IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 9, 2020, accepted June 17, 2020, date of publication June 29, 2020, date of current version August 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3005823

Text Sentiment Orientation Analysis Based on Multi-Channel CNN and Bidirectional GRU With Attention Mechanism

YAN CHENG[1], LEIBO YAO[1], GUOXIONG XIANG[2], GUANGHE ZHANG[1], TIANWEI TANG[3], AND LINHUI ZHONG[1]

[1]School of Computer Information Engineering, Jiangxi Normal University, Nanchang 330022, China
[2]School of Communication, Jiangxi Normal University, Nanchang 330022, China
[3]Center of Management Decision Evaluation Research, Jiangxi Normal University, Nanchang 330022, China

Corresponding authors: Yan Cheng (chyan88888@jxnu.edu.cn), Guoxiong Xiang (1529159315@qq.com), and Leibo Yao (yaoleibo_jxnu@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61967011, in part by the Key Research and Development Program of Jiangxi Province under Grant 20161BBE50086, in part by the Jiangxi Provincial Department of Education Science and Technology Key Project under Grant GJJ150299, and in part by the Humanities and Social Sciences Key (Major) Project of Education Department under Grant JD19056.

ABSTRACT Convolutional Neural Network(CNN) and Recurrent Neural Network(RNN) have been widely used in the field of text sentiment analysis and have achieved good results. However, there is an antero-posterior dependency between texts, although CNN can extract local information between consecutive words of a sentence, it ignores the contextual semantic information between words. Bidirectional GRU can make up for the shortcomings that CNN can't extract contextual semantic information of long text, but it can't extract the local features of the text as well as CNN. Therefore, we propose a multi-channel model that combines the CNN and the bidirectional gated recurrent unit network with attention mechanism (MC-AttCNN-AttBiGRU). The model can pay attention to the words that are important to the sentiment polarity classification in the sentence through the attention mechanism and combine the advantages of CNN to extract local features of text and bidirectional GRU to extract contextual semantic information of long text, which improves the text feature extraction ability of the model. The experimental results on the IMDB dataset and Yelp 2015 dataset show that the proposed model can extract more rich text features than other baseline models, and can achieve better results than other baseline models.

INDEX TERMS Convolutional neural network, bidirectional gated recurrent unit network, attention mechanism, text sentiment orientation analysis.

I. INTRODUCTION

Sentiment analysis, also known as opinion mining, refers to people's emotions, opinions, evaluations, attitudes and emotions about services, products, organizations, individuals, problems, events, topics and their attributes [1]. Emotion processing and emotion understanding is important for the closely related task of polarity detection [2]. In the field of natural language processing, sentiment orientation analysis is a basic task, which has attracted wide attention from scholars at home and abroad in recent years. With the rapid development of the Internet and mobile networks, users have more and more opinions and comments published on the Internet, it is worth mining the important emotional information from these contents. Sentiment analysis is mainly divided into two categories: subjective and objective sentiment recognition and sentiment classification, the former divides the given text into subjective or objective text, while the latter classifies subjective texts as positive, negative or neutral [3]. The sentiment analysis methods are mainly based on machine learning and sentiment lexicon, with the re-recovery of neural network research, deep learning-based methods widely applied to text sentiment analysis tasks. The method based on sentiment lexicon was first used in the field of sentiment analysis, it first requires us to construct an sentiment lexicon, then perform manual polarity and intensity annotation on the sentiment

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

lexicon, and finally achieve the text sentiment classification. Although this method can achieve text sentiment classification, it is not efficient, because it requires manual construction of an sentiment lexicon and manual labeling. Since the 1990s, machine learning methods have begun to emerge in the tasks of text sentiment analysis [4], [5]. Although the function of the machine learning model is simple, it usually requires complex feature engineering, and the quality of the feature can directly affect the subsequent result of classification. In addition, the generalization ability of machine learning methods is relatively low, with the rise and development of deep learning, many deep learning methods have been successfully applied to natural language processing tasks, especially in sentiment analysis tasks, such as CNN [6]–[8], LSTM [9], [10]. Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN) are two popular models in the field of deep learning. However, they also have certain defects, the CNN model can't consider the contextual semantic information, thus neglecting the interrelationship between the words within the sentence; the RNN model lacks the ability to extract features from the text, and can't extract local feature well. Aiming at the above problems, this paper uses CNN, Bidirectional GRU model and Attention mechanism, combined with word2vec [11] word vector model, and proposes a multi-channel CNN and Bidirectional GRU model with attention mechanism to solve the text sentiment classification problem. The model uses three CNNs for information extraction, and uses different convolution windows to extract emotional features with different granularities, so that richer emotion features of text can be extracted. The Bidirectional GRU network can consider the text context information between sentences, so that the feature information in the entire text can be more fully utilized, and the text sentiment polarity can be effectively recognized. At the same time, the attention mechanism is introduced in both the CNN channel and the Bidirectional GRU network channel, which can pay more attention to the words that are important to the sentiment polarity classification in the sentence, so as to further improve the classification accuracy.

The main contributions of this paper are as follows:

1). A new multi-channel convolutional neural network and bidirectional GRU model based on attention mechanism are proposed for text sentiment classification;

2). We uses multi-channel CNN model to extract the local features and bidirectional GRU model, our proposed model effectively solves the defect that single-channel CNN and single-channel bidirectional GRU can only extract single feature information;

3). Introduce attention mechanism on CNN and bidirectional GRU model to allow model automatically extract keyword information in the text, ignoring words that are not relevant to the text classification.

II. RELATED WORK

Sentiment analysis is the field of research on people's views, attitudes, and emotions. Sentiment orientation analysis has always been a research hotspot in the field of natural language processing. In the early years, sentiment lexicon-based and traditional machine learning-based methods were mainly used for sentiment classification. The method based on sentiment lexicon is to use the words that have been manually labeled in the sentiment lexicon and each word in the sentence to be matched, and then calculate the score of each sentence using the relevant formula, finally, the whole sentence score is greater than 0, divided into positive class, and less than 0 is divided into negative class, so as to conduct text sentiment orientation classification. For English sentiment lexicon, Kamps *et al.* [12] use the WordNet English sentiment lexicon to judge the sentiment orientation of English texts. Compared with the English sentiment lexicon, the normative Chinese sentiment lexicon is relatively lacking, the earliest and most commonly used is the large-scale sentiment analysis corpus library HowNet [13] marked by Dong *et al*. Although the method based on the sentiment lexicon is simple in operation, the construction of the sentiment lexicon is time-consuming and labor-intensive, and the scope of expansion is limited, which has domain limitations. Traditional machine learning-based sentiment classification methods are often seen as a type of supervised learning problem. In 2004, Pang and Lee [14] used maximum entropy, Naive Bayes and support vector machines to compare experiments in text sentiment analysis, it was found that using SVM to conduct text sentiment classification can achieve optimal results. Poirier *et al.* [15] used naive Bayesian classifier to classify the sentiment polarity of movie reviews. Lee and Renganathan [16] applied the maximum entropy classification to the polarity evalutation of a given electronic product review. Naz *et al.* [17] present an SVM based classifier that hybridizes n-gram based internal features with an external sentiment vector in order improve to standard n-gram based classifier, this method outperforms other machine learning methods on Twitter text sentiment analysis tasks. Although these machine learning methods can perform text sentiment classification, complex feature selection is needed in the process of classification, this process also needs manual design, which also leads to poor scalability and difficulty in adapting to different fields and application needs.

