International Journal of Data Warehousing and Mining
Volume 19 • Issue 1

Fusing Syntax and Semantics-Based Graph Convolutional Network for Aspect-Based Sentiment Analysis

Jinhui Feng, South China Normal University, China
Shaohua Cai, South China Normal University, China*
Kuntao Li, South China Normal University, China
Yifan Chen, South China Normal University, China
Qianhua Cai, South China Normal University, China*
Hongya Zhao, Shenzhen Polytechnic, China

ABSTRACT

Aspect-based sentiment analysis (ABSA) aims to classify the sentiment polarity of a given aspect in a sentence or document, which is a fine-grained task of natural language processing. Recent ABSA methods mainly focus on exploiting the syntactic information, the semantic information and both. Research on cognition theory reveals that the syntax an*/874d the semantics have effects on each other. In this work, a graph convolutional network-based model that fuses the syntactic information and semantic information in line with the cognitive practice is proposed. To start with, the GCN is taken to extract syntactic information on the syntax dependency tree. Then, the semantic graph is constructed via a multi-head self-attention mechanism and encoded by GCN. Furthermore, a parameter-sharing GCN is developed to capture the common information between the semantics and the syntax. Experiments conducted on three benchmark datasets (Laptop14, Restaurant14 and Twitter) validate that the proposed model achieves compelling performance comparing with the state-of-the-art models.

KEYWORDS

Aspect-Based Sentiment Analysis, Bert, Dependency Tree, Emotional Computing, Graph Convolutional Network, Multi-Head Self-Attention Mechanism, NLP, Syntactic and Semantics Information

INTRODUCTION

Aspect-based sentiment analysis (ABSA), a crucial task in fine-grained sentiment analysis, aims at automatically inferring the sentiment toward an aspect within its context. Generally, the sentiment of the given aspect is classified as positive, neural, or negative. Consider the following sentence: “I liked the atmosphere very much, but the food was not worth the price.” The sentiment of the “atmosphere” and “food” aspects are positive and negative, respectively.

DOI: 10.4018/IJDWM.319803
*Corresponding Author

So far, several state-of-the-art methods have been developed based on dual-channel graph convolutional networks (GCNs). These deal with syntactic and semantic information, which obtain satisfying results in ABSA tasks. Specifically, for both syntax and semantics processing, the fundamental idea is to reduce the distance between the aspect and its contextual words. In such a manner, the sentiment information of the aspect can be captured for sentiment classification. In the syntax-based approach, the dependency between the aspect and context is built and parsed to extract the syntactic information. Regarding long-term dependency, current models like ASGCN (Zhang et al., 2019) and CDT (Sun et al., 2019) exploit the GCN to establish adjacency matrices and derive the syntactic relation. However, a large amount of user-generated content involves informal grammatic style, such as text on Twitter. For this reason, exploiting semantic information for ABSA is highlighted.

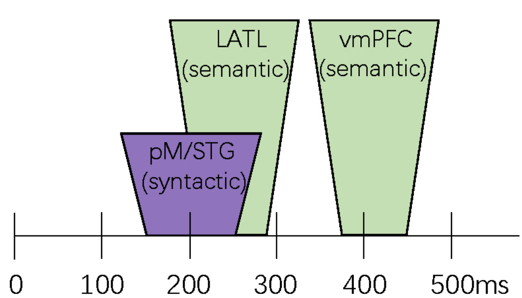
Most widely applied methods employ attention mechanisms to perform the interactions between aspect and its context. With the application of GCN, the attention matrix of the sentence is established and fed into GCN for semantic feature extraction (Guo et al., 2019). As such, the widespread use of syntactic- and semantic-GCN gives rise to the advances in dual-channel GCN methods. Generally, dual-channel GCNs are carried out in two ways. The first is to separately extract syntax and semantics before concatenating the syntactic and semantic representations (Pang et al., 2021). The second is to fuse these two categories of features during information encoding (Yan et al., 2021).

As with many facets of the natural language processing (NLP) task, a major challenge lies in teaching a computer to handle data that is distinctly human (Brooke, 2009). As such, the first step in ABSA method devising is to establish information flow, which directs the sentiment delivery from opinion words to the aspect. According to Pylkkänen (2020), the syntactic effects are performed earlier than the semantic effects during natural language comprehending. Concretely, measured by magnetoencephalography, the posterior middle/superior temporal gyrus (pM/STG), which processes the syntactic information, activates before the left anterior temporal lobe (LATL) and the ventromedial prefrontal cortex (vmPFC), whose purpose is to tackle the semantics (see Figure 1). Despite the order of precedence, the syntactic effects and semantic effects are difficult to distinguish from each other (Pylkkänen, 2019), indicated by the intersection of pM/STG and LATL in Figure 1.

In terms of recent ABSA approaches, two limitations can be observed. First, the semantics and syntax are generally processed in two separate channels. They do not consider the sequences. Second, in most cases, syntactic changes vary the meaning of the expression. The interaction between syntactic effect and semantic effect, referred to as the common information, remains neglected in ABSA tasks.

