

SSEGCN: Syntactic and Semantic Enhanced Graph Convolutional Network for Aspect-based Sentiment Analysis

Zheng Zhang, Zili Zhou*, Yanna Wang

School of Cyber Science and Engineering, Qufu Normal University, China

zhangzhengzz12@gmail.com,

zz1@qfnu.edu.cn, wyn@qfnu.edu.cn

Abstract

Aspect-based Sentiment Analysis (ABSA) aims to predict the sentiment polarity towards a particular aspect in a sentence. Recently, graph neural networks based on dependency tree convey rich structural information which is proven to be utility for ABSA. However, how to effectively harness the semantic and syntactic structure information from the dependency tree remains a challenging research question. In this paper, we propose a novel Syntactic and Semantic Enhanced Graph Convolutional Network (SSEGCN) model for ABSA task. Specifically, we propose an aspect-aware attention mechanism combined with self-attention to obtain attention score matrices of a sentence, which can not only learn the aspect-related semantic correlations, but also learn the global semantics of the sentence. In order to obtain comprehensive syntactic structure information, we construct syntactic mask matrices of the sentence according to the different syntactic distances between words. Furthermore, to combine syntactic structure and semantic information, we equip the attention score matrices by syntactic mask matrices. Finally, we enhance the node representations with graph convolutional network over attention score matrices for ABSA. Experimental results on benchmark datasets illustrate that our proposed model outperforms state-of-the-art methods.

1 Introduction

Aspect-based Sentiment Analysis (ABSA) aims to determine the sentiment polarity of a given aspect term in a sentence, where sentiment polarity includes positive, negative and neutral. For example, in Figure 1, ABSA determines the sentiment towards the aspects “*food*” and “*service*”. For aspect term “*food*”, the sentiment polarity is negative, but “*service*” is positive. That is, we need to discriminate sentiment polarities according to

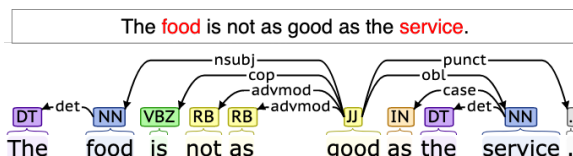


Figure 1: An example sentence with its dependency tree where aspect terms (highlighted in red) are connected with other words based on their syntactic dependencies.

different aspects. The main idea of most works is to model the dependency relation between aspects and their associated opinion words.

Prior studies exploit attention mechanism (Wang et al., 2016; Chen et al., 2017; Ma et al., 2017; Liu et al., 2018; Hu et al., 2018; Wang et al., 2018; Huang et al., 2018; Fan et al., 2018) to model the correlations between aspect term and the context. However, attention mechanism is vulnerable to noise in sentences, i.e., the irrelevant words.

Recent works on ABSA leverage Graph Neural Networks (GNNs) over the dependency tree of a sentence to exploit syntactic structure (Sun et al., 2019; Zhang et al., 2019; Liang et al., 2020; Wang et al., 2020; Zhao et al., 2020; Li et al., 2021; Tian et al., 2021). In these works, Zhang et al. (2019) employ Graph Convolutional Network (GCN) to integrate the syntactic information. Sun et al. (2019) propose a GCN model to enhance the feature representations of aspects learned by a Bi-directional Long Short Term Memory (Bi-LSTM). However, these two studies treat all the neighbor nodes of the current node in the graph equally, and are limited in lacking of efficient mechanisms to distinguish the importance of neighbor nodes. Accordingly, noisy nodes may cause the model to misjudge the sentiment polarity. To tackle this problem, Huang and Carley (2019) design a target-dependent graph attention network that updates each node representation by utilizing multi-head attention. The attention mechanism is used to consider semantic correlation of each neighbor node, but it is a lo-

*Corresponding author.

cal attention by only calculating the importance of neighboring nodes and neglects the global information of the sentence. Following this line, [Chen et al. \(2020\)](#) and [Li et al. \(2021\)](#) combine a syntactic structure graph with a latent semantic graph but the two graphs are constructed independently.