In recent years, deep learning methods have gradually become the mainstream method of sentiment analysis. It does not require manual intervention and is an end-to-end method that enables automatic selection of features, so it is more efficient than that base on sentiment lexicon and machine learning methods, and the field is more applicable. In 2011, Collobert *et al.* [18] first proposed the use of CNN to solve problems in the NLP field such as part-of-speech tagging. In 2014, Kim [7] proposed applying CNN to the English text sentiment classification task, and achieved a good classification result at that time. Kalchbrenner *et al.* [6] proposed a wide convolution model and replaced the max-pooling of traditional CNN with k-max pooling, which can retain more feature information, this model does not require any prior information input, nor does it require the construction of very complex artificial features. Yin and Schütze [19] used multi-channel convolutional neural networks of different sizes for sentence classification. However, CNN-based text sentiment classification cannot consider sentence context semantic information.

Compared with CNN, RNN introduces a memory unit to make the network have a certain memory and can capture long-distance dependencies between texts. However, it also has certain drawbacks, that is, the gradient vanishing problem during the training process. LSTM and GRU introduce a gate mechanism based on the traditional RNN, which better overcomes the disadvantages of gradient vanishing in RNN. Based on the LSTM and GRU models, many people have improved on them and achieved good classification results. Tang *et al.* [20] first use CNN/LSTM to implement single sentence representation, then use gated RNN to encode the internal relationship and semantic relationship between sentences, and finally implement the encoding of chapter-level text, this method can better capture semantic information between sentences. Wang *et al.* [21] proposed an RCNN model, which first used the bidirectional recurrent network model to obtain the context information of the word, then used the CNN convolution and pooling process to classify the text, and obtained the most result on several data sets at that time. Zhou *et al.* [22] proposed a C-LSTM model, which first uses the convolutional neural network to extract text features, and then uses the LSTM network replace the maximum pooling layer to obtain the final classification results. Tang *et al.* [23] used two LSTM to encode the sentences from the left and right sides of the aspect words, and then connect the outputs of the two networks as the sentiment related to the target aspect. Ruder *et al.* [24] uses a hierarchical bidirectional LSTM model for aspect-level text sentiment classification. Rao *et al.* [25] used two-layer LSTM for document-level sentiment classification. Sachin *et al.* [26] used LSTM, GRU, bidirectional GRU and bidirectional LSTM models to perform sentiment analysis of Amazon reviews, and achieved good results. Although the LSTM and GRU models are able to extract contextual semantic information, they do not extract the local features of a sentence as well as CNN.

In order to make full use of the respective advantages of CNN and RNN, more and more researchers have combined CNN and RNN to conduct text sentiment orientation analysis. Wang *et al.* [27] used one layer CNN and one layer RNN to form a fusion model for text sentiment analysis. The experiment proved that the effect is better than the simple use of CNN and RNN models. Zhou *et al.* [28] proposed a model combining bidirectional LSTM and CNN, and used 2-dimensional maximum pooling instead of the traditional maximum pooling at the pooling layer to obtain the final classification results. Zhang *et al.* [29] proposed a CNN-LSTM model for predicting the emotional intensity of Twitter text. Zhang *et al.* [30] proposed a multi-channel CNN-LSTM model for Twitter text sentiment classification task. Sun *et al.* [31] used a CNN-LSTM model to classify the Tibetan micro-blogs, and the hybrid deep learning algorithm obtains a good classification effect. Zhang *et al.* [32] used the Convolution-GRU model to perform emotional polarity discrimination on Twitter hate comment texts. Abd El-Jawad *et al.* [33] used a deep learning method that mixed CNN and RNN for text sentiment analysis, and achieved the best classification effect on the tweet data set.

In recent years, attention mechanism has been widely used in text classification tasks, because it allows models pay attention to important words in texts, for example, Yang *et al.* [34] combined the bidirectional recurrent neural network and attention mechanism, and constructed a new type of attention model applied in the text-level classification task, which achieved the best result of the text classification task at that time. Wang *et al.* [35] used a multi-layered attention mechanism and a convolutional neural network for the classification task. The experimental results of the model on multiple data sets show that the model using the attention mechanism have higher classification accuracy than models that do not use attention mechanism. Wang *et al.* [36] proposed an Attention-based Long Short-Term Memory Network for aspect-level sentiment classification, the model achieved state-of-the-art performance on aspect-level sentiment classification on SemEval 2014 dataset. Cheng *et al.* [37] proposed a HiErarchical ATtention (HEAT) network for aspect-level sentiment classification, the method achieves better performance on aspect-level sentiment classification than state-of-the-art models. Ma *et al.* [38] proposed the interactive attention networks (IAN) to interactively learn attentions in the contexts and targets, and generate the representations for targets and contexts separately. With this design, the IAN model can well represent a target and its collocative context, which is helpful to sentiment classification. Experimental results on SemEval 2014 Datasets demonstrate the effectiveness of the model. Han *et al.* [39] proposed a Multi-Attention Network (MAN) model which adopts several attention networks, the model solve the problem of the RNN-based model can't extract the potential correlation between relatively distant sentiment words and aspect words in complex statements, and the proposed model could achieve consistently superior results on three datasets. Gao *et al.* [40] propose a collaborative extraction hierarchical attention network, this proposed method achieves better performance than the methods which only use aspect features to extract sentiment feature for aspect-level sentiment classification on SemEval competition data set. Yuan *et al.* [41] proposed a sentiment analysis model based on multi-channel convolution and bidirectional GRU networks, and introduced an attention mechanism on the bidirectional GRU to automatically pay attention to features that have a strong influence on sentiment polarity.

