IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 13, 2020, accepted March 2, 2020, date of publication March 5, 2020, date of current version March 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2978511

Enhancing BERT Representation With Context-Aware Embedding for Aspect-Based Sentiment Analysis

XINLONG LI[1,2], XINGYU FU[1], GUANGLUAN XU[1], YANG YANG[3], JIUNIU WANG[1,2], LI JIN[1], QING LIU[1], AND TIANYUAN XIANG[1]

[1]Key Laboratory of Network Information System Technology (NIST), Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China
[2]School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100190, China
[3]PLA Unit 31008, Beijing, China

Corresponding author: Xingyu Fu (iecasfxy@163.com)

ABSTRACT Aspect-based sentiment analysis, which aims to predict the sentiment polarities for the given aspects or targets, is a broad-spectrum and challenging research area. Recently, pre-trained models, such as BERT, have been used in aspect-based sentiment analysis. This fine-grained task needs auxiliary information to distinguish each aspect. But the input form of BERT is only a words sequence which can not provide extra contextual information. To address this problem, we introduce a new method named GBCN which uses a gating mechanism with context-aware aspect embeddings to enhance and control the BERT representation for aspect-based sentiment analysis. Firstly, the input texts are fed into BERT and context-aware embedding layer to generate BERT representation and refined context-aware embeddings separately. These refined embeddings contain the most correlated information selected in the context. Then, we employ a gating mechanism to control the propagation of sentiment features from BERT output with context-aware embeddings. The experiments of our model obtain new state-of-the-art results on the SentiHood and SemEval-2014 datasets, achieving a test F1 of 88.0 and 92.9 respectively.

INDEX TERMS Aspect-based sentiment analysis, BERT network, context-aware embedding.

I. INTRODUCTION

Sentiment analysis (SA) is an important task in natural language processing and broadly used in industry. People analyze product sale, service strategies, and lyrical trend by exploring subjective sentiment information in articles and reviews. For example, Amazon and Tmall conduct automatic fine-grained review analysis services and NGP VAN provides candidates in the presidential election with emotional tendency information of social media.

However, the traditional work of predicting the sentence-level sentiment polarity does not satisfy all demands. For example, ''This phone has a large battery capacity, but the camera is not good.'' The review did not show the overall evaluation of the phone, so we only know whether the battery and the camera are suitable for customers. In this case, aspect-based sentiment analysis (ABSA) is employed to perform fine-grained sentiment polarity on specific aspects of interest. ABSA allows producers to gain a granular understanding of the users' requirements for specific aspects of their products or services.

Both SA and ABSA only conduct sentiment mining for single target. However in reality, a comment can refer to different targets. Therefore, targeted aspect-based sentiment analysis (TABSA) was introduced by Saeidi *et al.* [5], which is an extended task of ABSA. It aims to retrieve the corresponding aspects according to the specific targets and simultaneously predict the sentiment polarity of each target-aspect pair. As shown in Table 1. The polarity of target-aspect pairs [*location1, general*] and [*location2, price*] is positive, while the evaluation for [*location2, safety*] is negative.

The early work on (T)ABSA is mainly based on recurrent neural networks (RNNs), such as LSTM and GRU, to generate context, aspect and target representation.

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

TABLE 1. Example of TABSA task. Entity names are masked by location1 and location2.

example	I live comfortably in **location1**, even if **location2** rent is cheaper. Because public order is not good in **location2**.	
Target	Aspect	Sentiment
location1	general	positive
location2	price	positive
location2	safety	negative

They also use attention mechanism to transform context representation according to target and aspect representation [23], [1]. However, these studies have two problems to restrict their performance. The first problem is compared with transformer, RNNs are hard to parallelize and capture interactive semantics over longer time scales. Another problem is that the aspect and sentiment information have to encode context representation by measuring weight across all feature dimensions. So attention mechanism is time-consuming.

In language modeling [13]–[16], Gated Tanh Units (GTU) and Gated Linear Units (GLU) have shown effectiveness of gating mechanisms. Xue and Li [2] introduced a new kind of gating unit named Gated Tanh-ReLU Units (GTRU). The GTRU can select the sentiment features according to the given aspect or entity and the architecture of GTRU is much simpler than attention mechanism.

The pre-trained model, especially BERT [12], achieves state-of-the-art results in multiple NLP tasks, including text classification, reading comprehension, and named entity recognition in recent years. The main structure of BERT is multi-layer transformers (Vaswani *et al.* [6]), which can be trained in parallel. Sun *et al.* [4] constructed some simple auxiliary sentences for the context, such as *What do you think of the* ***price*** *of* ***location-1***. And then feed the context and the auxiliary sentence into BERT. The auxiliary sentences provide additional information about the target-aspect pairs for the (T)ABSA task. However, these auxiliary sentences cannot provide contextual and semantic information, because the only useful part in the auxiliary sentences is target and aspect words.

