Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/38/20230522

Sentiment analysis of hotel comments based on LSTM and GRU

Zeyu Xu

College of Artificial Intelligence, Anhui University of Science and Technology, Huainan, 232001, China

2020307090@aust.edu.cn

Abstract. Sentiment analysis, which tries to examine the emotional information in the provided text data, has always been a popular topic in the community of natural language processing. Sentiment analysis is currently used in many different contexts, including e-commerce platforms, social media platforms, public opinion platforms, and chatbots. These applications are crucial to the advancement of society and the domestic economy. However, due to the personalization of text data, especially comments, and the presence of acronyms, it is a challenging problem to obtain accurate sentiment information from large and complex unstructured text data. This study presents a comparative examination of various text sentiment analysis approaches, including LSTM, CNN, and GRU. These methods are employed to evaluate their respective performance on sentiment analysis tasks, specifically using a dataset of hotel reviews for training the models. The method presented in this research has been extensively validated through numerous experimental results, affirming its efficacy and its potential to offer novel perspectives for the practical implementation of sentiment analysis.

Keywords: Sentiment Analysis, Hotel Comments, LSTM, GRU.

1. Introduction

The Internet has experienced significant growth and widespread adoption in China, leading to the emergence of several interactive applications on various online platforms. Users can not only browse information on these platforms but also express their wishes and emotions on these platforms [1]. Therefore, the data in the form of text, pictures, voice, and video is growing exponentially, especially the unstructured text data. Because text data contains a lot of meaningful information, especially users' comments with emotional tendencies on certain things (network hot events, stocks, commodities, movies, hotels, etc.). Mining these text data and analyzing the emotional information in the data can not only help relevant businesses adjust the corresponding service strategies and improve the corresponding service quality, but also improve the service experience of users [2]. At now, sentiment analysis is utilized in various domains, including e-commerce platforms, social platforms, public opinion platforms, and Chatbot systems. This field has set off an upsurge of research on emotion analysis technology by experts at home and abroad. Various sectors, including government organizations, are displaying growing interest in emotion analysis and individuals' emotional inclinations. This phenomenon holds significant implications for societal progress and the national economy [3].

DOI: 10.54254/2755-2721/38/20230522

Text sentiment analysis is a subfield within the domain of natural language processing, which is commonly referred to as opinion mining. The term pertains to the process of extracting and evaluating the user's subjective attitudes, ideas, and feelings expressed in the text by means of developing and executing a hotel recommendation system that relies on sentiment analysis. This enables precise and efficient categorization of the sentiment conveyed in the text. The development of mainstream research methods in text sentiment analysis task mainly includes three stages. During the initial stages of study, a majority of scholars engaged in the manual construction of an emotional lexicon. Subsequently, they proceeded to match emotional terms within the text under consideration against the aforementioned emotional lexicon. This process facilitated the computation of an emotional score, which in turn aided in determining the emotional category to which the text belonged [4]. However, these dictionary-based methods rely too much on a manually constructed sentiment dictionary, which is very time-consuming and will be affected by people's prior knowledge. Finally, these factors will lead to the inefficiency and inaccuracy of this method [5]. Subsequently, following the advent and progression of machine learning, scholars have employed machine learning techniques in the domain of natural language processing. The present methodology demonstrates superior accuracy and efficiency in classification outcomes compared to prior approaches. However, it still necessitates manual preprocessing of the text and labeling of the text features. Subsequently, as computer hardware has undergone significant advancements, the field of deep learning has transitioned from a theoretical concept to practical implementation. Researchers have extended the application of deep learning methods from image processing to natural language processing, capitalizing on the remarkable accomplishments achieved in the former discipline. Thus far, this particular methodology has emerged as a prominent study approach within this domain.

In recent years, there has been significant progress in the field of natural language processing models, particularly those utilizing deep neural networks, including large-scale language models. These models have demonstrated great potential for development and implementation, as evidenced by their exceptional performance in a wide range of tasks, such as text sentiment analysis, question answering systems, and machine translation [6]. Within the domain of text sentiment analysis, the deep learning approach diverges from conventional machine learning and sentiment dictionary methodologies [7]. This approach minimizes the need for extensive manual involvement in feature construction. It operates as an end-to-end method, wherein the model solely requires the input of text data. The model then autonomously discerns the semantic aspects of the text for sentiment classification, ultimately producing the classification outcomes. However, this approach necessitates substantial empirical evidence to substantiate its claims [8]. The proposed methodology for emotion analysis utilizing deep learning comprises the following fundamental steps: Firstly, the task at hand involves the linguistic modeling of a text corpus. Neural network models, such as word2vec, glove (Global vectors for word representation), Elmo (embeddings from language models), and Bert (bidirectional encoder representation from transformer), are widely employed language models. Next, the text representation that has been trained is fed into the neural network model in order to acquire knowledge about the features. This is typically achieved through the utilization of various neural network architectures such as cyclic neural networks, convolutional neural networks, long-term and short-term memory neural networks, among others. Ultimately, the acquired feature vectors are utilized as input for the classifier in order to obtain the outcomes of emotion categorization.

