Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 2, No. 1, 2022

Chinese Text Sentiment Classification Based on ERNIE and BiLSTM-AT

Jianrong Wang, Naiyi Li*

School of Mathematics and Computer Science, Guangdong Ocean University, Zhanjiang, Guangdong, China
* **Corresponding author**: Naiyi Li (Email: linaiyi1979@163.com)

Abstract: Chinese text sentiment classification is a sub-task of natural language processing. However, when text representation is carried out, the polysemy of a word cannot be processed when using the traditional language model to construct the word vector, and the long-distance text information cannot be fully extracted when extracting text features. To solve this problem, this paper proposes a text sentiment classification model combining ERNIE and BiLSTM-AT. First, the pre-training model ERNIE is used to obtain the word vector representation of the fused statement context. Then, the bidirectional long-short-term memory neural network is used to extract the context information and depth semantic information of the text. Then, the attention mechanism is used to assign the corresponding weights to the hidden layer vectors of each time step output by the BiLSTM layer, and the weighted summation is integrated into the sentence features. Finally, the softmax function is used to calculate the probability distribution of the emotional category of the text in the output layer. The results show that the proposed model can achieve high accuracy on both hotel reviews and takeaway reviews. Based on the pre-training model, adding bidirectional long-term and short-term memory network and attention mechanism is beneficial to improve the classification effect of the model, and has certain practicability in text sentiment classification tasks.

Keywords: Text sentiment analysis; Pre-training model; Bidirectional long-term and short-term memory network; Attention mechanism.

1. Introduction

With the rapid development of information technology and mobile Internet, China has gradually stepped into the comprehensive Internet era. As early as the 49th Statistical Report on the Development of Internet in China released by China Internet Network Information Center (CNNIC) [1], it was pointed out that by December 2021, the number of Chinese netizens had reached 1.032 billion. The Internet penetration rate reached 73.0%. Users are used to publishing content on the Internet, including opinions on social hot topics, national policies, products, services, etc. All kinds of comments and opinions are more or less attached to users' emotional attitudes. Quick and accurate identification of emotion types plays an important role in many fields. For example, in terms of public opinion, it is possible to understand public opinions by paying attention to netizens' emotional attitudes toward social affairs, which can effectively prevent harmful events and facilitate the implementation of policies. In the business service, the analysis of user evaluation can help improve the quality of goods and services, and do a good job in customer relationship management; In the aspect of psychological and emotional counseling, through the analysis of the corpus expression of the object, it can quickly identify the object's emotional attitude, and conduct targeted counseling for users with psychological problems. Nowadays, according to different methods used in classification, text emotion classification can be roughly divided into three categories, namely emotion classification method based on emotion dictionary, emotion classification method based on traditional machine learning, and emotion classification method based on deep learning [2]. The text emotion classification method based on the emotion dictionary is based on the emotion dictionary. The pre-processed text data is used to match the emotion words and emotion polarity contained in the emotion dictionary, and the emotion polarity is classified according to different granularity. Li Yuqing [3] et al. built a bilingual multi-class emotion dictionary based on the bilingual dictionary method and conducted a multi-class emotion classification experiment, and the model experiment achieved good results. However, this classification method is limited by the scale of the sentiment dictionary, so the dictionary base should be continuously expanded. The sentiment analysis method based on machine learning needs to select the classification algorithm, obtain the model parameters through data training, and then use the trained model to predict the results. Tang Huifeng [4] et al. used common machine learning methods (SVM, KNN, etc.) to conduct the experiment of Chinese text sentiment classification, and to select appropriate feature representation and selection methods, SVM can achieve the optimal classification effect. However, this classification method fails to fully consider the position information of the words in the sentence, which will lose the text context information and affect the classification effect. Since the deep learning-based sentiment classification method will make use of the word order information of the text, extract the semantic information of the words, and take full account of the advantages of contextual information [2], many scholars have carried out relevant studies. Among the deep learning-based emotion classification methods, some scholars use a single neural network for emotion classification. For example, Jelodar [5] et al. used the LSTM model when analyzing comments on COVID-19, and the experimental results can provide data support for relevant decisions. Some scholars have considered the advantages and disadvantages of different neural network models, and then improved and mixed the models, and achieved good experimental results in the task of emotion classification. For example, Luo Fan et al. [6] proposed a multi-layer network H-RNN-CNN, combined

