RESEARCH ARTICLE

Aspect-Level Sentiment Analysis Using CNN Over BERT-GCN

HUYEN TRANG PHAN [1], **NGOC THANH NGUYEN** [2,3], **(Senior Member, IEEE), AND DOSAM HWANG** [1]

[1]Department of Computer Engineering, Yeungnam University, Gyeongsan 38541, South Korea
[2]Department of Applied Informatics, Wroclaw University of Science and Technology, 50-370 Wroclaw, Poland
[3]Faculty of Information Technology, Nguyen Tat Thanh University, Ho Chi Minh 70000, Vietnam

Corresponding authors: Dosam Hwang (dshwang@yu.ac.kr) and Ngoc Thanh Nguyen (ngoc-thanh.nguyen@pwr.edu.pl)

ABSTRACT Context-based GCNs have achieved relatively good effectiveness in the sentiment analysis task, especially aspect-level sentiment analysis (ALSA). However, the previous context-based GCNs for ALSA often used GCNs with the following limitations: (i) Using GCNs limited to a few layers (two or three) due to the vanishing gradient, limiting their performance. (ii) Not considering helpful information about the hidden context between the words. To solve these limitations, this paper proposes a novel CNN over the BERT-GCN model for ALSA. The contributions of the proposed method are summarized as follows: (i) Handling the disadvantage of limiting the GCN to a few layers by adding convolutional layers of the convolutional neural network (CNN) model after GCN layers. (ii) Considering further helpful information about the hidden context between the words by combining the Bidirectional Encoder Representations from Transformers (BERT) and Bidirectional Long Short Term Memory (BiLSTM) models. The proposed model includes the following steps: First, words in sentences are converted vectors using BERT. Second, the contextualized word representations are created based on BiLSTM over word vectors. Third, significant features are extracted and represented using the GCN model with multiple convolutional layers over the contextualized word representations. Finally, the aspect-level sentiments are classified using the CNN model over the feature vectors. Experiments on three benchmark datasets illustrate that our proposed model has improved the performance of the previous context-based GCN methods for ALSA.

INDEX TERMS Aspect-level sentiment analysis, GCN, CNN, BiLSTM, BERT-GCN.

I. INTRODUCTION

Since 2004, the number of users signing up on social networks has been growing more and more and increasing exponentially, and it has not ceased owing to wide popularity. The social networks statistics in 2021 show that 3.78 billion users worldwide have used social media in 2021 – and this number will continue growing over the next few years.[1] Opinions on social networks are significant sources of news and information for traders, governments, and researchers [27]. By extracting, evaluating, and analyzing hidden sentiments in the opinions of these data sources, traders, governments, and researchers can gain insight into trade, policies, and proposals and make better decisions.

The sentiment analysis (SA) of opinions is considered an exciting trend for artificial intelligence on social media. Today, as the economy is gradually transformed into a digital economy, SA is a significant step in several systems automatically, such as stance detection, recommendation systems, decision-making, and fake news detection on social media [15]. SA can be divided into three main levels: documents, sentences, and aspects. Document-level SA aims to determine the sentiment polarity of the entire text without dividing sentences, and sentence-level SA seeks to determine the sentiment polarity of separate sentences. In comparison, aspect-level SA aims at sentiment polarity regarding the aspects of entities appearing in the text. Aspect-level SA is the main objective of this study to classify the sentiments

---

The associate editor coordinating the review of this manuscript and approving it for publication was Nazar Zaki.

[1]https://www.statista.com/

of aspects in a sentence as positive, negative, or neutral. For instance, *"Phone color is not nice, but its style is so modern."* Here, the review focuses on ''color'' and ''style, which are two aspects of the entity ''Phone'', and the ''color'' is negative, while the ''style'' is positive. ALSA is increasingly used in the construction and development of numerous real-life applications. For example, producers can know which parts or aspects of their products are attractive to consumers to retain them and improve them in the opposite case.

In [13], the authors presented that *"Semantic information is compulsory to associate words with sentences according to their contextual information, which could be helpful in the extraction of explicit and implicit aspects."* Therefore, based on contextual information, relationships between different data objects could also be mapped to improve sentiment classification accuracy. e.g., *"I went to an Italian restaurant, and they provided me with eight choices of spaghetti"*, here, ''eight choices of spaghetti'' should be considered as a positive sentiment but normally taken as neutral''. This proves the context surrounding the aspect should be a significant impact on improving the efficacy of ALSA. Various approaches have been proposed for context-based ALSA, such as knowledge-based, machine learning-based, hybrid-based, and most recent GCN-based approaches. Context-based GCNs have achieved relatively good effectiveness in sentiment analysis task, especially ALSA, such as AHGCN-WIN [22], Hier-GCN [2], and GCNSA [4]. From this justification, we absolutely have the basis to believe that the contextualized-based GCN can improve the performance of the previous methods toward ALSA. However, the previous context-based GCN methods for ALSA often use GCNs in the following limitations:

- They use GCNs limited to a few layers (two or three) due to the vanishing gradient, limiting their performance.
- They have not considered helpful information about the hidden context between the words.

