Hindawi Wireless Communications and Mobile Computing Volume 2021, Article ID 5534615, 13 pages https://doi.org/10.1155/2021/5534615



Research Article

Aspect-Level Sentiment Analysis Approach via BERT and Aspect Feature Location Model

Guangyao Pang¹, Keda Lu, Xiaoying Zhu, Jie He, Zhiyi Mo, Zizhen Peng¹, and Baoxing Pu

¹School of Data Science and Software Engineering, Wuzhou University, Wuzhou 543002, China

Correspondence should be addressed to Zizhen Peng; pengzizhen@qq.com

Received 27 February 2021; Revised 13 June 2021; Accepted 16 July 2021; Published 10 August 2021

Academic Editor: Zhuojun Duan

Copyright © 2021 Guangyao Pang 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.

With the rapid development of Internet social platforms, buyer shows (such as comment text) have become an important basis for consumers to understand products and purchase decisions. The early sentiment analysis methods were mainly text-level and sentence-level, which believed that a text had only one sentiment. This phenomenon will cover up the details, and it is difficult to reflect people's fine-grained and comprehensive sentiments fully, leading to people's wrong decisions. Obviously, aspect-level sentiment analysis can obtain a more comprehensive sentiment classification by mining the sentiment tendencies of different aspects in the comment text. However, the existing aspect-level sentiment analysis methods mainly focus on attention mechanism and recurrent neural network. They lack emotional sensitivity to the position of aspect words and tend to ignore long-term dependencies. In order to solve this problem, on the basis of Bidirectional Encoder Representations from Transformers (BERT), this paper proposes an effective aspect-level sentiment analysis approach (ALM-BERT) by constructing an aspect feature location model. Specifically, we use the pretrained BERT model first to mine more aspect-level auxiliary information from the comment context. Secondly, for the sake of learning the expression features of aspect words and the interactive information of aspect words' context, we construct an aspect-based sentiment feature extraction method. Finally, we construct evaluation experiments on three benchmark datasets. The experimental results show that the aspect-level sentiment analysis performance of the ALM-BERT approach proposed in this paper is significantly better than other comparison methods.

1. Introduction

E-commerce is a thriving industry with increasing importance to the global economy. Particularly with the rapid development of social media, more and more users begin to express their sentiments on various online platforms. These comments reflect the sentiments of users and consumers and provide sellers and governments with a lot of valuable feedback on the quality of goods or services [1–3]. For example, before purchasing a product, the users can browse a large number of comments about the product on the e-commerce platform to determine whether the product is worth buying. Similarly, governments and companies can collect a large number of public comments directly from the Internet and analyze users' opinions and satisfaction from them, so as to meet their needs. Therefore, as a basic and key work of natu-

ral language processing (NLP), sentiment analysis has attracted widespread attention from the theoretical and practical circles [4]. However, the classic sentiment analysis task can only determine the users' sentiment polarities (e.g., positive, negative, and neutral) of the product or event from the entire sentences and cannot determine the sentiment polarity of a particular aspect of the sentence, let alone identify the multiple sentiments existing in a single sentence. In contrast, aspect-based sentiment analysis is a more fine-grained classification task, which can identify the sentiment polarities of multiple aspects in a sentence. Specifically, this scene is shown in Figure 1, where a sentence as a whole has an overall sentiment, and there are also multiple aspect-level sentiments. We can observe from the comment text: in the "It didn't come with any software installed outside of windows media, but for the price, I was very pleased with the condition

²Department of Mechanical and Electrical Engineering, Wuzhou Vocational College, Wuzhou 543002, China

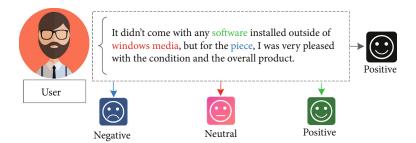


FIGURE 1: An example of consumer review with three aspect terms. Black represents sentence level sentiment analysis, while red, green, and blue represent the sentiment of corresponding aspect word, respectively.

and the overall product," the emotional polarity of "software" is negative, "Windows Media" is neutral, and "price" and "very satisfied" are positive. Among them, these different sentiment words are called aspect words.

In recent years, researchers have proposed various methods to complete aspect-level sentiment analysis. Among them, the supervised machine learning algorithm has the best effect [5–7]. However, such statistical-based methods rely on carefully designed manual features on large-scale datasets, resulting in a lot of waste of manpower and time [8, 9]. The neural network model can automatically learn the low-dimensional representation of reviews without relying on artificial feature engineering. This feature allows neural networks to be used for aspect-level sentiment analysis tasks and has attracted the attention of researchers [10, 11].

Unfortunately, the existing methods mainly use recurrent neural network (RNN) [12] or convolutional neural network (CNN) [6] to mine the semantic information of aspect word and its context, which is easy to ignore the fact that they are insensitive to the location of key components [10, 13]. Researchers have proved that the emotional polarity of the aspect word is highly correlated with the word order of the aspect word information [4], which means that the emotional polarity of aspect words is more easily affected by the context of aspect words with similar distance [14]. Besides, the neural network is difficult to capture long-term dependencies between aspect words and context, which causes a loss of valuable information. Even if the attention mechanism [15] can be positioned in the right context to alleviate this problem, but the problem still remains and limits their performance.