the advantages of the two models, used RNN to model text sequences, and used CNN to identify information across sentences. The model has obtained good experimental results. After seeing important achievements in the application of attention mechanisms in the field of visual images, some scholars tried to use attention mechanisms in the field of natural language processing. Bahdanau et al. [7] added an attention mechanism to the machine translation task, and the success of his experiment meant that the attention mechanism began to be applied in the field of natural language. Scholars have successively applied attention mechanisms to subtasks in the field, such as the two-layer CNN-BiLSTM proposed by Liu Fishing et al. [8], which added sentence emotion polarity ordering and attention mechanism. The model can fully extract text features and optimize input text features, and experiments show that the model has certain effectiveness and feasibility. With the emergence of the natural language pre-training model, scholars have been using the pre-training model in the task of emotion classification, and have achieved better results in emotion classification. Pre-training models [9] are divided into static models and dynamic models, among which the word2vec model [10] and glove model are static models, and the ELMO model [11], GPT model [12], BERT model and ERNIE model [13] are dynamic models. Because static word vectors can't represent polysemous words well, text representation is limited. With the introduction of dynamic word vectors, this problem has been effectively solved. The emergence of dynamic models GPT and BERT, both based on the Transformer model, provides a new way of thinking when dealing with natural language processing tasks in the future. The BERT model has a good performance in many tasks. It performs the vector representation of text according to the word level, extracts semantic information combined with context, and deals with polysemous words well. However, the BERT model does not make use of lexical, grammatical structure, and semantic information in sentences to learn modeling, which is difficult to provide a good vector representation of newly emerged words, while the ERNIE model [14] fully considers the lexical, grammatical structure and semantic information of text for modeling, improving the universality of semantic expression. In natural language processing tasks using the neural network method, how to convert text characters into digital features combined with text information will determine the upper limit of model performance. If a static pre-training model is used, it will only learn the static vector of words and ignore the multi-semantics of polysemy in the text. Therefore, the dynamic pre-training model ERNIE will be used in this paper. This model can effectively solve the polysemy representation problem through the sentence word vector generated after training. The overall process is to use each word segmentation vector of the output matrix of the ERNIE pre-training model as the input of the bidirectional long and short-term memory neural network, and then use the attention mechanism to weigh the output through the neural network. More weight is given to the time step output that has a greater impact on the emotional label, and more emphasis is placed on the words in the sentence to improve the overall performance of the model.

2. Text emotion classification model based on ERNIE and BiLSTM-AT

The text emotion classification model combined with ERNIE and BilSTM-AT is shown in Figure 1 below. The model has six layers, which are data preprocessing layer, word embedding layer, ERNIE layer, BiLSTM semantic extraction layer, attention mechanism layer and the final output layer. Among them, the data preprocessing layer will remove the useless data in the text, leaving only the text that can express the semantic information; In the word embedding layer, the word segmentation is mapped to the corresponding word vector in the word direction scale; In ERNIE layer, the text word vector is transformed into the dynamic vector representation of the sentence by combining the semantic information of the text. In the BiLSTM layer, the text context semantic information will be further extracted. At the level of attention mechanism, different weights are given according to the importance of each participle to the sentence emotion classification, and the weighted summation can get the semantic information at the sentence level. In the output layer, the fully connected network will be used to obtain the probability of each category, and then the classification results will be obtained.

