

Research Article

Aspect-Level Sentiment Analysis Based on Position Features Using Multilevel Interactive Bidirectional GRU and Attention Mechanism

Xiaodi Wang ¹, Xiaoliang Chen ^{1,2}, Mingwei Tang,¹ Tian Yang,¹ and Zhen Wang¹

¹School of Computer and Software Engineering, Xihua University, Chengdu 610039, China

²Department of Computer Science and Operations Research, University of Montreal, Montreal, QC H3C3J7, Canada

Correspondence should be addressed to Xiaoliang Chen; xdxlchen@gmail.com

Received 5 April 2020; Accepted 16 June 2020; Published 8 July 2020

Academic Editor: Francisco R. Villatoro

Copyright © 2020 Xiaodi Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The aim of aspect-level sentiment analysis is to identify the sentiment polarity of a given target term in sentences. Existing neural network models provide a useful account of how to judge the polarity. However, context relative position information for the target terms is adversely ignored under the limitation of training datasets. Considering position features between words into the models can improve the accuracy of sentiment classification. Hence, this study proposes an improved classification model by combining multilevel interactive bidirectional Gated Recurrent Unit (GRU), attention mechanisms, and position features (MI-biGRU). Firstly, the position features of words in a sentence are initialized to enrich word embedding. Secondly, the approach extracts the features of target terms and context by using a well-constructed multilevel interactive bidirectional neural network. Thirdly, an attention mechanism is introduced so that the model can pay greater attention to those words that are important for sentiment analysis. Finally, four classic sentiment classification datasets are used to deal with aspect-level tasks. Experimental results indicate that there is a correlation between the multilevel interactive attention network and the position features. MI-biGRU can obviously improve the performance of classification.

1. Introduction

Capturing and analyzing the sentiments implied in large-scale comment texts has become a central topic for natural language processing (NLP). The tasks of fine-grained sentiment classification of the target terms in a given context are called aspect-level sentiment analysis, which have received considerable attention compared with acquiring traditional comprehensive sentiment polarity [1, 2]. A growing number of prestigious researchers and engineers around the world have post their opinions and reports on topics of sentiment classification online and offer them for free. These technical contributions have been properly accepted and acclaimed for their obvious advantages in NLP tasks. However, further aspect-level sentiment analysis is quite sensitive to current researchers. There are many problems in sentiment classification of aspect-level, including classification, regression, and recognition. We mainly focus on classification issues [3].

Sentiment predictions for a target term in a text are important for our increased understanding of sentence semantics and user emotions behind the sentences. The typical feature of aspect-level sentiment analysis can be exemplified in studies using the follow sentence expression: “they use fancy ingredients, but even fancy ingredients do not make for good pizza unless someone knows how to get the crust right.” The sentiment polarity of the target terms “ingredients,” “pizza,” and “crust” were positive, negative, and neutral, respectively. However, one potential problem is that the predictive accuracy of polarity is much lower than expected by the application, which is limited to complex sentence features and language environments. Traditional methods of comprehensive sentiment evaluation do not meet the requirements of fine-grained aspect-level tasks based on the target terms [4]. There are few studies that have investigated the association between sentiment polarity and the position information of target terms. Hence, this paper

proposes a multilevel interactive bidirectional attention network model, integrating bidirectional GRU and position information to improve the accuracy of aspect-level sentiment predictions.

Traditional published methods for processing aspect-level tasks are limited to the selection of feature sets. The focus of these studies, such as bag-of-words and sentiment lexicons [5], is to manually label a large number of features. Scholars have long debated the waste of labour on manual marking. However, the existing studies indicate that the quality of training models largely depends on the constructed labelled featured set. Recently investigators have examined the effects of deep learning compared with traditional manual generation methods in NLP tasks [6, 7]. The former has a clear advantage.

Recurrent neural network (RNN) can extract the essential features of word embedding by using a multilevel recurrent mechanism and then generate a vector representation of the target sentences. Most of sentiment classification models using RNN can achieve acceptable results through well-established tuning steps [8]. More recent attention on the sentiment classification tasks has focused on the provision of RNN variants. The first method to improve the models is to adjust their structures. For example, target-dependent long short term memory (TD-LSTM) [9] can divide a context into left and right parts according to target terms. Then, the hidden states to deal with aspect-level tasks are generated by combining two LSTM models in structure. The second is characterized by a change in the input of the models. For example, there are methods to associate the target term vectors with the context vectors as the whole input of the LSTM model, which can realize aspect-level tasks by enhancing the semantic features of the words [10].

Research on the method of neural networks on aspect-level tasks has been mostly restricted to limited performance improvement. However, few studies have been able to draw on systematic research into the importance of the words in a sentence. In other words, we cannot effectively identify which words in a sentence are more indispensable and cannot accurately locate these key words in aspect-level tasks. Fortunately, attention mechanisms which are widely used in machine translation [11], image recognition [12], and reading comprehension [13, 14] can solve this problem. Attention mechanisms [15] can be utilized to measure the importance of each word in a context to the target terms, in which attentions are ultimately expressed as weight score. The model will focus more attention on the words with high weight score and extract more information from the words related to the target terms, thus improving the performance of classification. Some scholars have invested in this domain and achieved excellent results such as AE-LSTM, ATAE-LSTM [16], MemNet [17], and IAN [18]. However, the influence of position parameters of target terms on classification performance has remained unclear [19, 20]. This indicates a need to understand the actual contribution of the position information.

Researchers review that the sentiment polarity of a target term contained in a sentence is related to the context around it, but not to those with greater distant. A well-constructed

aspect-level model should allocate higher weight score to the contexts that possess closer relative distance to the target term. The idea can be illustrated briefly by the following sentence: “they use fancy ingredients, but even fancy ingredients do not make for good pizza unless someone knows how to get the crust right.” In this case, the polarity is positive, negative, and neutral when the target term is setting as the word “ingredients,” “pizza,” and “crust,” respectively. In order to decide the polarity, we should intuitively concentrate on the words that are close to the target one and then consider the other words far away from it. Hence, the word “fancy” in the case compared to other words such as “good” and “get” will make a greater contribution to determine the polarity of target term “ingredients.” Consequently, adding position features can enrich word semantics in the embedding process. This work attempts to illuminate the fact that a model with position information can learn more sentence features on aspect-level tasks.

This study proposes an improved aspect-level classification model by combining multilevel interactive bidirectional Gated Recurrent Unit, attention mechanisms, and position features (MI-biGRU). The function of the model consists of three parts:

- (1) Calculate the positional index of each word in the sentence based on the current target term and express it in an embedding
- (2) Extract semantic features of target words and context using multilevel bidirectional GRU neural network
- (3) Use bidirectional attention mechanism to obtain the weight score matrix of hidden states and determine relevance of each context to the target word

The model not only extracts the abstract semantic features of sentences but also calculates the position features of words in parallel by a multilevel structure. A vector representation with more features can be obtained according to the bidirectional attention mechanism, which can enhance the performance of sentiment classification tasks. On top of that, bidirectional embeddings can be brought together to tackle accurate sentiment classification at the fine-grained level. Finally, the effectiveness of this model will be evaluated by using four public aspect-level sentiment datasets. The experimental results show that the proposed model can achieve good sentiment discrimination performance at aspect-level on all datasets.

This paper is organized as follows. Section 2 introduces the related work. Section 3 formulates the improved model MI-biGRU that is composed of multilevel interactive bidirectional Gated Recurrent Unit, attention mechanisms, and position features. Section 4 deals with experiments and reporting results on aspect-level sentiment analysis. Conclusions and future work are presented in Section 5.

