Research on Chinese Emotion Classification using BERT-RCNN-ATT

FENG LI, YINTONG HUO, LINGLING WANG*
School of Management Science and Engineering,
Anhui University of Finance and Economics,
Bengbu 233030,
CHINA

Abstract: - Emotional classification is the process of analyzing and reasoning subjective texts with emotional color, that is, analyzing whether their emotional tendencies are positive or negative. Aiming at the problems of massive data and nonstandard words in the existing Chinese short text emotion classification algorithm, the traditional BERT model does not distinguish the semantics of words with the same sentence pattern clearly, the multi-level transformer training is slow, time-consuming, and requires high energy consumption, this paper proposes to classify users' emotions based on BERT-RCNN-ATT model, and extract text features in depth using RCNN combined with attention mechanism, Multi task learning is used to improve the accuracy and generalization ability of model classification. The experimental results show that the proposed model can more accurately understand and convey semantic information than the traditional model. The test results show that compared with the traditional CNN, LSTM, GRU models, the accuracy of text emotion recognition is improved by at least 4.558%, the recall rate is increased by more than 5.69%, and the F1 value is increased by more than 5.324%, which is conducive to the sustainable development of emotion intelligence combining Chinese emotion classification with AI technology.

Key-Words: Online; Comment Text; LSTM; Sentiment Analysis.

Received: May 18, 2022. Revised: January 9, 2023. Accepted: February 15, 2023. Published: March 17, 2023.

1 Introduction

According to the 49th Statistical Report on Internet Development in China [1] released by China Internet Network Information Center (CNNIC) in Beijing on February 25, 2022, as of December 2021, the number of Internet users in China has reached 1.032 billion, an increase of 42.96 million over December 2020, and the Internet penetration rate has reached 73.0%. The scale of Internet users in China has grown steadily. The Internet has penetrated into our lives like food, clothing, housing and transportation. For the massive amount of text information in the network, how to achieve automatic and efficient analysis of these comments and the emotions contained in the text has become a focus of attention [2]. The research on NLP (Natural Language Processing) came into being at the historic moment, but it still faces a series of difficulties and challenges. Information technology drives the paradigm shift of communication science, thus increasing the dependence of discipline research on text data mining technology [3]. Emotional classification is an important branch of the NLP field, which is widely used in many aspects, such as artificial customer service and emotional pacification, classification of depressive patients, and criminal investigation assisted psychological research [4].

Traditional emotion classification research is mainly based on emotion dictionary and machine learning. Early text emotion analysis work usually focused on building an emotion dictionary, establishing a direct mapping relationship between the dictionary and emotion, and then using statistical methods to extract features for analysis [5].Due to the lack of deep extraction of text information, neural networks as a way to achieve machine learning has been proposed. When the neural network gradually matures, researchers put into the method of deep learning and proposed "word vector" to solve the problem of data sparsity in high-dimensional space, and can even add more features [6].Pang (2002) et al. [7]were the first to apply machine learning methods to emotion orientation classification. The experiment shows that using word unary model as the feature and Bayesian and SVM as the classifier have achieved good results. Deep learning is considered as a new research field in machine learning, which has received more attention in recent years. Zhao and others described the challenges and opportunities they are facing in the future for multi-modal emotion recognition of deep learning [8]. Jacob Devlin pointed out in the pre training of BERT: Deep Bidirectional Converter for Language Understanding that BERT is simple in concept and rich in experience. It has obtained new and most advanced results on 11 natural language processing tasks [9].

In October 2018, Google AI Research Institute proposed a BERT (Bidirectional Encoder Representation from Transformers) pre training model [10], which is different from the traditional emotion classification technology in the past and has achieved the most advanced results in many popular NLP tasks. Bert model not only introduces the two-way coding mechanism of lstm, but also uses the Transformer in GPT for feature extraction. It has a very strong ability to extract text features, and can learn the potential syntactic and semantic information in sentences. Bai Qingchun and others invented a position gated recurrent neural network to dynamically integrate sentence level global and local information to achieve attribute based text emotion classification [11]. Duan et al. proposed a Chinese short text classification algorithm based on Transformer Bi directional Encoder Representation (BERT) [12].The research team of Chengdu University of Information Engineering put forward a time-series multimodal emotion classification model based on multiple perspectives to extract the key emotional information in a specific time period and multiple perspectives in view of the poor multimodal fusion effect and the inability to fully tap the key emotional information in a specific time period and multiple perspectives [13].Suzhou University proposed a small sample emotion classification method based on the distillation of large and small tutor knowledge, which reduced the frequency of visiting the big tutor model, reduced the distillation time in the process of training student models, reduced resource consumption and improved the accuracy of classification and recognition [14].In order to improve the existing Chinese comment emotion classification method based on deep learning network and improve the accuracy and efficiency of Chinese comment emotion classification, Fan Anmin and others improved the traditional BERT model based on Tensorflow framework; On the Nlpcc2014 [15] data set, each index is 1.30%, 0.54%, 2.32% and 1.44% higher than the BERT model. Research shows that this model performs well in the classification and processing of Chinese comments' emotions, and is better than previous deep learning network models [16].

On this basis, in order to further deal with the "emotional phenomenon" of subjectivity, emotion, mood, mood, attitude and feeling in the text [17], Lv Xueqiang, Peng Chen and others proposed multi label text classification based on TLA-BERT model, which integrates BERT and label semantic attention MLTC method. Different from multi category text classification, multi label text classification can refine the text center from multiple label perspectives [18];Zheng Yangyu and Jiang Hongwei fully control the emotional information implied in the context by using local context and gated convolutional network model [19];Literature [20] proposed a multi-channel emotion classification method integrating feature vectors and attention mechanisms of part of speech and word location, which achieved high accuracy on the crawled microblog dataset. Literature [21] added attention mechanism to multi-channel CNN and BiGRU for experiment, and its classification effect is better than that of single channel network model. The word vectors mentioned in the above research are static word vectors, which cannot represent rich emotional semantic information.

This paper analyzes the user's Chinese emotion through the user's emotion classification technology based on BERT-RCNN-ATT model, and gets inspiration from the research on news text classification based on improved BERT-CNN model [22] and medical information classification based on BERT-ATT-BiLSTM model [23].The use of relevant technologies, as well as the combination of Transformer to research BERT model, complete the collection of data sets and other technologies, so as to classify users' Chinese emotions, is conducive to improving the existing Chinese comment emotion classification methods based on deep learning networks, and improve the accuracy and efficiency of Chinese comment emotion classification.BERT model absorbs the design idea of unsupervised models such as auto encoder and word2vec, and combines the characteristics of information such as unordered relationship and sentence to sentence relationship to be captured, and proposes a new unsupervised objective function for the converter.From this contribution, BERT model is well deserved to be called the first pre training language representation model to capture the bidirectional relationship of text.