In this paper, we propose a new method to enhance BERT via refined novel embeddings for (T)ABSA. First, the input sentences are fed into BERT and context-aware embedding layer to generate BERT representation and refined context-aware embeddings separately. The context-aware embedding layer adjusts the representation of target and aspect (or only aspect vectors in ABSA) to make them contain features extracted from context. Then novel gating units are employed to dynamically control the flow of sentiment information between these context-aware vectors and the output of BERT encoding layer. We evaluate the effectiveness of our model on the SentiHood and SemEval-2014 Task 4 datasets. Compared to other baseline models, we achieve state-of-the-art performance on both aspect extraction and sentiment classification.

The main contributions of this work are presented as follows:

- We propose a new method for (T)ABSA without using any external knowledge. It investigates the potential of enhancing BERT via pre-trained context-aware target and aspect embeddings. These context-aware embeddings contain relevant information between target-aspect pair and context.
- We employ a new gating mechanism to control the path through which the sentiment information from context-aware embedding layer and BERT output towards the output layer.
- Experiment results show that our proposed method can substantially improve the performance of aspect detection and sentiment classification, where the F1 score achieves 88.0 on SentiHood dataset and 92.9 on SemEval-2014 dataset.

II. RELATED WORK

We present the relevant studies into following two categories, including aspect-based sentiment analysis and targeted aspect-based sentiment analysis.

A. ASPECT-BASED SENTIMENT ANALYSIS

In previous work, machine learning based methods [26], [25] were prevalent for aspect-based sentiment analysis. They mainly focus on extracting hand-craft features such as lexical, syntactic and semantic features [27]. And Vo and Zhang [28] proposed sentiment-specific word embedding for sentiment analysis. These studies which based on feature engineering require a professional foundation in linguistics and easily reach the performance bottleneck.

Neural networks, which have acquired high popularity on ABSA, can extract features dynamically due to their capability of encoding original features as continuous and low-dimensional vectors without feature engineering. Recurrent neural networks, such as LSTM and GRU, were used to model the context, and concatenated them as the representation for prediction. IAN [23] modeled both sentences and aspects by two LSTM networks separately. They calculated attentions by a pooling operation between sentences and aspects representation. RAM [30] employed multiple attentions with LSTM which strengthens the expressive power for handling more complications. Along these lines, Cabasc [31] used GRU with a novel content attention mechanism to deal with the syntactically structures of complex sentence.

B. TARGETED ASPECT-BASED SENTIMENT ANALYSIS

Neural networks have achieved high popularity on TABSA. Recursive networks such as Recursive deep model [9] and Tree-LSTM [22] used natural language syntactic structures to combine words to phrases naturally. Dong *et al.* [29] proposed to conduct semantic compositions on tree structures.

Attention-based LSTM with Aspect Embedding (ATAE-LSTM) [10] focused on the sentiment words in the text which were relatively correlative to the target or entity. The memory network, which is derived initially from reading comprehension task, has also proven to be successful for (T)ABSA. Liu *et al.* [33] employed a delayed memory to track and update the states by occupying external storage. Some external knowledge has been added into neural networks to enhance semantic representation. Lei *et al.* [34] proposed MEAN which integrates the information of the sentiment, negation, and intensity words into the deep neural network via attention mechanisms for sentiment prediction.

III. TASK DEFINITION

For TABSA, we assume a text sentence C consisting of a sequence of words: $\{c_1, \ldots, c_h\}$, and some of the words $\{c_{l1}, \ldots, c_{ln}\}$ are pre-identified targets $\{t_1, \ldots, t_n\}$. Each target t may correspond to several aspects a. The goal of TABSA can be regarded as a fine-grained sentiment expression as a tuple (t, a, p), where p refers to the polarity which is associated with a target-aspect pair (t, a). The TABSA task aims to detect the aspect $a \in A$ according to the given target and predict the sentiment polarity $p \in \{Positive, Negative, None\}$.

For ABSA, the aspects are considered only to substitute the target-aspect pairs (t, a). The rest settings are in stock with TABSA which aim to learn subtask 3(Aspect Category Detection) and subtask 4 (Aspect Category Polarity) of SemEval-2014 Task 4 at the same time.

IV. METHODOLOGY

In this section, we will introduce the proposed method for (T)ABSA. For simplicity, we describe GBCN model for TABSA in detail and only emphasize the main difference between TABSA and ABSA. The GBNCE model is a hierarchical multi-stage process and mainly consists of four components:

(I) **BERT Encoding Layer** which maps each word to a vector space using multi-layer bidirectional Transformers.

(II) **context-aware embedding layer** which generates refined target and aspect embeddings.

(III) **gating layer** which controls the sentiment information from BERT output flow according to the context-aware target and aspect embeddings.

(IV) **output layer** which predicts sentiment polarity toward corresponding target-aspect pair.

All of the framework can be demonstrated in Figure 1. In the following, we will introduce them respectively.

A. BERT ENCODING LAYER

This part of the model uses BERT encoder to model text sentences. BERT stands for Bidirectional Encoder Representations from Transformer and pre-trained from large-scale corpora. We finetune the pre-trained BERT model on (T)ABSA task.