Two above limitations motivated us proposing a novel aspect-level SA based on CNN over BERT-GCN to capture more effectively the context surrounding aspects and their sentiments. Why do we choose to combine these models? BERT is an embedding model that uses attention models as transformers to establish relationships between words via an encoder at the input and a decoder at the output. Unlike other embedding models that take the input as one word at a time, BERT can take the entire sentence as input based on transformers. Therefore, BERT can well learn the contexts between words in the sentence [5]. Unlike the basic grammar models that mainly depend on the statistical characteristics, BiLSTM is competent in learning context information by encoding the sentence according to both directions. Therefore, BiLSTM can capture the real meaning hidden between words from their context [3]. GCNs can effectively learn graph representations and have obtained satisfactory results in various applications and tasks, particularly classification tasks [25]. Meanwhile, CNNs have achieved superior perfor-

mance in different tasks and applications, including sentence SA and document classification [14]. Therefore, we have a significant basis to believe that a combination of these models can effectively enhance the performance of ALSA. The proposed method includes the following steps: First, words in sentences are converted into vectors using BERT [5]. Second, contextualized word representations are created based on BiLSTM [7] over word vectors. Third, significant features are extracted and represented using the GCN model with convolutional layers over contextualized word representations. Finally, aspect-level sentiments are classified using a CNN over the feature vectors. Experiments on three benchmark datasets illustrate that our proposed model has improved the performance of the previous context-based GCN methods. The contributions of our proposed method are summarized as follows:

- We handle the disadvantage of limiting the GCN to a few layers by adding convolutional layers of the CNN model after GCN layers.
- We consider further helpful information about the hidden context between the words by combining the BERT and BiLSTM models.

The remaining sections of this paper are organized as follows. Section 2 reviews the state-os-the-art works from which we have inherited the techniques and mechanisms. Section 3 elaborates on the research problem of the proposed method, including research problem definition and research questions. Section 4 presents a mathematical model to answer the research question. Section 5 offers a brief introduction to the datasets and experimental results of the proposed approach versus certain well-known methods. The conclusions and future directions are presented in Section 6.

II. RELATED WORKS

Various deep learning-based ALSA methods have been proposed by capturing the context information of aspects and their sentiments. TD-LSTM [19] constructs aspect-specific representations based on their right and left context and then uses two LSTMs to model them for ALSA with the highest *accuracy* and $F_1 score$ by 78.00% and 68.43%, respectively, for the Restaurant dataset. ATAE-LSTM [21] creates an attention vector by integrating aspect embedding into a hidden state vector. This attention vector is then appended to each word vector to better capture the aspect information. This model obtained the highest *accuracy* by 77.20% for the Restaurant dataset. IAN [12] is based on the attention to aspects and contexts by creating aspects and context representations separately. These representations were then concatenated to predict the sentiment polarity of these aspects. This model obtained the highest *accuracy* by 78.60% for the Restaurant dataset. AOA [8] focuses on learning aspects and sentence representations to capture automatically the crucial parts in sentences. This model obtained the highest *accuracy* and $F_1 score$ by 79.06% for the Restaurant dataset and 70.20% for the Twitter dataset.

Recently, graph neural networks have been widely applied for ALSA due to their adequate representation power, especially GCN, such as [22], [27], and [15]. In addition, to our best knowledge, the best accuracy and F_1score of previous ALSA models are all obtained from the GCNs-based methods. GCN [11] is a graphically structure-based deep learning method and is one category of graph neural network models. Given a graph $G = \{V, E, A\}$ where V and E are sets of nodes and edges, respectively, and A is the adjacency matrix. The GCNs focus well on learning node representations using convolutional network layers to integrate information from neighbors on the graph [23]. A GCN with one convolutional layer can only represent nodes using neighbor information. Otherwise, a GCN with multiple layers can represent nodes using more neighborhood information. We can make GCN layers more depth by extending the convolutional layers of the CNN model presented by Bijari et al in [1]. However, this CNN-GCN directly used CNN over GCN without the contextualized word representations and is not applied for ALSA. Various GCNs-based ALSA method also focus on considering the context information. AHGCN-WIN [22] is a model that focuses on explicit aspects based on contextual representations of graph nodes. Bi-LSTM is first used to capture the context of adjacent words. The GCN model with multiple layers is then used to capture sentiment features of aspect words and the remaining words in the opinion. Finally, a mask layer is applied to catch aspect-specific parts effectively. The AHGCN-WIN obtained the satisfactory performance with the highest *accuracy* and F_1score by 82.02% for the Restaurant dataset and 73% for the Laptop dataset. However, this method has two limitations presented in Section 1. Hier-GCN [2] introduces a model with two parallel graph convolutional layers that can encode both intrarelations among aspects and interrelations between aspects and sentiments. The Hier-GCN obtained the highest F_1score by 74.55% for the Restaurant-16 dataset. However, this method exists the first limitation presented in Section 1 and has not solved incomplete the second limitation. GCNSA [4] presented a combination of GCN and LSTM models, where convolution GCN layers were performed on the text graph to obtain hidden representations of full-text; meanwhile, LSTM was extended by an attention mechanism to capture the certain region information. The GCNSA obtained the highest F_1score by 78.12% for the Restaurant dataset. However, this method exists both limitations presented in Section 1.

