ALBERTC-CNN Based Aspect Level Sentiment Analysis

XINGXIN YE1, YANG XU1,2, AND MENGSHI LUO1

1College of Big Data Information Engineering, Guizhou University, Guiyang 550025, China
2Guiyang Aluminum Magnesium Design and Research Institute Company Ltd., Guiyang 550009, China

Corresponding author: Yang Xu (xuy@gzu.edu.cn)

This work was supported in part by the Guizhou University Talent Introduction under Grant 2015-12, and in part by the Guizhou Science and Technology Plan Project (Qkh LHz [2016]) under Grant 7429.

ABSTRACT In order to solve the problem that most aspect level sentiment analysis networks cannot extract the global and local information of the context at the same time. This study proposes an aspect level sentiment analysis model named Combining with A Lite Bidirection Encoder Representation from TransConvs and ConvNets(ALBERTC-CNN). First, the global sentence information and local emotion information in a text are extracted by the improved ALBERTC network, and the input aspect level text is represented by a word vector. Then, the feature vector is mapped to the emotion classification number by a linear function and a softmax function. Finally, the aspect level sentiment analysis results are obtained. The proposed model is tested on two datasets of the SemEval-2014 open task, the laptop and restaurant datasets, and compared with the traditional networks. The results show that compared with the traditional network, the classification accuracy of the proposed model is improved by approximately 4% and 5% on the two sets, whereas the F1 value is improved by approximately 4% and 8%. Additionally, compared with the original ALBERT network, the accuracy is improved by approximately 2%, and the F1 value is improved by approximately 1%.

INDEX TERMS ALBERT, aspect level, ConvNets, sentiment analysis.

I. INTRODUCTION

Sentiment analysis is one of the important research tasks in the field of natural language processing, and it has been a hot research direction. However, most of the current sentiment analysis methods are aimed at the document and sentence levels. These two types of granularity sentiment analysis methods have solved many problems to a certain extent, but they still cannot meet the needs of most of the current applications. Therefore, the development of fine-grained aspect level sentiment analysis methods has been widely studied in the vision field. Aspect level sentiment analysis refers to the sentiment tendency analysis of one or several aspects of a given sentence, such as “The menu is limited, but almost all of the dishes are excellent.” The emotional tendency of this sentence to the aspect item “menu” is negative, but the emotional tendency to the aspect item “dishes” is positive; thus, the aspect level emotional analysis is to analyze different emotional information of different aspects in a sentence.

By analyzing the emotional tendency of different aspects of a comment text in different fields, and applying the analysis results to the field, reasonable help and guidance to the consumer platform or consumers can be provided, such as monitoring the service attitude and consulting experience in an artificial service system, analyzing and controlling the current trend of public opinion, and recommending reasonable and personalized goods or experience items.

Aspect level sentiment analysis methods can be roughly divided into four method categories. The first category includes methods based on rules and dictionaries. Mahalakshmi et al. [1] improved the analysis effect by generating context-aware sentiment dictionaries. Alshari et al. [2] obtained more effective classification results by expanding the non-opinion words in sentiment vocabulary resources. These methods are easy to understand and use, but their accuracy is not high, and there are problems of dimension explosion and gradient disappearance, which limit their wide application in practice. The second category includes machine learning-based methods, such as those proposed by Álvarez-López [3], Mubarok [4], Rafeek [5], which...
use the support vector machine method [6], naive Bayes method [7], and logical regression method [8], respectively, to analyze each sentence part. These methods can improve the accuracy of emotion analysis to a certain extent but need to be combined with complex feature engineering for feature annotation and extraction, which requires much manpower, many material and financial resources, and the methods are not highly mobile, thus requiring heavy training for different tasks. All these problems in machine learning limit the further development of aspect level emotion analysis. The third category includes deep learning-based methods. With the development of convolutional neural networks (CNN) [9] and long short-term memory (LSTM) network [10], the neural networks have achieved excellent results in most tasks in the field of natural language processing. Chen [11] and Xue [12] applied these two network types to aspect level emotion analysis and achieved good results. These methods can effectively solve the problems of gradient explosion and gradient disappearance, but the accuracy needs further improvement. Also, the application of a pre-training model in image processing requires accuracy improvement. With the extensive use of a pre-training model in the field of image, many pre-training models have been developed in natural language processing, including the GPT (Generative Pre-Training) [13], ELMO (Embeddings from Language Models) [14], BERT (Bidirectional Encoder Representations from Transformers) [15], RoBERTa (A Robustly Optimized BERT Pretraining Approach) [16], ALBERT (A lite BERT) [17], and other pre-training language models. The performance of pre-training language models in various tasks has been surprisingly good, so the fourth method category, which includes methods based on a pre-training language model, was developed and has been widely studied and applied to many fields. Hoang et al. [18] used BERT to analyze aspect level emotion and achieved good results. Although this method can accurately extract the global emotion information in an aspect level text, it ignores the important local emotion information in the text. Therefore, this study explores using a pre-training model to extract global features and a convolution network to extract local features. The proposed model can make full use of the information in sentences and improves the accuracy of emotion analysis.