For the sake of solving the aforementioned problems, on the basis of Bidirectional Encoder Representations from Transformers (BERT) [16], this paper establishes an aspect-level sentiment analysis approach based on BERT and aspect feature location model (i.e., ALM-BERT). The core idea of the ALM-BERT approach is to recognize the emotion of different aspect words in the text, consider the contextual interaction information of aspect words, and reduce the interference of irrelevant words, thus forming an effective aspect-based sentiment analysis framework. The main contributions of this paper are as follows:

(i) Based on the pretrained general model BERT, we have constructed a multiangle text vectorization mechanism that can obtain high-quality contextual

information representation and aspect information representation. In addition, we also construct an aspect-based sentiment feature extraction method. This method utilizes an encoder based on the multihead attention mechanism to learn the expression features of the aspect words and the interactive information of the aspect word context, which can effectively distinguish different sentences and different contributions of different aspect words

- (ii) We construct an aspect feature location model to capture the aspect information when modeling sentences and integrate the complete information of the aspect words into the interaction semantics. This model can effectively reduce the influence of noise words that have nothing to do with aspect words and improve the integrity of aspect word information
- (iii) We conduct aspect-level sentiment analysis evaluation experiments on three benchmark datasets. The experimental results show that the accuracy and macro-F1 score of our proposed model (i.e., ALM-BERT) on the Restaurant dataset are 13.66% and 29.76% higher than those of the baseline MGAN models, respectively. At the same time, the accuracy of the ALM-BERT model on comment texts of different lengths is also better than other comparison methods. This shows that the ALM-BERT approach can better mine the users' aspect-level sentiments

We organize the remainder of this paper as follows: in Section 2, we introduce some related works on aspect-based sentiment analysis task briefly, the problem formulation is described in Section 3, we present the proposed model and its training process in detail in Section 4, experimental evaluation and result analysis are given in Section 5, and we conclude the paper and briefly discuss the future work in Section 6.

2. Related Works

The core goal of aspect-based sentiment analysis is to recognize the sentiment polarity of different aspect words in a given text, which means that it can mine more fine-grained sentiments, so it has become a research hotspot in the current sentiment analysis field. Currently, aspect-based sentiment analysis methods are mainly divided into two categories:

classic aspect-based sentiment analysis methods and neural network-based sentiment analysis methods.

2.1. Classical Aspect-Based Sentiment Analysis Methods. In the field of aspect-based sentiment analysis, early research mainly focused on traditional machine learning methods, including rule-based methods [17] and statistical-based methods [18]. These studies generally relied on laborious manual annotation and feature engineering and then employed traditional machine learning to establish a sentiment classifier [19]. For example, Qiu et al. [20] analyzed the relationship between aspect words and sentiment polarity according to the grammatical features. Analogously, Liu et al. [21] proposed a word alignment model to identify aspect words and sentiment polarity based on grammatical information. Subrahmanian and Reforgiato [22] proposed a comprehensive framework that fully considered the information of adjectives, verbs, and adverbs. Jing et al. [23] presented a topic modeling method and utilized grammatical features to help separate aspect words and sentiment words. Wu et al. [24] introduced the concept of phrase dependency parsing and took phrase fragments as an important part of identified polarity of sentiment. Zhao et al. [19] proposed a novel method, which decided the sentiment polarity of aspect words according to the grammatical features of the words related to aspect words. Kiritchenko et al. [25] adopted a support vector machine algorithm based on n-gram features, parse features, and lexical features.

Although these methods have achieved certain results, they rely too much on manual annotation and feature engineering, which means that there are performance bottlenecks that are difficult to break through.

2.2. Neural Network-Based Sentiment Analysis Methods. Different from the traditional methods mentioned above, the neural network can automatically learn continuous and low-dimensional representation features from the text without relying on manual feature engineering. In other words, the neural network can effectively solve the problems of excessive dependence on manual annotation and feature engineering in the above-mentioned traditional methods. Therefore, more and more researchers have constructed a series of aspect-based sentiment analysis methods based on neural networks. Tang et al. [26] constructed a Target-Dependent Long Short-Term Memory (TD-LSTM) model based on two LSTM networks, which concatenates the left context representation and right context representation of the aspect as the final context representation for predicting the sentiment. Moreover, neural network models based on attention mechanism, which was proposed in machine translation task, have been successfully applied in aspectbased sentiment analysis. Wang et al. [27] designed an LSTM model based on the attention mechanism, which can focus on the important parts related to aspect words in a sentence. Chen et al. [28] utilized a bidirectional LSTM and multiple attention mechanism to pick up important features to predict the final sentiment. Ma et al. [29] employed an interactive attention mechanism to obtain the context representation and aspect word representation. Ou et al.

[4] established a neural network with an attention-overattention model based on LSTM. The neural network models aspect words and context at the same time, which can mine important auxiliary information in aspect words and context.

Recently, the pretraining model BERT, which can not rely on labeled data, has attracted the attention of academia and industry. Specifically, the BERT model can train a general model with preliminary natural language features only by using a large amount of unlabeled text [16]. Of course, the BERT model needs to be further fine-tuned using labeled data to complete the training of the predictor. For instance, Song et al. [30] regarded the BERT model as the embedding layer to obtain the vector representation of context and have achieved good results. Qui et al. [31] proposed a novel auxiliary sentence construction method and transformed aspect-based sentiment classification task into a sentence-pair classification task. Gao et al. [32] constructed a BERT-based encoder to determine the sentiment polarity of aspect words.

The above-mentioned research has made some progress, but there are still many problems. For example, the standard BERT model only provides local context information [33], ignoring the differences in the emotional polarity and importance of words in different aspects. In addition, most of these existing studies do not explicitly model the complete information of the aspect words in a sentence. However, other researchers have indicated that the irrelevant information to aspect words would severely degrade the performance of the model [18]. Therefore, it remains a challenging task to identifying the sentiment polarity of different aspects.

3. Problem Formulation

Aspect-based sentiment analysis refers to the process of outputting the sentiment polarity of each aspect word in a sentence with a sentence and some predefined aspect words as input data. We will utilize some real comment examples to illustrate aspect-level sentiment analysis tasks.

Obviously, as shown in Table 1, each example sentence contains two aspect terms, and each aspect term has four different sentiment polarities (i.e., positive, neutral, negative, and conflict). The aspect-based sentiment analysis can be defined as follows:

Definition 1. Formally, we give a comment sentence $S = \{w_1, w_2, \dots, w_n\}$, where n is the total number of words in S. $A = \{a_1, \dots, a_i, \dots, a_m\}$ with length m represents an aspect vocabulary of length m, where a_i denotes the ith aspect word in aspect vocabulary A, and A is a subsequence of sentence S. $P = \{p_1, \dots, p_j, \dots, p_C\}$ denotes the candidate sentiment polarities, where C denotes the number of categories of sentiment polarity and the p_j is the jth sentiment polarity.