From above justifications, we have the basis for proposing this project and the motivation to improve the performance of GCNs-based ALSA methods. We wonder whether using CNN over BERT-GCN improves the performance of ALSA. This becomes a hypothesis that we need to solve and prove in this paper.

III. RESEARCH PROBLEM
A. PROBLEM DEFINITION
Given a finite set of n opinions $O = \{o_1, o_2, \ldots, o_n\}$ regarding the specific entity, for a specific opinion o_i: let a_{ij} be the

j-th aspect of the given entity, c_{ij} be the significant features related to aspect a_{ij}, and sen_{ij} refers to the sentiment of aspect a_{ij}. The objective of this proposal is to construct a CNN over the BERT-GCN model for ALSA. This objective can be formalized by finding a mapping function F as follows: $F : (c_{ij}) \rightarrow \{-1, 0, 1\}$ such that:

$$F(c_{ij}) = \begin{cases} 1, & \text{if } sen_{ij} \text{ is positive,} \\ 0, & \text{if } sen_{ij} \text{ is neutral,} \\ -1, & \text{if } sen_{ij} \text{ is negative} \end{cases} \quad (1)$$

B. RESEARCH QUESTIONS
The main aim of this study is to propose a CNN over the BERT-GCN method for aspect-level SA. Therefore, we attempted to answer the following research questions:
- How can words be converted into contextualized word representations by combining the BERT and BiLSTM models?
- How can we extract and represent significant features using a GCN over contextualized word representations?
- How to build the CNN over the BERT-GCN model using the CNN model over significant features?
- How can the CNN be used over the BERT-GCN model to analyze the sentiment of aspects?

IV. PROPOSED METHOD
In this section, we describe the concept and flow of our proposed CNN over the BERT-GCN model. The proposed method is illustrated in Figure 1.

The CNN over the BERT-GCN model includes the following steps:
- Convert the words of the input sentence into word vectors using the BERT model.
- Extract contextualized word representations using the BiLSTM model over word vectors.
- Extract significant features using the GCN model over contextualized word representations.
- Construct the CNN over the BERT-GCN model using the CNN over significant features.
- Use the CNN over the BERT-GCN model to analyze the aspect-level sentiment of the input sentence.

Given a set of n opinions $O = \{o_1, o_2, \ldots, o_n\}$. For $o \in O$, let $S = \{s_1, s_2, \ldots, s_e\}$ be a set of sentences in opinion o. For $s \in S$, let $s = \{w_1, w_2, \ldots, w_m\}$ be a set of words of one sentence s. The specifies of the approach are presented in the following sections.

A. CREATING WORD VECTORS BASED ON BERT
The BERT model automatically inserts a [CLS] symbol at the beginning of the sentence to indicate the beginning of each sentence and two [SEP] symbols at the beginning and end of the term to indicate the aspect. This implies that the word sequence in the given sentence is converted into the form $\hat{s} = [CLS] + s + [SEP] + a_i + [SEP]$. This sequence of words is then fed into the BERT model to obtain the vector for each
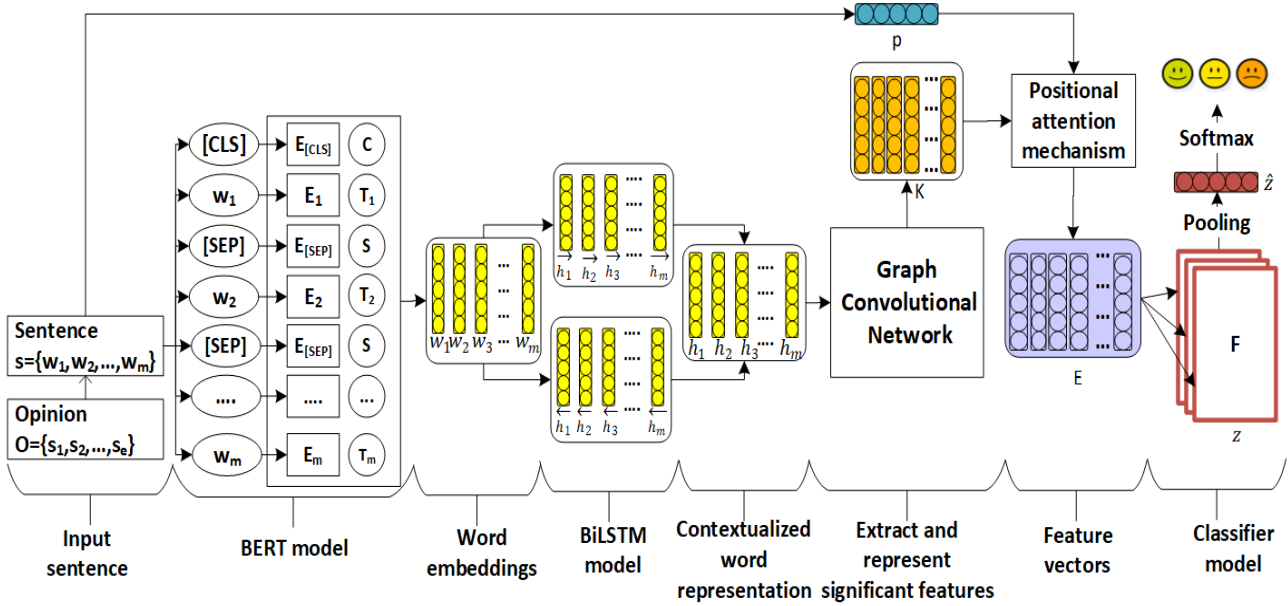
FIGURE 1. Overall framework for proposed method.