The main contributions of this paper can be summarized as follows.

1. The transformer feature extractor in the ALBERT network is improved by convolution networks, and the ALBERTC model is obtained. This model not only can accurately extract the global semantic information and location information from a text but also can effectively extract the local aspect information and emotional information so as to obtain a more comprehensive aspect level text information representation.

2. The output layer of the ALBERTC model is combined with convolution networks, and the obtained model is named the ALBERTC-CNN. This model can effectively improve the classification accuracy of aspect level emotion.

3. For the SemEval-2014 task, two open datasets, the restaurant and laptop datasets, are used in the experiments, and the competitive classification results are obtained.

The rest of the study is organized as follows. The proposed network model is introduced in Section II. The experimental verification of the proposed model and experimental results analysis are presented in Section III. In Section IV, the main conclusions are drawn, and future work directions are given.

II. PREPARATION PROPOSED NETWORK

The main problem to be solved in this study is the analysis of the aspect level emotional information. For given sentence information $L^a = \{a_1, a_2, a_3, \ldots, a_N\}$ of a length $N$ and aspect information $L^b = \{b_1, b_2, b_3, \ldots, b_M\}$ of a length $M$, the emotional tendency of an aspect item $l^a$ in a sentence $l^b$ can be effectively analyzed using the proposed model.

This study proposes an improved aspect level emotion analysis model based on the ALBERT network. As shown in Fig. 1, the proposed model is mainly composed of the input layer, TransConvs encoding layer, convolution layer, fully connected layer and softmax layer. The input layer mainly represents the aspect level text as a word vector, so that the later feature extractor can better extract the information from the text. The TransConvs feature extractor mainly extracts the global sentence feature information and position feature information, as well as the local aspect information and emotion information from the text vector. The output layer’s convolution layer mainly strengthens the feature extractor output. The main function of the fully connected and softmax layers is to map the extracted feature information with the corresponding sentiment categories of the text and obtain the sentiment tendency of all aspects.

A. INPUT LAYER

For the aspect level sentiment analysis task, the input sentence needs to be represented as a vector form before the next step of coding representation learning.

The input layer plays the role of converting the sentence into a vector, as shown in Fig. 2. This layer first represents...
the input sentence as [CLS] aspect words [SEP] sentence words [SEP]. Then, the input token semantic information, token location information and sentence segment information are represented in word vector form by Embedding.

After the input sentence passes through the input layer, the final output is shown in function (1). Where $E_i$ represents the vector information representation of the i-th word of the input sentence, $E_{token}$ represents the token semantic coding information of each word, $E_{segment}$ represents the segmentation information between aspect item and sentence, and $E_{position}$ represents the position information of each word.

$$E_i = E_{token} + E_{segment} + E_{position} \quad (1)$$

### B. TransConvs Layer

The ALBERT model represents a simplified BERT model. Compared with the BERT model, the ALBERT model has fewer training parameters and can achieve better performance, which is because it uses factorized embedding parameterization technology and cross-layer parameter sharing technology, which can decrease the number of training parameters, reduce the hardware requirements of training, and shorten the training time. At the same time, in the ALBERT model, simple NSP (Next Sentence Prediction) tasks are replaced into more complex SOP (Sentence-Order Prediction) tasks, which effectively prevents overfitting and increases the accuracy. The main structural part of the basic ALBERT model is the encoder part of the transformer feature extractor, which performs well in global information extraction and can extract the important position information, sentence information, and token semantic information in the whole sentence. However, Transformer is mainly composed of forward propagation module and multi-head self-attention mechanism module [32]. In [32], it is mentioned that the self-attention mechanism can capture long-distance dependencies by using KEY, QUERY and VALUE triples, so as to effectively extract the global information of context. However, this will inevitably lead to the loss of part of local information. As can be seen from the article [33], CNN is forced to capture local information through the use of local receiving domain, shared weight, etc. Therefore, CNN has a strong ability to capture local information. Therefore, our study uses CNN to improve Transformer to obtain TransConvs feature extractor. TransConvs combines the excellent ability of CNN to capture local features with the excellent ability of Transformer to extract global features, so as to improve the accuracy and F1 value of aspect level sentiment analysis. In general, TransConvs module is more suitable for the text aspect level emotional analysis task than Transformer, which is mainly composed of MHSAM (Multi-Headed Self Attention Module), FFNM (Feed Forward Network Module), and CNNM(Convolution Network Module).

As shown in Fig. 3, the first part of the TransConvs module is the multi-head self-attention module, which is composed of multi-head attention module, linear layer, normalization function between layers, and dropout function. The structure of the multi-head attention module is shown in Fig. 4, where it can be seen that when the word vector’s features are input into this module, the module first extracts the relationship between every two words in a sentence using the multi-head attention mechanism; then, the score matrix is randomly inactivated by the dropout function to get the output $u$ and the output $u$ is connected with the initial input residual to get the output information after the addition. Finally, the final output $y$ is obtained by the LayerNorm.