Problem 2. The goal of the aspect-based sentiment analysis model is to predict the most likely sentiment polarity of

Comments	Asmost	Sentiment polarities			
Comments	Aspect	Positive	Negative	Neutral	Conflict
All the constant into the interior decorate of the control of the	Interior decoration	✓			
All the money went into the interior decoration, none of it went to the chefs.	Chefs		✓		
Great Indian food and the service is incredible.	Indian food	✓			
	Service	✓			
The lobster sandwich is \$24, and although it was good,	Lobster sandwich				✓
it was not nearly enough to warrant that price.	Price		✓		

TABLE 1: Some examples of aspect-based sentiment analysis.

specific aspect word in a sentence, which can be formulated as follows:

input :
$$\begin{cases} S = \{w_1, w_2, \dots, w_n\}, \\ A = \{a_1, \dots a_i, \dots a_m\}, \end{cases}$$
output : $p_k = \phi_{\max} \left(a_i, p_j \mid S\right),$
constraints : $A \in S, m \in [1, N],$

$$(1)$$

where ϕ represents a function that quantifies the degree of matching between the aspect word a_i and the sentiment polarity p_j in the sentence S. Finally, the model outputs the sentiment polarity with the highest matching degree to be the classification result. The notation and their description in this model are summarized in Table 2.

4. Our Proposed Model

In word-level sentiment analysis and sentence-level sentiment analysis, the details of sentiment analysis will be covered up, and it also cannot accurately reflect people's fine-grained emotional expressions. In order to conduct a more complete sentiment analysis and discover the sentiment information expressed by different angles (i.e., aspects) of text reviews, this paper proposes an aspect-location model based on BERT for aspect-based sentiment analysis (i.e., ALM-BERT), which can mine different aspects of sentiment in comment details, to avoid incorrect results in real-world applications such as recommendation systems and question answering systems. The overall framework of the ALM-BERT approach is shown in Figure 2, which mainly includes four parts: multiangle text vectorization mechanism, important feature extraction model, fusion layer, and sentiment predictor.

Firstly, we employ the pretrained model BERT to generate a high-quality word vector of sequence, which provides effective support for subsequent steps (such as Section 4.1). Then, we build a new feature extractor (i.e., important feature extraction model) of multihead attention mechanism and position feedforward network to extract important context and target information (such as Section 4.2.1) and build an aspect feature location model, which can select information related to aspect words from context feature representation (such as Section 4.2.2). Finally, on the basis of fusing the

TABLE 2: The used symbols and their description.

Symbols	Description
S	The comment sentence
A	The aspect vocabulary
C	The number of categories of sentiment polarity
al	The length of aspect words
p_j	The j th alternative sentiment polarity
E_c	The context representation
E_a	The aspect representation
c_{cc}	The long-term dependent information of the context
t_{ca}	The context-aware information to aspect word
h_{af}	The important features of aspect word
h_{cm}	The final interaction hidden state of a context interaction
h_{am}	The final interaction hidden state of context and aspect words
$f_s(.)$	The attention score function
$f_e(.)$	The energy function

context and relevant important information related to the target, we use a sentiment predictor at the aspect level to predict the probability of different emotion polarities (such as Section 4.3).

4.1. Multiangle Text Vectorization Mechanism. The word embedding maps each word to a high-dimensional vector space, which mainly assists machines in understanding natural language. Its mainstream methods include Word2vec and Glove. Both of these methods belong to context-based word embedding models and have achieved good performance in aspect-level sentiment analysis tasks. However, previous research has already demonstrated that these two word embedding models cannot capture the enough information in the text [34], which leads to poor classification accuracy and reduces the performance of the aspect-based sentiment analysis model. Therefore, a high-quality word embedding model has an important influence on improving the accuracy of classification results [35].

The key of aspect-level sentiment analysis is to understand natural language processing effectively. This idea usually highly relies on large-scale high-quality annotation text. Fortunately, BERT is a language pretraining model that can

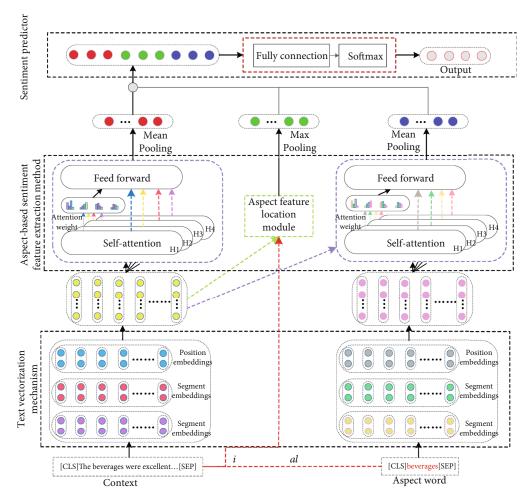


FIGURE 2: The overall framework of ALM-BERT.

effectively use unlabeled text. The model utilizes a method of randomly covering some words, utilizes a multilayer twoway converter encoder to extract a general natural language recognition model from a large amount of unlabeled text, and further uses a small amount of labeled data for finetuning to generate high-quality text feature vectors. Inspired by this idea, the ALM-BERT approach mentioned in this paper adds special word breaker tags [CLS] and [SEP] at the beginning and end of a given word sequence, respectively, and finally divides a given sequence into different segments. That is, the word embedding vector input in this way includes generating vectors such as token embeddings, segmentation embedding, and position embedding for different segments. In the ALM-BERT approach, we convert the comment text and aspect word into the form of "[CLS] + comment text + [SEP]" and "[CLS] + target + [SEP]", respectively. Finally, we obtain the context representation E_c and aspect representation E_a :

$$\begin{split} E_c &= \left\{ we_{\text{[CLS]}}, we_1, we_2, \cdots, we_{\text{[SEP]}} \right\}, \\ E_a &= \left\{ ae_{\text{[CLS]}}, ae_1, ae_2, \cdots, ae_{\text{[SEP]}} \right\}, \end{split} \tag{2} \end{split}$$

where $we_{[\text{CLS}]}$, $ae_{[\text{CLS}]}$ denotes the vector of classification mark [CLS], and the $we_{[\text{SEP}]}$ and $ae_{[\text{SEP}]}$ expressions the vector of separator [SEP].