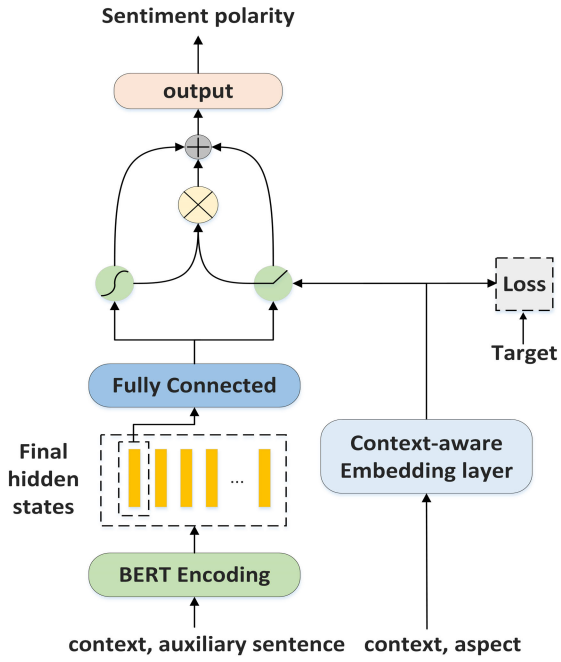


FIGURE 1. Overall architecture of GBCN.

TABLE 2. Example of constructing auxiliary sentences.

Target-aspect pair	(location2, price)
QA-M	what do you think of the price of location - 2?
NLI-M	location - 2 - price

1) INPUT REPRESENTATION

Given a context $C = \{c_i\}_{i=1}^{h}$, where h is the length of the sentence. Then we construct auxiliary sentence A for the context. The methods we construct the auxiliary sentence is the same as Sun *et al.* [4], which are called QA-M, NLI-M, QA-B, NLI-B respectively. Because the representations of auxiliary sentences have no difference mostly, we only use the first two. For each set of target-aspect pair in the context, the auxiliary sentences A we generate are showed in Table 2.

For ABSA, the target-aspect pair becomes aspect only, and the auxiliary sentence has to be changed correspondingly. Finally, we construct the input sequence S with both C and A,

$$S = [< CLS >, C, < SEP >, A, < SEP >], \tag{1}$$

where $< CLS >$ is a unique token of each sequence for classification, $< SEP >$ is the token separating C and A. For each given token $s_i \subset S$, its input representation is constructed as:

$$h_i^0 = s_i^{tok} + s_i^{pos} + s_i^{seg}, \tag{2}$$

where $s_i^{tok}, s_i^{pos}, s_i^{seg}$ are the corresponding token, segment, and position embeddings for s_i.

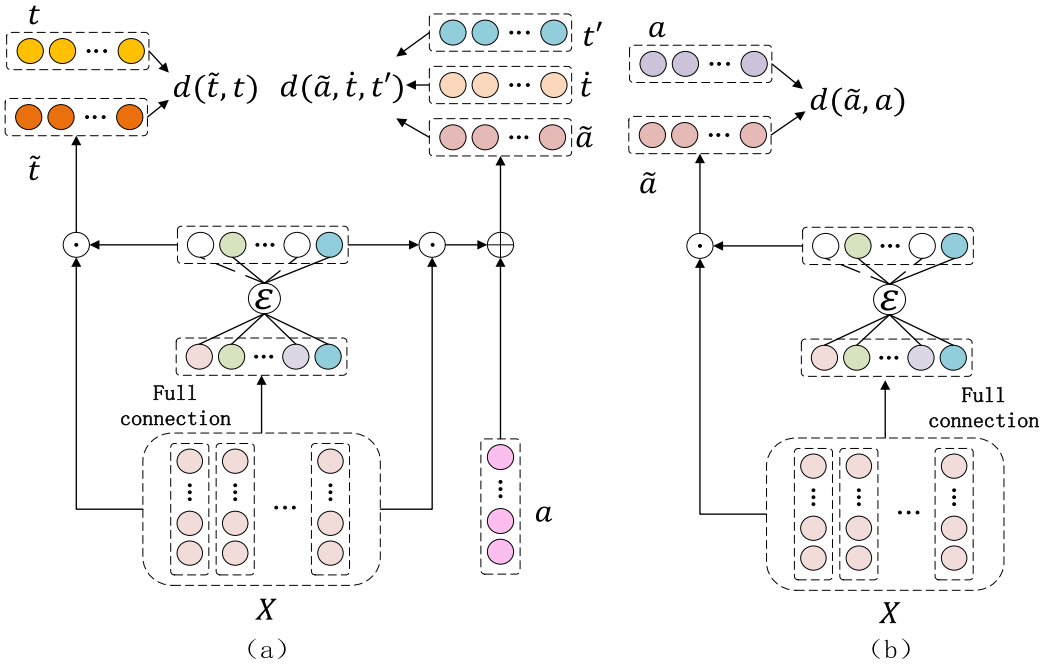


FIGURE 2. The framework of context-aware embedding layer. (a) context-aware embedding layer for TABSA, (b) context-aware embedding layer for ABSA, ⊙ is element-wise product, ⊕ is vector addition, ε is step function.