A fuse syntax and semantics-based graph convolutional network (FSSGCN) is developed for ABSA to mitigate the deficiencies of current ABSA methods. In the proposed model, the syntax structure of the sentence is resolved. Then, the semantic information is captured and fused with the syntactic information to enhance the sentiment delivery. Further, a common information module is built using GCN. It collects the information from both syntax and semantics to facilitate the sentiment classification. This work contains three main contributions as follows:

Figure 1.
MEG Results on Processing Stages of Language Comprehension



1. In line with the human cognition practice, a GCN-based method is established to deal with the syntactic information and semantic information in a cascading manner. The syntax dependency is extracted and integrated with the semantic information.
2. The parameter sharing scheme is carried out during the graph convolution. This is based on the common information of syntax and semantics.
3. Comprehensive experiments are performed on three benchmark datasets to evaluate the working performance of FSSGCN. Experimental results reveal that FSSGCN is a competitive alternative that achieves state-of-the-art performance.

GRAPH CONVOLUTIONAL NETWORKS

There are many non-Euclidean structure data in life, such as social networks, knowledge maps, and other data with rich information. Many are tree structures. In the past, CNN was often used to extract feature information, which can only be used for Euclidean structure data. As the graph convolutional network has been raised, the problem of extracting feature information from non-Euclidean structure data has been solved.

Graph convolution uses the topology graph in which nodes and edges are related to each other. It computes between nodes and edges to extract information from the graph. The input of the graph convolutional network includes the adjacency matrix and feature matrix. As the input information $G = \left(A, H^{(l)}\right)$, the operation process of the graph convolutional network can be expressed as:

$$H^{(l+1)} = \sigma\left(AH^{(l)}W^{(l+1)}\right) \tag{1}$$

where $A \epsilon R^{k\times k}$ is the adjacency matrix and $H^{(l)} = \left[h_1^{(l)}, h_2^{(l)}, \ldots, h_n^{(l)}\right]$ represents the set of feature matrices of all nodes. $W^{(l+1)}$ indicates the trainable parameter matrix of $\left(l+1\right) - th$ layer.

Through the calculation of graph convolutional network, the current node can fuse the information from neighbor nodes. By using the topological graph of specific relationship, such as syntax graph and semantic graph, the current node can fuse the information of specific relationship.

RELATED WORK

Many approaches have been proposed in the domain of ABSA. Among multiple ABSA methods, the focus lies in modeling the relation between the given aspect and its contexts. In this way, all approaches can be divided into the following three categories: (1) semantic-based models; (2) syntax-based models; and (3) semantic and syntax integration models.

Semantic-Based Models

The attention network, with its integration into deep neural networks, can establish the semantic relationship between the aspect and its contexts. As such, most semantic-based models are developed on the foundation of attention mechanism. Wang et al. (2016) applied the attention mechanism to long-short term memory (LSTM) to derive the aspect embedding and predict its sentiment polarity. Ma et al. (2017) proposed the interactive attention networks (IAN) for determining the attention weights of the contexts. Thus, it represented the aspect and its collocative context. More recently, the transformer model tackled sentence encoding with only attention mechanisms. It achieves a satisfying working performance in various tasks. The transformer's derived methods are also well-documented in ABSA tasks. The bidirectional encoder representations from transformer (BERT) model works on developing multi-layer bidirectional transformer encoders to comprehensively learn semantic information and obtain more precise representations. Similarly, target-dependent BERT aims to generate the accurate aspect representations with a simpler model structure (Gao et al., 2019). To exploit the superiorities of both RNN and multi-head attention mechanism, R-transformer is proposed to capture local structures and global long-term dependencies in sequences (Zhou & Wang, 2019).

Syntax-Based Models

Regarding syntax dependency tree, the syntax-based models are employed to reduce the long-distance dependencies between the aspect and its opinion words. Recent publications reveal that the application of GCN holds great promise in ABSA tasks. Both Zhang et al. (2019) and Sun et al. (2019) propose aspect-oriented GCN by using the syntax dependency tree, which aims to learn the syntactic information. Besides, in the BERT4GCN model, a lexical graph is constructed based on which bipolar interactive GCN is devised to learn the syntax and lexical information (Zhang & Qian, 2020).

Dual-Channel Models

The dual-channel models are proposed to integrate syntactic and semantic information to optimize the sentiment classification results. DGEDT uses both transformer and GCN as encoders to interactively learn the semantic information and syntactic information (Tang et al., 2020). Then, the sentence representation is generated by using two categories of features. DualGCN (Li et al., 2021) and DMGCN (Pang et al., 2021) build a semantic graph via multi-head self-attention mechanism. It takes dual-channel GCN to encode the sentence syntax and semantics, respectively. In the model SEGCN, an interactive dual-channel GCN is developed to process the syntactic dependency trees and semantic graphs using cosine similarity (Yan et al., 2021).