2 Related Work

Among the methods for studying Chinese emotion analysis, there are currently three categories:

methods based on emotion dictionary [24], methods based on machine learning [25] and methods based on deep learning [26]. The method based on emotion dictionary needs to establish emotion dictionary, and the classification mainly depends on the quality and size of emotion dictionary. However, it is difficult to build a complete emotional dictionary, and updating the dictionary requires a lot of manpower and financial resources [27]. Machine learning based methods require a lot of manual annotation, use machine learning models for training, and use the trained classifier to analyze the emotional orientation of the text.Early processing of emotion classification tasks mainly relies on manual intervention to formulate rules. Word vectors are characterized by exclusive hot coding, but this way of representing results has high dimensions and large redundancy.In order to further improve the representation ability of words, the pre training models Word2Vec [28] and Glove [29] based on neural network are proposed. Through a large number of corpus training and learning, the text is mapped into low dimensional vectors to automatically extract features.FastText [30 - 31] adds n-gram features. Compared with the input of Word2Vec, FastText is the context information of the whole sentence.

This paper uses BERT-RCNN Att model to study Chinese emotion classification.BERT model makes use of Transformer's bidirectional coding pre training model, so that each word can make a bidirectional prediction of the whole semantics. It can fully extract the emotional information between texts, and integrates the attention mechanism, which has a better effect in emotional information classification. While ALBERT is a pre training language model improved based on BERT model. Compared with BERT, this model reduces the number of parameters and also improves the running speed [32]. The ALBERT model decomposes the input vector into a low dimensional matrix, which is transferred to the hidden layer through vector mapping. This decomposition method can significantly reduce the scale of parameters used in data conversion of input text information [33]. The model can also realize parameter sharing. In ALBERT, Transformer uses the method of parameter sharing between layers to increase the depth of the model and reduce the number of model parameters, which significantly improves the speed of model training and reduces the occupation of memory space.For BERT to learn the correlation between statements by predicting NSP, ALBERT proposed SOP (sense order prediction) to replace NSP, which has improved the accuracy and efficiency [34]. Hu Shengli[35] and others used the ALBERT-CNN model to analyze takeout comments. First, they used ALBERT to extract the global features of the text vector, and the same word can distinguish different meanings in different contexts. Then, they used CNN to extract the local feature information of the text. The experimental results show that the accuracy of the model reaches 91.3%, which proves the effectiveness of the model.Since the CNN model needs to set the length of context dependence through the size of the window, the RNN model cannot retain long-term memory, so RCNN (recursive convolutional neural network) model is introduced for emotion classification detection [36].RCNN model replaces the convolution layer of the traditional convolution neural network with a recursive cyclic convolution layer. It combines the advantages of CNN and RNN models, can uniformly use the context information of words, and has better performance. Li Yuechen et al. [37] compared the experimental data. When the original data is less, the BERT-RCNN model has stronger semantic feature extraction ability than the traditional model.

In text analysis, RCNN combined with Attention can be used to link the expression of each word learning with the word needed for prediction, so as to obtain information. Its main function is to pay attention to the most critical information in many information and mine deeper semantic features. Zeng Ziming et al. [38] proposed a model integrating two-level attention to improve the performance of emotion analysis. They use BiLSTM and two-level attention to extract sentence level features and feature weight distribution of each level, and finally obtain the emotional classification of text, which proves that the model has achieved good results. This paper uses BERT, RCNN model and attention mechanism to construct BERT-RCNN Att model for Chinese emotion classification. This model has better advantages than other models.

3 Methodology

The overall architecture of this paper is as follows Fig1.

3.1 Word Embedment

BERT pre training model consists of input layer, coding layer and output layer.Google has provided two models based on bert2, which are respectively the base model with 12 layers of transformers, 12 layers of Attention Headers, 768 hidden layer units and 110 million parameters, and the large model with 24 layers of transformers, 16 layers of Attention Headers, 1024 hidden layer units and 340 million parameters. [39]

During the pre-training of BERT model, there are two types of pre training tasks: Task 1: Masked language modeling; Task 2: Next sentence prediction [40], that is, predict the next statement. The embedded value of BERT model is composed of word vector, position vector and sentence feature vector, which can ensure the correct order of words in the text and obtain sentence level representation ability, so as to enrich the vector representation information and facilitate the follow-up task.

(1) Word vector: the input text is converted into a real value vector through the word vector matrix.Suppose that the unique heat vector corresponding to the input sequence x is expressed as: $e^t \in R^{N\times|m|}$, then the corresponding word vector represents V^t.

$$V^t = e^t W^t$$

Where, $W^t \in R^{|m|\times e}$ represents the word vector matrix that can be trained, and |m| represents the vocabulary size; e represents the dimension of the word vector.

(2) Block Vector: its code is the block number of the current word (starting from 0); The input sequence is a single block (single sentence text classification), and the block code of all words is 0; The input is two modules (sentence to text classification). Each word in the first sentence corresponds to a block code of 0, each word in the second sentence corresponds to 1 [CLS], and the [SEP] start and end corresponding codes are both 0. The trainable block vector matrix is used.$W^s \in R^{|s|\times e}$ (|s| represents the number of blocks; e represents the dimension of block vector). Convert the block code $e^s \in R^{N\times|s|}$ into a real value vector to obtain the block vector V^s.