Notably, existing GCN-based approaches do not fully leverage syntactic structure, where only the information of the neighbor nodes is considered. Moreover, some hard cases obscurely express sentiment of aspect term, there is no direct syntactic relationship between aspect term and opinion words. How to capture the information of second-order nodes, third-order nodes and even the global syntactic structure in one shot is still a challenge. Recently, most methods apply multi-layer GCNs to derive the expression of opinion words which brings potential noise. For instance, consider the dependency tree as depicted in Figure 1, a dependency connection exists between the aspect “*food*” and the opinion word “*good*”. However, the “*good*” refers to another aspect “*service*”. For aspect term “*food*”, the representation of word “*not*” is obtained by utilizing two-layer GCN. At the same time, noisy word “*not*” is also obtained for aspect term “*service*”. As a result, simply considering the syntactic structure of dependency tree might be unsatisfactory. It is necessary to harness the aspect-related semantic information for different aspect terms.

In this paper, we propose a novel Syntactic and Semantic Enhanced Graph Convolutional Network (SSEGCN) model for integrating the syntactic and semantic information of a sentence to solve the above issues. Firstly, SSEGCN captures the contextualized word representations with sentence encoder. Secondly, to model particular semantic correlations for different aspect terms, we propose an aspect-aware attention mechanism to combine with self-attention. The aspect-aware attention learns aspect-related semantic information, while self-attention learns global semantic of the sentence. We take the obtained attention scores as the initial adjacency matrices for GCN. Besides, to fully utilize syntactic structure to complement semantic information, rather than just syntactic first-order neighbor node information, we construct the syntactic mask matrices calculated by the different distances between words in the syntactic dependency structure of the sentence to learn structure information from local to global. Then, we combine adjacency matrices with syntactic mask matrices

to enhance the conventional GCN. Finally, a multi-layer graph convolution operation is implemented to obtain aspect-specific features for aspect term sentiment classification. We evaluate our approach on three benchmark datasets. The results show that our model is more effective than a range of baselines and achieves new state-of-the-art performance.

Our contributions are summarized as follows:

- We propose a SSEGCN model that effectively integrates syntactic structure and semantic correlation for ABSA task.
- We propose an aspect-aware attention mechanism combined with self-attention to learn the aspect-related semantic correlations and the global semantic of the sentence. Meanwhile, we construct syntactic mask matrices to complement with semantic information.
- Experimental results on three benchmark datasets show that the SSEGCN model achieves the state-of-the-art performance. The code and datasets involved in this paper are provided on Github¹.

2 Related Work

Aspect-based sentiment analysis is a fine-grained sentiment analysis task and generally treated as a classification problem. Earlier methods ([Jiang et al., 2011](#); [Kiritchenko et al., 2014](#)) manually defined some syntactic rules to predict the sentiment polarity of aspect term.

Most recent researches solve aspect-based sentiment analysis by utilizing attention-based neural network to model semantic correlation between context and aspect term ([Wang et al., 2016](#); [Chen et al., 2017](#); [Ma et al., 2017](#); [Liu et al., 2018](#); [Hu et al., 2018](#); [Wang et al., 2018](#); [Huang et al., 2018](#); [Fan et al., 2018](#)). Among them, ([Wang et al., 2016](#)) utilized attention mechanism to concentrate on different parts of a sentence to generate an attention vector for aspect sentiment classification. [Chen et al. \(2017\)](#) proposed a multi-layer attention network to infer the sentiment polarity for the aspect. [Ma et al. \(2017\)](#) introduced an interactive attention mechanism to generate the representations for aspects and contexts separately. [Wang et al. \(2018\)](#) designed a hierarchical aspect-specific attention model for aspect sentiment classification. [Hu et al. \(2018\)](#) employed a constrained attention network

¹<https://github.com/zhangzheng1997/SSEGCN-ABSA>

with both orthogonal regularization and sparse regularization.

Another trend explicitly leverages dependency tree. Syntactical information can establish relation connections between aspect and corresponding opinion words, GCN based on dependency tree have achieved impressive performance in ABSA. (Zhang et al., 2019; Sun et al., 2019) stacked a GCN layer to extract rich representations over dependency tree. Liang et al. (2020) build aspect-focused and inter-aspect graphs to learn aspect-specific sentiment features. Zhang and Qian (2020) constructed a global lexical graph to capture the word co-occurrence relation and combined a global lexical graph and a syntactic graph. Liang et al. (2021) constructed a sentiment enhancement graph by integrating the sentiment knowledge from SenticNet to consider the affective information between opinion words and aspect term. Tian et al. (2021) utilized dependency types and distinguished different relations in the dependency tree. However, these approaches generally ignore the effective fusion of syntactic structure and semantic correlation to obtain richer information.