In view of the excellent performance of the neural network model with the attention mechanism, this paper introduces the attention mechanism for the text sentiment analysis task, which enables the network model to pay more attention to the words that contribute greatly to the emotional polarity of the text. Based on the advantages of CNN, RNN and attention mechanism, we propose a multi-channel CNN and bidirectional GRU model based on attention mechanism, this model uses 3-channel CNN to extract local feature information with different granularity and replaces traditional RNN with bidirectional GRU model, which extracts context semantic information of text, then introduces the attention mechanism on each channel, and finally concatenate them together to obtain a new multi-channel neural network model based on the attention mechanism. We comprehensively utilizes the advantages of CNN, bidirectional GRU and attention mechanism, it not only considers the local information of the text but also the contextual semantic information, avoids the loss of text information, and can pay attention to the words that contribute more to the sentiment classification in the sentence with the attention mechanism, and improve the performance of the model on the text sentiment classification.

III. MODEL ARCHITECTURES

A. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural network was first used in the field of computer vision, it has been widely used in natural language processing tasks in recent years and has achieved good results [7]. The common CNN model is mainly composed of convolutional layer, pooling layer and fully connected layer. Convolution layer is used to extract the features of input data, while pooling layer is to select and filter the features extracted by convolution layer, the fully connected layer is equivalent to the hidden layer of the traditional feed forward neural network, and is generally connected to the output layer at the end to achieve the final output. In our model, we represent the input sentence with a matrix $S \in R^{n\times d}$, where d represents the word embedding vector dimension of each word, and n represents the number of words in the sentence. Suppose the convolution kernel $W^c \in R^{h\times d}$, where c represents the number of convolution kernels and d represents the length of the convolution kernel, its size is the same as the dimension of the word embedding vector, and h is the width of the convolution kernel. For the input matrix $S \in R^{n\times d}$, the feature map vector $O = [o_0, o_1, \ldots, o_{n-h}] \in R^{n-h+1}$ is obtained by repeatedly applying a convolution kernel W to performing a convolution operation, where the calculation formula of each element in the feature vector O is as shown in Equation (1):

$$o_i = W \bullet S_{i:i+h-1} \tag{1}$$

where $i = 0, 1, 2, \ldots, n-h$, $(\bullet)$ denotes the point-wise multiplication operation of the matrix, and $S_{i:j}$ denotes the submatrix of the S matrix from i rows to j rows, ie the word embedding vector matrix of the i-th word to the j-th word. After the convolution operation is performed, each of the resulting feature map vectors O is sent to the pooling layer to generate potential features and filter the features. The common pooling strategy is max-pooling, the role of max-pooling is to capture the most important feature v after convolution, this strategy naturally handles a variable length sentence to a fixed length: $v = \underset{0\le i\le n-h}{max} \{o_i\}$.

We have described in detail the process of extracting a feature with a filter. The Fig. 1 model utilizes several filters (with different window sizes) to extract different features, and then max-pooling the features extracted by each filter to obtain a number of most important feature, these features are then concatenated and passed to the fully connected layer, we finally perform the softmax operation to get the final output, the output is the probability distribution on the label.

B. GRU AND BIDIRECTIONAL GRU

Cho *et al.* [42] originally proposed a gated recurrent unit in 2014 to enable each recurrent unit to adaptively capture dependencies of different time scales, the structure of the GRU is illustrated in Fig. 2:

Compared to LSTM, the GRU model is simpler, which consists of only the update gate z and the reset gate r, it has one less gate than the LSTM, so there are fewer parameters and faster convergence time during training, the parameters can be updated by Equations (2)-(5):

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \tag{2}$$
$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{3}$$
$$\tilde{h}_t = \tanh(W x_t + U(r_t \odot h_{t-1})) \tag{4}$$
$$h_t = (1 - z_t) h_{t-1} + z_t \tilde{h}_t \tag{5}$$

where W_*, U_* is the weight matrix of GRU, where σ represents the logical sigmoid function, $\odot$ represents the element multiplication operation, z_t represents the update gate, which can determine the update degree of the activation value in the GRU unit, which is determined by the current input state and the state of the previous hidden layer, r_t represents the reset gate, its update process is similar to the z_t, $\tilde{h}_t$ represents the candidate hidden layer, and h_t represents the hidden layer. GRU is actually a variant of LSTM, it mainly merges the input gate and the forget gate in LSTM into a single update gate, which reduces the training parameters and model convergence time, also reduces the training complexity, it is one of the popular recurrent neural network models at present.

The state of a unidirectional GRU is transmitted unidirectionally from front to back, in other words, it cannot take into account the influence of the following words, it is easy to ignore the influence of the following words, the bidirectional GRU is a variant of the unidirectional GRU, whose output depends on the double effects of forward and backward states, it solves the problem of unidirectional GRU, making the final output more accurate. The model structure of the bidirectional GRU is shown in Fig. 3:

C. ATTENTION MECHANISM

The attention mechanism was applied to the machine translation task as early as 2014 [43], and achieved the best results at that time, in 2017, Yang *et al.* [34] applied the attention mechanism to the sentiment classification task at the text chapter level, and made a good classification effect,