word w_i as follows:

$$x_i = BERT(w_i) \in R^{d_w \times 1} \quad (2)$$

where $BERT(w_i)$ is the word vector extracted using pretrained BERT[2] [20], and d_w is the dimension of the word vector. That means from the input sequence s, we obtain the sequence representation as $X = \{x_1, x_2, \ldots, x_m\}$.

B. CREATING CONTEXTUALIZED WORD REPRESENTATIONS

Contextualized word representations convert each word in a sentence into a vector, such that each vector is aggregated from the entire vector of the input sentence. In this study, we employed the BiLSTM model [9] over BERT embeddings for contextualized word representation. BiLSTM can learn the context information and latent meaning of words by reading the input sentence in two directions [3]. Contextualized representations were created according to the following steps:

Input layer: The sequence representation $X = \{x_1, x_2, \ldots, x_m\}$ is the input to the BiLSTM model.

BiLSTM layer: This layer integrates contextual information from the remaining words in the sentence in two directions for the target word [15]. BiLSTM includes two LSTMs, forward ($\overrightarrow{lstm}$) and backward ($\overleftarrow{lstm}$), to encode the sentence from left to right ($\overrightarrow{h}_i$ ($i = [1, m]$)) and otherwise ($\overleftarrow{h}_i$ ($i = [m, 1]$)) as follows:

$$\overrightarrow{h_i} = \overrightarrow{lstm}(W_{x\overrightarrow{h}} x_i + W_{\overrightarrow{h}\overrightarrow{h}} \overrightarrow{h}_{i-1} + b_{\overrightarrow{h}}) \in R^{d_h} \quad (3)$$

$$\overleftarrow{h_i} = \overleftarrow{lstm}(W_{x\overleftarrow{h}} x_i + W_{\overleftarrow{h}\overleftarrow{h}} \overleftarrow{h}_{i+1} + b_{\overleftarrow{h}}) \in R^{d_h} \quad (4)$$

[2]https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-uncased.tar.gz

$$h_i = W_{\overrightarrow{h}h} \overrightarrow{h_i} + W_{\overleftarrow{h}h} \overleftarrow{h_i} + b_h \quad (5)$$

where W is the weight matrix; for example, $W_{x\overleftarrow{h}}$ is the weight matrix between the input vector and forward hidden vectors. h_i is the contextualized representation. h_{i+1} is the next hidden vector of h_i and $h_{m+1} = 0$. h_{i-1} is the previous hidden vector of h_i and $h_0 = 0$.

Output layer: The contextualized word vectors matrix $H = (h_1, h_2, \ldots, h_m) \in R^{m \times d_h}$, where d_h is the size of the vectors in matrix H.

C. EXTRACTING AND REPRESENT SIGNIFICANT FEATURES

We build a sentence graph convolutional network which includes the following steps:

Building the sentence graph: A sentence graph $G = (V, E, A)$ includes a set V of nodes corresponding to m words in the sentence, a set E of edges indicating the dependencies of adjacent node pairs in the syntactic dependency tree,[3] and an adjacency matrix $A \in R^{|V| \times |V|}$ showing the nodes relations, and is defined as follows:

$$A_{ij} = \begin{cases} 1, & if \ v_i, v_j \in V, \ and \ e_{ij} \in E, \\ 1, & if \ v_i = v_j, \\ 0, & otherwise \end{cases} \quad (6)$$

In addition, graph G has a node feature matrix $Q = [H] \in R^{|V| \times d_h}$, where each row H_i represents the contextualized representation of word node $v_i \in V$.

Creating node embeddings: This step is realized as the following sub-steps:

Node representation: Matrices $K \in R^{|V| \times d_h}$ and $A \in R^{|V| \times |V|}$ are fed into a conventional GCN proposed by

[3]https://stanfordnlp.github.io/CoreNLP/

Kipf et al. [11] to create node representations as follows:

$$K = H^{(l)} = \alpha(DH^{(l-1)}W^{(l)} + b^{(l)}) \tag{7}$$

where l is the number of GCN layers, $K \in R^{|V| \times d_k}$; $H^{(0)} = Q$. α is a non-linear activation function as $ReLU$. $W^{(l)} \in R^{d_k \times |V|}$ are the transformation matrices in the l-th layer. $b^{(l)}$ are the biases of GCN layers, respectively.

