Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 2, No. 3, 2024

Study on text classification model combining BERT and convolutional neural network

Jia Wei *, Xiangguo Sun

Mechanical Engineering College, Sichuan University of Science and Engineering, Yibin, China

* **Corresponding author**: Jia Wei (Email: 18040480929@163.com)

Abstract: Text classification has been widely used in many practical scenarios such as sentiment analysis and topic classification, and is an important task in the field of natural language processing. In recent years, research on pre-training model based on Bidirectional Encoder Representations from Transformers (BERT) has made remarkable progress. However, at present, most studies are based on open source databases. In this study, a text classification model combining BERT model and convolutional neural network (CNN) is proposed to complete the text emotion classification task through the self-built dialect text database. This paper demonstrates the effectiveness and superiority of this model in Chinese text classification tasks by comparing experiments and performance evaluation.

Keywords: Text classification,BERT-CNN,Dialect database.

1. Introduction

Text Classification is one of the fundamental tasks in natural language processing (NLP), the goal of which is to automatically assign text into predefined categories. Traditional text classification methods often rely on feature engineering and machine learning algorithms such as support vector machines (SVMS) and Naive Bayes[1], which often exhibit limitations when dealing with complex linguistic features. In recent years, the development of deep learning technology has brought new breakthroughs for text classification, in which BERT is a powerful pre-trained model. However, the output sequence of BERT is still one-dimensional, and how to make full use of this information to improve the classification performance is an important direction of current research. At present, many scholars have achieved remarkable results in many NLP tasks. For example, Lu Jialai et al. [2]studied the Bert-TextCNN multi-label classification based on title and text. Compared with the text-based model only, this setting can provide a valuable reference for open source threat intelligence classification in terms of performance. Zou Liwei et al. [3]combined typhoon attribute data and multi-label classification method, and took BERT-BiLSTM as a classification model, proposed a typhoon disaster identification method based on Weibo text and deep learning, and identified the severe/super typhoon disasters that hit Guangdong Province from 2010 to 2019 through bidirectional context modeling and self-attention mechanism. It can capture richer semantic information. Ahmed R. Abas[4] et al used BERT to train the word semantic representation language model, and adopted the BERT-CNN model to achieve 94.7% accuracy and 94% F1 score in the semeval2019 task3 dataset. Achieved 75.8% accuracy and 76% F1 score in ISEAR dataset.

However, most of the current research is on open source data, and there are few studies on dialect texts. But most people still use dialects in daily life. Therefore, this paper proposes a model combining BERT and convolutional neural network (CNN) to extract local features from the self-constructed Sichuan dialect text database through BERT pre-training model and convolutional operations, so as to improve the performance of text classification.

2. Research Methods

2.1. BERT Model

BERT model is derived from Google's paper Pre-training of Deep Bidirectional Transformers for Language Understanding. In order to pre-train the model, the paper designs two tasks. The first task is to improve performance through language model training. Using the MaskLM method, randomly select a sentence to predict it, then randomly replace the original word with the symbol [Mask], and let the model learn and fill in the part replaced by [Mask]. The second task is to add a statement level continuity prediction task based on the bidirectional language model. That is, predict the text input for two paragraphs. By introducing this task, the model can learn the relationships between successive pieces of text. Unlike traditional RNNS and LSTMS, BERT can process data in parallel while extracting relational features between words in a sentence. By extracting multiple levels of relational features, BERT can more fully understand the meaning of the sentence[5]. In contrast to word2vec, BERT can understand the meaning of words based on context, thus avoiding ambiguity. BERT uses only Transformer's encoder module. In the original paper, the BERT model is built using 12-layer and 24-layer Transformer encoders, as shown in equation 1,2:

$$BERT_{BASE}: L = 12, H = 768, A = 12, TotalParameters = 110M \quad (1)$$

$$BERT_{LARGE}: L = 24, H = 1024, A = 16, TotalParameters = 340M \quad (2)$$

Where, L represents the number of Transformer encoder blocks, that is, the number of layers; H represents the dimension of the hidden layer; A is the number of self-attentional heads. In all cases, we set the dimensions of the feedforward network (i.e., the feedforward layer in the Transformer encoder) to 4 times the dimensions of the hidden layer, i.e. 4H. Specifically, when the dimension of the hidden layer H is 768, the dimension of the feedforward network is 3072; When H is 1024, the dimension is 4096. BERT is a pre-trained model based on Transformer that captures complex relationships between words through bidirectional context modeling. The core of BERT is a multi-layer bidirectional encoder with a self-attention mechanism and a feedforward neural network. BERT's pre-training includes mask language models and next sentence prediction tasks to help learn context and text structure[6].