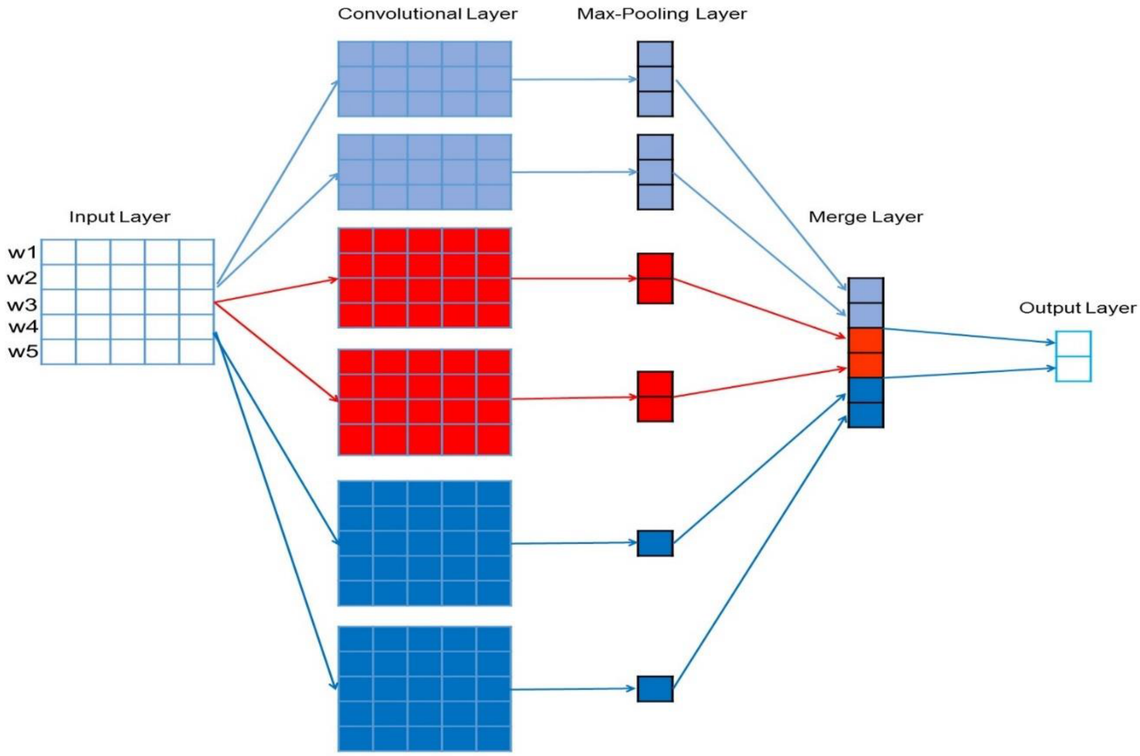

FIGURE 1. The architecture of the CNN.

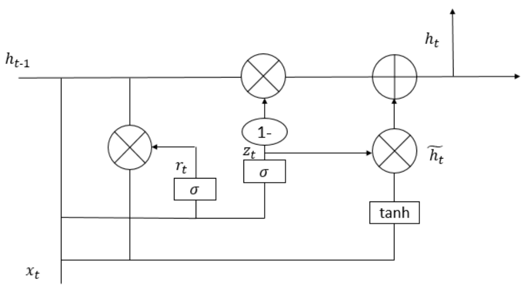

FIGURE 2. The architecture of the GRU.

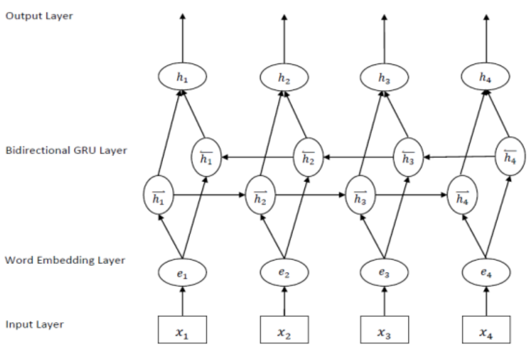

FIGURE 3. The architecture of the Bidirectional GRU.

Because each word has different degrees of importance for each sentence, we introduce attention mechanism to extract the semantic information of important words in the sentence. Attention mechanism can be abstracted into two modules: Encoder and Decoder. The Encoder is generally an encoder that performs a certain transformation on the input data to obtain a semantic vector, the Decoder is generally a decoder, and the output data is obtained after a certain transformation. The structure of the attention mechanism are shown in Fig. 4.

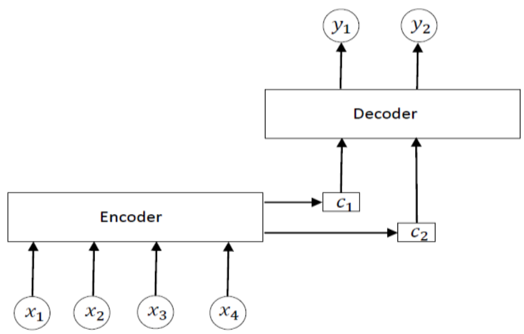

FIGURE 4. The architecture of attention mechanism.

Its main Equations are shown in (6)-(8):

$$u_i = \tanh(W_i h_i + b_i) \tag{6}$$

$$\alpha_i = \frac{\exp(u_i^T u_w)}{\sum_i \exp(u_i^T u_w)} \tag{7}$$

$$h_i = \sum_i \alpha_i h_i \tag{8}$$

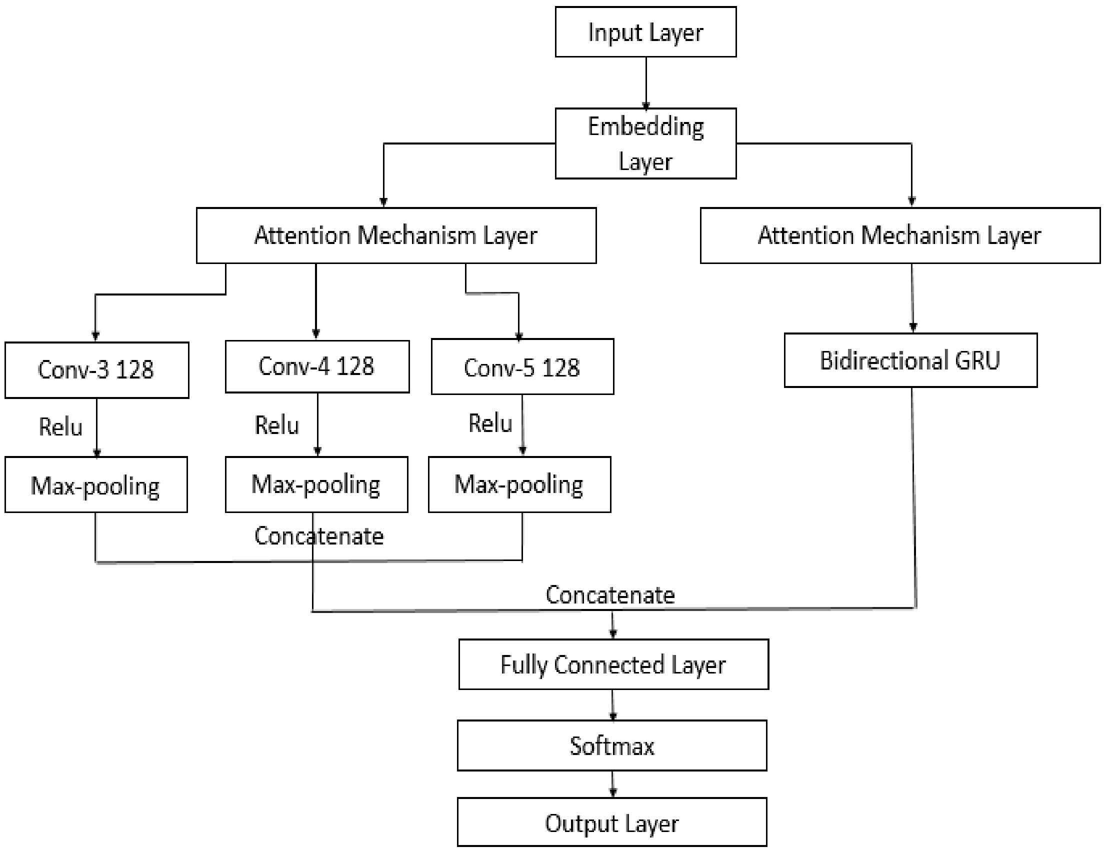

FIGURE 5. **The architecture of overall model.**

where: u_w is a randomly initialized context vector, which is updated during training; u_i is the result of a full connection operation of the hidden layer vector h_i; W_i and b_i are the weight matrix and bias term of attention calculation respectively; α_i is the attention score for the *i-th* word in the sentence.