In the year 2013, Socher et al. introduced a recursive tensor model for semantic combination. This model effectively captures the impact of negative words in tasks related to sentence classification. In 2014, Kim et al. applied CNN to the task of emotion analysis, input the pre trained word vector into CNN, get more abstract semantic information after multiple convolution and pooling, and finally classify the text. However, CNN cannot capture long-distance features, and the maximum pool layer loses word location information. In 2015, Tang et al. proposed to use LSTM to model the emotional relationship between sentences, which can obtain remote features and alleviate the problems of gradient disappearance and gradient explosion in the cyclic neural network [9-10]. Tai et al. proposed a tree LSTM model, which extended LSTM to tree structure network topology, and achieved good

DOI: 10.54254/2755-2721/38/20230522

results in sentiment analysis and semantic correlation tasks [11]. In addition, traditional deep learning methods, such as CNN and RNN. CNN uses the maximum pool layer to find the important salient features of words. This maximum pool method can improve the effect of text prediction and classification tasks. However, the maximum pool layer is only the salient feature corresponding to the maximum activation value, and it has no clear method to deal with the salient feature. Hence, to address this issue in a more significant manner, Bahdanau et al. introduced the notion of attention mechanism in the field of deep learning and applied it to machine translation. The attention mechanism is employed to prioritize significant characteristics by incorporating softmax into the model. This softmax function is typically utilized as the output system to determine the probabilities of various classes and categories.

However, due to the different expression habits of different users, the unstructured text data generated by them on the Internet platform is not standardized and personalized. The language expression of short text is short and casual, with sparse semantics and few features. In addition, there are always some “network nouns” and abbreviations on the Internet. The characteristics of text due to users’ expression habits will bring challenges to the task of emotion analysis. Hence, the issue in the process of emotional analysis lies in reliably and effectively identifying the emotional tendency of massive and complicated unstructured text data, and obtaining precise semantic information from it. This paper aims to examine the performance of various methods for text sentiment analysis by selecting representative LSTM, GRU, and CNN models. The sentiment of hotel reviews will be analyzed in order to offer novel insights for the practical implementation of sentiment analysis.

2. Relevant theoretical basis

2.1. LSTM

A fundamental concept behind recurrent neural networks is the establishment of connections between prior information and the ongoing activity, hence enabling the network to possess a memory function for past information. Because of the special nature of text context correlation, the use of recurrent neural networks in text tasks can get better results. As shown in Figure 1, the Long Short-Term Memory (LSTM) model is a distinct variant of the recurrent neural network (RNN) architecture, initially introduced by Hochreiter et al. in 1997. LSTM effectively addresses the issues of gradient vanishing and gradient exploding that are commonly seen in normal RNNs, resulting in significantly improved performance compared to conventional recurrent neural networks. One notable aspect of Long Short-Term Memory (LSTM) is that modifying the input threshold, forgetting threshold, and output threshold alters the magnitude of self-circulation weight. When the parameters of the model are held constant, the integration scale can be adjusted dynamically at various time points, thereby circumventing the issue of gradient vanishing or gradient expanding [12].