Assuming that the input of the module is denoted as $x_i$ and the output processed by the module is denoted as $y_i$, the calculation formula of the module is given by (2) and (3), where subscript $i$ refers to the sentence feature representation of the input $i$th input sentence.

$$x'_i = x_i + MHSAM \left( x_i \right) \quad (2)$$

$$y_i = \text{Layernorm} \left( x'_i \right) \quad (3)$$

The feed-forward module is connected to the multi-head attention module. The feed-forward module is composed of two layers having the linear transformation function and
nonlinear activation function. After the input $y$ passes through the two layers having the linear function and random deactivation, the output obtained by the glue activation function is connected with the input $y$ by residual, as shown in Fig. 5. Finally, the residual is added, and the final output $z$ of the module is obtained by inter-layer standardization.

Assuming that the initial input is denoted as $y_i$ and the output processed by the module is denoted as $z_i$, the calculation formula of the module is given by (4) and (5), where subscript $i$ refers to the text feature representation of the input $i$th input sentence.

$$y_i' = y_i + FFNM(y_i)$$

$$z_i = \text{Layernorm}(y_i')$$  \hspace{1cm} (4)

$$v_i = \text{Dropout}(z_i')$$  \hspace{1cm} (5)

### C. CONVOLUTION LAYER

Inspired by [19], the convolution module in this study uses a point-by-point convolution as a basic network layer. The point convolution means that the size of the convolution core is one, and the same conversion is performed for each input word.

In this study, a three-layer convolution network is connected in series. The numbers of convolution kernels in the first, second, and third layers are set to 64, 64, and 768, respectively. These parameter settings are used because they can provide an effective combination of the ALBERT network and the proposed multi-convolution module and can achieve satisfactory emotion analysis results. The activation function layer is placed behind each convolution layer in the network, and the activation function of this layer is the glue activation function proposed in [20], which introduces the idea of random regularization to the proposed model, so better results can be achieved than using the ReLU and ELU functions. The last layer of the multi-convolution module is connected with a layer with the random deactivation function dropout, which generates the final output of the convolution module. The structure of the convolution module is shown in Fig. 6.

Assuming that the initial input is denoted as $z_i$ and the output processed by the module is denoted as $v_i$, the calculation formula of the module is given by (6) and (7), where $PWconv$ represents the point convolution operation, $\delta$ is the activation function operation, $\text{Dropout}$ denoted the random deactivation operation, and subscript $i$ refers to the text feature representation of the input $i$th input sentence.

$$z_i' = \delta(PWconv_3(\delta(PWconv_2(\delta(PWconv_1(z_i))))))$$

$$v_i = \text{Dropout}(z_i')$$

### D. FULLY CONNECTED AND SOFTMAX LAYER

The full connection layer and the Softmax layer are connected in the last layer of the model, whose main role is to map the feature information extracted by the feature extractor to the text emotional tendency information. In this paper, the emotions of the text mainly include positive emotional tendencies, negative emotional tendencies and neutral emotional tendencies. Therefore, as shown in Fig. 7, the full connection layer and Softmax layer map the vector information represented by aspect items and sentence features into three-dimensional vector information.

### III. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental environment used in this study included Windows 10 64-bit operating system, NVIDIA GeForce GTX 1060Ti graphics card, Intel i7-9750H@2.60GHz, and 4G video memory.

### A. DATASETS

To verify the performance of the proposed model, two public comment datasets, laptop and restaurant, were used in the test. The laptop dataset referred to consumers’ sentiment toward notebook computers, and it included a total of 2328 training data samples and 638 test data samples. The restaurant dataset referred to the catering industry, and it included a total of 3608 training samples and 1120 test samples. The specific data distribution is shown in Table 1.

| Datasets  | Positive | Train | Negative | Test | Positive | Test | Negative |
|-----------|----------|-------|----------|------|----------|------|----------|
| Laptop    | 994      | 464   | 1560     | 341  | 169      | 173  |
| Restaurant| 2164     | 637   | 807      | 728  | 196      | 196  |

The data were in the XML format. To analyze the emotion from more aspects, the data were transformed to the TXT format, where zero indicated that the emotion tendency of the aspect item was negative, one indicated that the emotion tendency of the aspect item was positive, and two indicated that the emotion tendency of the aspect item was neutral. The processed data are shown in Figs. 8 and 9.
The experimental results are taken from the original paper. The networks are given in the following, where * indicates that the experimental results are taken from the original paper.

To compare the performance of the proposed model, the following networks were selected for comparison. The details of the networks are given in the following, where * indicates that the experimental results are taken from the original paper.

- (1) SVM *(Kiritchenko et al., 2014): This method is a traditional machine learning-based method. In [21], surface features, lexical features, and syntactic analysis features were combined, and the support vector machine was used to analyze the aspect level text emotion. The experiments were performed using two SemEval-2014 datasets, and the obtained results were competitive for that time.

