# Sentiment Analysis Algorithm Based on BERT and Convolutional Neural Network

2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC) | 978-1-7281-9018-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/IPEC51340.2021.9421110

Rui Man
School of Computer and Information Security
Guilin University of Electronic Technology
Guilin, China

Ke Lin
School of Computer and Information Security
Guilin University of Electronic Technology
Guilin, China
Linke@guet.edu.cn

*Abstract*—**The explosion of Internet information makes timely analysis and monitoring of Internet public opinion more and more important, and sentiment analysis of public opinion events is even more important. The traditional word2vec model cannot fully express the information contained in words. It is proposed to use the BERT model as the article feature extraction model; and use the deep convolutional neural network to extract the local information of the article, and then connect the fully connected network to classify the article, so as to achieve sentiment analysis purpose. Experimental results on public data sets show that sentiment analysis algorithms based on BERT and convolutional neural networks are better than traditional sentiment analysis algorithms.**

***Keywords—Sentiment analysis, BERT, CNN, public opinion analysis***

## I. Introduction

The rise of social apps has made the Internet an important channel for people to obtain information, and these platforms have also become platforms for people to express their personal opinions. Comments on social events, product reviews, and film reviews have produced a large number of emotionally inclined natural language texts. A better understanding of these texts can better capture public opinion and grasp user interests. With the rapid growth of information scale, it is impossible to complete this task only by manual processing, which promotes the development of text sentiment analysis technology. At present, the main research methods of text sentiment analysis are still based on traditional machine learning algorithms. Construct structured text information features by artificially designing features, and then use machine learning methods for analysis. Commonly used text sentiment analysis methods include naive Bayes, support vector machines, maximum entropy methods, etc. These methods can be divided into shallow learning methods. The shallow learning methods are less computationally intensive and easy to implement. However, the limited ability to express complex functions limits the generalization ability of complex classification problems. In order to make up for this shortcoming, artificially constructed functions are introduced into the model, such as sentiment dictionaries, grammar and grammatical analysis using artificial tags. Although these methods can effectively improve the accuracy of text sentiment analysis, they still require manual labeling of data, which is time-consuming and labor-intensive, and requires certain prior knowledge. Therefore, with the continuous development of the Internet, the continuous expansion of text data limits the development of these methods. This article uses a method based on BERT and convolutional neural network to avoid relying on artificially constructed features.

## II. Related research

Text sentiment analysis mainly analyzes the content of the text to judge the sentimental tendency expressed by the text, and finds the user's attention to a certain event. Since Pang and Lee [1] proposed the work on sentiment analysis in 2002, after many scholars' research, great development has been obtained. Sentiment analysis technology can be roughly divided into rule-based methods and statistics-based methods, among which the machine learning method based on sentiment dictionary is the main method at present. Pang and Lee used naive Bayes, support vector machines, and maximum entropy models based on text classification techniques in traditional natural language processing, and achieved good results in movie reviews; as in [2] used the PMI method to extend the basic Then use the semantic polarity (ISA) algorithm to analyze the emotional text, and the accuracy rate of processing general corpus data reaches 74%; as in [3] proposed a method for matching emotional words in a specific field to determine the emotional polarity ; Yang Xiaoping [4] used the word2vec tool to train a set of word vectors from a massive corpus, and obtained good experimental results on the general domain data set; as in [5] used the naive Bayes algorithm and K- The NN algorithm performs sentiment analysis on movie reviews and hotel reviews, and found that Naive Bayes is better than K-NN in movie reviews, but in hotel reviews, the accuracy of the two is not much different. Since deep learning was proposed by Hinton et al. in 2006, with the successful application of deep learning methods in the fields of computer vision and speech recognition, more and more deep learning techniques have also been applied in the direction of natural language processing. TANG et al. [6] constructed a neural network model to learn word vector representations containing emotional features and semantic features, and combined with manual features as composite features to train classifiers. Mikolov et al. [7] borrowed the idea of Log-Bilinear model and proposed the word2vec model, which realized the two frameworks of CBOW and Skip-gram, and as Google open sourced its code, word embedding was applied to many fields of natural language processing; With Kim [8] using convolutional neural networks to classify sentences; as in [9] proposed a semi-supervised learning model that combines convolutional neural networks with k-means algorithms, using a small amount of labeled data to aggregate short texts. Class; as in [10] proposed to use the attention model to achieve good results in machine translation, and was subsequently applied to the Google neural network translation system.

