TBLC-rAttention: A Deep Neural Network Model for Recognizing the Emotional Tendency of Chinese Medical Comment

QIBING JIN\textsuperscript{1}, XINGRONG XUE\textsuperscript{1}, WENJUAN PENG\textsuperscript{2,3}, WU CAI\textsuperscript{1}, YUMING ZHANG\textsuperscript{1}, AND LING ZHANG\textsuperscript{2,3}

\textsuperscript{1}Institute of Automation, Beijing University of Chemical Technology, Beijing 100029, China
\textsuperscript{2}Department of Epidemiology and Health Statistics, School of Public Health, Capital Medical University, Beijing 100069, China
\textsuperscript{3}Beijing Municipal Key Laboratory of Clinical Epidemiology, Beijing 100069, China

Corresponding author: Xingrong Xue (1976388602@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61673004 and Grant 61273132, and in part by the Fundamental Research Funds for the Central Universities of China under Grant XK1802-4.

ABSTRACT In the current paper, a hybrid depth neural network model, TBLC-rAttention, aiming at Chinese text emotion recognition, is proposed to identify the emotional tendency of the Chinese medical reviews. The model includes the following steps: acquiring and preprocessing the Chinese corpus; mapping the preprocessed text into the word vectors; using Bi-directional Long Short-Term Memory network (Bi-LSTM) with the attention mechanism to acquire the context semantic features of the text; using Convolutional Neural Network (CNN) to obtain local semantics features on the basis of the context semantic features; and inputting the final feature vectors into the classification layer to complete the task of emotion recognition and the classification of the Chinese medical reviews. In this experiment, the corpus data is the comments of 999 cold medicine on a large e-commerce platform. All corpus are divided into three types, including high praise, medium praise and bad review. Classical machine learning models (SVM, NB) and neural network models (CNN, LSTM, Bi-LSTM, BiLSTM-Attention and RCNN) are performed as the comparison benchmarks to assess the category performance of TBLC-rAttention model. All the results were obtained when the training accuracy and test accuracy were stable after 1000 cycles of repeated calculation. The results show that TBLC-rAttention can get better text feature than the reference models, and the text classification accuracy reaches to 99%. In conclusion, the TBLC-rAttention model can identify semantic feature information to the greatest extent. In addition, this study also completes the numerical quantification of the predicted results.

INDEX TERMS Attention mechanism, bi-directional long short-term memory network (Bi-LSTM), Chinese medical comment, Chinese text emotion recognition, convolutional neural network (CNN), deep learning, feature extraction, hybrid neural network, naive bayes (NB), natural language processing (NLP), support vector machine (SVM).

I. INTRODUCTION
In recent years, the Internet has become an important way for people to obtain information in their daily life. Statistics released by the China Internet Network Information Center (CNNIC) show that by June 2019, China’s Internet penetration rate had reached to 61.2%, with 854 million Internet users and 660 million online shoppers. The textual data generated by the network is increasing at an alarming rate, so how to effectively organize and utilize the textual data has become a problem that people can pay more and more attention to. The text classification technique of natural language processing (NLP) is an effective solution [1], which is a process of training classifiers using text data and then classifying new text based on the trained classifiers.

Text emotion recognition, also known as sentiment analysis [2], referred to making judgments on the opinions of the text through specific methods according to the meaning and the emotional information expressed by the text, and then dividing it into different categories of emotions [3]. So, it can be regarded as a classification task, which is one of the core research directions of NLP field and the core content...
of the network public opinion identification. Although traditional text emotion recognition methods have achieved good results in the acquisition, analysis and guidance of the network public opinion, problems such as time-consuming, low accuracy, poor timeliness and unstable performance also appeared in the processing of the massive network public opinion data with complex emotion characteristics [4]. However, the emerging of deep learning provides an effective way to solve these problems, which can better capture the hidden features of data and complete the task of emotional classification.

Recently, the amount of medical data has been increasing exponentially, and there is a mass of information contained in the online comments related to the medicine. However, few studies have applied the deep learning-based text emotion recognition technology to medical reviews. In this paper, an emotion recognition scheme, hybrid neural network (TBLC-rAttention) model, which based on deep learning method is proposed for Chinese text, and was verified in the corpus of Chinese. The corpus is consumers’ comments from e-commerce platforms after purchasing and using the cold medicine. This study utilized this model to identify consumers’ emotional tendency and aimed to further promote the application of NLP technology in the Chinese medical field [5]–[7].

The structure of this paper is as follows: Section 1 is the introduction. Section 2 is the related work, which briefly introduces the current typical methods of the emotion recognition and their shortcomings, points out the advantages of deep learning methods. Section 3 introduces in detail the Chinese medical text emotion recognition model proposed in this paper. Section 4 is the experimental part. Section 5 introduces and discusses the experiment results. Section 6 describes the conclusions and the future work.

II. RELATIVE WORK

The opinions expressed by netizens contain their own subjective emotional tendencies, which affect the trend of online public opinions and thus affect people’s lives [8]–[11]. Text emotion recognition technology provides the opportunity to mine hidden information from different angles. Consumers decide whether to buy a drug based on the others’ online comments. Drug developers can further improve drugs by analyzing consumer reviews of drug efficacy [12]–[14]. It indicates that the analysis of emotional tendency of the network comments plays an important role in public opinion guidance, marketing strategy formulation, government opinion survey and other aspects.

Generally, there are two mainstream ways to mine text emotion. One is machine learning-based method [15]–[17], the other is dictionary-based method [18], [19]. The dictionary-based method is to analyze the text emotion by contrast according to the established dictionary, so as to obtain the emotional polarity. This method has a relatively high accuracy, but relies heavily on the emotional dictionaries and costs considerable efforts on the dictionary construction. Moreover, it only calculated the positive or negative polarity values of the words [20], so this method not only fails to take the context meaning into account, but also is too simple to deal with the strength of emotions. The machine learning-based method is mainly to extract text features and classify emotions using classification algorithm. It firstly converts the text into digital information that computers can process, then uses artificial intelligence algorithms to train and categorize text emotion, and achieve better effect results than the dictionary-based approach. However, with the rapid increase of the network data, this method presents some problems such as text represents high dimension, high sparsity, weak feature expression ability and degraded classification performance [21], mainly because its performance is highly dependent on the extraction of classification features. Moreover, the massive text information and the emerging vocabulary on the Internet make it more difficult for traditional machine learning-based methods to fully extract accurate text features. However, with the development of artificial intelligence technology, deep learning provides a feasible solution to solve these problems.

Deep learning was first proposed by Hinton and other scholars in 2006 [22], which originated from the study of the neural networks. Traditional neural network learning algorithms, which are generally composed of single input layer, hidden layer, output layer and other low-level network structures, the features obtained after learning are monolayer. Deep learning adopted neural network learning algorithm, usually includes one input layer, several hidden layer and one output layer. It has high nonlinear operations level, thus can be regarded as a simulation of the nerve connections in the human brain structure. By transforming the original features layer by layer, the new feature space and the hierarchical features are obtained, and the classification is then realized. Therefore, the deep neural network can better capture the hidden features of corpus, eliminate useless information but retain useful information for classification task, and finally obtain better model performance. So far, deep learning has been widely studied in academia and applied in image classification, speech recognition and other directions [23]–[28]. It also been applied in NLP tasks such as automatic translation [29]–[31].

Many text emotion recognition methods based on neural network models emerged in recent years because of the rapid development of in-depth learning [32], [33]. Recurrent Neural Network (RNN) is good at capturing text information in time-series data [34]. However, it is time-consuming in processing data because of the complexity of the network structure caused by the recursive loop property of RNN itself. In addition, RNN also has the problem of gradient explosion and gradient disappearance [35]. Schuster et al. proposed a variant of RNN named Bi-directional Long Short-Term Memory network (Bi-LSTM) [36], which can not only process longer sequence information, but also capture the context information better through bi-directional structure. Although Bi-LSTM has alleviated the problem of gradient
explosion and the gradient disappearance, it makes calculated amount increase further. The Convolutional Neural Networks (CNN) [37] is characterized by time-saving because of sparse connection and parameter sharing. However, long sequence information cannot be obtained due to the fixed convolution kernel. Later, Kalchbrenner et al. used wide convolution instead of narrow convolution and adopted K-max pooling to solve the problem of sequence length limitation [38]. In a word, CNN and RNN show their respective advantages and disadvantages with their own network structure characteristics in the different tasks. Therefore, researchers put forward the TextRCNN model [39], which combines the advantages of CNN and RNN. Although the accuracy has only increased by 1%, the design idea of the model is worth learning. At present, Attention mechanism can distinguish the different importance of each word in a sentence when categorizing. The introduction of attention mechanism strengthens the ability of the neural network to mine data information, which opened its application in the NLP field [40]–[43].