2) FINE-TUNING PROCEDURE

Fine-tuning BERT is straightforward. The input representation described above are then fed into L successive Transformer encoder blocks,

$$h_i^{\ell} = \text{Transformer}(h_i^{\ell-1}), \ell = 1, 2, \cdots, L, \tag{3}$$

We denote the final hidden states (i.e., the output of the transformer) as $\{h_i^L\}_{i=1}^N \in \mathbb{R}^{d_1}$, where $N = m + 1$. To obtain a fixed-dimensional pooled representation of the input sequence, we use h_0^L which is the first token $< CLS >$ of final hidden state as the part of input for the next layer.

B. CONTEXT-AWARE EMBEDDING LAYER

Each word of the given context C in this layer can be represented as a d_2-dimensional embedding $x_i \in \mathbb{R}^{d_2}$, including the embedding of target $\mathbf{t} \in \mathbb{R}^{d_2}$ via random initialization and the embedding of aspect $\mathbf{a} \in \mathbb{R}^{d_2}$ which is an average of its constituting word embeddings or single word embedding. So the sequence is represented as an embedding matrix $\mathbf{X} \in \mathbb{R}^{m*d_2}$, where d_2 is the dimension of embedding. The method to get the context-aware embedding is inspired by Liang *et al.* [3]. The main idea is constructing a sparse coefficient vector to select highly correlated words from the context, and then adjust the representations of target and aspect to make them more valuable. The framework of our proposed method is demonstrated in Figure 2.

1) CONTEXT-AWARE EMBEDDING LAYER FOR TABSA

To extract the correlation between target and context for TABSA, the target representations are mainly reconstructed according to the highly correlated words in the context. So the target representation $\tilde{t}$ is computed as:

$$u = f(W_c \boldsymbol{X} + b_c) \tag{4}$$

$$\varepsilon(u_i) = \begin{cases} u_i, & u_i \geq mean(u) \\ 0, & u_i < mean(u) \end{cases} \tag{5}$$

$$u' = \varepsilon(u) \tag{6}$$

$$\tilde{t} = \boldsymbol{X}u' \tag{7}$$

where f is sigmoid function, $W_c \in \mathbb{R}^{d_2}$ and $b_c \in \mathbb{R}^m$ denote the weight matrix and bias respectively. The $mean(u)$ is arithmetic mean of $u = [u_1, u_2, \cdots, u_i, \cdots, u_m]$. The step function ε is designed to extra the correlation of words in the context. The irrelevant words are set to zero.

The context information can reflect the aspect representation. So the Xu' needs to be fed into $\tilde{a}$, which is the context-aware embedding of aspect. The aspect embedding a contain crucial semantic information. To this end, we add aspect itself into $\tilde{a}$. i.e.:

$$\tilde{a} = a + \alpha \boldsymbol{X}u' \tag{8}$$

where α is a hyper-parameter to control the influence between aspect and the context. We use two kinds of similarity measurement as the objective function between target embedding t and context-aware target representation $\tilde{t}$, including squared Euclidean and Pearson correlation coefficient. One is to calculate the distance between t and $\tilde{t}$, and the other is to measure the linear correlation. Moreover, in order to control

the sparseness of vector u', u'_i is added in the end.

$$d(\tilde{t}, t) = \sum_{i=1}^{m}(\sum_{j=1}^{d_2}(\tilde{t}_i^j - t_i^j)^2 + \text{Pear}(\tilde{t}_i^j, t_i^j) + \lambda u'_i) \quad (9)$$

The idea we construct the objective function $d(\tilde{a}, \dot{t}, t')$ of aspect representation is that each aspect should be moved closer to the homologous target and further away from the irrelevant one. The objective function is also divided into two parts: squared Euclidean and Pearson correlation coefficient.

$$\text{Dis}_i^j = (\tilde{a}_i^j - \dot{t}_i^j)^2 - \beta(\tilde{a}_i^j - t'^j_i)^2 \quad (10)$$

$$\text{Cor}_i^j = \text{Pear}(\tilde{a}_i^j, t'^j_i) - \gamma\text{Pear}(\tilde{a}_i^j, \dot{t}_i^j) \quad (11)$$

$$d(\tilde{a}, \dot{t}, t') = \sum_{i=1}^{m}(\sum_{j=1}^{d_2}\text{Dis}_j^i + \text{Cor}_j^i + \lambda u'_i) \quad (12)$$

where $\dot{t}$ is the homologous target and t' is the irrelevant one. β and γ are parameters that controls the distance from the irrelevant target.

2) CONTEXT-AWARE EMBEDDING LAYER FOR ABSA

For ABSA, we only consider how to construct context-aware aspect embedding from the related sentence. So the aspect representation $\tilde{a}$ is computed as:

$$\tilde{a} = Xu' \quad (13)$$

To make the reconstructed representation $\tilde{a}$ have high correlation with source aspect embedding a, the objection for ABSA is defined as:

$$d(\tilde{a}, a) = \sum_{i=1}^{n}(\sum_{j=1}^{m}(\tilde{a}_i^j - a_i^j)^2 + \text{Pear}(\tilde{a}_i^j, a_i^j) + \lambda u'_i) \quad (14)$$

C. GATING LAYER

This layer is designed to integrate the representation of BERT with context-based knowledge further. It takes representations $\{h_i^L\}$ from the BERT encoding layer as input and enriches them with relevant refined target-aspect embeddings, which makes the representations focus on the part associated with each target-aspect pair in the context more effectively.

