Expert Systems With Applications 186 (2021) 115712

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

GL-GCN: Global and Local Dependency Guided Graph Convolutional Networks for aspect-based sentiment classification

Xiaofei Zhu [a,*], Ling Zhu [a], Jiafeng Guo [b], Shangsong Liang [c], Stefan Dietze [d,e]

[a] College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China
[b] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[c] School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510275, China
[d] Knowledge Technologies for the Social Sciences, Leibniz Institute for the Social Sciences, Cologne 50667, Germany
[e] Institute of Computer Science, Heinrich-Heine-University Düsseldorf, Düsseldorf 40225, Germany

ARTICLE INFO

Keywords:
Graph convolutional networks
Aspect-based sentiment classification
Attention mechanism
Sentiment analysis

ABSTRACT

Aspect-based sentiment classification, which aims at identifying the sentiment polarity of a sentence towards the specified aspect, has become a crucial task for sentiment analysis. Existing methods have proposed effective models and achieved satisfactory results, but they mainly focus on exploiting local structure information of a given sentence, such as locality, sequentiality or syntactical dependency constraints within the sentence. Recently, some research works, which utilizes global dependency information, has attracted increasing interest and significantly boosts the performance of text classification. In this paper, we simultaneously introduce both global structure information and local structure information into the task of aspect-based sentiment classification, and propose a novel aspect-based sentiment classification approach, i.e., Global and Local Dependency Guided Graph Convolutional Networks (GL-GCN). In particular, we exploit the syntactic dependency structure as well as sentence sequential information (e.g., the output of BiLSTM) to mine the local structure information of a sentence. On the other hand, we construct a word-document graph using the entire corpus to reveal the global dependency information between words. In addition, an attention mechanism is leveraged to effectively fuse both global and local dependency structure signals. Extensive experiments are conducted on five benchmark datasets in terms of both Accuracy and F1-Score, and the results illustrate that our proposed framework outperforms state-of-the-art methods for aspect-based sentiment classification. The model is implemented using PyTorch and is trained on GPU GeForce GTX 2080 Ti.

1. Introduction

With the rapid growth of online review sites such as Amazon, Yelp and IMDB, aspect-based sentiment classification has become a crucial topic in recent years. The main challenge of aspect-based sentiment classification is to identify the corresponding sentiment polarity (e.g., positive, neutral, or negative) towards a specified target when multiple targets are available in a sentence. Fig. 1 shows an example of aspect-based sentiment classification with multiple sentiment polarities, where the sentiment polarity of food is positive, while for the atmosphere it is negative.

In recent years, a number of efforts have been made towards effectively modeling semantic relatedness between context words and the aspects within a sentence. Liu and Zhang (2017) and Wang, Huang, Zhao, and Zhu (2016) propose to utilize attention mechanisms Bahdanau, Cho, and Bengio (2015) together with Recurrent Neural Networks (RNN) (Bengio, Ducharme, Vincent, & Janvin, 2003; Hochreiter & Schmidhuber, 1997) for aspect-based sentiment classification. It assigns a positive weight for each context word, which reflects the importance of the word for determining the sentiment polarity of the specified target.

Fan et al. (2018) find that the sentiment of an aspect is usually determined by key phrases rather than individual words. Based on this observation, Li, Bing, Lam, and Shi (2018) and Xue and Li (2018) propose to employ Convolutional Neural Networks (CNNs) (Lecun, Bottou, Bengio, & Haffner, 1998) to capture multi-word phrases via the convolution operations over word sequences. As CNN and RNN prioritize locality and sequentiality (Battaglia, Hamrick, Bapst, et al., 2018), these models can effectively capture semantic and syntactic information

* Corresponding author.
E-mail addresses: zxf@cqut.edu.cn (X. Zhu), zhuling@2019.cqut.edu.cn (L. Zhu), guojiafeng@ict.ac.cn (J. Guo), liangshangsong@gmail.com (S. Liang), stefan.dietze@gesis.org (S. Dietze).

https://doi.org/10.1016/j.eswa.2021.115712
Received 7 December 2020; Received in revised form 2 June 2021; Accepted 1 August 2021
Available online 11 August 2021

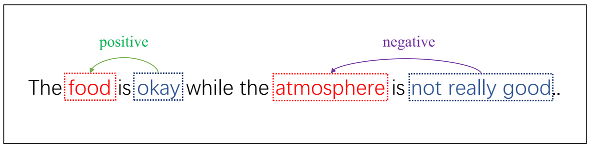

Fig. 1. An example of aspect-based sentiment classification with multiple sentiment polarities.

in local consecutive word sequences. However, they lack a mechanism to account for long-range word dependencies, and may result in identifying irrelevant clues for determining aspect sentiment. It is worth noting that the long-range word dependencies mentioned here is a comparative statement. Although CNN can model word dependencies within a sentence, they only capture local phrase-level dependencies detected by filters (e.g., a sliding-window). In recent years, dependency tree has received considerable attention since it can capture dependency relationship between two distant words (e.g., there is a syntactic relationship between them). Many research works propose to leverage the dependency tree to address the issue. For example, Zhang, Li, and Song (2019) propose to exploit syntactical dependency structures within a sentence. They build a Graph Convolutional Network (GCN) over the dependency tree of a sentence and exploit syntactical information to bridge the long-range word dependency.

Whereas the aforementioned works are promising and achieve satisfactory results, their limitation is that they mainly rely on exploiting local structure dependency information, such as locality, sequentiality or syntactical dependency constraints within a sentence, while global structure dependency information is largely ignored. To be specific, existing methods lack an explicit modeling of the global dependency signal, which is latent in the entire corpus to reveal the global relationships between words.

Recent research (Peng et al., 2018; Yao, Mao, & Luo, 2019) has shown that exploiting global structure dependency information can often significantly improve the performance of text classification. Peng et al. (2018) convert a document into a word co-occurrence graph, and then leverage graph convolution operations to convolve the graph. Yao et al. (2019) propose to build a text graph for an entire corpus, where nodes are words and documents. The edge between two word nodes relies on word co-occurrence, and the edge between a word node and a document node is using TFIDF. GCN is then used to capture high order neighborhood information.

In this paper, we propose Global and Local Dependency Guided Graph Convolutional Networks (GL-GCN), for aspect-based sentiment classification. In particular, we leverage two kinds of GCNs to learn different dependency structure information: (1) One GCN is leveraged to capture global dependency structure information via exploring the entire corpus; (2) Another GCN is utilized to model local dependency structure information given in each sentence. The framework of the proposed model is shown in Fig. 2.

We conducted extensive experiments on five datasets, i.e., TWITTER, LAPTOP, REST14, REST15, and REST16. All datasets are publicly available and have been widely used in the task of aspect-based sentiment classification. Experimental results demonstrate that the proposed GL-GCN approach can effectively model both global and local dependency structure information, and consistently outperforms the state-of-the-art baseline methods with a large margin.

In summary, this work makes the following main contributions:

- We propose a novel aspect-based sentiment classification approach exploiting both global and local dependency structure signals to better address the issue of long-range of multi-word dependency.
- We propose a novel architecture, which consists of two kinds of GCNs, to effectively encode both global and local structure signals. Moreover, a gating mechanism is leveraged to adaptively fuse these two kinds information.
- We conduct extensive experiments on five datasets demonstrating that GL-GCN is effective in improving the embedding quality by involving global structure information, thereby achieving the state-of-the-art performance

The rest of this paper is organized as follows. Section 2 presents the related work. Section 3 introduces the proposed approach GL-GCN to aspect-based sentiment classification. Section 4 presents extensive experiments to evaluate the effectiveness of our approach, and discusses the effectiveness of involving global dependency structure information. Section 5 concludes the paper.

2. Related work

In this section, we briefly review the related work in following two categories: Graph Convolutional Networks and Aspect-based Sentiment Classification.

2.1. Graph convolutional networks (GCNs)

Graph Convolutional Networks (GCNs), also known as Graph Neural Networks, have recently achieved promising advancement in various applications, including link prediction (Bordes, Usunier, García-Durán, Weston, & Yakhnenko, 2013; Lin, Liu, Sun, Liu, & Zhu, 2015), recommender systems (Wang, Zhao, Xie, Li, & Guo, 2019), node classification (Gong & Ai, 2019; Kipf & Welling, 2017), text classification (Liu, You, Zhang, Wu, & Lv, 2020; Yao et al., 2019). GCNs can be generally categorized as spectral and spatial methods.

