Research article

Word distance assisted dual graph convolutional networks for accurate and fast aspect-level sentiment analysis

Jiajia Jiao*, Haijie Wang, Ruirui Shen and Zhuo Lu

College of Information Engineering, Shanghai Maritime University, Shanghai, China

* Correspondence: Email: jiaojiajia@shmtu.edu.cn.

Abstract: Aspect-level sentiment analysis can provide a fine-grain sentiment classification for inferring the sentiment polarity of specific aspects. Graph convolutional network (GCN) becomes increasingly popular because its graph structure can characterize the words' correlation for extracting more sentiment information. However, the word distance is often ignored and cause the cross-misclassification of different aspects. To address the problem, we propose a novel dual GCN structure to take advantage of word distance, syntactic information, and sentiment knowledge in a joint way. The word distance is not only used to enhance the syntactic dependency tree, but also to construct a new graph with semantic knowledge. Then, the two kinds of word distance assisted graphs are fed into two GCNs for further classification. The comprehensive results on two self-collected Chinese datasets (MOOC comments and Douban book reviews) as well as five open-source English datasets, demonstrate that our proposed approach achieves higher classification accuracy than the state-of-the-art methods with up to 1.81x training acceleration.

1. Introduction

As a significant branch in the field of natural language processing, sentiment analysis aims to extract the sentiment polarity from an input text. Unlike the chapter-level or sentence-level sentiment analysis, the aspect-level sentiment analysis can identify the fine-grain polarity of different aspects in a sentence [1–3]. For example, in the sentence: "While the lesson content is substantial, the teacher's

speaking pace is excessively rapid!", the sentiment classification of "content" is positive and that of "speaking" is negative.

With the development of deep learning methods, the aspect-level sentiment analysis has achieved remarkable performance in the past few years. Wang et al. [4] proposed a long short-term memory (LSTM) network based on an attention mechanism to extract more information from different parts of a sentence. Ma et al. [5] proposed an interactive attention network to learn the attention weights of context and aspect words for good representations of target and context, respectively. However, the attention mechanism is not sufficient to capture the syntactic dependency between aspect words and context words. In order to solve this problem, a GCN based on dependency trees was proposed by Sun et al. [6]. Since, a variety of GCN variants are designed such as graph attention networks [7,8], multi-channel GCN [9], heterogenous GCNs [10] and dual GCNs [11–15] to capture more syntactic information and sematic information for accurate classification.

However, the existed works of advanced GCNs pay more attention on the sentiment knowledge information of individual words in a comment sentence and ignore the distance information between context words and aspect words. It is proven that the word distance is important for aspect-level sentiment analysis [16,17]. Therefore, it is significant to integrate word distance information into a graph network effectively facilitates the model in extracting dependency relationships between contextual words and specific aspects. Therefore, we first construct a conventional dependency graph for each sentence based on a dependency tree to capture the sentence's syntactic information. Then, through combining sentiment dependency relationships between contextual words and aspect words with distance relationships between words, all this information is fused into the dependency graph. Based on the syntactic dependency relationships and sentiment information of the sentence, a sentiment dependency graph for specific aspects is built. Moreover, considering the syntactic dependency relationships and distance information of the sentence, a distance-enhanced dependency graph for specific aspects is also constructed. Finally, the distance-enhanced and sentiment dependency graphs are input into a dual GCN model to obtain the graph representation of the comment sentence.

Therefore, we propose a word distance assisted dual graph convolutional network (DA2GCN) to characterize both sematic information and syntactic information well. The main contributions are as follows:

- We design a heterogenous dual-GCN to make good use of the word distance for characterizing the correlation of aspect words and context words. The word distance of a sentence is represented by a constructed matrix to establish the distance relationship between them. Then, the word distance information assists the dependency tree and sentiment graph to capture more information and are fed into two GCNs for further sentiment classification.

- We conduct a lot of experiments to verify the advances of our proposed DA2GCN on two self-collected Chinese dataset and five open-source English datasets. The comprehensive results and ablation study demonstrate that the proposed method DA2GCN can achieve higher accuracy and F1 with 1.69–1.81x training speedup over the latest dual-GCN work.

- We make the self-built Chinese datasets of MOOC and Douban, as well as the source code of DA2GCN publicly available on https://github.com/TJSL0715/DA2GCN under open-source licensing.

2. Related work

In this section, we introduce the typical deep learning methods and graph convolutional network assisted methods for aspect-level sentiment analysis.

2.1. Typical deep learning methods for aspect-level sentiment analysis

Most aspect-level sentiment analysis depends on extracting the sentiment information of sentences from the context for identifying the sentiment polarity of a specific aspect. Tang et al. [18] proposed a target-based long short-term memory network to predict the sentiment polarity by modeling the relationship between aspect words and context, and selecting the most relevant parts in the context. Huang et al. [19] proposed an attention-over-attention neural network to learn the representation of aspect words and sentences through bidirectional LSTM (Bi-LSTM), and automatically focuses on important parts of the sentence using the attention mechanism. Zhao et al. [20] proposed a knowledge-enabled language representation model BERT, which can inject the domain-specific emotional knowledge into language representation for aspect-based sentiment analysis. Xiao et al. [21] proposed an enhanced aspect-level sentiment analysis method based on both BERT and multi-attention. Through the interactive attention mechanism of text and aspect words, it captures the correlation between aspect words and the entire text sentence, thereby improving the accuracy of ABSA. An et al. [10] proposed a heterogeneous aspect graph neural network to learn structural and semantic knowledge from inter-sentence relationships to improve the sentiment classification performance. Ma et al. [22] proposed an aspect-context dense connection model to merge the deep semantic information from different aspects and contexts. Yan et al. [23] proposed a sentiment knowledge-based bidirectional encoder representation from transformers. This model utilizes the BERT pre-trained model to encode emotional knowledge vocabulary and contextual words separately, which are subsequently employed for sentiment classification. Tian et al. [24] proposed an attention-based multi-level feature aggregation network, which considers both local and global information by applying attention to convolutional filters. This model uses a multi-level self-attention module to effectively learn the feature information between aspect words and context. However, these typical models cannot consider the words dependency well from a long-distance perspective well so that the sentiment classification is limited.

2.2. GCN enhanced aspect-level sentiment analysis

To address the long-distance correlation issue, GCNs are increasingly applied into the field of aspect-level sentiment analysis. GCN [25] can obtain the information of adjacent nodes, so that it can better capture the local information and global structure of an input text by constructing a graph of words dependency. Zhang et al. [26] extracted the grammatical information of the sentence by constructing a syntactic dependency tree generated by the corresponding grammatical information adjacency matrix and used GCN to learn the obtained grammatical information for excellent classification results. Wang et al. [27] proposed a multi-oriented heterogeneous graph convolution network, which can aggregate multi-faceted information of sentences into a graph, and uses a GCN to update and represent nodes jointly. However, using a single graph to characterize the multiple kinds of information is limited. Wu et al. [28] proposed a phrase dependency relational graph attention network, which aggregates directed dependency edges and phrase information. Phan et al. [29] proposed an

aspect-level sentiment analysis of CNN Over BERT-GCN model. It uses BERT word embedding, Bi-LSTM to extract contextual features, GCN to extract grammatical information, and a CNN model is adopted on the feature vector to classify aspect-level emotions. Huang et al. [9] uses the multiple channels in subgraph structure in a novel scalable GCN for higher accuracy.