3 Proposed SSEGCN

Figure 2 gives an overview of SSEGCN. In this section, we describe the SSEGCN model which is mainly composed of three components: the Input and Encoding Layer, the Attention Layer, the Syntax-Mask Layer and the GCN Layer. Next, components of SSEGCN will be introduced separately in the rest of the sections.

3.1 Input and Encoding Layer

Given a sentence-aspect pair (s, a) , where $s = \{w_1, w_2, \dots, w_n\}$. $a = \{a_1, a_2, \dots, a_m\}$ is an aspect and also a sub-sequence of the sentence s . We utilize BiLSTM or BERT (Devlin et al., 2019) as sentence encoder to extract hidden contextual representations. We first map each word into a low-dimensional real-value vector with embedding matrix $E \in \mathbb{R}^{|V| \times d_e}$, where $|V|$ is the size of vocabulary and d_e denotes the dimensionality of word embeddings. Thus, the sentence s has corresponding word embeddings $x = \{x_1, x_2, \dots, x_n\}$. With the word embeddings of the sentence, BiLSTM is leveraged to produce hidden state vectors $H = \{h_1, h_2, \dots, h_n\}$, where $h_i \in \mathbb{R}^{2d}$. H contains the sub-sequence $h_a = \{h_{a_1}, h_{a_2}, \dots, h_{a_m}\}$ corresponding to the aspect term representation.

Take H as initial nodes representation in SSEGCN. For the BERT encoder, we adopt “[CLS] sentence [SEP] aspect [SEP]” as input.

3.2 Attention Layer

Attention mechanism is a common way to capture the interactions between the aspect and context words (Fan et al., 2018). In this subsection, we combine aspect-aware attention and self-attention for better semantic features. Figure 2 shows multiple attention adjacency matrices. Here, we construct p matrices and the p is a hyper-parameter.

3.2.1 Aspect-aware Attention

Unlike sentence level sentiment classification task, aspect-based sentiment classification aims at judging sentiments of one specific aspect term in its context sentence, and thus calls for modeling particular semantic correlation based on different aspect terms. We propose the aspect-aware attention mechanism, which regards aspect term as query to attention calculation for learning aspect-related features:

$$A_{asp}^i = \tanh \left(H_a W^a \times (K W^k)^T + b \right) \quad (1)$$

where K is equal to H produced by encoding layer. $W^a \in \mathbb{R}^{d \times d}$ and $W^k \in \mathbb{R}^{d \times d}$ are learnable weights. We apply mean pooling on the h_a and then copy it n times to obtain $H_a \in \mathbb{R}^{n \times d}$ as aspect representation. Note that we use p -head aspect-aware attention to obtain attention score matrices for a sentence, A_{asp}^i indicates that it is obtained through the i -th attention head.

3.2.2 Self-Attention

Similarly, here A_{self} can be constructed by utilizing p -head self-attention (Vaswani et al., 2017) that captures the interaction between two arbitrary words of a single sentence. The calculation involves a query and a key:

$$A_{self}^i = \frac{Q W^Q \times (K W^K)^T}{\sqrt{d}} \quad (2)$$

where Q and K are both equal to H produced by encoding layer. $W^Q \in \mathbb{R}^{d \times d}$ and $W^K \in \mathbb{R}^{d \times d}$ are learnable weights. Then, we integrate aspect-aware attention score with self-attention score:

$$A^i = A_{asp}^i + A_{self}^i \quad (3)$$

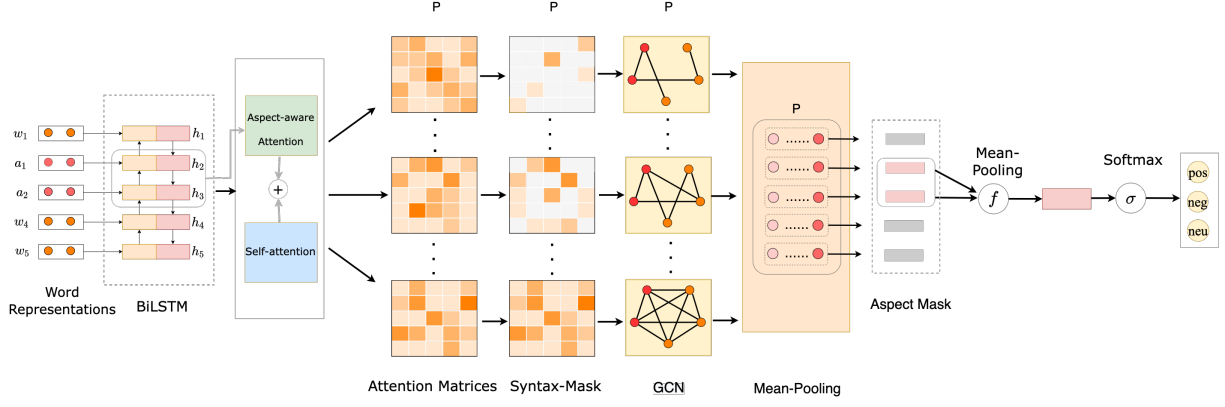


Figure 2: The overall architecture of SSEGCN.

where $A^i \in \mathbb{R}^{n \times n}$ is used as the input for the computation of the later Syntax-Mask Layer. For each A^i , it represents a fully connected graph.

3.3 Syntax-Mask Layer

In this section, we first introduce the syntactic mask matrix, and then mask each fully connected graph in terms of different syntactic distances. We treat the syntactic dependency tree as an undirected graph, and each token as a node. Then, we define the distance between node v_i and v_j as $d(v_i, v_j)$. Since there are multiple paths between nodes on the syntactic dependency tree, we define the distance of the shortest path as D :

$$D(i, j) = \min d(v_i, v_j) \quad (4)$$

In the previous part, the p -head attention mechanism can obtain p adjacency matrices. Therefore, we set the number of syntactic mask matrices based on different syntactic distances as the same as the number of attention heads. When syntactic distance is relatively small, our model can learn local information; on the contrary, if syntactic distance is relatively large, global structure information will be considered. The calculation of syntactic mask matrix M^k with threshold k can be formulated as:

$$M_{ij}^k = \begin{cases} 0, & D(i, j) \leq k \\ -\infty, & \text{otherwise} \end{cases} \quad (5)$$

where $k \in [1, p]$. To obtain global information and local feature, attention scopes are restricted by different syntactic distances.

$$M = \{M^1, \dots, M^k, \dots, M^p\} \quad (6)$$

$$A_{mask}^i = \text{softmax}(A^i + M^i) \quad (7)$$

Similarly, syntactic mask matrix based on the distance i is denoted as $A_{mask}^i \in \mathbb{R}^{n \times n}$.

3.4 GCN Layer

Since we have p different syntactic mask matrices, p graph convolution operations over $A_{mask} \in \mathbb{R}^{p \times n \times n}$ are required. If we denote h^{l-1} as the input state and h^l as the output state of the l -th layer, h^0 is the output of sentence encoding layer. Each node in the l -th GCN layer is updated according to the hidden representations of its neighborhoods:

$$h_i^l = \sigma \left(\sum_{j=1}^n A_{ij} W^l h_j^{l-1} + b^l \right) \quad (8)$$

where W^l is a linear transformation weight, b^l is a bias term, and σ is a nonlinear function. The final output representation of the l -layer GCN is $H^l = \{h_1^l, h_2^l, \dots, h_n^l\}$.

After aggregating node representation from each layer of SSEGCN, we obtain the final feature representation. We mask the non-aspect words of the output representation learned by the GCN layer to obtain aspect term representation. Moreover, an average pooling to retain most of the information in the aspect term representation h_a^l .

$$h_a^l = f(h_{a_1}^l, h_{a_2}^l, \dots, h_{a_m}^l) \quad (9)$$

where $f(\cdot)$ is a mean-pooling function applied over the enhanced aspect representation by GCN layer. Then, we feed h_a^l into a linear layer, followed by a softmax function to yield a probability distribution over polarity decision space:

$$p(a) = \text{softmax}(W_p h_a^l + b_p) \quad (10)$$

where W_p and b_p are the learnable weight and bias.

3.5 Training

Finally, the standard cross-entropy loss is used as our objective function:

$$L(\theta) = - \sum_{(s,a) \in \mathcal{D}} \sum_{c \in \mathcal{C}} \log p(a) \quad (11)$$

where \mathcal{D} contains all the sentence-aspect pairs and a represents the aspect appearing in sentence s . θ represents all the trainable parameters and \mathcal{C} is the collection of sentiment polarities.