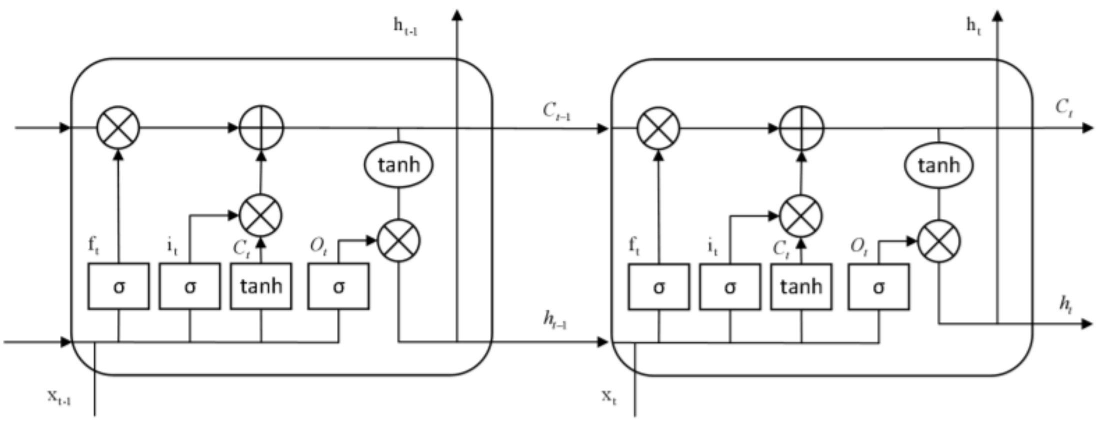


Figure 1. LSTM structure diagram.

2.2. GRU network

The extensive utilization of Long Short-Term Memory (LSTM) in the field of natural language processing, particularly in the context of text classification endeavors, has led researchers to increasingly recognize the drawbacks associated with LSTM. These limitations include prolonged training duration, a high number of parameters, and intricate internal computations. In 2014, Cho et al. introduced a modified Gated Recurrent Unit (GRU) model that incorporates the unit state and hidden layer state of Long Short-Term Memory (LSTM) while also implementing other modifications. The GRU model is a variant of the LSTM model that preserves the long short-term memory effect while offering a simpler architecture, reduced parameter count, and improved convergence properties.

As depicted in Figure 2, the GRU model is comprised of two distinct components, namely the update door and the reset door. The control over the impact of the prior time's output hidden layer on the present hidden layer is governed by the update gate. The influence of the output hidden layer from the previous time on the current hidden layer is directly proportional to the value of the update gate. The extent to which the information from the previous time's hidden layer is disregarded is determined by the reset gate. The significance of the reset gate is often overlooked in proportion to its apparentness [13].

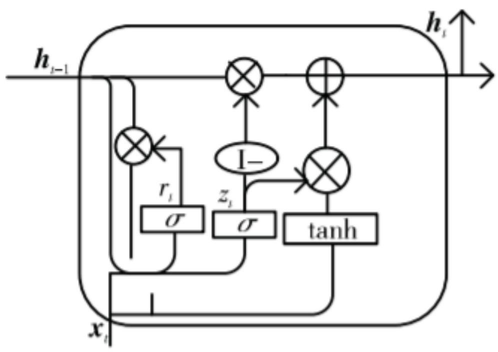


Figure 2. GRU structure diagram.

2.3. CNN network

CNN, or Convolutional Neural Network, is a type of feedforward neural network that has exceptional proficiency in the realm of large-scale pattern recognition. The architecture consists of five main layers: the input layer, convolution layer, pooling layer, fully connected layer, and output layer. It mainly processes the local features of the input through the convolution layer and pooling layer, extracts the important feature information, and makes it also has good performance in natural language processing. Thus, CNN can judge the sentiment (positive, negative, and neutral) of the obtained text comment data [14].

The technique of dropout is frequently employed in the training of Convolutional Neural Networks (CNNs). It involves selectively deactivating one or more individual neurons during the forward propagation process. This technique has been found to enhance the overall performance of the model by improving its ability to generalize and reducing its reliance on local features. Dropout first randomly deletes some hidden neurons in the network, and then modifies the network through forward and reverse transmission of the network. On this basis, the random gradient descent method is used to update the corresponding parameters (W, b) on the neuron without deletion, and this process will be recursive. Early stopping is a technique that leverages the convergence property of the model throughout the training process. When the model reaches a point of excessive simulation, its performance on the proof set ceases to improve and instead begins to deteriorate. Consequently, it is advisable to terminate the training process at this juncture in order to prevent overfitting.

3. Model construction

3.1. Original data and preprocessing

Based on the sentiment analysis of the Chinese text of the data set of hotel reviews by Tansongbo, the data set is labeled with POS and neg, with 2000 TXT texts respectively. The whole preprocessing of original text mainly includes word segmentation, short sentence completion, long sentence cutting, and indexing. Word segmentation is the basis of Chinese natural language processing. Commonly used open-source Chinese word segmentation tool libraries include Jieba and thulac. The result of word segmentation is to segment a comment into words one by one. In sentiment analysis, sometimes short sentences or long sentences are encountered, which need special treatment. The comment length histogram of our original dataset is shown in Figure 3. The short sentence can be completed by adding appropriate context before and after the sentence, while long sentences can be truncated by taking only part of them for emotional analysis. Some standard indexing schemes, such as word bag model, TF-IDF, etc., can be used to convert the text divided into words into numbers for processing. Among them, the bag of words model regards each word as an independent feature, and TF-IDF considers the feature weight of each word.

