Journal of Computers Vol. 34 No. 6, December 2023, pp. 163-171
doi: 10.53106/199115992023123406013

Strategies for Monitoring and Managing Online Public Opinion in Universities Under the Background of Big Data

Ji-Cheng Yang, Xue-Meng Du*

Tangshan Polytechnic College,
Tangshan City 063299, Hebei Province, China
{jlcheng20180023, xuemeng0826}@126.com

Received 1 October 2023; Revised 1 November 2023; Accepted 13 December 2023

Abstract. In response to the issue of some keywords not being logged in or having inaccurate semantics in current public opinion monitoring, this article uses an improved word segmentation method to extract semantic features. In response to the current issue of unable to control the emotional direction of public opinion comments in public opinion analysis, an emotion analysis model Bi_GRU is proposed for sentiment analysis. Finally, using students' commonly used Weibo as a verification scenario, sensitive information such as "food safety" and "campus bullying" is screened to control the emotional direction of college students. The final proof is that the method proposed in this article can effectively supervise public opinion in a centralized environment and provide effective means for student management.

Keywords: public opinion monitoring, big data, sentiment analysis

1 Introduction

The rapid development of information technology has brought great changes to human society. The process of informatization and networking in human society is becoming increasingly rapid, and the characteristics of virtualization and data are becoming more and more obvious. The important role of online ideological and political education is becoming increasingly prominent. A large number of vivid samples of college students' online behavior are gathered in online public opinion, and the public opinion information reflects the spiritual style and ideological dynamics of contemporary college students, which is beneficial for exploring the ideological and behavioral patterns of college students in contemporary online society.

With the development of online media, a large amount of online public opinion data has been generated, including information with emotional tendencies in this article. Common university management issues, such as public health issues, campus violence issues, food safety issues, employment issues, student management issues, etc., if not promptly addressed, often lead to the emergence of online public opinion within the university. Students express their attitudes, emotions, viewpoints, etc. through Weibo, short videos, and other means. These viewpoints and opinions are constantly spreading and fermenting on the internet, becoming widely concerned hot events, and thus forming online public opinion related to universities.

The generation of public opinion follows a regular pattern, generally including the formation period, peak period, stable period, and regression period. Analyzing the data generated by these public opinions can grasp the development laws of public opinion, understand students' emotional tendencies, provide strong basis for school management and guidance, improve the ability to handle public opinion events, ultimately calm students' emotions, maintain the order of teaching on campus, and ensure stable and stable student management.

However, current universities have the following shortcomings in supervising and managing their own public opinion:

(1) Most of the existing public opinion monitoring methods use word segmentation libraries or custom dictionaries for text segmentation, but these word segmentation methods have problems such as inaccurate segmentation of unlisted words and ambiguity.

(2) At present, most theme based monitoring methods are based on sensitive vocabulary dictionaries, which can effectively monitor comments containing sensitive vocabulary, but cannot monitor public opinion from an emotional perspective on comment texts.

(3) Based on emotional analysis, most methods use a combination of emotion dictionaries and machine learn-

* Corresponding Author

ing. Although it can effectively identify emotional tendencies in public opinion information, it is difficult to make good judgments when encountering expressions and emerging network words.

Therefore, in response to the above situation, the work done in this article is as follows:

(1) In response to the problem that existing public opinion monitoring segmentation methods cannot accurately segment unlisted words and ambiguities, an improved BP neural network is proposed for text segmentation. This model uses deep learning methods to combine multiple neural networks to form an attention convolutional neural network conditional random field segmentation model.

(2) In response to the inability of topic based monitoring methods to analyze comments from an emotional perspective, a sentiment analysis model Bi_GRU is proposed that integrates aspect information for sentiment analysis. The output vector is fused with the vector with aspect information and input into the attention layer. The attention layer assigns different weights to different aspect information, and finally completes sentiment classification through the classification layer, The experiment proves that the model has good performance in emotion classification tasks.

(3) In response to the issue of insufficient utilization of facial expressions and traditional sentiment dictionaries in the monitoring method of sentiment analysis, this article focuses on the analysis of public opinion instances after establishing a public opinion monitoring model.

Therefore, the structure of this article is as follows: Chapter 2 mainly introduces the relevant research results, Chapter 3 mainly introduces the construction of the public opinion monitoring network structure, Chapter 4 introduces the process of establishing a model for analyzing the emotional direction of public opinion, Chapter 5 is the experimental verification process, and Chapter 6 is the conclusion part.