$$V^S = e^s W_S$$

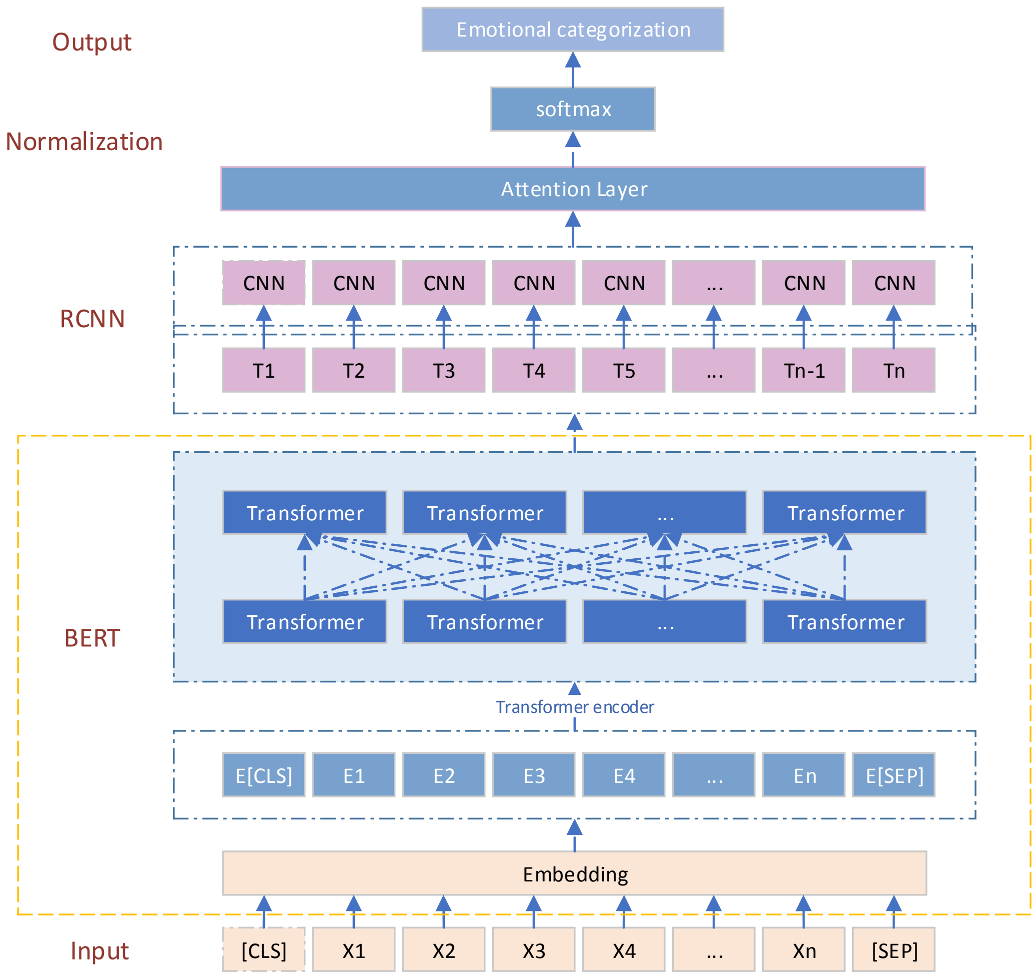

Fig. 1: Architecture of BERT-RCNN-ATT model

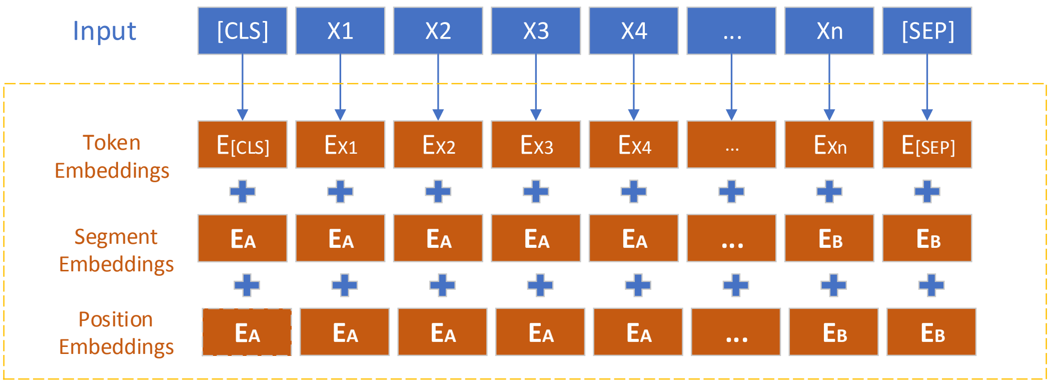

Fig. 2: Word embedding graph

(3) Position vector: The position vector is the absolute position of each word. Each word in the input sequence is converted into a position unique hot code according to its subscript order, and then the position vector matrix is used to convert the unique hot code into a real value vector to obtain the position vector V^P.

$$V^P = e^{PWP}$$

Where, $W^P \in R^{Nxe}$, N represents the maximum length, e represents the dimension of the location vector, e^P represents the unique hot code, and V^P represents the location vector.

3.2 Transformer Bidirectional Prediction

BERT only uses the Encoder part of Transformer, and its structure is shown in the figure.This part is connected by several Encoders of the Transformer model. Since each encoder has two residual connections, it ensures that the model effect will not deteriorate after each encoder. Finally, the semantic information of the sentence can be fully obtained under the action of multiple encoders, and then it is transmitted to downstream tasks for target task operation.Since the Self Attention mechanism does not have the ability to model the position information of the input sequence, and the position information reflects the logical structure of the sequence, which plays a vital role in the calculation, position coding is added to the input layer [41].

(1) Word vector and position coding: since the transformer model has no iterative operation of the cyclic neural network, the position information of each word must be provided to the transformer to identify the order relationship in the language. First, define the dimensions of inputs as [batch_size, sequence_length, embedding dimension], sequence_ Length refers to the length of a sentence or even the number of words contained in a sentence. Embedding division refers to the dimension of each word vector.

$$X = EmbeddingLookup(X) + PositionalEncoding$$

$$X \in R^{batch\_size*seq\_len*embed\_dim}$$

(2) Self-attention mechanism: first, we input sequence xi, where each xi can be considered as each word (word), and then we multiply xi by embedding by W to get the embedded input ai. For each ai, it has three matrices, namely query matrix (the matrix for querying other words), key matrix (the matrix for querying other words), and value matrix (the matrix for representing the extracted information value), Q, K, V can be obtained by multiplying ai with three matrices respectively. Finally, each Q pair and each key matrix can be multiplied by K point as an attention.

$$Q = X^i W_Q$$

$$K = X^i W_Q$$

$$V = X^i W_V$$

$$X_{attention} = Self\ Attention(Q, K, L)$$

(3) Residual connection and layer normalization: In the previous step, we obtained the V weighted by the attention matrix, that is, Attention (Q, K, V), and transposed it to make it consistent with the dimensions of Xembedding, that is, [batch_size, sequence_length, embedding dimension]. According to the consistency of the dimensions, we directly add the elements to make residual connection.

$$\boldsymbol{X_{attention}} = \boldsymbol{Self\ Attention(Q,K,L)} + \boldsymbol{X_{embedding}}$$

In the following operations, each module operation should add the value before and after the operation to get the residual connection. During training, the gradient can be directly backported to the initial layer by taking a shortcut:

$$\boldsymbol{X_{attention}} = \textbf{Layer Norm}(\boldsymbol{X_{attention}})$$

The output of BERT model includes character level vector and sentence level vector. This paper uses sentence level vector plus weight as semantic feature. Compared with traditional text representation methods, it reduces the steps of feature extraction and feature vector splicing, and has certain advantages.