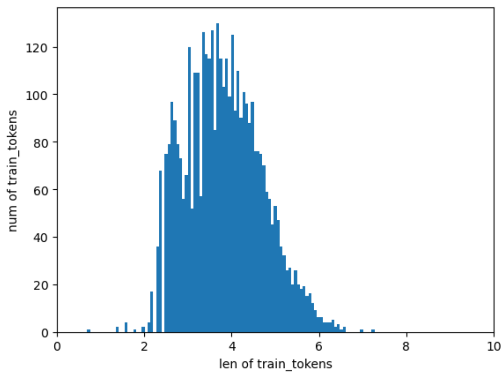


Figure 3. Comment length histogram.

3.2. Construct word vector

In this work, a pre-trained Chinese word vector model is adopted, which was trained using Baidu Baike, Wikipedia, Sogou news, tieba and other data training, including about 1.3 million Chinese words and 300 dimensional word vectors. Additionally, Keyedvectors is a class used to load pre-trained word vectors in gensim library, which can read and calculate the pre-trained word vectors. First, two variables num_Words and embedding_Dim were defined, representing the size of the vocabulary and vector dimension of each word, respectively. Then, a zero matrix embedding with the shape of (num_words, embedding_dim) is used to store the vector representation of each word. Next, we use a for loop to traverse the entire vocabulary and fill the vector representation of each word into embedding_ Matrix. Specifically, we use cn_ Model.index_ To_ Key[i] get the index of the current word in the model, then use the index to get the corresponding vector representation from the model, and assign it to embedding_ Matrix [i,:]. Finally, we will embed_ The data type of matrix is converted to 'float32' for subsequent processing in keras.

3.3. Model construction

The whole pipeline of model construction consists of the following steps: (1) Create a sequential model object, which means that our neural networks are stacked in the order of layers. (2) Add the embedding layer. The embedding layer is used to convert discrete text data into dense word embedded representation, whose parameters includes the num_Word, embedding_Dim, weights, input_Length, and trainable. num_Words is the size of the vocabulary, indicating how many different words can be used to encode the text. embedding_Dim is the dimension of word embedding, indicating the length of

DOI: 10.54254/2755-2721/38/20230522

the word vector. weights is used to set the word embedding matrix for pre training. Here, we embed the pre trained words as the initial weight. input_Length is the length of the input sequence, which is used to limit the shape of the input data. trainable indicates whether to train the word embedding layer. Here, set to false to indicate the weight of fixed word embedding. (3) Add a bidirectional LSTM layer. Bidirectional is a wrapper that encapsulates the LSTM layer as bidirectional. (4) Add GRU layer. Gru layer is a gated loop unit for further processing sequence data. (5) Add a fully connected output layer. This layer connects the output of the neural network to a single neuron for binary classification tasks. The activation function uses the sigmoid function to limit the output to [0, 1]. (6) Compile model. Compiling the model is to configure the loss function, optimizer and evaluation index.

4. Experimental results and analysis

4.1. experimental environment

We train the model and conduct all the experiments on Windows operate system, and all the configurations of our equipment are shown in Table 1.

Table 1. Configuration of experimental equipment.

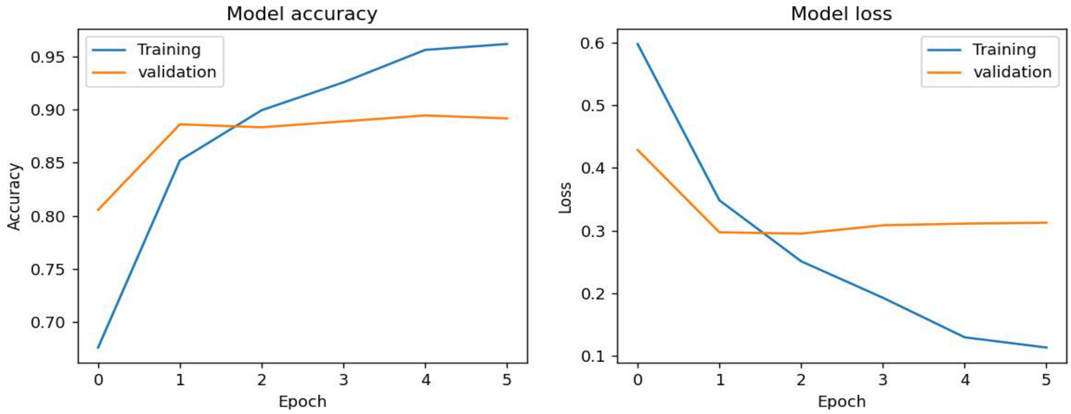
Software/hardware	allocation
operating system	Windows 11
Running memory	6g
development language	Python 3.7
Development framework	Tensorflow 2.11.0