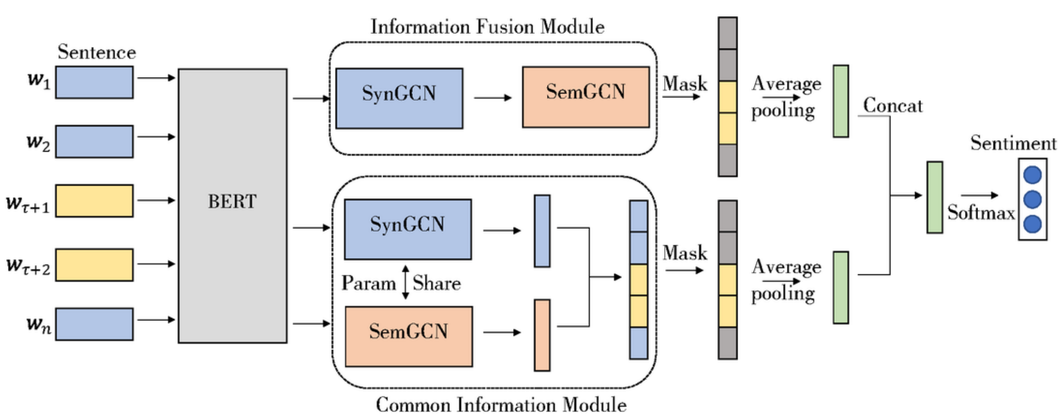
THE APPROACH

The architecture of FSSGCN is presented in Figure 2. The proposed model contains five major components: (1) sentence encoder; (2) information fusion module; (3) common information module; (4) feature fusion module; and (5) sentiment classifier. Details of each component are described as follows.

Sentence Encoder

Let $x = \left\{w_1, w_2, \ldots, w_{\tau+1}, \ldots, w_{\tau+m}, \ldots, w_n\right\}$ be a n-word sentence containing a specific aspect $a = \left\{w_{\tau+1}, \ldots, w_{\tau+m}\right\}$. The sequence [CLS] x [SEP] a [SEP] is sent to the pre-trained Bert model

Figure 2.
General Framework of the FSSGCN



to obtain the sentence embedding. This contributes to the explicit interaction between the aspect and its contexts. Thus, the aspect-oriented word representations can be derived. Then, the hidden states of the encoder are taken as the outcomes for further processing.

Information Fusion Module

The framework of the information fusion module is given in Figure 3. Notably, the module has two GCNs. The first is for syntax encoding (namely SynGCN). The second is for semantic learning (namely SemGCN). The word embeddings from BERT are first fed into SynGCN. Then, the word representations that incorporate the syntax are sent to the SemGCN for information fusion. By exploiting the syntax dependency tree, the distance between the aspect and its opinion words is effectively reduced. This is based on which syntactic information can be captured. In this way, the authors transform the syntax dependency tree of the sentence into the syntactic adjacency matrix $A_{syn} \in \mathbb{R}^{n\times n}$ (Zhang et al., 2019). Specifically, the authors take $A_{ij} = 1$ to characterize the connection between node i and node j (and $A_{ij} = 0$). With the input of sentence embedding, the syntax graph $G_{syn} = \left(A_{syn}, H^{c}\right)$ is established. At this stage, the GCN is employed to capture the syntactic information:

$$\tilde{A}_{syn} = \tilde{D}^{-\frac{1}{2}}\left(A_{syn} + I_{f}\right)\tilde{D}^{-\frac{1}{2}} \tag{2}$$

$$GCN\left(\tilde{A}, H^{(l)}, W^{(l+1)}\right) = ReLU\left(\tilde{A}H^{(l)}W^{(l+1)}\right) \tag{3}$$

$$H_{syn}^{(l+1)} = GCN\left(\tilde{A}_{syn}, H_{syn}^{(l)}, W_{syn}^{(l+1)}\right) \tag{4}$$

with:

$$H_{syn}^{(0)} = H^{c} \in R^{n\times d_{hid}} \tag{5}$$

where I_{f} is the identity matrix; $\tilde{A}_{syn}$ stands for A_{syn} with self-loop; $H_{syn}^{(0)}$ is derived from the word representation of Bert encoder; and $W_{syn}^{(l+1)} \in R^{\left(d_{hid}+l*d_{gcn}\right)\times d_{hid}}$ indicates the trainable parameter matrix of $\left(l+1\right)-th$ layer.