- (2) LSTM: This study uses the standard bidirectional LSTM model to train the corpus; the input is the same as the model in this paper, that is, the input text format is [CLS] [aspect] [SEP] [sentence] [SEP], and the number of output neurons is set to 768.

- (3) TD-LSTM (Tang et al., 2016a): This model first uses the forward and reverse LSTMs to model and train the left and the right texts of aspect level vocabulary, and then concatenates the two training matrices. Finally, the stitched feature vectors are used for aspect level emotion analysis.

- (4) ATAE-LSTM (Wang et al., 2016): This model is based on the LSTM, and it embeds the aspect vector into the input vector of each word so as to enhance the role of the aspect in the hidden state. Finally, the hidden state vector is used to generate the final vector representation.

- (5) MemNet (Tang et al., 2016): The MemNet model [23] uses a multi-hop attention network to represent the feature vector of the text, so the model can capture the important word information in the text.

- (6) IAN (Ma et al., 2017): In the IAN network, two LSTM networks are used to model the aspect items and text information at the aspect level. It uses the hidden information in sentences to generate the attention vector of aspect items and also uses the aspect information to generate the attention vector in sentences. Finally, the output of the two LSTM networks is cascaded and combined, and it is used for aspect level text sentiment analysis.

- (7) RAM (Chen et al., 2017): The RAM model is a multi-layer architecture, where each layer is composed of attention-based word feature aggregation units and GRU units, which can be used to learn the feature representation of aspect level sentences.

- (8) CABASC (Liu et al., 2018): The CABASC model has two attention enhancement mechanisms: sentence-level content attention mechanism and context attention mechanism. The former is mainly responsible for giving important information from a global perspective, while the latter is mainly responsible for considering the order of words and their relationship. The results presented in the original study also show that the two attention mechanisms play a key role.

### Table 2. Experimental parameter setting.

| Pre-Training model | Deep learning network |
|--------------------|----------------------|
| Parameter          | Value                | Parameter          | Value                |
| Embedding size     | 128                  | Embedding size     | 300                  |
| Hidden size        | 768                  | Initialization     | xavier_uniform_      |
| Learning rate      | 2e-5                 | Learning rate      | 1e-5                 |
| Batch size         | 16                   | Batch size         | 52                   |
| Epoch number       | 50                   | Epoch number       | 50                   |
| Optimizer          | Adam                 | Optimizer          | Adam                 |

### B. EXPERIMENTAL PARAMETERS AND EVALUATION INDEXES

In the experiment, the evaluation indexes that have been commonly used in natural language processing were adopted, and they were as follows: accuracy, macro average recall rate ($R_{\text{Macro}}$), and macro average F1 value ($F1_{\text{Macro}}$), and they were respectively calculated by:

$$\text{Accuracy} = \frac{T}{T+F}$$  \hspace{1cm} (8)

$$P_{\text{Macro}} = \frac{1}{N} \sum_{i=1}^{n} \frac{TP_i}{TP_i + FP_i}$$  \hspace{1cm} (9)

$$R_{\text{Macro}} = \frac{1}{N} \sum_{i=1}^{n} \frac{TP_i}{TP_i + FN_i}$$  \hspace{1cm} (10)

$$F1_{\text{Macro}} = \frac{2 \times P_{\text{Macro}} \times R_{\text{Macro}}}{P_{\text{Macro}} + R_{\text{Macro}}}$$  \hspace{1cm} (11)

where $T$ denoted the number of categories with the correct classification, $F$ denoted the number of categories with wrong classification, $P$ denoted the number of categories with the correct prediction, $N$ was the number of categories with a wrong prediction, and $P_{\text{Macro}}$ denoted the accuracy rate.

In this study, the parameter setting of the experiment included two parts: the parameter setting of the pre-training language model, and the parameter setting of deep learning model. In the pre-training language comparison models, the BERT and ALBERT models, the parameters of the official base models were set. The maximum sentence length was 128, the learning rate was set to $2 \times 10^{-5}$, the dimension of the word vector was 300, and the initialization model xavier_uniform_ was used to initialize the text vector. The hidden word vector dimension was also set to 300, the learning rate was set to $1 \times 10^{-3}$, and the random deactivation rate was set to 0.1. For all the models, the learning framework of Python 1.6.0 was used to build and train the model. The Adam optimizer was used to optimize model parameters, and the cross loss function was used as a loss function. The specific parameter settings are shown in Table 2.

### C. MODEL COMPARISON EXPERIMENT

To compare the performance of the proposed model, the following networks were selected for comparison. The details of the networks are given in the following, where * indicates that the experimental results are taken from the original paper.

