

Research on Sentiment Analysis of Government Short Video Comments based on BERT's Multi Strategy Combination Model

Yucong Zhang

School of Economics and Management Hebei University of Technology, Tianjin, China

Abstract: In the field of NLP, pre trained models greatly shorten development time, reduce usage difficulty, and improve model robustness. In this paper, the BERT pre training model is used to analyze the text emotion of Tiktok short video comments. Through experimental comparison, it is found that the BERT pre training model is better than CNN and BiLSTM in emotion classification. Based on the BERT model, the multi strategy hybrid method is used to further input the BERT output dynamic word vector into the CNN and BiLSTM models to further extract local features and features with long distance dependency, improve the accuracy of the model, and obtain a good classification effect on the government short video comment dataset, which is better than the single model.

Keywords: BERT; Short Video Comment; Sentiment Analysis.

1. Introduction

With the development of the Internet and mobile apps, various social media platforms have gradually flourished, with an increasing number of online users. Users use various communication media to comment on hot events and express their feelings, resulting in a large amount of emotional text information. These emotional tendencies are of great use to relevant government departments or businesses, but the forms of comments in these massive texts are complex and diverse, relying solely on manual processing has problems such as heavy workload and poor real-time performance. At present, most of the better performing machine learning methods rely on annotating a large amount of original text data and then processing test data. However, these methods are more difficult to conduct emotional analysis research on scene transfer or cross platform and cross domain public opinion events. Therefore, how to efficiently detect scene migration and cross platform emotional tendencies and trends will be the focus of future research, and it can reduce the effort spent on data annotation, making the model more robust and accurate. The text pre training model in the field of natural language processing has strong cross scene migration ability and high robustness, so it can be applied to various scenes requiring emotional orientation. This paper applies the BERT pre training model to the field of Tiktok government short video reviews, and explores its effect on emotion classification on the Tiktok review dataset.

2. Related Research

The measurement of social media emotions has been studied by scholars based on experiments [1] and scales [2]. These methods have certain limitations, are not suitable for large sample analysis, and have strong subjectivity. However, social media users generate a large amount of text content, and through text sentiment analysis, the public's emotional involvement can be obtained. The analysis of emotions is mainly divided into emotional tendency analysis [3], emotional polarity classification [4], and emotional intensity calculation [5]. The methods used mainly include sentiment

analysis based on emotion dictionaries, machine learning, deep learning, and scholars using existing software such as Baidu API for sentiment analysis [6]. The trend of improving the effectiveness of sentiment analysis is based on the combination of multiple strategies.

The sentiment analysis method based on sentiment dictionary mainly uses text segmentation to divide the document into several words, obtains the emotional polarity and intensity of the words through sentiment dictionary, and finally calculates the overall emotional tendency of the document by weighting. Wei Lai and Wang Weijie [5] divided emotions into seven categories based on the Dalian University of Technology Dictionary: joy, good, evil, shock, fear, anger, and sadness, and calculated positive and negative emotional values based on the emotional intensity and extreme emotions of each word. Qiu Jiangnan and Ge Yidi [7] calculated the emotional valence and emotional arousal intensity of Weibo information based on Dalian University of Technology Dictionary. Guo Xiuyuan et al. [8] used the SnowNLP sentiment analysis model to measure the emotional involvement of user comments in regional promotional videos. This method is easy to understand, but on the one hand, it relies on the construction of an emotional dictionary, and many new network terms, idioms, allegorical sayings, and other new words need to be supplemented to better recognize. On the other hand, the semantic relationship between contexts cannot be considered, and the emotional tendencies expressed by the same emotional word in different contexts are also different, so the effectiveness of cross domain emotional analysis is not good.

Machine learning based sentiment analysis relies on training models with annotated sentiment datasets, and then predicting the emotional tendencies of new sentences through the model. The core of this method is to construct a dataset, annotate the dataset, extract features based on the dataset, model, and then perform sentiment classification or prediction. Machine learning based learning algorithms include SVM [9], KNN [10], NB [11], etc. Wan Yanping et al. [12] achieved good classification performance in online comment datasets by improving the Stacking algorithm that

integrates multiple traditional machine algorithms. Machine learning technology can have stronger scalability and repeatability, improving classification accuracy. However, the classification accuracy of machine learning depends on high-quality annotated training sets. The annotation of training sets requires high labor costs on the one hand, and on the other hand, manual annotation has a certain degree of subjectivity.

