSTM (Wang et al., 2016)	77.20	-	68.70	-	-	-
IAN (Ma et al., 2017)	78.60	-	72.10	-	-	-
RAM (Chen et al., 2017)	80.23	70.80	74.49	71.35	69.36	67.30
MGAN (Fan et al., 2018)	81.25	71.94	75.39	72.47	72.54	70.81
TNet (Li et al., 2018b)	80.69	71.27	76.54	71.75	74.90	73.60
ASGCN (Zhang et al., 2019)	80.77	72.02	75.55	71.05	72.15	70.40
CDT (Sun et al., 2019b)	82.30	74.02	77.19	72.99	74.66	73.66
BiGCN (Zhang and Qian, 2020)	81.97	73.48	74.59	71.84	74.16	73.35
kumaGCN (Chen et al., 2020)	81.43	73.64	76.12	72.42	72.45	70.77
InterGCN (Liang et al., 2020)	82.23	74.01	77.86	74.32	-	-
R-GAT (Wang et al., 2020)	83.30	76.08	77.42	73.76	75.57	73.82
DGEDT (Tang et al., 2020)	83.90	75.10	76.80	72.30	74.80	73.40
Our DualGCN	84.27	78.08	78.48	74.74	75.92	74.29
BERT-SPC (Devlin et al., 2019)	86.15	80.29	81.01	76.69	75.18	74.01
R-GAT+BERT (Wang et al., 2020)	86.60	81.35	78.21	74.07	76.15	74.88
DGEDT+BERT (Tang et al., 2020)	86.30	80.00	79.80	75.60	77.90	75.40
Our DualGCN+BERT	87.13	81.16	81.80	78.10	77.40	76.02

Table 2: Experimental results comparison on three publicly available datasets.

Models	Restaurant		Laptop		Twitter	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
SynGCN-head	82.93	75.29	76.27	72.39	75.04	73.85
SynGCN	83.74	76.97	76.58	73.17	74.59	72.86
SemGCN	83.29	76.30	76.90	73.72	75.18	73.86
DualGCN w/o BiAffine	82.84	75.31	76.90	73.23	75.33	73.92
DualGCN w/o R_O & R_D	82.93	75.79	76.58	72.03	74.59	73.20
DualGCN w/o R_O	83.56	77.43	76.58	72.78	75.18	73.55
DualGCN w/o R_D	83.65	76.34	77.53	73.72	74.45	72.82
DualGCN	84.27	78.08	78.48	74.74	75.92	74.29

Table 3: Experimental results of ablation study.

attention score matrix are the attention probability distributions of “safari” and “browser”, respectively. The information to which “safari browser” pays attention is redundant and it does not pay more attention to the key opinion word “quick”. Thus, the DualGCN w/o R_O & R_D failed. In comparison, in Figure 3 (b), the attention score matrix produced by our DualGCN is relatively sparse. Both “safari” and “browser” are semantically related to “quick”, and their other attended items are also semantically reasonable. In addition, the attention scores of the related terms of each words tend to be distinct and precise due to the semantic constraints of these two regularizers. Therefore, our DualGCN model can readily predict the correct sentiment polarity of the aspect “safari browser”.

5.8 Impact of the DualGCN Layer Number

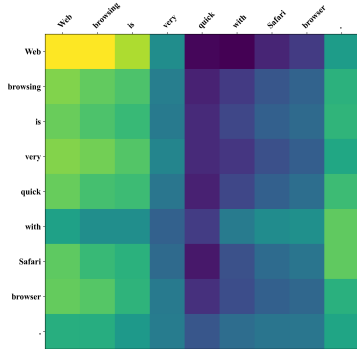
To investigate the impact of the DualGCN layer number, we evaluate our DualGCN model with one to eight layers on the Restaurant and Laptop datasets. As shown in Figure 4, our model with two DualGCN layers performs the best. On one the hand, node representations cannot propagate far when the number of layers is small. On the other hand, if the number of layers is excessive, the model will become unstable due to the vanishing gradient and information redundancy.