Spectral GCNs conduct convolution operation on graph spectral domains and apply spectral filtering operation on spectral domains. For example, Bruna, Zaremba, Szlam, and LeCun (2014) model the global structure of a graph with the spectrum of the graph-Laplacian to generalize the convolution operation. Kipf and Welling (2017) exploit the spectral structure of the graph, and propose a convolutional architecture via a localized first-order approximation of spectral graph convolutions. Levie, Monti, Bresson, and Bronstein (2019) employ an efficient spectral filtering scheme based on the new class of Cayley polynomials, which holds similar advantages of the Chebyshev filters such as localization and linear complexity in the number of edges.

Apart from spectral GCNs, spatial GCNs have recently attracted an increasing attention by researchers. Spatial approaches learn embeddings via aggregating a node's neighborhood information and multi-hop convolutional operations is used to model higher-order proximity information. For example, some recent works (Zhang et al., 2019; Zhang, Qi, & Manning, 2018) are devoted to applying GCN over dependency trees of sentences in order to exploit long-range multi-word relations. Zhang et al. (2018) encode the dependency structure over the input sentence with graph convolution operations, and then extracts entity-centric representations for relation extraction. Zhang et al. (2019) address aspect-based sentiment classification by applying a multi-layered graph convolution structure on top of the LSTM output, and then utilize a masking mechanism to obtain aspect-specific features. There also have been several recent efforts proposed to conduct GCN over a knowledge graph (KG). Wang et al. (2019) investigate the problem of KG-aware recommendation by capturing both higher-order structure and semantic information in the KG. To calculate the representation of a given entity in the KG, they sample a fixed-size neighborhood of each node as the receptive field and conduct convolutional operations over neighborhood information.

The most relevant literature to our work is TextGCN (Yao et al., 2019), which proposes to leverage GCN for text classification. Specifically, they construct a large and heterogeneous text graph which consists of word nodes and document nodes. Then a GCN is applied on the constructed text graph to explicitly model word-word and document-word relations. While TextGCN has shown state-of-the-art performance in some text classification tasks, the effectiveness in aspect-based sentiment classification still remains an open problem. Inspired by (Yao

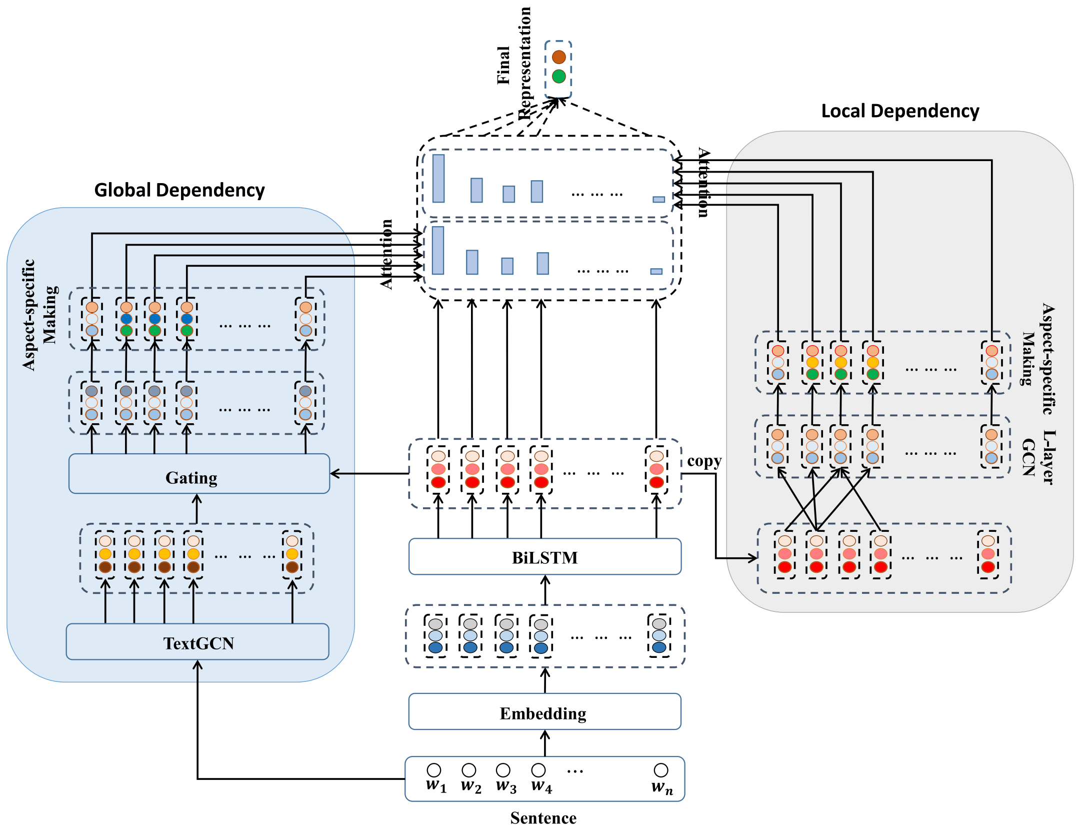

Fig. 2. The proposed architecture of GL-GCN.

et al., 2019), we aim to extract word global dependency information from the text graph constructed by the entire corpus. One limitation of TextGCN is that they have to incorporate all test documents for constructing the text graph in order to learn each test document representation during the GCN, which would limit its practical usage. In contrast to (Yao et al., 2019), our proposed model is more flexible as it only relies on the training dataset to learn informative word embeddings.

2.2. Aspect-based sentiment analysis

Early research works usually rely on hand-crafted features and traditional classification models such as SVM (Kiritchenko, Zhu, Cherry, & Mohammad, 2014; Wagner et al., 2014). Recently, neural network models are of growing interest due to their capacity to learn text representations from data without feature engineering. Tang, Qin, and Liu (2016b) develop a deep memory network for aspect-based sentiment classification. This model consists of multiple computational layers with shared parameters, and each layer is a content and location based attention model. Wang et al. (2016) propose attention-based LSTM for aspect-level sentiment classification. It introduces an attention mechanism to enforce the model to attend to the important part of a sentence concerning the given aspect. Zhang, Zhang, and Vo (2016) further use a gated recurrent neural network layer to model syntactic and semantic information of the input text. The gated neural network is used to explicitly model the interaction between the left context, the right context and the target. As LSTM based models treat each context word equally (with the same operation), it cannot explicitly reveal the importance of each context word. To deal with this issue, Chen, Sun, Bing, and Yang (2017) adopts a multiple-attention mechanism, which are non-linearly combined with a recurrent neural network, to capture sentiment features separated by a long distance. Ma, Li, Zhang, and Wang (2017) adopted two attention networks to interactively learn attentions in the context and targets, and generate the representations for target and contexts separately. Some recent works also propose to adopt convolutional neural networks for aspect-based sentiment classification, which has been shown that competitive performance can be achieved by capturing multi-word phrases. Xue and Li (2018) employed convolutional neural networks and gating mechanisms. In particular, it has two separate convolutional layers on the top of the embedding layer, and a gating unit is then used to combine the output of the two convolutional layers. Works mentioned above cannot sufficiently determine sentiments depicted by long-range multi-word relations. To account for this limitation, very recently, Zhang et al. (2019) applied a GCN over the dependency tree of a sentence to extract exploit long-range multi-word relations and syntactical information. Specifically, they propose a multi-layered graph convolution structure on top of the LSTM output, and then apply a masking mechanism to keep high-level aspect-specific features. These aspect-specific features will be fed back to the LSTM output for retrieving informative features according to the aspect.

One major limitation of existing methods is that they mainly focus on exploiting local structure information of a given sentence, such as locality, sequentiality or syntactical dependency constraints within the sentence.