People are paying more and more attention to health, previous study has found that more than 81 percent of the Internet users in the United States search for health information online [2], which indicated that mining the hidden information in the comments of these social network sites has become an urgent problem. However, NLP technology has not been widely applied in the Chinese medical field. One of the reasons is because most of the research results of emotion recognition study in medical field have had been written in English. The second reason is because although medical text contains a lot of valuable information, more difficult to handle. Medical text contains a large amount of unstructured free text and image information. The doctors themselves input the text may lead to spelling mistakes. Abbreviations for medical terms and idioms used by different doctors in different regions can lead to the obtained information cannot be effectively recognized by the computers. Therefore, deep learning and NLP may play an important role in the analysis and mining of medical texts. This study uses a hybrid deep neural network based on supervised learning to realize text emotion recognition in the field of Chinese medicine.

III. METHOD

Using existing information to better understand corpus and find out the effective features is the core problem in the task of text emotion recognition. This study proposed a hybrid deep neural network model, TBLC-rAttention, aiming at Chinese text emotion recognition (Fig. 1). The model combines Bi-LSTM and CNN with attention mechanism. In a gradual way, it first obtains contextual semantic features, then extracts the local semantic features, at the same time give feature information at each moment \( r \) different weights, thus get better text feature representation and further improve the accuracy of the text classification.

The model consists of seven parts: (1) Input layer: obtaining text data. (2) Preprocessing layer: word segmentation and removal of the irrelevant data. (3) Embedding layer: mapping text data into word vectors. (4) Bi-LSTM layer: extracting global semantic features of text data. (5) Attention mechanism: generating weighted global semantic feature with probability distribution of attention. (6) CNN layer: extracting local semantic features based on weighted global semantic features. (7) Output layer: achieving text classification. Fig. 2 is the process Chinese text emotion recognition using TBLC-rAttention model, and steps are summarized as follows:

1) Obtaining text data.
2) Preprocessing text data.
3) Mapping the pre-processed text data into word vectors.
4) Building a hybrid neural network text emotion recognition model.
5) Establishing the cost function and training the text emotion recognition model using the random gradient descent method.
6) Calculate the emotional polarity value of the text.
7) Validating the effect of the model based on validation set text data.
8) Application of the TBLC-rAttention model.
TABLE 1. Partial medical terms used in the word segmentation procedure.

| Chinese   | English       | Chinese   | English       | Chinese   | English     |
|-----------|---------------|-----------|---------------|-----------|-------------|
| 青霉素     | Benzylpenicillin | 匹美西林   | Pimecillin    | 头孢噻肟   | Ceftizoxime |
| 甲氧西林   | Meticillin    | 亚美西林   | Azimicillin   | 头孢拉定   | Latamoxef   |
| 萘夫西林   | Naficillin    | 茜美西林   | Ternicillin   | 头孢匹罗   | Cefpirime   |
| 两味西林   | Oxacicillin   | 双美西林   | Azodicillin   | 头孢他肟   | Cefepime    |
| 乳酸西林   | Cloxacillin   | 乳酸西林   | Pheneticillin | 头孢噻肟   | Cefuzonam   |
| 双氯西林   | Dichloxacillin | 丙氯西林   | Propicillin   | 头孢甲肟   | Cefmenoxime |
| 氯氨西林   | Flucloxacillin | 吩布西林   | Furubicillin  | 头孢拉宗   | Cefuberazone |
| 氯苯西林   | Biflocillin   | 吡阿西林   | Aplicillin    | 头孢匹胺   | Cefpiramide |
| 氯唑西林   | Ampicillin    | 仑氨西林   | Lenampicillin | 头孢替坦   | Cefotetan   |

TABLE 2. Partial stop words used in the participle process.

| Partial stop words |
|--------------------|

A. PREPROCESSING LAYER
The obtained text data is preprocessed in the following way:
1. Cleaning the text data, including deleting irrelevant data and duplicate data, processing outliers and missing data, such as HTML web page tags, punctuation marks, special emoticons, etc. Thus, the information irrelevant to classification is preliminarily filtered out.
2. Categorizing and labeling the text data.
3. Using jieba to segment words, during which the user dictionary is loaded. Some medical terms are added to the dictionary, as shown in Table 1. In order to improve the accuracy of the word segmentation, we use the fastText neology-finding algorithm to identify the unregistered words and add them to the dictionary. Then, the emotion polarity of the text is determined by the emotion extremum calculation algorithm used, and then it is added to the semantic emotion dictionary. In addition, some common stop words are removed in the process, as shown in Table 2. In order to make the model suitable for long text data, we used TextRank to remove the stop words.
algorithm to remove irrelevant data from the text that ends in the segmentation.

4) Dividing the pre-processed data into three parts: training set, test set and verification set.

**B. EMBEDDING LAYER**

The premise of using NLP to process text is to transform text data into a vectorized form that can be recognized and processed by computer [44]. The text vectorization mapping process is shown in Fig. 3.

By mapping the word embedding matrix $E_w$, the batch text data with labels is mapped to a three-dimensional word vector matrix $M$, which contains two parts: comment content $D$ and label content $L$. The word embedding matrix can be obtained by Word2Vec or NNLM method. Then, a text containing $n$ words $D_j = \{x_1, x_2, \ldots, x_n\}$ can be expressed in the following form:

$$e_i = b_ix_i E_w$$

$$M = (e_1, e_2, \ldots, e_1, \ldots, e_n)$$

where $M \in \mathbb{R}^{\text{batch} \times n \times d}$ and $E_w \in \mathbb{R}^{V_w \times d}$ denote the word embedding matrix, $V_w$ denotes the dictionary size, $\text{batch}$ is the number of text data read in batches, $d$ denotes the word vector dimension. Each word in the word embedding matrix $E_w$ has a unique index $b_{ix}$ for retrieving its corresponding word vector, which is a binary vector whose value is 0 or 1. All positions, except the $x_i$ position, are zero. These vectors represent the most primitive information in the corpus and have a significant impact on the next steps.

**C. BILSTM LAYER**

LSTM can overcome the shortcomings of RNN [34], and its unit structure is shown in Fig. 4.

LSTM adjusts information through the gate structure and stores historical information through the storage unit. It is mainly composed of four parts: input gate $i_t$, forgetting gate $f_t$, output gate $o_t$, candidate gate $g_t$. Using LSTM to extract semantic features, the hidden layer state $h_t$ in a single direction at time step $t$ is updated as follows:

$$i_t = \sigma(W_{ei}e_t + W_{hi}h_{t-1} + b_i)$$

$$f_t = \sigma(W_{ef}e_t + W_{hf}h_{t-1} + b_f)$$

$$g_t = \tanh(W_{ec}e_t + W_{hc}h_{t-1} + b_c)$$

$$c_t = i_t g_t + f_t c_{t-1}$$

$$o_t = \sigma(W_{eo}e_t + W_{ho}h_{t-1} + b_o)$$

$$h_t = o_t \tanh(c_t)$$

where $W_{ei}, W_{hi}, W_{ef}, W_{hf}, W_{ec}, W_{hc}, W_{eo}, W_{ho}, W_{ec}, W_{hf}$ represent the corresponding weights, respectively. $b_i, b_f, b_o, b_c$ represent offset values. $\sigma$ and $\tanh$ represent activation functions. Each part will influence data at the next moment. Using the current word vector $e_t$ and the hidden state $h_{t-1}$ of the previous moment as input, the unit structure determines whether the current state uses these inputs, and whether to forget the previously stored partial data, and whether to output the new generated state. Thus, the current unit state $c_t$ is determined by calculating the data of the unit state $c_{t-1}$ at the previous moment and the data currently generated by the unit.
Nevertheless, LSTM network only considers the temporal information while ignoring the contextual information. Bi-LSTM network expands the one-way LSTM network by two layers of network structure (forward and reverse), and guarantees that the past and future information can be taken into account in the time series [35]. Therefore, the global semantic features of the text can be fully captured. As shown in Fig. 1, \( C_{f0} \) and \( C_{r0} \) represent the forward and reverse initial unit state, respectively, \( C_{fn} \) and \( C_{rn} \) store the forward and reverse final unit state information, respectively.