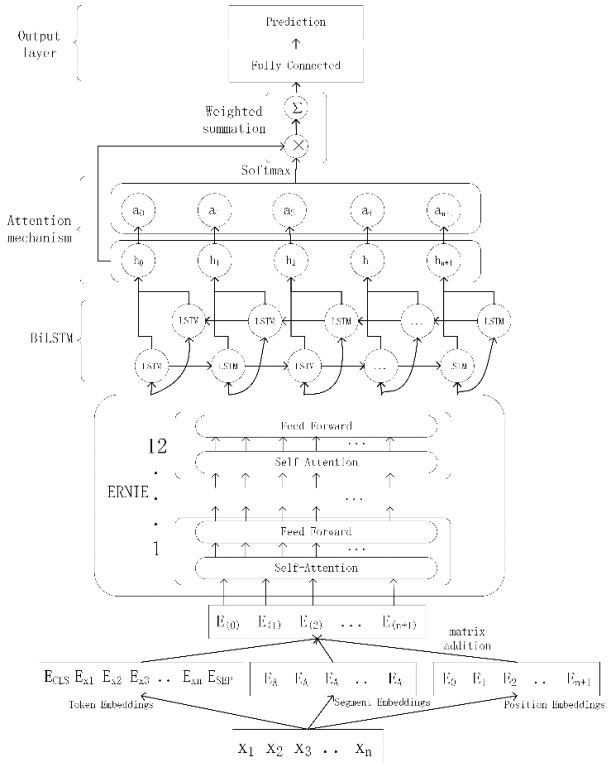

Figure 1. Text emotion classification model based on ERNIE and BiLSTM-AT

2.1. Data preprocessing and word static vector representation

Due to the large amount of noise in the comment text data, it will affect the extraction of subsequent sentence features, so it is necessary to de-noise the text before the extraction of sentence features. Regular expressions can be used to remove punctuation marks and expressions, etc., and then word segmentation is carried out to limit the sentence length to no more than the maximum length minus 2, and then [CLS] and [SEP] are added at the beginning and end of the sentence respectively. Then, each character in the sentence is converted into a corresponding vector. Without input into the model for training, these vectors are just static vectors that cannot solve the polysemous problem. After the training, the dynamic vector representation of the text that integrates the sentence information will be obtained, which can better solve the polysemous problem.

2.2. ERNIE pre-training model

ERNIE model is a kind of enhanced presentation pre-training model which realizes knowledge integration through mask mechanism. It includes two parts: text encoder and knowledge integration. The former has the same structure as the encoder in transformer. It captures the context information of each mark in the sentence through the self-attention mechanism, inputs the result into the feed-forward neural network, and generates the corresponding word vector representation at last. The encoder structure is shown in Figure 2 below.

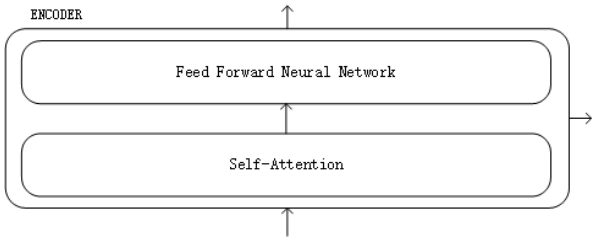

Figure 2. Transformer encoder structure diagram

Knowledge integration is accomplished through the multi-stage knowledge mask strategy to obtain the language representation of the fusion phrase level and entity level [15]. The ERNIE model structure is shown in Figure 3 below.

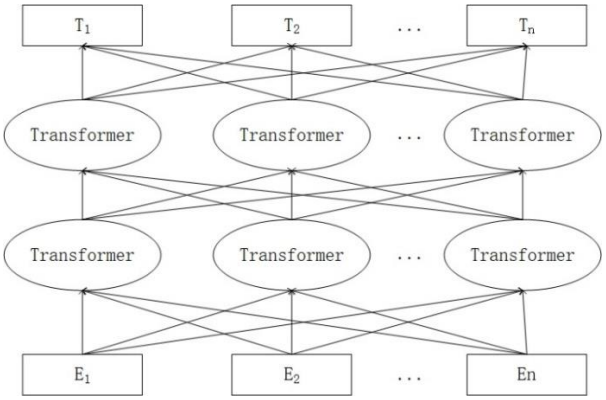

Figure 3. Network structure of ERNIE model

The input vector is formed by the corresponding addition of the position vector, sentence vector, and word embedding vector, as shown in Figure 4 below.