3. Our approach

The main novelty of our proposed approach GL-GCN is to exploit both global and local dependency structure signals to better address the issue of long-range of multi-word dependency. In this section, we present the details of GL-GCN. In particular, we first formulate the problem of aspect-based sentiment classification as well as the graph convolutional networks, then present the framework of our GL-GCN model and introduce the global dependency module as well as the local dependency module. At last, we present how to obtain final representation and model training.

Table 1
Notations and their definitions.

Notation	Definition
S	A sentence with n-words
n	Length of S
m	Length of the aspect
w_i	The ith word in S
M	An embedding matrix
V	The vocabulary (the set of nodes)
d_e	Dimension of the word embedding
E	Word embedding sequence
e_{w_i}	The embedding of the ith word in S
H	The output hidden state vectors of BiLSTM
h_i	The ith hidden state vector of BiLSTM
G	A graph
A	The adjacency matrix of G
L	The maximum layer of GCN
$W^{(l)}$	Trainable weight matrix for the lth layer of GCN
$b^{(l)}$	Trainable bias for the lth layer of GCN
$h_i^{(l)}$	The output vector of the ith layer for node i
σ	A nonlinear function
$\tilde{A}$	The adjacency matrix with self-loops of G
I	A $n \times n$ identity matrix
d_i	The degree of the ith node in G
$g(i)$	The position weight of the ith word in S
τ	A position indicator related to the aspect term
A'	Weighted matrix of the text graph
Y	The ground-truth label matrix
E^g	Global word embedding matrix
Z^g	Word embedding matrix fused with global and local signals
$\tilde{Z}^g$	Position-aware transformation of Z^g
$\tilde{Z}^g_{mask}$	Aspect masked $\tilde{Z}^g$
β_t	Global dependency attention weight of the tth word
δ_t	Fused global and local attention weight of the tth word
r	Aspect-oriented hidden representation of sentence S

3.1. Problem formulation

Given a n-words sentence $S = (w_1, \dots, w_{\tau+1}, \dots, w_{\tau+m}, \dots, w_n)$ as well as an m-words aspect starting from the $(\tau + 1)$-th word, the goal of aspect-based sentiment classification is to identify the sentiment polarity of the sentence S given the specified aspect. To this end, we first embed the sentence with an embedding matrix $M \in \mathbb{R}^{|V|\times d_e}$, where $|V|$ denotes the vocabulary size and d_e is the embedding dimension. In particular, we embed each word in the sentence into a low-dimensional real-valued word embedding, and obtain $E = (e_{w_1}, \dots, e_{w_{\tau+1}}, \dots, e_{w_{\tau+m}}, \dots, e_{w_n})$. Then we use a bi-directional LSTM (BiLSTM), which is leveraged for modeling the sequential property of a sentence. We concatenate the hidden vectors from both directions of LSTM and construct hidden state vectors $H = (h_1, \dots, h_{\tau+1}, \dots, h_{\tau+m}, \dots, h_n)$, where $h_t \in \mathbb{R}^{2d_h}$ is the hidden state at time step t, and d_h represents the dimensionality of a hidden state vector of a non-direction LSTM. Finally, the output of BiLSTM will be fed into two separate modules, where two kinds of GCN are utilized to learn attention weights from global and local dependency structures, respectively. The framework of GL-GCN is illustrated in Fig. 2 and the key notations used throughout the rest of this paper are summarized in Table 1.

3.2. Graph Convolutional Networks

We first give a brief introduction of GCN, which is a multi-layer neural network that adapts convolution operations over nodes in a graph. Formally, we consider a graph $G = (V, E)$, where $V(|V| = n)$ is the set of nodes and E is the set of edges. Denote $A \in \mathbb{R}^{n\times n}$ as the corresponding adjacency matrix of G, where $A_{ij} = 1$ if there is an edge between word i and word j. In a L-layer GCN, denote $h_i^{(l)}$ $(l \in \{1, \dots, L\})$ as the output of the lth layer for node i, the graph convolution operation can be formally defined as:

$$h_i^{(l)} = \sigma(\sum_{j=1}^{k} A_{ij} W^{(l)} h_j^{(l-1)} + b^{(l)}) \tag{1}$$

where $W^{(l)}$ is a linear transformation weight, $b^{(l)}$ is a bias term, and σ is a nonlinear function (e.g., ReLU). Through the above graph convolution operation, each node can aggregate information from its immediate neighbors, and a L-layer GCN then can be used to gather information from neighboring nodes within L steps.

3.3. Local dependency attention

In this section, we introduce how to obtain aspect-oriented attention from local dependency structure. To address this issue, inspired by the work in (Zhang et al., 2019), we perform a graph convolution network over dependency trees of sentences, and apply it on top of the BiLSTM mentioned above. Through this process, long-range multi-word relations at sentence level can be effectively captured. To be specific, we first construct a dependency tree for each given sentence, and obtain its adjacency matrix $A \in \mathbb{R}^{n\times n}$ according to the words in the sentence. Then we apply the graph convolution operation to model dependency trees. As suggested in (Zhang et al., 2019, 2018), we add self-loops to each node and normalize the activations in the graph convolution before converting it through the nonlinearity, which can be formalized as follows:

$$h_i^{(l)} = \sigma(\sum_{j=1}^{n} \tilde{A}_{ij} W^{(l)} h_j^{(l-1)} / d_i + b^{(l)}) \tag{2}$$

where $\tilde{A} = A + I$ with I being the $n \times n$ identity matrix, $d_i = \sum_{j=1}^{n} \tilde{A}_{ij}$ denotes the degree of the ith node, $W^{(l)}$ and $b^{(l)}$ are layer-specific trainable parameters.

Similar to (Li et al., 2018; Zhang et al., 2019), a position-aware transformation is leveraged to reveal the importance of context words with respect to the given aspect. Specifically, we define the position weight to the ith word as follows:

$$g(i) = \begin{cases} 1 - \frac{\tau+1-i}{n} & 1 \le i < \tau + 1 \\ 0 & \tau + 1 \le i \le \tau + m \\ 1 - \frac{i-\tau-m}{n} & \tau + m < i \le n \end{cases} \tag{3}$$

Therefore, the graph convolution operation in Eq. (2) will be reformulated as:

$$h_i^{(l)} = \sigma(\sum_{j=1}^{n} g(j) \tilde{A}_{ij} W^{(l)} h_j^{(l-1)} / d_i + b^{(l)}) \tag{4}$$

The final output of the L-layer GCN over dependency trees is $H^{(L)} = (h_1^{(L)}, \dots, h_{\tau+1}^{(L)}, \dots, h_{\tau+m}^{(L)}, \dots, h_n^{(L)})$. We further apply an aspect-specific masking layer on top by masking out non-aspect word states with zeros, and the output of masking layer is $H_{mask}^{(L)} = (0, \dots, h_{\tau+1}^{(L)}, \dots, h_{\tau+m}^{(L)}, \dots, 0)$. Then, the local dependency attention weights α are computed as follows:

$$e_t = \sum_{i=\tau+1}^{\tau+m} h_t{}^T h_i^{(L)} \tag{5}$$

$$\alpha_t = \frac{exp(e_t)}{\sum_{i=1}^{n} exp(e_i)} \tag{6}$$

3.4. Global dependency attention

In this section, we aim at learning global dependency attention weights by extracting word global dependency structure information. There are many works devoted to extracting word global dependency structures, such as word co-occurrence graph (Peng et al., 2018) or text graph (Yao et al., 2019). Compared with constructing a word co-occurrence graph, a text graph consists of both words and sentences as nodes. With sentence node as bridge, long-range word relations over the entire corpus can be captured, which will be used to further enhance the word representation learning. Thus, in this work, we pretrain a TextGCN over the text graph to learn informative word representations from a global perspective.