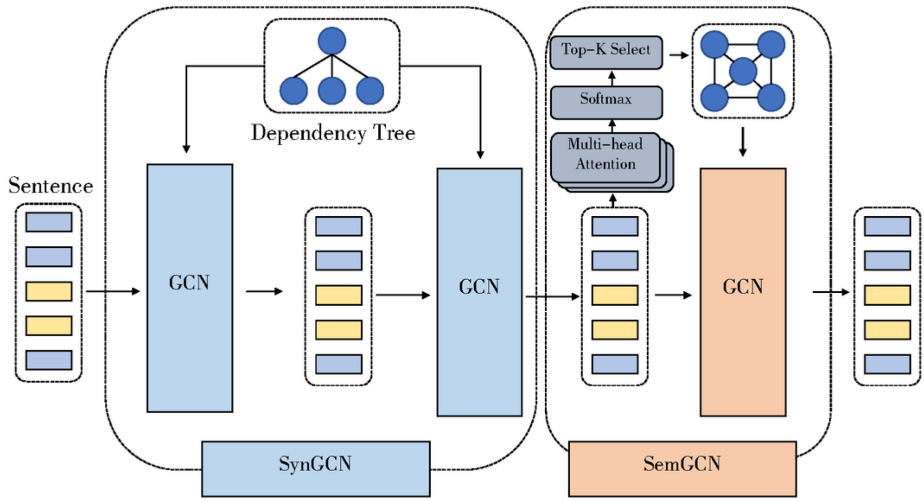
The syntactic information is incorporated into the node representation through graph convolution. Conforming to the cognition procedure, the hidden layer state of SynGCN is sent to the SemGCN for information fusion. Notably, the multi-head self-attention network is employed to identify valuable contextual information (Guo et al., 2019). This is further integrated into the aspect representation. The sentence is characterized by a fully connected graph; therefore, the application of the multi-head self-attention mechanism can build the relationship between the aspect and its contexts in a soft pruning manner.

For semantic information processing, the input of semantic-learning GCN is derived as:

$$H_{sem}^{(0)} = \left[H_{syn}^{(0)}; H_{syn}^{(1)}\right] \tag{6}$$

On this occasion, a k-head attention network is used to construct the semantic adjacency matrices $A_{sem,i}^{(0)}\left(i = 1,\ldots,k\right)$ via multi-head attention:

Figure 3.
Framework of Information Fusion Module



$$A_{sem,i}^{(0)} = \frac{\left(H_{sem}^{(0)} W_{sem,k}^{(0)}\right)\left(H_{sem}^{(0)} W_{sem,q}^{(0)}\right)^{T}}{\sqrt{d_{head}}} \tag{7}$$

with:

$$d_{head} = \frac{d_{hid}}{K} \tag{8}$$

where $W_{sem,k}^{(0)}$ and $W_{sem,q}^{(0)}$ are trainable weight matrices.

The Softmax function is carried out to compute the probability of each semantic adjacency matrix, which is:

$$A_{sem}^{(0)} = argmax\left[softmax\left(A_{sem,1}^{(0)}, \ldots, A_{sem,K}^{(0)}\right)\right] \tag{9}$$

Subsequently, the top-k selection is employed to pick the semantic adjacency matrix of the largest probability. Thus, it maintains the substantial contextual information within the matrix:

$$A_{sem}^{(0)} = top - k(A_{sem}^{(0)}) \tag{10}$$

In this way, the matrix $A_{sem}^{(0)}$ is the attentive matrix that most relates to the sentence semantics. The authors construct the semantic graph as $G_{sem} = \left(A_{sem}^{(0)}, H_{sem}^{(0)}\right)$. The GCN is taken to learn the semantic information, which is written as:

$$H_{sem}^{(1)} = GCN\left(\tilde{A}_{sem}^{(0)}, H_{sem}^{(0)}, W_{sem}^{(1)}\right) \tag{11}$$

Note that the aspect node within the graph generated by the multi-head attention mechanism connects to every other word. In turn, the information of all other nodes is inevitably incorporated. The noise can also be introduced during information fusion. For this reason, the authors use the top-k selection mechanism to sparse the fully connected graph based on the k edges with the largest attention weights preserved. If k = 2, the top-k mechanism is performed as presented:

$$\left[0.2, 0.5, 0.7, 0.1\right] \xrightarrow{top-k} \left[0, 1, 1, 0\right] \tag{12}$$

To simplify the information fusion, the authors re-write the processing within this module as:

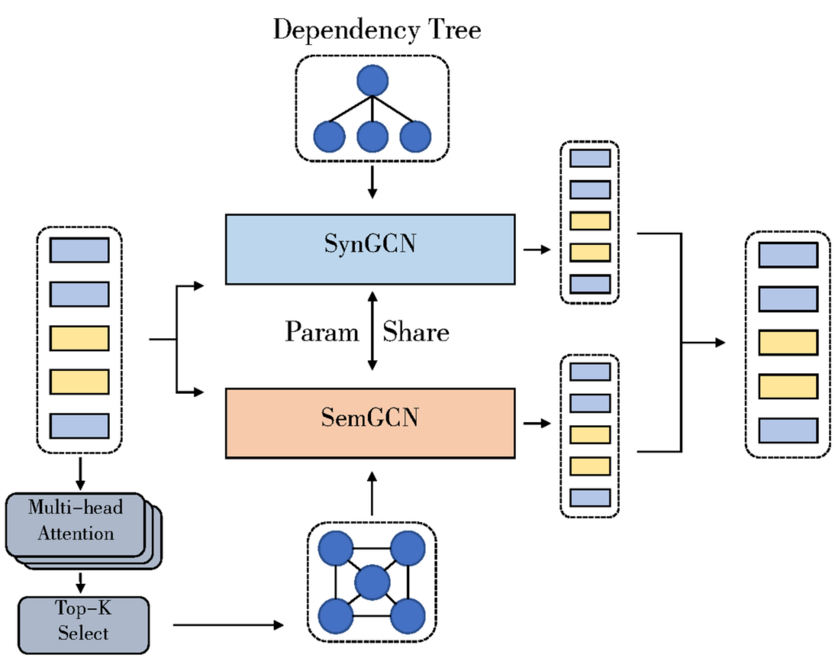
$$H_{syn}^{(l)} = SynGCN\left(A_{syn}, H_{syn}^{(l-1)}, W_{syn}^{(l)}\right) \tag{13}$$

$$H_{sem}^{(l+1)} = SemGCN\left(A_{sem}, H_{syn}^{(l)}, W_{sem}^{(l+1)}\right) \tag{14}$$