D. OVERALL MODEL

In our paper, a multi-channel text sentiment analysis model combining CNN and Bidirectional GRU with attention mechanism is proposed on the above basic model. The model consists of convolutional neural network, bidirectional gated recurrent network and attention mechanism. The model structure is shown in Fig. 5:

The overall model of our paper is composed of multiple channels, the main body is 3 CNN model channels and one bidirectional GRU model channel, the three CNN channels are mainly used to extract different local features of words between sentences, while the bidirectional GRU channel is used to get sentence context semantic information. The input of the CNN channel is each word in the sentence, the first layer is the embedding layer, its function is to map each input word into a vector representation; the second layer is the attention mechanism layer, mainly to extract important word information between sentences; the third layer is the convolution layer, which is mainly used to extract the local features between words, the dimension of word embedding vector is set to 300 in our paper, the three selected filter sizes are 3,4,5 with 128 feature maps each, the stride is set to 1, the padding is set to valid (no need to perform zero-padding operation), the local features of the sentence can be obtained after the convolution operation; the fourth layer is the pooling layer, this layer mainly performs the max pooling operation on the local features obtained by the convolutional layer, extracts the most important features between sentences, discards some irrelevant and useless features, and generates feature vectors of fixed dimensions, then three features output through the max pooling operation are concatenated as part of the input of the fully connected layer, the three different features obtained through three different filters can be concatenated to provide richer feature information for the subsequent sentence sentiment classification.

The first layer of the bidirectional GRU channel is also the word embedding layer, and the dimension of the word embedding vector is also set to 300 dimensions, the second layer is also a layer of attention mechanism, the attention mechanism used in the CNN model is used to extract important word information in a sentence; the third and fourth layers of bidirectional GRU are forward GRU and backward GRU structures respectively, and their hidden layer sizes are set to 128, because the current input words have a certain relationship with the preceding and succeeding words, therefore, the input sequence is input into the GRU model from the front and back directions, as shown in Fig.3, and then the hidden layer is used to save the information of the front and back direction text, and finally the output of the two hidden layers is concatenated to obtain the final output of the bidirectional GRU, the bidirectional GRU output code is as follows: BiGRU_output=keras.layers.Bidirectional(GRU (hidden_dim,return_sequences=True))(input).

The bidirectional GRU model is mainly used to extract the contextual semantic information of the words in the sentence, in order to extract the global features of the words in the text, then we combine the features of the CNN channel that are concatenated together by the pooling operation and the output of the bidirectional GRU, the code is as follows: output=keras.layers.merge.concatenate([cnn_output, BiGRU_output]).

The concatenated feature is used as the input of the fully connected layer, the activation function in the fully connected layer uses Relu, after the fully connected layer, the dropout mechanism is added, the purpose of this is to prevent over-fitting, and finally input it into the softmax classifier to get the final classification result, we use the negative cross-entropy loss function as the final loss function of the model in softmax layer, the loss function of the model is shown in Equation 9:

$$Loss = -\frac{1}{n}\sum_{i=1}^{n} y_i ln y_i' + (1-y_i)ln(1-y_i') \tag{9}$$

where y_i is the true value of sample, y_i' is the predicted value of the neural network, n is the number of sample, and the parameters in the network are updated by gradient descent of the loss function.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL ENVIRONMENT

The experimental environment of our paper is as follows: the operating system is Windows10, the CPU is Intel Core i7-6700U, the GPU is GeForce GTX 1080, the memory size is DDR3 8GB, the development environment is Keras 2.1.0, and the development tool uses PyCharm.

B. EXPERIMENTAL DATASET

The dataset used in our paper is public IMDB review dataset and Yelp 2015 review dataset, both datasets are briefly introduced below.

Yelp2015 [20] reviews: The dataset is obtained from Yelp Dataset Challenge in 2015. There are five levels of ratings from 1 to 5.

IMDB [44] reviews: The dataset is obtained from IMDB, which includes 25,000 positive reviews and 25,000 negative reviews.

MR [45]: Movie reviews with one sentence per review. Classification involves binary categories of reviews (positive and negative)

CR: This dataset, consists of reviews of five electronics products downloaded from Amazon and Cnet. The sentences have been manually labeled as to whether an opinion is expressed, and if so, what feature from a pre-defined list is being evaluated.

Each comment in the above four datasets has been artificially set with emotional tags. In order to reduce the random effects brought by the training process, we performed a 10-fold cross-validation experiment on these two datasets.

C. EXPERIMENTAL PREPROCESSING AND MODEL HYPERPARAMETER SETTING

Word vector plays a important role in natural language processing, pre-training with word vector helps to improve the classification accuracy of the model [8]. This paper uses large-scale English Wikipedia data to train the CBOW model to obtain English word vectors. In the training process, we set the word vector dimension to 300. Then divide each sentence into words by space, and then remove the stop words, in this paper, we set the maximum length of the sentence to 100. When the length is less than 100, the sentence is zero-padded, when the length is greater than 100, the sentence is truncated. This paper is based on the Keras deep learning framework. The model's optimization function is the Adam [46], because it can design independent adaptive learning rates for different parameters and speed up network convergence.

In our experiment, in order to obtain more rich emotional feature information, the convolution kernel window size we selected in multi-channel CNN is 3, 4, 5, and the number of each convolution kernel is 128. To prevent over-fitting, we used the L2 regularization and Dropout mechanisms in our experiments. The detailed hyperparameter settings for this model are shown in Table 1.

D. EXPERIMENTAL COMPARISON

In order to verify the classification performance of the proposed model, we use the accuracy as the evaluation metrics, the multi-channel CNN and bidirectional GRU (MC-AttCNN-AttBiGRU) model with the attention mechanism were compared with the following 10 model methods on two datasets.

