Journal of Artificial Intelligence and Capsule Networks (ISSN: 2582-2012)
www.irojournals.com/aicn/

Personalized Explainable Recommendation based on BERT

WANG Yuan-mei[1], WANG Ya-jun[1*], ZHAO Shi-bo[1]

School of Electronic and Information Engineering, Liaoning University of Technology, Jinzhou, 121001, China.

E-mail: [1]2276325238@qq.com, [2]wyj_lg@163.com, [3]891687864@qq.com

Abstract

With the proliferation of text information on the Internet, the rapid evolution of artificial intelligence, and the wide application of machine learning and deep learning, text emotion analysis has been widely concerned by the academic community. Personalized recommendation system has emerged as the circumstances demanded, and has quickly gained interest in both industry and academic circles. It has gradually become an extremely important part of people's life and work in many fields, such as e-commerce, short video content, take-out service, online advertising push and so on. First, the Chinese comment text of an item is analyzed in this paper. Since the absolute standard of binary text sentiment analysis cannot meet the need of the recommendation system to recommend the item to users for interpretation, a ternary text sentiment analysis method based on BERT model is used to combine with the idiosyncrasies of text data is proposed to solve the problems of poor Chinese text representation, low exactitude, and inability to precisely comprehend the semantic information expressed in the text, which are caused by polysemy of Chinese version. The proposed method can generate interpretable recommendations for items that users are interested in. The relevant properties of the text are captured by the Transformer encoder in the BERT model, meanwhile the attention framework is used to weight the information recovered from the pattern to highlight the hinge information in the comment on the text. Second, the SoftMax function is used for categorizing the text aspect data of users' reviews of items, as well as finally the recommendation system recommends interested items to users and produces emotional and coloration reasons for recommendation that are accorded with users' justifications. The method is applied to real datasets, and the results show that text breakdown effect has been achieved, which greatly improves the interpretability of recommendation system, which is more in line with users' ideas.

DOI: https://doi.org/10.36548/jaicn.2023.1.003
Received: 26.01.2023, received in revised form: 26.02.2023, accepted: 13.03.2023, published: 24.03.2023

Keywords: Personalized recommendation, Transformer, BERT, Emotional analysis

1. Introduction

People's daily life is gradually getting connected with the Internet more and more. China Internet Information Center announced the report of the 50th report of China Internet development in Beijing, which shows that by June 2022, the Internet penetration rate has reached 74.4% [1]. Therefore, the Chinese netizens play an increasingly important role on the Internet. The more copies, pictures, and comments they publish, the higher the value and importance of the information they contain. The consumption evaluation of these goods can effectively help users to understand the information about the services of businesses, goods, brand reputation and other aspects of information, and help users make the most effective choice and judgment when selecting businesses and goods. Therefore, it is significant research to analyze the user's comments on goods. Text emotion analysis has proved to be a major research task in the recommendation system.

Internet enables people to buy what they need without leaving home, and they can choose other alternative items according to the input content, which greatly improves people's selectivity. However, when users browse major websites, they are often overwhelmed by many miscellaneous news, and it isn’t easy to find the resources or items that best match them in the fastest and most efficient way. So, to deal with the problem, the recommendation field rose rapidly since 1990s and has gradually developed to the present. The essence of recommendation system can be regarded as a kind of information filtering system, which aims to recommend the matching items for users and avoid users from spending a lot of time or energy looking for the desired resources or items in numerous network resources. Therefore, the recommendation system not only improves the user's experience, but also increases the sales profit. Accurate recommendation results can transform the user's click-through rate into purchase behavior, and then improve the distribution rate to make the turnover surge. Judging from this, the task of text emotion classification has been more and more significant in investigated value and broad market application prospects, which can create huge social and economic value for the socialist modernization construction. Therefore, based on the deep learning knowledge, Chinese texts with emotion can be classified accurately and quickly by using word vector, natural language processing and other related technologies. The research of this paper is of great significance and value.