Dual graphs are proposed for a more effective aspect-level sentiment analysis. Among them, the combinations of dual GCN models based on syntax and semantics have achieved excellent results in emotion classification. Zhu et al. [11] proposed a global dependency and local dependency mixed GCN, which can make good use of both syntactic dependency structure and contextual information to mine the local structure information of sentences, and constructs a word-document graph through the entire corpus to reveal global dependency information between words. Zhu et al. [12] developed a text sequence graph and refined a dependency graph to uncover valuable structural insights. Two graph convolutional networks are used to effectively extract and enhance the understanding of this structural information. Wei et al. [13] proposed a new method GP-GCN, which aimed at to reduce noise by constructing a simplified global feature structure of text, and use the local structure and global features obtained by orthogonal feature projection for the final aspect-level sentiment classification. Jin et al. [30] proposed a knowledge-enhanced dual-channel graph neural network. The model integrates external emotional knowledge into both semantic and syntactic channels, and then utilizes a dynamic attention mechanism to fuse the diverse information from these channels. Wu et al. [14] fused two parallel graph convolutional networks to simultaneously learn different relationship features between sentences, and added a gate mechanism to GCN to filter the related noise during aggregating information. Although this method considers both the grammatical information and the sentiment information of the word, it does not consider the distance information between the aspect word and other words.

Inspired by these works, we also adopt a dual-GCN framework to capture more heterogeneous feature information. More importantly, unlike the existed work, our proposed DA2GCN exploits the word distance information to rebuild the grammatical dependency tree and syntactic sentiment knowledge and feed them into two GCNs for accurate and fast aspect-level sentiment analysis.

3. Proposed model DA2GCN

In this section, we detail the proposed method DA2GCN for accurate and fast aspect-level sentiment analysis.

3.1. Overall framework

The proposed DA2GCN consists of five parts: (1) the word embedding layer and Bi-LSTM layer, (2) dual graph convolution layers, (3) graph convolution fusion layer with aspect masking, (4) interactive attention layer, and (5) output layer as illustrated in Figure 1. Given a sentence S containing n words, $S = \{\omega_1, \omega_2, \cdots, \omega_{\tau+1}, \cdots, \omega_{\tau+m}, \cdots, \omega_{n-1}, \omega_n\}$, where $\{\omega_{\tau+1}, \cdots, \omega_{\tau+m}\}$ represents the aspect word or aspect phrase in the sentence. The input sentence proceeds to the five parts orderly and the sentiment classification result can be output finally.
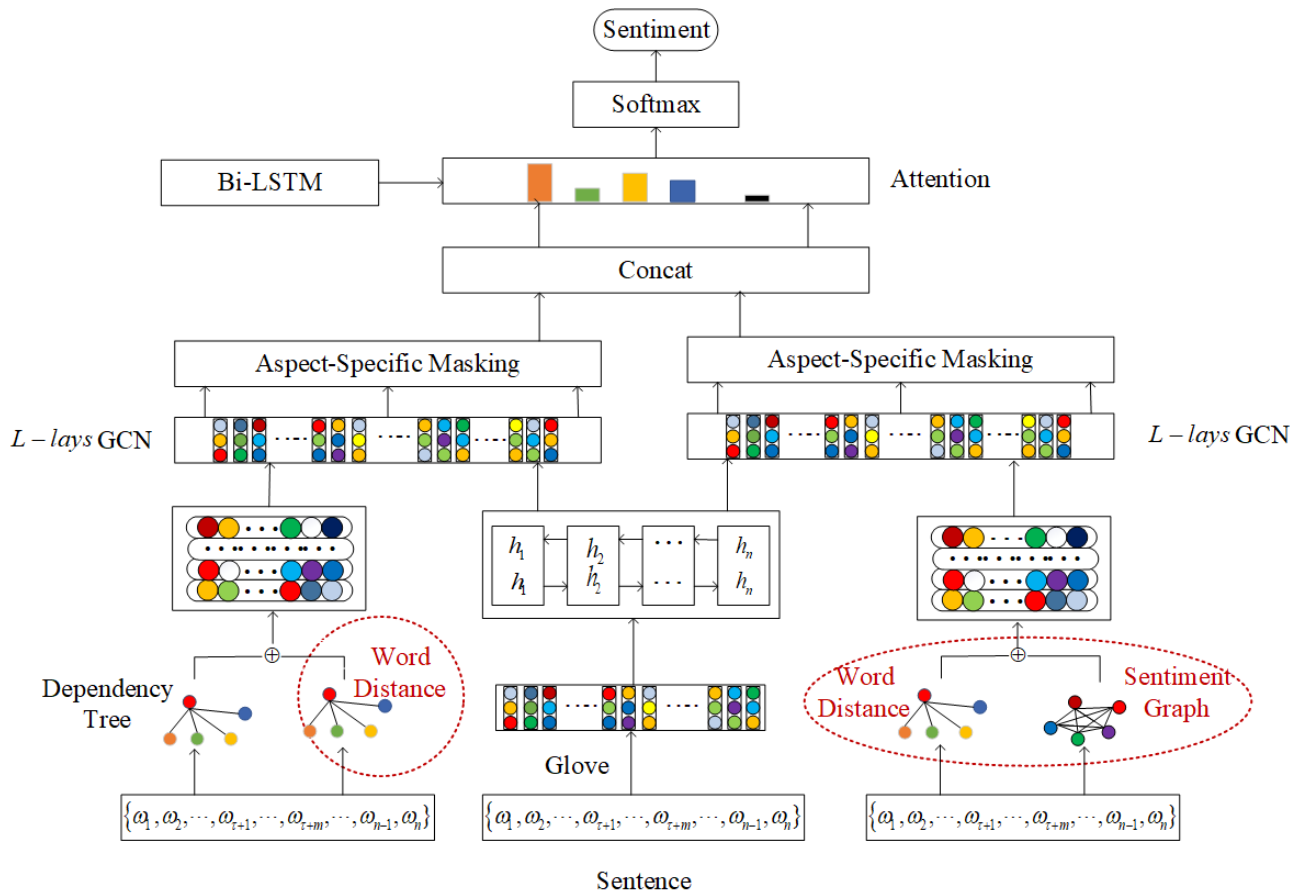
Figure 1. Overall framework of proposed model DA2GCN.

3.2. Word embedding and Bi-LSTM

This paper utilizes a 300-dimensional pre-trained word vector to transform the input sentence S into a word embedding matrix $V \in R^{|n| \times d_e}$, where d_e represents the dimensionality of the word embeddings. As the experiments in this paper involve datasets in both Chinese and English, we have utilized pre-trained word vectors from Chinese Wikipedia [31] and GloVe [32]. Subsequently, the word embedding matrix V is fed into a Bi-LSTM, resulting in the hidden layer state vector $H = \{h_1, h_2, \cdots, h_{\tau+1}, \cdots, h_{\tau+m}, \cdots, h_n\}$ for the sentence. Here, $h_i \in R^{2d_h}$ denotes the hidden state vector for the i word, and d_h represents the output dimension of the unidirectional LSTM.

3.3. Word distance module

Typically, the aspect words and the sentiment words in a sentence are relatively close to each other. The distance between the sentiment words and the aspect words probably contains some important information. Therefore, we incorporate the distance between the aspect words and other words in a sentence into a sentiment classification model to exploring the key information. The distance between the aspect words and other words is shown in Figure 2. For example, "Substantial" is closer to "content" than "speaking", while its distance information is 2 and 4 for the content aspect and the speaking aspect, respectively.

| Aspect word: | Content speaking | While the lesson content is substantial , the teacher's speaking pace is excessively rapid . |

Aspect word:		While the lesson content is substantial , the teacher's speaking pace is excessively rapid .														
Aspect	Content	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12
word:	speaking	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5

Figure 2. An example of word distance representation for different aspects.