4 Experiments

4.1 Datasets

We conduct experiments on three benchmark datasets for aspect-based sentiment analysis, including Restaurant and Laptop reviews from SemEval 2014 Task 4 (Pontiki et al., 2014) and Twitter (twitter posts) from Dong et al. (2014). Each aspect is labeled by one of the three sentiment polarities: positive, neutral and negative. The statistics for three datasets are reported in Table 1.

4.2 Implementation Details

For our experiments, we initialize word embeddings with 300-dimensional Glove vectors provided by Pennington et al. (2014). Additionally, we use 30-dimensional Part-of-Speech (POS) embeddings and 30-dimensional position embeddings which is the relative position of each word with respect to the aspect term in the sentence. Then, word embeddings, POS embeddings and position embeddings are concatenated as input word representations. All sentences are parsed by the Stanford parser². The batch size of all model is set as 16 and the number of GCN layers is 2. Besides, dropout function is applied to the input word representations of the BiLSTM and the dropout rate is set as 0.3. Our model is trained using the Adam optimizer with a learning rate of 0.002 to optimize the parameters. For SSEGCN+BERT, we employ the bert-base-uncased³ English version.

4.3 Baseline Comparisons

To comprehensively evaluate the performance of our model, we compare with state-of-the-art baselines:

²<https://stanfordnlp.github.io/CoreNLP/>

³<https://github.com/huggingface/transformers>

Dataset	Division	# Positive	# Negative	# Neutral
Rest14	Train	2164	807	637
	Test	727	196	196
Laptop	Train	976	851	455
	Test	337	128	167
Twitter	Train	1507	1528	3016
	Test	172	169	336

Table 1: Statistics for the three experimental datasets.

- 1) **IAN** (Ma et al., 2017) interactively learns the relationship between aspect and their context.
- 2) **RAM** (Chen et al., 2017) proposes a recurrent attention memory network to learn the sentence representation.
- 3) **TNet** (Li et al., 2018) employs a CNN model to extract salient features from target-specific embeddings by transformed BiLSTM embeddings.
- 4) **ASGCN** (Zhang et al., 2019) proposes to build GCN to learn syntactical information and word dependencies for ABSA.
- 5) **CDT** (Sun et al., 2019) utilizes a convolution over a dependency tree model to learn the representations of sentence features.
- 6) **TD-GAT** (Huang and Carley, 2019) proposes a target-dependent graph attention network for aspect level sentiment classification, which explicitly utilizes the dependency relationship among words.
- 7) **BiGCN** (Zhang and Qian, 2020) builds a concept hierarchy on both the syntactic and lexical graphs for sentiment prediction.
- 8) **kumaGCN** (Chen et al., 2020) combines information from a dependency graph and a latent graph to learn syntactic features.
- 9) **R-GAT** (Wang et al., 2020) proposes a relational graph attention network to encode the new tree reshaped by an ordinary dependency parse tree.
- 10) **DGEDT** (Tang et al., 2020) proposes a dependency graph enhanced dual-transformer network by jointly considering the flat representations from Transformer and graph-based representations from the dependency graph.
- 11) **DualGCN** (Li et al., 2021) designs a SynGCN module and a SemGCN module with orthogonal and differential regularizers.
- 12) **BERT** (Devlin et al., 2019) is the vanilla BERT model, which adopts “[CLS] sentence [SEP] aspect [SEP]” as input.
- 13) **R-GAT+BERT** (Wang et al., 2020) is the R-GAT model based on pre-trained BERT.
- 14) **DGEDT+BERT** (Li et al., 2021) is the DGEDT

model based on pre-trained BERT.

15) **BERT4GCN** (Zhang and Qian, 2020) integrates the grammatical sequential features from the PLM of BERT and the syntactic knowledge from dependency graphs.

16) **T-GCN** (Tian et al., 2021) utilizes dependency types to distinguish different relations in the graph and uses attentive layer ensemble to learn the contextual information from different GCN layers.