The organization details of the article are as follows: Section 1 mainly introduces the background and meaning of this paper and briefly summarizes the structure of this paper. Section 2 contains the related works which introduces the present state of research on text emotion analysis and its enormous challenge. The third section is the main work, where the content and method of this research are elaborated, and the Bert model and the attention mechanism are briefly explained. Section 4 includes the experimental results and analysis. The model is tested on the actual data, and the prediction accuracy and descriptive quality of the model are compared with other algorithms. The fifth section is the conclusion and the perspective, where the main contents of the body are summarized, and the work and research of the next step are briefed.

2. Related Works

In 2015, Chung et al.,[2] introduced the bottom-up article representation method and realized the single sentence representation through the circular neural network. When dealing with the internal relations and semantics of sentences, the gated RNN was used to code and establish connections. In 2016, Liu et al.,[3] introduced attention into the LSTM model, which effectively applied its importance of obtaining different language environment information for a given aspect and solved the question of aspect level in sentiment categorization tasks. In 2017, Google proposed a Transformer model for solving version sequence problems combined with the attention mechanism, which completely abandoned the network structure of CNN and RNN[4]. In 2018, Sang et al.,[5] introduced a bidirectional GRU framework connected with perceptual attention, which completes the emotional classification of text through sequence modeling and word properties recognition.

In 2019, Zhang et al.,[6] proposed a fine-grained LSTM-CNN classification model based on the attention mechanism. In 2020, Zhao et al.,[7] proposed a hybrid model which used CNN and Bi-LSTM to extract different properties of the data, and finally blend these properties into a SoftMax classifier to obtain classification results. In 2021, Gan [8] proposed a model of hybrid with attention mechanism. The model used Bi-LSTM to excerpt the language environment properties of the version, used graph perplexity neural network to capture the local properties of the version, then fused the two properties. After that, the attention mechanism model was added to deal with the situation that a specific aspect containing multiple words, to accurately obtain the language environment representation. In 2022, Zhou[9] came up with a Chinese text sentiment analysis method with BGRU. First, text information was converted into

a computer recognizable language, then BGRU obtained the input excerpt semantic news, finally, the emotional tendency of text was analyzed through the classifier.

Emotional analysis of text, also known as view mining, refers to the process of extracting, processing, summarizing, and analyzing views with personal feelings, and reasoning and judging the emotional trend of information. Due to the popularization of the Internet and the increasing number of users, massive information and complicated information resources have been generated. How to effectively, accurately, and quickly obtain the emotion contained in the user text from the massive comment information has become the focus and difficulty of the current text emotion classification task.

3. Proposed Work

In this paper, a three-way emotional analysis of the user's reviews in the article is carried out, and the problem of polysemy is effectively solved by using the dynamic word vector in BERT. Then, BERT is used to deal with this appearance of two-way text, to obtain the appearance of this text and add interest mechanism to weight important information to ensure more language environmental properties and more accurate information. The results show that the method can effectively upgrade the exactitude of text emotion classification.

3.1 Model of BERT

The BERT can obtain term vectors by calculating the information of vocabularies in different time series and spatial sequences, which is helpful to solve the problem of polysemy in text data. Therefore, this model adds the attention mechanism to BERT, to make version emotion classification more accurate. First, the version is processed using the advance training model, and the word vector, word vector and position vector of the text are acquired, and then the vector information is added. Next, the properties information of the text are accurately obtained by using advance training model with BERT, and then the properties news extracted by the model is weighted by introducing attention mechanism to highlight the key information. Finally, SoftMax is used to normalize information and complete version emotion classification task. The structure of the interpretable recommendation model based on BERT is shown in Fig. 1. By combining with the Transformer model, this model realizes the multi-level and two-way function of Transformer encoder, and effectively extracts the text properties of item comments.