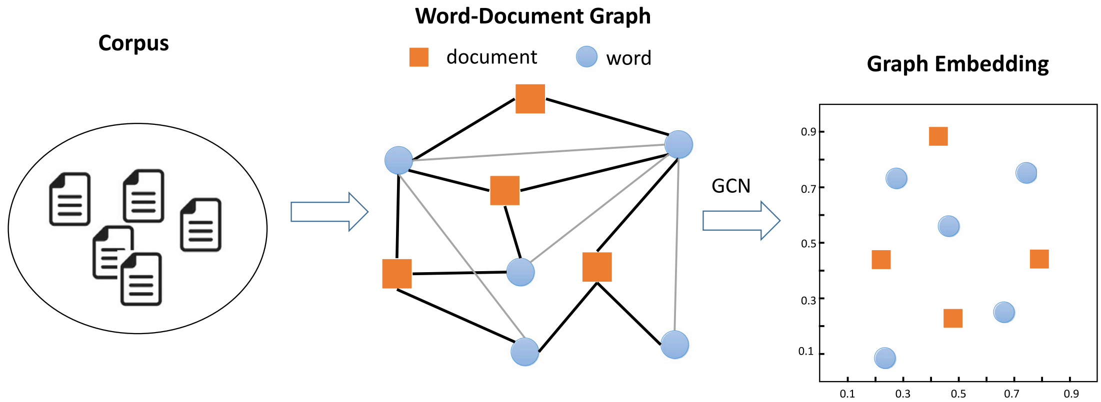

Fig. 3. Illustration of the TextGCN Framework. The bold line denotes the document-word edge weighted by the TF-IDF value, and the light line depicts the word-word edge weighted by the PMI value.

TextGCN: It is worth noting that the original work TextGCN approach (Yao et al., 2019) learns sentence[1] representations for the test data during the training phase, which will restrict its real applications. In our work, we only leverage sentences in the training data to pre-train a word embedding matrix from a global dependency view, which will be easily incorporated into our proposed model. In particular, a text graph is constructed from the training corpus consisting of words and sentences as nodes. The edge between two word nodes is built by utilizing global word co-occurrence information, and the edge between a word node and a sentence node is built by using the TF-IDF of the word in the document. Formally, the weight of an edge between node i and node j is defined as

$$A'_{ij} = \begin{cases} PMI(i,j) & i,j \text{ are words}, PMI(i,j) > 0 \\ TF-IDF(i,j) & i \text{ is document}, j \text{ is word} \\ 1 & i = j \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The PMI value of a word pair i, j is computed as

$$PMI(i,j) = \log \frac{p(i,j)}{p(i)p(j)} \tag{8}$$

$$p(i,j) = \frac{\#N(i,j)}{\#N} \tag{9}$$

$$p(i) = \frac{\#N(i)}{\#N} \tag{10}$$

where $\#N(i)$ is the number of sliding windows in a corpus that contain word i, $\#N(i,j)$ is the number of sliding windows that contain both word i and j, and $\#N$ is the total number of sliding windows in the corpus. Here, we only consider edges between word nodes with positive PMI values.

We then apply a two-layer GCN, named TextGCN, over the text graph, which is formulated as:

$$\hat{Y} = softmax(\bar{A}'\sigma(\bar{A}'XW_0)W_1) \tag{11}$$

where $\bar{A}' = D^{-1/2}A'D^{-1/2}$, $\hat{Y}$ is the output, X is a matrix containing all n nodes with their features, W_0 and W_1 are trainable weight matrix. The loss function is defined as the cross-entropy error over all labeled sentences:

$$\mathcal{L}_G = -\sum_{d \in \mathcal{Y}_D} \sum_{f=1}^{F} Y_{df} \log \hat{Y}_{df} \tag{12}$$

where $\mathcal{Y}_D$ is the set of sentences with labels and F is the dimension of the output features, i.e., the number of classes. Y is the ground-truth label matrix. The framework of the TextGCN is illustrated in Fig. 3.

[1] In TextGCN, they focus on long text scenarios(i.e., document), while in our work, we consider short-text(i.e., sentence).

We pre-train a TextGCN over the entire corpus, and obtain word representations which will be used as the word embedding matrix. The learnt word representations reflect word semantic relations from a global dependency perspective. With the learnt word embedding matrix in hand, we then embed the original input sentence and obtain $E^g = (e^g_{w_1}, \dots, e^g_{w_{\tau+1}}, \dots, e^g_{w_{\tau+m}}, \dots, e^g_{w_n})$. After that, we combine E^g with the output of BiLSTM H using a gate mechanism. Formally, the fused hidden representation Z^g are computed by:

$$\begin{aligned} G &= sigmoid(W_g e^g_{w_i} + W_h h_i + b_g), \\ z^g_i &= G \odot e^g_{w_i} + (1 - G) \odot h_i \end{aligned} \tag{13}$$

where $W_g \in \mathbb{R}^{2d_h \times 2d_h}$, $W_h \in \mathbb{R}^{2d_h \times 2d_h}$, and $b_g \in \mathbb{R}^{2h}$ are the parameters in the gating layer. In order to learn the aspect-oriented global dependency attention weights, we further involve a position-aware transformation layer as well as an aspect-specific masking layer as conducted in the local dependency attention module. In particular, we apply a position-aware transformation of Z^g and obtain $\tilde{Z}^g = (\tilde{z}^g_1, \dots, \tilde{z}^g_{\tau+1}, \dots, \tilde{z}^g_{\tau+m}, \dots, \tilde{z}^g_n)$ where $\tilde{z}^g_i = g(i)z^g_i$. Then an aspect-specific masking layer is applied on top of it and we obtain $\tilde{Z}^g_{mask} = (0, \dots, \tilde{z}^g_{\tau+1}, \dots, \tilde{z}^g_{\tau+m}, \dots, 0)$. Finally, we obtain the global dependency attention weights β, which are computed as follows:

$$e_t = \sum_{i=\tau+1}^{\tau+m} {h_t}^T \tilde{z}^g_i \tag{14}$$

$$\beta_t = \frac{exp(e_t)}{\sum_{i=1}^{n} exp(e_i)} \tag{15}$$

3.5. Obtaining final representation and model training

The final representation of the sentence with a given aspect will be calculated by leveraging both aspect-oriented global and local dependency attention information. Formally, we simply add both global and local attention weights, and obtain the fused attention weights $\delta_t = \alpha_t + \beta_t$, $t = (1, \dots, n)$. We also apply others attention weight fusion strategies, such as multiply or maximum, and the results show that a simple add operation demonstrates a superior performance. Finally, the aspect-oriented hidden representation of sentence r are computed as follows:

$$r = \sum_{t=1}^{n} \delta_t h_t \tag{16}$$

We then feed r into a fully-connected layer and a softmax layer to produce a probability distribution $p \in \mathbb{R}^{d_p}$ as follows:

$$p = softmax(W_p r + b_p) \tag{17}$$

where d_p equals to the number of sentiment labels, $W_p \in \mathbb{R}^{d_p \times 2d_h}$ and $d_p \in \mathbb{R}^{d_p}$ are trainable weights and bias.

Table 2
Statistics of the datasets.

Dataset		#Positive	#Neutral	#Negative
TWITTER	Train	1561	3127	1560
	Test	173	346	173
LAPTOP	Train	994	464	870
	Test	341	169	128
REST14	Train	2164	637	807
	Test	728	196	196
REST15	Train	912	36	256
	Test	326	34	182
REST16	Train	1240	69	439
	Test	469	30	117

For model training, we use the standard gradient descent algorithm with the cross-entropy loss and a L_2-regularization:

$$\mathcal{L} = -\sum_{d\in\mathcal{Y}_D}\sum_{f=1}^{F} Y_{df}\log p_{df} + \lambda\|\Theta\|_2 \tag{18}$$

where $\mathcal{Y}_D$ is the set of sentences with labels, F is the number of classes. Y is the ground-truth label matrix, p_{df} is the predicted probability of the document d to the fth class, and λ is the coefficient of L_2-regularization.

4. Experiments

In this section, we compare our method with a range of competitive baselines on five real-world datasets.

4.1. Datasets

We conduct our evaluation on five real-world datasets: TWITTER , LAPTOP, REST14 , REST15 and REST16. The TWITTER dataset contains twitter posts, which are built by Dong, Wei, Tan, Tang, Zhou, and Xu (2014). The other four datasets (LAPTOP, REST14, REST15, REST16) are built by Pontiki et al. (2016), Pontiki, Galanis, Papageorgiou, Manandhar, and Androutsopoulos (2015), Pontiki, Galanis, Pavlopoulos, et al. (2014)). In our experiments, similar to Tang et al. (2016b), we remove instances which have conflicting polarities or no explicit aspects. The statistics of all datasets are summarized in Table 2.