Given a sentence S containing n words, $S = \{\omega_1, \omega_2, \cdots, \omega_{\tau+1}, \cdots, \omega_{\tau+m}, \cdots, \omega_{n-1}, \omega_n\}$, where $\{\omega_{\tau+1}, \cdots, \omega_{\tau+m}\}$ represents the aspect word or aspect phrase in the sentence. According to the distance between each word and a specific aspect word, an $n \times n$ diagonal matrix is constructed, where $D_n = \text{diag}(y_1, y_2, \cdots, y_n)$ represents the distance between each word and the aspect word. Due to the excessive length of certain sentences leading to a significant disparity in distances between the aspect words and the other words, and considering that words closer to the aspect words carry greater weight in terms of information, the diagonal matrix can be modified by Eq (1):

$$D = diag\left(1 - \frac{y_i}{2*y_{max}}\right) \tag{1}$$

Among them, $1 \leq i \leq n$, y_{max} refers to the maximum value from y_1 to y_n.

3.4. Dual graph convolutional networks

3.4.1. Distance assisted sentiment knowledge for a grammatical GCN

In order to utilize the sentiment information between the words in a given sentence, we use different polices for Chinese datasets and English datasets. As for a Chinese dataset, we use the Chinese sentiment word extreme value table from the Department of Chinese Language and Literature of Tsinghua University to build a sentiment knowledge module for Chinese dataset. English datasets are handled by the sentiment dictionary SenticNet [33] to build a sentiment knowledge module. When the word is positive, the extreme sentiment value is > 0; when the word is negative, the extreme sentiment value is < 0; when the word is neutral, the extreme sentiment value is = 0.

For any two words, ω_i and ω_j in the sentence S, their corresponding weight is calculated by Eq (2):

$$S_{ij} = |TsNet(\omega_i)| + |TsNet(\omega_j)| \tag{2}$$

Among them, $TsNet(\omega_i)$ represents the weight of the word ω_i in the sentiment dictionary. When the word ω_i is not present in the sentiment polarity lexicon, $TsNet(\omega_i) = 0$. Consequently, the sentiment information matrix $E \in R^{n \times n}$ for the sentence can be obtained.

In addition, the word distance matrix is combined with the sentiment knowledge matrix to rebuild a new sentiment knowledge matrix in Eq (3).

$$D_s = D + E \tag{3}$$

A GCN is further used to extract the distance and the sentiment features, the adjacency matrix D_s of the new distanced assisted sentiment knowledge, along with the context representations H, which is generated by Bi-LSTM. The update procedure for each node in the grammatical GCN is as follows:

$$h_i^l = \sum_{j=1}^n D_s W^l g_j^{l-1} \tag{4}$$

$$h_i^l = \text{ReLU}\left(h_i^l/(d_i + 1) + b^l\right) \tag{5}$$

$$g_i^l = f\left(h_i^l\right) \tag{6}$$

where $g_j^{l-1} \in R^{2d_h}$ is the hidden representation of the j-th node convolved in the $l-1$ layer graph, $h_i^l \in R^{2d_h}$ is the hidden representation of the i-th node convolved in the l-th layer graph, $d_i = \sum D_s$ is the degree of the node, W^l and b^l represent the weight matrix and bias term of the l-th layer graph convolutional network, respectively, and $f(\cdot)$ is a position-aware transformation function.

This grammatical graph convolutional network, which is augmented with sentiment knowledge through layer-wise distances over L layers, can ultimately provide the following representation in Eq (7):

$$h_s^L = \{h_1^L, h_2^L, \cdots, h_n^L\} \tag{7}$$

3.4.2. Distance assisted dependency tree for a syntactic GCN

To leverage the word dependency relationships within a sentence, we employ the SpaCy [34] toolkit to construct a sentence dependency tree. Considering the positions of words in a dependency tree, an adjacency matrix $G \in R^{n \times n}$ of the sentence dependency tree can be derived, where n represents the number of words in the sentence. When $G_{ij} = 1$, it signifies a connection between the word i and the word j in the dependency tree. When $G_{ij} = 0$, it indicates the absence of a relationship between the word i and the word j in the dependency tree. Using the self-loop concept [25] to retain more information of word nodes, a self-loop operation is added to all the word nodes in the dependency tree, when $i = j$, $G_{ij} = 1$.

To take advantage of the relationship between different words in the dependency tree, this paper integrates the word distance information into a sentence's dependency tree. A new distance dependency tree can be constructed by considering both the dependency relationships between sentiment words and aspect words as well as the distance information between them. This approach takes into account both the syntactic dependencies and the spatial relationships between sentiment words and aspect words in Eq (8).

$$D_g = D + G \tag{8}$$

Similar to the above grammatical GCN, a syntactic GCN is used to extract the distance information and the sentiment features, the adjacency matrix D_g of the distance-enhanced dependency tree, along with the context representations H, which is generated by Bi-LSTM. Each node is updated in the graph convolutional network as follows:

$$h_i^l = \sum_{j=1}^n D_g W^l g_j^{l-1} \tag{9}$$

$$h_i^l = \text{ReLU}\left(h_i^l/(d_i + 1) + b^l\right) \tag{10}$$

$$g_i^l = f\left(h_i^l\right) \tag{11}$$

where $g_j^{l-1} \in R^{2d_h}$ is the hidden representation of the j-th node convolved in the $l-1$ layer graph, $h_i^l \in R^{2d_h}$ is the hidden representation of the i-th node convolved in the l-th layer graph, $d_i = \sum D_g$

is the degree of the node, W^l and b^l represent the weight matrix and bias term of the l-th layer graph convolutional network, respectively, and $f(\cdot)$ is a position-aware transformation function.

This syntactic GCN is enhanced by layer-wise distance for the dependency tree over L layers, and gives the following information in Eq (12):

$$h_g^L = \{h_1^L, h_2^L, \cdots, h_n^L\} \tag{12}$$

3.5. Graph convolution fusion layer and Aspect-specific Masking layer

After the input sentence is processed by the two GCNs, we concatenate the outputs of two distinct graph convolutional networks to obtain a more comprehensive feature representation, denoted as h_{sg}.

$$h_{sg} = [h_s; h_g] \tag{13}$$

To emphasize the feature information of the aspect words in a sentence, this proposed method DA2GCN masks the hidden state vectors of non-aspect words while keeping the state of aspect words unchanged in Eq (14):

$$h_t^L = 0 \tag{14}$$

Where the condition should be satisfied $1 \leq t < \tau + 1, \tau + m < t \leq n$

Then, the output of the aspect masking layer is obtained in Eq (15):

$$H_{mask}^L = \{0, \cdots, h_{\tau+1}^L, \cdots, h_{\tau+m}^L, \cdots, 0\} \tag{15}$$

3.6. Aspect-specific masking and Aspect-aware Attention

The idea of the interactive attention layer is to learn the weights for interactions between different elements in the input, which enables a more effective capture of the interdependence among elements. Thus, in the interactive attention layer, our proposed model learns the correlations between aspect words and other words, which facilitates a more comprehensive understanding of the relationships and patterns in the input data.

The calculation of interactive attention weights is as follows:

$$\beta_t = \sum_{t=1}^{n} h_t h_t^L = \sum_{t=\tau+1}^{\tau+m} h_t h_t^L \tag{16}$$

$$\alpha_t = \frac{exp(\beta_t)}{\sum_{i=1}^{n} exp(\beta_i)} \tag{17}$$

$$r = \sum_{t=1}^{n} \alpha_t h_t \tag{18}$$

where, h_t refers to the output of the Bi-LSTM, h_t^L corresponds to the output of the aspect masking layer.