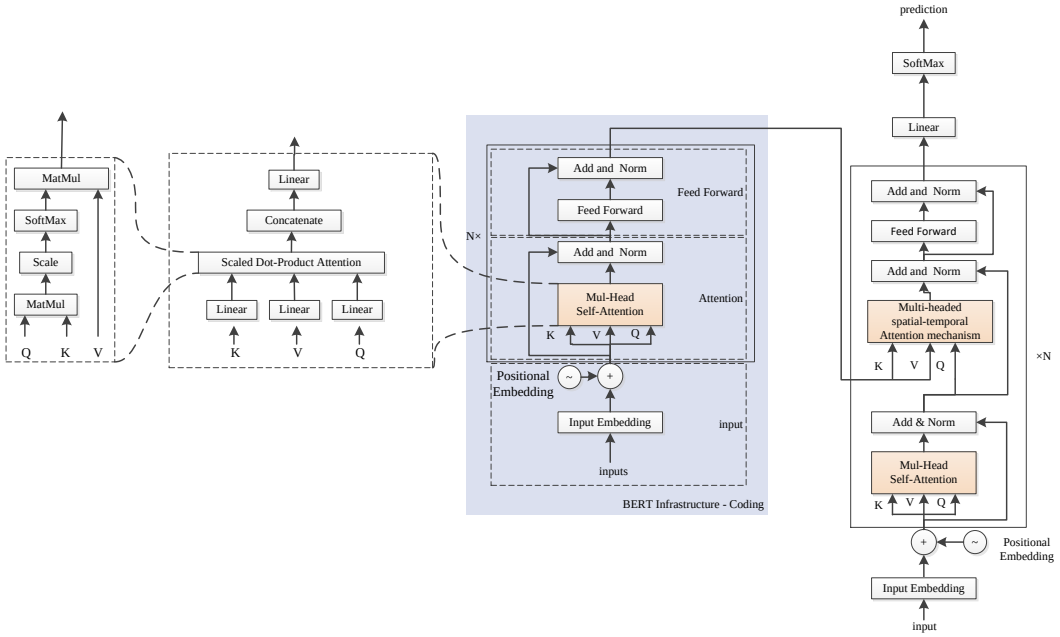

Figure 1. Explanation recommendation model based on BERT

The BERT advance training model extracts the version news of all layers from the top to bottom through the training of many corpora to realize the expression of version [10]. The vector manifestation of the ultimate input is gained by adding their conforming positions. This word vector obtained by summing the three dimensions is used as the input of the model. The model structure is shown in Fig. 2.