2 Related Work

Regarding the issue of sentiment analysis, Sharat Sachin attempted to research different deep learning technologies that have been applied to sentiment classification and analysis, and combined them with a recursive neural network structure, ultimately conducting validation experiments on the Amazon platform [1]. Hocheol Lee used a cross-sectional study method to analyze Twitter posts from South Korea and Japan from February 1, 2020 to April 30, 2020, in order to determine their impact on COVID-19. We collected Twitter data from major social media platforms in South Korea and Japan [2]. Ping Huang collected hot topics on campus for preprocessing and used the Word2vec model to generate word vectors. Then, she used convolutional neural networks to extract features from them and perform sentiment orientation classification. Through comparison of experimental data, the sentiment orientation classification based on convolutional neural networks (CNN) achieved an accuracy of 89.76% [3]. Ruidan Zhao utilized the Vector Space Model (VSM) algorithm in semantic analysis to further filter the collected webpage data, fully ensuring the quality of the collected public opinion data, and developed a public opinion monitoring system based on this method [4]. Chuanying Zhang, a customized system architecture based on Java language and open source software, filters out public opinion information related to the determined theme from complex content, and then processes and transforms the source data to achieve accurate and efficient monitoring of the social and public opinion of universities across the entire network platform [5]. Jundi Zhang, based on natural language processing algorithms, constructed a public opinion monitoring and warning model. Text features were extracted using the TF-IDF algorithm, and data was trained using a neural network model based on the path vector function to achieve the functions of public opinion analysis and warning. The experimental results demonstrated the efficiency of the method [6].

3 Construction of Network Detection Structure

The standard neural network may forget the old samples that have already been learned when learning new samples, and the noisy data will be adjusted to the network parameters after learning, resulting in unstable learning process of the network parameters. Therefore, it is proposed to improve the network structure of the standard neural network by adding an input adapter layer to the input layer of the neural network. Its function is to record the historical information of the input layer and delay the data of the input layer by one step. After adding an input adapter layer, when noisy data enters the network for learning, the previous historical input information will have a limited impact on the current noisy data, making the network's parameter learning more stable and improving the performance of the neural network. The improved neural network structure is shown in Fig. 1.

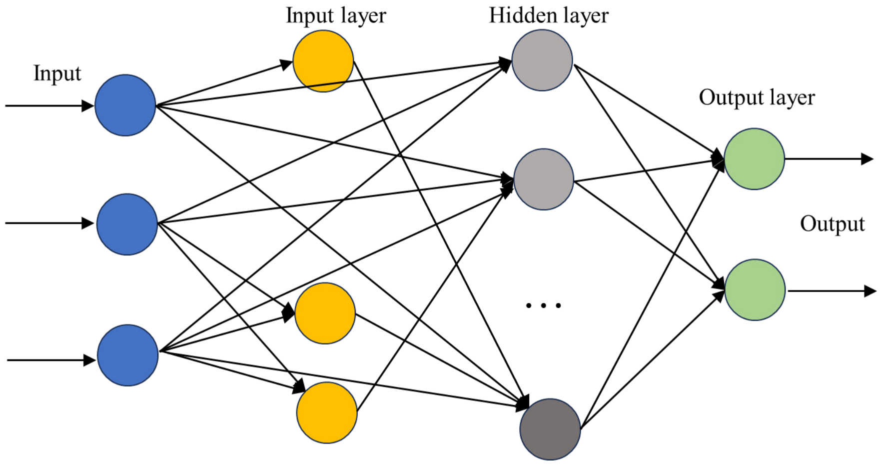

Fig. 1. Improved neural network structure

The state space expression of the improved BP neural network [7] is:

$$u^{(k)} = f\left[\beta\omega_1 m^{t-1} + (1-\beta)\omega_3 n^{t-1}\right]. \tag{1}$$

$$n^t = m^{t-1}. \tag{2}$$

$$v^t = g\cdot(\omega_2 u^k). \tag{3}$$

Among them, m^{t-1} is the input sample data, u^k is the input of the hidden layer, n^t is the data of the input layer, v^t is the output result, ω_1 is the weight connecting the input layer and the hidden layer, ω_2 is the weight connecting the hidden layer and the output layer, ω_3 is the weight connecting the input layer and the hidden layer, and β is the balance factor between the input layer and the input layer. Function I is the transfer function of the hidden layer using a linear transfer function, while function $f(x)$ is the transfer function of the output layer using a logarithmic transfer function.