As previously stated, we take the first taken of BERT output $\{h_1^L\} \in \mathbb{R}^{d_1}$, context-aware target representation $\tilde{t} \in \mathbb{R}^{d_2}$ and aspect representation $\tilde{a} \in \mathbb{R}^{d_2}$ as input. Then we employ the gating mechanism GTRU to select the most relevant context adaptively. The output of GTRU is connected to a fully connected layer. Explicitly, we compute the features o as:

$$c = \tanh(Wh_1^L + b) \quad (15)$$

$$e = \text{ReLU}(c + \tilde{t}T_a + \tilde{a}A_a + b_a) \quad (16)$$

$$s = \tanh(c + b_s) \quad (17)$$

$$o = e \odot s + e + s \quad (18)$$

TABLE 3. Statistics of the datasets.

Dataset	Positive		Neural		Negative	
	Train	Test	Train	Test	Train	Test
SentiHood	1626	834	-	-	810	406
SemEval-2014	2164	728	637	196	807	196

where $\tilde{t}$ and $\tilde{a}$ are the pre-trained embedding vectors of the given target and aspect category. The ReLU gate receives the refined target and aspect information to control the propagation of sentiment features. The outputs of two gates are element-wise multiplied for the output layer. To enrich the representation, e and s are add to the output layer too.

For ABSA, the information of the targets does not exist. So we define e as follow without $\tilde{t}$:

$$e = \text{ReLU}(c + \tilde{a}A_a + b_a) \quad (19)$$

D. OUTPUT LAYER

To ease overfitting, a dropout layer is used for the output of gating layer. The final fully connected layer with softmax function uses the vector o to predict the sentiment polarity $\hat{y}$. The model is trained by minimizing the cross-entropy loss between the ground-truth y and the predicted value $\hat{y}$ for all data samples.

$$\mathcal{L} = -\sum_i \sum_j y_i^j \log \hat{y}_i^j \quad (20)$$

where i is the index of a data sample, j is the index of a sentiment class.

V. EXPERIMENTS

A. DATASETS

We evaluate our method on the SentiHood [5]. The entire dataset consists of 5215 sentences with 3862 sentences containing a single target and the remainder containing two targets. Location1 and location2 cover all the location target names. Following [5], we only consider the four most frequent aspects: general, price, safety and transit-location. Ultimately, We need to retrieval aspect through the target in each sentence and predict sentiment polarity for each target-aspect pair.

We also evaluate our model for ABSA on SemEval-2014 Task 4 [20] dataset, which consists of 4033 sentences for the training dataset and 800 for testing dataset about restaurant. The difference from the SentiHood is that SemEval-2014 only has aspects instead of target-aspect pairs. Following [18], we only consider the top five aspects: ambience, anecdotes, food, price and service. Subtask 3 (Aspect Category Detection) and subtask 4 (Aspect Category Polarity) are jointly evaluated.

The statistics of the SentiHood and SemEval datasets are shown in Table 3.

TABLE 4. Performance on SentiHood Dataset. We boldface the score with the best performance across all models. we use the results reported in Ma *et al.* [32], Liu *et al.* [31], and Sun *et al.* [4]. "-" means not reported."GBGN" means GBCN without context-aware embedding.

Models	Aspect			Sentiment	
	Acc	F1	AUC	Acc	AUC
LR [5]	-	39.3	92.4	87.5	90.5
LSTM+TA+SA [32]	66.4	76.7	-	86.8	-
SenticLSTM [32]	67.4	78.2	-	89.3	-
Dmu-Entnet [33]	73.5	78.5	94.4	91.0	94.8
BERT [4]	73.7	81.0	96.4	85.5	84.2
BERT-pair-QA-M [4]	79.4	86.4	97.0	93.6	96.4
BERT-pair-QA-M with GBGN	79.5	86.5	96.8	93.9	97.1
BERT-pair-QA-M with GBCN	**81.9**	87.6	**97.3**	**94.5**	**97.5**
BERT-pair-NLI-M [4]	78.3	87.0	97.5	92.1	96.5
BERT-pair-NLI-M with GBGN	78.6	86.9	97.3	92.5	96.9
BERT-pair-NLI-M with GBCN	81.3	**88.0**	97.2	93.8	97.2

B. EXPERIMENT SETTING

In context-aware embedding layer, all word embeddings are initialized by Glove [17] which are pre-trained on unlabeled data of 840 billion tokens except target embeddings. Location1 and location2 are randomly initialized in the model. These groups of embeddings are 300-dimension. $W, W_c, S_a, b, b_a, b_s, b_c, A_a, T_a$ are randomly initialized. The hyperparameters of α, β, γ are set to 1, 1 and 0.5 respectively.

Throughout the training process, the entire model is divided into two parts: context-aware embedding layer and the rest of the model named gating BERT networks. When the number of 0 in the sparse coefficient vector u' does not exceed 4, it is considered that the context-aware embedding layer has converged and we fix the parameters of it.