When the word vector is input into the forward Bi-LSTM network, the forward hidden layer semantic feature \( \tilde{h}_i \) will be obtained. Similarly, the backward hidden layer semantic feature \( \tilde{h}_i \) can be obtained if the word vector is input to the reverse Bi-LSTM network. The context embedded representation \( h_i \) at the i-th moment in Bi-LSTM network is a concatenation of the forward output \( \tilde{h}_i \) and the inverted output \( \tilde{h}_i \). The global semantic features \( H \) are obtained by splicing the semantic states of each time step, as shown in formula (10).

\[
\begin{align*}
  h_i &= [\tilde{h}_i \oplus \tilde{h}_i] \\
  H &= (h_1, h_2, \ldots, h_i, \ldots, h_n)
\end{align*}
\]

where \( H \in \mathbb{R}^{\text{batch} \times n \times 2d} \), \( \oplus \) denotes concatenation. \( n \) denotes time step, which is equal to the maximum sequence length of text.

### D. ATTENTION MECHANISM

Attention mechanism is a resource allocation system, which simulates the characteristics of human brain attention and pays more attention to the important information. The introduction of attention mechanism in NLP can highlight the impact of input on output [17]. In this paper, attention mechanism is introduced after Bi-LSTM network to generate weighted global semantic features \( V \) with attention probability distribution \( a \), so that highlights the influence of different features in input global semantic features on text categories. The attention mechanism model used in the current paper is shown in Fig. 5.

\[
\begin{align*}
  m &= W_{a2} \sigma(W_{a1}H + b_a) \\
  a_i &= \frac{\exp(m_i)}{\sum_{j=0}^{n} \exp(m_j)} \\
  a &= [a_1, a_2, \ldots, a_i, \ldots, a_{rn}]
\end{align*}
\]

where \( a \in \mathbb{R}^{\text{batch} \times n \times d} \), \( m_i \in m, m_j \in m, r \) denotes the number of attention schemes for each context. \( W_{a1} \in \mathbb{R}^{d \times n} \) is the global attention weight matrix, \( b_a \) is the global attention bias matrix, \( W_{a2} \in \mathbb{R}^{n \times d} \) denotes the different attention scheme matrices for each context, and \( a_i \) is the global semantic feature attention probability distribution at the i-th moment. The larger value of \( m \), the more important global semantic features of the moment are.

After obtaining \( a_i \) and \( h_i \) of each time, \( V_i \) is calculated by multiplying \( a_i \) and \( h_i \). Then, \( V \) can be obtained by splicing all \( V_i \).

\[
\begin{align*}
  V_i &= a_i h_i \\
  V &= [V_1, V_2, \ldots, V_i, \ldots, V_{rn}]
\end{align*}
\]

where \( V \in \mathbb{R}^{\text{batch} \times n \times 2d} \).

### E. CNN LAYER

By inputting \( V \) into CNN, the local features can be extracted. The wide CNN model used in text categorization is shown in Fig. 6. Each convolution generates a new feature \( \hat{C}_j \) through a fixed size window. After convolution, the j-th text contains local and global semantic features \( C_j \):

\[
\begin{align*}
  \hat{C}_j &= \sigma(W_{vi}V_{ih} + b_{vi}) \\
  C_j &= [\hat{C}_1, \hat{C}_2, \ldots, \hat{C}_i, \ldots, \hat{C}_{rn-h+1}]
\end{align*}
\]
where $C_j \in \mathbb{R}^{(r \times n-h+1) \times 2d}$, $W_v \in \mathbb{R}^{2d \times h}$ denotes the convolution kernel vector used in convolution operation, $h$ and $2d$ denote the height and width of the convolution kernel window respectively. $V_{i:h}$ denotes the weighted global semantic eigenvalues of lines i-th through h-th, corresponding to the feature i-th through h-th. $b_v$ denotes bias.

Then, the maximum pooling method is used to get the final feature representation $\tilde{c}_j$ of each text. After obtaining all of $\tilde{c}_j$, the final text feature vector representation $C$ of the batch of text data is obtained.

$$\tilde{c}_j = \max(C_j)$$  \hspace{1cm} (18)  
$$C = [\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_j, \ldots, \tilde{c}_{\text{batch}}]$$  \hspace{1cm} (19)

where $C \in \mathbb{R}^{\text{batch} \times 2d}$.

**F. OUTPUT LAYER**

Finally, $C$ is classified into the categories of emotional tendency through the Softmax classifier. Taking $C$ as the input of the classification layer, the classification layer connects the best feature $C_d$ to the Softmax classifier by dropout method, and calculates the output vector $p(y)$:

$$p(y) = \text{soft max}(W_c C_d + b_c)$$  \hspace{1cm} (20)
$$y = \text{arg max}(p(y))$$  \hspace{1cm} (21)

where $p(y) \in \mathbb{R}^{\text{batch} \times \text{classes}}$, $W_c \in \mathbb{R}^{2d \times \text{classes}}$ and $b_c$ denote the weight parameters and bias terms of the Softmax classifier, respectively. The classes represent the number of text categories. $C_d \in \mathbb{R}^{\text{batch} \times 2d}$, which is the best feature generated by $C$ through dropout.

Softmax classifier is used to calculate the probability vector $p(y)$ of text data belonging to every category. It is a vector whose dimension is the number of the categories. Each dimension of the vector is a number in the range of 0 to 1, representing the probability that the text belongs to each category. Then, the type corresponding to the maximum probability $y$ is selected as the predictive output of text classification.

Through the classification of classifier layer, the whole model achieves the task of categorizing text data.

**G. THE DIGITAL REPRESENTATION OF THE PREDICTIVE RESULTS**

In order to measure the text classification results digitally, this study used a dictionary-based unsupervised method to calculate the emotion value of text classification results. The method first uses various dictionaries to identify emotional words in the text, and then calculates the overall emotional value of the text according to the judgment rules. Its emotional calculation process is shown in Fig. 7.
1) BASIC EMOTION DICTIONARY CONSTRUCTION

Emotional words referred to the words that can express emotional tendency. They are the basis of emotion calculation process, including nouns, verbs, adjectives, adverbs and idioms. This study used the NTUSD [17], the ontology database of Chinese emotion vocabulary, and the Hotnet emotion dictionary (www.keenage.com) as the basic emotion dictionaries, and combined the positive and negative emotion words. Then, we filtered out emotion words whose polarity is not obvious, emotional words with ambiguous emotion tendency, as well as repeated emotion words. Finally, the basic emotion dictionaries of positive and negative emotional tendency were obtained, respectively. The specific construction process of the dictionary is shown in Fig. 8.

![Algorithm chart of constructing the basic emotion dictionary.](image)

2) EMOTION DICTIONARY EXPANSION

The extension methods of emotion dictionary mainly include manual construction and automatic construction. Since the manual construction method consumes more manpower, material resources and time, the automatic construction method is adopted in this study. The unregistered words are used to expand the emotion dictionary, and the process is shown in Fig. 9. Unregistered words recognized by the fastText neology-finding algorithm in the pre-processing.