4.2. Experimental settings

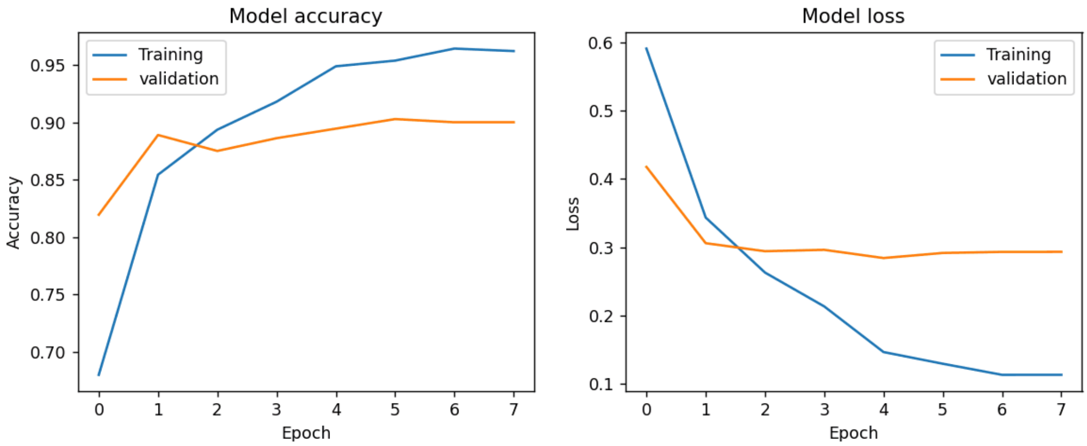
The learning rate used to control the speed of updating model parameters is set to 1e-2. For the test_size indicating the a part of the data set used as test data is set to 0.1. Similar to test_size, the parameter of validation_ split is the split ratio of validation set that controls the proportion of validation sets in the training set, which is set to 0.1. The model is trianed through the Optimizer of Nadam for 5 epochs, with a batch size of 128, a dropout rate of 0.2. The loss used to supervised the training process is binary cross entropy loss.

4.3. Results analysis

To analyze the model performance, we first report the accuracy and training loss of different models in Figure 4. For the first model trained by combining the BiLSTM and GRU, the training accuracy can increase up to 95% with the process of training stage, while the validation result can also achieve a accuracy of 88%. The model trained by combining CNN, BiLSTM and GRU can finally a best training accuracy of 96% and a best validation accuracy of 89%. Compared with the first model, it can be observed that introducing the CNN can benifit the performance of sentiment analysis, which demonstrates that CNN can helps to improve the feature quality as well as the final accuracy. What's more, we also observe that the loss of model based on CNN, BiLSTM and GRU show a better stability. All the results vary the effectiveness of our proposed methods.



(a) Accuracy and loss of BiLSTM+GRU model.



(b) Accuracy and loss of CNN+BiLSTM+GRU model.

Figure 4. The comparison of different models in accuracy and loss.

As shown in Table 2, we also discuss the model performance of proposed models for different scenes on the test set. For the first model, it can achieve the accuracy of 88.75%, 89.99% and 87.00% for three test scenes, respectively. When introducing the CNN, the accuracy can increase to 90.50%, 90.00%, and 90.00%. We also give some demo in Figure 5. All the results show the effectiveness of our method.

Table 2. Accuracy of the model on the test set.