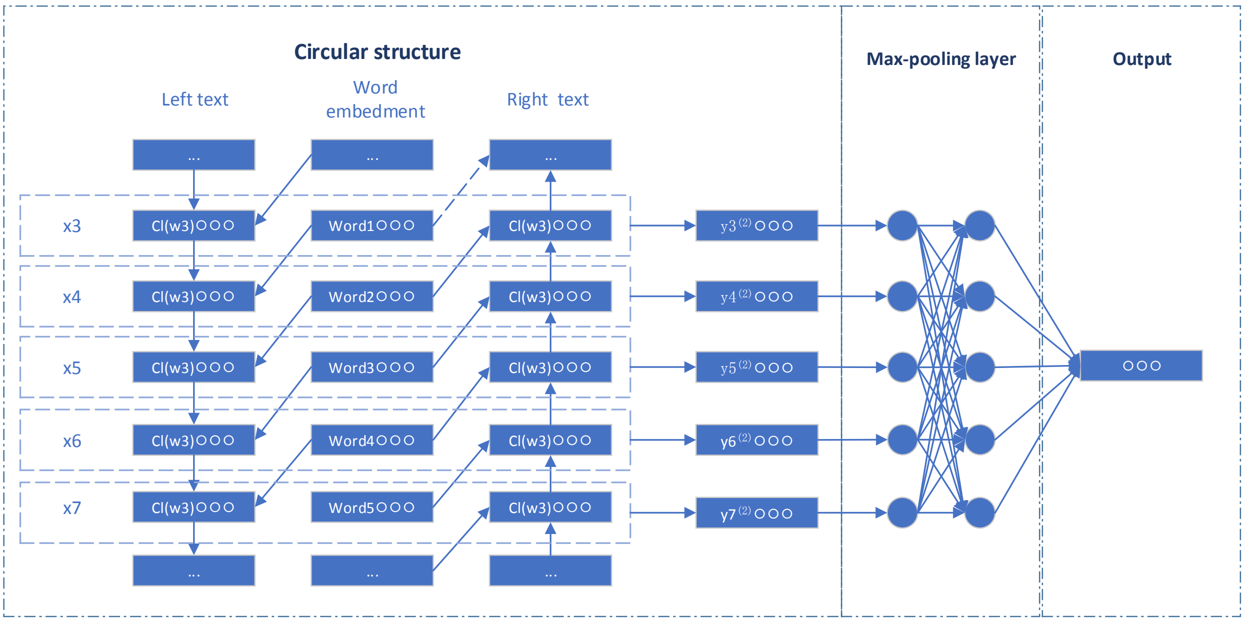

Fig. 3: RCNN structure diagram

3.3 RCNN Model

RCNN model includes six parts: input layer, convolution layer, splicing layer, pooling layer, full connection layer and output layer. In the convolution layer, multiple convolutions are used to detect the input matrix, and then the convolution operation is carried out. Finally, the features of the convolution core are obtained by detecting some of the features obtained. In the CNN model, the key link is the pooling layer. In order to extract the feature vector from the convolution, the pooling layer will extract the important feature vector from the feature vector. In this way, the pooling layer will produce a matrix with a fixed size. Finally, the data is transferred to the entire connection level for data processing and classification. The model structure is shown in the figure.

This paper uses RCNN depth to extract local features, gives weight to the sequence features (E1, E2... En) output from the last layer of BERT text as the word embedding of convolution operation, and extracts the local features of each entity in the feature sequence. The calculation process of convolution operation of convolution features is as follows:

(1) Convolution layer: if a sentence has n words and the word vector length of each word is m, the input matrix is nxm, which is similar to the "image" with channel 1. Then, after one-dimensional convolution of different sizes of convolution kernels (region_size) with the number of k, let the number of convolution kernels with different sizes (filters) be t, the width of convolution kernels and the dimension of word vector are the same as m, and the height h is a super parameter. A total of k * t feature maps [42] were obtained. By combining the information of the upper left and lower right words in the recursive structure, the upper left and embedded vectors of words are connected and the hidden semantic feature vectors of words are calculated.

The calculation formula is as follows, where ⊕ is concatenated by lines, f is a nonlinear activation function, and b is used to represent the partial term.

$$\mathbf{c_l} = f(\mathbf{W^{(1)}} \mathbf{c_l}(\mathbf{w_{i-1}}) + \mathbf{W^{(sl)}} E(\mathbf{w_i} - 1))$$

$$\mathbf{c_r} = f(\mathbf{W^{(1)}} \mathbf{c_r}(\mathbf{w_{i-1}}) + \mathbf{W^{(sr)}} E(\mathbf{w_i} + 1))$$

Where $c_l(w_i)$ represents the upper left word of a word, $c_r(w_i)$ represents the upper right word, $E(w_i)$ represents the embedding vector, and W(l) is the weight matrix. It is transferred from the upper left word $c_l(w_i)$ of the previous word to the upper left word $c_l(w_{i+1})$ of the next word, $W^{(sl)}$ indicating that it is transferred to the upper left word of the next word by combining the semantics of the current word $E(w_i)$.

$$\mathbf{X_i} = \mathbf{c_l}(\mathbf{w_i}) \oplus E(\mathbf{w_i}) \oplus \mathbf{c_r}(\mathbf{w_i})$$

$$\boldsymbol{Y_i^{(2)}} = f(\mathbf{W^{(2)}} \cdot \mathbf{X_i} + \mathbf{b^{(2)}})$$

Each row in the matrix represents the extraction result of T convolution kernels at the same position in the sentence matrix. Since the extraction results of T convolution kernels are collected, the row vector vi in S represents all the convolution features extracted for a certain position of the sentence.

