A New Approach to Testing BGRU Text Affective Analysis

Tian Li

Department of Foreign Language, Shaoyang University, Shaoyang, China

Keywords: Bidirectional Gated Recurrent Unit (BGRU); Deep learning; Affective analysis

Abstract: As an indispensable part of social life, affective analysis of text data generated by social networks has become a research hotspot in the field of natural language processing. In view of the fact that depth learning technology can automatically construct text features, CNN (Conventional Neural Network) and BLSTM (Bidirectional Long Short-term Memory) have been proposed to solve the problem of text emotion analysis. However, BGRU (Bidirectional Gated Recurrent Unit) can memorize the context information of the sequence, and has simple structure and fast training speed. This thesis proposes a BGRU-based affective analysis method for English texts. Firstly, the text is converted into a sequence of word embeddings, then the contextual affective features of the text are obtained by using BGRU, and finally the affective tendencies of the text are given by the classifier. Experiments on ChnSentiCorp corpus show that this method achieves 90.61% F1 value, which is superior to CNN and BLSTM models, and the training speed is 1.36 times as fast as BLSTM.

1. Introduction

There are two methods of text emotion analysis: Semantic-based affective dictionary approach and machine learning-based approach. Semantic-based affective dictionary method needs to construct affective dictionary firstly, then design the algorithm of affective bias calculation, and finally determine the threshold value to judge the affective bias of the text. For example, Zhao Yanyan and others used massive microblogging data to build a large-scale emotion dictionary of 100,000 words to improve the performance of emotion classification. The advantage of this method is that it is easy to implement and does not need to label the training set manually, but its effect depends on the size and quality of the affective dictionary. The method based on machine learning firstly needs manual labeled text emotional tendencies training set, then extracts text emotional features through manual design features, and finally uses machine learning method to construct classifier to classify text emotional tendencies. The text emotional tendencies are classified into three categories: text emotional tendencies, text emotional tendencies, text emotional tendencies, text emotional tendencies, text emotional tendencies and text emotional tendencies. Naive Bayesian Maximum Entropy Model and Support Embedding Machine are commonly used in machine learning. For example, Wang Chun improved the way of selecting emotion words and calculating weights, and combined ensemble learning method and voting method with traditional machine learning method. This method has the characteristics of small computation and easy implementation, but its generalization ability is insufficient in the face of complex classification problems. In this thesis, we propose a BGRU (Bidirectional Gated Recurrent Unit)-based approach to affective analysis of English texts. This method does not need to construct emotion features artificially, but extracts text emotion features directly by deep neural network training, and then classifies text emotion tendencies. The experimental results show that the proposed method is effective in solving the problem of affective analysis of English texts.

2. Relevant work

Affective analysis, also known as bias analysis, has attracted the attention of many scholars since it was put forward in 2002. At present, most of the researches in affective analysis are divided into two parts: praise and disparagement. For example, Pang et al. took the meta-grammar and part of

  DOI: 10.25236/aisct.2019.084

speech of text as their emotional features by using the bag of words model, and then classified the emotional features of text by using Naive Bayesian, Maximum Entropy Model and Support Embedding Machine respectively. Follow-up research is mainly focused on the emotional characteristics of the text to optimize. Yao Tianfang proposed to extract the topic of the sentence, and then through syntactic analysis, to identify the topic and emotional descriptors, so as to determine the emotional tendencies. Ding et al. proposed a method of affective word pairing in a specific domain to judge affective tendencies. By using synonyms and antonyms in WordNet, Hu et al. Get the emotional polarity of words, and then determine the emotional inclination of sentences according to the dominant emotional polarity of words in sentences. By expanding HowNet, Zhu Yanran and others used semantic similarity and semantic related words to judge the emotional tendency of words.

3. Affective Analysis Model Based on BGRU

3.1 Word embedding

Word embedding is a low-dimensional real-number embedding representation of words through neural network model. By mapping discrete words into dimensional real number space, it can express abundant semantic information, solve embedding sparse problem of pouch model, and mine semantic association between words at the same time. In this thesis, the open source word embedding tool word2vec provided by Google is used to train the word embedding based on skip-gram model. The skip-gram model is shown in Figure 1.