Model type	Experiment 1	Experiment 2	Experiment 3
Bilstm+Gru model	88.75%	89.99%	87.00%
CNN+bilstm+Gru model	90.50%	90.00%	90.00%

```
请输入需要预测的中文评论文本（输入 stop 退出）：酒店真好
1/1 [==============================] - 1s 663ms/step
[[0.6862695]]
情感极性为：正面
请输入需要预测的中文评论文本（输入 stop 退出）：不好
1/1 [==============================] - 0s 24ms/step
[[0.1526114]]
情感极性为：负面
请输入需要预测的中文评论文本（输入 stop 退出）：天气好
1/1 [==============================] - 0s 24ms/step
[[0.7408332]]
情感极性为：正面
请输入需要预测的中文评论文本（输入 stop 退出）：服务差
1/1 [==============================] - 0s 23ms/step
[[0.18679054]]
情感极性为：负面
```

Figure 5. User input module test results 1.

5. Conclusion

This research introduces two deep learning models and evaluates their performance using a hotel review dataset. In summary, the inclusion of a CNN layer in the CNN+bilstm+Gru model has resulted in a modest improvement of approximately 1% in the accuracy of emotion categorization when compared to the bilstm+Gru model. The utilization of multi-model joint training is observed to effectively amalgamate the strengths of diverse models, hence enhancing the overall generalization performance of the model. We should improve these models by mixing these models with other machine learning algorithms to obtain better performance. According to the basic principle of this research, we can further expect to use the same technology in various deep learning applications, and study how this method of combining various models works in a small amount of data, and combine appropriate deep learning strategies to obtain more effective results. In addition, I am considering combining three or more different algorithms to take advantage of each method. In addition, in order to make these models robust, I plan to implement the model in other relevant data sets. To sum up, the model can be further explored through a variety of experiments, and make it more accurate and faster to obtain the results of emotion classification. It can also enable businesses to obtain real customer feedback in real time through these comments, understand the problems existing in the hotel, and formulate reasonable operation strategies to achieve greater economic benefits.

References

[1] Su Yan. Research on the reform of information technology teaching in Higher Vocational Colleges integrated with craftsman Spirit -- Taking the course of “Fundamentals of marketing” as an example [j]. Xiamen science and technology, 2022 (03): 49-52

[2] Wang Tao Research on text sentiment analysis based on deep learning and CTM model [d]. central China Normal University, 2020.doi:10.27159/d.cnki.ghzsu.2020.001319

[3] Jiwang Sentiment analysis of e-commerce commodity reviews [d]. Jiangsu University of science and technology, 2021.doi:10.27171/d.cnki.ghdcc.2021.000499

[4] Fengzhiji Research on new energy vehicle market based on text mining [d]. Hebei University of economics and trade, 2020.doi:10.27106/d.cnki.ghbju.2020.000551

[5] Zhou Kai Research on text sentiment analysis algorithm based on deep learning [d]. Guizhou University, 2019

[6] GUI Tao, xizhiheng, Zheng Rui, et al. Research review on robustness of naturallanguageprocessing based on deep learning [j/ol]. Acta computerica Sinica: 1-26[2023-08-29] http://kns.cnki.net/kcms/detail/11.1826.tp.20230727.0855.002.html.

[7] Chenzhengyu Research on text emotion analysis based on deep learning [d]. Chengdu University of information engineering, 2019.doi:10.27716/d.cnki.gcdxx.2019.000204
[8] Chengkangxin Emotional feature extraction algorithm of Chinese restaurant reviews based on LSTM and CNN [d]. Beijing University of Posts and telecommunications, 2020.doi:10.26969/d.cnki.gbydu.2020.002649
[9] Wangting, yangwenzhong. A review of text sentiment analysis methods [j]. computer engineering and applications, 2021,57 (12): 11-24
[10] Chen Qian, Che Miaomiao, Guo Xin, et al. A text sentiment classification method based on cyclic convolutional attention model [j]. computer science, 2021,48 (02): 245-249
[11] Chen Tao Emotional analysis of Chinese microblog based on deep learning [d]. Chongqing University of Posts and telecommunications, 2018.doi:10.27675/d.cnki.gcydx.2018.000874
[12] Huronglei, ruilu, Qi Xiao, et al. Text sentiment analysis based on recurrent neural network and attention model [j]. computer application research, 2019,36 (11): 3282-3285.doi:10.19734/j.issn.1001-3695.2018.05.0300
[13] Wang Wei, sunyuxia, qiqingjie, et al. Text emotion classification model based on bigru attention neural network [j]. computer application research, 2019,36 (12): 3558-3564.doi:10.19734/j.issn.1001-3695.2018.07.0413
[14] Qin Yao Research on the recommendation method of traditional Chinese medicine prescriptions based on machine learning [d]. Beijing University of technology, 2020.doi:10.26935/d.cnki.gbjgu.2020.000608