Common Information Module

As discussed, the sentence semantics can vary in line with the variation of syntactic structure. In other words, the syntax and semantics are not independent from each other. The interaction between them must be considered during processing. Thereby, the common information is highlighted to enhance the sentiment delivery. Figure 4 exhibits the framework of common information module.

Figure 4.
Framework of Common Information Module



The authors design a common information module with reference to Wang et al. (2020). They use a parameter-sharing GCN to extract the common information between the syntactic space and semantic space. The schematic diagram of the module is as follows.

The common information module is established on the foundation of the SemGCN and SynGCN. The sentence hidden states of Bert H^c are taken as the common inputs of both GCNs, which can be written as:

$$H_{c-syn}^{(l+1)} = SynGCN\left(A_{syn}, H^c, W_{com}^{(l+1)}\right) \tag{15}$$

$$H_{c-sem}^{(l+1)} = SemGCN\left(A_{sem}, H^c, W_{com}^{(l+1)}\right) \tag{16}$$

where $W_{com}^{(l+1)} \in R^{\left(d_{hid}+l*d_{gcn}\right)\times d_{hid}}$ is the parametric matrix of the l-th GCN-layer. With l times iteration, the common information is integrated into both the syntax representation $H_{c-syn}^{(l+1)}$ and semantic representation $H_{c-sem}^{(l+1)}$. Then, the sentence representation, with common information incorporated, can be computed as:

$$H_{com}^{(l+1)} = \frac{H_{c-syn}^{(l+1)} + H_{c-sem}^{(l+1)}}{2} \tag{17}$$

Feature Fusion

The outputs of the information fusion module and the common information module are utilized to obtain a precise feature representation of the aspect. The mask operation is separately conducted on the sentence representations from both modules to preserve the aspect words. Then, the average pooling is performed on the masked representations to generate two corresponding aspect vectors:

$$h_{sem} = f\left(mask\left(H_{sem}^{(l+1)}\right)\right) \tag{18}$$

$$h_{com} = f\left(mask\left(H_{com}^{(l+1)}\right)\right) \tag{19}$$

where $f\left(\cdot\right)$ stands for the average pooling and $mask\left(\cdot\right)$ is the mask operation. Specifically, the masking indicates the setting of all words to 0 (except for the aspect word):

$$mask = \begin{cases} 0, 1 \le t < \tau + 1, \tau + m < t < n \\ 1, \tau + 1 \le t \le \tau + m \end{cases} \tag{20}$$

The final aspect representation is the concatenation of h_{sem} and h_{com}:

$$h_a = \left[h_{sem}; h_{com}\right] \tag{21}$$

The authors, thus, send h_a to the Softmax classifier. The sentiment distribution of the given aspect can be classified as:

$$\hat{y} = softmax\left(h_a W_1^T + b_1\right) \tag{22}$$

where W_1^T and b_1 denote the trainable weight matrix and the bias, respectively.

Model Training

The training of the model is carried out to minimize the cross-entropy loss between the predicted outcome and real result. The loss function is given by:

$$L = -\sum_i \sum_{j=1}^{P} y_i^j log\hat{y}_i^j \tag{23}$$

where the subscript *i* represents the *i*th sample and the subscript *j* represents the *j*th sentiment polarity. P stands for the sentiment classes. y is the real sentiment distribution and $\hat{y}$ is predicted.

EXPERIMENT

Experimental Setting

Experiments are carried out on several benchmark datasets, including Restaurant and Laptop from Semeval (Kirange et al., 2014) and Twitter (Dong et al., 2014). Each sample from the experimental dataset is labeled as positive, neutral, or negative. Table 1 provides details of each dataset.

The word embeddings are initialized by using the pre-trained Bert model with a lexicon size of 30,522 and a word-embedding dimension of 768. For the multi-head attention network, the head number is 12. The hidden layer dimension is 12. The learning rate is set to 0.00001. Moreover, the Adam optimizer is adopted with the weight of the regularization term, set as 0.00001.

For each method in the experiment, the result is averaged over three runs. All parameters are randomly initialized for every run. In addition, the authors take accuracy and macro-F1 score as evaluation metrics.

The process of calculating accuracy is expressed as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{24}$$

where *TP*, *TN*, *FP*, and *FN* represent the number of true positive samples, true negative samples,incorrectly identified positive samples, and incorrectly identified negative samples, respectively.