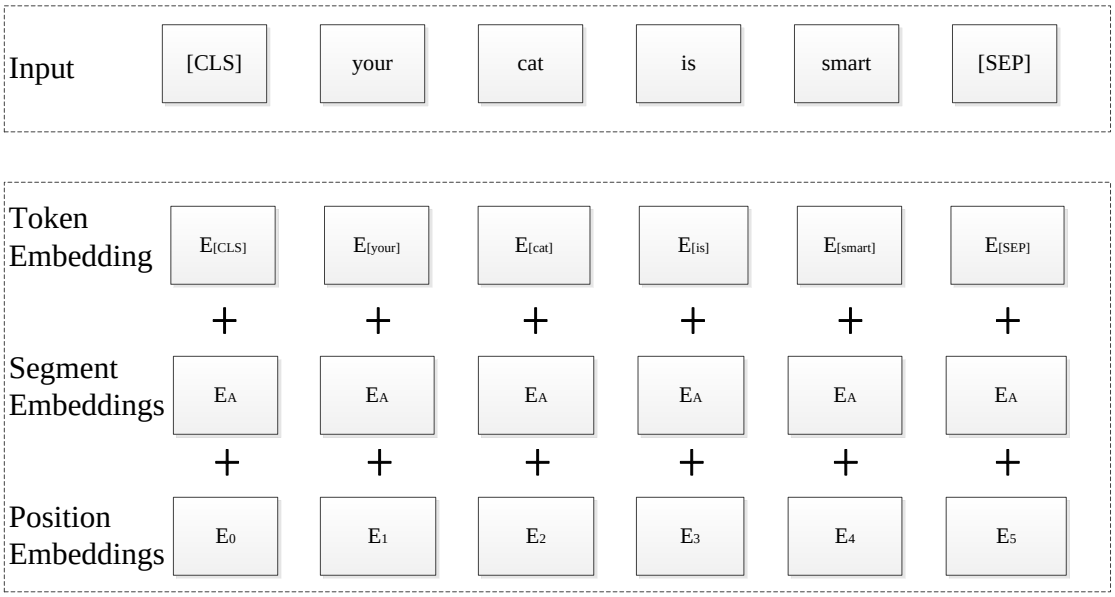

Figure 2. BERT advance training model input diagram

The addition process of word vectors in the model can be represented by the following methods: for example, the matrix dimension of word embedding is [4768], the matrix dimension of fragment embedding is [3768], and the matrix dimension of position embedding is [2768]; for a word, the word's unique hot code is [1,0,0,0], the location's unique hot code is [1,0.0], and the segment's unique hot code is [1,0]. The combined properties of the three unique hot codes is [1,0,0,0,0,0,0,0,1,0], and the word vector obtained by passing it through the full connection layer with the dimension of [4+3+2768]=[9768] is equivalent to the word vector obtained by adding the three matrix dimensions. Therefore, the BERT model uses the word vector obtained by summing the three dimensions as the input of the model.

The BERT model sums the three matrix dimensions. Its essence can be seen as properties fusion. It is effective for model training by learning the meaning of information of the fusion properties to the cart model and can upgrade the exactitude of the model. Since the BERT advance training model trains the text, the closer the distance between the texts, the greater the correlation; hence, it is very necessary to encode the position embedding. It mainly provides position information for the model through the linear change of sine and cosine functions. Formulas are shown in equations (1) and (2).

$$P_{E\ positive,2i)} = sin(positive / 10000^{2i/d_m}) \quad (1)$$

$$P_{E\ positive,2i+1)} = \cos(positive / 10000^{2i/d_m}) \quad (2)$$

The advance training process of BERT model mainly has two unsupervised tasks: Masked Language Model and Next Sentence Prediction.

(1) Masked Language Model

The traditional language model trains the target by predicting the probability of the next word, which has obvious effect in one-way coding. However, during bidirectional coding, it is possible that the word to be predicted may have appeared, which is meaningless for prediction. Therefore, masking language model is introduced in BERT to train bidirectional coding.

The masking language model was first proposed by Taylor in 1953. Unlike the traditional language model, which uses the given words to predict the next vocabulary, this masking language is similar in form to the cloze form commonly used in exams. By randomly blocking k% of these vocabularies in the sentence, then using the language environment to predict these words, its purpose is to train the ability of two-way representation of the model.

For example, in the training data, 15% of the token positions are randomly selected as the prediction information. If the token of the ith position is selected, it will be replaced according to the following strategy:

a) There is 80% possibility to use [MASK] token

b) There is 10% possibility to use random token

c) There is 10% possibility to keep the token unchanged

Subsequently, the output T will be used for the prediction of the original token.

(2) Next Sentence Prediction

In the emotion classification task, the text does not appear in the form of a single sentence, but often accompanied by a large amount of language environment information related to the sentence[11]. Therefore, it is very necessary to predict the information in the next sentence. In short, the prediction process is a two-class task. When a sample sentence A and a sample sentence B are selected as the training sentence pair, the probability of using IsNext tag to show that the sample sentence B is the real last sentence of sample sentence A which is 1/2, NotNext is used to indicate that example sentence B has a 1/2 probability of coming from a random sentence in the corpus.

3.2 Attention mechanism

The attention mechanism is developed in the image. It is a way to solve problems by simulating human attention. It focuses on key keys in large amounts of information, selects important information from large amounts of information and ignores unimportant information. After that, the attention mechanism is gradually applied in text processing and marks remarkable results in machine translation and version analysis [12].

Attention mechanism can be expressed as mapping data pairs composed of Q (Query), K (Key) and V (Value) from the same input vector information into the output. Specifically, it calculates the similarity between Q and K as the weight of V corresponds to K, and then weights the obtained weight with V. The calculation formula is shown in equation (3).

$$Attention(Q,K,V) = softmax(QK^T)\ V \tag{3}$$

Generally, there are other methods such as cos similarity, point multiplication method, Q and K series connection, multi-layer perceptron, etc. for calculating the similarity. The calculation formula is shown below.

$$\text{sim}(q,k) = \frac{q^T k}{\|q\| \cdot \|k\|} \tag{4}$$

$$s(q,k) = q^T k \tag{5}$$

$$s(q,k) = W[q;k] \tag{6}$$

$$s(q,k) = v_\alpha^T \tanh(W_q + U_k) \tag{7}$$

For the emotion analysis task, each word has an impact on the sentence, especially the Chinese breadth and depth. Each vocabulary has an impact on the emotion categorization, so it is necessary to grasp the key information and effectively obtain the emotion information. This model with the BERT advance training extracts the text properties through the introduction of attention mechanism and weights the acquired properties information.