Deep learning belongs to a subset of machine learning. Deep learning has the ability of active learning and automatic feature extraction, which solves the problem that traditional machine learning cannot make full use of the context of the context text when conducting emotion analysis. Deep learning can use contextual information to infer the meaning of the next word, eliminating the need for artificially labeled corpus. Through deep learning for text analysis, it can be found that its advantage lies in the ability to obtain rich information from large-scale corpora and use pre trained models for sentiment analysis.

Deep learning sentiment analysis is mostly based on CNN [13], RNN [14], and LSTM [15] models. CNN extracts local features through different convolutional kernels, down sampling and dimensionality reduction through pooling layers, and outputting abstract features through full connectivity, achieving good classification results. However, it only considers the impact of the previous input and cannot preserve the previous information. RNN considers temporal features and preserves inputs from other times. As a special structure of RNN, LSTM can extract features with long-distance dependencies, solve the problems of gradient vanishing and explosion, improve training speed, and save costs and time. BiLSTM solves the problem where LSTM only considers the preceding text and cannot consider the following text. It extracts features from both positive and negative directions.

After fully considering the advantages of different neural networks, some scholars have combined methods to complement each other's strengths. Rehman et al. [16] conducted emotional analysis on film reviews based on LSTM and CNN models. The CNN model effectively extracted features of higher dimensions through the convolution layer and pooling layer. The LSTM model can capture long-term dependencies between word sequences, which is better than the single model. Du Yongping et al. [4] used the CNN-LSTM model for short text sentiment analysis and achieved good results on three Chinese and English datasets. Some scholars have introduced attention mechanisms to improve the efficiency of models, focusing on key information in massive amounts of information. Wei et al. [17] proposed a combination of BiLSTM network and multipolar orthogonal attention mechanism to assign higher weights to important word vectors, achieving implicit sentiment analysis of text.

Pre trained models refer to models that have already been trained using a dataset. If these trained models are retained, they can be directly used in similar situations by simply adjusting the ground parameters, which will save a lot of time and cost. So, the pre trained model gained the attention of scholars. As an application of transfer learning, the pre training model can transfer the knowledge learned from the open domain to the downstream tasks to improve the execution effect of low resource tasks. It has achieved the best results in almost all NLP tasks. Among them, the BERT pre training model has been widely used and achieved good classification results. Duan Dandan et al. [18] used the BERT

pre training model to classify Chinese short text news on Sohu and found that its F1 value was as high as 93%, 6% higher than TextCNN. Li Tiefei et al. [19] proposed the BERT-TECNN model to further optimize the feature extraction ability of the BERT model. The output of BERT was further input into CNN models with different scale convolution kernels, and good classification results were achieved in three different datasets. Chen Zhiqun and Ju Ting [20] used the BERT-BiLSTM model to input the dynamic vector features of BERT generated text into a bidirectional LSTM network for sentiment analysis of Weibo comments. The F1 value reached 91.45%.

In summary, machine learning is more accurate than emotion dictionaries and no longer relies on the construction of artificial emotion dictionaries. It can use databases to update vocabulary in a timely manner, but its classification effect depends on the quality of manually annotated datasets. While deep learning has the ability of active learning, which can fully consider the semantics of the context, preserve the sequence of sentences, effectively identify polysemy, and use the data features extracted by multi-layer neural network, so the model has better effect. By pre training the model, it is possible to retain the already trained model, greatly reducing development time and usage difficulty, and performing better in the NLP field. However, it requires a large amount of data support to achieve optimal results.