The vector representation of post-word segmentation corpus and basic emotion dictionary is obtained by using the open source tool word2vec of Google. Then, the emotion tendency of the unregistered words is identified by calculating the semantic similarity distance of corpus and basic emotion dictionary. After that, the unregistered words are classified into the nearest basic emotion dictionary (positive or negative) by polarity judgment algorithm, so as to automatically expand the dictionary of basic emotion dictionary. The semantic similarity distance can be calculated by calculating the cosine value of two word-vectors. The larger the cosine value is, the more similar the semantics of two words are. The formula for cosine of two N-dimensional vectors \( a = (x_1, x_2, \ldots, x_n) \) and \( b = (y_1, y_2, \ldots, y_n) \) is calculated as follows:

\[
\cos(x, y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}}
\]  \( (22) \)

The polarity judgment algorithm is described as follows:
(a) Word2vec is used to train the preprocessed corpus and the basic emotion dictionary respectively, and the corresponding vectorized representation files are obtained.
(b) If the word \( w \) in the corpus can be found in the basic emotion dictionary, directly transferred to (f). Otherwise, go to (c).
(c) Getting the \( n \) words closest to the unregistered word \( w \) in the corpus vectorization representation file by word2vec.
(d) If there are positive (negative) emotion words in these \( n \) words, the words are labeled \( \text{pos}(\text{neg}) \). Then, calculating the emotion tendency of \( w \) according the formula (23), when \( EO(w) \) is greater than 0, the emotional polarity of \( w \) is positive. Otherwise, it is negative.

\[
EO(w) = \sum_{\text{pos}} \cos(w, \text{pos}) - \sum_{\text{neg}} \cos(w, \text{neg})
\]  \( (23) \)
(e) Storing \( w \) into the corresponding emotional dictionary (positive or negative).
(f) Analyzing the emotion polarity of the next unregistered word.

3) CONSTRUCTION OF OTHER DICTIONARIES

The dictionary-based text emotion calculation method failed to consider the function of other categories of words, such as the negative words. So, it is necessary to construct other dictionaries. This article mainly through the CNKI and manual working to construct other dictionaries. Negative words, for example, searching the semantic elements that contains negative meaning, such as \{neg | 否\}, \{deny | 否认\}, \{BeUnable | 无能\}, \{unable | 庸\} etc., extracting concepts that containing negative semantic elements, and collecting them to obtain
the final Negative word dictionary. The final information of each dictionary in this study is shown in Table 3. Then, the emotional value of the corpus is calculated based on these constructed dictionaries.

### TABLE 3. The final collation information of the dictionaries.

| Dictionary          | Size  |
|---------------------|-------|
| Sentiment words     | 9171  |
| Adverb of degree    | 221   |
| Negative word       | 29    |
| Personal pronoun    | 32    |
| Demonstrative pronoun | 64   |
| Interrogative pronoun | 26   |
| Modal verb          | 29    |
| Conjunction         | 93    |
| Interjection        | 139   |
| Modal particle      | 17    |
| State word          | 49    |
| Auxiliary           | 719   |

4) TEXT EMOTION VALUE CALCULATION

#### a: CALCULATING THE EMOTIONAL VALUE OF EACH WORD IN THE SENTENCE

In the process of emotion value analysis, the subjective sentence should be distinguished from the objective sentence. The subjective sentence refers to the sentence with certain emotional tendency or some emotional color, whereas the objective sentence refers to the sentence without any emotional color, which needs to be filtered out. Matching each sentence after word segmentation with emotion dictionary. If emotion word appears in the sentence, the sentence could be judged as subjective sentence. Then, emotion values are assigned to each phrase in the subjective sentence. The weight of emotion values in various dictionaries is shown in Table 4.

### TABLE 4. Emotional value weights of various dictionaries.

| Dictionary          | Weight |
|---------------------|--------|
| Sentiment words     | 0.98   |
| Adverb of degree    | 0.75   |
| Negative word       | 0.51   |
| Personal pronoun    | 0.95   |
| Demonstrative pronoun | 0.75  |
| Interrogative pronoun | 0.90  |
| Modal verb          | 0.81   |
| Conjunction         | 0.88   |
| Interjection        | 0.75   |
| Modal particle      | 0.75   |

In addition, if sentiment words are dynamic sentiment words, it is necessary to judge whether there are collocation words nearby. If so, the polarity of the sentiment word is need to be inverted, and consider the word as a new sentiment word for polarity calculation of the text. At this point, the emotional value of each word in each sentence is obtained.

#### b: CALCULATING THE AVERAGE EMOTIONAL VALUE OF EACH SENTENCE IN THE TEXT

The weighted sum method is used to calculate the average emotional value of each sentence. After preprocessing, each sentence is represented as $S = \{W_1, W_2, \ldots, W_z, \ldots, W_n\}$.

The average emotional value of each sentence can be expressed as:

$$E(S_i) = \frac{(-1)^{\text{Neg}} \sum_{z=1}^{n} EO(W_z) \prod_{j=1}^{m} \text{mod}_j}{n}$$

where $\text{Neg}$ is the number of negative words in the sentence, $n$ is the number of sentiment words in the sentence, $EO(W_z)$ represents the emotional value of each word, $m$ is the modifier of the i-th emotional word, and mod$_j$ is the weight of the corresponding j-th modifying word.

#### c: CALCULATING THE AVERAGE EMOTIONAL MEAN OF THE TEXT

Text $D$ is generally composed of multiple sentences, $D = \{S_1, S_2, \ldots, S_z, \ldots, S_n\}$. The text $D$ is segmented by punctuation that can break sentences. Only comma "", full stop "", exclamation mark "!", semicolon ";", ellipsis "...", and question mark "?" were considered. Due to "!" and "?" can affect the emotional tendency of the sentence, it is important to determine the punctuation at the end of the sentence. If "!" and "?" conform to the judgment of tag words of rhetorical questions, the emotional score of the sentence is multiplied by 1.5. If "?" and it conforms to the judgment of tag words of rhetorical questions, the emotional score of the sentence is multiplied by $-1$. If other punctuation marks, the sentence score remains the same.

In order to eliminate the effect of the text length on scores, the sum of scores is divided by the total sentence $L_S$, and final be the emotional score of the whole review text. In this paper, the threshold $T$ is set to 0. When the emotional $\text{Score}(D)$ is greater than $T$, the text is positive. When less than $T$, it is negative. otherwise, it is neutral. The formula is as follows:

$$\text{Score}(D) = \frac{\sum_{i=1}^{L} E(S_i)}{L_S}$$

### IV. EXPERIMENTS

#### A. EXPERIMENTAL CORPUS

In this experiment, crawling technology was used to get corpus data, which is the comments of users on a large e-commerce platform after purchasing and using 999 cold medicine. The experimental corpus is shown in Table 5. Fig. 10 presents the sentence length distribution of corpus. Fig. 11 shows some key information of corpus in the form of word clouds.

#### B. EXPERIMENTAL SETTINGS

The experiment was completed in Python language under the Tensorflow deep learning framework. The experimental environment configuration information is shown in Table 6.
Table 6. Experimental environment configuration.

| Environmental configuration | Value                        |
|-----------------------------|------------------------------|
| Operating system            | 64-bit Windows 10 operating system |
| Processor                   | Intel(R) Core(TM) i7-8550U @ 1.80GHZ 1.99 GHZ |
| Memory (RAM)                | 8.00 GB                      |
| Programming language        | Python 3.7                    |
| Deep Learning Framework     | TensorFlow 1.14.0            |

Figure 10. Sentence length distribution of corpus.

Figure 11. Corpus word cloud illustration.

Word vector dimension represents the size of text features. The more features, the more accurately words can be distinguished from words. We set the size of word vector to 100. As Fig. 10 shown, the length of the text sequence is mostly distributed within 50 words, so the maximum length of the text sequence can be set to 50. We use three types of comment text data in this paper, including high praise, medium praise and bad review. The total number of Chinese characters has exceeded 80,000, while only 3500 words are commonly used, including 1000 secondary common words. In order to preserve the content of text data as much as possible, the size of dictionary is set to 5000.

One layer of Bi-LSTM neural network was used, and the number of forward and reverse neurons were both set to 100. After the processing of Bi-LSTM layer, the eigenvalue of corpus increases from the initial 100 to 200. Therefore, the width and height of convolution core is set to 200 and 5, respectively. Different attention schemes can learn sentence representation with different emphasis points, in this experiment, three attention schemes were taken for each text data.

Table 7. Settings of experimental parameters.

| Parameter                         | Value |
|-----------------------------------|-------|
| Word vector dimension            | 100   |
| Maximum length of text sequence  | 50    |
| Number of text categories        | 3     |
| Corpus dictionary size           | 5000  |
| Bi-LSTM layer number             | 1     |
| Number of Bi-LSTM unidirectional neural units | 100   |
| Convolutional kernel width       | 200   |
| Height of convolution nucleus    | 5     |
| Dropout retention ratio          | 0.5   |
| Learning rate                    | 1e-3  |
| Number of attention schemes      | 3     |
| Optimizer                        | Adam  |

In order to avoid the phenomenon of over-fitting in the process of model training, we used dropout in the experiment and set the value to 0.5. The learning rate of training was set to 1e-3. In addition, to constantly understand the training situation, the training results were saved once every 10 cycles and were displayed every 100 cycles in the training session. Table 7 summarized the experimental parameters in this experiment.