Optimizing the topology structure of a neural network mainly involves obtaining the number of hidden layers and the number of neurons on each hidden layer. Binary encoding is performed on the topology structure of the neural network, assuming that the number of hidden layers is or. The form of binary encoding is shown in Fig. 2.

L	l_1	l_2
$\alpha_1\alpha_2$	$\alpha_{11}\alpha_{12}\cdots\alpha_{1j}$	$\alpha_{21}\alpha_{22}\cdots\alpha_{2j}$

Fig. 2. Principle of binary encoding

The binary encoding form is divided into three parts, and the code string L represents the number of hidden layers, where the value of α is 1 or 0. The method for representing the number of hidden layers is as follows:

$$L = \begin{cases} 1 & \alpha_1\alpha_2 = 0 \\ 2 & \alpha_1\alpha_2 \neq 0 \end{cases}. \tag{4}$$

The code string l_1 represents the number of neurons in the first hidden layer, while the code string represents the number of neurons in the second hidden layer. There is a calculation relationship between them as follows:

$$\begin{cases} l_1 = 2^0\alpha_{11} + 2^1\alpha_{12} + \cdots + 2^{j-1}\alpha_{1j} \\ l_2 = 2^0\alpha_{21} + 2^1\alpha_{22} + \cdots 2^{j-1}\alpha_{2j} \end{cases}. \tag{5}$$

Among them, j represents the length encoded by the number of hidden layer nodes.

Steps to optimize the network structure of a neural network [8] using genetic algorithms:

(1) Set the number of hidden layers in the neural network and the range of the number of neurons in each layer. Binary encode the number of hidden layers, neurons in the first and second layers, and randomly generate N chromosomes with the same encoding. Encode N chromosome into a corresponding neural network

(2) Set different initial connection weights to learn training steps (1) to form a network

(3) Calculate the fitness of each individual in the initial state, and the fitness function is the error function of the neural network

(4) Select individuals with high fitness values as parents and perform genetic operations

(5) Using crossover and mutation operations in genetic algorithms to deal with contemporary populations and generate new ones.

(6) Repeat steps (2) - (5) until an individual in the population can meet the end condition, and the resulting individual is approximately the optimal solution of the neural network structure. From this, determine the number of hidden layers in the neural network and the number of neurons on each hidden layer.

4 Establishment of a Sentiment Analysis Model for College Student Social Platform Data

This article uses an emotion analysis model that integrates aspect information. *GRU* network is used instead of *LSTM* to learn text features, and attention mechanism is used to assign weight to aspect information to increase its importance. The feature input of aspect information weight is learned into the classification layer, and finally, sentiment analysis is achieved through *soft*max classifier classification. The flowchart is shown in Fig. 3:

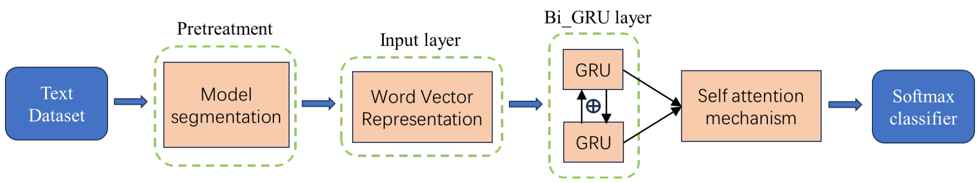

Fig. 3. Model flowchart

The preprocessing part is described in Chapter 3, which mainly introduces the input layer, *Bi_GRU* layer, attention layer, and classification layer.

4.1 Input layer

This layer uses the Word2vec model to train the data to obtain a word vector matrix $X, X \in R^{d\times l}$, where the text is represented as $V_i = \{v_{i1}, v_{i2}, ..., v_{il}\}$, where d is the word vector dimension, i is the i -th text in the training intonation, l is the maximum number of valid words in the sentence, the training set $M = \{V_1, V_2, ..., V_n\}$, n is the size of the training corpus, and $y = \{y_1, y_2, ..., y_n\}$ is the corresponding training label. The input layer updates the parameters of the V_i input *Bi_GRU* obtained from the training set N training.