3. Experimental Process and Model Construction

The emotional intensity of video comments reflects the emotional involvement of users in watching Tiktok short videos. Tiktok short video comments have the function of likes ranking, and the comments with high likes rank first in the video comment area. The public expresses their cognition and emotions through comments, and popular videos have a large number of comments. Based on the usefulness of the comments, this article selects comments with a likes volume greater than zero for sentiment analysis. In the past, scholars generally used questionnaire surveys to measure emotional engagement, which had a certain degree of subjectivity. Some scholars measured emotional engagement based on emotional dictionaries, for example, Guo Xiuyuan et al. [8] used emotional dictionaries to measure the emotional engagement of user comments in regional promotional videos. However, sentiment dictionaries cannot take into account the contextual semantic relationships of sentences and the polysemy of a word. In order to more accurately measure the emotional involvement of the audience after watching government short videos, this article uses a large-scale training-based BERT training model to conduct text sentiment analysis on video comments in annotated government short video datasets. Divide the emotional tendencies of video comments into positive and negative categories, and use the BERT pre training model to measure the probability values of positive and negative emotions in the text as the positive and negative emotional involvement of the comment. The result of the positive emotional involvement of the video is between [0,1], and the closer it is to 1, the more positive the sentiment of the comment text is. Conversely, the sentiment of the comment text is more negative. the experimental process is shown in Figure 1

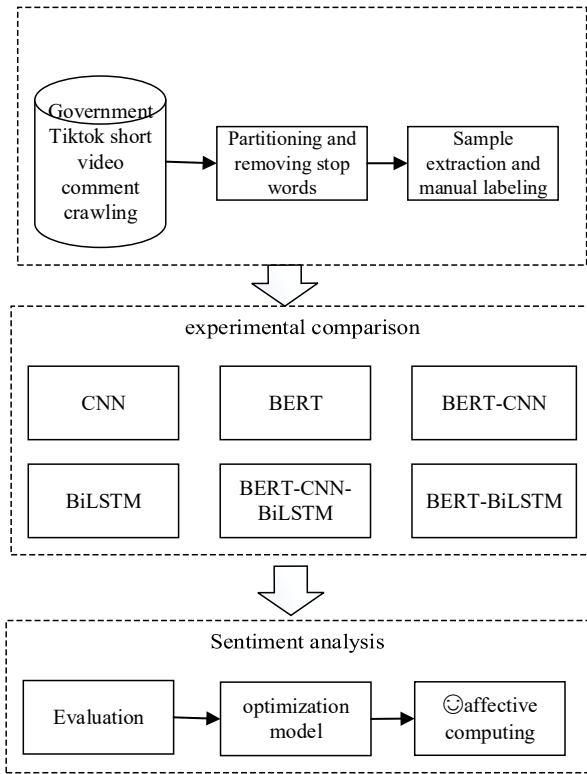


Fig 1. Emotional analysis of video comments

3.1. Data Sources

Grey Porpoise Data is a professional short video data

Table 1. Example of comment data annotation

User	comment	time	comment likes	Emotional tendencies
Pear***nest	China's strongest diplomat, domineering [likes] [likes] [likes]	2022/3/18 21:24	410 133	Positive
Long***Heart	Chinese team, come on! The Chinese team is the best!	2022/3/4 21:02	71 212	Positive
2***8	She was intentional!!! [sobbing uncontrollably] It's really disgusting	2022/4/6 6:03	522	negative
Ha***Ha	This kind of rascal needs to be severely punished!!!	2022/4/6 10:17	1	negative

3.4. Experimental Comparison

3.4.1. Model Introduction

(1) Building a BERT sentiment analysis model

The core of the BERT (Bidirectional Encoder Representations from Transformers) language pre training model based on the bidirectional transformer structure is to train the semantic vectors of input text words through multi-layer transformers. Its network structure [21] is shown in Figure 2. It effectively solves the problem of polysemy and ignoring the context in text analysis. Using attention mechanism and parallel processing of information can effectively obtain Semantic information of longer sentences. Text classification can be carried out by fine-tuning the parameters of the pre training model, with strong generalization ability [21].

analysis platform. By accessing the big data API to track the trend of short video traffic, users are provided with visual short video data, which is rich in content and has the function of classifying and ranking short videos, accurately and comprehensively summarizing various types of short video materials. This article used the Grey Porpoise Third Data Platform on April 6, 2022 to crawl the comment data of government related popular short video frequency for nearly 90 days from January 6, 2022 to April 6, 2022.