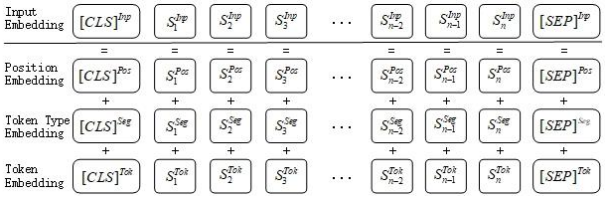

Figure 4. Input structure of the ERNIE model

As can be seen from the ERNIE model structure diagram, the output part is the word vector combined with the sentence context. Because the bidirectional language model will use contextual information when predicting words, and the transformer has its self-attention, when predicting other words, the information of the word will also be included in the network parameters of the previous layer, resulting in leakage of information. To solve this problem, the BERT model randomly masks out about 15% of the words and hides part of the input sequence. ERNIE model improves the BERT model, which can only cover words. In addition to covering sequence words, the Ernie model also uses a multi-stage knowledge mask strategy to cover phrases and entities of sentences.

2.2.1. Transformer Encoder

The basic structure of the ERNIE model is a multi-layer bidirectional transformer encoder stack, where the encoder structure is the same, and the weight is not shared. The self-attention mechanism will be used in the encoder to integrate the sentence information according to the importance of the words in the sentence, to improve the utilization of the features. The specific calculation formula is as follows:

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

Where: Q,K,V - Input word vector matrix; d_k - Input vector dimensions.

2.2.2. Knowledge Integration

Different from BERT's pre-training model, ERNIE [14] can make good use of the lexical, grammatical structure, and semantic information in the training data and propose multi-stage Knowledge Masking Strategies to integrate the phrase and entity level knowledge into the language representation. Instead of directly adding knowledge embedding, this greatly enhances the syntactic and grammatical representation ability of word vectors. In addition to the BERT model which only provides Basic-level Masking, it adds Phrase-based Masking and Entity-level Masking. The example of the knowledge mask policy used in ERNIE is shown in Figure 5[17].

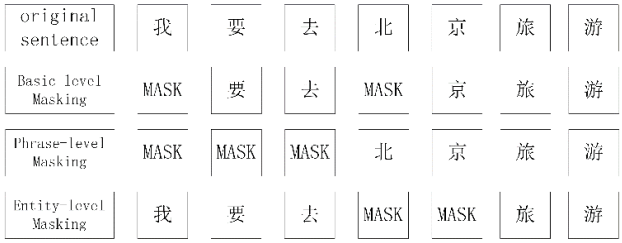

Figure 5. Example of ERNIE knowledge mask strategy

2.3. Text semantic extraction BiLSTM layer and attention mechanism

The main function of the BiLSTM layer is to extract text features and construct two LSTM networks that are good at dealing with long dependencies to obtain forward and reverse text information, which is easier to extract the deep semantic expression hidden in the text. Compared with traditional RNN, LSTM can deal with long-term dependence and gradient disappearance mainly due to its internal three gate functions: input gate, forget gate, and output gate. These three functions are used to control information retention. The model structure is shown in Figure 6 below.

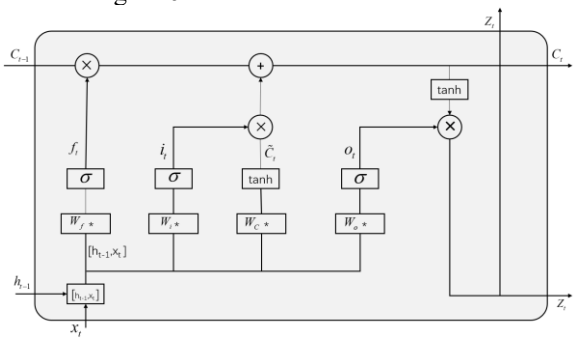

Figure 6. LSTM model structure

The forward calculation process of a single LSTM memory unit at the moment is as follows:

(1) Input unit, processing the input of the current sequence position:

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \times [h_{t-1}, x_t] + b_C), C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (3)$$

(2) Forgetting unit, forgetting information:

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (4)$$

(3) Update the unit, update the status of the unit after the abandoned information:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (5)$$

(4) Output unit, determine the output value:

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (6)$$

$$Z_t = o_t \times \tanh(C_t) \quad (7)$$

Where: $\{W_*, b_*\}$ the parameters to be trained; h_{t-1} : the previous time of t hides the state of the layer; x_t : the input at time t.