4.2. Baselines

We compare the proposed approach to the following 7 state-of-the-art methods :

- SVM (Kiritchenko et al., 2014): This model won the SemEval-2014 Task 4. It is based on a traditional support vector machine and relies on conventional feature extraction for aspect-level sentiment classification.
- LSTM (Tang, Qin, Feng, & Liu, 2016a): This is an effective LSTM model for target-dependent sentiment classification, which used the last hidden state vector for the prediction.
- MemNet (Tang et al., 2016b): Unlike feature-based models SVM and sequential neural models such as LSTM, this model considered contextual information as external memories for inferring the sentiment polarity of an aspect.
- AOA (Huang, Ou, & Carley, 2018): It captures the interaction between aspects and context sentences, and jointly learn the representations for both aspects and sentences. In particular, the target representation and text representation generated from LSTMs interact with each other through an attention-over-attention module (Cui, Chen, Wei, Wang, Liu, & Hu, 2017).
- IAN (Ma et al., 2017): It proposes to interactively learn attentions in the contexts and targets, and generate the representations for targets and contexts separately. To be specific, it computes the context representation via utilizing the attention mechanism associated with a target, and then uses the interactive information from context to supervise the modeling of the target.
- TNet-LF (Li et al., 2018): It proposes target specific transformation component to integrate target information into the word representation, and employs a CNN layer to extract salient features from the transformed word representations originated from previous LSTM-based layers.
- ASGCN (Zhang et al., 2019): To exploit the long-range word dependencies, this work considers syntactical information by applying a Graph Convolutional Network (GCN) over the dependency tree of a sentence. Specifically, it starts with a BiLSTM layer to capture contextual information regarding word orders. Then it implements a multi-layered graph convolution structure to obtain aspect-specific features, which is followed by a masking mechanism to keep high-level aspect specific features.
- ASCNN (Zhang et al., 2019): This is simplified model of ASGCN, which replaces 2-layer GCN with a 2-layer CNN in ASGCN. To some extent, ASCNN shares a similar spirit with TNet-LF, where both methods account for target specific transformation as well as leverage CNN layer to capture multi-word phrase signals.

4.3. Evaluation metrics

For evaluation, we adopt two metrics, accuracy (ACC) and macro-averaged F1-score (F1), to measure the performance of aspect-based sentiment classification.

- ACC: Accuracy measures the percentage of correct predicted samples in all samples. Formally, it is defined as:

$$Acc = \frac{T}{N} \tag{19}$$

where T denotes the number of correctly predicted samples, and N denotes the total number of samples. A higher accuracy indicates a better performance.

- F1: In this work, we use Macro-averaged F1 (Peng et al., 2018) which evaluates averaged F1 of all different class-labels. It gives equal weight to each label. Formally, Macro-averaged F1 is defined as:

$$F1 = \frac{1}{|C|}\sum_{t\in C}\frac{2P_tR_t}{P_t+R_t} \tag{20}$$

where $P_t = \frac{TP_t}{TP_t+FP_t}$, $R_t = \frac{TP_t}{TP_t+FN_t}$, and TP_t, FP_t, FN_t denote the true-positives, false-positives, and false-negatives for the tth label in a label set C, respectively.

4.4. Experimental settings

For GL-GCN, we initialize word embeddings using a 300-dimensional pretrained GloVe vectors (Pennington, Socher, & Manning, 2014). The dimension for hidden states of TextGCN is set to 600, which is equal to the dimension of the output hidden states of BiLSTM. The number of GCN layers for syntactical dependency tree and text graph are set to 2, which is the best setting as reported in previous studies. We use Adam as the optimizer with a learning rate of 0.003. The coefficient of L_2-regularization is 10^{-5} and batch size is 32. We run the experiments 3 times with random initialization and report the averaged performance. We also perform paired t-test to verify whether the improvements achieved by our methods over the baselines are significant. We implement GL-GCN using PyTorch. All the experiments are conducted on the hardware with Intel Core CPU I7-9700K 3.6 GHz and NVIDIA GeForce GTX 2080TI.

Table 3
The performance comparison of all approaches in terms of Accuracy (ACC) and F1-score (F1). The best two performing approaches are shown in bold.

Model	TWITTER		LAPTOP		REST14		REST15		REST16	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
SVM	63.40	63.30	70.49	N/A	80.16	N/A	N/A	N/A	N/A	N/A
LSTM	69.56	67.70	69.28	63.09	78.13	67.47	77.37	55.17	86.80	63.88
MemNet	71.48	69.90	70.64	65.17	79.61	69.64	77.31	58.28	85.44	65.99
AOA	72.30	70.20	72.62	67.52	79.97	70.42	78.17	57.02	87.50	66.21
IAN	72.50	70.81	72.05	67.38	79.26	70.09	78.54	52.65	84.74	55.21
TNet-LF	**72.98**	**71.43**	74.61	70.14	80.42	71.03	78.47	59.47	**89.07**	**70.43**
ASCNN	71.05	69.45	72.62	66.72	**81.73**	**73.10**	78.47	58.90	87.39	64.56
ASGCN-DG	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**

*Indicates statistical significance at p-value < 0.05 using the paired t-test with regard to the strongest baseline ASGCN-DG.
**Indicates statistical significance at p-value < 0.01 using the paired t-test with regard to the strongest baseline ASGCN-DG.

Table 4
Ablation study results in terms of Accuracy (ACC) and F1-score (F1). The best two performing approaches are in bold.

Model	TWITTER		LAPTOP		REST14		REST15		REST16	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
G-GCN(w/o pos.)	69.56	67.96	73.67	69.25	80.27	71.40	78.11	60.53	86.85	64.91
G-GCN(w/o mask)	70.86	69.08	71.42	66.21	79.70	68.59	76.94	54.49	85.82	62.69
G-GCN	72.30	**70.65**	73.93	69.68	79.99	69.21	78.53	59.13	87.12	66.09
L-GCN(w/o pos.)	**72.69**	70.59	73.93	69.63	**81.22**	**72.94**	79.58	61.55	88.04	66.63
L-GCN(w/o mask)	72.64	70.63	72.05	66.56	79.04	68.29	77.80	57.51	86.36	61.41
L-GCN	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**

Table 5
Performance of GL-GCN with respect to different numbers of aspects and polarities.

Datasets	Model	Multi-Aspect & Multi-Polarity		Multi-Aspect & Same-Polarity		Single-Aspect	
		Acc	F1	Acc.	F1	Acc.	F1
LAPTOP	ASCNN	**64.95**	**63.42**	84.47	77.72	75.34	71.44
	ASGCN	60.82	60.50	86.21	**80.42**	75.00	71.14
	GL-GCN	61.85	61.00	**86.74**	79.43	**76.02**	**72.60**
REST14	ASCNN	66.2	65.54	88.77	77.44	82.03	**75.18**
	ASGCN	**67.13**	**65.94**	89.47	78.68	81.42	73.61
	GL-GCN	66.67	65.79	**89.81**	**79.31**	**82.59**	75.12
REST15	ASCNN	60.18	50.19	92.52	59.27	80.60	65.28
	ASGCN	63.43	53.57	91.61	58.81	81.32	64.08
	GL-GCN	**65.28**	**53.62**	**93.20**	**60.31**	**81.63**	**66.98**
REST16	ASCNN	**74.70**	52.26	**96.80**	62.10	86.56	68.94
	ASGCN	72.22	60.44	96.13	61.28	86.46	71.89
	GL-GCN	74.07	**67.29**	96.27	**64.62**	**87.47**	**73.17**

4.5. Overall performance

The performance comparison results are shown in Table 3. With the exception of REST16, the proposed model GL-GCN achieves the best performance on all datasets with respect to both evaluation metrics ACC and F1, which demonstrates the superiority of our model. Specifically, we observe that the method SVM obtains an uncompetitive performance as compared with these deep neural network based approaches. The method LSTM outperforms SVM on the dataset TWITTER, and achieves comparable results over LAPTOP dataset. The three methods MemNet, AOA and IAN share a similar strategy of leveraging the attention mechanism to capture the interaction between contexts and aspects, and they demonstrate better or competitive performance compared to LSTM. The method TNet-LF, which employs a CNN layer as well as considers target information into the word representation, presents a better performance than all above mentioned baselines. It reveals that multi-word phrase can be effectively modeled by the CNN module. ASCNN is a simplified version of ASGCN-DG and it uses a CNN layer to replace the GCN layer in ASGCN-DG. The different performance between ASCNN and ASGCN-DG indicates the effectiveness of exploiting the dependency tree to preserve the long-range multi-word syntactic relations. Our proposed method GL-GCN shows superior performance compared with all baselines. The results verify that it is critical to incorporate both global and local dependency structure signals as they can compensate each other, and GL-GCN can effectively capture both structure signals by two kinds of GCNs and fuse two dependency structure signals via a gating mechanism.