**TABLE 2. Experimental parameter setting.**

| Pre-Training model | Deep learning network |
|--------------------|----------------------|
| Parameter          | Value                | Parameter          | Value                |
| Embedding size     | 128                  | Embedding size     | 300                  |
| Hidden size        | 768                  | Initialization     | xavier_uniform_      |
| Learning rate      | 2e-5                 | Learning rate      | 1e-5                 |
| Batch size         | 16                   | Batch size         | 52                   |
| Epoch number       | 50                   | Epoch number       | 50                   |
| Optimizer          | Adam                 | Optimizer          | Adam                 |

**FIGURE 8.** Data display of the laptop dataset.

1. room 2. Good
2. room 3. The room is spacious, but the furniture is outdated.
3. bed 4. The bed is comfortable, but the sheets are thin.
4. comfort 5. The comfort level is acceptable, but the noise is too loud.

**FIGURE 9.** Data display of the restaurant dataset.

1. food 2. The food was great, but the service was slow.
2. service 3. The service was friendly, but the prices were high.
3. price 4. The price is reasonable, although the service is poor.
4. quality 5. The quality of the food was high, but the service was slow.

In the experiment, the evaluation indexes that have been commonly used in natural language processing were adopted, and they were as follows: accuracy, macro average recall rate ($R_{\text{Macro}}$), and macro average F1 value ($F1_{\text{Macro}}$), and they were respectively calculated by:

$$\text{Accuracy} = \frac{T}{T+F}$$  \hspace{1cm} (8)

$$P_{\text{Macro}} = \frac{1}{N} \sum_{i=1}^{n} \frac{TP_i}{TP_i + FP_i}$$  \hspace{1cm} (9)

$$R_{\text{Macro}} = \frac{1}{N} \sum_{i=1}^{n} \frac{TP_i}{TP_i + FN_i}$$  \hspace{1cm} (10)

$$F1_{\text{Macro}} = \frac{2 \times P_{\text{Macro}} \times R_{\text{Macro}}}{P_{\text{Macro}} + R_{\text{Macro}}}$$  \hspace{1cm} (11)

where $T$ denoted the number of categories with the correct classification, $F$ denoted the number of categories with wrong classification, $P$ denoted the number of categories with the correct prediction, $N$ was the number of categories with a wrong prediction, and $P_{\text{Macro}}$ denoted the accuracy rate.

In this study, the parameter setting of the experiment included two parts: the parameter setting of the pre-training language model, and the parameter setting of deep learning model. In the pre-training language comparison models, the BERT and ALBERT models, the parameters of the official base models were set. The maximum sentence length was 128, the learning rate was set to $2 \times 10^{-5}$, the dimension of the word vector was 128, and the dimension of the hidden layer vector was 768. In other comparison models, the initial word vector dimension was 300, and the initialization model xavier_uniform_ was used to initialize the text vector. The hidden word vector dimension was also set to 300, the learning rate was set to $1 \times 10^{-3}$, and the random deactivation rate was set to 0.1. For all the models, the learning framework of Python 1.6.0 was used to build and train the model. The Adam optimizer was used to optimize model parameters, and the cross loss function was used as a loss function. The specific parameter settings are shown in Table 2.

**FIGURE 8.** Data display of the laptop dataset.

1. food 2. Not only was the food outstanding, but the little "perks" were great.
3. price 4. The price is reasonable, although the service is poor.
5. service 6. The service was friendly, but the prices were high.
7. quality 8. The quality of the food was high, but the service was slow.

**FIGURE 9.** Data display of the restaurant dataset.

1. food 2. The food was great, but the service was slow.
3. service 4. The service was friendly, but the prices were high.
5. price 6. The price is reasonable, although the service is poor.
7. quality 8. The quality of the food was high, but the service was slow.
(9) TNet-LF (Li et al., 2018): In [27], two aspect dependent long-term and short-term memory networks were constructed, and the aspect information in the target sentence was actively considered.

(10) AOA (Huang et al., 2018): This model mainly models aspects and sentences in a joint way. It can learn aspect information and sentence information at the same time and can automatically focus on important parts of sentences.

(11) MGAN (fan et al., 2018): The multi-granularity attention network (MGAN) model was proposed in [29], and this model can capture the character level interaction information between aspect items and context using the fine-grained attention mechanism.

(12) ASGCN (Zhang et al., 2019): In [30], the graph neural network was used to capture syntactic information and long-distance word dependence information to solve the problem that most models cannot effectively use syntactic constraints and long-distance word dependence so as to improve the accuracy of aspect level sentiment analysis.

(13) ASEGCAN(Xiao et al., 2020): In [31], the syntactic edge enhancement graph convolutional network is adopted to realize the sentiment analysis at the aspect level of sentences. With full consideration of different types of domains and boundary constraints, the network can efficiently learn better the representation of aspect information and subjective words, thus improving the accuracy of the sentiment analysis at the aspect level to some extent.

(14) BERT: This study adopts the BERT-based-uncased model released by Google, which follows the basic setting of the official website, and the learning rate is set to $2 \times 10^{-5}$.

(15) ALBERT: The ALBERT-BASE-V2 model is also released by Google, and the parameter settings of this model are the same as that of the Google official settings.