3.2. Text Preprocessing

Text preprocessing mainly includes word segmentation and removing stop words. Common word banks for discontinued words include Harbin Institute of Technology, Baidu, and other discontinued words. This article uses jieba segmentation to segment short video comments, and uses the Harbin Institute of Technology's discontinued word database to remove discontinued words such as "@", username, link, "de", "le", and "#" from comments.

3.3. Data Annotation

Based on the usefulness of the comments, 32 222 comments with likes greater than zero were selected for artificial sentiment annotation as model training. In order to eliminate the impact of data imbalance, we selected an equal number of positive and negative samples, among which 15000 positive and 15000 negative samples with better quality were selected. And divide the training set, testing set, and validation set into 80%, 10%, and 10%, as shown in Table 1 for annotated comments.

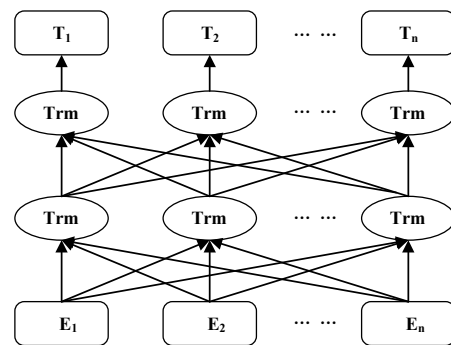


Fig 2. BERT network structure[21]

Like other models, the word (Token) needs to be converted into a vector input model through the Token embeddings layer. However, the difference between BERT is that it requires the addition of segment embeddings and position embeddings.

Segment embeddings are used to distinguish sentences, and BERT uses NSP (Next Sentence Perspective) to predict whether two sentences are adjacent. The Position Embedding vector is used to record the position order of words in input text, with the longest input being 512. Due to the difference between BERT and traditional RNN in processing word vectors according to sentence order, it is based on attention mechanism for parallel processing. In order to avoid the same sentence being output with the same result after disordering the order, temporal information needs to be added for differentiation. The calculation formulas are Formula (1) and Formula (2). The superposition of the above three vectors serves as the final input result. The Chinese corpus uses each word as the granularity of word segmentation, with the beginning of the sentence marked with [CLS] and the end marked with [SEP].

$$p_E(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \quad (1)$$

$$p_E(pos, 2i + 1) = \cos\left(\frac{pos}{10000^{\frac{2i+1}{d}}}\right) \quad (2)$$

In the formula, the vector representing the output position refers to the position of the unit word in the target sentence, represents the dimension of the feature vector, and represents the even and odd dimensions of the token embeddings, respectively.

After Token embeddings, the attention mechanism and feedforward neural network of the Transformer encoder will be entered in the next step. As shown in Figure 4.7, residual connection and normalization processing will be added to each layer of attention mechanism and feedforward neural network. The residual connection avoids the disappearance of gradients, completely transmitting the information trained from the previous layer to the next layer, and accelerating its training speed through layer normalization. Finally, the linear change is connected to Softmax for sentiment classification.

The self-attention mechanism effectively extracts the weight associated with other feature words by calculating the weight of each feature word. Each word vector is transformed to obtain three matrices: Query, key, and Value, represented by, respectively. Query is used to find the correlation with other words, while key is used to match and calculate the correlation with other word vectors to obtain the correlation score Value. After calculating the correlation score between each word and other words in the sentence, the weighted sum is used to obtain the correlation between this word and the sentence. The specific calculation formula is shown in formula (3), where d is the dimension of the word vector.

$$Attention(Q, K, V) = \text{soft max}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

The multi head attention mechanism can simultaneously focus on different positions of sentences, improving the model's feature extraction ability. The vectors output by each self-attention mechanism are concatenated using formulas such as (4) and (5) to form a linear variation matrix.

$$Attention_i = Attention(QW_q^i, kW_k^i, vW_v^i) \quad (4)$$

$$Multihead = \text{concat}(Attention_1, Attention_2, \dots, Attention_n)W \quad (5)$$

(2) BERT-CNN model

BERT can obtain semantic features over long distances through Transformer, CNN obtains local features through

different convolutional kernels, extracts important features through maximum pooling layer, and concatenates them into abstract features through full connectivity. The dynamic word vector trained by BERT is further input into CNN to extract local features of sentences, improving the accuracy of text sentiment analysis. The dimension of dynamic word vector is 768. In this paper, three different convolution kernels are used with the size of 3, 4, and 5, and the step size is 1. The activation function is the number of ReLU filters set to 256. To prevent overfitting, a drop out layer with a drop rate of 0.5 is added in front of the full connection layer, and then the SoftMax classification layer is accessed for emotion classification, as shown in Figure 3.