$$D = M^{-0.5}AM^{-0.5} \tag{8}$$

D is the normalized symmetric matrix of A; M is the degree matrix of A, where:

$$M_{ii} = \sum_j A_{ij} \tag{9}$$

Position-aware transformation: This step reduces noise and bias when processing the GCN. In this study, we use a positional attention mechanism to capture essential parts in the sentence regarding this aspect as follows:

$$E_i = \sum_{i=1}^{|V|} \alpha_i K_i \tag{10}$$

where

$$\alpha_i = \frac{exp(u^A tanh(W^A[K_i \oplus p_i]) + b^A)}{\sum_{k=1}^{|V|} exp(u^A tanh(W^A[K_k \oplus p_k]) + b^A)} \tag{11}$$

where $W^A \in R^{(d_h+d_p) \times (d_h+d_p)}$, $b^A \in R^{d_h \times d_p}$, and $u^A \in R^{d_h \times d_p}$ are learnable matrices. $\oplus$ is a concatenation operator. $p_i \in R^{d_p}$ is the position weight of the i-th word and calculated as

$$p_i = \begin{cases} \frac{|i - start|}{m}, & i \leq start, \\ 0, & start \leq i \leq end, \\ \frac{|i - end|}{m}, & i > end \end{cases} \tag{12}$$

where *start* and *end* are the beginning and ending order of the aspect in the sentence, respectively.

D. BUILDING THE CLASSIFIER MODEL

1) CONVOLUTIONAL LAYER

The convolutional layer decreases the dimension of the node embeddings by creating feature vectors z_i that are determined by sliding a filter $F \in R^{f \times d_h}$ of length f from i to $i + f - 1$ and extracting important information [15] as follows:

$$z_i = ReLU(F \ominus E_{i:i+|V|-1} + b) \tag{13}$$

where $i = [1, |V|]$ is the order of the node representations E; $\ominus$ is the convolution operator; $ReLU$ is an activation function. b is a bias term. Therefore, the feature vectors are created from node representations as follows:

$$z = [z_1, z_2, .., z_{|V|}] \tag{14}$$

2) MAX-POOLING LAYER

The max-pooling layer creates feature vectors of the same size by selecting the maximum number from each vector z_i. The main reason is that the size of the feature vectors z_i depends on the dimensions of both matrices E and F. Therefore, the dimensions of vectors $z_i \in z$ will differ if the sentence length and filter size are different. New feature vectors $\hat{z}$ are defined as follows:

$$\hat{z} = [\hat{z}_1, \hat{z}_2, .., \hat{z}_{|V|}] \tag{15}$$

where $\hat{z}_i = Max(z_i)$.

3) FULLY CONNECTED LAYER

The fully connected layer fine-tunes the characteristics of the previous layers to determine the aspect sentiment as follows:

$$\hat{y} = Softmax(W^E \cdot \hat{z} + b) \tag{16}$$

where $W^E \in R^{l \times |V|}$ and $b \in R^l$ are a weight matrix and a bias of this layer. l is the number of sentiment classes.

E. TRAINING MODEL

Training the CNN over BiLSTM-GCN model is to minimize the cross-entropy loss function as follows:

$$L = -\sum_i^l y_i log(\hat{y}_i) + \lambda \|\theta\|^2 \tag{17}$$

where l is the number of sentiment classes in the training set, y_i is the distribution of the true sentiment of the i-th class, and $\hat{y}_i$ is the distribution of the predicted sentiment of the i-th class. λ represents the L_2 regulation coefficient. θ is the parameter set from the previous layers.

The steps to train the CNN over BERT-GCN model are illustrated as Figure 2:

V. EXPERIMENTAL EVALUATION

A. DATASET AND EXPERIMENTAL SETUP

In this study, to prove the efficacy of the proposed model and ensure a fair comparison of the proposed method with other methods, we used three benchmark datasets: Laptop, Restaurant[4] [16], and Twitter[5] [6]. The restaurant and laptop datasets contain opinions on restaurants and laptops. These opinions were divided into separate sentences that included at least one aspect and sentiment. The tweet dataset consists of opinions regarding celebrities, products, and companies. These opinions were also divided into separate sentences that contained only one aspect and their sentiments. Detailed information on the databases is shown in Table 1.

We implemented the proposed model using pre-trained BERT[6] with dimension of 300 and a learning rate of 5e-05. We initialized all model weights with a uniform distribution. The dimensions of the hidden vectors were set to 204. The learning rate of the Adam optimizer [10] was 0.001. The
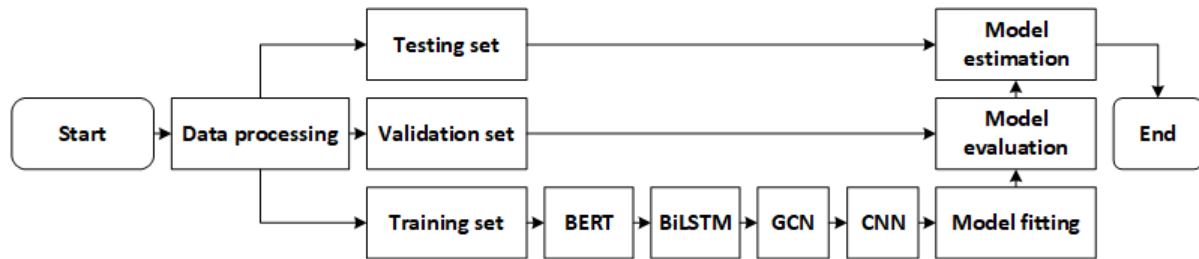
---

[4]http://alt.qcri. org/semeval2014/task4
[5]http://goo.gl/5Enpu7
[6]https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-uncased.tar.gz

FIGURE 2. The training process of the CNN over BERT-GCN model.