4.2. Aspect-Based Sentiment Feature Extraction Method. In order to extract the implicit features of the aspect words and their context and to consider the auxiliary information contained in the aspect words, we design an aspect-based sentiment feature extraction method inspired by a transformer encoder [36]. The basic idea of this method is to integrate the information of aspect words and context and to model the interaction between context and target words. Furthermore, we hold the opinion that the accuracy of sentiment classification can be improved by capturing the feature information of aspect words in context.

4.2.1. Important Feature Extraction Model. A transformer encoder is a novel feature extractor based on multihead attention mechanism and position-wise feed-forward networks, which can learn different important information in different feature representation subspaces. Not only that, the transformer encoder can also directly capture the long-term dependencies in the sequence, and it is easier to parallelize than recurrent neural network and convolutional neural

Require: the context representation e_c ; the position i of aspect words in a sentence; the length al of aspect words; the batch size bs;

- 1: repeat
- 2: **for** each $e_c \in bs$ **do**
- 3: Select lines (i + 1 and i + 1 + al) of e_c to obtain aspect feature af;
- 4: Calculate the most important features *AF* according to Eq. (8);
- Apply the dropout operation to all the important features to get the h_{af} ;
- 6: end for;
- 7: until Accuracy and macro-F1 tend to be stable.

ALGORITHM 1: Aspect feature location algorithm.

networks, which greatly reduces the training time. Based on the same principle, we design the important feature extraction model as shown below.

Specifically, we first construct a multihead attention mechanism composed of multiple self-attention mechanisms. This mechanism employs different heads to capture the implicit information of the text from different aspects and can achieve high-performance parallel computing independently of RNN and CNN. Among them, the different aspects include query sequence (Q), key-value pairs (K and V). The attention score $f_s(.)$ in the self-attention mechanism is calculated as follows:

$$f_s(Q, K, V) = \sigma(f_e(Q, K))V, \tag{3}$$

where $\sigma(.)$ stands for the normalized exponential function, and $f_e(.)$ is the energy function to learn the correlation features between K and Q, which can be calculated by using the following formula:

$$f_e(Q, K) = \frac{QK^T}{\sqrt{d_k}},\tag{4}$$

where $\sqrt{d_k}$ denotes the scale factor, and the d_k is the dimension of the query and key vectors.

The attention score of multihead attention mechanism $f_{\it mh}(.)$ is obtained by concatenating attention score of self-attention mechanism:

$$f_{mh}(Q, K, V) = \left[a^1; a^2; \dots; a^i; \dots; a^{n-\text{head}}\right] W_d,$$

$$a^i = f_c^i(Q, K, V),$$
(5)

where a^i represents the *i*th attention score, [:] denotes concatenates of the vector, and W_d is the weight matrix.

Secondly, we input the context representation and aspect representation into the multihead attention mechanism to capture the long-term dependencies of the context and decide which context is crucial for determining the sentiment of the aspect word, which is shown in the following:

$$c_{cc} = f_{mh}(E_c, E_c),$$

$$t_{ca} = f_{mh}(E_c, E_a),$$
(6)

where c_{cc} and t_{ca} denote the long-term dependent information of the context and the context-aware information to aspect word, respectively.

Then, we utilize the transform encoder to take c_{cc} and t_{ca} as the input of the position-wise feed-forward network and dig out the hidden states h_c and h_a . Formally, the position-wise feed-forward networks PFN, h_c , and h_a are defined as follows:

$$\begin{split} h_c &= \mathrm{PFN}(c_{cc}), \\ h_a &= \mathrm{PFN}(t_{ca}), \end{split} \tag{7} \\ \mathrm{PFN}(h) &= \zeta(hW_1 + b_1)W_2 + b_2, \end{split}$$

where $\zeta(.)$ expressions the rectified linear unit, b_1 and b_2 represent biases, and W_1 and W_2 denote learnable weights.

Finally, after the mean pooling operation of h_c and h_a , we get the final hidden states h_{cm} and h_{am} .

4.2.2. Aspect Feature Location Model. The above-mentioned important feature extraction model captures the long-term dependence of the context and also generates the interactive semantic information between the aspect word and the context. On this basis, in order to further highlight the importance of different aspect words, we build an aspect feature positioning model based on the maximum pooling function (which is shown in Algorithm 1). This model divides the extracted aspect words and their context hiding features into multiple regions (i.e., line 3) and selects the maximum value in each region to represent the region (i.e., lines 4-5). In this way, the model can also locate core features and reduce the influence of noise words that are not related to aspect words, thereby improving the integrity of aspect word information. In other words, capturing aspect features and the different importance of aspect features can further improve the accuracy of aspect-level emotion classification.

Specifically, combining the characteristics of the position and length of the aspect word, the feature location algorithm extracts the most important relevant information of the aspect word af from the context representation e_c . Moreover, We applied max-pooling to af to get the most important features AF.

$$AF = Maxpooling(af, dim = 0).$$
 (8)

Afterwards, we perform a dropout operation on AF and obtain the important features h_{af} of the aspect word in the context representation.

4.3. Sentiment Predictor. One of the cores of ALM-BERT is to utilize multiple self-attention mechanisms to obtain multiangle text hidden expression features, and after processing by aspect feature positioning models, we have obtained a wealth of aspect-level auxiliary features and contextual interaction of aspect word information. In order to effectively utilize these complete and rich features, this paper uses fully connection layer to fuse and preprocess the features in advance and uses the softmax function to map the features to the [0,1] interval, so as to achieve effective mapping from features to sentiment classification. Specifically, we concatenate the h_{cm} , h_{am} , and h_{af} first to obtain the comprehensive representation r, which is shown as follows:

$$r = \left[h_{cm} ; h_{am} ; h_{af} \right]. \tag{9}$$

Subsequently, we use a linear function to preprocess the data of r, as shown in the following:

$$x = W_{u}r + b_{u}, \tag{10}$$

where W_u represents the weight matrix, and b_u denotes the bias.

At last, we utilize a softmax function to compute the probability Pr that the sentiment polarity of the aspect word *a* in a sentence is *p*, as shown in the following:

$$\Pr\left(a=p\right) = \frac{\exp\left(x_p\right)}{\sum_{i=1}^{C} \exp\left(x_i\right)},\tag{11}$$

where *C* denotes the number of categories of sentiment polarity.