C. MODEL TRAINING

In the training process of various deep learning network models, network parameters are adjusted according to the characteristics of network input in an attempt to make the predictive label as consistent as possible with the input label. Suppose the network input is \((x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_n, y_n)\), where \(x_i\) is the input text, \(y_i\) is the corresponding label, \(i = 1, 2, \ldots, n\). \(W_i\) is the weight that connecting \(i\)-th layer neurons and \(i\)-th layer neurons in the network, \(b_i\) is the bias of the \(i\)-th layer neurons. \(W_i\) and \(b_i\) are training update parameters, and the training process is divided into two stages: forward and back propagation stage. In the forward propagation stage, the predicting output \(y_{out}\) is calculated through \(x_i\), \(W_i\) and \(b_i\) of the current network. In back propagation stage, the error between the predicting output \(y_{out}\) and \(y_i\) is calculated and
the cost function $E(W, b)$ is calculated according to the error. Then, the cost function is optimized to update the parameters $W$ and $b$. It is usually optimized by gradient descent method.

$$W \leftarrow W - \eta \frac{\partial E(W, b)}{\partial W} \quad (26)$$

$$b \leftarrow b - \eta \frac{\partial E(W, b)}{\partial b} \quad (27)$$

where $\eta$ is the learning rate parameter used to control the strength of the error propagation.

During model training, the cross-entropy $J(\theta)$ between the predicted results and the real text categories is regarded as the cost function of model training. The random gradient descent method of Adam optimizer is then used to optimize the model.

$$\tilde{\theta} = \theta + \alpha \frac{J(\theta)}{\partial \theta} \quad (28)$$

$$p_1 = p(L_1) \log(y; \theta) \quad (29)$$

$$p_2 = (1 - p(L_1)) \log(1 - y; \theta) \quad (30)$$

$$E(D, L; \theta) = -\sum_{i=0}^{N} (p_1 + p_2) \quad (31)$$

$$J(\theta) = \frac{1}{N} E(D, L; \theta) + \frac{\lambda}{2} ||\theta||^2 \quad (32)$$

where the $E(D, L; \theta)/N$ is the mean error between the actual output and the predicted output, the $\lambda$ is scale factor, $(\lambda, ||\theta||^2)/2$ is norm square of model parameters, which is used to control the impact of overfitting. $\theta$ is the current parameter of TBLC-rAttention model. $\alpha$ denotes learning rate. $N$ is the size of training sample. $D$ is the training sample. $L$ is the real class label (High praise, Medium praise, Bad review) corresponding to sample $D, L_i \in L, y$ is the predicted classification result of Softmax classifier. $p(L_i)$ denotes the correct classification result. Cross-entropy can express the degree of difficulty of $y$ by $p(L_i)$. The smaller the cross-entropy, the more similar $p(L_i)$ and $y$ are.

Adam method is used to minimize the $J(\theta)$ while training the model. It mainly uses the first-order moment estimation and second-order moment estimation of gradient to dynamically adjust each parameter of the model. After the bias correction of Adam, every iteration learning will have a certain range, which makes the parameter change more stable. The training process of the TBLC-rAttention network is shown in Fig. 12.

D. EVALUATION INDICATORS

In order to facilitate the evaluation of the model proposed in this paper, the commonly evaluation indicators of text categorization, Precision ($P$), Recall ($R$) and $F_1$, are used. The calculation formulas are as follows:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (33)$$

$$R = \frac{TP}{TP + TN} \times 100\% \quad (34)$$

$$F_1 = \frac{2 \times P \times R}{P + R} \times 100\% \quad (35)$$

where $TP$ means true positive, $TN$ means true negative, $FN$ means false negative, $FP$ means false positive. The relationship is shown in Table 8.

V. RESULTS AND DISCUSSION

The performance of a new model can generally be evaluated from the accuracy of the results and the complexity of the model. In this study, when the cycle reaches about 2000 times, the accuracy and loss gradually stabilized, with a value around 99.00% and 0.01, respectively. In order to assess the performance of TBLC-rAttention model, the results of another five categorization models (CNN, LSTM, Bi-LSTM, BiLSTM-Attention and RCNN) are used as comparison benchmarks. Fig. 13 shows the accuracy and loss of the whole model training process, where (a) The accuracy value...
during the training process of TBLC-rAttention, (b) The loss value during the training process of TBLC-rAttention, (c) The accuracy value during the training process of all neural network models, (d) The loss value during the training process of all neural network models. Fig. 14 gives the schematic diagram of the TensorFlow training model.
TABLE 9. Comparison of classification effects of text classification models.

| Number | Model                                | accuracy | recall | F1  |
|--------|--------------------------------------|----------|--------|-----|
| 1      | SVM                                  | 0.80     | 0.78   | 0.79|
| 2      | SVM + sentiment word                 | 0.50     | 0.51   | 0.50|
| 3      | SVM + sentiment word + negative word | 0.62     | 0.65   | 0.63|
| 4      | SVM + sentiment word + negative word + adverbs | 0.82 | 0.72  | 0.76|
| 5      | SVM + sentiment word + negative word + adverbs + part of speech | 0.88   | 0.78   | 0.81|
| 6      | SVM + sentiment word + negative word + adverbs + part of speech + conditional word | 0.89   | 0.91   | 0.90|
| 7      | SVM + sentiment word + negative word + adverbs + part of speech + conditional word + ? + ! | 0.84   | 0.86   | 0.87|
| 8      | SVM + Word2vec                       | 0.83     | 0.97   | 0.87|
| 9      | NB                                   | 0.76     | 0.74   | 0.75|
| 10     | NB + sentiment word                  | 0.48     | 0.50   | 0.49|
| 11     | NB + sentiment word + negative word  | 0.60     | 0.63   | 0.61|
| 12     | NB + sentiment word + negative word + adverbs | 0.81   | 0.71   | 0.78|
| 13     | NB + sentiment word + negative word + adverbs + part of speech | 0.86   | 0.77   | 0.80|
| 14     | NB + sentiment word + negative word + adverbs + part of speech + conditional word | 0.85   | 0.81   | 0.83|
| 15     | NB + sentiment word + negative word + adverbs + part of speech + conditional word + ? + ! | 0.79   | 0.84   | 0.74|
| 16     | NB + Word2vec                        | 0.77     | 0.92   | 0.84|
| 17     | CNN                                  | 0.83     | 0.82   | 0.82|
| 18     | LSTM                                 | 0.87     | 0.87   | 0.87|
| 19     | Bi-LSTM                              | 0.90     | 0.90   | 0.90|
| 20     | BiLSTM-Attention                     | 0.94     | 0.94   | 0.94|
| 21     | RCNN                                 | 0.93     | 0.93   | 0.93|
| 22     | TBLC-rAttention                      | 0.99     | 0.99   | 0.99|

All the results were obtained when the training accuracy and test accuracy were stable after 1000 cycles of repeated calculation. The code and data access link of this study is: “https://github.com/xingyu068/tblc-ration”.

In Table 9, as show in the experiment 1-16, the accuracy of traditional machine learning method, SVM and Naive Bayes (NB), can only reach about 80% and 76%, respectively. When SVM and NB are combined with feature extraction method, the method of “SVM (or NB) + sentiment word + negative word + adverbs + part of speech + conditional word” achieves the highest accuracy of 89% and 85%, respectively. Meanwhile, it can be found that the classification accuracy is gradually improved with the addition of “sentiment word”, “negative word”, “part of speech”, and “conditional word”, among which “negative word” and “adverbs” have the greatest influence. However, the addition of “?” and “!” actually reduced the classification accuracy, which indicate that the symbols could not be used as a feature of traditional machine learning methods, and only the best feature combination can guarantee the accuracy and reduce the cost of computation. The combination of Word2Vec and SVM (or NB) can also achieve a better classification effect, which indicates that the word vectors obtained by Word2Vec contain rich semantic information and can be directly used as the feature of emotion classification. Moreover, the result of SVM is better than that of NB.