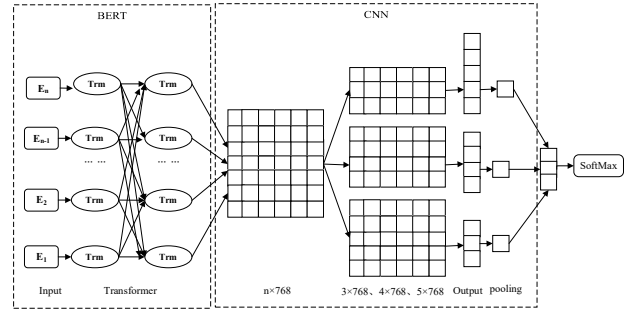


Fig 3. Structure diagram of BERT-CNN model

(3) BERT-BiLSTM model

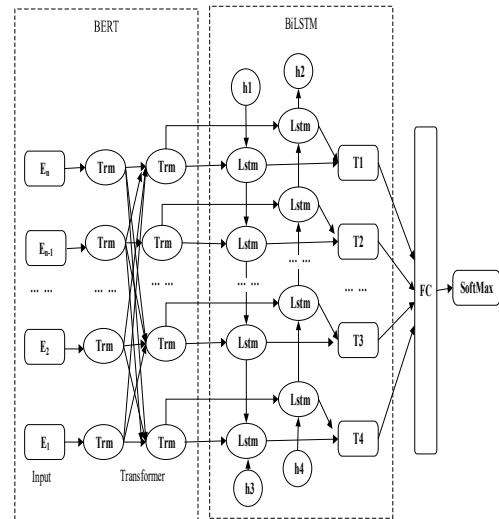


Fig 4. Structure diagram of BERT-LSTM model

BiLSTM extracts contextual semantic features from the front and back directions of a sentence, thereby obtaining dependency relationships for longer sentences. The output at time t depends on the product of the outputs in both positive and negative directions and their respective weights, plus the bias vector at time t . The positive output at time t depends on the previous positive output and the current input, similarly, the reverse output depends on the reverse output at the previous time and the current input. The BERT pre training model is not based on input timing, but rather on parallel operations through a multi head attention mechanism, and adopts a masking language model (MLM) training method, such as "cloze filling", masking words, and improving the generalization ability of the model through contextual prediction. However, its effectiveness is also affected to some extent. The dynamic semantic vector obtained through BERT

pre training model is further input into BiLSTM to obtain long-distance dependent features with continuous word order, improving the overall robustness of the model, as shown in Figure 4.

(4) BERT-CNN-BiLSTM model

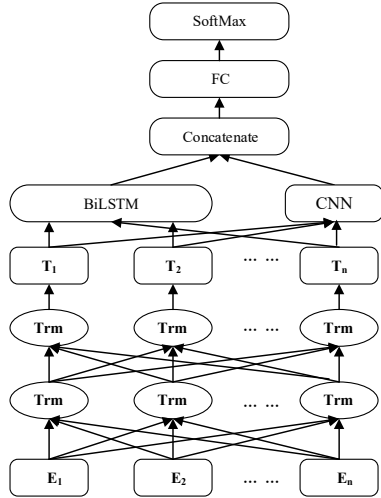


Fig 5. Structure diagram of BERT-CNN-LSTM model

In this paper, the dynamic word vector fused with context semantics obtained through BERT pre training model is input to CNN and BiLSTM to further extract local features and context long-distance semantic features to obtain more accurate Semantic information. Then, the features of CNN and BiLSTM are connected with Concatenate, and then through a full connection layer, Dropout processing is added to prevent overfitting. Finally, Softmax is used for prediction. The overall model diagram is shown in Figure 5.