On the whole, the ALM-BERT approach, which is proposed in this paper, is an end-to-end computing process. Moreover, in order to optimize the parameters of the ALM-BERT approach, so as to minimize the loss between the predicted sentiment polarity y and the correct sentiment polarity \hat{y} , we adopt cross-entropy with L2 regularization as the loss function to train our model, which is defined as

$$loss = -\sum_{i=1}^{D} \sum_{i=1}^{C} y_i^j \log \hat{y}_i^j + \lambda \|\theta\|^2,$$
 (12)

where D means all training data, and j and i denote the index of a training data sample and a sentiment class, respectively. λ represents the factor for L2 regularization, and θ denotes the parameter set of the model.

5. Experimental Evaluation

For the sake of evaluating the rationality and effectiveness of the ALM-BERT approach, this section describes the details of experiment settings and designs comparative experiments. Moreover, we also analyze the experimental results.

TABLE 3: The statistical information of datasets.

Dataset	Overall	Positive	Neutral	Negative	Conflict
Laptop-train	2373	994	870	464	45
Laptop-test	654	341	128	169	16
Restaurant-train	3699	2164	807	637	91
Restaurant-test	1134	728	196	196	14
Twitter-train	6248	1561	3127	1560	_
Twitter-test	692	173	346	173	_

- 5.1. Datasets. For our experiments, we conduct experiments on three public English review datasets. The statistical information of these datasets is illustrated in Table 3. Among them, in the Restaurant and Laptop datasets provided by SemEval 2014 [37], each sentence contains some aspect words and the corresponding emotional polarity (polarity is marked as positive, negative, neutral, and conflict); in the twitter dataset collected by Tan et al. [38], users' comments are marked with emotional polarity, and the emotional polarity is positive, negative, and neutral, respectively. These three datasets are currently popular review datasets, which have been widely used in aspect-based sentiment analysis tasks.
- 5.2. Baselines and Evaluation Metrics. In order to verify the effectiveness of our model, we compare the ALM-BERT approach with many popular aspect-based sentiment analysis models, as listed in the following:
 - (i) TD-LSTM [26] is a classic model, which improves the accuracy of classification by integrating the correlation information between aspect words and context into the LSTM-based classification model
 - (ii) ATAE-LSTM [27] is a classification model that attaches the embedded representation of aspect words to the embedded representation of sentence as input and then applies the attention mechanism to calculate the weight
 - (iii) *MemNet* [39] is a data-driven model that utilizes multiple attention-based computational layers to capture the importance of each context word
 - (iv) IAN [29] proposes interactive attention networks to model aspect words and context separately and generate the representations for targets and contexts
 - (v) RAM [28] constructs a framework based on multiattention mechanism, as to capture the long-distance features in the text and enhance the representation ability of the model
 - (vi) TNet [40] utilizes bidirectional LSTM to generate the hidden representation of context and aspect words and then utilizes a CNN layer to extract important features from the hidden representation instead of the attention mechanism

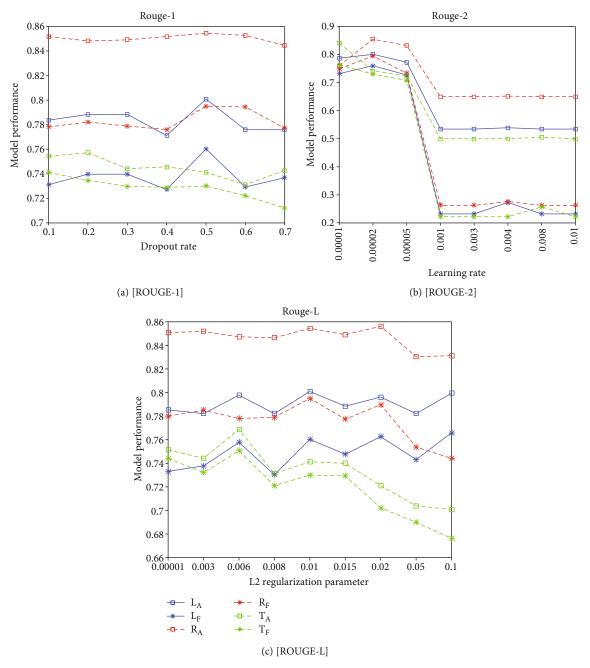


FIGURE 3: The scores of ROUGE on parameter optimization.

- (vii) Cabasc [41] proposes two kinds of attention enhancement mechanisms to focus on aspect words and context, respectively, and comprehensively considers the relevance between context and aspect words
- (viii) AOA [4] constructs an attention-over-attention model to associate sentiment words with aspect words. Moreover, the attention-over-attention model automatically generates mutual attentions from aspect-to-text and text-to-aspect
- (ix) MGAN [42] proposes a multigrained attention model to capture the interactive information

- between aspect words and context from coarse to fine
- (x) AEN-BERT [30] is a model based on attention mechanism and BERT and shows good performance in aspect-based sentiment analysis tasks
- (xi) *BERT-base* is an aspect-based sentiment analysis model based on pretrained BERT, which has a full connection layer and a softmax layer for classification

For the sake of measuring the performance of the model fairly, we extend the AOA, IAN, and MemNet models by replacing the embedding layer of these models with BERT,

Word embedding	Models	Laptop		Restaurant		Twitter	
		Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
Embedding	LSTM	0.6144	0.4401	0.7304	0.533	0.6474	0.6058
	ATAE-LSTM	0.6019	0.4909	0.7375	0.5725	0.6864	0.6501
	Cabasc	0.6301	0.5297	0.7241	0.5245	0.6171	0.5657
	IAN	0.6191	0.4671	0.7268	0.4897	0.6618	0.6251
	MGAN	0.5878	0.4264	0.7179	0.4974	0.6373	0.5856
	MemNet	0.7915	0.7576	0.8241	0.7313	0.6936	0.6772
	RAM	0.5956	0.4308	0.7152	0.4656	0.6358	0.5998
	TNet	0.7022	0.6404	0.7688	0.6269	0.6994	0.6827
BERT	IAN-BERT	0.7696	0.7179	0.808	0.722	0.7269	0.7048
	AOA-BERT	0.7774	0.7407	0.7341	0.7173	0.7341	0.7173
	MemNet-BERT	0.7915	0.7576	0.8241	0.7313	0.6936	0.6772
	AEN-BERT	0.7947	0.7544	0.8125	0.7069	0.7168	0.7052
	BERT-base	0.768	0.7288	0.8375	0.7613	0.7442	0.7271
	LCF-bert	0.7837	0.7441	0.8509	0.7894	0.7254	0.7113
	ALM-BERT	0.8009	0.7603	0.8545	0.795	0.7413	0.73

Table 4: Some examples of aspect-based sentiment analysis.

to obtain AOA-BERT, IAN-BERT, and MemNet-BERT models. The structure of the rest models is consistent with those described in the corresponding paper.