The BERT model and deep learning method proposed in 2018 have achieved good results. This paper proposes Chinese text sentiment analysis based on BERT and CNN. After experimental verification, the method proposed in this paper is effective.

## III. Sentiment Analysis Algorithm Based on BERT and CNN

### A. BERT

The BERT model implements the Encoder-Decoder structure based on the Transformer framework, the structure of the Transformer is shown in Figure 1.

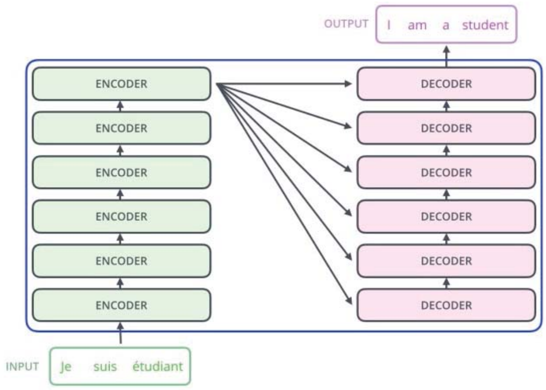


Fig. 1. Transformer framework

Transformer also uses the Encoder-Decoder framework, which is the same as the Attention model, but its structure is much more complicated than Attention. Its Encoder module is composed of 6 encoders, and the Decoder module is composed of 6 decoders. Each encoder contains a self-attention layer and a feed-forward neural network. Self-attention allows the current node not only to focus on the current word, but also to capture the semantics of the context. Each decoder node contains a self-attention layer, an attention layer and a feedforward neural network. The attention layer can help the current node obtain the current key content that needs to be paid attention to.

Figure 2 shows Self-attention, each Self-attention needs to calculate the three vectors of Query, Key, and Value. The dimensions of each vector are 512 dimensions. They are the result of multiplying the input vector embedding by a matrix. This matrix is initialized randomly at the beginning, and then the specific content of the matrix is obtained through training. The dimension of the matrix is (64,512), and the value of the matrix is always updated during the back propagation process. . In order to improve the calculation speed, we use the matrix method to directly calculate the matrix of Query, Key, Value, and then directly multiply the embedding value with the three matrices, and multiply the new matrix Q and K obtained by A constant, do the softmax operation, and finally multiply the V matrix.

$$Attention(Q,K,V)=softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

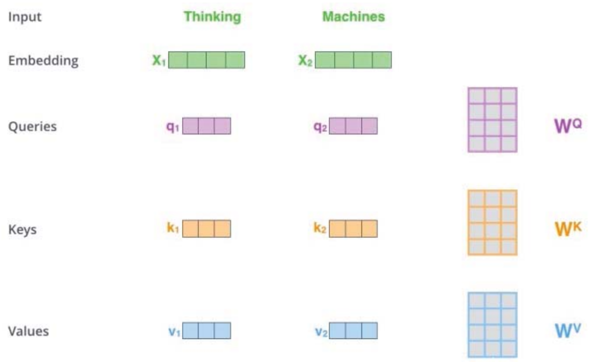


Fig. 2. The Self Attention

The transformer model also lacks a way to explain the order of words in the input sequence. In order to deal with this problem, the transformer adds an additional vector Positional Encoding to the input of the encoder layer and the decoder layer. The dimension is the same as the dimension of the embedding. This vector uses a very unique method to let the model learn this value. This vector Can determine the position of the current word, or the distance between different words in a sentence. The specific calculation method of this position vector is as follows:

$$PE(pos,2i)=sin(\frac{pos}{10000^{2i/d_{model}}}) \tag{2}$$

$$PE(pos,2i+1)=cos(\frac{pos}{10000^{2i/d_{model}}}) \tag{3}$$

### B. CNN

Overall architecture of convolutional neural network: Convolutional neural network is a multi-layer supervised learning neural network. The convolutional layer and pool sampling layer of the hidden layer are the core modules to realize the feature extraction function of the convolutional neural network. The network model adopts the gradient descent method to minimize the loss function to reversely adjust the weight parameters in the network layer by layer, and improves the accuracy of the network through frequent iterative training. The low hidden layer of the convolutional neural network is composed of a convolutional layer and a maximum pool sampling layer alternately, and the upper layer is a hidden layer and a logistic regression classifier corresponding to the traditional multi-layer perceptron. The input of the first fully connected layer is the feature image obtained by feature extraction of the convolutional layer and the sub-sampling layer. The final output layer is a classifier, which can use logistic regression, Softmax regression or even support vector machine to classify the input image.

The convolutional neural network structure includes: convolutional layer, down sampling layer, and fully-linked layer, as shown in Figure 3. Each layer has multiple feature maps, each feature map extracts a feature of the input through a convolution filter, and each feature map has multiple neurons.

Convolutional layer: The reason for using the convolutional layer is that an important feature of the convolution operation is that through the convolution operation, the original signal characteristics can be enhanced and the noise can be reduced.

Down sampling layer: The reason for using downsampling is that, according to the principle of image local correlation, sub-sampling the image can reduce the amount of calculation while maintaining the image rotation invariance.

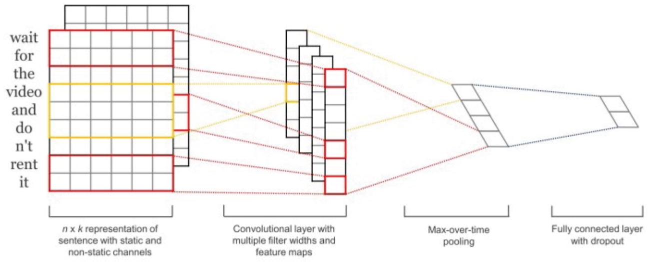


Fig. 3. CNN network

### C. BERT-CNN

After using the BERT model to obtain the characteristics of the word, it is connected to the CNN network for text classification. Obtain the final token-level output of the BERT model, and use this as embedding_inputs into the convolutional network CNN.The structure of the BERT-CNN is shown in Figure 4.

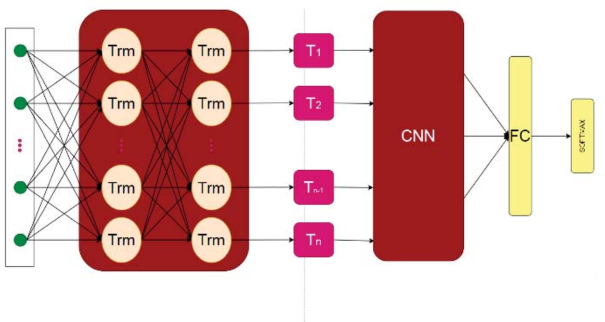


Fig. 4. BERT-CNN

## IV. EXPERIMENT

The experimental data set selected the public data set, which is the hotel review corpus (ChnSentiCorp) compiled by Dr. Tan Songbo of the Chinese Academy of Sciences. The total number of corpus scales is 7766, of which 5322 are positive comments and 2,444 are negative comments. The corpus is divided into a training set and a verification set. In the training set, there are 4360 positive comments, 1996 negative comments, and the rest of the corpus is used as a verification set.

An example of the data is as table I (where the negative label is 0 and the positive label is 1):

TABLE I. WEIBO DATA SET EXAMPLE

| id | label | content |
|---|---|---|
| 1 | 1 | It is close to Chuansha Highway, but the bus directions are wrong. If it is "Cailu Line", it will be very troublesome. It is recommended to use another route. The room is relatively simple. |
| 2 | 1 | Business big bed room, the room is very big, the bed is 2M wide, the overall feeling is economical and affordable! |
| 3 | 1 | The breakfast is too bad, no matter how many people go there, no food is added there. The hotel should pay attention to this issue. The room itself is good. |
| 4 | 0 | This is the worst hotel I booked with Ctrip. It is recommended to cancel the cooperation with this hotel. |
| 5 | 0 | Like the slippers in the warehouse, the bathroom is good and the breakfast is poor. |
| 6 | 0 | Breakfast is 30, cannot be refunded, nor can it be expected, it is too reasonable. I will no longer hotel in the future. The room is smelly, repair too. |