6 Conclusion

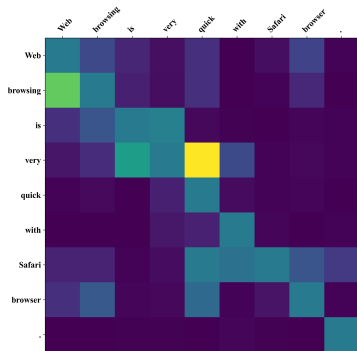
In this paper, we propose a DualGCN architecture to address the disadvantages of attention-based and dependency-based methods for ABSA tasks. Our

#	Review	ATAE-LSTM	IAN	SynGCN	SemGCN	DualGCN
1	Great food but the service was dreadful!	(N_X, N_{\checkmark})	(N_X, N_{\checkmark})	$(P_{\checkmark}, N_{\checkmark})$	$(P_{\checkmark}, N_{\checkmark})$	$(P_{\checkmark}, N_{\checkmark})$
2	Works well, and I am extremely happy to be back to an apple OS .	$(P_{\checkmark}, P_{\checkmark})$	$(P_{\checkmark}, P_{\checkmark})$	(P_{\checkmark}, O_X)	$(P_{\checkmark}, P_{\checkmark})$	$(P_{\checkmark}, P_{\checkmark})$
3	Did not enjoy the new Windows 8 and touchscreen functions .	(O_X, P_X)	(O_X, N_{\checkmark})	(P_X, O_X)	$(N_{\checkmark}, N_{\checkmark})$	$(N_{\checkmark}, N_{\checkmark})$
4	I never tried any external mics with that iMac.	O_{\checkmark}	N_X	N_X	N_X	O_{\checkmark}
5	In mi burrito, here was nothing but dark chicken that had that cooked last week and just warmed up in a microwave taste .	(N_{\checkmark}, P_X)	$(N_{\checkmark}, N_{\checkmark})$	(N_{\checkmark}, O_X)	(N_{\checkmark}, O_X)	$(N_{\checkmark}, N_{\checkmark})$

Table 4: Case studies of our DualGCN model compared with state-of-the-art baselines.



(a) The attention score matrix of DualGCN w/o R_O & R_D



(b) The attention score matrix of DualGCN

Figure 3: An illustration on how orthogonal and differential regularizers contribute to the self-attention layer.

DualGCN model integrates syntactic knowledge and semantic information by means of the SynGCN and SemGCN modules. Moreover, to effectively capture the semantic correlation between words, we propose orthogonal and differential regularizers in the SemGCN module. These regularizers can attend to the semantically related items with less overlap of each word and capture feature representations that differ from the syntactic structure. Extensive experiments on benchmark datasets show that our DualGCN model outperforms baselines.

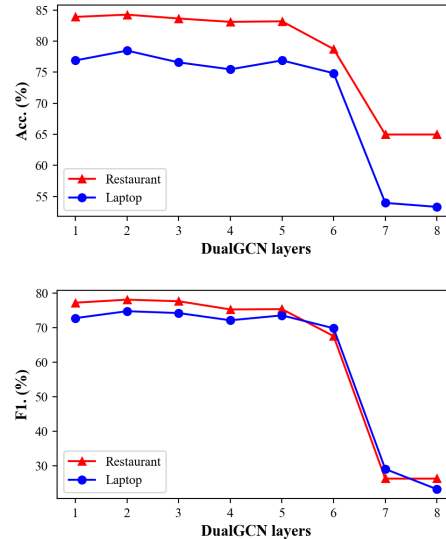


Figure 4: Effect of the number of DualGCN layers.

Acknowledgments

This work was supported in part by the National Key R&D Program of China under Grant 2019YFF0303300 and Subject II under Grant 2019YFF0303302, in part by the National Natural Science Foundation of China under Grants 61906018 and 62076032, in part by the 111 Project under Grant B08004, and in part by the Fundamental Research Funds for the Central Universities under Grant 2021RC36.

References

- Chenhua Chen, Zhiyang Teng, and Yue Zhang. 2020. [Inducing target-specific latent structures for aspect sentiment classification](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5596–5607, Online. Association for Computational Linguistics.
- 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, Copen-