In addition, in order to objectively evaluate the performance of the ALM-BERT model, similar to existing aspect-level sentiment analysis tasks, we use macro-F1 score (F1) and accuracy (Acc) as evaluation indicators.

Accuracy (Acc) is defined as

$$Acc = \frac{SC}{N}, \tag{13}$$

where SC denotes the number of samples correctly classified, and *N* represents the total number of samples. Generally, the higher the accuracy, the better the performance of the model.

In addition, macro-F1 is used to truly reflect the performance of the model, which is the weighted average of precision and recall. The macro-F1 is calculated according to the following formula:

$$\Pr e_{C_{i}} = \frac{T_{C_{i}}}{T_{C_{i}} + \operatorname{FP}_{C_{i}}},$$

$$R_{C_{i}} = \frac{T_{C_{i}}}{T_{C_{i}} + \operatorname{FN}_{C_{i}}},$$

$$\operatorname{macro-}F1 = \frac{1}{C} \left\{ \sum_{i=1}^{C} \left\{ \frac{\left(2 * \operatorname{Pr} e_{C_{i}} * R_{C_{i}}\right)}{\left(\operatorname{Pr} e_{C_{i}} + R_{C_{i}}\right)} \right\} \right\},$$
(14)

where T represents the number of samples correctly classified as sentiment polarity i, FP denotes the number of samples incorrectly classified as sentiment polarity i, FN represents the number of samples whose sentiment polarity i is misclassified as other sentiment polarities, C denotes the number of categories of sentiment polarity, $\Pr{e_{C_i}}$ indicates

the precision of sentiment polarity i, and R_{C_i} denotes the recall of sentiment polarity i. In our experiment, for a more comprehensive evaluation of the performance of our model, we divided the categories of sentiment polarity into $3C = \{\text{positive, neutral, negative}\}$ and $4C = \{\text{positive, neutral, negative}\}$ and $4C = \{\text{positive, neutral, negative, conflict}\}$.

5.3. Parameter Optimization. The training process of the ALM-BERT model mainly introduces BERT to generate vector representations of context and aspect words. Therefore, we use BERT's standard parameter BERT_{BASE} to complete the model training, that is, the number of conversion models, the number of hidden neurons, and the number of self-attention heads are 12, 768, and 12, respectively. Furthermore, we have optimized the training process of the model as follows.

The dropout [43] refers to the probability of discarding some neurons during the training process of neural network, which is used to enhance the generalization ability of the model. We initialize the value of dropout to 0.3 and then search for the optimal value at intervals of 0.1. As shown in Figure 3(c), the experimental results demonstrate that when the dropout is 0.5, the ALM-BERT has the best accuracy and F1 value on the three datasets.

The learning rate determines whether and when the objective function converges to the local minimum. In our experiments, we use the Adam optimization algorithm to update the parameters of the model and explore the best learning rate parameters in the range of $[10^{-5}, 0.1]$. As shown in Figure 3(c), when the learning rate is $2 * 10^{-5}$, ALM-BERT has the best performance.

The L2 regularization parameter is a hyperparameter, which can prevent the model from overfitting. According to the results of Figure 3(c), the ALM-BERT performs best when the value of L2 regularization parameter is set to 0.01. Meanwhile, we initialize model weights by Glorot

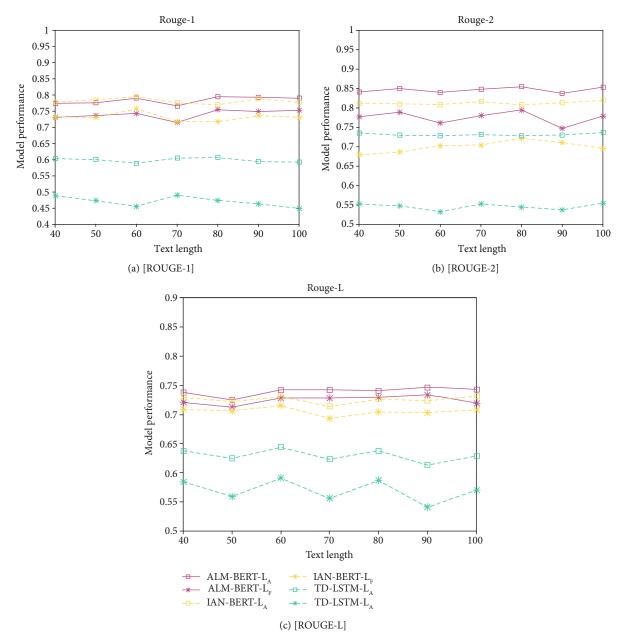


FIGURE 4: The scores of ROUGE on different lengths of the source text.

initialization [44] and set the batch size to 16 and train a total of 10 epochs.

5.4. Evaluation Experiment of All Comparison Methods. As shown in Table 4, the results of sentiment classification when sentiment polarity C=3. We can easily observe from the experimental results that the accuracy and macro-F1 of BE RT_{BASE} are significantly higher than those of Glove and Word2vec based models. Particularly for Restaurant dataset, the accuracy and macro-F1 of ALM-BERT are 12.77% and 30.97% higher than those of the classical IAN model, respectively. This shows that in the field of NLP, the introduction of BERT to build a pretrained word embedding model can indeed better express the semantic and grammatical features

of the text. Meanwhile, we find that the ALM-BERT approach presented in this paper achieves the best classification performance on the three datasets.