In language expression, the semantic meaning of words in sentences is affected by the context, so it is necessary to extract sentence features in both positive and negative directions. BiLSTM uses LSTM to extract features in both positive and negative directions of text, and finally integrates and outputs the hidden layer results of both directions. The calculation process is as follows:

$$H_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (8)$$

$$H = \{H_1, ..., H_t, ..., H_T\} \quad (9)$$

Where: $\vec{h}$: Forward hidden layer vector; $\overleftarrow{h}$: Backward hidden layer vector; H_t : Hidden layer vector at time t; H : Hidden layer vector at all times.

Passing through the BiLSTM layer will obtain the hidden layer vector at all times as input to the attention mechanism layer. Initialize W_g , multiply W_g by each H_t respectively, and then obtain u_t through the tanh function. Then obtain the corresponding weight value a_t of each hidden layer through weight calculation. Finally, carry out weighted summation of the hidden layer vectors at all times to obtain the sentence feature representing R . The calculation process is as follows:

$$u_t = \tanh(W_g \cdot H_t + b_g) \quad (10)$$

$$\alpha_t = \frac{exp(u_t)}{\sum_t exp(u_t)} \quad (11)$$

$$R = \sum_t \alpha_t H_t \quad (12)$$

Where, W_g is the parameter matrix of the attention mechanism layer, and b_g is the bias vector. Both of them need to be learned during training and $exp(\ )$ is an exponential function.

The weighted sentence representation is obtained through the attention mechanism layer, and the sentence vector is input into the fully connected network for spatial mapping. Then, the sentence emotion classification probability distribution P_R is obtained through the softmax function, and the formula is as follows:

$$P_R = Softmax(W_S R + b_S) \quad (13)$$

Among them, W_S and b_S are the parameter matrix and bias in the fully connected network respectively, which need to be learned in the training process. Finally, the category corresponding to the maximum probability is selected from the output as the prediction result [16].

3. Experiment and result analysis

3.1. Experimental environment and data set

The development environment for the experiment in this paper is the paddle, and the development tool is AI Studio based on Baidu's deep learning open-source platform Fei Paddle. The development language is python, and the free GPU-accelerated program provided by the online platform is used. The data used in the experiment are user review data set and hotel review data set of a food delivery platform, among which there are 11,987 food review data sets, including 4,000 positive and 7,987 negative evaluation data. There are 7766 pieces of hotel evaluation data set, including 5322 pieces of positive evaluation data and 2444 pieces of negative and negative evaluation data. In the experiment, the data set will be divided into the training set and the test set according to 9:1.

Table 1. Partition of data set

Dataset	Training set		Test set	
	Positive	Negative	Positive	Negative
Takeaway review date set	3600	7189	400	798
Hotel review data set	4730	2172	592	272

3.2. Experimental evaluation criteria

To verify the effectiveness of the model in text classification experiments, some evaluation indexes commonly used in classification models are introduced to measure the model. The evaluation indexes used in the binary experiment include Precision (P-value), Recall (R-value), and F1 value. The equivalent values of TP, FP, FN, and TN used in the calculation are as follows:

Table 2. Classification discriminant confusion matrix

Confusion matrix		Real label	
		Positive	Negative
Predictive label	Positive	TP	FP
	Negative	FN	TN

Accuracy, recall rate and value are calculated as follows:

$$P = \frac{TP}{TP + FP} \quad (14)$$

$$R = \frac{TP}{TP + FN} \quad (15)$$

$$F_1 = \frac{2 * P * R}{P + R} \quad (16)$$

3.3. Selection of experimental parameters

The model experiment results will be related to the parameter Settings. After several comparative experiments, the model parameter values are set as follows: the maximum length of the input text is 128; The pre-training model output word vector dimension is 768 dimensions; The output dimension of BiLSTM is 256 dimensions. dropout set to 0.5; Choose the cross entropy as the loss function; AdamW was selected as the optimizer, and the learning rate was set to 5E-5; Set the epoch of training set to 5 and Batch_size to 32.