(2) Pooling layer: The feature sizes obtained by convolution kernels of different sizes are different. Use pooling functions for each feature map to make their dimensions the same, and then splice them to obtain the final k * m dimension column vector. In this experiment, the maximum value of the convolved column vector is extracted by using the maximum pooling layer. After pooling, we will get a num_ The row vector of the filters dimension, that is, the maximum value of each convolution kernel is connected to eliminate the difference in the length of sentences.

$$\mathbf{Y^{(3)}} = \mathbf{max} \boldsymbol{Y_i^{(2)}} \ \mathbf{(i=1,2,3...)}$$

(4) Output layer: obtain the most representative key features in the text from the above max pooling layer, then output the full connection layer, and

finally obtain the classification results through the softmax function [43].

$$\mathbf{Y^{(4)}} = \mathbf{W^{(4)}Y^{(3)}} + \mathbf{b^{(4)}}$$

$$\boldsymbol{P} = \frac{\boldsymbol{EXP}(\boldsymbol{Y_j^{(4)}})}{\sum_{k=1}^{n} \boldsymbol{EXP}(\boldsymbol{Y_k^{(4)}})}$$

3.4 Attention

Attention mechanism can help the model give different weights to each part of the input Xi, extract more critical and important information, and make more accurate judgment on the machine polarity of emotion classification. At the same time, it will not bring more overhead to the calculation and storage of the model. This is why this paper uses Attention mechanism many times. The model mainly uses the attention mechanism to fuse the emotional feature vector h of the text and the directional feature vector b of the political entity, and calculate their attention scores a. Attention mechanism comes from that human beings will selectively focus on the key parts of all information while ignoring other visible information.The training set of training neural network is composed of Q, K and V. The constituent elements in the Source can be imagined to be composed of a series of<Key, Value>data pairs. At this time, an element Query in a given Target is calculated to obtain the weight coefficient of the corresponding Value of each Key by calculating the similarity or correlation between Query and each Key, and then the value is weighted and summed to obtain the final Attention value. Therefore, Essentially, the Attention mechanism is a weighted sum of the value values of the elements in the Source, while Query and Key are used to calculate the weight coefficients of the corresponding values. That is, its essence can be rewritten into the following formula:

$$\mathrm{Attention(Query, Source)} = \sum_{i=1}^{Lx} \boldsymbol{Similarity}(\boldsymbol{Query}, \boldsymbol{Key_i}) * \boldsymbol{Value_i}$$

In this paper, the output Ht after RCNN depth extraction of text context information is used as the input of the Attention layer. The model structure is shown in Figure 4. Assumed word vectors x1, x2,..., xn learn to derive the context vector gi when focusing on specific important words. When predicting sentence categories, the mechanism should focus on important words in the sentence, weighting and combining words with different weights:

$$\mathbf{g_i} = \sum_{j=i}^{i=1} \boldsymbol{\alpha_{i,j}} \cdot \boldsymbol{x_j}$$

Among $\alpha_{i,j}$ is called attention weight, requiring $\alpha \geq 0$ and $\alpha_{i,j} \cdot x_j = 1$, which is realized through softmax normalization. The formula describing the attention mechanism is as follows:

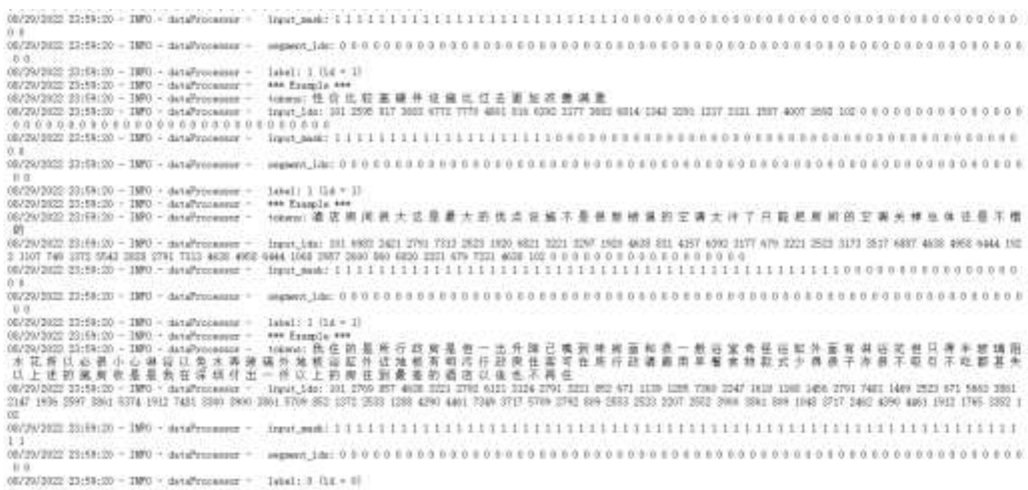

Fig. 4: Experiment dataset

$$\alpha_{i,j} = \frac{exp(score(x_i, x_j))}{\sum_j = exp(score(x_i, x_j))}$$

$$\text{Score}(x_i, y_i) = v_a^T tanh(W_a[x_i \oplus x_j])$$

The score value is calculated according to RCNN, which is used to simulate the correlation of words. The x with a larger score value has more weight in the context.

4 Experimental Evaluation

4.1 Dataset

This experiment extracts information texts about hotels, takeout, microblog and other user reviews, including more than 7000 hotel review data, more than 5000 positive reviews and more than 2000 negative reviews; There are more than 4000 positive and 8000 negative user reviews collected by a takeout platform; More than 100000 Sina Weibo texts with emotional annotation, about 50000 positive and negative comments each.

There are two columns of data, one is the review column, representing the input x of the model, and the other is the label column, representing the input y of the model. From the label perspective, there are two types of data, so the model is classified into two categories.

4.2 Evaluation Criteria

In order to evaluate the proposed method, this paper uses four indicators, namely, the accuracy rate Acc (accuracy), the accuracy rate P (precision), the recall rate R (recall) and the F1 value (f-score), to illustrate the efficiency of the experiment.Accuracy refers to the proportion of the number of positive and negative comment samples predicted by the model to the total number of samples; The accuracy rate refers to the proportion of correctly classified negative comments among all the samples predicted as negative comments. It is to show how many of the samples predicted as positive are real positive samples from the perspective of prediction results; Recall rate refers to the proportion of correctly classified negative comments among all samples that are true negative comments. It describes how many positive examples in the real sample are correctly predicted from the perspective of the original sample [44]; In order to evaluate different algorithms, the concept of F1 value is proposed based on Precision and Recall to evaluate Precision and Recall as a whole.

$$\textbf{Accuracy} = \frac{\textbf{TP} + \textbf{TN}}{\textbf{TP} + \textbf{TN} + \textbf{FP} + \textbf{FN}}$$

$$\textbf{Precision} = \frac{\textbf{TP}}{\textbf{TP} + \textbf{FP}}$$

$$\textbf{Recall} = \frac{\textbf{TP}}{\textbf{TP} + \textbf{FN}}$$

F1=2*(Precision*Recall)/(Precision+Recall)

Among them, TP means actually negative comments and identified by the model as negative comments, FP means actually positive comments but identified by the model as negative comments, FN means actually negative comments but identified by the model as positive comments, and TN means actually positive comments and identified by the model as positive comments.