4.6. Ablation study

To investigate the impact of each component, such as the global dependency attention module and the local dependency attention module, we perform comparison between GL-GCN models and its ablations. In Table 4, L-GCN represents a model which removes the global dependency attention module from GL-GCN. It is worth noting that L-GCN is equivalent to the best performing baseline ASGCN-DG. G-GCN denotes a model which removes the local dependency attention module from GL-GCN. From Table 4, we observe that GL-GCN achieves a superior performance compared with both L-GCN and G-GCN on all datasets in term of both ACC and F1 metrics. This reveals that global

Table 6
Case Study. Visualization of attention scores from ASCNN, ASGCN and GL-GCN on testing samples, along with their corresponding predictions and gold labels.

Model	Aspect	Attention visualization	Prediction	TrueLabel
ASCNN	windows 8	did not enjoy the new windows 8 and touchscreen functions.	positive×	negative
	ssd	performance is much much better on the pro, especially if you install an ssd on it.	neutral×	positive
	drinks	a beautiful atmosphere, perfect for drinks and/or appetizers.	positive×	neutral
	pialla	the food is just okay, and it's almost not worth going unless you are getting the pialla, which is the only dish that is really good.	negative×	positive
ASGCN	windows 8	did not enjoy the new windows 8 and touchscreen functions.	negative√	negative
	ssd	performance is much much better on the pro, especially if you install an ssd on it.	neutral×	positive
	drinks	a beautiful atmosphere, perfect for drinks and/or appetizers.	positive×	neutral
	pialla	the food is just okay, and it is almost not worth going unless you are getting the pialla, which is the only dish that is really good.	neutral×	positive
GL-GCN	windows 8	did not enjoy the new windows 8 and touchscreen functions.	negative√	negative
	ssd	performance is much much better on the pro, especially if you install an ssd on it.	positive√	positive
	drinks	a beautiful atmosphere, perfect for drinks and/or appetizers .	neutral√	neutral
	pialla	the food is just okay, and it is almost not worth going unless you are getting the pialla, which is the only dish that is really good.	positive√	positive

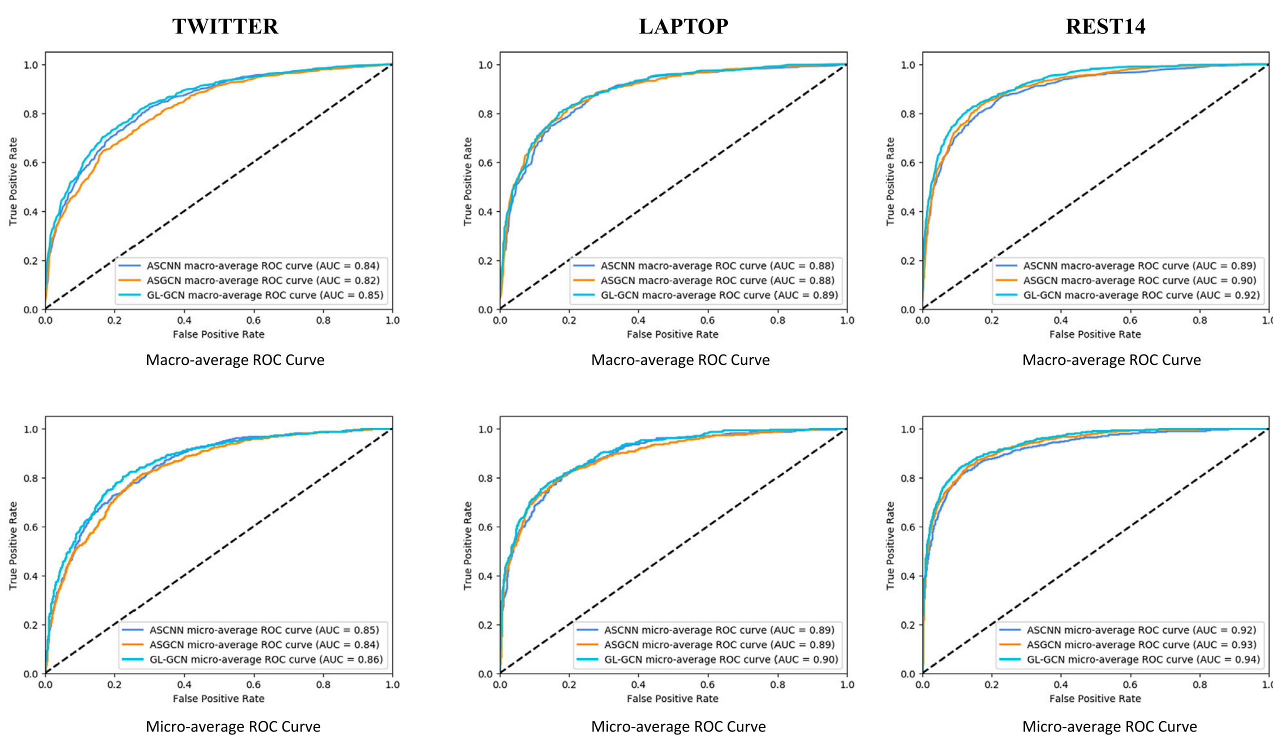

Fig. 4. ROC curves of the proposed model GL-GCN, and two most competitive baselines (i.e., ASGCN and ASCNN) on three datasets (i.e., Twitter, Lap14, Rest14).

dependency structure and local dependency structure can contribute complementary information.

We also investigate the influences of position weights and aspect-specific masking in G-GCN and L-GCN, respectively. Table 4 shows that position weights and aspect-specific masking play a critical role in G-GCN. In particular, removal of position weights from G-GCN, i.e., G-GCN (w/o pos.), results in performance decline on TWITTER and REST16, while does not affect the performance on the remaining three datasets. One reason would be that position weights would not helpful to reduce noise when modeling the aspect-specific representation. In addition, removal of aspect-specific masking G-GCN (w/o mask) from G-GCN will cause a considerable drop of performance on all datasets, which indicates the significance of aspect-specific masking in G-GCN. Similar results can also be observed in L-GCN.

4.7. Performance over different numbers of aspects and polarities

In this section, we further investigate the effectiveness of our proposed method considering different numbers of aspects and polarities. For simplicity, we manually group all test samples into three categories: 1) Single-Aspect, sentences in this category only have one aspect; 2) Multi-Aspect & Multi-Polarity, sentences in this category have more than two aspects and the polarity of these aspects are different; 3) Multi-Aspect & Same-Polarity, sentences in this category have more