The results of LSTM and Bi-LSTM (Experiment 18 and 19) show that the accuracy of Bi-LSTM is about 3% higher than that of LSTM. This is due to the fact that Bi-LSTM network extends one-way LSTM network through two layers of forward and reverse network structure, which can fully extract text context information. As a result, the accuracy will be improved, but the space complexity is twice that of LSTM. Compared to Bi-LSTM model (Experiment 19), the accuracy of BiLSTM-Attention model (Experiment 20) is improved by about 4%, which indicates that attention mechanism can effectively identify the feature information that has great impact on classification. Due to the feature of parameter sharing, the accuracy of CNN (Experiment 17) is not very high, but the training time can be greatly saved. As show in the experiment 17-21, RCNN absorbs the advantages of RNN (and its variants) and CNN, and the classification effect is better than that of RNN (and its variants) or CNN, which is close to the result of BiLSTM + Attention model.

Compared with the deep learning method (Experiment 17-22), the traditional machine learning method has a lower classification effect. Its classification accuracy largely depends on the extraction of features of training corpus. It is difficult to identify emotional information in complex sentence patterns because it ignores the features of Chinese sentence structure. It’s worth noting that the TBLC-rAttention model has the highest accuracy among all the models in this experiment. It introduces rAttention into RCNN and used TextRank algorithm in the pretreatment process. As a result, the accuracy of the model reaches to 99% and loss value is as low as 0.035.

Model complexity can be analyzed from time complexity and space complexity. In general, the number of basic repeated operations of the algorithm is a function $f(n)$ of module $n$. So, the time complexity of the algorithm can be recorded as $T(n) = O(f(n))$. $T(n)$ is directly proportional to $f(n)$, so with the increasing of module $n$, the smaller $f(n)$ is, the lower the time complexity of the algorithm is, and the higher the efficiency of the algorithm is. In order to calculate the $T(n)$, first calculate the $f(n)$, and then get the order of magnitude $\xi(n)$ of $f(n)$ by ignoring the constant and the coefficient of the lower power and the highest power.

VOLUME 8, 2020

96823
FIGURE 15. The relationships between common time complexity.

If \( \lim_{n \to \infty} \frac{f(n)}{\xi(n)} \neq 0 \), \( O(f(n)) \) can be represented by \( \xi(n) \), which retains only the highest order and contains no coefficient terms. For example, the time complexity of \( f(n) = 5n^3 + 2n^2 \) is \( O(n^3) \). The common time complexity is shown in Table 10, and \( O(1) < O(\log n) < O(n) < O(n \log n) < O(n^2) < O(n^3) < O(2^n) < O(n!) < O(n^n) \) (Fig. 15). Note that only the runnable statements could affect the time complexity, so all runnable statements in the program except for loops have a time complexity of \( O(1) \). In the code of this experiment, only two layers of loop structure are used at most, in which the loop number of the outer loop is \( n \) and the loop number of the inner loop is \( \text{batch}_{-}\text{train} = 100 \), so \( f(n) = 100n \). As shown in Table 10, the time complexity of TBLC-rAttention is of linear order. In Fig. 16, it is also verified by parameter “time”, which is the change of time spent in the training process of TBLC-rAttention. It can be seen that it is almost linear.

Space complexity is a measure of the amount of storage space temporarily occupied by an algorithm during its operation, denoted as \( S(n) \). If the data size is independent of the number of temporary variables, the spatial complexity is always \( S(1) \) as long as the number of defined variables does not change, regardless of the size of the data. In general, if the algorithm does not involve dynamically allocated space and recursive operations, the space complexity is \( S(1) \). In this study, all the defined variables are the same, and there is little use of temporary variables, so the space complexity of TBLC-rAttention is \( S(1) \). In total, the space complexity of TBLC-rAttention is not very high, and it shows a good classification accuracy, so it has high practical value.

Table 11 shows the validation result of TBLC-rAttention model. It categories the comment data with almost 100% accuracy. Of note, the e-commerce system will default to

| Execution degree function \( T(n) \) | Time complexity \( O(f(n)) \) | Name |
|--------------------------------------|-----------------------------|------|
| Constant, for example, 8             | \( O(1) \)                  | Constant order |
| \( 5n + 3 \)                         | \( O(n) \)                  | Linear order |
| \( 5n^2 + 3n + 2 \)                  | \( O(n^2) \)                | Square order |
| \( 5n^3 + 3n^2 + 6n + 9 \)          | \( O(n^3) \)                | Square order |
| \( 5 \log_2^a + 8 \)                | \( O(\log^a) \)             | Logarithmic order |
| \( 3n + 5n \log_2^a + 7 \)          | \( O(n \log^a) \)           | \( n \log^a \) order |
| \( 2^a + 3n + 9 \)                  | \( O(2^a) \)                | Exponential order |

FIGURE 16. Partial parameter data in training process of TBLC-rAttention model.

| Iter: | 100, Train Loss: 0.91, Train Acc: 62.00%, Val Loss: 1.11, Val Acc: 53.69%, Time: 00:48 |
|-------|--------------------------------------------------------|
| Iter: | 200, Train Loss: 0.7, Train Acc: 73.00%, Val Loss: 0.89, Val Acc: 60.40%, Time: 01:34 |
| Iter: | 300, Train Loss: 0.54, Train Acc: 67.00%, Val Loss: 0.72, Val Acc: 61.74%, Time: 02:22 |
| Iter: | 400, Train Loss: 0.51, Train Acc: 73.00%, Val Loss: 0.68, Val Acc: 63.09%, Time: 03:12 |
| Iter: | 500, Train Loss: 0.44, Train Acc: 76.00%, Val Loss: 0.6, Val Acc: 71.14%, Time: 04:01 |
| Iter: | 600, Train Loss: 0.48, Train Acc: 78.00%, Val Loss: 0.57, Val Acc: 71.14%, Time: 04:49 |
| Iter: | 700, Train Loss: 0.46, Train Acc: 78.00%, Val Loss: 0.49, Val Acc: 76.51%, Time: 05:40 |
| Iter: | 800, Train Loss: 0.42, Train Acc: 79.00%, Val Loss: 0.52, Val Acc: 78.52%, Time: 06:26 |
| Iter: | 900, Train Loss: 0.4, Train Acc: 81.00%, Val Loss: 0.51, Val Acc: 77.18%, Time: 07:16 |
| Iter: | 1000, Train Loss: 0.27, Train Acc: 90.00%, Val Loss: 0.44, Val Acc: 84.56%, Time: 08:01 |
| Iter: | 1100, Train Loss: 0.31, Train Acc: 84.00%, Val Loss: 0.36, Val Acc: 87.92%, Time: 08:48 |
| Iter: | 1200, Train Loss: 0.21, Train Acc: 89.00%, Val Loss: 0.36, Val Acc: 87.25%, Time: 09:32 |
| Iter: | 1300, Train Loss: 0.14, Train Acc: 94.00%, Val Loss: 0.31, Val Acc: 89.93%, Time: 10:19 |
| Iter: | 1400, Train Loss: 0.15, Train Acc: 94.00%, Val Loss: 0.38, Val Acc: 91.28%, Time: 11:04 |
| Iter: | 1500, Train Loss: 0.13, Train Acc: 96.00%, Val Loss: 0.24, Val Acc: 91.95%, Time: 11:51 |
| Iter: | 1600, Train Loss: 0.095, Train Acc: 96.00%, Val Loss: 0.25, Val Acc: 92.62%, Time: 12:37 |
| Iter: | 1700, Train Loss: 0.14, Train Acc: 95.00%, Val Loss: 0.29, Val Acc: 87.92%, Time: 13:24 |
| Iter: | 1800, Train Loss: 0.2, Train Acc: 90.00%, Val Loss: 0.2, Val Acc: 91.95%, Time: 14:13 |
| Iter: | 1900, Train Loss: 0.14, Train Acc: 91.30%, Val Loss: 0.16, Val Acc: 94.63%, Time: 15:00 |
praise when consumers do not comment, but the model treats such data as a medium review, which is more objective. In addition, when the predicted digitalized emotion value is positive, the review is high praise. When it is negative, it means that the review is bad, and when it is equal to 0, it means that the review is medium praise. Emotion value can not only further determine the accuracy of classification, but also realize the digital measurement of the text emotion.