4.3 Implementation Process

First, train CNN, LSTM and GRU models, read review column and label column respectively, and perform jieba segmentation on data, while removing stop words and punctuation; Train word2vec, build vocabulary and embedding matrix.B

The models to be compared and initialize the models. When training the model, input each data into the model and get the output. Finally, calculate the cross entropy with the label to get the final loss. Update the parameters through gradient back propagation.

Second, Bert model configuration environment:NVIDIA RTX A4000-24G， CPU : E5-2680 v4， CUDA v11.2， PyTorch v1.10

Use the existing and trained Bert model to fine tune. The fine tuning process is as follows:

1. Read data, read review column and label column respectively

2. Initialize the word breaker, and the bert (BERT: Pre training of Deep Bidirectional Transformers for Language Understanding) uses wordpiece word segmentation.

3. Segmenting all review columns and converting them into input_ Ids and attention_ mask ；

4. Create a dataset and input_ Ids and attention_ Masks are integrated together, and then a data read iterator is constructed through the dataloader.

5. Load the bert model, and then overlay the full connection classification layer.

6. Training model, input each data_ Ids and attention_ Mas inputs the model to get the output, calculates the cross entropy with label to get the final loss, and finally performs gradient back propagation and updates the parameters.

4.4 Result and Discussion

First, train on the training set, and then test on the test set. As shown in the loss graph of this model on the training set, it can be observed that the loss is decreasing and the model is gradually converging.

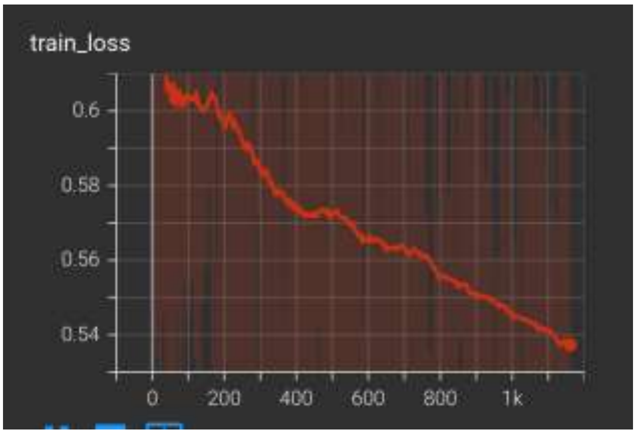

Fig. 5: Loss function

In order to make a more objective evaluation of the model in this paper, for three different types of data sets, under each data set, the model proposed in this paper is compared with the previous traditional models, and then the evaluation parameters of the model are analyzed.First, observe the loss chart. Due to a series of problems such as excessive complexity, excessive noise data interference in the sample, or inconsistent distribution of the characteristics of the training set and the test set, the model in the verification set may change with the change of the model, showing a trend of "first decreasing, then slightly increasing", which leads to the risk of over fitting, Therefore, we should jointly observe the dynamic changes of accuracy and loss value to judge