TABLE 1. Databases used in experiments.

Data	Restaurant		Laptop		Twitter	
	Train	Test	Train	Test	Train	Test
Positive	2164	727	976	337	1507	172
Neutral	637	196	455	167	3016	336
Negative	807	196	851	128	1528	169
Total	3611	1119	2282	632	6051	677

TABLE 2. Performance of the CNN over BERT-GCN model (%).

Dataset	Restaurant	Laptop	Twitter
Accuracy	85.15	77.76	74.38
F_1	77.38	74.19	73.19

learning rate of the L_2-regularization was 1e-05, and the batch size was 32. Moreover, the number of CNN and GCN layers was set to 1-3, which is the best performing depth for the proposed method. The filter sizes used in the CNN were set up as (5,3,1).

The experimental results were obtained by averaging five runs with random initialization, where F_1 and *accuracy* [17] were adopted as the evaluation metrics. We also executed a paired test on F_1 and *accuracy* to verify whether the improvements achieved by our models over the baselines are significant.

B. BASELINE METHODS

To prove that the performance of our model is better than that of other models, we deployed three different methods, including our proposal and two baselines, on two datasets.

- TD-LSTM [19] represents the aspect-specific using both the aspect's left-side context and the aspect's right-side context and then uses two LSTMs to model them for ALSA.
- ASGCN-DT and ASGCN-DG [24] represents the aspect-specific using both the aspect's left-side context and the aspect's right-side context and then uses two LSTMs to model them for ALSA.
- ASP-BiLSTM, ASP-GCN [18] are aspect-level SA methods by constructing a convolution over a dependency tree model with a BiLSTM to take advantage of sentence feature representations and a GCN that operates directly on the dependency tree of the sentence to enhance the role of embeddings further.
- SDGCN-A, SDGCN-G, and SDGCN-BERT [26] are aspect-level SA methods based on modeling aspect-specific representations from its context words with a bidirectional attention mechanism and position encoding and capturing the sentiment dependencies between different aspects in one sentence with GCN over the attention mechanism.

C. RESULTS AND DISCUSSION

The performance of the sentiment analysis methods over the testing sets of the given datasets is presented in Tables 2, 3, and 4.

From Table 2, we can see that our proposed method achieves the best performance, including the *accuracy* and F_1 score on the restaurant dataset, and the lowest performance on the Twitter dataset. Why is there a difference in the performance when using the same model on different datasets? We can easily see that the results obtained when training a model are better if the training datasets are more extensive. This is correct for the restaurant and laptop datasets. However, in the case of the restaurant and Twitter datasets, although the number of opinions used to train in the Twitter dataset is almost twice as large as that of the restaurant dataset, the results obtained on the restaurant dataset are better. The main reason for this is the quality of the samples in the two datasets. The proposed method can significantly improve this result by carefully preprocessing a dataset before using it.

A performance comparison of the models is presented in Table 3.

It can be seen that our model consistently outperforms all the compared models on the restaurant and Twitter datasets and achieves promising results on the laptop dataset compared with most of the baselines, except for SDGCN-BERT on the laptop dataset. The results demonstrate the effectiveness of our model and the sufficiency of using CNN layers after the GCN layers to effectively capture the essential features and use a combination of BiLSTM and BERT to capture the contextualized representations. Why can the proposed method enhance the *accuracy* and F_1 score of the baseline methods? In this study, the BERT model captured the semantics of the text well, and the BiLSTM model accurately extracted the context of the sentence. Moreover, the GCN model can best represent contextual text. In addition, the GCN model is currently one of the algorithms that achieve good accuracy for sentiment analysis. The results confirmed

TABLE 3. Performance comparison of models (%).

Method	Restaurant		Laptop		Twitter	
	Accuracy	F_1	*Accuracy*	F_1	*Accuracy*	F_1
TD-LSTM	78.00	68.43	71.80	68.46	69.89	66.21
ASGCN-DT	80.86	72.19	74.14	69.24	71.53	69.68
ASGCN-DG	80.77	72.02	75.55	71.05	72.15	70.40
ASP-BiLSTM	80.95	72.38	74.22	69.35	73.66	72.32
ASP-GCN	81.30	73.18	74.53	69.78	70.91	69.07
GCNSA	-	78.12	-	70.53	-	70.09
AHGCN-WIN	82.02	72.67	76.80	73.00	73.12	71.94
SDGCN-A	82.14	73.47	75.39	70.04	-	-
SDGCN-G	82.95	75.79	75.55	71.35	-	-
SDGCN-BERT	83.57	76.47	**81.35**	**78.34**	-	-
CNN over BERT-GCN	**85.25**	**78.76**	78.59	74.76	**74.03**	**72.39**

TABLE 4. Performance of ablation studies (%).