As shown in Table 3, the accuracy, macro average F1 value, and macro average recall rate of the proposed model were better than those of the other models on both datasets. On the laptop dataset, the proposed model achieved the accuracy of 80.56%, the macro average F1 value of 76.97%, and the macro average recall rate of 79.13%. The performances of the proposed model on the restaurant dataset were as follows: the accuracy was 87.23%, the average F1 value of macro was 81.01%, and the average recall rate of macro was 80.44%. The experimental results of the proposed model were better than those of the traditional machine learning network and the deep learning network. Compared with them, the accuracy of the proposed model was improved by 4%–10% on the laptop dataset and by 4%–7% on the restaurant dataset. The presented experimental results demonstrate the superiority and expressiveness of the proposed ALBERTC-CNN model over the other models in aspect level emotion analysis task.

D. MODEL ABLEATION EXPERIMENT

Compared to the original ALBERT model, the proposed model has three main improvements. The first one is that using the ConvNets, the Transformer feature extractor is improved, and the ALBERTC is obtained. The second is that by adding the ConvNets to the output layer of the ALBERT, the ALBERT model is improved, and the ALBERT-CNN model is obtained. The third is that both the transformer feature extractor and the output layer of ALBERT are improved by adding the ConvNets, and the ALBERT-CNN model is obtained. The mentioned improvements were tested experimentally to determine their effects on model performance. The experimental results are shown in Table 4.

As shown in Table 4, after adding the ConvNets, the model performance was improved compared to the original ALBERT model, and both the accuracy and the macro average recall were improved on both datasets. The experimental performance of the ALBERT-CNN model, which was obtained by improving the output layer of the ALBERT by adding the ConvNets, was better than that of the original ALBERT model. Compared with the original ALBERT model, the accuracy of the ALBERTC model was improved by 0.78% and 0.63% on the laptop and restaurant datasets, respectively, which indicated that the multi-convolution networks could improve model performance. When the transform feature extractor and the output layer of the ALBERT model were both improved by adding the multi-convolution network (the ALBERTC-CNN model), the best performance was achieved. The results showed that the accuracy, macro average F1 value, and macro average recall rate of the proposed model were improved by approximately 1.5% and 1% on laptop and restaurant datasets, respectively. The experimental results effectively illustrate the superiority of the

| Laptop | Restaurant |
|--------|------------|
| SVM*   | 70.49      | 80.16    | 65.72 |
| LSTM   | 69.28      | 63.08    | 64.63 |
| TD-LSTM| 69.59      | 63.86    | 77.05 |
| ATAE-LSTM| 70.06    | 64.62    | 77.76 |
| MemNet | 73.67      | 69.24    | 70.85 |
| IAN    | 72.88      | 68.46    | 70.01 |
| RAM    | 74.45      | 70.53    | 71.33 |
| CABASC | 75.23      | 69.59    | 71.92 |
| TNet-CDF | 76.18    | 71.22    | 73.78 |
| AOA    | 74.76      | 70.26    | 73.08 |
| MGAN   | 75.54      | 72.79    | 72.42 |
| ASGCN  | 75.39      | 69.75    | 72.07 |
| ASEGCN*| 76.49      | 72.64    | 81.61 |
| BERT   | 78.36      | 73.89    | 75.49 |
| ALBERT | 78.84      | 75.41    | 77.51 |
| ALBERT-CNN | 80.56    | 76.97    | 79.13 |

| Laptop | Restaurant |
|--------|------------|
| SVM*   | 70.49      | 80.16    | 65.72 |
| LSTM   | 69.28      | 63.08    | 64.63 |
| TD-LSTM| 69.59      | 63.86    | 77.05 |
| ATAE-LSTM| 70.06    | 64.62    | 77.76 |
| MemNet | 73.67      | 69.24    | 70.85 |
| IAN    | 72.88      | 68.46    | 70.01 |
| RAM    | 74.45      | 70.53    | 71.33 |
| CABASC | 75.23      | 69.59    | 71.92 |
| TNet-CDF | 76.18    | 71.22    | 73.78 |
| AOA    | 74.76      | 70.26    | 73.08 |
| MGAN   | 75.54      | 72.79    | 72.42 |
| ASGCN  | 75.39      | 69.75    | 72.07 |
| ASEGCN*| 76.49      | 72.64    | 81.61 |
| BERT   | 78.36      | 73.89    | 75.49 |
| ALBERT | 78.84      | 75.41    | 77.51 |
| ALBERT-CNN | 80.56    | 76.97    | 79.13 |

TABLE 3. Model comparison with experimental results.

TABLE 4. Ablation results.
proposed model and the necessity of improving the ALBERT model by the introduction of an convolution networks. Compared with the original ALBERT model, the model proposed in this paper performs better, mainly because the model in this paper combines the ability of Transformer to extract long-dependent information [32] and the ability of CNN to extract fine-grained information [33], which enables the text to extract more comprehensive text information in the feature extraction stage. So that accuracy and F1 worth to be effectively improved.