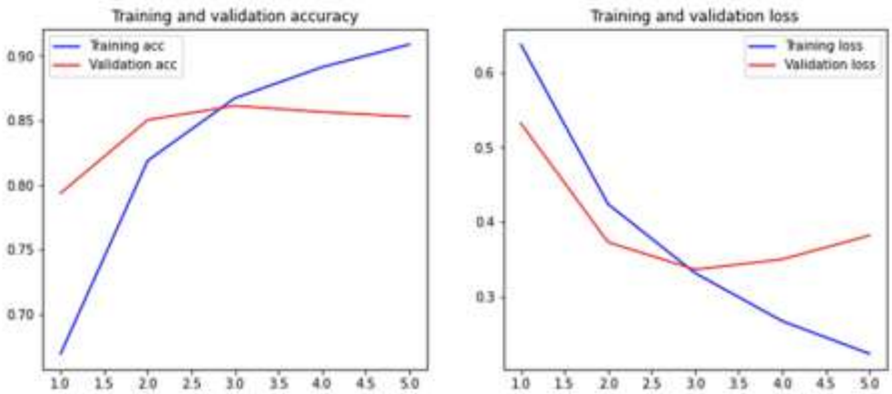

Fig. 6: Loss diagram of CNN model

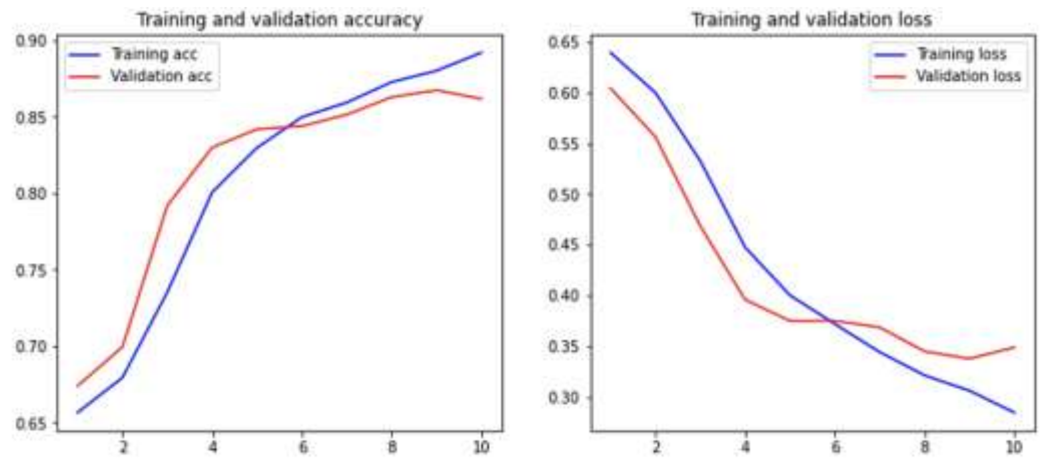

Fig. 7: Loss diagram of GRU

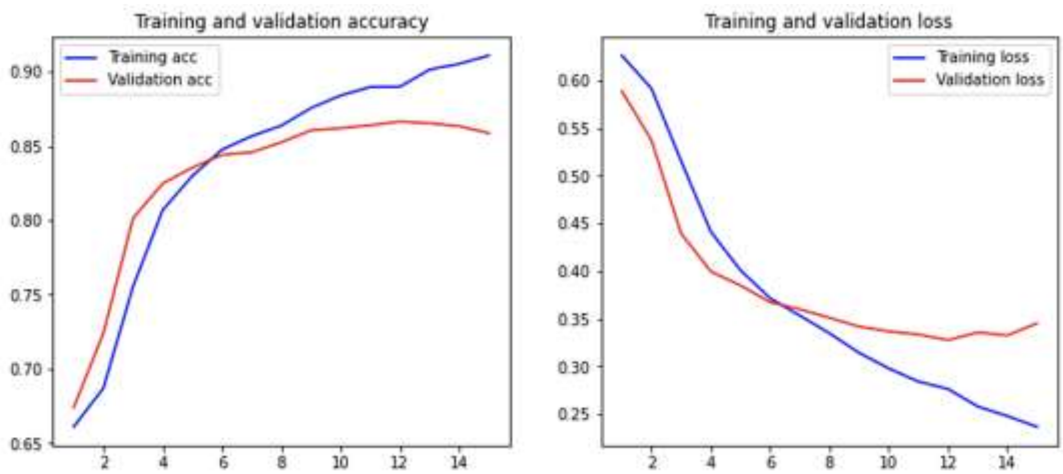

Fig. 8: Loss diagram of Lstm

Table 1. Experiment results on ChnSentiCorp_htl_all

Model	Accuracy	Precision	Recall	F1
BERT-RCNN-Att	0.93834	0.93799	0.93815	0.93807
CNN	0.79047	0.78705	0.71168	0.72960
GRU	0.86389	0.84379	0.84638	0.84506
LSTM	0.87334	0.86016	0.84607	0.85246

Table 2. Experiment results on waimai_10k

Model	Accuracy	Precision	Recall	F1
BERT-RCNN-Att	0.91271	0.90016	0.90255	0.90134
CNN	0.86117	0.84250	0.84082	0.84165
GRU	0.86788	0.85900	0.83407	0.84454
LSTM	0.86620	0.85338	0.83703	0.84429

Table 3. Experiment results on weibo_senti_100k

Model	Accuracy	Precision	Recall	F1
BERT-RCNN-Att	0.96191	0.96248	0.96188	0.96191
CNN	0.92161	0.92235	0.92168	0.92159
GRU	0.93460	0.93473	0.93463	0.93460
LSTM	0.92322	0.92341	0.92326	0.92322

It can be seen from the figure that as the network is more complex, the calculation amount of the lstm model is relatively larger, and its training convergence is slower, but on the whole, while the loss of the training set continues to decrease and the accuracy continues to improve, the loss function of the verification set is also decreasing, and the accuracy of the verification set is also rising.The training set is used to train the model, evaluate the training effect, and use the test set to evaluate the accuracy of the model. In order to prevent accidents, this experiment iterates the RCNN model 40 times to obtain various experimental results and evaluation values. Each iteration will produce multiple models, evaluate their efficiency respectively, and then use the test set to test their efficiency after optimization. The accuracy of the initial test set is about 85%, After multiple iterations, the model can be basically stable above 90%.

Common evaluation indicators include Accuracy, Precision, Recall and F1 score. For three different types of data sets on this test set, we get the following comparison indicators:

This paper analyzes the internal structure and principle of BERT and RCNN models, in which the recursive structure of the model and max pooling play a key role, retaining and deeply capturing a wide range of text information, which tests the model effect on text classification tasks.From the experimental results, it can be seen that the accuracy of the model introduced RCNN is about 7.6% higher than that of the CNN model, which shows that unlike CNN, which cannot store memory for a long time, RCNN model, as the combination of RNN and CNN models, achieves high accuracy in extracting context information, improves classification accuracy, and occupies a certain advantage in text classification.The attention mechanism can improve the ability of the model to focus on more important sequence information. The weight of each position relative to another position can be calculated in parallel, which is much faster than the lstm under the premise of sufficient computing resources, and further improves the model accuracy.Through the design of pre training tasks, the above persuasive data sets are used to fine tune the model on the basis of the trained model. The model effect is better. The accuracy rate and accuracy rate of the model itself are improved slightly. The accuracy rate of the model in this paper is about 5.5% higher than that of LSTM and GRU models, and it exceeds the existing methods in many Chinese text data sets of text classification.At the same time, it is found that compared with the traditional window based neural network, the noise of RCNN concept experiment is less, which shows that the model has strong universality.

5 Conclusion

In this paper, when solving user sentiment analysis, we use BERT model to classify Chinese emotions, extract information through transformer, and make multi-layer and two-way prediction on sentences to better understand the deep meaning of sentences.At the same time, the RCNN model combined with attention mechanism is used for in-depth extraction, which can effectively analyze the positive and negative emotions of users according to the user's comments and the tendency of public opinion. It is helpful for enterprises and the government to take timely measures through relevant analysis to dig out greater social value.In addition, the Chinese emotion analysis in this paper also involves the integration of artificial intelligence and computer science, which promotes the development of artificial intelligence. Emotional analysis is a research work with broad application prospects. I believe there will be more achievements in the near future. At the same time, this study also has shortcomings. In the follow-up study, we will expand the scope of data for in-depth research to provide better suggestions for Chinese emotion analysis.

Acknowledges:

This work was supported in part by the Innovation and entrepreneurship training program for college students under Grant No. 202210378351.

References:

[1] Started in November 1997, it is one of the most authoritative reports on Internet development data released by China Internet Network Information Center (CNNIC).

[2] Zhang Xiaoyan. Sentiment Analysis of Chinese online Comments based on weighted fusion word vector [J]. Application Research of Computers,2022,39(01):31-36.

[3] Shi Hao. Application, Challenge and Opportunity of natural language processing in computational communication research [J]. Transmission and copyright,2021(04):55-58.

[4] Chen Guowei, Zhang Pengzhou, Wang Ting, Ye Qiankun. A review of multi-modal sentiment analysis.Journal of Communication University of China (Natural Science Edition), 2022,29(02):70-78.

[5] Wang Suge. Research on Web-based sentiment classification of comment text [D]. Shanghai: Shanghai University ,2018.

[6] Wang Yingjie, Zhu Jiuqi, Wang Zumin, Bai Fengbo, Gong Jian. A survey on the application of natural language processing in text sentiment analysis [J]. Journal of Computer Applications, 2022,42(04):1011-1020.

[7] Pang B. Lee L. Vaithyanathan S. Thumbs u p?: sentiment classification using machinelear ning techniques[C]//Proceedings of the ACL-02 conference on Empirical methods in natura l language processing-Volume 10. Association for Computational Li nguistics, 2002:79-86.

[8] Zhao Xiaoming, Yang Yijiao, Zhang Shiqing.Research progress of multi-modal emotion recognition for deep learning [J]. Computer Science and Exploration, 2022,16(07):1479-1503.

[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, et al.BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,2019.