3.7. Sentiment classification and model training

The output γ of the interactive attention layer is the input into a fully connected layer, so that the classification output is obtained through a softmax normalization layer.

$$p = \text{softmax}\left(W_p r + b_p\right) \tag{19}$$

where, W_p is the weight matrix, and b_p is the bias term.

This paper uses the cross-entropy function and L_2 regularization as loss function:

$$\text{Loss} = -\sum_{i=1}^{C} y_i \log_2 p_i + \lambda \parallel \theta \parallel^2 \tag{20}$$

Here, C represents the number of sentence classifications, y_i represents the real sentiment category of the sentence, p_i represents the predicted sentence sentiment category, λ represents the weight of the L_2 regularization term, and θ represents all trainable parameters.

4. Experiential results and analysis

In this section, we detailed the experimental configuration and results analysis to verify the advances of our proposed method DA2GCN.

4.1. Experimental configuration

In order to verify the effectiveness of the proposed DA2GCN in Chinese sentiment analysis, we collect an extensive range of review data from the Chinese University MOOC and Douban Book Review web pages, preprocess and annotate them. The joint dataset of MOOC and Douban is also used to evaluate the generalization of DA2GCN. Each dataset has three sentiment categories: Positive, negative, and neutral. Finally, the Chinese datasets are divided for training and test as Table 1 shows.

Table 1. Chinese Experimental dataset.

Dataset	Positive		Negative		Neutral	
	Train	Test	Train	Test	Train	Test
MOOC	2645	1065	580	283	275	152
Douban	747	334	287	142	566	198
Joint dataset	3392	1399	967	425	841	350

Table 2. English Experimental dataset.

Dataset	Positive		Negative		Neutral	
	Train	Test	Train	Test	Train	Test
Twitter	1561	173	1560	173	3127	346
Lap14	994	341	870	128	464	169
Rest14	2164	728	807	196	637	196
Rest15	912	326	256	182	36	34
Rest16	1240	469	439	117	69	30

Moreover, we conduct the experiments on five public English datasets: one (Twitter) is originally

built by Dong et al. [35] containing twitter posts, restaurants (Rest14), and laptops (Lap14) domain of SemEval 2014 task 4 [36], and restaurants (Rest15, Rest16) domain of SemEval 2015 task 12 [37] and SemEval 2016 task 5 [38]. The datasets configuration is detailed in Table 2.

The typical classification criteria Accuracy and F1 score are used to measure the sentiment analysis results in this paper. The accuracy and F1 are defined as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{21}$$

$$F1 = \frac{2PR}{P+R} \tag{22}$$

TP represents the number of successfully predicted positive samples, FP represents the number of incorrectly predicted negative samples, TN represents the number of successfully predicted negative samples, and FN represents the number of incorrectly predicted positive samples. $Accuracy$ represents the proportion of correctly classified samples among the total samples. P refers to the proportion of samples classified as positive among the true positive samples. R represents the proportion of correctly classified samples among the total samples. The $F1$ score is the weighted harmonic mean of precision and recall.

Table 3. Experimental configuration.

Parameter	Value
Embed_dim	300/768
Batch_size	32
Num_Epoch	100
Learning_rate	0.001/0.00002
GCN_layers	2
Optimizer	adam
l2reg	0.00001

As for non-BERT based models, we use the Chinese Wikipedia 300-dimensional pre-trained word vector the Chinese datasets as the initial word embedding, and the GloVe vector to map each word to 300 dimensions for the English datasets in Table 3. The coefficient λ of L_2 regularization item is 0.00001. The dimension of the hidden state vector is set to 300. The model parameters are optimized and updated using Adam with a learning rate of 0.001. If BERT is used for the word embedding to assist the aspect-level sentiment analysis, the word embedding dimension is configured to 768 for the pre-trained uncased BERT-base model. The corresponding learning rate is set to 0.00002.

To evaluate the effectiveness of our proposed method, we compare it with a series of the state-of-the-art (SOTA)methods. Due to the limited number of studies on Chinese sentiment analysis, we select the following five SOTA methods for comparison:

- TD-LSTM [18]: A bidirectional LSTM model is proposed to automatically extract target information and perform sentiment analysis based on the correlation between aspect words and contextual words.
- ASCNN [26]: Bi-LSTM is combined with an undirected dependency tree for the sentence and CNN is used to extract information from syntactic dependency relationships.

- ASTCN [26]: GCN takes the place of CNN in the above ASCNN to extract syntactic information for a more accurate sentiment analysis.
- ASGCN [26]: Unlike ASTCN, GCN further exploits contextual features and syntactic information between words, which is combined with attention mechanisms for sentiment analysis.
- DSSGCN [15]: A dual-channel semantic learning graph convolutional network is proposed to extract semantic information obtained through cosine similarity and structural information acquired through co-occurrence words for sentiment analysis.

In order to further demonstrate the effectiveness of the proposed DA2GCN model, the following six GCN based SOTA methods are used on five English datasets:

- GL-GCN [11]: A local graph based on the syntactic information and sentence order, and a word-document global graph are used to construct two GCNs for the aspect-level sentiment analysis.
- SEDC-GCN [12]: The two graphs for text sequence enhanced dependency are constructed to characterize the more structural information, while a dual-channel graph encoder is designed to model them jointly.
- GP-GCN [13]: A dual-graph convolutional network is proposed using the global feature structure of the text, and the local dependency structure of the sentence for the aspect-level sentiment analysis.
- PFGGCN [14]: It integrates two parallel GCNs to learn the distinct relational features between sentences, and a gating mechanism to filter out the noise.
- TD-BERT [39]: A BERT model is used for assisting the aspect-level sentiment analysis.
- SK-GCN [40]: A new graph convolutional network model based on grammar and knowledge to make good use of both syntactic dependency trees and common-sense knowledge through dual GCNs for the aspect-level sentiment classification.

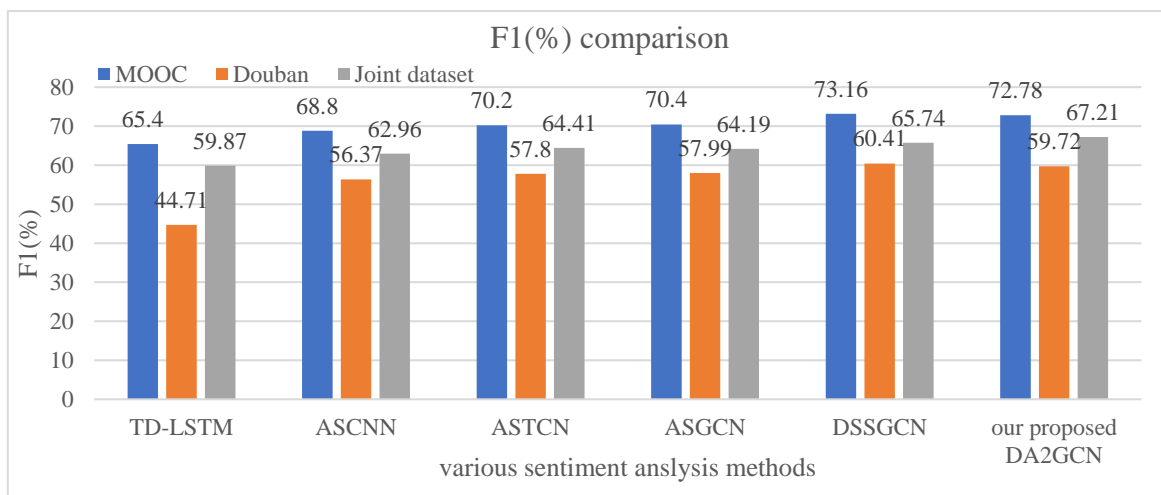## 4.2. Accuracy and F1 comparison with SOTA methods