2. Related Work

This section introduces the development of sentiment analysis in recent years. The general research can be divided into three parts: traditional sentiment analysis methods,

neural network-based methods, and applications of attention mechanism in aspect-level tasks.

2.1. Traditional Sentiment Analysis Methods. Existing traditional methods on sentiment classification are extensive and focus particularly on machine learning technologies, which solve two problems: text representation and feature extraction. First of all, several studies have used support vector machines (SVM) to deal with text representation research studies in sentiment classification tasks [21]. According to the formulation of SVM, all the words of the text do not make a distinction between target terms and normal context. There are other relatively text representation methods of the literature that is concerned with sentiment words [22, 23], tokens [24], or dependency path distance [24]. Above methods are called coarse-grained classifications. On top of that, the majority of studies on feature extraction have obtained the sentiment lexicon and bag-of-words features [25–27]. These methods have been playing an increasingly important role in improving the performance of classification. Yet, these existing types of approaches have given rise to a lot of heated debate. The model training is heavily dependent on the features we extract. Manually labelling features will inevitably take a lot of manpower and time resources. As a result, the classification performance is low because of the high dimension of useful information if the features are obtained from an unlabelled text.

2.2. Neural Network-Based Sentiment Analysis Methods. Using neural network-based methods has become increasingly popular among the sentiment classification tasks for their flexible structure and pleasant performance [28]. For example, all the models such as Recursive Neural Networks [29], Recursive Neural Tensor Networks [30], Tree-LSTMs [31], and Hierarchical LSTMs [32], more or less enhance the accuracy of sentiment classification by different constructive model structures. The models above improve the accuracy compared with traditional machine learning. However, researchers have come to recognize their inadequacies. Ignoring to distinguish the target terms of a sentence will greatly decrease the classification effect. Therefore, some scholars in academia have turned their attention to the target terms. Jiang et al. performed a similar series of experiments to show the significance of target terms for sentiment classification tasks [5]. Tang et al. reviewed the literature from the period and proposed two improved models TD-LSTM and TC-LSTM, which can deal with the problem of automatic target term extraction, context feature enhancement, and classification performance improvement. Zhang et al. conducted a series of trials on sentiment analysis in which they constructed a neural network model with two gate mechanisms [33]. The mechanisms implement the functions of extracting grammatical and semantic information and the relationship information between the left and right context for a target term, respectively. Finally, information extracted by two gate mechanisms is aggregated for sentiment classification. Overall, these studies highlight

the need for target terms, but there is no reference to the position of such terms and the relationship between position information and classification performance.

2.3. Application of Attention Mechanism in Aspect-Level Sentiment Analysis. Deep learning technologies are originally applied in the field of images, which have gradually turned to NLP area and achieved excellent results. Attention mechanisms in deep learning are necessarily serve as an effective way to highly accurate sentiment classification. A few NLP researchers, who have surveyed the intrinsic relevance between the context and the target terms in sentences, have been found. For example, Zeng et al. designed an attention-based LSTM for aspect-level sentiment classification [10], which processes target term embedding and word embedding in the pretraining step, simultaneously. Then, the target term vector will be put into an attention network to calculate the term weight. More recent attention has focused on aspect-level tasks with similar attention mechanisms. Tang et al. designed a deep memory network [17] with reporting multiple computational layers. Each layer is a context-based attention model, through which the relationship weight from context to target terms can be obtained. Ma et al. suggest the deep semantic association between context and target terms by proposing an interactive attention model [18]. This model obtains the two-way weight and combines them to perform aspect-level sentiment classification.

Most of the improved neural network models can achieve better results compared with the original one. However, these methods ignore the role of position relationship between context and target terms. As a result, the polarity of the target terms must be affected under certain positional relationships. The study of Zeng et al. subsequent offers some important insights into the application of position information in classification tasks [34]. For example, understanding the distance between context and target term and how to present such distance by embedding representation will help our aspect-level work. The work of Gu et al. uses a position-aware bidirectional attention network to investigate aspect-level sentiment analysis [20], which provides rich semantic features in embedding words expression.

As noted above, interactive network or position information is particularly useful in studying aspect-level sentiment analysis. So far, little attention has been paid to both of them, simultaneously. Hence, a combination of both interactive concept and position parameter was used in this investigation. This study proposes an improved aspect-level sentiment analysis model by combining multilevel interactive bidirectional Gated Recurrent Unit, attention mechanisms, and position features (MI-biGRU). First of all, the distance from the context to target term in our model will be prepared according to the similar procedure used by the scholars Zeng et al. [34]. On top of that, the word embedding with position information will be trained by using a multilevel bidirectional GRU neural network. Finally, a bidirectional interactive attention mechanism is used to compute the weight matrix to identify the context possessing the semantic association with the target term. MI-biGRU can perform a more classification accuracy for aspect-level tasks than previous models, which will be shown in Section 4.

3. Model Description

This section presents the details of the model MI-biGRU for aspect-level sentiment analysis. In the previous work, several definitions with symbolic differences have been proposed for sentiment analysis. Hence, we need to provide the basic concepts and notations of MI-biGRU classification involved in this paper.

A sentiment trcategory problem represented in MI-biGRU is associated with a three-tuple polarity set (positive, negative, and neutral). Given a sentence with n words, including context and target terms. A target term is usually denoted by the word group composed of one or more adjacent words in context, where the position of the first and last word in the word group is called the start and end position, respectively. The target term embedding sequence can be denoted by $[e_a^1, e_a^2, \dots, e_a^m]$ with m predetermined target terms. Notation $[p_1, p_2, \dots, p_n]$ represents the relative distance embedding from each word w_c^i , $i \in 1, 2, \dots, n$, of a sentence to a target term.

The overall architecture of MI-biGRU can be illustrated in Figure 1. This model possesses two inputs. One is a word embedding sequence $[w_1, w_2, \dots, w_n]$ that consists of a context embedding sequence $[e_c^1, e_c^2, \dots, e_c^n]$ and a position embedding sequence $[p_1, p_2, \dots, p_n]$. The other is a target term embedding sequence $[e_a^1, e_a^2, \dots, e_a^m]$. The goal of the model is to extract enough semantic information from two embedding sequences and combine them to perform aspect-level sentiment classification. Notations employed to represent the components of MI-biGRU are described in Table 1. The details of the model are divided into six steps based on their execution order.

3.1. Position Representation. Aspect-level tasks have benefited a lot from position embedding representation for more valuable word features [34]. The concept of relative distance between words serves to quantify the relevance of a sentence word to a target term. We are required to represent the position information by a embeddable vector pattern, which can be formalized as an integer vector or a matrix depending on whether there concerns unique or multiple target terms.

First of all, the word position index of a target term in a sentence will be marked as the cardinal point “0.” Discrete spacing from the i th word w_c^i in a sentence to the cardinal point is called the relative distance of w_c^i , which is denoted by p_i and can be calculated by the formula:

$$p_i = \begin{cases} |a_s - i|, & i < a_s, \\ 0, & a_s < i < a_e, \\ |a_e - i|, & i > a_e. \end{cases} \quad (1)$$

Extending this concept in a sentence with n words gives us the following position index list $\mathcal{P} = [p_1, p_2, \dots, p_n]$. This can be illustrated briefly by the following two examples. Firstly, if the unique word “quantity” is applied as the target term in the sentence “the quantity is also very good, you will come out satisfied,” we can develop the position index list

$[1, \underline{0}, 1, 2, 3, 4, 5, 6, 7, 8, 9]$ by setting a cardinal point “0” for the second word “quantity” and deriving an increasingly positive integer from the left or right direction for other words. Secondly, if the target term contains more than one adjacent word, all the internal words are assigned as the cardinal point “0.” Other words in the sentence will obtain an increasingly positive integer from the start position of the term to left direction or from the end position of the term to right direction. Therefore, the position index list $[6, 5, 4, 3, 2, 1, \underline{0}, \underline{0}, 1, 2, 3, 4, 5, 6, 7]$ will be obtained if we set the case sentence and the target term with “all the money went into the interior decoration, none of it went to the chefs” and “interior decoration,” respectively.