Table 1.
Statistics of Dataset

Datasets		Classes	Positive	Neutral	Negative
Twitter	Trian	3	1,561	3,127	1,560
	Test	3	173	346	173
Laptop14	Train	3	994	464	870
	Test	3	341	169	128
Restaurant14	Train	3	2,164	637	807
	Test	3	728	196	196

The macro-F1 score represents the average of the F1 scores of the three types of sentiments (positive, neutral, and negative):

$$Macro - F1 = \frac{F1 - score_1 + F1 - score_2 + F1 - score_3}{3} \tag{25}$$

where $F1 - score_1$, $F1 - score_2$ and $F1 - score_3$ are the F1 scores of the three types of sentiments. It means that the weight of F1 value of three types of sentiments is equal.

Baseline Methods

To confirm the working performance of the proposed model, seven state-of-the-art approaches are taken as baselines. Notably, all word embeddings of these methods are generated via the Bert encoder.

- **BERT-SPC:** The basic BERT model is devised based on bidirectional transformer. The sentence, together with the aspect, is fed into a finetuned Bert model for sentiment classification (Song et al., 2019).
- **R-GAT+BERT:** An aspect-oriented dependency tree is constructed and pruned. This is further encoded via a relational graph attention network (Wang & Shen et al., 2020).
- **DGEDT+BERT:** A mutual bi-affine model structure is developed to fuse the flat representations learned by the traditional transformer and the graph-based representations learned via the dependency graph (Tang et al., 2020).
- **TDGAT+BERT:** A multi-layer graph attention network is established. This is capable of propagating sentiment features from important syntax neighboring words to the aspect (Huang & Carley, 2019).
- **InterGCN+BERT:** Considering the relation among distinguishing aspects, a graph-aware model, based on interactive graph convolution, is proposed to learn the syntactic representation of the given aspect (Liang et al., 2020).
- **DMGCN+BERT:** A multi-channel GCN method is designed to encode the syntax, semantics, and correlated information from the generated graph (Pang et al., 2021).
- **DualGCN+BERT:** With the transformation of the syntax adjacency matrix, a dual-channel GCN is proposed to deal with both the syntactic information and semantic information (Li et al., 2021).

Result and Analysis

Experimental Results

Table 2 shows the experimental results on the three datasets. Among all the evaluation settings, the proposed model is a competitive alternative that outperforms most baselines.

When comparing with the syntactic-based models (i.e., R-GAT, DGEDT, TDGAT, and InterGCN), a considerable improvement is obtained in both accuracy and macro-F1 score. The main reason is that the integration of semantic information contributes to the sentiment delivery to a large extent. Likewise, in contrast to BERT-SPC, the model shows superiority in building the syntax structure for sentiment classification.

Compared to the dual-channel GCN approaches, the model is best-performing on both Restaurant and Twitter. The proposed method can exploit and fuse the semantics and syntax in a more dedicated manner based on which sentiment information of the aspect can be extracted. Regarding cognition practice, the aspect representation in the model is both informative and accurate. This optimizes the working performance.

However, the accuracy on Laptop fails to overperform DualGCN. A possible explanation is that the syntactic probability matrix constructed in DualGCN performs better in fusing the syntactic information on the Laptop dataset.

The model provides a state-of-the-art result by integrating the syntax and semantics of the sentence. Therefore, it is reasonable to expect a more satisfying working performance in ABSA.

Table 2.
Experimental Results

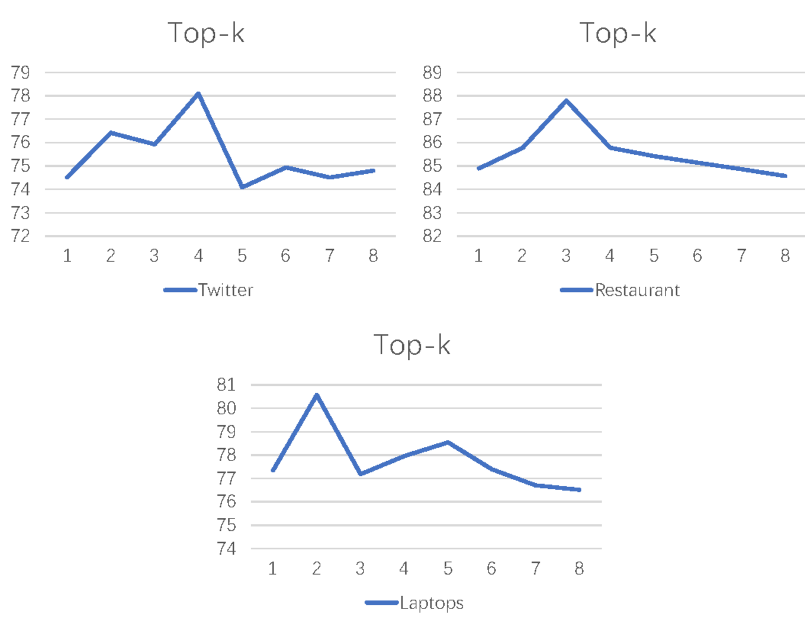
Model	Restaurant14		Laptop14		Twitter	
	Accuracy	F1	Accuracy	F1	Accuracy	F1
BERT-SPC	86.15	80.29	78.48	74.74	75.92	74.29
R-GAT+BERT	86.60	81.35	78.21	74.07	76.15	74.88
DGEDT+BERT	86.30	80.00	79.80	75.60	77.90	75.40
TDGAT+BERT	83.0	--	80.1			
InterGCN+BERT	87.12	81.02	82.87	79.32		
Dual-Channel Models						
DMGCN+BERT	87.66	82.79	80.22	77.28	78.06	77.36
DualGCN+BERT	87.13	81.16	81.80	78.10	77.40	76.02
FSSGCN+BERT	**87.79**	**82.72**	**80.58**	**77.75**	**78.09**	**76.37**