4. Results and Discussion

4.1 Experimental environment

In the experiment, the Python version used is version 3.7, and the in-depth learning framework uses Python. The detailed software and hardware environment settings are shown in Table 1.

Table 1. The software and hardware settings

Experimental environment	Specific configuration
CPU	11th Gen Intel CoreTM i5-11400H @ 2.70GHz
GPU	NVIDIA GeForce RTX 3050 Ti Laptop GPU
Internal storage	16G
Python	3.8
Pytorch	1.7

4.2 Dataset

In this paper, four datasets are used from different fields, which are commonly used by Amazon in the recommendation field. The datasets include user's historical behavior data. The datasets are Automotive, Digital Music (referred to as Music), Cell phone and Accessories (referred to as Phone), Clothing, Shoes, and Jewelry (referred to as Clothing). Table 2 describes the datasets in detail, including the number of users, the number of items, the rating score (i.e., the user's rating or comment on the item) and the intensity of the datasets.

Table 2. Dataset

Name	Number of users	Number of items	Number of scores	Density
Automotive	3025	1725	21423	0.382
Music	5687	3457	65842	0.328
Phone	28936	12875	195463	0.068
Clothing	38267	25634	285689	0.032

From Figure 3, it can be determined that "negative emotion" is the least, and "neutral emotion" is the most, while "positive emotion" is between "negative emotion" and "neutral emotion". This kind of emotional data distribution will not have a significant impact on the exactness of the categorization of the recommendation system [13]. Therefore, it is significant to process the unbalanced news with this under-sampling method, so that each marked data will be kept in a relatively balanced state, and the impact on the training model of the recommendation system will be minimized.

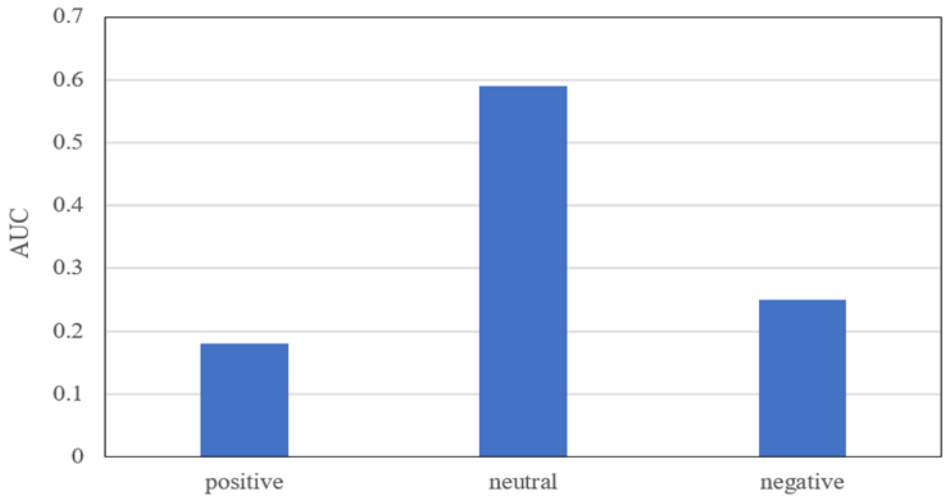

Figure 3. Comparison of emotional distribution of data

In the dataset, 80% of its data information is randomly selected as the training set of the model, 10% of its data information is selected as the test set of the model, and the remaining 10% of its data information is selected as the verification set of the model. Finally, the method of multiple tests on the model will use a 50% cross validation to effectively improve the stability of the model. In the data of different labels, the distribution of text length (as shown in Fig. 4) is not significantly different.