4.2 Bi_GRU Layer

Unit *GRU* consists of an update gate and a reset gate, where the update gate is formed by merging the forget gate and input gate of Unit *LSTM*, controlling the degree to which the state information at time $t-1$ is brought into the current state. Specifically, the larger the value, the more state information is brought into the previous state. The reset gate controls how much state information at time $t-1$ is written into the current state candidate set, and the smaller the value, the less previous state information is written [9]. The unidirectional *GRU*-unit ignores contextual information, while adding a negative hidden state in *Bi_GRU* enables full learning of reverse semantics, allowing the *Bi_GRU* -layer to learn complete contextual information and have better performance in sentiment analysis. At time *t*, the output of the *Bi_GRU*-layer is $h_t^{'}$, and the expression is:

$$h_t^{'} = \left[\overrightarrow{h_t}, \overleftarrow{h_t}\right]. \tag{6}$$

4.3 Attention layer

The attention mechanism is used to assign different weights to different parts to reflect the different contributions of different aspect words to emotional tendencies. The output of layer *Bi_GRU* is $h_t^{'}$, which serves as the input for that layer. Then, the hidden vector $h_t^{'}$ is passed through a multi-layer perceptron to obtain a new hidden vector $u_t^{'}$. The new hidden vector $u_t^{'}$ and the context vector $u_v^{'}$ are calculated to obtain the weight value a_t . $u_v^{'}$ is a high-dimensional vector used to determine the importance of words in sentences. The expression is as follows:

$$u_t^{'} = \tanh(W_w h_t^{'} + b_w). \tag{7}$$

$$a_t = \frac{\exp(u_t^{T} u_v^{'})}{\sum_t \exp(u_t^{T} u_v^{'})}. \tag{8}$$

$$C^{'} = \sum_t a_t h_t^{'}. \tag{9}$$

4.4 Classification Layer

The classification layer inputs the features captured by the attention layer into the *soft*max classifier for classification, and the activation function uses the *soft*max function [10]. The calculation formula is as follows:

$$\tilde{y} = soft\max\left(W_C C + b_C\right). \tag{10}$$

In the formula, W_C is the weight matrix, b_C is the bias, and $\tilde{y}$ is the classification result. The loss function used in this chapter's model is the cross entropy loss function with the L_2 regularization term, which is defined as follows:

$$L = -\frac{1}{N}\sum_{r=1}^{m}\sum_{q=1}^{C} y_{rq} \log\left(\tilde{y}_{rq}\right) + \lambda\left\|\theta\right\|_2. \tag{11}$$

Among them, y and $\tilde{y}$ are the actual label values compared to the predicted label values, C is the number of label categories, and λ is the regularization coefficient of L_2. The improved model flowchart is shown in Fig. 4.

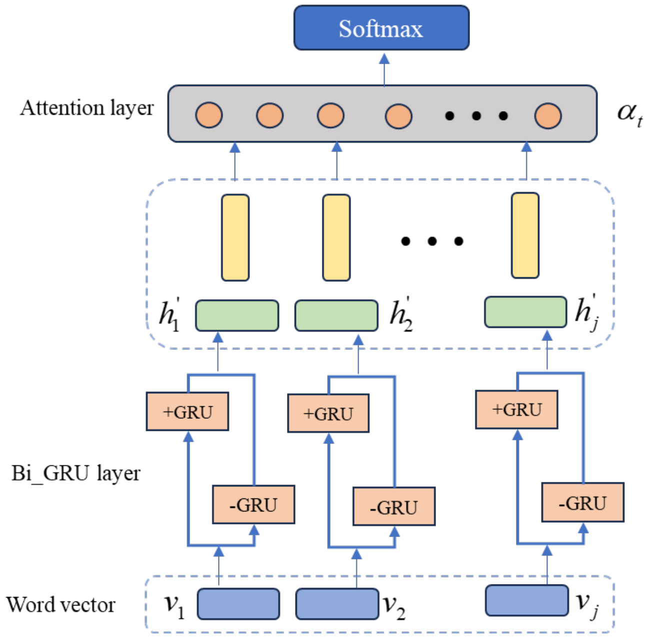

Fig. 4. Model *Bi_GRU* framework diagram

5 Experimental Results and Analysis

In order to verify the feasibility of the above model and emotion classification, based on selecting the same dataset and objective evaluation indicators, three sets of parameter comparison experiments were set up to select the optimal parameters. The emotion classification model in this article was used to classify the results and compare them.