Note: Results are cited from the original paper of the baseline where "-" the result is not available.

Hyperparameter Settings

To investigate the effect of the hyperparameter on working performance, the authors optimized the value of k for top-k selection. Figure 5 presents the setting of k on different datasets. According to Figure 5, accuracy reaches its highest level when the k equals 3, 2, and 4 for Restaurant, Laptop, and Twitter, respectively. With the increment of k, the working performance declines to a certain extent. The main reason is that the node number in the semantic graph increases along with k, which can lead to the incorporation of irrelevant information. In such a manner, the generated aspect representation contains semantic noise, impacting the sentiment classification.

Figure 5.
Setting of k on Different Datasets



Ablation Study

To verify the effect of the information fusion module and common information module, the authors conduct an ablation study based on the FSSGCN. First, removing the information fusion module leads to the accuracy drop of 3.35%, 2.82%, and 2.38% on Rest14, Lap14, and Twitter, respectively. This indicates that the information fusion module proposed by the authors can capture the syntactic and semantic information. It also demonstrates the significance of the information fusion module. Second, due to the loss of the common information module, the accuracy drops 1.29%, 1.05%, and 2.63% on Rest14, Lap14, and Twitter, respectively. This shows that the common information of syntax and semantics is useful for sentiment classification. Overall, the results of the ablation study showed that both modules of the approach contribute to the accuracy. See Table 3.

Case Study

The effectiveness of the model is further validated by visualizing the attentive weights of context words in Figure 6. The aspect for the first sentence is "waiter." The second sentence uses "entrees" and "duck." In Figure 6(a), two words that convey positive sentiment ("rude" and "disinterested") are recognized. In Figure 6(b), the sentiment word toward "entrees" is "great." Regarding the word "duck," the attention weights are assigned to both "suggest" and "great" (with the integration of syntax and semantics). Thus, the proposed model is distinct in identifying the sentiment of the aspect, which facilitates the task of ABSA.

CONCLUSION

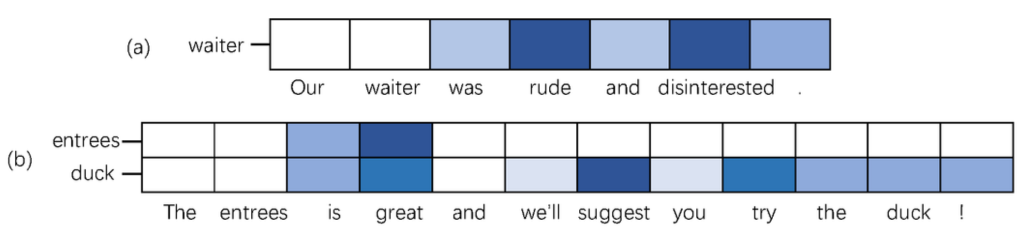
On the task of ABSA, the authors propose a GCN-based method that integrates the syntax and semantics in line with the human cognition principle. The sentiment information fusion process is

Table 3.

Model	Restaurant14	Laptop14	Twitter
	Accuracy	Accuracy	Accuracy
FSSGCN w/o h_{sem}	84.43	77.76	75.71
FSSGCN w/o h_{com}	86.50	79.53	75.46
FSSGCN	**87.79**	**80.58**	**78.09**

FSSGCN w/o h_{sem} : removing the information fusion module
FSSGCN w/o h_{com} : removing the common information module.

Figure 6.
Visualization of Attention Mechanism



carried out to encode the syntactic information and semantic information in sequence. Furthermore, regarding parameter sharing, the common information of syntax and semantics can be precisely captured. Experiments are conducted on three publicly available datasets to evaluate the effectiveness of the proposed model. The experimental results verify that FSSGCN is a competitive alternative. It achieves advanced performance compared with the baselines.

ACKNOWLEDGMENT

This work was supported by the Characteristic Innovation Projects of Guangdong Colleges and Universities (Nos. 2018KTSCX049), the Science and Technology Plan Project of Guangzhou under Grant Nos. 202102080258 and 201903010013.

CONFLICT OF INTEREST

The authors of this publication declare there is no conflict of interest.

FUNDING AGENCY

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Brooke, J. (2009). *A semantic approach to automated text sentiment analysis* [Doctoral dissertation, Dept. of Linguistics-Simon Fraser University].

Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., & Xu, K. (2014, June). 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, pp. 49–54). ACL. doi:10.3115/v1/P14-2009

Gao, Z., Feng, A., Song, X., & Wu, X. (2019). Target-dependent sentiment classification with BERT. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 154290–154299. doi:10.1109/ACCESS.2019.2946594

Guo, Z., Zhang, Y., & Lu, W. (2019, July). Attention guided graph convolutional networks for relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 241–251). ACL. doi:10.18653/v1/P19-1024

Huang, B., & Carley, K. M. (2019, November). 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)* (pp. 5469-5477). ACL. doi:10.18653/v1/D19-1549

Kirange, D. K., Deshmukh, R. R., & Kirange, M. D. K. (2014). *Aspect based sentiment analysis semeval-2014 task 4. Asian Journal of Computer Science and Information Technology* (Vol. 4). AJCSIT.

Li, R., Chen, H., Feng, F., Ma, Z., Wang, X., & Hovy, E. (2021, August). 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, pp. 6319–6329). ACL. doi:10.18653/v1/2021.acl-long.494

Liang, B., Yin, R., Gui, L., Du, J., & Xu, R. (2020, December). 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* (pp. 150–161). ACL. doi:10.18653/v1/2020.coling-main.13

Ma, D., Li, S., Zhang, X., & Wang, H. (2017, August). Interactive attention networks for aspect-level sentiment classification. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence* (pp. 4068–4074). doi:10.24963/ijcai.2017/568

Pang, S., Xue, Y., Yan, Z., Huang, W., & Feng, J. (2021, August). Dynamic and multi-channel graph convolutional networks for aspect-based sentiment analysis. In *Findings of the Association for Computational Linguistics* (pp. 2627–2636). ACL-IJCNLP. doi:10.18653/v1/2021.findings-acl.232

Pylkkänen, L. (2019). The neural basis of combinatory syntax and semantics. *Science*, *366*(6461), 62–66.

Pylkkänen, L. (2020). Neural basis of basic composition: what we have learned from the red-boat studies and their extensions. *Philosophical Transactions of the Royal Society B, 375*(1791), 20190299.

Song, Y., Wang, J., Jiang, T., Liu, Z., & Rao, Y. (2019). Attentional encoder network for targeted sentiment classification. *arXiv preprint arXiv:1902.09314*.

Sun, K., Zhang, R., Mensah, S., Mao, Y., & Liu, X. (2019, November). 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)* (pp. 5679–5688). ACL.

Tang, H., Ji, D., Li, C., & Zhou, Q. (2020, July). Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 6578–6588). ACL.

Wang, K., Shen, W., Yang, Y., Quan, X., & Wang, R. (2020, July). Relational graph attention network for aspect-based sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 3229–3238). ACL.

Wang, W., Pan, S. J., Dahlmeier, D., & Xiao, X. (2016, November). Recursive neural conditional random fields for aspect-based sentiment analysis. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 616–626). ACL.

Wang, X., Zhu, M., Bo, D., Cui, P., Shi, C., & Pei, J. (2020, August). Am-gcn: Adaptive multi-channel graph convolutional networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 1243–1253). ACL.

Yan, Z., Pang, S., & Xue, Y. (2021, October). Semantic Enhanced Dual-Channel Graph Communication Network for Aspect-Based Sentiment Analysis. In *CCF International Conference on Natural Language Processing and Chinese Computing* (pp. 531–543). Springer.

Zhang, C., Li, Q., & Song, D. (2019, January). Aspect-based sentiment classification with aspect-specific graph convolutional networks. In EMNLP/IJCNLP, (1).

Zhang, M., & Qian, T. (2020, November). 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)* (pp. 3540–3549). ACL.

Zhou, Z., & Wang, Q. (2019). R-transformer network based on position and self-attention mechanism for aspect-level sentiment classification. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 127754–127764.

Jinhui Feng received the B.S. degree from South China Normal University, Guangzhou, China, in 2020 and currently working toward the M.S. degree in electronic information from South China Normal University, Foshan, China.

Cai Shaohua, senior experimenter/chairman of department trade union, Master of Science of Jinan University, majoring in computer science, educational technology and e-commerce.

Kuntao Li received the B.S. degree from Jilin Agricultural University, Changchun, China, in 2022 and currently working toward the M.S. degree in Communication Engineering from South China Normal University, Foshan, China,in 2025.

Yifan Chen received the B.S. degree in Communication Engineering from Fujian Normal University, Fujian, China, in 2022 and the M.S. degree in Communication Engineering from South China Normal University, Foshan, China, in 2025.

Qianhua Cai received her master degree in the school of computer science of South China Normal University. After graduation, she joined the School of Electronics and Information Engineering of South China Normal University. Research interests: data mining, text data analysis.

Hongya Zhao received her Ph.D. degree from Hong Kong Baptist University in 2006. And then she worked at City University of Hong Kong and M.D. Andercen Cancer Center UT. Currently she is the Professor of the Industrial Center at Shenzhen Polytechnic. Her current research interests include statistics, pattern recognition, and machine learning.