The experimental model selected the following models:

word2vec-svm: Use word2vec to extract the features of words, and the classification model adopts the svm model;

word2vec-cnn: Use word2vec to extract the features of words, and the classification model adopts the cnn model;

word2vec-Att-cnn: Use word2vec to extract the features of words, and the classification model uses the Attention mechanism combined with the cnn model;

And the bert-cnn used in this article, the results use accuracy, recall and F1 value as evaluation indicators. The experimental results are as table II:

TABLE II. EXPERIMENTAL RESULTS

| algorithm | Accuracy | Recall rate | F1 |
|---|---|---|---|
| word2vec-svm | 0.813 | 0.811 | 0.812 |
| word2vec-cnn | 0.852 | 0.849 | 0.850 |
| word2vec-Att-cnn | 0.873 | 0.871 | 0.872 |
| bert-cnn | 0.905 | 0.901 | 0.903 |

From the experimental results in the above table, it can be seen that if word2vec is used as the feature extraction model, CNN is better than SVM for the classification model; at the same time, if the attention mechanism is added, the effect will increase by 2%; and bert is used as the feature the effect of the extraction model is the best, the F1 value can reach 0.903, and the accuracy and recall are the best.

The learning ability of CNN is stronger than the learning ability of SVM, and it can better grasp the local characteristics of the data; compared with the Attention mechanism, the learning ability of the Multi-Head Attention mechanism is also stronger.

## V. CONCLUSION

This paper proposes a sentiment analysis algorithm that combines BERT and CNN. It uses BERT to extract the features of each word and uses it as the input of CNN. After convolution and pooling, it is connected to the Softmax layer for classification. Experiments show that the model is feasible Sex and effectiveness. At present, in the public opinion industry, more and more text information need to be automatically processed. The method proposed in this article still has a lot of room for improvement. In the future, the two-way LSTM and convolutional neural network will be combined to improve the effect of sentiment analysis.

## REFERENCES

[1] Pang Bo,Lee L.Opinion mining and sentiment analysis[J].Joumal Foundations and Trends in Information Retrieval,2008,2(2):1-135.

[2] TURNEY P D, LITTMAN M L. Measuring praise and criticism: Inference of semantic orientation from association[J]. Acm Transactions on Information Systems ,2003 ,21(4) :315-346.

[3] Ding Xiaowen,Liu Sing,Yu P S.A holistie lexicon—based approach to opinion mining[C]//Proc of Conference on Web Search and Web Data Mining.New York:Association for Computing Machinery, 2008:231—240.

[4] Yang Xiaoping, Zhang Zhongxia, Wang Liang, et al. Automatic construction and optimization of emotional dictionary based on Word2Vec [J]. Computer Science, 2017, 44(1): 42-47.

[5] DEY L, CHAKRABORTY S, BISWAS A, et al. Sentiment analysis of review datasets using Naive Bayes and K-NN classifier [J].

Information Retrieval ,2016 ,8(4) :54-62.

[6] TANG Duyu, WEI Furu, QIN Bing, et al. Coooolll: A deep learning system for twitter sentiment lassification[C]// Proceedings of the 8th International Workshop on Semantic Evaluation. Dublin, Ireland: ACL, 2014: 208-212.

[7] Mikolov T,Sutskever I,Chen Kai,et a1.Distributed representations of words and phrases and their compositionality[C]//proc of Intemational Conference on Neural Information Processing Systems. USA:Curran Associates Inc.,2013:3111—3119.

[8] Kim Y.Convolutional neural networks for sentence classification [EB/OL].(2014-08—25).https://arxiv.org/abs.1408.5882.

[9] WANG Zhiguo, MI Haitao, ITTYCHERIAH A. Semisupervised clustering for short text via deep epresentation learning [EB/OL]. 2016-02-22. http://arxiv.org/abs/1602.06797.

[10] Bahdanau D,Cho K,Bengio Y.Neural machine translation by joinfly learning to aliguand translate[C]//Proc of International Conference on Learning RepresentationS: 2015.