5.1 Experimental Environment

The computer is installed on the PC side of the Win10 system, with an Intel Core i7-8750H processor, NVIDIA's 1080Ti GPU, 8GB CPU memory, Google's Colab compilation software, Python 3.0 compilation language, and Keras deep learning framework.

5.2 Dataset Selection

To ensure the objectivity of the results, the dataset for sentiment analysis tasks needs to be sourced from representative websites or social media platforms. Weibo is a commonly used textual social platform for college students, constantly reflecting their life dynamics. Simplifyweibo_4_moods dataset is a standard dataset based on the Weibo platform, which contains over 360000 Weibo data with 4 types of emotional labels. Firstly, integrate the data labeled with disgust, anger, and low emotions in the dataset, and then remove or reprocess the data with certain ambiguities to label these negative emotions as negative emotion data. Similarly, ambiguity processing is also performed on data labeled as joyful emotion labels in the dataset, and these data containing positive emotions are labeled as positive emotion data. Among them, positive emotional data is labeled as 1, and negative emotional data is labeled as 0.

Four commonly used model evaluation indicators were selected: Accuracy expressed as A, precision expressed as P, Recall expressed as R, and F_1 value as the evaluation criteria for this experiment. Before introducing each evaluation indicator, first define some representation methods. The representation methods for each evaluation indicator are as follows:

$$A = \frac{TP + TN}{TP + FP + TN + FN}. \tag{12}$$

$$P = \frac{TP}{TP + FP}. \tag{13}$$

$$R = \frac{TP}{TP + FN}. \tag{14}$$

$$F_1 = 2 \times \frac{PR}{P + R} = \frac{2TP}{2TP + FP + FN}. \tag{15}$$

5.3 Parameter Settings

Firstly, the model iteration count is set to 15, the optimizer uses Adam, and the learning rate is set to 0.001. Due to the insufficient sample size for learning, the dropout is set to 0.2, batch_ Set the size to 64 and the L2 regularization coefficient to 0.0001. Train the opinions expressed on employment issues, student management, etc. based on the above parameter settings, and the training results are shown in Fig. 5.

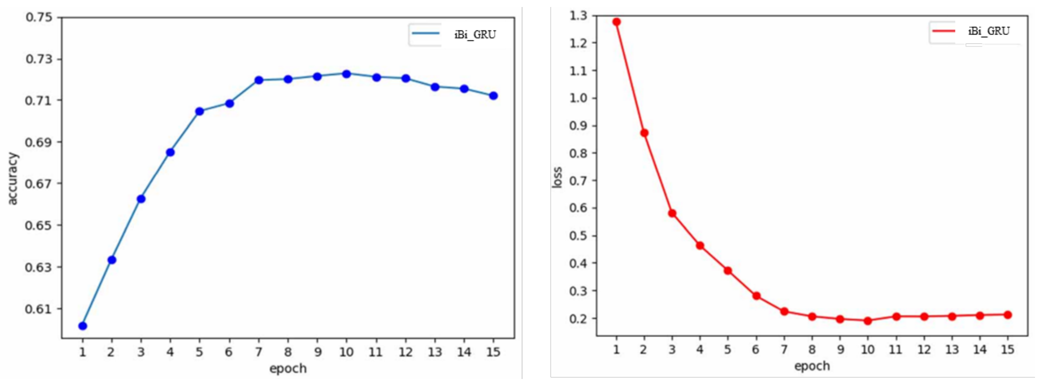

Fig. 5. Model training results

For the convenience of describing, this article describes the improved model as iBi_ GRU, as shown in the figure, the accuracy of the model gradually stabilizes when the number of iterations reaches 8 or more. At 10, the classification accuracy of the model is the highest, while the accuracy of the model decreases as the number of iterations continues to increase. When the number of iterations is 10, the loss rate of the model reaches its lowest, and as the number of iterations continues to increase, the loss rate of the model increases. The goal of setting epoch parameters is to expect the model to have high accuracy and low loss rate in classification tasks. Based on the two results, the overall performance of the model is best when epoch is 10.In order to demonstrate the effectiveness of the improved algorithm proposed in this article, we selected and compared it, and the comparison results are shown in Fig. 6.