3.4. Experimental comparison and analysis

To verify the effectiveness of the model in this paper, the same data set and the same experimental environment are used to compare the emotion classification model in this paper with other emotion classification models, train on the hotel evaluation training set, and finally test on the corresponding test set. The test results are taken as the macro average value of the evaluation index, as shown in Table 3:

Table 3. Comparison of experimental results of hotel review test set

Number of the model	Model	Precision	Recall	F1
1	word2vec-LSTM	0.7930	0.7224	0.7382
2	Word2vec-LSTM-AT	0.8549	0.8288	0.8399
3	word2vec-BiLSTM	0.8521	0.8398	0.8455
4	Word2vec-BiRNN-AT	0.8590	0.8360	0.8462
5	Word2vec-BiLSTM-AT	0.8818	0.8527	0.8647
6	ERNIE	0.8952	0.8995	0.8973
7	ERNIE-BiLSTM	0.8979	0.8972	0.8976
8	ERNIE-BiLSTM-AT	0.9058	0.8869	0.8953

In the table, Model 1 introduces the word vector obtained by using word2vec in the external corpus training, then inputs the text sentence represented by the word vector in the word embedding layer into the LSTM model to extract the text depth feature, and finally gets the prediction result through the output of the full connection layer. In Model 2, the same external word vector model is introduced, and then the sentence vector is input into LSTM-AT, and the prediction result is obtained through the full-connection layer output. In Model 3, the pre-trained word vector model was introduced, and then the text sentence vector was input into the model BiLSTM to extract the depth features, and the prediction results were output in the full-connection layer. In model 4, the pre-trained word vector is used to input the obtained text segmentation vector into BiRNN-AT, and then the prediction result is obtained through the full-connection layer output. In Model 5, the same pre-trained word vector was used to input the embedded word vector into the downstream BiLSTM-AT, and the prediction result was obtained through the full-connection layer output. Model 6 inputs the first label [CLS] vector of the ERNIE model output sentence into the full connection layer and softmax to get the prediction result; In model 7, the hidden state sequence at the last layer of the ERNIE model is used as the word vector of the sentence, and the word vector is input into the BiLSTM model to get the prediction result. Model 8 used ERNIE pre-training technology to get the word vector of the text, and then input the word vector into the downstream model BiLSTM-AT to get the prediction result. As can be seen from the table, comparing experiments 1, 3, 2, and 4, it can be found that bidirectional LSTM can achieve a better classification effect than one-way LSTM, mainly because the bidirectional structure can extract text context information better. By comparing experiments 1, 2,3, 5 and 7, and 8, it can be found that by adding the attention mechanism to the original model, the evaluation index of classification results is improved, mainly because the introduction of the attention mechanism can make the model focus on words that are more important for emotion analysis. In contrast, in experiment 6, the first sentence flag [CLS] was directly used, and in experiment 7, the text word vector obtained by ERNIE pre-training was used as text representation and input into BiLSTM for further feature extraction. Then, classification results were output through the full connection layer. The accuracy rate and F1 value of the classification index were slightly improved. In experiment 8, an attention mechanism was added to increase the weight of more important words for emotional labels in text statements, and the accuracy rate of the evaluation index was improved to a certain extent.