Masked Language Model (MLM)

Masked language models learn the contextual representation of each word. Segmentation of input text into words by input text, usually using "word segmentation", which splits words into smaller sub-word units.

The masking operation randomly selects some of the words being typed and replaces them with a special [Mask] label. For example, "i love natural language processing" might be masked as "i love [mask] processing".

Predictive mask words :BERT uses bidirectional context understanding to predict the actual content of mask words. This process includes not only the before and after parts, but also all the words in the context. BERT inputs include Segment Embeddings, Position Embeddings, and Token Embeddings. The following diagram shows the representation in the BERT model of text sentence input, as shown in Figure 1:

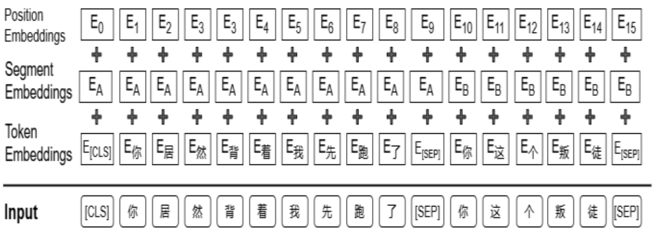

Figure 1: Input representation of BERT

2.2. Convolutional Neural Networks (CNN)

Convolutional neural networks (CNNS) perform well in tasks such as image classification and object detection. CNN extracts local features by convolutional kernel and reduces dimensions by pooling layer. Although CNN is mainly used for image processing, its local feature extraction capability is also applicable to text data, which can capture local semantic features between words.

2.3. Model structure

The BERT-CNN model proposed in this study combines the advantages of BERT and CNN[7][8], and its specific structure is as follows:

The model mainly includes the following four parts: Specifically, the structure of the model is as follows: BERT model: The pre-trained BERT model 'Bert-base-Chinese' is used to process chinese text and can provide context-rich word embeddings. Convolution layer: Three convolution cores of different sizes (3, 4, 5) are set up to capture local features of different ranges. Pooling layer: An adaptive maximum pooling layer is used to unify the convolution results to a fixed size and reduce the dimension. Full connection layer: The pooled feature vectors are imported into the classifier for final category prediction [9][10][11].

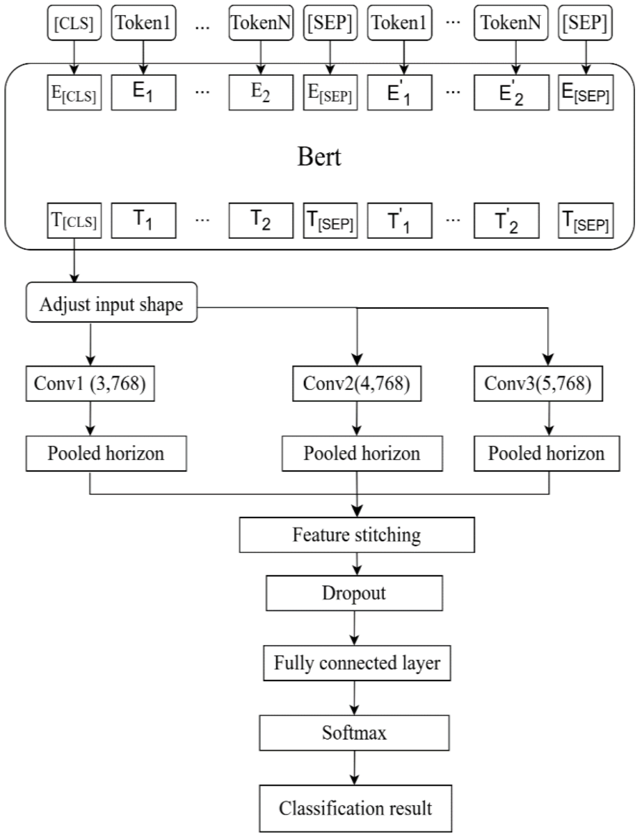

Figure 2: The model structure of this paper

3. Experiment

3.1. Data set and preprocessing

The text data set of Sichuan dialect used in this experiment comes from films and TV dramas of Sichuan dialect, and the text data is extracted. Each data set contains 6 categories, and each category contains several text samples. The specific categories and quantities are shown in Table 1. The data set is divided into a training set and a validation set. The training set is used for model training and the validation set is used for model evaluation.