The above experimental results indicate that the proposed model is effective. Compared with the other models, it achieves the best results. This also shows that the combination of Bi-LSTM and CNN can more effectively mine text semantic features and also shows the necessity of introducing attention mechanism. However, although TBLC-rAttention improves the accuracy, it meanwhile increases the computational complexity and results in a slight increase in the computational time. In conclusion, the method proposed in this study has the following characteristics:

1. It can effectively mine the hidden information in the comments of the drug, help the staff analyze the problems existed in each link from the medicine sale to use. For example, for managers, by analyzing the emotional tendency of evaluators, they can provide improvement strategies for products. For drug developers, it is possible to analyze the user's evaluation of the drug to understand the therapeutic effect of the drug and further improve the drug.

2. It proposed an automatic emotion dictionary extension method based on the knowledge of medicine in medical field. Firstly, vector representation is performed on the corpus after word segmentation and the basic emotional dictionary. Then, the semantic similarity distance between the corpus and the basic emotional dictionary is calculated to judge the emotion tendency of unregistered words in the corpus. Lastly, polarity judgment algorithm is used to classify the unregistered words into the basic emotional dictionary. The automatic expanded emotion dictionary is applied in the preprocessing to verify

TABLE 11. Verification results of TBLC-rAttention model.

| Corpus for verification | Emotion value | Model results | Real results |
|-------------------------|---------------|---------------|--------------|
|                         | 0.260         | High praise   | High praise  |
|                         | 0.213         | High praise   | High praise  |
|                         | 0.334         | High praise   | High praise  |
|                         | 0.325         | High praise   | High praise  |
|                         | 0.232         | High praise   | High praise  |
|                         | 0.213         | High praise   | High praise  |
|                         | 0.334         | High praise   | High praise  |
|                         | 0.325         | High praise   | High praise  |
|                         | 0.232         | High praise   | High praise  |
|                         | 0.149         | Bad review    | Bad review   |
|                         | 0.396         | Bad review    | Bad review   |
|                         | 0.383         | Bad review    | Bad review   |
|                         | 0.536         | Bad review    | Bad review   |
|                         | 0.241         | Bad review    | Bad review   |
the effect of the expanded dictionary on the performance of TBLC-rAttention. The experiment shows that the expanded dictionary can better identify new words in the medical field, further improve the effect of Chinese word segmentation in the medical field and improve the recognition accuracy of the model.

(3) TBLC-rAttention deep neural network model is designed for Chinese medical text emotion recognition based on deep learning technology. It fully draws on the advantages of Bi-LSTM, attention mechanism and CNN. Firstly, Bi-LSTM is used to extract the global semantic features of the text, and rAttention mechanism is introduced in the process. Secondly, CNN is applied to extract the local semantic features and further mining text information, and obtain the final semantic feature vectors. Thirdly, r attention schemes are used instead of traditional single attention scheme during calculating weighted global semantic features. Different attention scheme can learn sentence representations with different emphases, thus more valuable feature information can be extracted. Fourthly, TextRank algorithm is used in the pre-processing to let the model can also process long text data and minimize the impact of interference information on classification. Finally, the classification task of Chinese medical reviews is completed by using the obtained semantic feature vectors. In this way, the model can mine semantic feature information to the greatest extent, and meanwhile effectively solve the problems existing in traditional methods, such as high dimension, high sparse, weak feature expression ability, inapplicability in large data sets and so on. In addition, the model is implemented by Google’s TensorFlow framework and trained to adjust parameters. Experiments show that the model has good convergence speed and accuracy.

(4) A dictionary-based unsupervised method was used to calculate the emotional value of the results predicted by the TBLC-rAttention model. At first, various dictionaries are used to identify the sentiment words in the text, and then the overall emotional value of the text is calculated according to the judgment rules. On the one hand, the digital measurement of the predicted results is realized; on the other hand, the positive and negative values of the calculated results can verify the predicted results. And the absolute value of the calculated results indicates the emotional intensity.

(5) The model is easy to use. Deep learning can handle the problem of text understanding, and don’t need to understand any language related to the other syntactic or semantic structure. After the training is completed, deep learning can directly analyze and infer the high-level goals, so as to solve the problems of inaccurate word segmentation or inadequate understanding of semantic features caused by a large number of professional terms in the medical field, and improve the classification effect of the model.

(6) The TBLC-rAttention model solves the difficulty of little information in short phrase data. For example, “Logistics is very fast, cure cold, very responsible, bought two bags at one time, 999 is trustworthy.” In the absence of context learning, it is difficult to judge what “999” means in “999 is trustworthy” only by using the literal meaning of the text. This phrase is too short to contain particularly obvious information words that can be used as subject judgments. The key word “999” is more likely to be understood as simple numbers in machine. TBLC-rAttention model can link “999” to disease after learning from the context of the text, such as “cure”, “wind-cold” and “cold” appear in the text.

VI. CONCLUSION
In this paper, a hybrid deep neural network model, TBLC-rAttention, aiming at Chinese text emotion recognition, is proposed. This model can be used for knowledge extraction, modeling, classification and mining of the medical texts. In this study, the performance of the TBLC-rAttention is evaluated and verified on the collected data set. Although the training time is slightly longer, it is within an acceptable range.

Deep learning is an effective technique for analyzing and processing medical texts. Although this study has made some achievements in this respect, the classification and mining of medical text using models with better performance and higher intelligence level remains a challenging field. Thus, In the future, our major work includes: (1) considering how to carry out a better preprocessing operation on the corpus data, such as reducing noise to a greater extent, segmenting words more accurately, and recognizing medical terms better. And it is necessary to further optimize the efficiency and accuracy of dictionary and medical terms recognition. (2) Trying other algorithms and models, especially the new method just proposed, and carrying out effective fusion and improvement, so as to further improve the accuracy of the model and improve its time efficiency. For example, more convolution cores can be designed in the convolutional neural network model for parallel convolution calculation. (3) Increasing the diversity of corpus data, like using as many kinds of drug comment data as possible, even text data in other fields. Therefore, the model will be more versatile. It can be used for text categorization tasks in long/short text and other arbitrary fields.

REFERENCES
[1] M. Yang, W. Zhao, L. Chen, Q. Qu, Z. Zhao, and Y. Shen, “Investigating the transferring capability of capsule networks for text classification,” Neural Netw., vol. 118, pp. 247–261, Oct. 2019.
[2] S. M. Jiménez-Zafrí, M. T. Martín-Valdivia, M. D. Molina-González, and L. A. Ureña-López, “How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain,” Artif. Intell. Med., vol. 93, pp. 50–57, Jan. 2019.
[3] A. Gangemi, V. Presutti, and D. R. Recupero, “Frame-based detection of opinion holders and topics: A model and a tool,” IEEE Comput. Intell. Mag., vol. 9, no. 1, pp. 20–30, Feb. 2014.
[4] M. Krishnan, “Against interpretability: A critical examination of the interpretability problem in machine learning,” Philosophy Technol., pp. 1–6, Aug. 2019.
[5] M. C. Díaz-Galiano, M. T. Martín-Valdivia, and L. A. Ureña-López, “Query expansion with a medical ontology to improve a multimodal information retrieval system,” Comput. Biol. Med., vol. 39, no. 4, pp. 396–403, Apr. 2009.
[6] M. Gridach, “Character-level neural network for biomedical named entity recognition,” J. Biomed. Inform., vol. 70, pp. 85–91, Jan. 2017.
[7] C. Ye, T. Fu, S. Hao, Y. Zhang, O. Wang, B. Jin, M. Xia, M. Liu, X. Zhou, Q. Wu, Y. Guo, C. Zhu, Y.-M. Li, D. S. Culver, S. T. Alfreds, F. Stearns, K. G. Sylvester, E. Widen, D. McElhinney, and X. Ling, “Prediction of incident hypertension within the next year: Prospective study using statewide electronic health records and machine learning,” J. Med. Internet Res., vol. 20, no. 1, p. e22, Jan. 2018.

[8] Y. Rao, Q. Li, L. Wenying, Q. Wu, and X. Quan, “Affective topic model for social emotion detection,” Neural Netw., vol. 58, pp. 29–37, Oct. 2014.