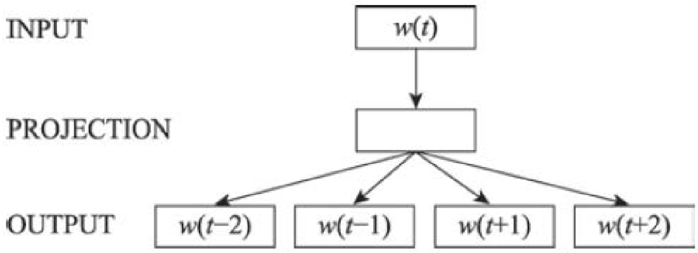

Fig.1 Model of skip-gram

The model predicts the context words from the middle words $w(t)$ of the window, and the objective function is shown in Equation (1):

$$f = \frac{1}{N}\sum_{i=1}^{N}\sum_{-t\leqslant j\leqslant t}\ln P\left(w_{i+j} \mid w_i\right) \tag{1}$$

Where t is the size of the context window at the time of training. This thesis sets the size of t to 5 and the word embedding dimension to 100.

3.2 Cyclic neural network RNN

In the traditional neural network model, it is generally composed of input layer, hidden layer and output layer. The nodes between layers are connected completely, but the nodes between each layer are not connected. This traditional model cannot describe the correlation between the input sequence and the output sequence, which leads to the poor performance in dealing with the sequence problem. RNN (Recurrent Neural Networks) connects the nodes between hidden layers on the basis of the traditional network. For a given input sequence $\left(x_1, x_2, \cdots, x_n\right)$, RNN can effectively learn the sequence characteristics of the data dynamically and has certain memory ability. The so-called memory ability means that the network will save the output information in $t-1$ of the network and apply it to the calculation of the time t. That is, the input of the implicit layer at the time t includes the output of the input layer at the time t and the output of the implicit layer at the time $t-1$

RNN is applied to affective analysis of English texts, and its network structure is shown in Figure 2. In the figure, the input text "Good weather today" is changed to "Good/weather/today" after word segmentation. First, each word is converted into the corresponding word embedding (x_1, x_2, x_3, x_4), and then input into the RNN network in turn. U is the weights connected from the input layer to the implicit layer, W is the weights connected from the implicit layer to itself, and V is the weights from the implicit layer to the output layer.

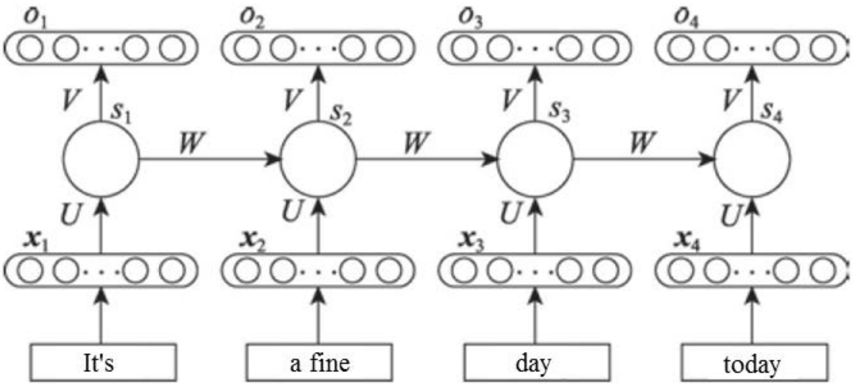

Fig.2 Sentiment analysis network based on

The RNN is calculated as follows:

(1) At time t, input x_t to hidden layer.

(2) s_t is the output of step t of the implicit layer, s_t is obtained from the output x_t of the current input layer and the state s_{t-1} of the implicit layer at the previous time, $s_t = f\left(Ux_t + Ws_{t-1}\right)$, in which f is a nonlinear function, such as tanh or ReLU.