On top of that, if multiple target terms are applied to a sentence, we can obtain a position index list sequence that is called the position matrix. Assuming that a sentence has n words and m target terms, let notation \mathcal{P}_i denote the list index of the i th target term in a sentence. Position matrix G is defined as

$$G = \begin{bmatrix} \mathcal{P}_1 \\ \mathcal{P}_2 \\ \vdots \\ \mathcal{P}_m \end{bmatrix} \in \mathbb{R}^{m \times n}, \quad (2)$$

where m refers to the number of target terms and n is the number of words in the sentence. Then, we use a position embedding matrix $P \in \mathbb{R}^{d_p \times n}$ to convert position index sequence into a position embedding, where d_p refers to the dimension of position embedding. P is initialized randomly and updated during the model training process. The matrix is further exemplified in the same example “all the money went into the interior decoration, none of it went to the chefs,” which contains two target terms “interior decoration” and “chefs.” We can obtain the position matrix:

$$G = \begin{bmatrix} \mathcal{P}_1 \\ \mathcal{P}_2 \end{bmatrix} = \begin{bmatrix} 6, & 5, & 4, & 3, & 2, & 1, & \underline{0}, & \underline{0}, & 1, & 2, & 3, & 4, & 5, & 6, & 7 \\ 14, & 13, & 12, & 11, & 10, & 9, & 8, & 7, & 6, & 5, & 4, & 3, & 2, & 1, & \underline{0} \end{bmatrix}. \quad (3)$$

The position matrix has been playing an increasingly important role in helping researchers get a better sense of aspect-level tasks. We can first observe the polarity of the emotional words which are near the target term and then consider the other words which are far away to judge whether a sentence is positive or not. For example, as presented in the case “the quantity is also very good, you will come out satisfied,” the distance of “good” (4) is closer than “satisfied” (9) by a simple numerical comparison according to the index list $[1, \underline{0}, 1, 2, 3, 4, 5, 6, 7, 8, 9]$, and the approach will give priority to “good” instead of “satisfied” when we judge the sentiment polarity of the subject “quantity” of the sentence. This study suggests that adding position information to initialize word embedding can provide more features to perform aspect-level sentiment classification.

3.2. Word Representation. One of the basic tasks of sentiment analysis is to present each word in a given sentence by embedding operation. A feasible approach is to embed each

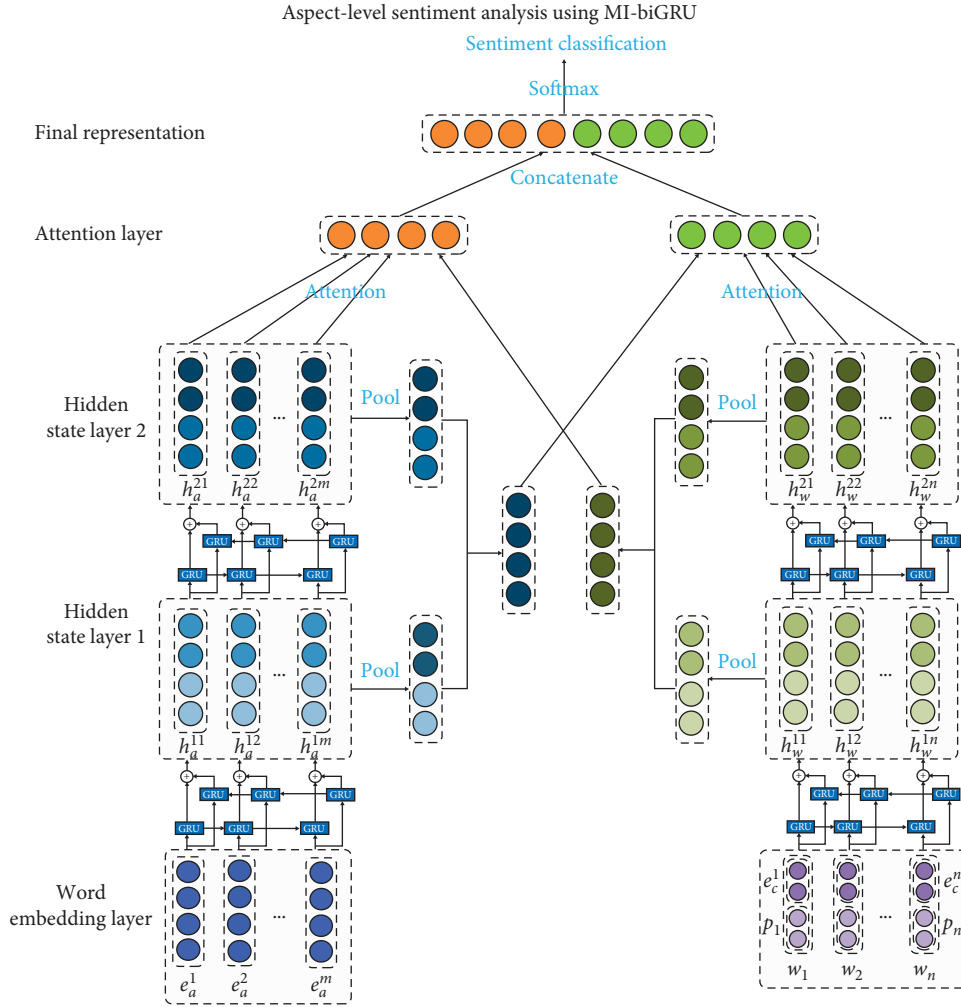


FIGURE 1: Architecture of MI-biGRU.

TABLE 1: Component description of the proposed model.

Notation	Description
w_i^i	The i th word in a context
w_a^j	The j th word in a target term
p_i	Relative distance from the i th word to a target term in a sentence
w_i	Concatenation of the i th context and position embedding
a_s	Start position of a target term
a_e	End position of a target term
e_c^i	Embedding of the i th context in a sentence
e_a^i	Embedding of the i th target term in a sentence
h_w^{1i}	First hidden layer vector formed from w_i
h_a^{1i}	First hidden layer vector formed from e_a^i
h_w^{2i}	Second hidden layer vector formed from h_w^{1i}
h_a^{2i}	Second hidden layer vector formed from h_a^{1i}

word in a low-dimensional real value vector through the word embedding matrix $E \in \mathbb{R}^{d_w \times v}$, where d_w represents the dimension of word embedding and v denotes the size of vocabulary. Matrix E is generally initialized by random number generation operation. Then, matrix weight will be updated to reach a stable value in the process of model

training. Another feasible method to obtain the matrix E is to pre-train it through the existing corpus [35].