3.4.2. Adjusting Parameters

Table 2. Model parameter configuration

Parameter	Value	Description
loss	cross_entropy	Cross entropy loss function
optimizer	Adam	optimizer
epochs	10	The number of iterations for network training
batch_size	16	The size of data selected for each iteration
max_length	512	Maximum sequence length
learning_rate	2e-5	Learning rate
hidden_size	768	Bert hidden layers
num_hidden_layers	2	Number of layers of encoder
dropout	0.5	Random deactivation to prevent overfitting
t_lstm	384	Number of LSTM hidden layers
filters	256	Number of convolutional kernels
kernel_size	3, 4, 5	Convolutional window size

During the model training process, the setting of parameters affects the effectiveness of the final model. This article draws on the relevant experience of parameter setting in existing research papers and uses GPU processors on the Kaggle official website for deep learning training and prediction. GPU is a parallel computing system composed of thousands of smaller and more efficient cores designed to handle multiple tasks simultaneously, such as 3D applications, video and image rendering, deep learning, and more. These are large-scale parallel tasks that can be processed more efficiently and quickly with GPU. This article is based on the fine-tuning parameters of the pre trained BERT model in Chinese "Chinese_L_12H-768_A-12". According to the characteristics of the model itself, the hardware requirements for model training, and the iterative training results, set the model parameters as shown in Table 2.

3.4.3. Model Evaluation

In depth learning, there are generally four cases of emotion dichotomization results, and the confusion matrix is shown in Table 3.

Table 3. Confusion matrix

parameter	Predictions are positive and	predictions are negative
Actual positive	TP	FN
Actual negative	FP	TN

Among them, TP is the number of comments predicted to be positive when the actual emotional tendency is positive; FN represents the number of comments with positive actual emotional tendencies and negative predictions; FP represents the number of comments with negative emotional tendencies and positive predictions; TN is the actual negative emotion, predicted as the number of negative comments, TP+FN+FP+TN=the total number of samples.

Precision is the proportion of correct classification, while Recall represents the proportion of correct validation in actual correct classification. Accuracy and recall are mutually constrained, and generally higher accuracy leads to lower recall. The F1 (F-measure) indicator is widely used in machine learning to measure the effectiveness of models, consisting of Precision and Recall, which can reflect both the accuracy and correctness of the model. The definition formulas for accuracy, recall, and F1 are shown in formulas (6), (7), and (8).

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = \frac{2 \times Recall \times Precision}{Precision + Recall} \quad (8)$$

This article selects the open-source version of BERT on Github and fine-tuning it based on the characteristics of government short video comments to improve BERT's classification performance in downstream tasks. CNN, BiLSTM, BERT, BERT BiLSTM, BERT-CNN, and BERT-CNN BiLSTM were selected for experimental comparison. This article selects the optimal value of epochs iteration to evaluate the emotional classification performance of each model. The classification performance of each model test set is shown in Table 4.

Table 4. Comparison results of different models

Data samples	Precision(%)	recall(%)	F1(%)
CNN	81.03	72.03	76.05
BiLSTM	82.99	75.07	78.75
BERT	85.94	82.63	84.18
BERT-CNN	90.00	87.81	88.88
BERT-BiLSTM	89.60	88.26	88.93
BERT-CNN-BiLSTM	92.90	91.96	92.43

4. Summary

The results show that the classification performance of the BERT pre trained model is significantly better than traditional deep learning algorithms, and BERT-CNN and BERT-BiLSTM are better than BERT. This paper uses a multi strategy algorithm fusion method, which improves the F1 of the BERT model by 8.76% compared to a single algorithm. BERT-CNN BiLSTM is better than a single algorithm, which increases the F1 value of BERT-CNN by 3.55% compared to BERT-BiLSTM, and improves the text feature extraction by 3.50% compared to BERT-BiLSTM. The accuracy of the model is 92.90%, The recall rate is 91.96%, and F1 is 92.43%.