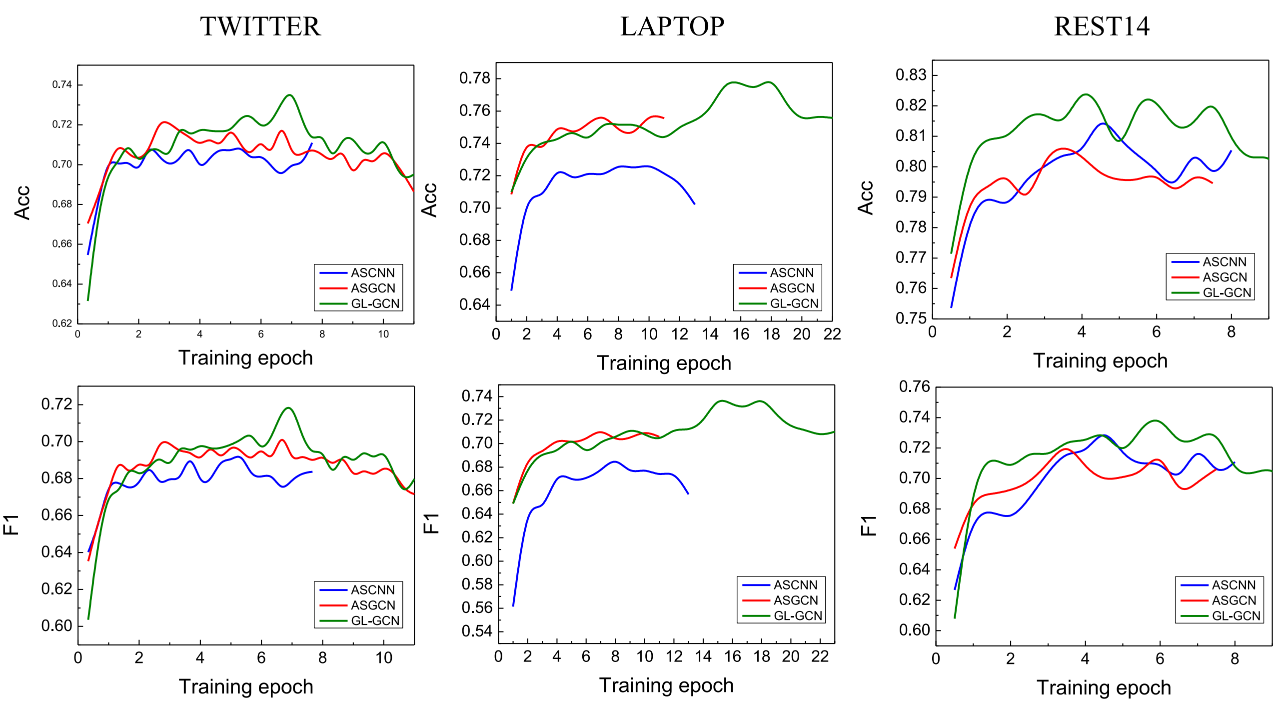

Fig. 5. Learning curves of the proposed model GL-GCN, and two most competitive baselines (i.e., ASGCN and ASCNN) on three datasets (i.e., Twitter, Lap14, Rest14).

than two aspects and all aspects have the same polarity. Table 5 shows the results of different comparing methods on four datasets. Please note that all samples of the TWITTER dataset only have one aspect, and we do not take it into consideration.

From Table 5, on the Single-Aspect, we can observe that GL-GCN considerably outperforms the two most competitive baselines (i.e., ASCNN and ASGCN) on three datasets (i.e., LAPTOP, REST15 and REST16). For example, GL-GCN achieves a F1 score of 72.60 on the LAPTOP dataset, which is 1.12 and 1.46 higher than that of ASCNN and ASGCN, respectively. We observe similar results on REST15 and REST16. On the dataset REST14, GL-GCN obtains a competitive F1 score as compared to that of the two baselines. On the Accuracy metric, GL-GCN consistently outperforms both ASCNN and ASGCN.

On the Multi-Aspect & Same Polarity, our proposed method GL-GCN has a superior or competitive performance as compared with two baselines. It is interesting that on the Multi-Aspect & Multi-Polarity, GL-GCN does not demonstrate a substantially improvement over the two baselines. The reason would that the task becomes more complicated when sentences with multiple aspects and multiple polarities. We leave this issue as the future work.

4.8. Case study

To better understand how GL-GCN works, in this section, we present a case study with several testing samples. In particular, we visualize the attention scores offered by ASCNN, ASGCN and GL-GCN in Table 6, along with their predictions on these samples and the corresponding ground truth labels. The first sample is "did not enjoy the new windows 8 and touchscreen functions "., which contains negation in the sentence and can easily lead models to make wrong predictions. The second sample is "performance is much much better on the pro , especially if you install an ssd on it ". The aspect word in this sentence is far away from the corresponding opinion word, thus it is difficult for the model to capture its true emotional polarity. The third sample is "a beautiful atmosphere , perfect for drinks and/or appetizers ". For sentences with neutral sentimental polarity and positive words in the sentence, the model is generally difficult to make a correct prediction. The last sample is "the food is just okay , and it's almost not worth going unless you are getting the pialla , which is the only dish that's really good"., in which there are two aspect words, and the model can easily mix up the corresponding opinion words.

From Table 6, we can observe that ASCNN fails in all four samples. Although ASGCN can integrate syntactic relevance information into rich semantic representations, it still fails in most of the cases as it faces difficult to distinguish sentences with neutral terms and sentimental polarity. Our model GL-GCN handles all four samples correctly, which reflects that the local and global dependency information modeled by GL-GCN plays a crucial role in recognizing sentences sentimental polarity.

4.9. Simulation performance

To further investigate the performance of our proposed approach GL-GCN, we compare the Receiver Operating Characteristic (ROC) curves of GL-GCN with that of two most competitive baselines (i.e., ASGCN and ASCNN) on three datasets (i.e., Twitter, Lap14, Rest14). As ROC curve is usually used for the binary classification task, we leverage two variants of ROC curve, named macro-average ROC curve and micro-average ROC curve, for our multi-class classification task.

Fig. 4 shows the ROC curves of our method GL-GCN and two baselines. We can observe that on all three datasets, GL-GCN consistently demonstrates superior performance as compared with ASGCN and ASCNN with respect to both macro-average ROC curves (top row of Fig. 4) and micro-average ROC curves (bottom row of Fig. 4). In addition, we also give the Area Under the Curve (AUC) value of each method on three datasets in Fig. 4. The results show that GL-GCN obtains a higher AUC value as compared with that of ASGCN and ASCNN.

4.10. Learning curve

We further conduct experiments on three datasets (i.e., TWITTER, LAPTOP, REST14) to investigate the convergence speed of the proposed methods. It is worth noting that we apply early stopping for training the model, which stops training once the model performance stops improving on the validation set. Fig. 5 shows the learning curve of GL-GCN and two most competitive baselines (i.e., ASGCN and ASCNN). From Fig. 5, we can observe that: First, our proposed model achieves its best performance at a small epoch. For example, GL-GCN obtains the best performance at epoch 7, 15 and 4 on the TWITTER, LAPTOP, and REST14, respectively. Second, our model consistently outperforms both ASGCN and ASCNN, which verifies the effectiveness of the proposed model.

5. Conclusion

In this paper, we investigate the aspect-based sentiment classification problem and propose a novel model, Global and Local Dependency Guided Graph Convolutional Network (GL-GCN), to deal with it. Based on the text graph built on the entire corpus, we apply a graph convolutional network to mine word global semantic dependency relations. Further, a dependency tree built on a sentence is leveraged to extract word local syntactic dependency relations. Extensive experiments are conducted on five real-word datasets, and experimental results show that our proposed method achieves superior performance compared to the state-of-the-art methods.

CRediT authorship contribution statement

Xiaofei Zhu: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Ling Zhu:** Software, Investigation, Writing – original draft. **Jiafeng Guo:** Supervision, Project administration. **Shangsong Liang:** Writing – review & editing. **Stefan Dietze:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [grant number 61722211]; the Federal Ministry of Education and Research, Germany [grant number 01LE1806A]; the Beijing Academy of Artificial Intelligence, China [grant number BAAI2019ZD0306]; and the Technology Innovation and Application Development of Chongqing, China [grant number cstc2020jscx-dxwtBX 0014].

References

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd international conference on learning representations*.

Battaglia, P. W., Hamrick, J. B., Bapst, V., et al. (2018). Relational inductive biases, deep learning, and graph networks. CoRR. abs/1806.01261.

Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, *3*, 1137–1155.

Bordes, A., Usunier, N., García-Durán, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In *Proceedings of the 26th international conference on neural information processing systems* (pp. 2787–2795).

Bruna, J., Zaremba, W., Szlam, A., & LeCun, Y. (2014). Spectral networks and locally connected networks on graphs. In *Proceedings of the 2nd international conference on learning representations*.

Chen, P., Sun, Z., Bing, L., & Yang, W. (2017). Recurrent attention network on memory for aspect sentiment analysis. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 452–461).