To test the effectiveness of the model in this paper, the takeout review data set was used for training and testing under the same environment. Compared with Model 1 and model 2, various classification evaluation indexes were improved by adding BiLSTM. Model 3 incorporated an attention mechanism on the basis of Model 2, and the indexes were further improved. Specific experimental results are shown in Table 4:

Table 4. Three Scheme comparing

Number of the model	Model	Precision	Recall	F1
1	ERNIE	0.8929	0.8841	0.8883
2	ERNIE-BiLSTM	0.8993	0.8890	0.8937
3	ERNIE-BiLSTM-AT	0.9061	0.9020	0.9040

4. Conclusion

In this paper, we propose a text emotion classification model combining ERNIE and BiLSTM-AT to achieve sentence-level emotion classification. We use the ERNIE pre-training model to get word vector representation integrating text context. We use neural networks and attention mechanisms to extract text word vector features to better understand text semantic information. The model in this paper is tested on the test set, and some evaluation indexes are improved to a certain extent, indicating that the model has a certain effect. Only binary task is considered in this model. In future work, multi-emotion text classification experiments can be considered, and the application of the latest ERNIE3.0 model in emotion classification tasks can also be tried.

Acknowledgements

This work was partially supported by Natural Science Foundation of China (12161074), Natural Science Foundation of Guangdong Province (2022A1515010978) and Natural Science Foundation of Guangdong Ocean University (R17083, C17201, P16091).

References

[1] CNNIC Internet Research. The 49th Statistical Report on the Development of Internet in China [R]. Beijing: China Internet Network Information Center, 2022.

[2] Wang Ting, Yang Wenzhong. Review on Text Sentiment Analysis Methods [J]. Computer Engineering and Applications,2021,57(12):11-24.

[3] Li Yuqing, Li Xin, Han Xu, Song Dandan, LIAO Lejian. A Bilingual Lexicon-Based Multi-class Semantic Orientation Analysis for Microblogs[J]. Acta Electronica Sinica,2016,44(09):2068-2073.

[4] Tang Huifeng, Tan Songbo, Cheng Xueqi. Research on Sentiment Classification of Chinese Reviews Based on Supervised Machine Learning Techniques., 2007, 21(6):88-94.

[5] H Jelodar, Wang Y , Orji R , et al. Deep Sentiment Classification and Topic Discovery on Novel Coronavirusor COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach[J]. arXiv,2020.

[6] Luo Fan, Wang Houfeng. Chinese text sentiment classification based on hierarchical network of RNN and CNN. Journal of Beijing University (Natural Science), 2018, v.54; No.287(03):4-10.

[7] Bahdanau, Dzmitry et al. “Neural Machine Translation by Jointly Learning to Align and Translate[J].” CoRRabs/1409.0473 (2015): n. pag.

[8] Liu Fasheng, Xu Minlin, Deng Xiaohong. Research on Emotion Analysis Combining Attention Mechanism and Sentence Ordering [J]. Computer Engineering and Applications,20,56(13):12-19.

[9] Li Zhou-jun, Fan Yu, WU Xian-jie. A review of pre-training techniques for Natural Language Processing [J]. Computer Science,20,47(03):162-173.

[10] Tomas Mikolov et al. Efficient Estimation of Word Representations in Vector Space[J]. CoRR, 2013, abs/1301.3781

[11] Peter M E, Neumann M, Iyyer M, et al. Deep contextualized word representations[J]. arXiv：1802. 05365，2018.

[12] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[J/OL].[2020-07-01].https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf.

[13] Gao Z J, Feng A, Song X Y, et al. Target-dependent sentiment classification with BERT[J]. IEEE Access, 2019,7: 154290-154299.

[14] Yu Sun et al. ERNIE: Enhanced Representation through Knowledge Integration.[J]. CoRR, 2019, abs/1904.09223

[15] Jingsheng Lei, Ye Qian." Chinese-Text classification method based on ERNIE-BiGRU." Journal of Shanghai University of Electric Power 36.04(2020):329-335+350.

[16] Zemin Huang, Xiaoling Wu, Yinggang Wu, Jie Ling. Analysis of Chinese text emotions combining BERT and BiSRU-AT[J]. Computer Engineering and Science, 201,43(09):1668-1675.

[17] Chen Jie, Ma Jing, Li Xiaofeng. Data Analysis and Knowledge Discovery, 201,5(09):21-30.