Table 1: Data set categories and quantities

Category	Angry	Happy	Fear	Neutral	Sad	Surprise
Number	1102	1098	1008	1112	1011	1002

Data preprocessing is a key step in model training. The data processing process of this study is as follows:data read: Read text data from the specified directory and assign labels according to the folder name. Data segmentation: 'train_test_split' was used to divide the data set into the training set and the verification set, the ratio was 80:20%.BERT word segmentation is used to encode the text, including filling, truncating and generating attention masks to adapt to model input requirements.

3.2. Experimental Setup

In the process of model training, the following Settings were adopted: In terms of hardware environment, the PyTorch framework was used and the training was carried out on the GPU with CUDA support to accelerate the model training. In terms of hyperparameter Settings, the maximum sequence length is 256, batch size is 16, learning rate is 2e-5, and dropout rate is 0.2. In terms of loss function, the cross entropy loss function (CrossEntropyLoss) is adopted, which is suitable for multi-classification tasks. In terms of optimizer, AdamW optimizer is selected, which has weight attenuation capability and is suitable for large-scale pre-training models. In the evaluation process, the classification performance is measured by calculating the accuracy of the model on the verification set.

3.3. Experimental Results

Three sets of experiments were designed, taking CNN, BERT and BerT-CNN as models, to record the loss value of the model in the training process and the accuracy of the verification set respectively, as shown in Table 2 below:

Table 2: Accuracy rate and loss of different models

Model	Loss	Accuracy
CNN	0.8398	0.73
BERT	0.5301	0.77
BERT-CNN	0.4765	0.78

4. Conclusion

In this paper, a text classification model combining BERT and CNN is proposed, and an experiment is carried out in Chinese text classification task. The experimental results show that the model performs well in text emotion classification and can be applied to robot text understanding. Future research will focus on optimizing the model structure and exploring more application scenarios. Specific research directions include: finding more efficient model structure, reducing computational overhead and improving training speed; Multi-task learning is combined with other NLP tasks to improve the universality of the model. The data enhancement technique is used to increase the training samples and improve the robustness and generalization ability of the model.

References

[1] Zhou Jun, He Li, Han Weihong, et al. Empirical research on propensity classification of Chinese comments based on Machine learning [J]. Information Network Security, 2013, (10):164-166.

[2] Lu Jiali. Multi-label classification method of open source threat intelligence text based on Bert-TextCNN [J]. Research on Information Security, 2018,10(08):760-768.

[3] Zou Liwei, He Zhi, Zhou Chengle. Typhoon disaster identification method based on Weibo and deep learning: A case study of the transit typhoon in Guangdong Province from 2010 to 2019 [J]. Tropical geography, 2024, 44 (6) : 1079-1089. The DOI: 10.13284 / j.carol carroll nki RDDL. 003882.

[4] Abas R A ,Elhenawy I ,Zidan M , et al.BERT-CNN: A Deep Learning Model for Detecting Emotions from Text[J]. Computers, Materials & Continua,2022,71(2):2943-2961.

[5] An Xin, Dai Zibiao, Li Yang, et al. End-to-end speech synthesis method based on BERT [J]. Computer Science, 202,49(04): 221-226. (in Chinese)

[6] Palace success. Sentiment analysis unsupervised adaptation research [D]. Nanjing university of science and technology, 2021. The DOI: 10.27241 /, dc nki. Gnjgu. 2021.001700.

[7] Xu Jianfei, Wu Yuecheng. Research on News text Classification based on BERT-BiLSTM-CNN model [J]. Software engineering, 2023, 26 (6) : 11-15. DOI: 10.19644 / j.carol carroll nki issn2096-1472.2023.006.003.

[8] Zhi Shiyao, Wu Zhenru, Chen Tao, et al. Research on Sentiment Analysis of microblog comments based on ALBERT-BiLSTM-Att [J]. The computer age, 2022, (02) : 19-22. DOI: 10.16644 / j.carol carroll nki cn33-1094 / tp. 2022.02.006.

[9] Yanling L ,Jiaye W ,Xudong L .BERT-CNN based evidence retrieval and aggregation for Chinese legal multi-choice question answering[J].Neural Computing and Applications, 2024,36(11):5909-5925.

[10] Bi X ,Zhang T .Pedagogical sentiment analysis based on the BERT-CNN-BiGRU-attention model in the context of intercultural communication barriers.[J].PeerJ. Computer science,2024,10e2166.

[11] Hamed K A ,Saleh F A ,Akhil K , et al.A novel multi-layer feature fusion-based BERT-CNN for sentence representation learning and classification[J].Robotic Intelligence and Automation,2023,43(6):704-715.