Specifically, compared with the performance of the AEN-based model on Restaurant dataset, the ALM-BERT can improve the accuracy and macro-F1 by 4.2% and 8.81%. In addition, it is not difficult to find that the classification accuracy and macro-F1 of the ALM-BERT on the Laptop dataset are 3.29%, 3.15% higher than those of the BERT-base model. This proves that our aspect feature location model plays a positive role in aspect-based sentiment analysis.

5.5. Evaluation Experiment for Mining Long-Term Dependencies. For the sake of verifying the performance of

different methods to capture long-term dependencies, we construct a series of verification experiments in texts of different lengths.

As shown in Figures 4(a)–4(c), the ALM-BERT approach obtains higher accuracy and macro-F1 than TD-LSTM on the whole, which means that our transform encoder can simulate the implicit relationship between contexts better than LSTM-based encoder. In addition, compared with AEN, as shown in Figure 1, the prediction accuracy and macro-F1 of the ALM-BERT model in different length sentences are improved by 3.1% and 6.56%, respectively. This shows that ALM-BERT makes better use of aspect information than AEN and reduces the interference of aspect independent information.

To sum up, these experiments reveal that the ALM-BERT can get higher accuracy and macro-F1, which further verifies that the BERT and aspect information is feasible and effective in the task of aspect-based sentiment analysis.

6. Conclusion

In this paper, we establish a transformer encoder based on BERT to capture the long-term dependencies of the context and generate the interactive semantic information between aspect words and context. Then, we propose an aspect feature location model to extract more aspect features from context information. Experiments on several datasets demonstrate that our proposed approach (i.e., ALM-BERT) is superior to other methods. In addition, with the increase of text length, our proposed approach continues to maintain excellent performance. In other words, the ALM-BERT approach is better able to handle long text data and better excavate the users' aspect-level sentiment.

In our proposed approach, we mainly focus on utilizing natural language texts to identify users' sentiment. However, people's way of expression on social platforms has become more abundant. Therefore, we are interested in combining with image processing technology to analyze multimodal data in the future.

Data Availability

For our experiments, we conduct experiments on three public English review datasets. Among them, the Restaurant and Laptop datasets are provided by SemEval 2014; each sentence in those datasets contains some aspect words and corresponding sentiment polarity, which are labeled with positive, negative, neutral, and conflict. The last datasets consist of user comments collected from twitter; the sentiment polarity is labeled with positive, negative, and neutral. These three datasets are currently popular review datasets, which have been widely used in aspect-based sentiment analysis tasks.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Scientific Research Basic Ability Promotion Foundation of Guangxi Universities' Young and Middle-aged Teachers (Grant Nos. 2020KY17018, 2021KY1492, and 2019KY0686), the National Natural Science Foundation of China (Grant No. 61961036), the Guangxi Innovation-Driven Development Special Fund Project (Guike AA18118036), the Industry-University-Research Project of Wuzhou High-tech Zone and Wuzhou University (Grant No. 2020G003), and the Guangxi Natural Science Foundation (No. 2020GXNSFAA238013).

References

- [1] M. E. Mowlaei, M. Saniee Abadeh, and H. Keshavarz, "Aspect-based sentiment analysis using adaptive aspect-based lexicons," *Expert Systems with Applications*, vol. 148, article 113234, 2020.
- [2] Z. Cai and Z. He, "Trading private range counting over big iot data," in 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), pp. 144–153, Dallas, TX, USA, 2019.
- [3] Y. Lin, X. Wang, F. Hao et al., "Dynamic control of fraud information spreading in mobile social networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 6, pp. 3725–3738, 2021.
- [4] O. Yanglan, B. Huang, and K. M. Carley, "Aspect level sentiment classification with attention-over-attention neural networks," in *Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2018*, R. Thomson, C. Dancy, A. Hyder, and H. Bisgin, Eds., vol. 10899 of Lecture Notes in Computer Science, pp. 197–206, Springer, Cham, 2018.
- [5] S. Cortes, U. B. Dasha, J. F. Bogdanova, J. Wagner, P. Arora, and L. Tounsi, "Dcu: aspect-based polarity classification for semeval task 4," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 223–229, Dublin, Ireland, 2014.
- [6] H. Xu, P. W. W. L. Z. Cai, Z. Xiong, and Y. Pan, Generative adversarial networks: a survey towards private and secure applications, ACM Computing Surveys (CSUR), 2021.
- [7] Z. Cai and Z. Xu, "A private and efficient mechanism for data uploading in smart cyberphysical systems," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 766–775, 2018.
- [8] A. Nazir, Y. Rao, L. Wu, and L. Sun, "Issues and challenges of aspect-based sentiment analysis: a comprehensive survey," *IEEE Transactions on Affective Computing*, 2020.
- [9] Y. Lin, Z. Cai, X. Wang, F. Hao, L. Wang, and A. M. V. V. Sai, "Multi-round incentive mechanism for cold startenabled mobile crowdsensing," *IEEE Transactions on Vehicular Tech*nology, vol. 70, no. 1, pp. 993–1007, 2021.
- [10] N. Liu and B. Shen, "Aspect-based sentiment analysis with gated alternate neural network," *Knowledge-Based Systems*, vol. 188, article 105010, 2020.
- [11] Y. Liang, Z. Cai, J. Yu, Q. Han, and Y. Li, "Deep learning based inference of private information using embedded sensors in smart devices," *IEEE Network*, vol. 32, no. 4, pp. 8–14, 2018.
- [12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