This study uses pretrained Glove (pretrained word vectors of Glove can be obtained from <http://nlp.stanford.edu/projects/glove/>) from Stanford University to get word embeddings. Four sets of sequence symbols are applied to

distinguish texts and their embedding form. In a sentence, the context and its embedding with n words can be denoted by $[w_c^1, w_c^2, \dots, w_c^n]$ and $[e_c^1, e_c^2, \dots, e_c^n]$, respectively. The m target terms and their embeddings can be denoted by $[w_a^1, w_a^2, \dots, w_a^m]$ and $[e_a^1, e_a^2, \dots, e_a^m]$, respectively. On top of that, context vector $[e_c^1, e_c^2, \dots, e_c^n]$ and obtained position embedding $[p_1, p_2, \dots, p_n]$ will be concatenated to get final word embedding representation of each word in a sentence as $[w_1, w_2, \dots, w_n]$.

3.3. Introduction of Gate Recurrent Unit (GRU).

Recurrent neural network is a widespread network employed in natural language processing in recent years. One advantage of RNN is the fact that it can process variable-length text sequence and extract key features of a sentence. However, model performance of traditional RNN in the case of long sentences has been mostly restricted by the problems of gradient disappearance and explosion during the training process. As a result, RNN was unable to send vital information in text back and forth.

A great deal of previous research into RNN has focused on model variants. Much of the current scholars pays particular attention to LSTM models and the GRU models. Both LSTM and GRU provide a gate mechanism so that the neural network can reserve the important information and forget those that are less relevant to the current state. It has been universally accepted that GRU has the advantages of fewer necessary parameters and lower network complexity compared with LSTM. Therefore, this paper plans to use the GRU model to extract the key features of word embedding.

Details of GRU are illustrated according to the network structure shown in Figure 2. The GRU simplifies four gate mechanisms of LSTM, i.e., input gate, output gate, forget gate, and cell state, into two gates that are called reset gate and update gate. At any time step t , GRU includes three parameters: reset gate r_t , update gate z_t , and hidden state h_t . All the parameters are updated according to the following equations:

$$\begin{aligned} z_t &= \sigma(U^z \cdot x_t + W^z \cdot h_{t-1}), \\ r_t &= \sigma(U^r \cdot x_t + W^r \cdot h_{t-1}), \\ \tilde{h}_t &= \tanh(U^h \cdot x_t + W^h \cdot (h_{t-1} \odot r_t)), h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t. \end{aligned} \quad (4)$$

The symbolic meaning are described as follows. x_t denotes the input word embedding at time t . h_{t-1} represents the hidden state at time $t-1$. U^z, U^r, U^h and $W^z, W^r, W^h \in \mathbb{R}^{d_h \times d_h}$ denote weight matrices, where d_h indicates the dimension of hidden state. σ and \tanh denote the sigmoid and tanh function, respectively. Notation \cdot is the dot product and \odot is the elementwise multiplication.

In this study, we choose bidirectional GRU to obtain the vector representation of a hidden layer for the target term and context, which can extract more comprehensive features compared with normal GRU. The bidirectional GRU consists of a forward hidden layer state $h_t^f \in \mathbb{R}^{d_h}$ and a backward hidden layer state $h_t^b \in \mathbb{R}^{d_h}$. First of all, this work concatenates the hidden states from opposite directions to get the

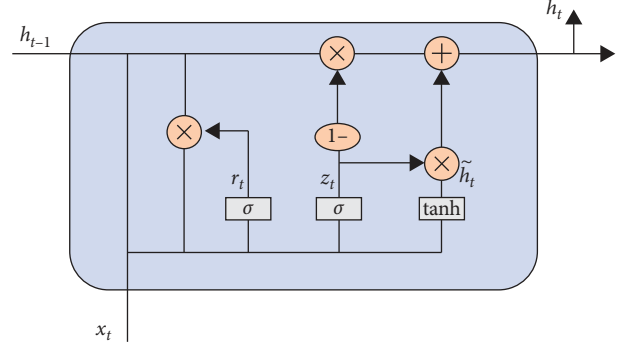


FIGURE 2: Representation of GRU structure diagram.

hidden layer embedding $h_t^f = [\vec{h}_t^f, \overleftarrow{h}_t^f] \in \mathbb{R}^{2d_h}$. Some vital features of a sentence that we required are included in h_t^f . On top of that, two hidden layer embeddings h_a^{1i} ($1 \leq i \leq m$) and h_w^{1j} ($1 \leq j \leq n$) can be obtained by processing the target term and context with bidirectional GRU, where m and n denotes the number of target terms and context in a sentence, respectively. Let notations h_a^{2i} ($1 \leq i \leq m$) and h_w^{2j} ($1 \leq j \leq n$) denote the hidden layer embeddings induced by h_a^{1i} and h_w^{1j} , respectively. Finally, we can iteratively put obtained vectors $[h_a^{11}, h_a^{12}, \dots, h_a^{1m}]$ and $[h_w^{11}, h_w^{12}, \dots, h_w^{1n}]$ into bidirectional GRU and generate the following hidden layer embeddings $[h_a^{21}, h_a^{22}, \dots, h_a^{2m}]$ and $[h_w^{21}, h_w^{22}, \dots, h_w^{2n}]$, respectively. The latter carries higher level features than previous embeddings.

3.4. Position-Based Multilevel Interactive Bidirectional Attention Network. Interaction information between target terms and context is particularly useful in studying sentiment classification. This study performs mean pooling on hidden layer embeddings $[h_a^{11}, h_a^{12}, \dots, h_a^{1m}]$, $[h_a^{21}, h_a^{22}, \dots, h_a^{2m}]$, $[h_w^{11}, h_w^{12}, \dots, h_w^{1n}]$, and $[h_w^{21}, h_w^{22}, \dots, h_w^{2n}]$ in order to get interactive information, respectively. The pooled results are then averaged as a part of input for the attention mechanism, and the final averaged embedding contains most of semantic features. The formulas used for average pooling are described as follows:

$$\begin{aligned} w_{\text{avg1}} &= \frac{\sum_{i=1}^n h_w^{1i}}{n}, \\ a_{\text{avg1}} &= \frac{\sum_{i=1}^m h_a^{1i}}{m}, \\ w_{\text{avg2}} &= \frac{\sum_{i=1}^n h_w^{2i}}{n}, \\ a_{\text{avg2}} &= \frac{\sum_{i=1}^m h_a^{2i}}{m}, \\ w_{\text{avg}} &= \frac{(w_{\text{avg1}} + w_{\text{avg2}})}{2}, \\ a_{\text{avg}} &= \frac{(a_{\text{avg1}} + a_{\text{avg2}})}{2}. \end{aligned} \quad (5)$$

The bidirectional GRU can extract more information carried by words in a given sentence and convert them into hidden states. However, the words differ in their importance to the target term. We should capture the relevant context for different target terms and then design a strategy to improve the accuracy of classification by increasing the intensity of the model's attention to these words. Weight score can be used to express the degree of model concern. The higher the score, the greater the correlation between target terms and context. Hence, an attention mechanism is developed to calculate the weight score between different target terms and context. If the sentiment polarity of a target term is determined by the model, we should pay greater attention to the words that have a higher score.

This study calculates the attention weight score in opposite directions. One is from target terms to context, and the other is from context to target terms. Therefore, the two-way approach was chosen because we can get two weight score matrices in order to improve the performance of our model. The process of using the attention mechanism in the model are described, as shown in Figure 1. First of all, the target term embedding matrix $[h_a^{21}, h_a^{22}, \dots, h_a^{2m}]$ and the averaged context embedding w_{avg} are calculated to obtain the attention vector α_i :

$$\alpha_i = \frac{\exp(f(h_a^{2i}, w_{\text{avg}}))}{\sum_{j=1}^m \exp(f(h_a^{2j}, w_{\text{avg}}))}, \quad (6)$$

$$f(h_a^{2i}, w_{\text{avg}}) = \tanh(w_{\text{avg}}^T \cdot W_m \cdot h_a^{2i} + b_m),$$

where f is a score function that calculates the importance of h_a^{2i} in the target term, $W_m \in \mathbb{R}^{2d_h \times 2d_h}$ and $b_m \in \mathbb{R}^{1 \times 1}$ are the weight matrix and the bias, respectively, notation \cdot is the dot product, w_{avg}^T is the transpose of w_{avg} , and \tanh represents a nonlinear activation function.