Number of layers		Restaurant		Laptop		Twitter	
CNN	GCN	*Accuracy*	F_1	*Accuracy*	F_1	*Accuracy*	F_1
1	1	**85.25**	**78.76**	**78.59**	**74.76**	74.03	72.39
1	2	85.16	78.15	77.81	74.32	72.67	71.92
1	3	84.89	77.14	77.60	73.12	**77.13**	**74.64**
2	1	83.19	74.67	77.24	73.32	73.86	71.03
2	2	83.72	74.39	78.54	74.03	73.10	71.61
2	3	83.73	76.35	76.62	73.29	73.58	71.65
3	1	84.43	76.52	77.55	74.34	73.61	72.60
3	2	84.80	77.47	77.19	73.29	72.39	70.46
3	3	84.71	78.31	77.65	73.34	69.86	68.65

that the use of a GCN for contextual text representation significantly impacts the accuracy of sentiment analysis methods. Why can the proposed method not obtain a performance comparison with SDGCN-BERT over the laptop dataset? One main reason for this phenomenon is that the laptop dataset is not as sensitive to context information; however, it is too responsive to syntactic information.

The impact of the number of GCN and CNN layers is presented in Table 4. The number of GCN and CNN layers is a significant parameter that needs to be set in our model. To determine the reasonable value for GCN and CNN layers, we experimented with different GCN layers from 1 to 5 and different CNN layers from 1 to 5 for the proposed model. In the experimental process, we observe that if the number of CNN is greater than three or the number of GCN layers is greater than three, overfitting occurs. To our knowledge, the main reason for this phenomenon of the performance decrease may be that as the parameters increase, the model is more challenging to train. Therefore, we only calculated the performance for a maximum of 3 CNN layers and 3 GCN layers. It can be seen that the proposed method obtained the best *accuracy* and F_1 over the restaurant and laptop datasets when number of CNN and GCN layers is 1, too. However, it achieved the best performance over the Twitter dataset when the number of CNN and GCN layers were 1 and 3, respectively. Why did the difference in performance occur between the datasets for the same number of GCN and CNN layers? The main reason is that opinions in the laptop and restaurant datasets are more syntactically written than those in the Twitter dataset. Therefore, the proposed model can best capture the context information with only one GCN layer.

VI. CONCLUSION AND FUTURE WORKS

This study introduced a method to improve the performance of aspect-level sentiment analysis based on a GCN model by using CNN over BERT-GCN to capture contextualized information more effectively. Experimental discussions show that the proposed method significantly improves the performance of sentence-level sentiment analysis over the three benchmark datasets. However, this proposal does not consider all contextual factors, semantic relations, and emotional knowledge simultaneously when building text-representation graphs. In the future, we will focus on building graphs that simultaneously represent contextual factors, semantic relations, and sentimental knowledge to enhance the performance of sentiment analysis methods.

REFERENCES

[1] K. Bijari, H. Zare, E. Kebriaei, and H. Veisi, "Leveraging deep graph-based text representation for sentiment polarity applications," *Expert Syst. Appl.*, vol. 144, Apr. 2020, Art. no. 113090.

[2] H. Cai, Y. Tu, X. Zhou, J. Yu, and R. Xia, "Aspect-category based sentiment analysis with hierarchical graph convolutional network," in *Proc. 28th Int. Conf. Comput. Linguistics*, 2020, pp. 833–843.

[3] R. Cai, B. Qin, Y. Chen, L. Zhang, R. Yang, S. Chen, and W. Wang, "Sentiment analysis about investors and consumers in energy market based on BERT-BiLSTM," *IEEE Access*, vol. 8, pp. 171408–171415, 2020.

[4] J. Chen, H. Hou, Y. Ji, and J. Gao, "Graph convolutional networks with structural attention model for aspect based sentiment analysis," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–7.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[6] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "Adaptive recursive neural network for target-dependent Twitter sentiment classification," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, 2014, pp. 49–54. [Online]. Available: https://aclanthology.org/P14-2009

[7] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, nos. 5–6, pp. 602–610, Jul. 2005.

[8] B. Huang, Y. Ou, and K. M. Carley, "Aspect level sentiment classification with attention-over-attention neural networks," in *Proc. 11th Int. Conf. Social, Cultural, Behav. Modeling (SBP-BRiMS)*. Washington, DC, USA: Springer, 2018, pp. 197–206.

[9] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," 2015, *arXiv:1508.01991*.

[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[11] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[12] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," 2017, *arXiv:1709.00893*.

[13] A. Nazir, Y. Rao, L. Wu, and L. Sun, "Issues and challenges of aspect-based sentiment analysis: A comprehensive survey," *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 845–863, Apr. 2020.

[14] M. Niepert, M. Ahmed, and K. Kutzkov, "Learning convolutional neural networks for graphs," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 2014–2023.

[15] H. T. Phan, N. T. Nguyen, and D. Hwang, "Convolutional attention neural network over graph structures for improving the performance of aspect-level sentiment analysis," *Inf. Sci.*, vol. 589, pp. 416–439, Apr. 2022.

[16] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, 2014, pp. 27–35. [Online]. Available: https://aclanthology.org/S14-2004

[17] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to Information Retrieval*, vol. 39. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[18] K. Sun, R. Zhang, S. Mensah, Y. Mao, and X. Liu, "Aspect-level sentiment analysis via convolution over dependency tree," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, Hong Kong, 2019, pp. 5679–5688. [Online]. Available: https://aclanthology.org/D19-1569

[19] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective LSTMs for target-dependent sentiment classification," 2015, *arXiv:1512.01100*.