[10] DEVLIN J,CHANG M W,LEE K,et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]// Proceedings

of the 2019 Conference of the North-American Chapter of the Association for Computational Linguistics:Human Language Technologies.
Stroudsburg,PA:ACL,2019:4171-4186.

[11] Bai Qingchun, Xiao Jun, Wang Lamei.An attribute-level text sentiment classification method based on location-gated recurrent neural network [P]. Shanghai: CN114996454A,2022-09-02.

[12] Duan Dandan, Tang Jiashan, Wen Yong, Yuan Kehai. A new method for Chinese text classification based on BERT model [J].Computer Engineering. 2021,47(01)

[13] Tao Quanhui, An Junxiu, Dai Yurui, Chen Hongsong, Huang Ping.Research on temporal multi-model sentiment Classification based on multi-view Learning [J]. Computer Application Research.10.19734/j.issn.1001-3695.2022.06.0298.

[14] Li Shoushan, Chang Xiaoqin, Zhou Guodong. A small sample sentiment classification method based on knowledge distillation of large and small mentors [P]. Jiangsu Province: CN114722805A,2022-07-08.

[15] NLPCC, short for Natural Language Processing and Chinese Computing, is the first international conference in the field of NLP natural language processing in China, and also the first choice in the field of Chinese computing.

[16] Fan Anmin , Li Chunhui ,Improved BERT model for Chinese comment sentiment classification [J]. Software Guide, 2022,21(02):13-20.

[17] A New Perspective of Sentiment analysis Research, Caroline Brun(CSDN),2020.3.10, https://europe.naverlabs.com/blog/new-horizons-in-sentiment-analysis-research/

[18] Lv Xueqiang, Peng Chen, Zhang Le, Dong Zhian, You Xindong. A Text Multi-Label Classification Method Combining BERT and Label Semantic Attention [J]. Journal of Computer Applications, 2022,42(01):57-63.

[19] Zheng Yangyu, Jiang Hongwei.Aspect-level sentiment classification model based on local context and GCN [J]. Journal of Background Information Science and Technology University (Natural Science Edition),2022,37(01):76-81.

[20] Han Pu, Zhang Wei, Zhang Zhanpeng, et al. Sentiment Analysis of Public Health Emergency in Micro-Blog Based on Feature Fusion and Multi-Channel[J]. Data Analysis and Knowledge Discovery, 2021, 5(11): 68-79.

[21] Cheng Y, Yao L, Xiang G, et al. Text Sentiment Orientation Analysis Based on Multi-Channel CNN and Bidirectional GRU With Attention Mechanism[J]. IEEE Access, 2020, 8: 134964-134975.

[22]Zhang Xiaowei, Shao Jianfei.Research on news text classification based on improved BERT-CNN model [J]. Television Technology,2021,45(07),146-150.

[23] Yu Zhangxian, Hu Kongfa. A new model for the classification of medical information based on BERT-Att-BiLSTM model [J]. Computer Age,2020,(03),1-4.

[24] Wu Jiesheng, Lu Kui, Wang Shibing.Sentiment Analysis of Movie reviews based on multi-sentiment dictionary and SVM. Journal of Fuyang Teachers University (Natural Science Edition),2019,36(02):68-72.

[25] Cheng Zhengshuang, Wang Liang.Sentiment analysis method of online reviews based on support Vector machine. Electronic Technology and Software Engineering,2019,36(02):68-72.

[26] Cui Weijian. Text sentiment analysis based on deep learning [D]. Jilin University,2018.

[27] Xu Minlin. Research on text Sentiment Analysis based on Sentiment Dictionary and Neural Network [D].Jiangxi University of Science and Technology,2020.

[28] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[J]. arXiv preprint arXiv: 1301.3781, 2013.

[29] Pennington J, Socher R, Manning C. Glove: Global Vectors for Word Representation[C]. In: Conference on Empirical Methods. 2014.

[30] Bojanowski P , Grave E , Joulin A , et al. Enriching Word Vectors with Subword Information[J]. 2016.
[31] Choi J, Lee S W. Improving FastText with Inverse Document Frequency of Subwords[J]. Pattern Recognition Letters, 2020, 133: 165-172.
[32] Zhi Shiyao, Wu Zhenru, Chen Tao, Li Shengda, Peng Dong. Research on sentiment analysis of Micro-blog comments based on ALBERT-BiLSTM-Att[J]. Computer Age,2022(02):19-22.
[33] Cai Lei. Text sentiment analysis based on ALBERT-BIGRUATT[D]. Xinjiang University,2021.
[34] Gao Ying. Text sentiment analysis based on ALBERT-SABL model [D]. Shenyang Normal University,2022.
[35] Hu Shengli, Zhang Liping. Sentiment analysis of takeaway comments based on ALBERT-CNN [J]. Modern Information Technology, 2022,6(10):157-160.
[36] Wu Hao, Pan Shanliang.A new approach to Chinese comment recognition based on BERT-RCNN [J]. Journal of Information Science and Technology,2019,36(01):92-103.
[37] Li Yuechen, Qian Lingfei, Ma Jing.Early rumor detection based on BERT-RCNN model [J]. Information Theory & Practice,2021,44(07):173-177+151.
[38] Zeng Ziming, Wan Pinyu.A novel micro-blog sentiment analysis for public security events based on Bi-level attention and Bi-LSTM [J]. Information Science, 2019,37(06):23-29.
[39] Zhang Yanhua, Yang Shuo, Liu Chao. Training models of BERT based education equipment supply chain[J]. Journal of public opinion report system and innovation of science and technology, 2022 (16) : 48-51.DOI:10.15913/j.cnki.kjycx.2022.16.015.
[40] Sun Dandan, Zheng Ruikun.Application of BERT-DPCNN model in network Public opinion sentiment analysis [J]. Network Security Technology and Application,2022(08):24-27.
[41] Zhao Hong, Fu Zhaoyang, Zhao Fan.Micro-blog sentiment analysis based on BERT and hierarchical Attention [J]. Computer Engineering and Applications, 2022,58(05):156-162.
[42] Bai Jing, Li Fei, Ji Donghong.A new model for the detection of Chinese micro-blog position-based on BiLSTM-CNN[J]. Computer Applications and Software,2018,35(03):266-274.
[43] Wang Haochang, Sun Mingze.Chinese short text classification based on ERNIE-RCNN model [J]. Computer Technology and Development, 2022,32(06):28-33.
[44] Liu Siqin, Feng Xurui.Text sentiment classification based on BERT [J]. Information Security Research,2020,6(03):220-227.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This work was supported in part by the Innovation and entrepreneurship training program for college students under Grant No. 202210378351.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.