4.4 Main Results

To demonstrate the effectiveness of SSEGCN, we compare our model with previous works using accuracy and macro-averaged F1 as evaluation metrics, and report results in Table 2. Experimental results show that our SSEGCN model achieves best performance on the three datasets. In particular, GCN-based models take into account the syntactic structure of a sentence and capture long-term syntactic dependencies between the aspect word and the opinion word, hence outperform all attention-based methods. In GCN-based models, our SSEGCN learns structure information from local to global and considers aspect-related semantic information, and performs significantly better than the previous GCN-based models (i.e., CDT, TD-GAT, BiGCN, KuamGCN, R-GAT, DGEDT and DualGCN) that verifies the effectiveness of fusing syntactic and semantic information. On the other hand, we can observe that the basic BERT has been significantly better than most ABSA models. Combining our SSEGCN with BERT, the results show that the effectiveness of this powerful model is further improved, justifying that SSEGCN learns more syntactic and semantic knowledge can empower ABSA.

4.5 Ablation Study

As illustrated in Table 3, we further conduct an ablation study to examine the effectiveness of different modules in SSEGCN. The basic SSEGCN is regarded as a baseline model. First, we observe that removal of self-attention degrades the performance, verifying that global semantics of the sentence is necessary for ABSA. We can also notice that model without aspect-aware attention performs unsatisfactory, which indicates that the model lacks of the ability to capture aspect-related semantics, resulting in 1.70%, 1.58% and 1.47% reductions in accuracy on Restaurant, Laptop and Twitter, respectively. It indicates that aspect-aware attention is essential to capture the correlated semantic informa-

tion between aspect and contextual words. Second, removing syntactic mask matrix leads to dropping 0.90%, 1.42% and 0.88% in accuracy on Restaurant, Laptop and Twitter respectively, which indicates that syntactic mask matrix can assist GCNs to learn better syntactic structure information in original dependency trees. In addition, the removal of syntactic mask matrix and aspect-aware attention leads to a significant performance drop, which further indicates that these two components play crucial roles in SSEGCN for ABSA task. In summary, the ablation experimental results show that each component contributes to our entire model.

4.6 Case Study

To examine whether SSEGCN is able to capture syntax and semantic information for improving ABSA, we conduct case study with a few of sample sentences. Particularly, we compare SSEGCN with ATAE-LSTM, IAN and CDT in Table 4 which contain their predictions and the corresponding truth labels on these sentences. The notations P, N and O in the table represent positive, negative and neutral sentiment, respectively. The first sample “*great food but the service was dreadful!*” has two aspects (“*food*” and “*service*”) with contrast sentiment polarities, which may interfere with the prediction of the attention models. The second sample “*Biggest complaint is Windows 8.*” has the interfering word “*Biggest*”, which may neutralize the polarity of the word “*complaint*”. In the third sample, the key point is capturing the negated semantics which is most methods tend to ignore and easily make wrong predictions. The last two examples have no explicit sentiment expression. For the sentence “*Not as fact as I would have expect for an i5*”, CDT does not obtain the representation of the keyword “*not*” from a syntactic point of view and thus produces wrong prediction. Thus, the ability to learn integral semantics of a sentence is a significant factor for ABSA task. Our SSEGCN correctly predicts all the samples, and the results suggest that SSEGCN effectively combines syntactic structure and semantic information. Additionally, when dealing with complex sentences with implicit sentiment expressions, our SSEGCN can achieve better performance.

4.7 Visualization

To further demonstrate how our SSEGCN improves ABSA task, we use the two test examples in Figure 3 to visualize attention scores. For the sentence

Models	Restaurant		Laptop		Twitter	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
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
TNet (Li et al., 2018)	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., 2019)	82.30	74.02	77.19	72.99	74.66	73.66
TD-GAT (Huang and Carley, 2019)	81.2	-	74.0	-	-	-
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
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
DualGCN (Li et al., 2021)	84.27	78.08	78.48	74.74	75.92	74.29
Our SSEGNCN	84.72	77.51	79.43	76.49	76.51	75.32
BERT (Devlin et al., 2019)	85.97	80.09	79.91	76.00	75.92	75.18
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
BERT4GCN (Xiao et al., 2021)	84.75	77.11	77.49	73.01	74.73	73.76
T-GCN+BERT (Tian et al., 2021)	86.16	79.95	80.88	77.03	76.45	75.25
Our SSEGNCN+BERT	87.31	81.09	81.01	77.96	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
SSEGNCN	84.72	77.51	79.43	76.49	76.51	75.32
<i>w/o</i> self-attention	82.93	75.25	78.32	74.73	75.32	73.22
<i>w/o</i> aspect-aware attention	83.02	75.51	77.85	75.16	75.04	73.29
<i>w/o</i> syntactic mask matrix	83.82	75.60	78.01	74.33	75.63	74.34
<i>w/o</i> aspect-aware attention and syntactic mask matrix	82.75	75.02	77.37	73.35	74.89	72.40

Table 3: Experimental results of ablation study.