In the beginning of the training period, each context C with one or multiple lists of labels (t, a, p) are Given. Context is provided to context-aware embedding network first to general refined target and aspect representations. Then the context and embeddings are feed into the gating BERT network. Follow Sun *et al.* [4], the pre-trained language model we used is uncased BERT-base model published by Google. The number of Transformer blocks is 12, the hidden layer size is 768, the number of self-attention heads is 12, and the total number of parameters for the pre-trained model is 110M. The hyperparameters are the same as Sun *et al.* [4] set for fairness.

C. EXP-I: TABSA

To demonstrate the effectiveness of our model, we compare GBCN against the following methods.

- **LR** [5] uses a logistic regression classifier with n-gram and pos-tag features.
- **LSTM+TA+SA** [32] uses a biLSTM model to general final representation with both target-level and sentence-level attention mechanisms.
- **SenticLSTM** [32] is similar to LSTM + SA + TA, but uses SenticLSTM to extract external features introduced from SenticNet [7].
- **Dmu-Entnet** [33] uses "memory chains" which is inspired from reading comprehension tasks. The core of this model is a delayed memory update mechanism to track entities.
- **BERT with auxiliary sentence (BERT-pair-QA-M, BERT-pair-NLI-M)** [4] uses BERT as base model and constructs auxiliary sentences for BERT input.

Following Liu *et al.* [33], we use strict accuracy, Macro-F1 as the evaluation indices in aspect detection, and we also report AUC. When evaluating the sentiment classification, we use accuracy and macro-average AUC.

1) RESULTS

The result of our experimental on SentiHood is shown in Table 4. The classifiers based on our proposed methods (GBCN, BERT-pair-QA-M with GBCN, BERT-pair-NLI-M with GBCN) achieve better performance than competitor models on both aspect detection and sentiment classification. We can see that BERT-pair with GBCN goes beyond the previous best performing models on all evaluation indices. Compared with BERT-pair-QA-M [4], which is the best performing model we have mentioned, our model (GBCN) significantly improves the performance of aspect detection (by 2.5% in strict accuracy, 1.2% in macro-average F1 and 0.3% in AUC) and sentiment classification (by 0.9% in strict accuracy and 1.1% in AUC) on SentiHood. This is because context-aware targets and aspects representations provide additional structural and semantic information than the auxiliary sentences. When using GBGN on the SentiHood, the consequent of BERT improves not as significant as using GBCN. Compared with BERT-pair, some evaluation indices of BERT-pair with GBGN even lower. This phenomenon proves the importance of context-aware embeddings component.

TABLE 5. Test set results for Semeval-2014 Task 4 Subtask 3: Aspect category detection. We use the results reported in XRCE [18], NRC-Canada [19] and BERT with auxiliary sentences [4].

Models	P	R	F1
XRCE	83.23	81.37	82.29
NRC-Canada	91.04	86.24	88.58
BERT	92.78	89.07	90.89
BERT-pair-QA-M	92.87	90.24	91.54
BERT-pair-QA-M-GBCN	93.59	91.32	92.44
BERT-pair-NLI-M	93.15	90.24	91.67
BERT-pair-NLI-M-GBCN	**94.26**	**91.55**	**92.89**

TABLE 6. Test set accuracy(%) for Semeval-2014 Task 4 Subtask 4: Aspect category polarity. We use the results reported in XRCE [18], NRC-Canada [19], ATAE-LSTM [10] and BERT with auxiliary sentences [4]. "-" means not reported.

Models	4-way	3-way	Binary
XRCE	78.1	-	-
NRC-Canada	82.9	-	-
LSTM	-	82.0	88.3
ATAT-LSTM	-	84.0	89.9
BERT	83.7	86.9	93.3
BERT-pair-QA-M	85.2	89.3	95.4
BERT-pair-QA-M-GBCN	**86.4**	**90.8**	**96.5**
BERT-pair-NLI-M	85.1	88.7	94.4
BERT-pair-NLI-M-GBCN	86.0	89.8	96.2

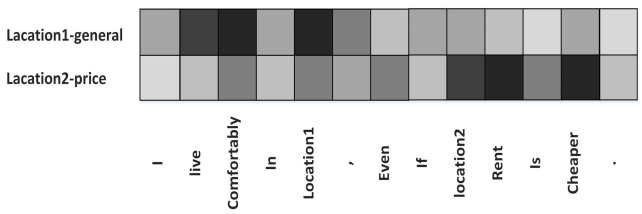


FIGURE 3. The output value of the gating units *o* for TABSA.

D. EXP-II: ABSA

We evaluate our methods for SemEval-2014 Task on two baselines, include two best-performing systems in Pontiki *et al.* [20], ATAE-LSEM [10] and BERT with auxiliary sentences [4]. Following Pontiki *et al.* [20], aspect category detection and polarity are evaluated by Micro-F1 and accuracy respectively.

1) RESULTS

The results of our experimental on SemEval-2014 is shown in Table 5 and Table 6. All models based on BERT achieve better performance. Comparing these BERT-based models, the increase of GBCN on both subtasks is relatively significant. The BERT-pair-NLI-B with GBCN achieves the best performance for aspect category detection. For aspect category polarity, BERT-pair-QA-B with GBCN performs best on all 4-way, 3-way, and binary settings.