IV. CONCLUSION

Using the aspect level sentiment analysis can provide improvements in many fields, so it is of great significance to study the aspect level sentiment analysis. The global semantic information and local emotional information cannot be extracted simultaneously by most networks, so this study proposes an improved aspect level emotion analysis model based on the ALBERT. The proposed model is verified by comparison experiments, and the obtained experimental results show that the improved feature extractor TransConvs of the ALBERT can effectively extract the global semantic information, location information, local aspect information, and emotional information from a text at the same time. Introducing the convolution network in the output layer of the ALBERT can enhance the aspect and emotional information representation ability of a network, and the text aspect level emotion analysis ability. Compared with the traditional network and the baseline network ALBERT, the proposed model can achieve higher accuracy and F1 value and better convergence, which shows that it can be used for the emotion analysis in various fields of aspect level emotion analysis due to its advantages and competitiveness.

In future work, we will further consider the applicability of an ALBERT-CNN and further improve the structure of the network. In addition, we will try to increase the feature extraction ability of the proposed model and apply it to other tasks of natural language processing, such as sentence-level sentiment analysis, named object recognition, machine translation, and opinion extraction for the purpose of further evaluation.

ACKNOWLEDGMENT

The authors would like to thank LetPub for its linguistic assistance during the preparation of this manuscript.

REFERENCES

[1] C. Mahalakshmi and A. M. Abirami, “Aspect—Based sentiment analysis using context-aware learning for online reviews,” in Proc. Int. Conf. Innov. Eng. Technol. (ICIEET), Tokyo, Japan, 2016. [Online]. Available: http://iirset.com/upload/2016/icieet/11_ICIEET16_paper_138.pdf
[2] E. M. Aalshari, A. Azman, S. Doraisamy, N. Mustapha, and M. Alkehr, “Effective method for sentiment lexical dictionary enrichment based on Word2Vec for sentiment analysis,” in Proc. 4th Int. Conf. Inf. Retr. Knowl. Manage. (CAMP), Mar. 2018, pp. 1–5.
[3] T. Álvarez-López, J. Juncal-Martínez, M. Fernández-Gavilanes, E. Costa-Montenegro, and F. J. González-Castaño, “GTI at SemEval-2016 task 5: SVM and CRF for aspect detection and unsupervised aspect-based sentiment analysis,” in Proc. 10th Int. Workshop Semantic Eval. (SemEval), 2016, pp. 306–311.
[4] M. S. Mubarak, Adiwijaya, and M. D. Aldhi, “Aspect-based sentiment analysis to review products using Naïve Bayes,” in Proc. AIP Conf., vol. 1867, 2017, Art. no. 020060, doi: 10.1063/1.4994463.
[5] R. Rafieek and R. Rennya, “Detecting contextual word polarity using aspect based sentiment analysis and logistic regression,” in Proc. IEEE Int. Conf. Smart Technol. Manage. Comput., Commun., Controls, Energy Mater. (ICSTM), Aug. 2017, pp. 102–107.
[6] C. Cortes and V. Vapnik, “Support-vector networks,” Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.
[7] J. Zhu and T. Hastie, “Kernel logistic regression and the import vector machine,” in Proc. Int. Conf. Neural Inf. Process. Syst., Natural Synth., 2001, pp. 1081–1088.
[8] N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian network classifiers,” Mach. Learn., vol. 29, no. 2, pp. 131–163, Nov. 1997.
[9] Y. Kim, “Convolutional neural networks for sentence classification,” in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2014, pp. 1746–1751.
[10] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, “Attention-based bidirectional long short-term memory networks for relation classification,” in Proc. 5th Annu. Meeting Assoc. Comput. Linguistics (Short Papers), vol. 2, 2016, pp. 207–212.
[11] P. Chen, Z. Sun, L. Bing, and W. Yang, “Recurrent attention network on memory for aspect sentiment analysis,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2017, pp. 452–461.
[12] W. Xue and T. Li, “Aspect based sentiment analysis with gated convolutional networks,” 2018, arXiv:1805.07043. [Online]. Available: http://arxiv.org/abs/1805.07043
[13] A. Radford et al. (2018). Improving Language Understanding by Generative Pre-Training. [Online]. Available: https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf
[14] A. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, “Deep contextualized word representations,” in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol., (Long Papers), vol. 1, 2018, pp. 2227–2237.
[15] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” in Proc. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol., Minneapolis, MI, USA: The NAACL-HLT Press, May 2019, pp. 4171–4186.
[16] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, “RoBERTa: A robustly optimized BERT pretraining approach.” 2019, arXiv:1907.11692. [Online]. Available: http://arxiv.org/abs/1907.11692
[17] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, “ALBERT: A lite BERT for self-supervised learning of language representations.” 2019, arXiv:1909.11942. [Online]. Available: http://arxiv.org/abs/1909.11942
[18] M. Hoang, “Aspect-based sentiment analysis using BERT,” in Proc. NODILDA, 2019, pp. 187–196.
[19] Z. Wu, Z. Liu, J. Lin, Y. Lin, and S. Han, “Lite transformer with long-short range attention.” 2020, arXiv:2004.11886. [Online]. Available: http://arxiv.org/abs/2004.11886
[20] D. Hendrycks and K. Gimpel, “Gaussian error linear units (GELUs),” 2016, arXiv:1606.08415. [Online]. Available: http://arxiv.org/abs/1606.08415
[21] S. Kiritchenko, X. Zhu, C. Cherry, and S. Mohammad, “NRC-Canada-2014: Detecting aspects and sentiment in customer reviews,” in Proc. 8th Int. Workshop Semantic Eval. (SemEval), 2014, pp. 437–442.
[22] D. Tang, B. Qin, X. Feng, and T. Liu, “Effective LSTMs for target-dependent sentiment classification,” 2015, arXiv:1512.01100. [Online]. Available: http://arxiv.org/abs/1512.01100
[23] Y. Wang, M. Huang, X. Zhu, and L. Zhao, “Attention-based LSTM for aspect-level sentiment classification,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2016, pp. 606–615.
[24] D. Tang, B. Qin, and T. Liu, “Aspect level sentiment classification with deep memory network,” 2016, arXiv:1605.08900. [Online]. Available: http://arxiv.org/abs/1605.08900
[25] D. Ma, S. Li, X. Zhang, and H. Wang, “Interactive attention networks for aspect-level sentiment classification,” 2017, arXiv:1709.00893. [Online]. Available: http://arxiv.org/abs/1709.00893
[26] Q. Liu, H. Zhang, Z. Huang, and Z. Wu, “Content attention model for aspect based sentiment analysis,” in Proc. World Wide Web Conf. World Wide Web (WWW), 2018, pp. 1023–1032.
[27] X. Li, L. Bing, W. Lam, and B. Shi, “Transformation networks for target-oriented sentiment classification,” 2018, arXiv:1805.01086. [Online]. Available: http://arxiv.org/abs/1805.01086