“*The staff should be a bit more friendly.*”, our model correctly identifies the sentiment of aspect term “*staff*” as negative. Our SSEGNCN model considers the semantics of the sentence and reduce the attention weight on the word “*friendly*” through syntactic distance mask and aspect-aware attention. For the harder example of multiple aspects with different sentiment polarities, our model also performs well. For sentence “*great food but the service was dreadful!*”, our model considers the aspect-related semantic correlations by introducing aspect-aware attention and combining syntactic mask matrices. Thus, SSEGNCN can accurately find the opinion words corresponding to each aspect term.

4.8 Effect of SSEGNCN Layers

In this section, we investigate the effect of the layer number ranging from 1 to 5 on the Restaurant and Laptop datasets. As shown in Figure 4, experimental results show that our model achieves the best performance with 2 layers. First, if the number

of GCN layer is set to 1, SSEGNCN can only learn local node information with syntactic distance of 1. Second, node representation will be over-smooth and obtain more redundancy information when the number of GCN layers is large, thus making model in poor performance.

4.9 Effect of Syntax-Mask

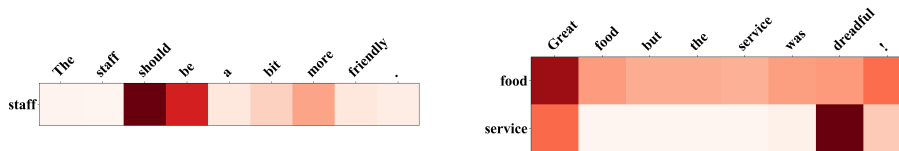
In this section, we further analyze the effect of the multiple different syntactic mask matrices on the performance of SSEGNCN in Restaurant and Laptop datasets. We conduct different numbers of the syntactic mask matrices from 1 to 7 and the results are demonstrated in Figure 5. We observe that the proposed SSEGNCN shows an upward trend with the increase of the number when the number of syntactic mask matrices is less than 5. One main reason may be that SSEGNCN can learn structure information from local to global when the syntactic distance becomes larger and SSEGNCN achieves remarkable results on five syntactic mask matrices. However, increasing the number of syntactic

Aspect	Sentence	ATAE-LSTM	IAN	CDT	SSEGNCN	Label
{food, service}	Great food but the service was dreadful!	(N, N)	(N, N)	(P, N)	(P, N)	(P, N)
{Windows 8}	Biggest complaint is Windows 8 .	(O)	(O)	(N)	(N)	(N)
{settings}	The settings are not user-friendly either.	(N)	(P)	(N)	(N)	(N)
{i5}	Not as fact as I would have expect for an i5	(P)	(P)	(O)	(N)	(N)
{Tequila}	If you are a Tequila fan you will not be disappointed.	(N)	(N)	(N)	(P)	(P)

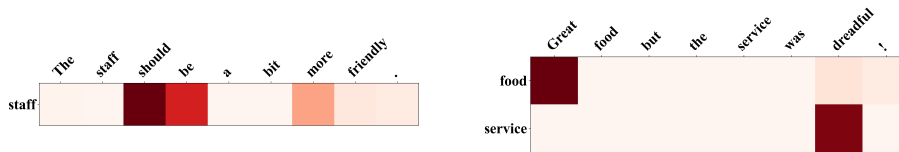
Table 4: Case study of our SSEGCN model compared with state-of-the-art baselines.



(a) The attention score of self-attention.



(b) The attention score of aspect-aware attention.



(c) The attention score of attention layer with syntax-mask layer.

Figure 3: Two visualized examples on how aspect-aware attention and syntactic mask matrix contribute to the attention layer.

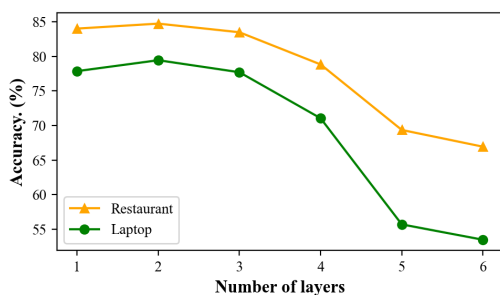


Figure 4: Effect of the number of SSEGCN layers.