Support Vector Machine (SVM): Compared with other machine learning methods, SVM has a better effect on sentiment classification [14], in this paper, each word in a sentence is represented by a word vector, then these word vectors are weighted average as input to the SVM for classification.

Fasttext [47]: Fasttext is Facebook's open source text categorization tool. In our experiment, we set the model learning rate to 0.001 and the word vector to 300.

CNN [7]: The common single-channel CNN classification, the word vectors obtained by embedding each word in the sentence are used as the input of the CNN, and then through the convolutional layer, the pooling layer, the fully connected layer, and the final softmax output layer.

MC-CNN: This is a three-channel convolutional layer with convolution kernel window sizes of 3,4 and 5 respectively,

TABLE 1. Hyperparameter setting.

Hyperparameter	Value
Word vector dimension	300
Convolution kernel size	(3,4,5)
Bidirectional GRU hidden layer size	128
Batch_size	128
Epochs	50
Adam learning rate	0.001
L2 regularization penalty λ	0.001
Dropout rate	0.5

that is, 3-channel CNN sentiment classification. The process is the same as the single CNN classification, but finally, a concatenation of the three largest features obtained through the pooling layer is performed to obtain more rich and different granularity of feature information.

RCNN [21]: This model uses CNN and RNN for text classification, and fine-tunes the input word vector during training.

C-LSTM [22]: This is a single-channel model. First, the word vector of each word in the sentence is used as the input of CNN, then the extracted feature is used as the input of LSTM, and finally the final result is obtained through the full connection and softmax layer.

Convolution-GRU [32]: This is a hybird model that combines CNN and GRU. First, the word vector obtained by the method word embedding is used as the input of the CNN, and then the extracted features are used as the input of the GRU, and finally the full connection and softmax layer to get the final classification result.

MC-CNN-LSTM [30]: This is a multi-channel CNN-LSTM model. The model consists of two parts: one is a multi-channel CNN model; the other is an LSTM model. First, multi-channel CNN is used to extract different n-gram features, then the obtained feature is used as the input of LSTM, and finally get the final classification result by the full connection and softmax layer.

BiGRU: Sentence sentiment classification is performed through a bidirectional GRU network model using word vectors after word embedding as input.

ATT-MCNN-BGRUM [41]: This is a single-channel CNN-BiGRU model. First, the multi-channel CNN model is used to extract different n-gram features of the text, then the obtained feature is input into the BiGRU model based on the attention mechanism, finally, Maxout neurons are used to obtain the final classification results.

MC-CNN-BiGRU: This is the concatenation of three-channel CNN and bidirectional GRU channel, the input of three-channel CNN and bidirectional GRU channel are the word vectors of words after word embedding, but three-channel CNN and bidirectional GRU channel do not add attention mechanism.

MC-AttCNN-BiGRU: This is the concatenation of three-channel CNN and bidirectional GRU channel, and attention mechanism is added to all channels of CNN, while the bidirectional GRU channel does not add attention mechanism, and other parameter settings are consistent with MC-CNN-BiGRU.

MC-CNN-AttBiGRU: This is the concatenation of three-channel CNN and bidirectional GRU channel, and the attention mechanism is added on the bidirectional GRU channel, while the three-channel CNN do not add attention mechanism, and other parameter settings are consistent with MC-CNN-BiGRU.

MC-AttCNN-AttBiGRU: This is the final model proposed in this paper, this is the concatenation of three-channel CNN and bidirectional GRU channel, and the attention mechanism is added to the bidirectional GRU channel and the three-channel CNN, taking the word vector obtained after the word embedding as input, the features are extracted through the CNN channel and the bidirectional GRU channel respectively, and finally the extracted two features are merged to perform the final sentiment classification.

E. ANALYSIS OF EXPERIMENTAL RESULTS

In this paper, the proposed final model method is compared with the other neural network model methods on the IMDB, Yelp2015, MR and CR. The experimental results are shown in Table 2:

It can be seen from Table 2 that the proposed MC-AttCNN-AttBiGRU model in this paper has achieved best classification results than the other models on the four public datasets, the classification accuracy on the four datasets Yelp2015, IMDB, MR and CR reached 92.90%, 91.70%, 83.89% and 87.23% respectively. It indicates the superiority of the proposed model in this paper.

It can be seen from Table 2 that the neural network model method(CNN, etc.) is significantly superior to the traditional machine learning model method SVM in terms of classification effect, because the SVM model simply

TABLE 2. Comparison of the models on two public data sets.

Model	Yelp 2015 (Accuracy)	IMDB ((Accuracy)	MR (Accuracy)	CR (Accuracy)
SVM[14]	81.16%	79.86%	71.42%	74.29%
Fasttext[47]	83.96%	81.69%	73.38%	76.28%
CNN[7]	90.36%	88.82%	81.31%	84.32%
MC-CNN	91.06%	89.38%	81.88%	84.76%
RCNN[21]	90.78%	89.20%	81.57%	84.53%
C-LSTM[22]	90.80%	89.13%	81.76%	84.74%
Convolution-GRU[32]	90.91%	90.01%	81.94%	84.97%
MC-CNN-LSTM[30]	91.13%	90.11%	82.27%	85.35%
BiGRU	90.50%	89.00%	81.48%	84.42%
ATT-MCNN-BGRUM[41]	91.30%	90.22%	82.83%	85.87%
MC-CNN-BiGRU	91.60%	90.38%	82.39%	85.43%
MC-AttCNN-BiGRU	92.07%	91.09%	83.24%	86.35%
MC-CNN-AttBiGRU	91.82%	90.82%	82.76%	85.78%
MC-AttCNN-AttBiGRU (Proposed Model)	92.90%	91.70%	83.89%	87.23%