On top of that, the averaged target term embedding a_{avg} and context embedding matrix $[h_w^{21}, h_w^{22}, \dots, h_w^{2n}]$ are applied to calculate the attention vector β_i for the context. The equation is described as follows:

$$\beta_i = \frac{\exp(f(h_w^{2i}, a_{\text{avg}}))}{\sum_{j=1}^n \exp(f(h_w^{2j}, a_{\text{avg}}))}. \quad (7)$$

If we obtain the attention weight vectors α_i and β_i , the target term representation a and the context representation w can be deduced by the following equations:

$$a = \sum_{i=1}^m \alpha_i h_a^{2i}, w = \sum_{i=1}^n \beta_i h_w^{2i}, \quad (8)$$

where $a \in \mathbb{R}^{2d_h}$ and $w \in \mathbb{R}^{2d_h}$ denote the final word embedding representations of the target term and context, which will be processed in the output layer.

3.5. Output Layer. Target term and context representations described in Section 3.4 will be concatenated as $d = [a; w] \in \mathbb{R}^{4d_h}$ at the output layer. Then, a nonlinear transformation layer and a softmax classifier are prepared

according to equation (9) to calculate the sentiment probability value:

$$z = \tanh(d \cdot W_n + b_n), \quad (9)$$

where $W_n \in \mathbb{R}^{4d_h \times d_c}$ and $b_n \in \mathbb{R}^{d_c}$ are the weight matrix and the bias, respectively, and d_c represents the number of classifications. Probability P_k of category was analyzed by using the following softmax function:

$$P_k = \frac{\exp(z_k)}{\sum_{j=1}^k \exp(z_k)}. \quad (10)$$

3.6. Model Training. In order to improve the model performance for the tasks of aspect-level sentiment classification, this approach deals with the optimization of training process, including word embedding layer, bidirectional GRU neural network layer, attention layer, and nonlinear layer. The crossentropy with L2 regularization is applied as the loss function, which is defined as follows:

$$L = - \sum_{i=1}^S y_i \log(\tilde{y}_i) + \frac{1}{2} \lambda \|\theta\|^2, \quad (11)$$

where λ is a regularization coefficient, $\|\theta\|^2$ represents the L2 regulation, y_i denotes the correct sentiment polarity in training dataset, and \tilde{y}_i denotes the predicted sentiment polarity for a sentence by using the proposed model. The parameter Θ is updated according to the gradient calculated by using a backpropagation method. The formula is as follows:

$$\Theta = \Theta - \lambda \frac{\partial L(\Theta)}{\partial \Theta}, \quad (12)$$

where λ is the learning rate. In the training process, the method designs dropout strategy to randomly remove some features of the hidden layer in order to avoid overfitting.

4. Experiments

Section 3 has shown the theoretical formulas and operational steps for the MI-biGRU model. This section has attempted to provide a series of experiments relating to four public aspect-level sentiment classification datasets from different domains. The aim of the experiments is to test the feasibility of applying MI-biGRU to deal with aspect-level tasks and evaluate the effectiveness of the proposed MI-biGRU model.

4.1. Experimental Setting

4.1.1. Dataset and Preprocess. SemEvals are one of the most popular aspect-level comment dataset series, which report a large amount of customer comments from the domain of restaurant and laptop sales. We extract SemEval 2014 (<http://alt.qcri.org/semEval2014/task4/>), SemEval2015 (<http://alt.qcri.org/semEval2015/task12/>), and SemEval2016 (<http://alt.qcri.org/semEval2016/task5/>) from the series as the experimental data. SemEval 2014 includes two datasets for restaurant and laptop sales, respectively. SemEvals 2015 and

2016 are related to the restaurant. Each piece of data in the datasets is a single sentence which contains comments, target terms, sentiment labels, and position information. We remove the sentences which the target term is labelled with “null” or “conflict” from the datasets, and the remaining sentences possess a corresponding sentiment label for each target term. The statistics of the datasets are provided in Table 2.

4.1.2. Pretraining. This section presents the pretraining process for word embedding matrix E that is generally set by a random initialization operation, and the weight score is then updated during the training process. However, E can be pretrained on some existing corpora. The benefit of this approach is that we can obtain the optimal parameters of the model from the high-quality datasets. Hence, a pretrained Glove from Stanford University was adopted in order to improve the model performance. As a result, the parameters of word embedding and bidirectional GRU in this study are initialized with the same parameters in corresponding layers.

4.1.3. Hyperparameters Setting. Other parameters except pretraining of word embedding are initialized by the sampling operations from uniform distribution $U(-0.1, 0.1)$, in which all bias are set to zero. Both dimensions of word embedding and bidirectional GRU hidden state are set to 300. The dimension of position embedding is considered as 100. The size of batch is placed at 128. We take 80 as the max length of a sentence. The coefficient of $L2$ regulation and the learning rate is set to 10^{-5} and 0.0029, respectively. This experiment uses dropout strategy with the dropout rate 0.5 in order to avoid suffering from overfitting. It is important to emphasize that using the same parameters for different datasets may not yield the best results. However, there will be a series of parameters that will optimize the execution of the model on each dataset from a global perspective. Therefore, we confirm the above parameters as the final hyperparameters of the model through a large number of experiments. In addition, we use the Adam optimizer to optimize all parameters. Experiments have shown that the Adam optimizer performs better than other optimizers such as SGD and RMSProp on our classification task.

4.1.4. Evaluation. This section presents the performance evaluation indicators for all baseline methods that are mentioned in Sections 4.2 and 4.3. The definition of “Accuracy” is as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (13)$$

As outlined in Table 3, the symbol TP is the abbreviation for “True Positive,” which refers to the fact that both the sentiment label and the model prediction are positive. The symbol FP is short for “False Positive,” which means that the sentiment label is negative and the model prediction is positive. Similarly, “False Negative” and “True Negative” are presented as the symbols FN and TN, respectively.

A broader set of estimation indicators precision, recall, and F1-score [36, 37] is also adopted in this study:

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ F_1 &= \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \end{aligned} \quad (14)$$

Historically, the term “Precision” has been used to describe the ratio of correctly predicted positive observations to total predicted positive observations, which is generally understood as the ability to distinguish the negative samples. The higher the “Precision,” the stronger the model’s ability to distinguish negative samples. Previous studies mostly defined “Recall” as the ratio of correctly predicted positive observations to all observations in the actual class [38], which reflects the model’s ability to recognize positive samples. The higher the recall, the stronger the model’s ability to recognize positive samples. The term “F1-score” combines “Precision” and “Recall.” The robustness of the classification model is determined by “F1-score.”