References

- [1] Liu Luchuan, Sun Yilu The impact of negative online comments with different emotional intensities on consumer attitudes: an experimental study based on regulatory orientation theory [J] Intelligence Theory and Practice, 2020, 43 (5): 163-169.
- [2] Wang Chaoyang, Yu Huilin The Emotional Preference Effect in News Short Video Communication: An Empirical Study Based on the Social Segment of Pear Video [J] News and Communication Review, 2019, 72 (3): 42-55.
- [3] Qian Chen, Zhang Tao Analysis of transmission characteristics and influencing factors of AIDS related short videos [J] AIDS and STD in China, 2021, 27 (12): 1394-1399.
- [4] Du Yongping, Zhao Xiaozheng, and Pei Bingbing Short text sentiment classification based on CNN-LSTM model [J] Journal of Beijing University of Technology, 2019, 45 (7): 662-670.
- [5] Wei Lai, Wang Weijie Emotional Analysis of Bullet Screen Information on Video Websites Based on the Theory of Interactive Ritual Chain - Taking Bilibili Health Science Popularization Video as an Example [J] Intelligence Theory and Practice, 2022, 45 (9): 119-126.
- [6] Yang Dashen, Li Shixuan, Cong Yingnan Research on the influencing factors of short video transmission effect of Tiktok reading promotion [J] Library science Research, 2021 (23): 34-44.
- [7] Qiu Jiangnan, Ge Yidi The Impact of Social Media Emotions on Information Behavior: A Comparative Study Based on Two Types of Disaster Events [J] Management Science, 2020, 33 (1): 3-15.
- [8] Guo Xiuyuan, Xue Shipai, Xu Jiayao Research on the audience engagement of regional promotional videos [J] China Television, 2022 (11): 68-74.
- [9] Wang Haiyan, Li Jianhui, Yang Fenglei Overview of Support Vector Machine Theory and Algorithm Research [J] Computer Application Research, 2014, 31 (5): 1281-1286.
- [10] Zhang Lei, Xiao Siyao, Yang Zelai, etc Research on Public Opinion Processing Based on Improved KNN Algorithm and SIR Model [J] Computer Simulation, 2021, 38 (5): 477-483.
- [11] Su Ying, Zhang Yong, Hu Po, etc Emotional analysis based on the combination of naive Bayes and potential Dirichlet distribution [J] Computer Applications, 2016, 36 (6): 1613-1618.
- [12] Wan Yanping, Gu Jiazhen, Zhang Fang Integrating Improved Stacking and Rules for Text Emotional Analysis [J] Small Micro Computer Systems, 2021, 42 (7): 1389-1395.
- [13] He Xueqin, Yang Wenzhong, Wu Shouer Slam, et al Integrating syntactic rules and CNN for emotional analysis of tourism reviews [J] Computer Engineering and Design, 2019, 40 (11): 3306-3312.
- [14] Shang Rongxuan, Zhang Bin, Mi Jianing An end-to-end aspect level sentiment analysis method for government app comments based on BRNN [J] Data Analysis and Knowledge Discovery, 2022, 6 (1): 364-375.
- [15] Zhang Yangsen, Zheng Jia, Huang Gaijuan, et al Weibo sentiment analysis method based on dual attention model [J] Journal of Tsinghua University (Natural Science Edition), 2018, 58 (2): 122-130.
- [16] Rehman A U, Malik A K, Raza B, et al. A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis[J]. Multimedia Tools and Applications, 2019, 78(18): 26597-26613.
- [17] Wei J, Liao J, Yang Z, et al. BiLSTM with multi-polarity orthogonal attention for implicit sentiment analysis[J]. Neurocomputing, 2020, 383: 165-173.
- [18] Duan Dandan, Tang Jiashan, Wen Yong, et al Chinese short text classification algorithm based on BERT model [J] Computer Engineering, 2021, 47 (1): 79-86.
- [19] Li Tiefei, Sheng Long, Wu Di Research on Text Classification Method of BERT-TECNN Model [J] Computer Engineering and Applications, 2021, 57 (18): 186-193.
- [20] Chen Zhiqun, Ju Ting Research on Weibo Comment Tendency Analysis Based on BERT and Bidirectional LSTM [J] Intelligence Theory and Practice, 2020, 43 (8): 173-177.
- [21] Zhang Mingquan, Zhou Hui, Cao Jingang Research on Double BERT Directed Emotional Text Classification Based on Attention Mechanism [J] Journal of Intelligent Systems, 2022, 17 (6): 1220-1227.