Chinese datasets results and analysis. It is observed that the proposed DA2GCN in this paper can achieve the best accuracy 78.54% and F1 67.21%, and outperforms the SOTA methods on the joint dataset in Figure3. This primarily results from the unique characteristics of word distance consideration and dual GCNs in DA2GCN. It is interesting that DA2GCN achieves the highest accuracy, 1.41% and 1.11% higher accuracy, but a slightly lower F1 than DSSGCN for the two Chinese datasets of MOOC and Douban, respectively. Because the imbalance of the training datasets of three sentiment categories (positive, negative, and neutral) probably causes a higher accuracy and a bit lower F1 in DA2GCN. On the MOOC dataset, the DA2GCN performs a 2.04% higher accuracy and higher 2.38% F1 than ASGCN. Since the Douban data set mostly contains long and difficult sentences, the accuracy and f1 value of the model are both lower than the MOOC dataset. In comparison to the ASGCN model, our model exhibits an improvement of 2.22% in accuracy and 1.73% in F1 value. It can be shown that considering both sentiment knowledge information and distance information between words in our proposed DA2GCN is beneficial to improve the performance aspect-level sentiment analysis.

English dataset results and analysis. When GloVe is used for word embeddings on the English dataset, our proposed DA2GCN perform better than TD-LSTM, ASGCN, GL-GCN, and GP-GCN, but a bit worse than SEDC-GCN, FPGGCN, and DSSGCN as shown in Table 4. The primary reason

is that the significant differences in the tokenization format between Chinese and English sentences in GloVe, while our proposed method is mainly customized for Chinese datasets. More importantly, when the popular BERT is used to enhance the word embeddings, our proposed model can achieve outstanding classification results on the Rest14, Rest15, and Rest16 datasets over the SOTA methods of SK-GCN, TD-BERT and GP-GCN. In a summary, our proposed model not only effectively enhances the sentiment classification performance on English datasets but also achieves the superior performance on Chinese datasets.
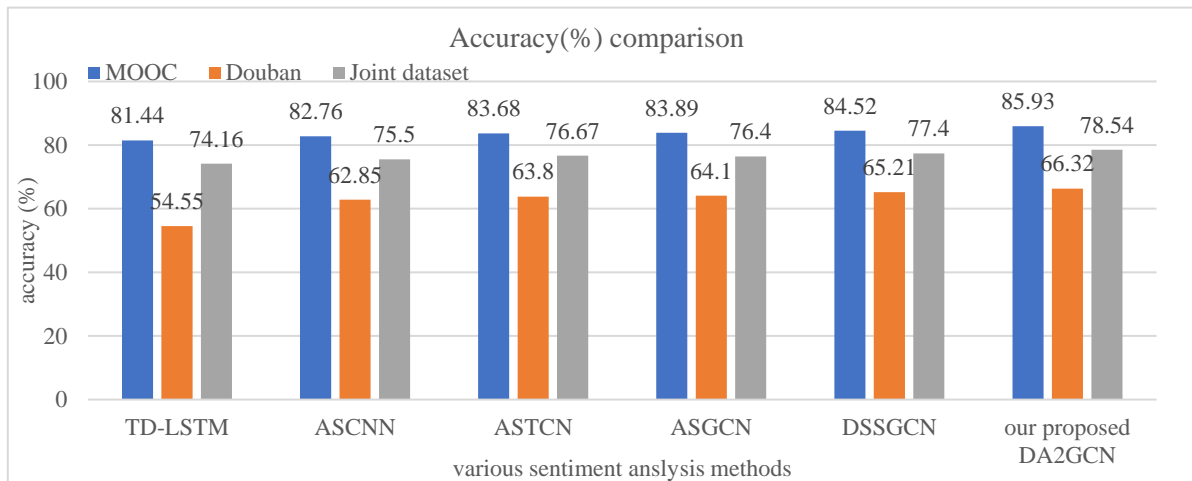
(a)

(b)

Figure 3. Accuracy and F1 results comparison between different aspect-level sentiment analysis methods on Chinese datasets. (a) Accuracy comparison; (b) F1comparison.

Table 4. Accuracy and F1 results comparison between different aspect-level sentiment analysis methods on five English datasets.

Embedding	Models	Twitter(%)		LAP14(%)		REST14(%)		REST15(%)		REST16(%)	
		Acc.	F1.	Acc.	F1.	Acc.	F1.	Acc.	F1.	Acc.	F1.
Glove	TD-LSTM	68.64	66.60	68.88	63.93	78.60	67.02	78.48	62.84	83.77	61.71
	ASGCN	72.15	70.40	75.55	71.05	80.77	72.02	79.89	61.89	88.99	67.48
	GL-GCN	73.26	71.26	76.91	72.76	82.11	73.46	80.81	64.99	88.47	69.64
	GP-GCN	71.67	69.45	73.90	68.67	80.89	70.90	79.89	61.78	83.90	64.67
	SEDC-GCN	74.42	73.37	77.74	74.68	83.30	77.51	81.73	66.23	90.75	73.84
	PFGGCN	-	-	78.06	74.52	83.78	76.55	82.15	66.73	90.92	75.26
	DSSGCN	**75.25**	**73.71**	**78.49**	**74.63**	**84.36**	**77.35**	**82.62**	**66.39**	**91.38**	**75.43**
	DA2GCN	72.83	71.17	76.96	73.68	83.30	76.52	80.26	64.57	88.47	70.75
BERT	SK-GCN	75.00	73.01	79.00	75.57	83.48	75.19	83.20	66.78	87.19	72.02
	TD-BERT	**76.69**	**74.28**	78.87	74.38	85.1	78.35	-	-	-	-
	GP-GCN	75.90	73.90	**79.90**	**75.89**	83.89	75.09	83.90	66.89	87.78	**72.89**
	DA2GCN	74.28	72.87	78.2	74.66	**85.8**	**79.66**	83.21	**68.4**	90.26	72.32

4.3. Training time and model size comparison with SOTA methods

The model size and training time are also significant for evaluating the training speed. The comparative results of model size, training time and speedup are given in Figure 4(a)–(c), respectively.

The model size is measured by the number of parameters and our method has smaller parameters than the effective dual-GCN method DSSGCN. Therefore, the training speedup of our DA2GCN is up to 1.81x over DSSGCN. It is noted that the complex DSSGCN brings the higher performance on English datasets at the expense of a large model and long training time. Our DA2GCN can achieve the good tradeoff between accuracy and speed.

Compared with the similar-size models ASCNN, ASTCN and ASGCN, the training time of our DA2GCN is significantly reduced because of its fast convergence. In a summary, the proposed DA2GCN can be used for the fast and accurate aspect-level sentiment analysis using fewer parameters and training time.