4.2. Compared Models. To demonstrate the advantage of our method on aspect-level sentiment classification, we compared it with the following baselines:

- (i) LSTM [10]: LSTM is a classic neural network, which learns the sentiment classification labels by the transformation from word embeddings to hidden states, an average operation of the states, and a softmax operation. LSTM has only been carried out in coarse-grained classification tasks and has not dealt with the aspect-level tasks.
- (ii) AE-LSTM [10]: AE-LSTM is a variant of LSTM, which adds the connection operation between hidden states and target term to generate an attention weight vector. Existing studies use this vector representation to determine the sentiment polarity of a sentence.
- (iii) ATAE-LSTM [10]: the model structure of ATAE-LSTM is similar to AE-LSTM except for the step of word embedding initialization. Adding the target word embedding into each context embedding at the initialization step can highlight the status of the target term in LSTM and obtain more sufficient features.
- (iv) IAN [18]: INA is a neural network with interactive structure. Firstly, the model calculates the two-way attention weights between the context and target word to obtain their rich relational features. Secondly, the information of two directions is concatenated to perform aspect-level sentiment classification.
- (v) MemNet [17]: MemNet is an open model to perform aspect-level sentiment classification by using

TABLE 2: Statistics of aspect-based sentiment analysis datasets.

Dataset	Train			Test		
	Positive	Negative	Neural	Positive	Negative	Neural
Restaurant14	2164	807	637	728	196	196
Laptop14	994	870	464	341	128	169
Restaurant15	1178	382	50	439	328	35
Restaurant16	1620	709	88	597	190	38

TABLE 3: Description of TP, FP, TN, and FN.

Actual class	Predict class	
	Positive	Negative
Positive	TP (True Positive)	FN (False Negative)
Negative	FP (False Positive)	TN (True Negative)

the same attention mechanism multiple times. The weight matrix is optimized through multilayer interaction, which can extract high-quality abstract features.

- (vi) PBAN [20]: in contrast to above baselines, PBAN introduces the relative distance between the context and the target word in a sentence to perform aspect-level sentiment classification. The model focuses more on the words that are close to the target term. The attention mechanism is also used in PBAN to calculate the weight matrix.
- (vii) MI-biGRU: the model is proposed in this paper. MI-biGRU combines the concept of relative distance and the improved GRU with interactive structure to perform aspect-level tasks.

4.3. Comparison of Aspect-Level Sentiment Analysis Model.

This section presents an application result for all baseline methods that are mentioned in Section 4.2. We evaluated the effectiveness of our method in terms of aspect-level sentiment classification results on four shared datasets. The experiment chooses *Accuracy* and *F1-score* to evaluate all these methods because *Accuracy* is the basic metric and *F1-score* measures both *precision* and *recall* of the classification results.

As we can see from Table 4, the *Accuracy* and *F1-score*, i.e., 74.28 and 60.24, under dataset Restaurant14, of LSTM is the lowest of all the models. A common low score under different datasets indicated that LSTM lacks the mechanism to process multiple target terms in a sentence, although LSTM averages the hidden states. Developing a model to process more than one target term in a sentence contributes a lot to improve classification performance. An improvement over the baseline LSTM is observed in AE-LSTM and ATAE-LSTM. Particularly, we can notice that the results on the four datasets are better than the baseline LSTM by approximately 2–7% since AE-LSTM and ATAE-LSTM add judgments on target terms.

Introducing interactive structure or attention mechanism to models, i.e., IAN and MemNet in Table 4, is

necessarily served as an effective way to improve the assessment scores since the abstract features and relationships between target terms and context play a positive role in model performance. In the process of initializing word embedding, abundant features can be learned if the relative distance is considered. As we can see from the model PBAN in Table 4, with the added position information, its scores surpass most baseline methods except for 79.73 and 80.79 in datasets Restaurant14 and Restaurant16.

The proposed model MI-biGRU combines the concept of relative distance, interactive model structure, and the word embedding initialization involving context and target terms to perform aspect-level tasks. Linking the position vector with the context can enrich the input features of the model.

Recently IAN and MemNet models have examined the positive effects of the attention mechanism between context and target terms on aspect-level feature extraction. Improvement model MI-biGRU of bidirectional GRU neural network generates the attention weight matrix from the target term to context and from the context to target term by using interactive attention mechanism twice. Degree found to be influencing classification effect of MI-biGRU is presented in Table 4. We can notice that the classification performance on MI-biGRU is much better than other baseline methods.

We achieve the best *Accuracy* and *F1-score* in datasets Restaurant14, Laptop14, and Restaurant15. We can notice that ATAE-LSTM obtained the best *Accuracy* 85.17 in Restaurant16. The reason why the accuracy of Restaurant16 is not optimal on MI-biGRU may be the imbalance of data. In dataset SemEval 2016, the amount of data with sentiment polarity is very small in the restaurant domain for testing, so the model does not classify some comments correctly. However, the optimal *F1-score* is also obtained by our model, which illustrates that MI-biGRU is capable of distinguishing positive and negative samples. Therefore, we can come to the conclusion that our model can achieve the best performance in four public datasets.

4.4. Model Analysis of MI-biGRU. This section illustrates the rationality of each component in MI-biGRU by contrast experiments. The dependence of model accuracy on the amount of data is described according to a regular data change. Experiments of different technical combinations are shown in Table 5.

In order to assess the contribution of the three technical elements, i.e., bi-GRU structure, position, and interaction, in the process of word embedding, measuring accuracy by

TABLE 4: Performance of different models on sentiment analysis datasets.

Dataset	Restaurant14		Laptop14		Restaurant15		Restaurant16	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1
LSTM	74.28	60.24	66.45	60.23	75.36	48.69	83.12	52.55
AE-LSTM	76.12	61.08	68.83	62.08	75.94	50.11	84.46	56.28
ATAE-LSTM	77.23	62.97	68.61	62.99	76.54	53.29	85.17	59.32
IAN	77.14	64.53	71.94	66.51	77.28	52.41	81.49	62.57
MemNet (9)*	80.57	65.49	71.49	68.37	77.18	53.36	81.42	60.49
PBAN	79.73	71.08	74.61	70.82	78.11	56.73	80.79	61.85
MI-biGRU	81.66	73.86	75.25	70.93	79.17	59.97	82.54	63.48

The number (9) after MemNet refers to nine computational layers that are adopted.

TABLE 5: Model accuracy of different technical combinations.

	Restaurant14	Laptop14	Restaurant15	Restaurant16
Single-level interactive bi-GRU	81.38	72.49	77.99	81.96
Multilevel interactive bi-GRUs	79.03	72.78	78.82	82.07
Position + single-level interactive bi-GRU	81.41	74.38	78.58	82.07
Position + multilevel interactive bi-GRUs	81.66	75.25	79.17	82.54

element combinations of four datasets was used. First of all, it can be seen from the data in Table 5 that the accuracy advantages of the model with two bidirectional GRUs (bi-GRUs) are not obviously compared with only one bi-GRU structure, especially in laptop domain of SemEval 2014 and restaurant domain of SemEval 2015. Therefore, increasing the number of bi-GRU structures may not significantly improve model performance. On top of that, it is apparent from this table in which position information plays an increasingly important role in model accuracy since the performance of a bi-GRU structure with position is greatly enhanced compared to the independent bi-GRU model. Finally, what is highlighted in the table is the model MI-biGRU that adds a multilevel interactive GRU layer based on position information. MI-biGRU achieved the best performance on four aspect-level sentiment classification datasets. The test findings in Table 5 illustrate the feasibility of our ideas on aspect-level sentiment analysis.

Simple random data extraction experiments were utilized to analyze the dependence of model accuracy on the quantity of data. We randomly take 20%, 40%, 60%, and 80% of the data from the dataset of SemEval 2014 in restaurant and laptop domain. The trends of *Accuracy* and *F1-score* of baselines LSTM, IAN, PBAN, and MI-biGRU under datasets Restaurant14 and Laptop14 are shown in Figures 3–6.