Cui, Y., Chen, Z., Wei, S., Wang, S., Liu, T., & Hu, G. (2017). Attention-over-attention neural networks for reading comprehension. In *Proceedings of the 55th annual meeting of the association for computational linguistics* (pp. 593–602).

Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., & Xu, K. (2014). Adaptive recursive neural network for target-dependent Twitter sentiment classification. In *Proceedings of the 52nd annual meeting of the association for computational linguistics* (pp. 49–54).

Fan, C., Gao, Q., Du, J., Gui, L., Xu, R., & Wong, K.-F. (2018). Convolution-based memory network for aspectbased sentiment analysis. In *Proceedings of the 41st international ACM SIGIR conference on research and development in information retrieval* (pp. 1161–1164).

Gong, P., & Ai, L. (2019). Neighborhood adaptive graph convolutional network for node classification. *IEEE Access*, *7*, 170578–170588.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Huang, B., Ou, Y., & Carley, K. M. (2018). Aspect level sentiment classification with attention-over-attention neural networks. In *SBP-BRiMS 2018* (pp. 197–206).

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th international conference on learning representations*.

Kiritchenko, S., Zhu, X., Cherry, C., & Mohammad, S. (2014). Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In *SemEval* (pp. 437–442).

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Levie, R., Monti, F., Bresson, X., & Bronstein, M. M. (2019). CayleyNets: Graph convolutional neural networks with complex rational spectral filters. *IEEE Transactions on Signal Processing*, *67*(1), 97–109.

Li, X., Bing, L., Lam, W., & Shi, B. (2018). Transformation networks for target-oriented sentiment classification. In *Proceedings of the 56th annual meeting of the association for computational linguistics* (pp. 946–956).

Lin, Y., Liu, Z., Sun, M., Liu, Y., & Zhu, X. (2015). Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the 29th AAAI conference on artificial intelligence* (pp. 2181–2187).

Liu, X., You, X., Zhang, X., Wu, J., & Lv, P. (2020). Tensor graph convolutional networks for text classification. CoRR.

Liu, J., & Zhang, Y. (2017). Attention modeling for targeted sentiment. In *Proceedings of the 15th conference of the european chapter of the association for computational linguistics*.

Ma, D., Li, S., Zhang, X., & Wang, H. (2017). Interactive attention networks for aspect-level sentiment classification. In *Proceedings of the 26th international joint conference on artificial intelligence* (pp. 4068–4074).

Peng, H., Li, J., He, Y., Liu, Y., Bao, M., Wang, L., et al. (2018). Large-scale hierarchical text classification with recursively regularized deep graph-CNN. In *Proceedings of the 27th international conference on world wide web* (pp. 1063–1072).

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing* (pp. 1532–1543).

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., et al. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *SemEval* (pp. 19–30).

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., & Androutsopoulos, I. (2015). Semeval-2015 task 12: Aspect based sentiment analysis. In *SemEval 2015* (pp. 486–495).

Pontiki, M., Galanis, D., Pavlopoulos, J., et al. (2014). Semeval-2014 task 4: Aspect based sentiment analysis. In *SemEval 2014* (pp. 27–35).

Tang, D., Qin, B., Feng, X., & Liu, T. (2016). Effective lstms for target-dependent sentiment classification. In *Proceedings of the 25th international conference on computational linguistics* (pp. 3298–3307).

Tang, D., Qin, B., & Liu, T. (2016). Aspect level sentiment classification with deep memory network. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 214–224).

Wagner, J., Arora, P., Cortes, S., Barman, U., Bogdanova, D., Foster, J., et al. (2014). DCU: aspect-based polarity classification for SemEval task 4. In *SemEval* (pp. 223–229).

Wang, Y., Huang, M., Zhao, L., & Zhu, X. (2016). Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*.

Wang, H., Zhao, M., Xie, X., Li, W., & Guo, M. (2019). Knowledge graph convolutional networks for recommender systems. In *Proceedings of the 28th international conference on world wide web* (pp. 3307–3313).

Xue, W., & Li, T. (2018). Aspect based sentiment analysis with gated convolutional networks. In *Proceedings of the 56th annual meeting of the association for computational linguistics* (pp. 2514–2523).

Yao, L., Mao, C., & Luo, Y. (2019). Graph Convolutional networks for text classification. In *Proceedings of the 33th AAAI conference on artificial intelligence* (pp. 7370–7377).

Zhang, C., Li, Q., & Song, D. (2019). Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing* (pp. 4567–4577).
Zhang, Y., Qi, P., & Manning, C. D. (2018). Graph convolution over pruned dependency trees improves relation extraction. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 2205–2215).
Zhang, M., Zhang, Y., & Vo, D. (2016). Gated neural networks for targeted sentiment analysis. In *Proceedings of the 30th AAAI conference on artificial intelligence* (pp. 3087–3093).

Prof. Dr. Xiaofei Zhu is a full professor at College of Computer Science and Engineering, Chongqing University of Technology. He received his Ph.D. degree at the Institute of Computing Technology, Chinese Academy of Science (ICT-CAS) in 2012. Then he spent four years as a Postdoctoral Research Fellow at the L3S Research Center, Leibniz University Hannover. His research interests include web search, data mining and machine learning, and he has published more than 30 papers in international conferences and journals, including the top conferences like SIGIR, WWW, CIKM, TKDE, etc. He has won the Best Paper Awards of CIKM (2011). He serves as area chair, program committees and editorial board of numerous international conferences and journals, including SIGIR, AAAI, CIKM, etc.

Ling Zhu is a Master Candidate at College of Computer Science and Engineering, Chongqing University of Technology. She received the B.S. degree in Computer Science from the Neijiang Normal University in 2014. Her main research interest focuses on machine learning, text mining and sentiment analysis.

Jiafeng Guo is currently a Professor in Institute of Computing Technology, Chinese Academy of Sciences. He received his Ph.D. in Computer Software and Theory from the University of Chinese Academy of Sciences, Beijing, China, in 2009. He has worked on a number of topics related to web search and data mining, including query representation and understanding, learning to rank, and text modeling. His current research focuses on representation learning and neural models for information retrieval and filtering. He has published more than 80 papers in several top conferences/journals such as SIGIR, WWW, CIKM, IJCAI, and TKDE. His work on information retrieval has received the Best Paper Award in ACM CIKM (2011), Best Student Paper Award in ACM SIGIR (2012) and Best Full Paper Runner-up Award in ACM CIKM (2017). Moreover, he has served as the PC member for the prestigious conferences including SIGIR, WWW, KDD, WSDM, and ACL, and the associate editor of TOIS.

Shangsong Liang is an associate professor at the School of Data and Computer Science, Sun Yat-sen University. He received a Ph.D. degree from the University of Amsterdam in 2014. His research interests lie in the field of Information Retrieval, Data Mining, Artificial Intelligence and Deep Learning. He has published over 50 peer-reviewed papers, most of which are in top-tier venues such as SIGIR, KDD, WSDM, AAAI, CIKM, ECIR, IEEE TKDE, ACM TOIS and Information Processing and Management. He received various awards/honors such as the SIGIR 2017 Outstanding Reviewer Award, Outstanding Contribution for instructing Data Mining course from the International Petroleum Engineers, the Kingdom of Saudi Arabia Section.

Prof. Dr. Stefan Dietze is a professor at the University of Dsseldor and the scientific director of Knowledge Technologies for the Social Sciences (WTS), GESIS . He received his Ph.D. from the Institute for Computer Science of the University of Potsdam (Ph.D./Dr. rer. nat. in Applied Computer Science), Germany, in 2004. His professional career led him to the Fraunhofer Institute for Software and Systems Engineering (ISST) in Berlin. Then, he spent five years as a Postdoctoral Research Fellow at the Knowledge Media Institute (KMi) of The Open University in Milton Keynes, UK. Before joining GESIS, he led a research group at the L3S research center at Leibniz University Hannover. His research interests include semantic web technologies, web intelligence, semantic search, and information retrieval. He has published numerous papers in prestigious journals and international conferences, including SIGIR, WWW, CIKM, ESWC, CHIIR and so on. He has been general/poster/track chair, program committee and editorial board member of numerous international conferences and journals.