[20] I. Turc, M.-W. Chang, K. Lee, and K. Toutanova, "Well-read students learn better: On the importance of pre-training compact models," 2019, *arXiv:1908.08962*.

[21] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 606–615.

[22] K. Xu, H. Zhao, and T. Liu, "Aspect-specific heterogeneous graph convolutional network for aspect-based sentiment classification," *IEEE Access*, vol. 8, pp. 139346–139355, 2020.

[23] L. Yao, C. Mao, and Y. Luo, "Graph convolutional networks for text classification," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 7370–7377.

[24] C. Zhang, Q. Li, and D. Song, "Aspect-based sentiment classification with aspect-specific graph convolutional networks," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 4568–4578. [Online]. Available: https://aclanthology.org/D19-1464

[25] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: A comprehensive review," *Comput. Social Netw.*, vol. 6, no. 1, pp. 1–23, Dec. 2019.

[26] P. Zhaoa, L. Houb, and O. Wua, "Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification," 2019, *arXiv:1906.04501*.

[27] J. Zhou, J. X. Huang, Q. V. Hu, and L. He, "SK-GCN: Modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification," *Knowl.-Based Syst.*, vol. 205, Oct. 2020, Art. no. 106292.

HUYEN TRANG PHAN received the M.S. degree in computer science from the University of Science and Technology–The University of Da Nang, Vietnam, in 2015, and the Ph.D. degree in computer engineering from Yeungnam University, in 2020. She is currently a Research Professor with the Department of Computer Engineering, Yeungnam University, South Korea. She is the author of four journal articles and 13 conference papers. Her research interests include sentiment analysis, decision support systems, machine learning, and deep learning.

NGOC THANH NGUYEN (Senior Member, IEEE) is currently a Full Professor with the Wroclaw University of Science and Technology, and the Head of Information Systems Department, Faculty of Computer Science and Management. He is the author or coauthor of five monographs and more than 350 journal and conference papers. He has given 22 plenary and keynote speeches for international conferences, and more than 40 invited lectures in many countries. His research interests include collective intelligence, knowledge integration methods, inconsistent knowledge processing, and multi-agent systems. He served as a member for the Council of Scientific Excellence of Poland, a member for Committee on Informatics of the Polish Academy of Sciences, an expert for National Center of Research and Development, and European Commission in evaluating research projects in several programs like Marie Sklodowska-Curie Individual Fellowships, FET, and EUREKA. He was a General Chair or a Program Chair of more than 40 international conferences. He also served as the Chair for IEEE SMC Technical Committee on Computational Collective Intelligence. He is also the Honorary Chair of the Scientific Board with Nguyen Tat Thanh University. He has been edited more than 30 special issues in international journals, 52 books, and 35 conference proceedings. He serves as an Editor-in-Chief for *International Journal of Information and Telecommunication* (Taylor&Francis), *Transactions on Computational Collective Intelligence* (Springer), and *Vietnam Journal of Computer Science* (World Scientific). He is also an Associate Editor-in-Chief for several prestigious international journals, among others, *Journal of Intelligent and Fuzzy Systems*, and *Applied Intelligence*. In 2009, he was granted the title Distinguished Scientist of ACM. He was also a Distinguished Visitor of IEEE and a Distinguished Speaker of ACM.

DOSAM HWANG received the Ph.D. degree from Kyoto University, Kyoto, Japan. He is a Full Professor with the Department of Computer Engineering, Yeungnam University, South Korea. He has served as the Head of Yeungnam University, Department of Computer Engineering for five years, from 2005 to 2009. He has also held a position as a Principal Researcher at the Korea Institute of Science and Technology (KIST) and also been a Visiting Professor at (KAIST). He has more than 50 publications. His research interests include natural language processing, ontology, knowledge engineering, information retrieval, and machine translation. He has been steering committee member of ICCCI and ACIIDS, and MISSI international conferences. He was awarded the Prize for Good Conduct from Kyunghee High School, in 1973. He has been the Co-Chair of several international conferences. He has been the Assistant Secretary of ISO/TC37/SC4 for Language Resource Management from 2005 to 2007, and also the Secretary of Korean TC for ISO/TC37/SC4. In 2006, he was the Director of the Korean Society for Cognitive Science (KSCS) and the Korean Information Science Society (KISS). He has been serving as the Society's Director and the Mentor for a knowledge engineering study group, since 2007. In addition to this, he has also participated in several Korean national research projects, such as a project on machine translation system from 1985 to 1990, and the national IT ontology infrastructure and technology development project called "CoreOnto from 2006 to 2009," and "Exobrain, from 2013 to 2014," the project focused on the construction of deep knowledge base and question-answering platform. He is currently an intelligent service integration based on the IoT Big Data as a part of Korea's another principal national research project "BK+ (since 2014)." He has been honored as a Distinguished Researcher of KIST, in 1988, by Korea's Ministry of Science and Technology (MoST).

• • •