VI. DISCUSSION

To better understand what the model has learned, we visualise the output value of the gating units *o* in Figure 3, where colour intensity indicates how much information is gained for each word. Observe that, it shows that the GBCN would control the magnitude of the output.

The overall results show that our model can substantially improve the performance of aspect detection and sentiment classification by controlling the propagation of sentiment features from BERT output and refined embeddings with the gating mechanism. This indicates that the refined representation is more learnable, and the gating mechanism can extract the interdependence between aspect and the corresponding target in the context. We have tried LSTM with attention layer instead of the gated mechanism, but the performance is not better, and training speed is slower.

Why can GBCN achieve a better result than the state-of-the-art model before? On the one hand, we refine the target and aspect embeddings by constructing a sparse coefficient vector. This makes these new embeddings contain context structural and semantic information. It is clearly seen when comparing the increase of BERT-pair. On the other hand, the ReLU activation function in GTRU can selectively output the sentiment features according to the given refined embeddings. This is because that GTRU use ReLU instead of the sigmoid function in GTU and GLU, which has the upper bound $+1$ and may not be able to distill sentiment features effectively. The experience also proves that our model can effectively extract features at two sources: one is from BERT representation, the other is context-aware embedding layer based on Glove embeddings.

VII. CONCLUSION

In this paper, we employ GBCN which uses a gating mechanism with context-aware embeddings to contral the BERT representation for (T)ABSA task. GBCN is designed to construct context-aware target and aspect vectors effectively, and dynamically control the flow of sentiment information through GTRU. We prove the performance is obviously improved compared with other neural models by extensive experiments on SentiHood and Semeval-2014 dataset. How to promote our model to other subtasks of sentiment analysis would be our future work.

REFERENCES

[1] X. Li, L. Bing, W. Lam, and B. Shi, "Transformation networks for target-oriented sentiment classification," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, Melbourne, VIC, Australia, 2018, pp. 946–956.
[2] W. Xue and T. Li, "Aspect based sentiment analysis with gated convolutional networks," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, Melbourne, VIC, Australia, 2018, pp. 2514–2523.
[3] B. Liang, J. Du, R. Xu, B. Li, and H. Huang, "Context-aware embedding for targeted aspect-based sentiment analysis," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, Florence, Italy, 2019, pp. 4678–4683.
[4] C. Sun, L. Y. Huang, and X. P. Qiu, "Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence," in *Proc. NAACL*, Minneapolis, MN, USA, 2019, pp. 380–385.
[5] M. Saeidi, G. Bouchard, M. Liakata, and S. Riedel, "SentiHood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods," in *Proc. COLING*, Osaka, Japan, 2016, pp. 1546–1556.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and L. Kaiser, "Attention is all you need," in *Proc. NIPS*, Long Beach, CA, USA, 2017, pp. 260–351.

[7] E. Cambria, S. Poria, R. Bajpai, and B. Schuller, "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives," in *Proc. COLING*, Osaka, Japan, 2016, pp. 2666–2677.

[8] F. Liu, T. Cohn, and T. Baldwin, "Recurrent entity networks with delayed memory update for targeted aspect-based sentiment analysis," in *Proc. NAACL*, New Orleans, LA, USA, 2016, pp. 278–283.

[9] R. Socher, A. Perelygin, Y. J. Wu, J. Chuang, C. D. Manning, Y. A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proc. EMNLP*, Seattle, WA, USA, 2013, pp. 611–619.

[10] 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.*, Austin, TX, USA, 2016, pp. 606–615.

[11] Z. Lei, Y. Yang, M. Yang, and Y. Liu, "A multi-sentiment-resource enhanced attention network for sentiment classification," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, Melbourne, VIC, Australia, 2018, pp. 758–763.

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*. [Online]. Available: http://arxiv.org/abs/1810.04805

[13] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," in *Proc. ICML*, Sydney, NSW, Australia, 2017, pp. 933–941.

[14] N. Kalchbrenner, L. Espeholt, K. Simonyan, A. V. D. Oord, A. Graves, and K. Kavukcuoglu, "Neural machine translation in inear time," *CoRR*, 2016.

[15] A. V. D. Oord, N. Kalchbrenner, L. Espeholt, K. Kavukcuoglu, O. Vinyals, and A. Graves, "Conditional image generation with PixelCNN decoders," in *Proc. NIPS*, Barcelona, Spain, 2016, pp. 4790–4798.

[16] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," in *Proc. ICML*, Sydney, NSW, Australia, 2017, pp. 1243–1252.

[17] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Doha, Qatar, 2014, pp. 1532–1543.

[18] C. Brun, D. N. Popa, and C. Roux, "XRCE: Hybrid classification for aspect-based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, Dublin, Ireland, 2014, pp. 838–842.

[19] S. Kiritchenko, X. Zhu, C. Cherry, and S. Mohammad, "NRC-Canada-2014: Detecting aspects and sentiment in customer reviews," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, Dublin, Ireland, 2014, pp. 437–442.