- [13] Z. Xiong, Z. Cai, D. Takabi, and W. Li, "Privacy threat and defense for federated learning with non-i.i.d. data in aiot," *IEEE Transactions on Industrial Informatics*, p. 1, 2021.
- [14] P. Zhao, L. Hou, and O. Wu, "Modeling sentiment dependencies with graph convolutional networks for aspect- level sentiment classification," *Knowledge-Based Systems*, vol. 193, article 105443, 2020.
- [15] Y. B. J. F. D. Bahdanau, K. H. Cho, and L. Tounsi, "Neural machine translation by jointly learning to align and translate," in *Proceedings of the 3rd international conference on learning* representations, pp. 1–15, San Diego, USA, 2015.
- [16] M.-W. Chang, L. K. Devlin, and J. K. Toutanova, "Bert: pretraining of deep bidirectional transformers for language understanding," in *Proceedings of the Conference of the North American Chapter of the Association for Computational Lin*guistics, pp. 1–16, Minneapolis, USA, 2018.
- [17] B. L. X. Ding and S. Y. Philip, "A holistic lexicon-based approach to opinion mining," in *Proceedings of the Interna*tional Conference on Web Search and Web Data Mining -WSDM '08, pp. 231–240, New York, NY, USA, 2008.
- [18] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification," in *Proceedings of the 49th annual meeting of the Association for Computational Linguistics: human language technologies*, pp. 151–160, Portland, Oregon, USA, 2011.
- [19] X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid," in *Proceedings of* the 2010 Conference on empirical methods in natural language processing, pp. 56–65, Cambridge, MA, 2010.
- [20] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," *Computational Linguistics*, vol. 37, no. 1, pp. 9–27, 2011.
- [21] Y. Liu, K. Liu, H. L. Xu, and J. Zhao, "Opinion target extraction using partially-supervised word alignment model," in *Proceed*ings of the international joint conferences on artificial intelligence, pp. 2134–2140, Beijing, China, 2013.
- [22] V. S. Subrahmanian and D. Reforgiato, "Ava: adjective-verb-adverb combinations for sentiment analysis," *IEEE Intelligent Systems*, vol. 23, no. 4, pp. 43–50, 2008.
- [23] Jing, Y. Hongfei, X. Zhao, Jiang, and X. Li, "Jointly modeling aspects and opinions with a maxent-lda hybrid," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 56–65, MIT, Massachusetts, USA, 2010.
- [24] Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing Volume 3 EMNLP '09, pp. 1533–1541, Singapore, 2009.
- [25] C. C. S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Nrc-Canada-2014: detecting aspects and sentiment in customer reviews," in *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, pp. 437–442, Dublin, Ireland, 2014.
- [26] X. F. D. Tang, B. Qin, and T. Liu, "Effective lstms for target-dependent sentiment classification," in *Proceedings of the 26th International Conference on Computational Linguistics*, pp. 3298–3307, Osaka, Japan, 2016.
- [27] Y. Wang, M. Huang, L. Zhao, and X. Zhu, "Attention-based lstm for aspect-level sentiment classification," in *Proceedings of the 2016 Conference on empirical methods in natural language processing*, pp. 606–615, Austin, Texas, 2016.

- [28] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *Pro*ceedings of the 2017 Conference on empirical methods in natural language processing, pp. 452–461, Copenhagen, Denmark, 2017
- [29] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," in *Proceed*ings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI), pp. 4068–4074, Melbourne, Australia, 2017.
- [30] Y. Song, J. Wang, T. Jiang, Z. Liu, and Y. Rao, "Targeted sentiment classification with attentional encoder network," in *Artificial Neural Networks and Machine Learning ICANN 2019: Text and Time Series. ICANN 2019*, I. Tetko, V. Kůrková, P. Karpov, and F. Theis, Eds., vol. 11730 of Lecture Notes in Computer Science, pp. 93–103, Springer, Cham, 2019.
- [31] X. Q. C. Sun and L. Huang, "Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence," in Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics, pp. 1–6, Minneapolis, USA, 2019.
- [32] Z. Gao, A. Feng, X. Song, and X. Wu, "Targetdependent sentiment classification with bert," *IEEE Access*, vol. 7, pp. 154290– 154299, 2020.
- [33] Z. Lu, D. Pan, and J.-Y. Nie, "VGCNBERT: augmenting BERT with graph embedding for text classification," in *Advances in Information Retrieval. ECIR* 2020, J. Jose, Ed., vol. 12035 of Lecture Notes in Computer Science, pp. 369–382, Springer, Cham, 2020.
- [34] L.-C. Yu, J. Wang, K. R. Lai, and X. Zhang, "Refining word embeddings using intensity scores for sentiment analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Pro*cessing, vol. 26, no. 3, pp. 671–681, 2018.
- [35] S. Rida-E-Fatima, A. Javed, A. Banjar et al., "A multi-layer dual attention deep learning model with refined word embeddings for aspect-based sentiment analysis," *IEEE Access*, vol. 7, pp. 114795–114807, 2019.
- [36] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," in *Advances in Neural Information Processing Systems* 30: Annual Conference on Neural Information Processing Systems 2017, pp. 5998–6008, Long Beach, CA, USA, 2017.
- [37] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "Semeval-2014 task 4: aspect based sentiment analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 27–35, Dublin, Ireland, 2015.
- [38] C. Tan, D. Tang, M. Zhou, L. Dong, F. Wei, and X. Ke, "Adaptive recursive neural network for target-dependent twitter sentiment classification," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 49–54, Baltimore, Maryland, USA, 2014.
- [39] B. Qin, D. Tang, and T. Liu, "Aspect level sentiment classification with deep memory network," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, pp. 214–224, Austin, Texas, USA, 2016.
- [40] W. Lam, X. Li, L. Bing, and B. Shi, "Transformation networks for target-oriented sentiment classification," in *Proceedings of the 56th annual meeting of the Association for Computational Linguistics*, pp. 946–956, Melbourne, Australia, 2018.
- [41] Q. Liu, H. Zhang, Y. Zeng, Z. Huang, and Z. Wu, "Content attention model for aspect based sentiment analysis," in

- *Proceedings of the 2018 World wide web conference*, pp. 1023–1032, Lyon France, 2018.
- [42] F. Fan, Y. Feng, and D. Zhao, "Multi-grained attention network for aspect-level sentimentclassification," in *Proceedings of the 2018 Conference on empirical methods in natural language processing*, pp. 3433–3442, Brussels, Belgium, 2018.
- [43] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [44] Y. Bengio and X. Glorot, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth Interna-tional Conference on Artificial Intelligence and Statistics*, pp. 1–8, Chia Laguna Resort, Sardinia, Italy, 2010.