weights the average of all word vectors in a sentence without considering context semantics between sentences and some deeper information, so the classification effect is not good, the accuracy of the neural network based model method is about 10% higher than that the traditional machine learning based method SVM, indicating that the deep learning method is still more effective in text classification tasks; in addition, the classification accuracy of Fasttext model is better than the SVM model, which proves the excellent performance of the Fasttext model in text sentiment analysis, because it is simple and the operation speed is fast, therefore, it is used as a baseline model in this paper; it can be seen from the three comparative experiments of MC-CNN and single-channel CNN, C-LSTM and MC-CNN-LSTM, ATT-MCNN-BGRUM and MC-AttCNN-AttBiGRU, compared with the single-channel model, the multi-channel model can extract abundant emotional feature information, which helps to improve the performance of sentiment classification; by comparing C-LSTM and Convolution-GRU, it can be seen that the classification effect of the GRU model is better than the LSTM model on the four datasets; by comparing MC-CNN-BiGRU, MC-CNN-LSTM and RCNN, C-LSTM, Convolution-GRU, it can be seen that the multi-channel CNN-RNN model is still superior to the single-channel CNN-RNN model; by comparing MC-CNN-BiGRU with CNN and BiGRU, it can be seen that the multi-channel model combining CNN and BiGRU is still much better than the CNN and BiGRU classification alone, because MC-CNN-BiGRU combines the advantages of CNN and BiGRU, which can extract local information of consecutive words between sentences and obtain sentence context semantic information, thus obtaining higher classification accuracy; by the comparison of MC-AttCNN-BiGRU, MC-CNN-AttBiGRU and MC-CNN-BiGRU, it can be seen that the multi-channel model with attention mechanism can significantly improve the classification accuracy, because it allows the model to pay more attention to the words in the sentence that contribute more to the sentiment polarity classification; compared with MC-AttCNN-AttBiGRU and MC-AttCNN-BiGRU, MC-CNN-AttBiGRU, the introduction of attention mechanisms in both the CNN channel and the BiGRU channel can further improve the accuracy of text sentiment classification.

The proposed MC-AttCNN-AttBiGRU model in this paper can not only extract the local feature information of consecutive words in sentences, but also capture the semantic information of sentence context, give full play to the advantages of CNN and BiGRU, and introduce attention mechanism to make the model pay more attention to the more important emotional word in the sentence in the process of feature extraction, and reduce the influence of words that are not important to classification, therefore, the proposed model in this paper can achieve the best results compared to other baseline models.

F. ATTENTION MECHANISM ANALYSIS

In order to more intuitively demonstrate the effect of attention mechanism on feature selection, this group of experiments randomly selected two texts from the Yelp2015 and IMDB

TABLE 3. Text attention mechanism visualization of Yelp2015 dataset.

Dataset:Yelp2015 Category: Positive
This place was great! The room was super clean with a comfortable chair. The staff was friendly and room service was on point. Close enough to the action of the casino floor and food areas but far away enough to just relax.
I've stayed here twice and love it. The rooms are nice, the service is wonderful, and the walk off of the strip is worth saving a bit of money and still being in a really nice hotel.

TABLE 4. Text attention mechanism visualization of IMDB dataset.

Dataset: IMDB Category: Negative
Robert DeNiro plays the most unbelievably intelligent illiterate of all time. This movie is so wasteful of talent, it is truly disgusting. The script is unbelievable. The dialog is unbelievable.
I saw the capsule comment said "great acting." In my opinion, these are two great actors giving horrible performances, and with zero chemistry with one another, for a great director in his all-time worst effort.

datasets for visualization as shown in Table3 and Table4, the attention layer of the model can get the weight of each word, red color indicates that the word with a high attention weight, that is, a word that contributes a large amount to the text classification.

As can be seen from Table 3 and Table 4, the adjective phrases (adjectives + nouns) and adjectives are considered more important words when training. For example, in the Yelp 2015 dataset, "comfortable chair", "nice hotel"; in IMDB dataset, "horrible performances", "worst effort"; in addition, turning words, noun phrases, and comparative words also have high weight values, such as "but", "most", "worst".

This group of experiments proves that the word attention layer can indeed make the model find keywords in the training process and selectively extract data features. In the sentiment classification task, emotional words, adjective phrases, comparative words, and turning words all have a greater impact on sentiment orientation judgment.

V. CONCLUSION AND FUTURE WORK

We propose a network model of multi-channel CNN and bidirectional GRU with attention mechanism for text sentiment orientation analysis in this paper, the model can not only use CNN to extract the local features of adjacent words between texts, but also use bidirectional GRU to capture the global semantic information of sentence contexts, and the introduction of the attention mechanism can better pay attention to the words in the sentence that contribute greatly to the sentiment classification. The results of several experiments on two public datasets fully demonstrate the validity of the proposed model, this multi-channel hybrid model performs better than the single CNN and bidirectional GRU in extraction features, and the proposed model achieves the best classification results on two public datasets compared to other baseline models. Although the proposed model can extract local features by CNN and extract certain context semantic information by bidirectional GRU, this is not enough for the sentiment classification task. The traditional sentiment classification method generally adds some syntactic structure features, and the proposed model in this paper does not use any syntactic structure features, so the future work we will explore how to combine deep learning methods with traditional methods, and consider introducing more text features to further improve classification accuracy.

REFERENCES

[1] B. Liu and S. analysis, *Mining Opinions, Sentiments, and Emotions*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[2] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, Mar. 2016.

[3] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.

[4] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proc. ACL-Conf. Empirical Methods Natural Lang. Process., Assoc. Comput. Linguistics*, Jul. 2002, pp. 79–86.

[5] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*. New York, NY, USA: ACM, Aug. 2004, pp. 168–177.

[6] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," 2014, *arXiv:1404.2188*. [Online]. Available: http://arxiv.org/abs/1404.2188

[7] Y. Kim, "Convolutional neural networks for sentence classification," 2014, *arXiv:1408.5882*. [Online]. Available: http://arxiv.org/abs/1408.5882

[8] Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," 2015, *arXiv:1510.03820*. [Online]. Available: http://arxiv.org/abs/1510.03820