(a)

(b)

©

Figure 4. Training time and model size comparison between different aspect-level sentiment analysis methods on Chinese datasets. (a) Model size comparison; (b) Training time comparison; and (c) Speedup over SOTA.
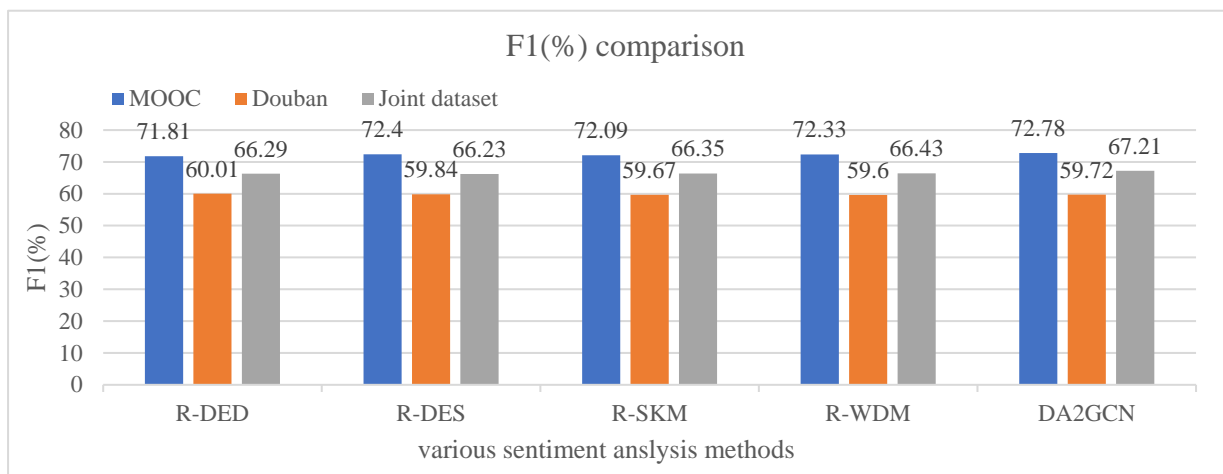
4.4. Ablation study

To further examine the impacts of each component of DA2GCN on the sentiment classification performance, we conduct the ablation study as follows:

- R-DED: When the distance-enhanced dependency tree module is removed, and the model only considers the feature information of distance-enhanced sentiment knowledge. It employs a single graph convolutional network to extract features of distance-enhanced sentiment knowledge.
- R-DES: When the distance-enhanced sentiment knowledge module is removed, the model only considers the feature information of the distance-enhanced dependency tree. A single graph convolutional network is employed to extract features of the distance-enhanced dependency tree.
- R-SKM: When the sentiment knowledge information is removed, the model only considers the feature information of the dependency tree and word distance. It uses the dual graph convolutional networks, features of the word distance matrix, and the distance-enhanced dependency tree matrix are extracted separately.
- R-WDM: When the word distance information is removed, the model considers only the feature information of the dependency tree and sentiment knowledge. It uses the dual graph convolutional networks, features of the sentiment knowledge matrix, and the dependency tree adjacency matrix are extracted separately.
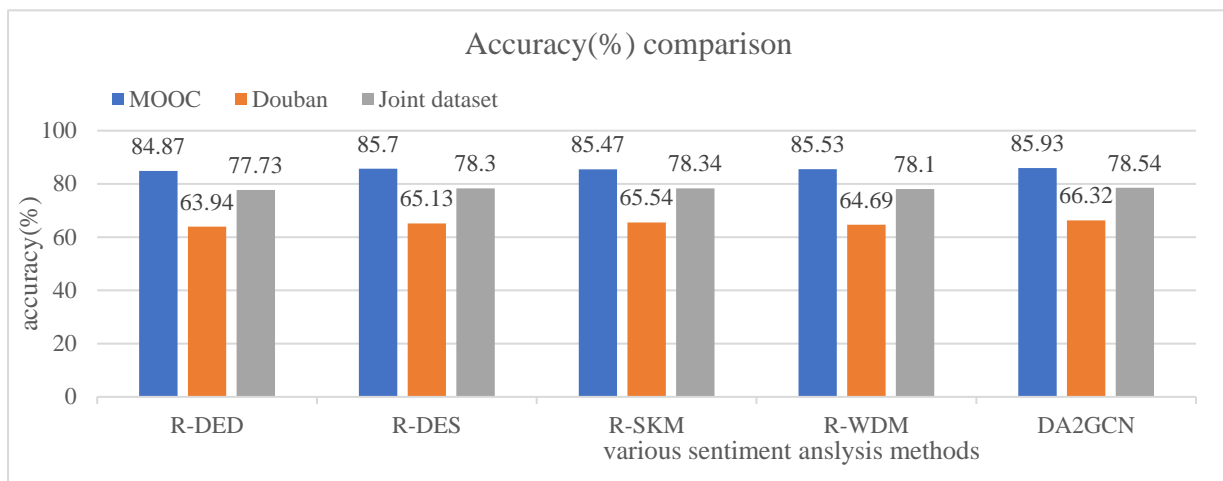
(a)

(b)

Figure 5. Accuracy and F1 results comparison of ablation study. (a) Accuracy comparison; (b) F1 comparison.

As is listed in Figure 5, the ablation study demonstrates that the complete DA2GCN model surpasses any variant lacking a single module in terms of classification performance. This confirms the importance of incorporating word distance and sentiment knowledge information into the model. When the word distance module is removed in R-DED, there is a noticeable decrease in accuracy and F1 score compared to the complete DA2GCN model, which highlights the significance of the word distance information in the accurate aspect-level sentiment analysis. In the case of the Douban dataset, the methods of R-DED and R-DES has little better F1 than our proposed method. Because the comments from Douban are mostly long sentences, the sentiment classification depends less distance information. Therefore, our DA2GCN achieves good performance on MOOC dataset and the joint dataset well.

In addition, we conducted a statistical analysis of the parameter count and training time for the models used in the ablation experiments on the MOOC dataset and Douban dataset, as detailed in Table 5. It is noteworthy that the table indicates an equivalent parameter count for each model. This is attributed to our exclusive modification of the values within the input parameter matrices, without altering the matrices' dimensions. Consequently, the total number of parameters in the models remains constant throughout this process. However, the training time varies among the models. This discrepancy arises because, despite the consistent parameter count, altering the initial values of the weights may influence the training and convergence processes, leading to differences in training times.

Table 5. Parameters and training time of the ablation experimental model.

| Model	MOOC		Douban	
	Time(s)	Parameters(M)	Time(s)	Parameters(M)
R-DED	123	7.2	266	5.6
R-DES	153	7.2	244	5.6
R-SKM	154	7.2	339	5.6
R-WDM	164	7.2	414	5.6
DA2GCN	146	7.2	365	5.6

4.5. The number of GCN layers impacts on DA2GCN

The number of GCN layers decides how long the neighbor's information can be collected and is significant for GCN. More importantly, we use the effective dual GCNs. Therefore, the number of GCN layers in the two GCNs in our proposed DA2GCN is discussed. The homogeneous and heterogenous GCNs are both considered, where the same number or different values are set to search the optimal configuration.

The experimental results are illustrated in Figure 6, where the x-axis represents the number of layers in the dual graph convolution networks and the y-axis depicts the corresponding accuracy. The first digit signifies the layers for the word distance assisted grammatical GCN, while the second digit represents the layers for the word distance assisted syntactic GCN. It is noted that when the number of GCN layers are both set to 2, the highest accuracy is obtained. If there is only one layer in the two GCNs, the model has the lowest accuracy due to a smaller receptive field, failing to capture sufficient sentence feature information. On the other hand, as the number of layers increases for the two GCNs, the model becomes more complex and also introduces more noise to decrease the accuracy. Furthermore, an excessive number of network layers may lead to the overfitting so that the model's