94754 VOLUME 9, 2021
[28] B. Huang, Y. Ou, and K. M. Carley, “Aspect level sentiment classification with attention-over-attention neural networks,” in Proc. Int. Conf. Social Comput., Behav.-Cultural Modeling Predict. Behav. Represent. Modeling Simulation, Cham, Switzerland: Springer, 2018, pp. 197–206.

[29] F. Fan, Y. Feng, and D. Zhao, “Multi-grained attention network for aspect-level sentiment classification,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2018, pp. 3433–3442.

[30] C. Zhang, Q. Li, and D. Song, “Aspect-based sentiment classification with aspect-specific graph convolutional networks,” 2019, arXiv:1909.03477. [Online]. Available: http://arxiv.org/abs/1909.03477

[31] Y. Xiao and G. Zhou, “Syntactic edge-enhanced graph convolutional networks for aspect-level sentiment classification with interactive attention,” IEEE Access, vol. 8, pp. 157068–157080, 2020, doi: 10.1109/ACCESS.2020.3019277.

[32] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” 2017, arXiv:1706.03762. [Online]. Available: https://arxiv.org/abs/1706.03762

[33] H. Wu, B. Xiao, N. Codella, M. Liu, X. Dai, L. Yuan, and L. Zhang, “CvT: Introducing convolutions to vision transformers,” 2021, arXiv:2103.15808. [Online]. Available: http://arxiv.org/abs/2103.15808

XINGXIN YE was born in Guiyang, Guizhou, China, in 1995. She received the bachelor’s degree in electronic and information engineering from Zhejiang Normal University, in 2019. She is currently pursuing the master’s degree in electronic and communication engineering with the College of Big Data and Information Engineering, Guizhou University.

Her research interests are mainly natural language processing technology, including sentence level sentiment analysis and aspect level sentiment analysis; in addition, she is also studying big data information processing foundation and Internet technology. During her bachelor’s degree, she won the First Prize in the electronics and design competition, the third-class academic scholarship, and the excellent graduate. While studying for her master’s degree, she won a second-class scholarship and a merit student of the university.

YANG XU received the bachelor’s degree from Guizhou University, in 2003, and the Ph.D. degree in engineering from the Institute of Modern Materials and Mechanics, Chinese Academy of Sciences, in 2008.

He serves as a master’s tutor with Guizhou University, where he is currently an Associate Professor. His current research areas and directions include big data acquisition, the Internet of Things technology, and embedded technology.

MENGSHI LUO was born in Guiyang, Guizhou, China, in 1995. She received the bachelor’s degree in information engineering from Kaili University, in 2019. She is currently pursuing the master’s degree in electronic and communication engineering with Guizhou University.

Her research interests include data mining, human pose estimation and analysis, and Internet technology. The awards she won include the academic scholarship for her bachelor’s degree, the title of provincial merit student, and the second-class academic scholarship for her master’s degree.