[9] X. Wang, Y. Liu, C. Sun, B. Wang, and X. Wang, "Predicting polarities of tweets by composing word embeddings with long short-term memory," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, vol. 1, 2015, pp. 1343–1353.
[10] P. Liu, S. Joty, and H. Meng, "Fine-grained opinion mining with recurrent neural networks and word embeddings," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Sep. 2015, pp. 1433–1443.
[11] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*. [Online]. Available: http://arxiv.org/abs/1301.3781
[12] J. Kamps, M. Marx, R. J. Mokken, and M. D. Rijke, "Words with attitude," in *Proc. Belgian-Netherlands Conf. Artif. Intell.*, 2002, pp. 332–341.
[13] Z. Dong, Q. Dong, and C. Hao, *HowNet and the Computation of Meaning*. Singapore: World Scientific,2006.
[14] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proc. 42nd Annu. Meeting Assoc. Comput. Linguistics-ACL*, Jul. 2004, pp. 271–278.
[15] D. Poirier, C. Bothorel, E. G. D. Neef, and M. Boullé, "Automating opinion analysis in film reviews: The case of statistic versus linguistic approach," in *Proc. Lang. Resour. Eval.*, 2011, pp. 125–140.
[16] H. Y. Lee and H. Renganathan, "Chinese sentiment analysis using maximum entropy," in *Proc. Workshop Sentiment Anal. AI Meets Psychol. (SAAIP)*, Nov. 2011, pp. 89–93.
[17] S. Naz, A. Sharan, and N. Malik, "Sentiment classification on Twitter data using support vector machine," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Dec. 2018, pp. 676–679.
[18] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12 pp. 2493–2537, Aug. 2011.
[19] W. Yin and H. Schütze, "Multichannel variable-size convolution for sentence classification," 2016, *arXiv:1603.04513*. [Online]. Available: http://arxiv.org/abs/1603.04513
[20] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Sep. 2015, pp. 1422–1432.
[21] R. Wang, Z. Li, J. Cao, T. Chen, and L. Wang, "Convolutional recurrent neural networks for text classification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 2267–2273.
[22] C. Zhou, C. Sun, Z. Liu, and F. C. M. Lau, "A C-LSTM neural network for text classification," 2015, *arXiv:1511.08630*. [Online]. Available: http://arxiv.org/abs/1511.08630
[23] 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
[24] S. Ruder, P. Ghaffari, and J. G. Breslin, "A hierarchical model of reviews for aspect-based sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 999–1005.
[25] G. Rao, W. Huang, Z. Feng, and Q. Cong, "LSTM with sentence representations for document-level sentiment classification," *Neurocomputing*, vol. 308, pp. 49–57, Sep. 2018.
[26] S. Sachin, A. Tripathi, N. Mahajan, S. Aggarwal, and P. Nagrath, "Sentiment analysis using gated recurrent neural networks," *Social Netw. Comput. Sci.*, vol. 1, no. 2, pp. 1–13, Mar. 2020.
[27] X. Wang, W. Jiang, and Z. Luo, "Combination of convolutional and recurrent neural network for sentiment analysis of short texts," in *Proc. 26th Int. Conf. Comput. Linguistics, Tech. Papers (COLING)*, Dec. 2016, pp. 2428–2437.
[28] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu, "Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling," 2016, *arXiv:1611.06639*. [Online]. Available: http://arxiv.org/abs/1611.06639
[29] Y. Zhang, H. Yuan, J. Wang, and X. Zhang, "YNU-HPCC at EmoInt-2017: Using a CNN-LSTM model for sentiment intensity prediction," in *Proc. 8th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal.*, 2017, pp. 200–204.
[30] H. Zhang, J. Wang, J. Zhang, and X. Zhang, "YNU-HPCC at SemEval 2017 task 4: Using a multi-channel CNN-LSTM model for sentiment classification," in *Proc. 11th Int. Workshop Semantic Eval. (SemEval)*, 2017, pp. 796–801.
[31] B. Sun, F. Tian, and L. Liang, "Tibetan micro-blog sentiment analysis based on mixed deep learning," in *Proc. Int. Conf. Audio, Lang. Image Process. (ICALIP)*, Jul. 2018, pp. 109–112.
[32] Z. Zhang, D. Robinson, and J. Tepper, "Detecting hate speech on Twitter using a convolution-GRU based deep neural network," in *Proc. Eur. Semantic Web Conf.*, Jun. 2018, pp. 745–760.
[33] M. H. Abd El-Jawad, R. Hodhod, and Y. M. K. Omar, "Sentiment analysis of social media networks using machine learning," in *Proc. 14th Int. Comput. Eng. Conf. (ICENCO)*, Dec. 2018, pp. 174–176.
[34] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, Jun. 2016, pp. 1480–1489.
[35] L. Wang, Z. Cao, G. de Melo, and Z. Liu, "Relation classification via multi-level attention CNNs," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, Aug. 2016, pp. 1298–1307.
[36] 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.
[37] J. Cheng, S. Zhao, J. Zhang, I. King, X. Zhang, and H. Wang, "Aspect-level sentiment classification with HEAT (HiErarchical ATtention) network," in *Proc. ACM Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 97–106.
[38] 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
[39] H. Han, X. Li, S. Zhi, and H. Wang, "Multi-attention network for aspect sentiment analysis," in *Proc. 8th Int. Conf. Softw. Comput. Appl. (ICSCA)*, 2019, pp. 22–26.
[40] Y. Gao, J. Liu, P. Li, and D. Zhou, "CE-HEAT: An aspect-level sentiment classification approach with collaborative extraction hierarchical attention network," *IEEE Access*, vol. 7, pp. 168548–168556, 2019.
[41] H. Yuan, X. Zhang, W. Niu, and K. Cui, "Sentiment analysis based on multi-channel convolution and bi-directional GRU with attention mechanism," *J. Chin. Inf. Process.*, vol. 33, no. 10, pp. 109–118,2019.
[42] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: http://arxiv.org/abs/1406.1078
[43] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*. [Online]. Available: http://arxiv.org/abs/1409.0473
[44] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1, Jun. 2011, pp. 142–150.
[45] B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," in *Proc. 43rd Annu. Meeting Assoc. Comput. Linguistics-ACL*, 2005, pp. 115–124.
[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[47] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," 2016, *arXiv:1607.01759*. [Online]. Available: http://arxiv.org/abs/1607.01759

YAN CHENG received the Ph.D. degree from the School of Electronic and Information Engineering, Tongji University. She was a Postdoctoral Researcher with the Computer Science and Technology Postdoctoral Station, Tongji University. She is currently a Professor with the School of Computer Information Engineering, Jiangxi Normal University. Her research interests include artificial intelligence and intelligent information processing, deep learning, and sentiment analysis.

LEIBO YAO received the bachelor's degree from the Nanchang Institute of Technology. He is currently pursuing the master's degree with the School of Computer Information Engineering, Jiangxi Normal University. His research interests include deep learning and sentiment analysis.

GUOXIONG XIANG received the bachelor's degree from Jiangxi Normal University. He is currently a Professor with the School of Journalism and Communication, Jiangxi Normal University. His research interests include machine learning and deep learning.

GUANGHE ZHANG received the M.E. degree from Jiangxi Normal University, Nanchang, China, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences (CAS), China. He is currently an Assistant Teaching Professor with Jiangxi Normal University. His current research interests include body sensor network security, biosignal processing, and development of software for education.

TIANWEI TANG received the Ph.D. degree from Beijing Normal University. He was a Postdoctoral Researcher with Tsinghua University. He is currently a Professor with Jiangxi Normal University. His research interests include machine translation and sentiment analysis.

LINHUI ZHONG received the Ph.D. degree from the School of Electronics Engineering and Computer Science, Software Engineering Institute, Peking University. He is currently an Associate Professor with the School of Computer and Information Engineering, Jiangxi Normal University. His main research interests include software architecture, software evolution analysis and management, and software engineering.