- hagen, Denmark. Association for Computational Linguistics.
- 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, Minneapolis, Minnesota. Association for Computational Linguistics.
- Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. 2014. [Adaptive recursive neural network for target-dependent Twitter sentiment classification](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2)*, pages 49–54, Baltimore, Maryland. Association for Computational Linguistics.
- 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, Brussels, Belgium. Association for Computational Linguistics.
- Shuqin Gu, Lipeng Zhang, Yuexian Hou, and Yin Song. 2018. [A position-aware bidirectional attention network for aspect-level sentiment analysis](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 774–784, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2018. [Effective attention modeling for aspect-level sentiment classification](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1121–1131, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- 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, Hong Kong, China. Association for Computational Linguistics.
- Binxuan Huang, Yanglan Ou, and Kathleen M. Carley. 2018. [Aspect level sentiment classification with attention-over-attention neural networks](#). *CoRR*, abs/1804.06536.
- 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, Portland, Oregon, USA. Association for Computational Linguistics.
- Thomas N. Kipf and Max Welling. 2017. [Semi-supervised classification with graph convolutional networks](#). In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017*.
- 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, Dublin, Ireland. Association for Computational Linguistics.
- Jiwei Li and Eduard Hovy. 2017. [Reflections on sentiment/opinion analysis](#). In *A Practical Guide to Sentiment Analysis*, pages 41–59, Cham. Springer International Publishing.
- Lishuang Li, Yang Liu, and AnQiao Zhou. 2018a. [Hierarchical attention based position-aware network for aspect-level sentiment analysis](#). In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 181–189, Brussels, Belgium. Association for Computational Linguistics.
- Xin Li, Lidong Bing, Wai Lam, and Bei Shi. 2018b. [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, Melbourne, Australia. Association for Computational Linguistics.
- Bin Liang, Rongdi Yin, Lin Gui, Jiachen Du, and Ruifeng Xu. 2020. [Jointly learning aspect-focused and inter-aspect relations with graph convolutional networks for aspect sentiment analysis](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 150–161, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Bing Liu. 2012. [Sentiment analysis and opinion mining](#). *Synthesis lectures on human language technologies*, 5(1):1–167.
- Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. [Interactive attention networks for aspect-level sentiment classification](#). In *Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI’17*, page 4068–4074. AAAI Press.
- Diego Marcheggiani and Ivan Titov. 2017. [Encoding sentences with graph convolutional networks for semantic role labeling](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1506–1515, Copenhagen, Denmark. Association for Computational Linguistics.
- Khalil Mrini, Franck Deroncourt, Trung Bui, Walter Chang, and Ndapa Nakashole. 2019. [Rethinking self-attention: An interpretable self-attentive encoder-decoder parser](#). *arXiv preprint arXiv:1911.03875*.

- 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, Doha, Qatar. Association for Computational Linguistics.
- Minh Hieu Phan and Philip O. Ogunbona. 2020. **Modelling context and syntactical features for aspect-based sentiment analysis**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3211–3220, Online. Association for Computational Linguistics.
- 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, Dublin, Ireland. Association for Computational Linguistics.
- Chi Sun, Luyao Huang, and Xipeng Qiu. 2019a. **Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence**. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*, pages 380–385, Minneapolis, Minnesota. Association for Computational Linguistics.
- Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. 2019b. **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, Hong Kong, China. Association for Computational Linguistics.
- Xingwei Tan, Yi Cai, and Changxi Zhu. 2019. **Recognizing conflict opinions in aspect-level sentiment classification with dual 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 3426–3431, Hong Kong, China. Association for Computational Linguistics.
- Duyu Tang, Bing Qin, and Ting Liu. 2016a. **Aspect level sentiment classification with deep memory network**. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 214–224, Austin, Texas. Association for Computational Linguistics.
- 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, Austin, Texas. Association for Computational Linguistics.
- Hao Tang, Donghong Ji, Chenliang Li, and Qiji Zhou. 2020. **Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6578–6588, Online. Association for Computational Linguistics.
- Ivan Titov and Ryan McDonald. 2008. **Modeling online reviews with multi-grain topic models**. In *Proceedings of the 17th International Conference on World Wide Web*, page 111–120, New York, NY, USA. Association for Computing Machinery.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. **Attention is all you need**. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.
- Duy-Tin Vo and Yue Zhang. 2015. **Deep learning for event-driven stock prediction**. In *Proceedings of IJ-CAI*, Buenos Aires, Argentina.
- Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020. **Relational graph attention network for aspect-based sentiment analysis**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3229–3238, Online. Association for Computational Linguistics.
- Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. **Attention-based LSTM for aspect-level sentiment classification**. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615, Austin, Texas. Association for Computational Linguistics.
- Hu Xu, Bing Liu, Lei Shu, and Philip Yu. 2019. **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*, pages 2324–2335, Minneapolis, Minnesota. Association for Computational Linguistics.
- 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*, pages 4568–4578, Hong Kong, China. Association for Computational Linguistics.
- Mi Zhang and Tiejun Qian. 2020. **Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis**. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3540–3549, Online. Association for Computational Linguistics.