[9] S. R. Das and M. Y. Chen, “Yahoo! For amazon: Sentiment extraction from small talk on the Web,” Manage. Sci., vol. 53, no. 9, pp. 1375–1388, Sep. 2007.

[10] S. Hassanpour, C. P. Langlotz, T. J. Amrhein, N. T. Befera, and M. P. Lungren, “Performance of a machine learning classifier of knee MRI reports in two large academic radiology practices: A tool to estimate diagnostic yield,” Amer. J. Roentgenology, vol. 208, no. 4, pp. 750–753, Apr. 2017.

[11] Y. Li, W. Jiang, L. Yang, and T. Wu, “On neural networks and learning systems for business computing,” Neurocomputing, vol. 275, pp. 1150–1159, Jan. 2018.

[12] R. Kavuluru, A. Rios, and T. Tran, “Extracting drug-drug interactions with word and character-level recurrent neural networks,” in Proc. IEEE Int. Conf. Healthcare Inform. (ICHI), Aug. 2017, pp. 5–12.

[13] J. Frankovich, C. A. Longhurst, and S. M. Sutherland, “Evidence-based Y. Li, W. Jiang, L. Yang, and T. Wu, “On neural networks and learning systems for business computing,” Neurocomputing, vol. 275, pp. 1150–1159, Jan. 2018.

[14] D. M. Blei and J. D. McAuliffe, “Supervised topic models,” in Proc. Int. Conf. Neural Inf. Process., 2010, pp. 121–128.

[15] B. E. Chapman, S. Lee, H. P. Kang, and W. W. Chapman, “Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm,” J. Biomed. Inform., vol. 44, no. 5, pp. 726–737, Oct. 2011.

[16] L.-W. Ku and H.-H. Chen, “Mining opinions from the Web: Beyond relevance retrieval,” J. Amer. Soc. Inf. Sci. Technol., vol. 58, no. 12, pp. 1838–1850, Oct. 2007.

[17] S. Bao, S. Xu, L. Zhang, R. Yan, Z. Su, D. Han, and Y. Yu, “Mining social emotions from affective text,” IEEE Trans. Knowl. Data Eng., vol. 24, no. 9, pp. 1658–1670, Sep. 2012.

[18] S. Poria, A. Gelbukh, D. Das, and S. Bandypadhyay, “Fuzzy clustering for semi-supervised learning-case study: Construction of an emotion lexicon,” in Proc. Mex. Int. Conf. Artif. Intell., 2012, pp. 73–86.

[19] E. Cambria, B. Schuller, B. Liu, H. Wang, and C. Havasi, “Knowledge-based approaches to concept-level sentiment analysis,” IEEE Intell. Syst., vol. 28, no. 2, pp. 12–14, Mar. 2013.

[20] X. Cheng, X. Yan, Y. Lan, and J. Guo, “BTM: Topic modeling over short texts,” IEEE Trans. Knowl. Data Eng., vol. 26, no. 12, pp. 2928–2941, Dec. 2014.

[21] G. E. Hinton, S. Osindero, and Y. W. Teh, “A fast learning algorithm for deep belief nets,” Neural Comput., vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[22] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, vol. 313, no. 5786, pp. 504–507, Jul. 28, 2006.

[23] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, “Exploring strategies for training deep neural networks,” J. Mach. Learn. Res., vol. 1, pp. 1–40, Jan. 2009.

[24] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, “Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion,” J. Mach. Learn. Res., vol. 11, no. 12, pp. 3371–3408, Dec. 2010.

[25] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, “Deep convolutional neural network for inverse problems in imaging,” IEEE Trans. Image Process., vol. 26, no. 9, pp. 4509–4522, Sep. 2017.

[26] L. Deng, “A tutorial survey of architectures, algorithms, and applications for deep learning—ERRATUM,” AFSIPA Trans. Signal Inf. Process., vol. 3, p. 22, Jan. 2014.

[27] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[28] O. Abdel-Hamid, A. R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, “Convolutional neural networks for speech recognition,” IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 22, no. 10, pp. 1533–1545, Oct. 2014.

[29] E. Cambria and B. White, “Jumping NLP curves: A review of natural language processing research [Review Article],” IEEE Comput. Intell. Mag., vol. 9, no. 2, pp. 48–57, May 2014.

[30] R. Colloberti, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, “Natural language processing (almost) from scratch,” J. Mach. Learn. Res., vol. 12, pp. 2493–2537, Aug. 2011.

[31] T. Cai, A. A. Giannopoulos, S. Yu, T. Kelil, B. Ripley, K. K. Kumamaru, F. J. Rybicki, and D. Mitsouras, “Natural language processing technologies in radiology research and clinical applications,” Radiographics, vol. 36, no. 1, pp. 91–176, Jan/Feb. 2016.

[32] M. Long, Y. Cao, J. Wang, and M. I. Jordan, “Learning transferable features with deep adaptation networks,” Comput. Sci., vol. 37, pp. 97–105, Jul. 2015.

[33] B. T. Ong, K. Sugiura, and K. Zettou, “Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM2.5,” Neural Comput. Appl., vol. 27, no. 6, pp. 1553–1566, Aug. 2016.

[34] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” IEEE Trans. Neural Netw., vol. 5, no. 2, pp. 66–67, Mar. 1994.

[35] M. Schuster and K. Paliwal, “Bidirectional recurrent neural networks,” IEEE Trans. Signal Process., vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[36] Y. Kim, “Convolutional neural networks for sentence classification,” in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2014, pp. 1746–1751.

[37] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, “A convolutional neural network for modelling sentences,” in Eprint Arxiv, Baltimore, Maryland, 2014, pp. 655–665.

[38] S. Lai, L. Xu, K. Liu, and J. Zhao, “Recurrent convolutional neural networks for text classification,” in Proc. 29th AAAI Conf. Artif. Intell., Feb. 2015, pp. 2267–2273.

[39] Y. Li, Y. Cai, H.-F. Leung, and Q. Li, “Improving short text modeling by two-level attention networks for sentiment classification,” in Proc. DASFAA, May 2018, pp. 878–890.

[40] H. Geng, L. Li, Y. Liu, S. Qian, and A. Zhou, “Aspect-based sentiment analysis based on dynamic attention GRL,” Scientia Sinica Informationis, vol. 49, no. 8, pp. 1019–1030, Aug. 2019.

[41] Y. Niu, X. Zhu, J. Li, and G. Hirst, “Analysis of polarity information in medical text,” in Proc. AMIA Annu. Symp. Proc., 2005, pp. 570–574.

[42] C. Tung and W. Lu, “Analyzing depression tendency of Web posts using an event-driven depression tendency warning model,” Artif. Intell. Med., vol. 66, pp. 53–62, Jan. 2016.

[43] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” Jan. 2013, arXiv:1301.3781. [Online]. Available: https://arxiv.org/abs/1301.3781
WENJUAN PENG received the bachelor’s degree in preventive medicine from Capital Medical University, in 2017, where she is currently pursuing the Ph.D. degree. She was admitted to the Department of Epidemiology and Health Statistics, School of Public Health, Capital Medical University, Beijing, China. Her research directions are molecular epidemiology and bioinformatics.

WU CAI received the B.S. degree from the Beijing University of Chemical Technology, Beijing, China, in 2014, where she is currently pursuing the Ph.D. degree with the Department of Information Science and Engineering. Her research interests are in decoupling and disturbance rejection of multivariable systems.

YUMING ZHANG received the B.Sc. degree in automation from the Tianjin University of Commerce, Tianjin, China, in 2011, where he is currently pursuing the Ph.D. degree. In 2013, he entered the Institute of Automation, Beijing University of Chemical Technology, Beijing, China. His main research interests are modern control and system identification, including active disturbance rejection control, disturbance observer, model predictive control, and intelligent modeling.

LING ZHANG received the Ph.D. degree in molecular epidemiology from Tokyo Medical and Dental University, Tokyo, Japan, in 2007. She joined Capital Medical University, Beijing, China, in 2007. She is currently a Full Professor with the Department of Epidemiology and Health Statistics, School of Public Health. In recent years, she was awarded several prizes for science and technology progress. Her main research interests include molecular epidemiology, clinical epidemiology, and evidence-based medicine.