The trends of Figures 3–6 reveal that there has been a steady increase of *Accuracy* and *F1-score* of all the baseline methods with the rise of data amount. The *Accuracy* and *F1-score* of MI-biGRU is low in the initial data amount. However, the obvious finding to emerge from the analysis is that the *Accuracy* and *F1-score* of MI-biGRU, at 60% of the data amount, has reached an approximate peak. This may be because rich data is provided to the model, and the advantages of position information and multilevel interaction mechanism are realized. Compared with using a small amount of data, rich data can allow the model to learn more complete semantic features. Therefore, the performance of

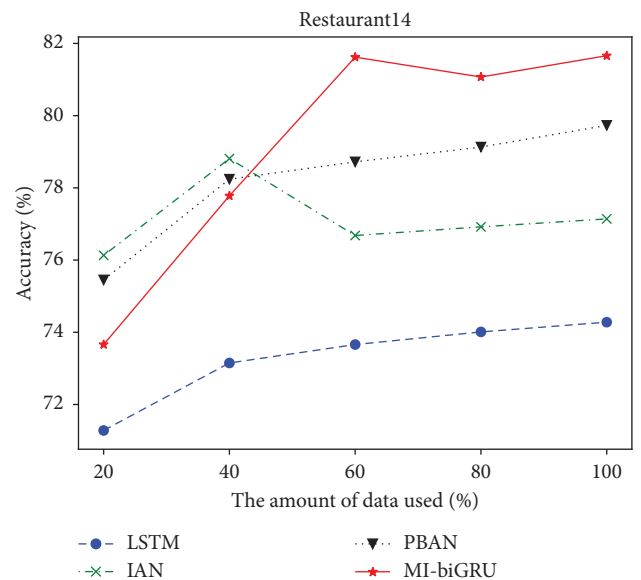


FIGURE 3: Accuracy trend of baseline models in restaurant domain.

sentiment analysis achieves better results than baseline models. When the amount of data exceeds 60%, our model's performance enhances rapidly and finally reaches the highest score. Therefore, from a holistic perspective, MI-biGRU can achieve the best results in classification tasks compared to baseline models if we can reach a critical mass of data.

4.5. Case Study. Working principle and high performance of the novel model can be further illustrated by a case study that visualizes the attention weight between target terms and context according to color shades. The darker color of the words in Figure 7 indicates the greater attention weight,

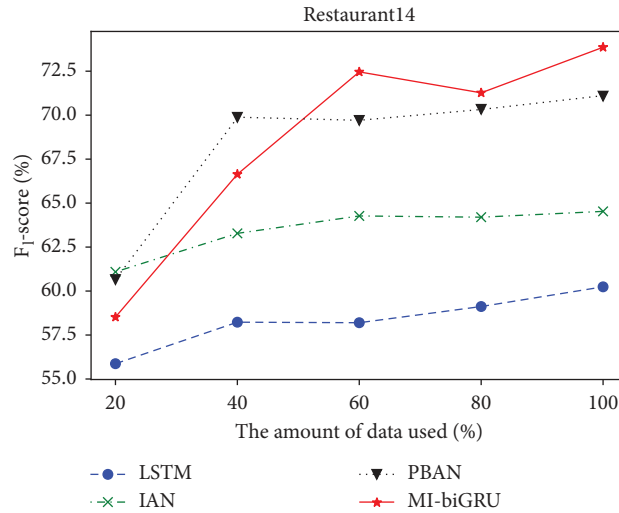


FIGURE 4: F1-score trend of baseline models in restaurant domain.

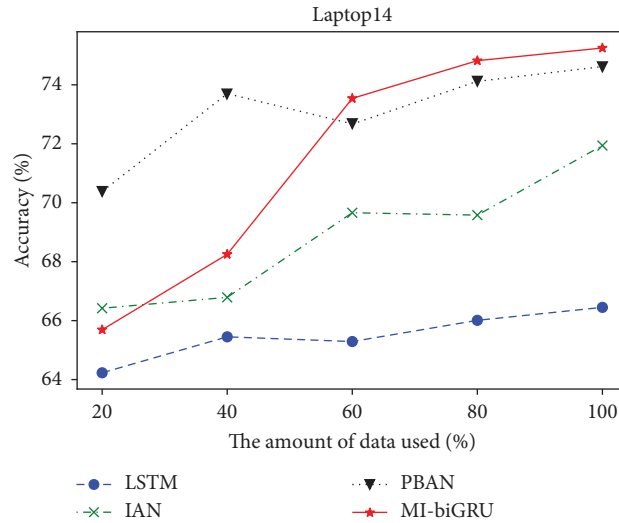


FIGURE 5: Accuracy trend of baseline models in laptop domain.

which is more essential to judge the polarity of a target term. Thus, the model will pay more attention to these words.

This study confirms that when the model judged the polarity of a target term, it paid more attention to the words around it. Target term “weekends” and “French food” are perfect examples in Figure 7. When we distinguish the polarity of “weekends,” the words “bit” and “packed” are more critical than the words with bigger relative distances. The words such as “best” that are close to the target term “French food” will have a greater impact on polarity judgment.

One interesting finding was that the model may give some words with small distance a low attention weight. For example, the model actually gives the words “but” and “vibe” a low weight when the target term is “weekends.” This phenomenon is normal since the sentiment contribution of the same words in the sentence to different target terms is varied. MI-biGRU will automatically select surrounding words that are worth giving more attention according to the

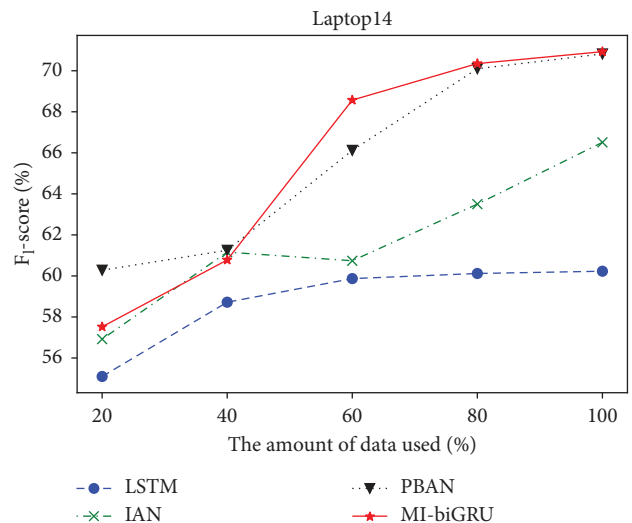


FIGURE 6: F1-score trend of baseline models in laptop domain.

Target term	Sentence	Polarity
weekends	It may be a bit packed on weekends , but the vibe is good and	Negative
	it is the best French food you will find in the area .	
French food	It may be a bit packed on weekends , but the vibe is good and	Positive
	it is the best French food you will find in the area .	

FIGURE 7: Visualized attention weights.

specific target term and then judge the sentiment polarity to perform aspect-level sentiment classification.

5. Conclusion and Future Work

This paper puts forward a novel descriptive multilevel interactive bidirectional attention network (MI-biGRU) which involves both bidirectional GRUs and position information for aspect-level sentiment analysis. We refine the traditional aspect-level models by considering the context, target terms, and relative distance in word embeddings. In addition, two bidirectional GRUs and interactive attention mechanisms are combined to extract abstract deep features in aspect-level tasks. The experimental results on restaurant and laptop comment tasks demonstrate our advantage over the traditional sentiment classification methods.