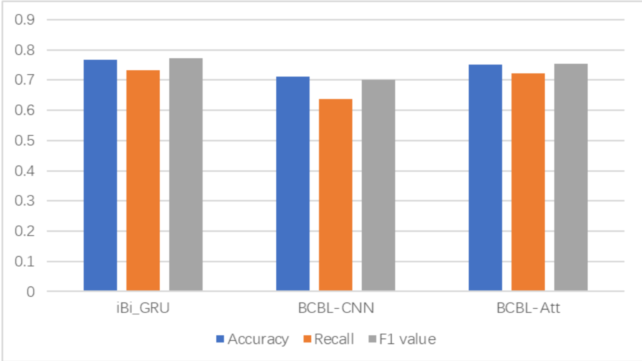

Fig. 6. Model comparison

The accuracy rate obtained by using iBi_GRU for sentiment classification under the standard dataset is 0.7681, the recall rate is 0.7329, and the F1 value is 0.7727. The accuracy, recall, and F1 values of the BCBL-CNN model with the best neutral performance in the comparison model were 0.7124, 0.6375, and 0.7004, respectively, under the standard dataset. The results of the three evaluation indicators of BCBL-CNN were all lower than iBi_GRU. It can be seen that the improved iBi_GRU sentiment classification model has improved compared to the previous model in terms of accuracy, recall, and F1 value. The experimental results validate the effectiveness of the improved A model by introducing attention mechanism, and due to the dynamic adjustment of the weights of vocabulary containing different emotional levels by introducing attention mechanism, the BCBL-Att model performs better than the BCBL-CNN model in text sentiment classification tasks on standard datasets.

6 Conclusion

After analysis, the speech recognition model proposed in this article can accurately recognize public opinion vocabulary in the network, and can also make judgments on emotional tendencies. The effectiveness of the algorithm has been proven in simulation experiments. At the same time, this article also has shortcomings:

1) There are shortcomings in public opinion monitoring for short video platforms, and it is also the main direction of future research;

2) The control of emotions in public opinion comments is not precise enough.

References

[1] S. Sachin, A. Tripathi, N. Mahajan, S. Aggarwal, P. Nagrath, Sentiment Analysis Using Gated Recurrent Neural Networks, SN Computer Science 1(2)(2020) 74.

[2] H. Lee, E.-B. Noh, S.-H. Choi, B. Zhao, E.-W. Nam, Determining Public Opinion of the COVID-19 Pandemic in South Korea and Japan: Social Network Mining on Twitter, Healthcare informatics research 26(4)(2020) 335-343.

[3] P. Huang, H.-J. Zhu, L.-L. Chen, Application of Emotion Classification Technology based on Deep Learning in University Public Opinion Analysis, Software Engineering 24(11)(2021) 59-62.

[4] R.-D. Zhao, X. Zhu, Design of online public opinion collection system based on crawler technology and semantic analysis, Electronic Design Engineering 29(14)(2021) 56-60.

[5] C.-Y. Zhang, S.-Y. Wang, D. Dong, College Network Public Opinion Monitoring System Based on Distributed Data Acqcuisition and Natural Language Processing, China-Arab States Science and Technology (3)(2021) 138-140.

[6] J.-D. Zhang, Research on public opinion monitoring and warning model based on natural language processing and intelligent semantic recognition, Electronic Design Engineering 30(17)(2022) 165-169.

[7] B. Zhang, Application of BP Neural Network Based on Wavelet Denoising in Deformation Monitoring, Beijing Surveying and Mapping 35(12)(2021) 1592-1596.

[8] Z. Wu, H.-L. Wu, UAV route planning based on the improved genetic algorithm, Electronic Measurement Technology 44(24)(2021) 52-58.

[9] Z. Li, B.-Y. Yang, Y. Li, X.-L. Li, Multi-tag Disaster Information Prediction Based on ALBERT and Bidirectional GRU, Science Technology and Engineering 21(35)(2021) 15284-15289.
[10] P. Zhang, B.-W. Sun, W.-S. Li, J.-F. Xu, Y.-W. Sun, Phishing Mail Detection System Based on LSTM Neural Network, Transactions of Beijing Institute of Technology 40(12)(2020) 1289-1294.