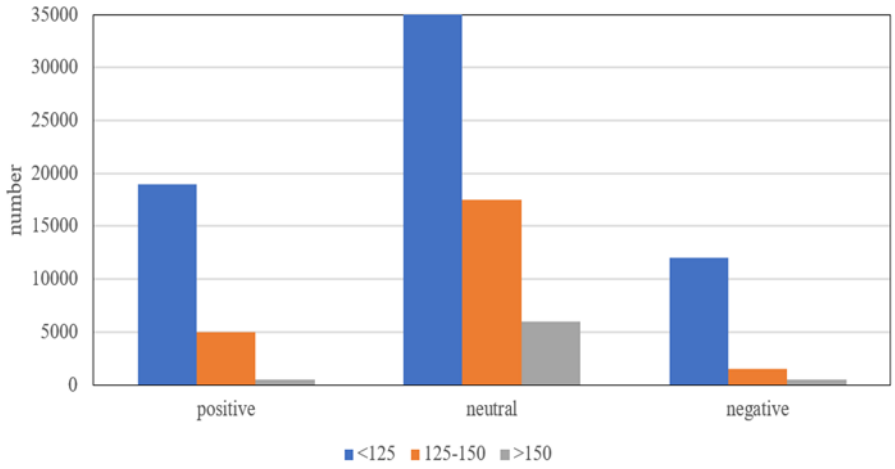

Figure 4. The comparative analysis chart of data length

Therefore, it can be found that the emotional distribution of the text is relatively uniform. To cover some of the data information and avoid the loss of lots of text properties information of the user's comments on the item, the sentence length is vectorized into 200 characters to make the properties information comprehensive during the training of the model. The selected dataset is analyzed according to the space-time sequence, as shown in Fig. 5. From the beginning of January to the end of January, the number of users' comments on items is small; in late January, the number of users' comments on items has sharply increased, and the distribution of emotional information in various texts changed with changes in different time and space.

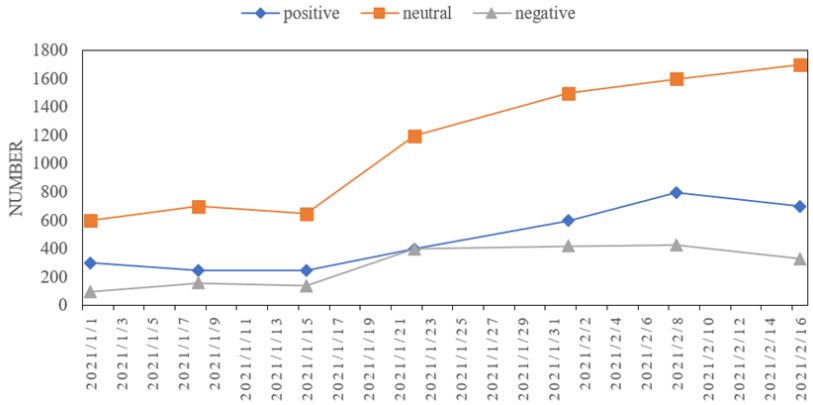

Figure 5. Distribution of emotional information in text with time and space

4.3 Evaluating Indicator

Text emotion analysis is a basic and important task in the field of natural language processing, which can be defined as short text classification to a certain extent. In this experiment, three different common evaluation indicators of recommendation system, namely accuracy P@P, recall rate R@P and average accuracy MAP@P are used.

(1) Accuracy P@P: It is the ratio of the predicted accurate items of P among the recommended top-P items, and is as shown in Equation (8):

$$P@P = \frac{TP}{TP + FP} \tag{8}$$

(2) Recall rate R@P: It is the ratio of accurately predicted items in the recommended top-P items to the number of positive sample items of all users in the test set, and it is as shown in Equation (9):

$$R@P = \frac{TP}{TP + FN} \tag{9}$$

(3) Average accuracy MAP@P: The ranking information are added to calculate the accuracy, and it is as shown in Equation (10):

$$MAP@P = \frac{TP + TN}{P + N} \tag{10}$$

where, TP (True Positives) represents the correctly retrieved positive items, FN (False Negatives) represents the incorrectly retrieved negative items, FP (False Positives) represents the incorrectly retrieved positive items, and TN (True Negatives) represents the correctly retrieved negative items.