As a sentiment classification method, MI-biGRU performs very well on comment context, especially when a critical mass of aspect-level sentences is reached. Extracting the attention weight of words from the position-labelled context according to information interaction makes MI-biGRU more effective than other regular methods.

Our future work focuses on finding the embedding conclusions of the words with semantic relationships. Furthermore, we will figure out phrase-level sentiment method with position information.

Data Availability

The data used to support the findings of this study are included within the article by reporting a website link <http://alt.qcri.org/semEval2014/task4/>, <http://alt.qcri.org/semEval2015/task12/>, and <http://alt.qcri.org/semEval2016/task5/>.

Conflicts of Interest

The authors declare that they have no conflicts of Interest.

Acknowledgments

This work was supported by the Chunhui Plan Cooperation and Research Project, Ministry of Education of China (no. Z2015100), National Natural Science Foundation (no. 61902324), Scientific Research Funds project of Science and Technology Department of Sichuan Province (nos. 2016JY0244, 2017JQ0059, 2019GFW131, 2020JY, and 2020GFW), Fund Project of Chengdu Science and Technology Bureau (no. 2017-RK00-00026-ZF), Fund of Sichuan Educational Committee (nos. 15ZB0134 and 17ZA0360) and

Foundation of Cyberspace Security Key Laboratory of Sichuan Higher Education Institutions (no. sjzz2016-73), and University-sponsored Overseas Education Project of Xihua University.

References

- [1] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, Cambridge University Press, Cambridge, UK, 2015.
- [2] K. Schouten and F. Frasinca, "Survey on aspect-level sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 813–830, 2016.
- [3] M. Tsytarau and T. Palpanas, "Survey on mining subjective data on the web," *Data Mining and Knowledge Discovery*, vol. 24, no. 3, pp. 478–514, 2012.
- [4] Y. Tay, A. T. Luu, and S. C. Hui, "Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis," 2017, <https://arxiv.org/abs/1712.05403>.
- [5] L. Jiang, M. Yu, M. Zhou et al., "Target-dependent twitter sentiment classification," in *Proceedings of the Meeting of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Portland, OR, USA, June 2011.
- [6] K. C. Cho, B. Merriënboer, C. Gulcehre et al., "Learning phrase representations using rnn encoder-decoder for statistical machine translation," 2014, <https://arxiv.org/abs/1406.1078>.
- [7] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
- [8] S. W. Lai, L. H. Xu, K. Liu et al., "Recurrent convolutional neural networks for text classification," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, vol. 333, pp. 2267–2273, Austin, TX, USA, January 2015.
- [9] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective LSTMs for target-dependent sentiment classification," 2015, <https://arxiv.org/abs/1512.01100>.
- [10] J. Zeng, X. Ma, and K. Zhou, "Enhancing attention-based LSTM with position context for aspect-level sentiment classification," *IEEE Access*, vol. 7, pp. 20462–20471, 2019.
- [11] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proceedings of the ICLR*, San Diego, CA, USA, May 2015.
- [12] V. Mnih, N. Heess, A. Graves et al., "Recurrent models of visual attention," in *Proceedings of the Advances in Neural Information Processing Systems*, Montreal, Canada, December 2014.
- [13] Y. M. Cui, Z. P. Chen, S. Wei et al., "Attention-over-attention neural networks for reading comprehension," 2016, <https://arxiv.org/abs/1607.04423>.
- [14] K. M. Hermann, T. Kočiský, E. Grefenstette et al., "Teaching machines to read and comprehend," 2015, <https://arxiv.org/abs/1506.03340>.

- [15] O. Wallaert and F. Frasincar, "A hybrid approach for aspect-based sentiment analysis using a lexicalized domain ontology and attentional neural models," in *The Semantic Web*, Springer, Cham, Switzerland, 2019.
- [16] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Austin, TX, USA, November 2016.
- [17] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 214–224, Austin, TX, USA, November 2016.
- [18] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, Australia, August 2017.
- [19] B. Liu, X. An, and J. X. Huang, "Using term location information to enhance probabilistic information retrieval," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Santiago, CL, USA, August 2015.
- [20] S. Gu, L. Zhang, Y. Hou et al., "A position-aware bidirectional attention network for aspect-level sentiment analysis," in *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 774–784, Santa Fe, New Mexico, USA, August 2018.
- [21] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proceedings of the Empirical Methods in Natural Language Processing*, pp. 79–86, Philadelphia, PA, USA, July 2002.
- [22] D. T. Vo and Y. Zhang, "Target-dependent twitter sentiment classification with rich automatic features," in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, AAAI Press, Buenos Aires, Argentina, July 2015.
- [23] M. S. Akhtar, D. Gupta, A. Ekbal, and P. Bhattacharyya, "Feature selection and ensemble construction: a two-step method for aspect based sentiment analysis," *Knowledge-Based Systems*, vol. 125, no. 1, pp. 116–135, 2017.
- [24] J. Wagner, P. Arora, S. Cortes et al., "DCU: aspect-based polarity classification for SemEval task 4," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval-2014)*, Dublin, Ireland, August 2014.
- [25] N. Kaji and M. Kitsuregawa, "Building lexicon for sentiment analysis from massive collection of HTML documents," in *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, Prague, Czech Republic, June 2007.
- [26] V. Prez-Rosas, "Learning sentiment lexicons in Spanish," in *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, Istanbul, Turkey, May 2012.
- [27] S. M. Mohammad, S. Kiritchenko, and X. Zhu, "Nrc-Canada: building the state-of-the-art in sentiment analysis of tweets," 2013, <https://arxiv.org/abs/1308.6242>.
- [28] H. H. Dohaiha, P. Prasad, A. Maag, and A. Alsadoon, "Deep learning for aspect-based sentiment analysis: a comparative review," *Expert Systems with Applications*, vol. 118, pp. 272–299, 2018.
- [29] R. Socher, J. Pennington, E. H. Huang et al., "Semi-supervised recursive autoencoders for predicting sentiment distributions," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing. DBLP*, Edinburgh, UK, July 2011.
- [30] R. Socher, A. Perelygin, Y. W. Jean et al., "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1631–1642, Seattle, WA, USA, October 2013.
- [31] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term, memory networks," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, vol. 5, no. 1, p. 36, Beijing, China, July 2015.
- [32] S. Ruder, P. Ghaffari, and J. G. Breslin, "A hierarchical model of reviews for aspect-based sentiment analysis," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Austin, TX, USA, November 2016.
- [33] M. Zhang, Y. Zhang, and D. T. Vo, "Gated neural networks for targeted sentiment analysis," in *Proceedings of the AAAI*, pp. 3087–3093, Phoenix, AZ, USA, February 2016.
- [34] D. Zeng, K. Liu, S. Lai et al., "Relation classification via convolutional deep neural network," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 2335–2344, Dublin, Ireland, August 2014.
- [35] J. Pennington, R. Socher, and C. Manning, "Glove: global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, October 2014.
- [36] B. Huang, Y. Ou, and K. M. Carley, "Aspect level sentiment classification with attention-over-attention neural networks," in *Social, Cultural, and Behavioral Modeling*, Springer, Cham, Switzerland, 2018.
- [37] S. Wu, Y. Xu, F. Wu, Z. Yuan, Y. Huang, and X. Li, "Aspect-based sentiment analysis via fusing multiple sources of textual knowledge," *Knowledge-Based Systems*, vol. 183, Article ID 104868, 2019.
- [38] C. D. Manning, P. Raghavan, and H. Schtze, *Introduction to Information Retrieval 1*, Cambridge University Press, Cambridge, UK, 2008.