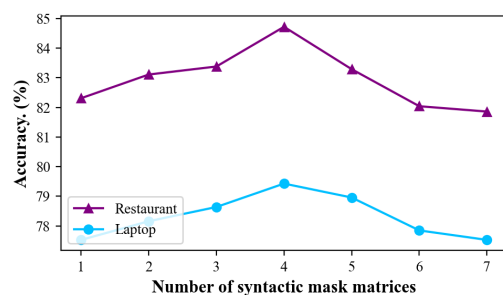


Figure 5: Effect of the number of syntactic mask matrices.

mask matrices from 5 to 7 leads to the performance degradation of SSEGCN. When the syntactic distance is greater than five, multiple syntactic mask matrices are fully connected matrices, and leads to

the introduction of noise and decline of the model performance.

5 Conclusion

In this paper, we propose a SSEGCN architecture which integrates semantic information along with the syntactic structure for ABSA task. Specifically, we first design an aspect-aware attention mechanism, which is responsible for learning aspect-related semantic information. Then, we combine it with the self-attention to compose the attention layer. Furthermore, we construct syntactic mask matrices of a sentence according to the different syntactic distances that learn local to global structure information. Consequently, we combine attention score matrices with syntactic mask matrices to fuse the semantic and syntactic information. Experimental results demonstrate the effectiveness of our approach on three public datasets.

Acknowledgements

This work was supported in part by the Shandong Provincial Natural Science Foundation (No.ZR2020MF149), in part by the Science and Technology Commission of Shanghai Municipality (21511100302).

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.
- 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, 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: Short papers)*, pages 49–54.
- 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.
- Mengting Hu, Shiwan Zhao, Li Zhang, Keke Cai, Zhong Su, Renhong Cheng, and Xiaowei Shen. 2018. Can: Constrained attention networks for multi-aspect sentiment analysis. *arXiv preprint arXiv:1812.10735*.
- Binxuan Huang and Kathleen M Carley. 2019. Syntax-aware aspect level sentiment classification with graph attention networks. *arXiv preprint arXiv:1909.02606*.
- 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. Springer.
- 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.
- 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.
- Ruifan Li, Hao Chen, Fangxiang Feng, Zhanyu Ma, Xiaojie Wang, and Eduard Hovy. 2021. Dual graph convolutional networks for aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6319–6329.
- Xin Li, Lidong Bing, Wai Lam, and Bei Shi. 2018. Transformation networks for target-oriented sentiment classification. *arXiv preprint arXiv:1805.01086*.
- Bin Liang, Hang Su, Lin Gui, Erik Cambria, and Ruifeng Xu. 2021. Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. *Knowledge-Based Systems*, page 107643.
- 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.
- Qiao Liu, Haibin Zhang, Yifu Zeng, Ziqi Huang, and Zufeng Wu. 2018. Content attention model for aspect based sentiment analysis. In *Proceedings of the 2018 World Wide Web Conference*, pages 1023–1032.

- Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*.
- Jeffrey Pennington, Richard Socher, and Christopher D 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, and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. *Proceedings of International Workshop on Semantic Evaluation at*.
- 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.
- 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.
- Yuanhe Tian, Guimin Chen, and Yan Song. 2021. Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2910–2922.
- 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 *Advances in neural information processing systems*, pages 5998–6008.
- Jingjing Wang, Jie Li, Shoushan Li, Yangyang Kang, Min Zhang, Luo Si, and Guodong Zhou. 2018. Aspect sentiment classification with both word-level and clause-level attention networks. In *IJCAI*, volume 2018, pages 4439–4445.
- Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020. Relational graph attention network for aspect-based sentiment analysis. *arXiv preprint arXiv:2004.12362*.
- 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.
- Zeguan Xiao, Jiarun Wu, Qingliang Chen, and Congjian Deng. 2021. Bert4gcn: Using bert intermediate layers to augment gcn for aspect-based sentiment classification. *arXiv preprint arXiv:2110.00171*.
- Chen Zhang, Qiuchi Li, and Dawei Song. 2019. Aspect-based sentiment classification with aspect-specific graph convolutional networks. *arXiv preprint arXiv:1909.03477*.
- 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.
- Pinlong Zhao, Linlin Hou, and Ou Wu. 2020. Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification. *Knowledge-Based Systems*, 193:105443.