4.4 Selection of comparison model

(1) Transformer[14]: This model is mainly based on the attention mechanism. The coding position is generally the position of the words to get the sequential data information of the comment text. At the same time, the model uses the two methods of attention mechanism and multi-head attention mechanism to mine the long-distance dependence information, to ensure that the information of each word can be obtained more comprehensively.

(2) textCNN[15]: This model uses convolution neural network to classify text. First, the user's comment text properties are extracted and processed through convolution layer and maximum pooling layer, and then the vector of text properties is obtained. Then, the activation function in the full connection layer is used to output it. Finally, the SoftMax function is used to classify text.

(3) Bi-LSTM[16]: Because LSTM can only predict from front to back in the sequence, in order to predict from front to back and from back to front, the model combines forward LSTM and backward LSTM to solve this problem and better mine the language environment information in the comment text.

(4) Bi-LSTM+Attention[17]: This model is based on the Bi-LSTM and adds the attention mechanism, so that the current text properties information is added with self-attention weight, and then collates and analyzes it, and then normalizes it through the SoftMax function, and finally outputs the matrix with attention weight through the full connection layer.

By comparing the above models, the proposed approach has the best effect, with obvious improvement in three indicators, reaching the best in the comparison model, and has the highest average accuracy, thus achieving the most stable performance. The results are shown in Table 3 and Fig. 6.

Table 3. Comparison of models

Name	Precision	Recall	Average accuracy
Transformer	0.8531	0.8215	0.8370
textCNN	0.8164	0.7708	0.7929
Bi-LSTM	0.8320	0.7838	0.8072
Bi-LSTM+Attention	0.8522	0.7923	0.8212
Proposed	0.8789	0.8286	0.8530

By comparing the proposed approach with textCNN, Bi-LSTM, Bi-LSTM+Attention and Transformer models, the precision of the suggested model has increased by 7.66%, 5.64%, 3.13% and 3.02% respectively. Meanwhile, comparing the average accuracy rate of models, it can be noticed that the average accuracy rate of the proposed technique reaches 0.8530, which proves that the model has strong stability.

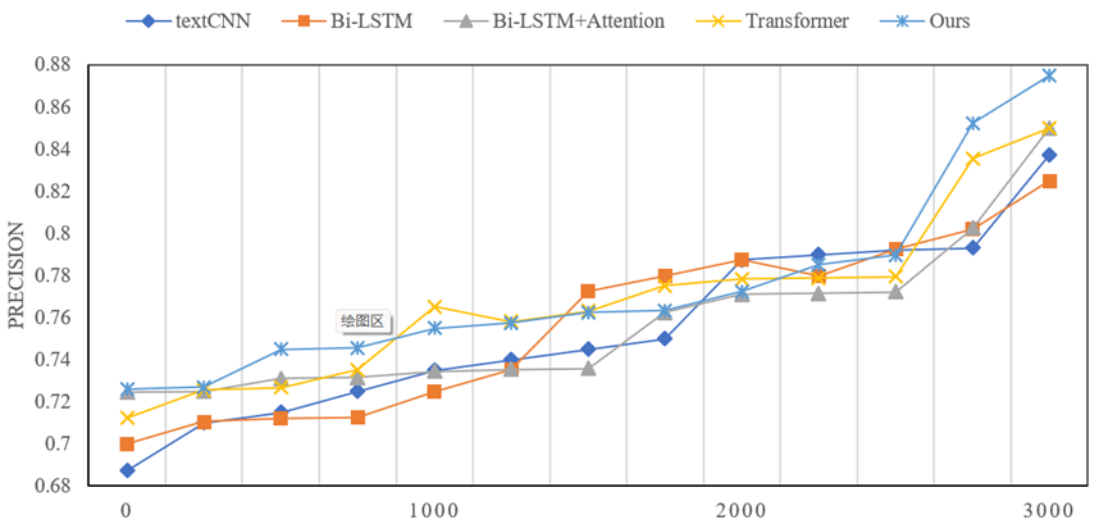

Figure 6. Precision comparison chart

The proposed approach uses the attention mechanism in the Transformer encoder to capture the negative and antonymous information in the text semantics.