[20] 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)*, Dublin, Ireland, 2014, pp. 27–35.

[21] B. Wang, M. Liakata, A. Zubiaga, and R. Procter, "TDParse: Multi-target-specific sentiment recognition on Twitter," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, Valencia, Spain, 2017, pp. 483–493.

[22] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *Comput. Sci.*, vol. 1, no. 1, pp. 232–241, May 2015.

[23] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Melbourne, VIC, Australia, Aug. 2017, pp. 421–423.

[24] J. J. Lin, W. J. Mao, and D. J. Zeng, "Semi-supervised polarity lexicon induction," in *Proc. EACL*, Athens, Greece, 2009, pp. 1256–1266.

[25] J. Wagner, P. Arora, S. Cortes, U. Barman, D. Bogdanova, J. Foster, and L. Tounsi, "DCU: Aspect-based polarity classification for SemEval task 4," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, Dublin, Ireland, 2014, pp. 223–229.

[26] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," in *Proc. EMNLP*, Philadelphia, PA, USA, 2002, pp. 269–278.

[27] D. Rao and D. Ravichandran, "Semi-supervised polarity lexicon induction," in *Proc. 12th Conf. Eur. Chapter Assoc. Comput. Linguistics (EACL)*, Athens, Greece, 2009, pp. 675–682.

[28] D. T. Vo and Y. Zhang, "Target-dependent Twitter sentiment classification with rich automatic features," in *Proc. IJCAI*, Buenos Aires, Argentina, 2015, pp. 25–31.

[29] 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*, Baltimore, MD, USA, 2014, pp. 49–54.

[30] P. Chen, Z. Q. Sun, L. D. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *Proc. ACL*, Vancouver, BC, Canada, 2017, pp. 452–461.

[31] Q. Liu, H. Zhang, Y. Zeng, Z. Huang, and Z. Wu, "Content attention model for aspect based sentiment analysis," in *Proc. World Wide Web Conf. World Wide Web (WWW)*, Buenos Aires, Argentina, 2018, pp. 412–423.

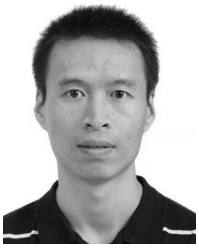
[32] Y. K. Ma, H. Y. Peng, and E. Cambria, "Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM," in *Proc. AAAI*, New Orleans, LA, USA, 2018, pp. 5876–5883.

[33] D. Sorokin and I. Gurevych, "Modeling semantics with gated graph neural networks for knowledge base question answering," 2018, *arXiv:1808.04126*. [Online]. Available: http://arxiv.org/abs/1808.04126

XINLONG LI received the B.S. degree in electronic information engineering from Tianjin University, Tianjin, China, in 2018. He is currently pursuing the M.A.Sc. degree with the Institute of Electronics, Chinese Academy of Sciences, Beijing, China. His research interests include deep learning, natural language processing, sentiment analysis, and question answering.

XINGYU FU received the B.S. degree from the Ocean University of China, Qingdao, China, in 2007, and the Ph.D. degree from the Institute of Electronics, Chinese Academy of Sciences, Beijing, China, in 2012.

He is currently an Associate Professor with the Institute of Electronics, Chinese Academy of Sciences. His research interests include geospatial data intelligent processing, geospatial multisource heterogeneous data comprehensive retrieval technology, and geospatial data cognitive computing technology.



GUANGLUAN XU received the B.S. degree from Beijing Information Science and Technology University, Beijing, China, in 2000, and the Ph.D. degree from the Institute of Electronics, Chinese Academy of Sciences, Beijing, in 2005.

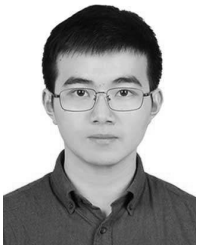
He is currently a Professor with the Institute of Electronics, Chinese Academy of Sciences. His research interests include remote-sensing image understanding, and geospatial data mining and visualization.

YANG YANG is currently the Chief Engineer with PLA Unit 31008.

JIUNIU WANG received the B.S. degree in electrical engineering from the Beijing Institute of Technology, Beijing, China, in 2016. He is currently pursuing the Ph.D. degree with the Institute of Electronics, Chinese Academy of Sciences. His current research interests are in computer vision, natural language processing, and deep neural networks.

LI JIN received the B.S. degree from Xidian University, Xi'an, China, in 2012, and the Ph.D. degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 2017.

He is currently an Assistant Professor with the Institute of Electronics, Chinese Academy of Sciences. His research interests include machine learning, knowledge graph, and geographic information processing.



QING LIU received the B.S. degree from the Wuhan University of Technology, Wuhan, China, in 2016, and the M.D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2019. He is currently an Intern Researcher with the Institute of Electronics, Chinese Academy of Sciences. His research interests include machine learning, knowledge graph, and natural language processing.

TIANYUAN XIANG received the M.D. degree from Shanghai Jiaotong University, Shanghai, China. He is currently pursuing the Ph.D. degree with the University of Chinese Academy of Sciences. His research interest is mainly in the areas of swarm intelligence and task allocation.