generalization is diminishing. Therefore, selecting an appropriate number of GCN layers is crucial for extracting effective feature information for higher accuracy.
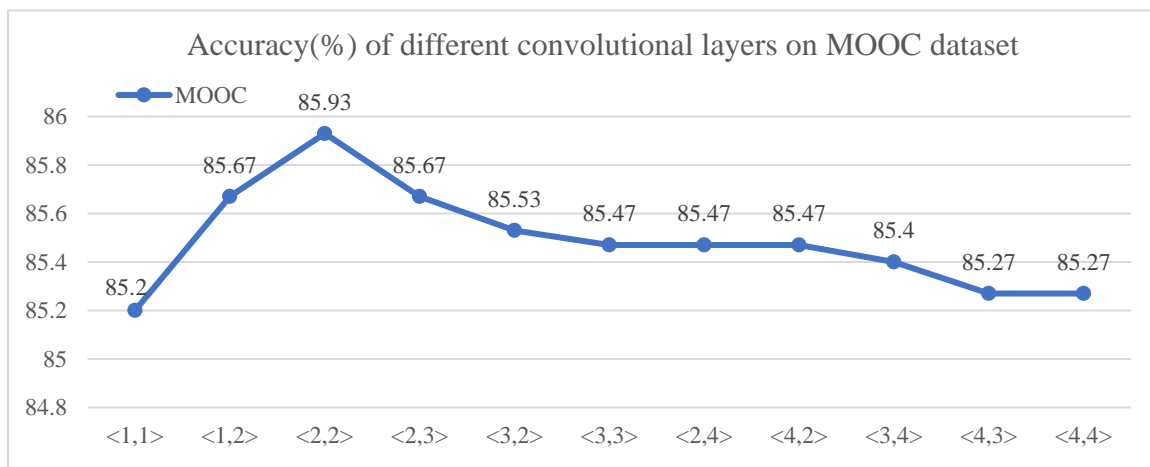
Figure 6. The impact of varying dual-GCN layers.

4.6. Visualizations

In order to qualitatively demonstrate the improved performance of the proposed DA2GCN model in predicting aspect-based sentiment polarity, we visualize the attention weight through representative examples in Figure 7. It is observed that, in comparison to the R-WDM model (the model without the word distance module), the complete DA2GCN model pays more attention to crucial sentiment words and successfully extracts emotion features corresponding to specific aspects. This indicates that strengthening the word distance relationship based on specific aspects allows for a more accurate characterization of sentiment features, thereby enhancing the performance in the aspect-based sentiment analysis tasks.
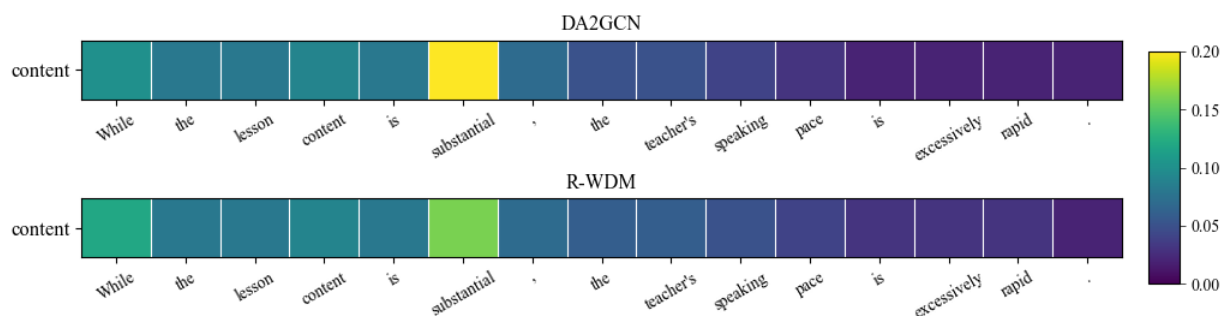
Figure 7. The attention visualizations of different models.

4.7. Discussion

The above experiments demonstrate that the proposed DA2GCN make good use of the dual GCN framework to consider the word distance information completely for rebuilding the grammatical

knowledge and syntactic dependency tree and achieving higher performance of aspect-level sentiment analysis. More importantly, this approach not only works well on each single dataset from two Chinese datasets and five English datasets, but also has scalability and performance even larger datasets (joint MOOC and Doban up to more than 7000 comments). Furthermore, it is noted that DA2GCN has shorter training time than other dual-GCN methods because of the smaller number of parameters. Therefore, the lightweight DA2GCN can be used in the real-time applications. In summary, our proposed DA2GCN performs good scalability in the real-time applications.

Besides the aspect-level sentiment analysis, DA2GCN inspire other directions of natural language processing (NLP) such as spanning question answering, relation extraction, and machine reading comprehension to exploits diverse information in a dual-GCN framework for high performance text processing. Thus, the proposed method has good generalizability to other NLP applications.

However, it is noted that due to the significant differences in the tokenization formats between Chinese datasets and English datasets, our proposed method uses the Chinese customized tokenization format to extracts the distance information for all the datasets. Therefore, we will focus on the customized representation of English word distance for more accurate and fast aspect-level sentiment analysis in future work.

5. Conclusions

In this paper, a novel dual-GCN method DA2GCN is proposed to make good use of word distance information for rebuilding the grammatical knowledge and syntactic dependency tree so that the aspect-level sentiment classification can be improved well. The comprehensive results on two self-built Chinese datasets and five open-source English datasets demonstrate our DA2GCN can achieve higher accuracy and F1 over the SOTA methods. Moreover, DA2GCN has fewer parameters and consumes less training time than the latest dual-GCN methods. It is noted that due to the significant differences in the tokenization formats between Chinese datasets and English datasets, our proposed method uses the Chinese customized tokenization format to extracts the distance information for all the datasets. Consequently, DA2GCN can represent the distance information for Chinese words more effectively and performs better than English. Therefore, we will focus on the customized representation of English word distance for more accurate and fast aspect-level sentiment analysis in future work.

To further improve the performance of aspect-level sentiment classification, multimodal (text, video and audio) is increasingly attractive, unlike the pure textual data, the correlations between multimodal data are also more intricate. The complex hypergraph neural networks [41] can comprehensively depict complex higher-order data correlations using hyperedge convolution operations. In the future work, we will develop more effective hypergraph neural networks to characterize more heterogeneous information from multimodal and achieve higher performance of sentiment classification.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Conflict of interest

The authors declare that there are no conflict of interest.

References

1. H. T. Phan, N. T. Nguyen, D. Hwang, Aspect-level sentiment analysis: A survey of graph convolutional network methods, *Inf. Fusion*, **91** (2023), 149–172. https://doi.org/10.1016/j.inffus.2022.10.004

2. R. Das, T. D. Singh, Multimodal sentiment analysis: A survey of methods, trends and challenges, *ACM Comput. Sur.*, (2023). https://doi.org/10.1145/3586075

3. J. Cui, Z. Wang, S. Ho, E. Cambria, Survey on sentiment analysis: Evolution of research methods and topics, *Artif. Intell. Rev.*, **56** (2023), 8469–8510. https://doi.org/10.1007/s10462-022-10386-z

4. Y. Wang, M. Huang, X. Zhu, L. Zhao, Attention-based LSTM for aspect-level sentiment classification, in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, (2016), 606–615. https://doi.org/10.18653/v1/D16-1058

5. D. Ma, S. Li, X. Zhang, H. Wang, Interactive attention networks for aspect-level sentiment classification, preprint, arXiv: 1709.00893. https://doi.org/10.48550/arXiv.1709.00893

6. K. Sun, R. Zhang, S. Mensah, Y. Mao, X. Liu, Aspect-level sentiment analysis via convolution over dependency tree, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, (2019), 5679–5688. https://doi.org/10.18653/v1/D19-1569

7. B. Huang, K. Carley, Syntax-aware aspect level sentiment classification with graph attention networks, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, (2019), 5469–5477. https://doi.org/10.18653/v1/D19-1549

8. K. Wang, W. Shen, Y. Yang, X. Quan, R. Wang, Relational graph attention network for aspect-based sentiment analysis, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (2020), 3229–3238. https://doi.org/10.18653/v1/2020.acl-main.295