5. Conclusion

The traditional emotion classification model cannot accurately understand the content expressed by the text, such as true words, polysemy, etc., resulting in the low accuracy of the version emotion categorization model and the weak generalization ability. In this paper, the three-element text emotion classification method with BERT has been proposed. First, the text is processed by using the BERT advance training model to obtain the word vector, word vector and position vector of the text, and then the vector information of the word is obtained by summing them. Then the information of the text properties is accurately obtained by using the BERT advance training, then the properties information extracted by this model is weighted by introducing attention mechanism to highlight the key information. Finally, the SoftMax function is used to normalize the information and complete the task of text emotion categorization. The availability of the proposed approach is proved by experimental comparison.

References

[1] China Internet Information Center. The 50th Statistical Report on China's Internet Development.https://baike.baidu.com/reference/8278359/93615UVtwE_6yiAuBi_S-JJwU71ft5yhzv5673Q_w1it41TuuO5iPXwLSA3o_ndA7689KuBbdS_6IxnK-kqzBZNkilOJno--a3AxaT4y

[2] Chung J, Gulcehre C, Cho K H. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[J]. Eprint Arxiv, 2015, arXiv.1412.3555.

[3] 刘高军,王小宾. 基于CNN+LSTM Attention的营销新闻文本分类[J]. 计算机技术与发展, 2020, 30(11): 59-63.

[4] Vaswani A, Shazeer N, Parmar N. Attention Is All You Need[J]. Eprint ArXiv, 2017, arXiv.1706.03762.

[5] Sang H, Chen Z, He D. Human Motion Prediction Based on Bidirectional-GRU and Attention Mechanism Model[J]. Journal of Computer-Aided Design and Computer Graphics, 2018, 31(7):1166-1174.

[6] Zhang W, Li L, Zhu Y. CNN-LSTM neural network model for fine-grained negative emotion computing in emergencies[J]. AEJ - Alexandria Engineering Journal, 2019, 61(9): 6755-6767.

[7] Zhao C, Huang X, Li Y. A Double-Channel Hybrid Deep Neural Network Based on CNN and BiLSTM for Remaining Useful Life Prediction[J]. Sensors, 2020, 20(24):7109-7124.

[8] Gan C, Feng Q, Zhang Z. Scalable multi-channel dilated CNN-BiLSTM model with attention mechanism for Chinese textual sentiment analysis[J]. Future Generation Computer Systems, 2021, 118(1–2): 297-309.

[9] Zhou Q, Zhou C, Wang X. Stock prediction based on bidirectional gated recurrent unit with convolutional neural network and properties selection[J]. PLoS ONE, 2022, 17(2): 0262501-0262521.

[10] Hu P, Dong L, Zhan Y. BERT Advance training Acceleration Algorithm Based on MASK Mechanism[J]. Journal of Physics: Conference Series, 2021, 2025(1):012038-012048.

[11] Song X, Petrak J, Roberts A. A Deep Neural Network Sentence Level Classification Method with Language environment Information[J]. Eprint Arxiv, 2018, arXiv:1809.00934.

[12] Zheng Z, Feng L, Liu R. Ultra-short-term Forecast of Multi-energy load for Integrated Energy System based on Attention Mechanism and BiLSTM[J]. IOP Publishing Ltd, 2022, 2271(01): 1742-1749.

[13] Liu X, Li T, Tang C. Emotion Recognition and Dynamic Functional Connectivity Analysis Based on EEG[J]. IEEE Access, 2019, PP(99):143293-143302.

[14] Wang Z, Wu J, Liu L. Analysis of current characteristics of transformer bias caused by subway stray current based on measured data[J]. 2021, 2087(01): 4325-4333.

[15] Lei T, Barzilay R, Jaakkola T. Molding CNNs for text: non-linear, non-consecutive convolutions[J]. Indiana University Mathematics Journal, 2015, 58(3): 1151-1186.

[16] Bollmann M, Sgaard A. Improving historical spelling normalization with bi-directional LSTMs and multi-task learning[J]. Eprint Arxiv, 2016, arXiv:1610.07844 .

[17] Mahato N K, Dong J, Song C. Electric Power System Transient Stability Assessment Based on Bi-LSTM Attention Mechanism[C]. 2021 6th Asia Conference on Power and Electrical Engineering (ACPEE). 2021: 777-783.