9. Z. Huang, W. Zhou, K. Li, Z. Jia, SGCN: A scalable graph convolutional network with graph-shaped kernels and multi-channels, *Knowl. Based Syst.*, **279** (2023), 110923. https://doi.org/10.1016/j.knosys.2023.110923

10. W. An, F. Tian, P. Chen, Q. Zheng, Aspect-based sentiment analysis with heterogeneous graph neural network, *IEEE Trans. Comput. Soc. Syst.*, **10** (2022), 403–412. https://doi.org/10.1109/TCSS.2022.3148866

11. X. Zhu, L. Zhu, J. Guo, S. Liang, S. Dietze, GL-GCN: Global and local dependency guided graph convolutional networks for aspect-based sentiment classification, *Expert Syst. Appl.*, **186** (2021), 115712. https://doi.org/10.1016/j.eswa.2021.115712

12. L. Zhu, X. Zhu, J. Guo, S. Dietze, Exploring rich structure information for aspect-based sentiment classification, *J. Intell. Inf. Syst.*, **60** (2023), 97–117. https://doi.org/10.1007/s10844-022-00729-1

13. S. Wei, G. Zhu, Z. Sun, X. Li, T. Weng, GP-GCN: Global features of orthogonal projection and local dependency fused graph convolutional networks for aspect-level sentiment classification, *Connect. Sci.*, **34** (2022),1785–1806. https://doi.org/10.1080/09540091.2022.2080183

14. Y. Wu, G. Deng, A parallel fusion graph convolutional network for aspect-level sentiment analysis, *Big Data Res.*, **32** (2023), 100378. https://doi.org/10.1016/j.bdr.2023.100378

15. A. Dai, X. Hu, J. Nie, J. Chen, Learning from word semantics to sentence syntax by graph convolutional networks for aspect-based sentiment analysis, *Int. J. Data Sci. Anal.*, **14** (2022), 17–26. https://doi.org/10.1007/s41060-022-00315-2

16. B. Liang, Q. Liu, J. Xu, Q. Zhou, P. Zhang, Aspect-based sentiment analysis based on multi-attention CNN, *J. Comput. Res. Develop.*, **54** (2017), 1724–1735.

17. P. Chen, Z. Sun, L. Bing, W. Yang, Recurrent attention network on memory for aspect sentiment analysis, in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, (2017), 452–461. https://doi.org/10.18653/v1/D17-1047

18. D. Tang, B. Qin, X. Feng, T. Liu, Effective LSTMs for target-dependent sentiment classification, in *Proceedings of COLING 2016*, (2016), 3298–3307. https://doi.org/10.48550/arXiv.1512.01100

19. B. Huang, Y. Ou, K. M. Carley, Aspect level sentiment classification with attention-over-attention neural networks, in *Social, Cultural, and Behavioral Modeling: 11th International Conference*, (2018), 197–206. https://doi.org/10.1007/978-3-319-93372-6_22

20. A. Zhao, Y. Yu, Knowledge-enabled BERT for aspect-based sentiment analysis, *Knowl. Based Syst.*, **227** (2021), 107220. https://doi.org/10.1016/j.knosys.2021.107220

21. J. Xiao, X. Luo, Aspect-level sentiment analysis based on BERT fusion multi-attention, in *2022 14th International Conference on Intelligent Human-Machine Systems and Cybernetics*, (2022), 32–35. https://doi.org/10.1109/IHMSC55436.2022.00016

22. G. Ma, X. Guo, Dense concatenation memory network for aspect level sentiment analysis, *IEEE Access*, **11** (2023), 20486–20493. https://doi.org/10.1109/ACCESS.2023.3248639

23. H. Yan, B. Yi, H. Li, D. Wu, Sentiment knowledge-induced neural network for aspect-level sentiment analysis, *Neural Comput. Appl.*, **34** (2022), 22275–22286. https://doi.org/10.1007/s00521-022-07698-0

24. D. Tian, J. Shi, J. Feng, A self-attention-based multi-level fusion network for aspect category sentiment analysis, *Cogn. Comput.*, **15** (2023), 1372–1390. https://doi.org/10.1007/s12559-023-10160-5

25. T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, preprint, arXiv: 1609.02907. https://doi.org/10.48550/arXiv.1609.02907

26. C. Zhang, Q. Li, D. Song, Aspect-based sentiment classification with aspect-specific graph convolutional networks, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, (2019), 4568–4578. https://doi.org/10.18653/v1/D19-1464

27. S. Wang, G. Zhang, J. Cao, Aspect-based sentiment analysis with multi-aspects heterogeneous graph convolutional networks, in *Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering*, (2021), 915–920. https://doi.org/10.1145/3501409.3501574

28. H. Wu, Z. Zhang, S. Shi, Q. Wu, H. Song, Phrase dependency relational graph attention network for Aspect-based Sentiment Analysis, *Knowl. Based Syst.*, **236** (2022), 107736. https://doi.org/10.1016/j.knosys.2021.107736

29. H. T. Phan, N. T. Nguyen and D. Hwang, Aspect-level sentiment analysis using CNN over BERT-GCN, *IEEE Access*, **10** (2022), 110402–110409. https://doi.org/10.1109/ACCESS.2022.3214233

30. H. Jin, Q. Zhang, X. Liang, Y. Zhou, W. Li, Dual channel graph neural network enhanced by external affective knowledge for aspect level sentiment analysis, in *ICONIP 2023*, (2023), 257–274. https://doi.org/10.1007/978-981-99-8082-6_20

31. S. Li, Z. Zhao, R. Hu, W. Li, T. Liu, X. Du, Analogical reasoning on Chinese morphological and semantic relations, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, (2018), 138–143. https://doi.org/10.18653/v1/P18-2023

32. J. Pennington, R. Socher, C. Manning, Glove: Global vectors for word representation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, (2014), 1532–1543. https://doi.org/10.3115/v1/D14-1162

33. E. Cambria, Q. Liu, S. Decherchi, F. Xing, K. Kwok, SenticNet 7: A commonsense-based neurosymbolic AI framework for explainable sentiment analysis, in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, (2022), 3829–3839. https://aclanthology.org/2022.lrec-1.408

34. Explosion AI. (2022). spaCy 3.2.0. https://spacy.io/

35. L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, K. Xu, Adaptive recursive neural network for target-dependent twitter sentiment classification, in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, (2014), 49–54. https://doi.org/10.3115/v1/P14-2009

36. M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, Semeval-2014 task 4: Aspect based sentiment analysis, in *Proceedings of the 8th International Workshop on Semantic Evalution*, (2014), 27–35. https://doi.org/10.3115/v1/S14-2004

37. M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, I. Androutsopoulos, Semeval-2015 task 12: Aspect based sentiment analysis, in *Proceedings of the 9th International Workshop on Semantic Evaluation*, (2015), 486–495. https://doi.org/10.18653/v1/S15-2082

38. M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, et al., Semeval-2016 task 5: Aspect based sentiment analysis, in *Proceedings of the 10th International Workshop on Semantic Evaluation*, (2016), 19–30. https://doi.org/10.18653/v1/S16-1002

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

40. J. Zhou, J. X. Huang, Q. V. Hu, L. He, SK-GCN: Modeling Syntax and Knowledge via graph convolutional network for aspect-level sentiment classification, *Knowl. Based Syst.*, **205** (2020), 106292. https://doi.org/10.1016/j.knosys.2020.106292

41. K. Li, Z. Huang, Z. Jia, RAHG: A role-aware Hypergraph neural network for node classification in graphs, *IEEE Trans. Network Sci. Eng.*, **10** (2023), 2098–2108. https://doi.org/10.1109/TNSE.2023.3243058