(3) The output is given o_t, $o_t = \text{softmax}\left(Vs_t\right)$.

For the emotional analysis of text, the output of each word is not needed.Only the output of the last word of the text is needed as the emotional feature representation of the sequence, which is input into the classifier to classify the emotional information of the text.

3.3 Threshold Cycle Unit GRU

When the RNN model is expanded according to time series, it is a multi-layer feedforward neural network, and there will be gradient disappearance and gradient explosion in the training process. When dealing with long sequences, RNN cannot solve the problem of long-time dependency, and it is difficult to learn long-distance information. The LSTM memory cell proposed by Hochreiter et al is based on RNN memory cell with a threshold mechanism. The structure of the LSTM memory cell is shown in Figure 3. As can be seen from the figure, LSTM memory cells are mainly composed of cell state and gate structure. The cell state is responsible for storing historical information, and the gate structure is responsible for protecting and controlling the cell state. A memory sheet has three gate structures, namely an input gate i, an output gate o, and a forgetting gate f. The forgetting gate f determines the information to be discarded from the cell state, which can achieve the effect of filtering historical information and solve the problem of gradient disappearance.

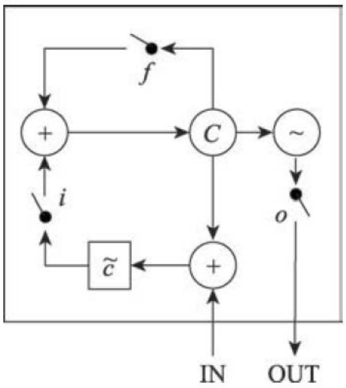

Fig.3 Memory unit of LSTM

Because of the complex structure of LSTM memory cells, there is a problem of long training time. Cho et al. proposed a variant model GRU of LSTM, the structure of which is shown in Figure 4.

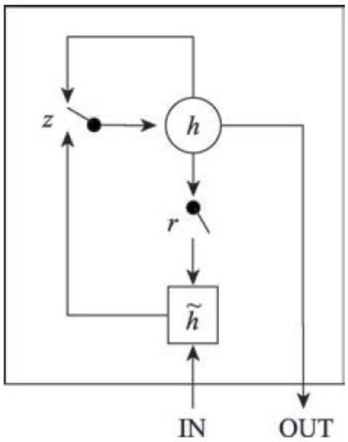

Fig.4 Memory unit of GRU

As can be seen from Figure 4, GRU memory unit combines forgetting gate f and input gate i in LSTM into an update gate z, which not only preserves important features and solves long dependency problem, but also has a simpler structure than LSTM. At the moment t, for a given input x_t, the hidden layer output h_t of the GRU is calculated as follows:

(1) $z_t = \sigma\left(W_z \cdot \left[h_{t-1}, x_t\right]\right)$

(2) $r_t = \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right]\right)$

(3) $\tilde{h}_t = \tanh\left(W \cdot \left[r_t * h_{t-1}, \mathbf{x}_t\right]\right)$

(4) $h_t = \left(1 - z_t\right) * h_{t-1} + z_t * \tilde{h}_t$

Where W is the weight matrix connecting the two layers, σ and tanh are activation functions, z and r are update gates and reset gates respectively.

3.4 BGRU

The standard RNN makes use of the past information but does not take the future information into account when dealing with the sequence problem according to the forward input sequence. To solve this problem, the BRNN (Bidirectional Recurrent Neural Network) model proposed by Schuster et al,not only memorizes the above information, but also memorizes the following information. The basic idea is to use two RNNs to process forward and reverse sequences respectively, and then connect their outputs to the same output layer, so that the bidirectional context information of feature sequences can be recorded. On the basis of BRNN, the hidden layer neurons in BRNN can be replaced with GRU memory cells, and the BGRU model can be obtained. In this thesis, BGRU is applied to affective analysis of English texts, and its network structure is shown in Figure 5.

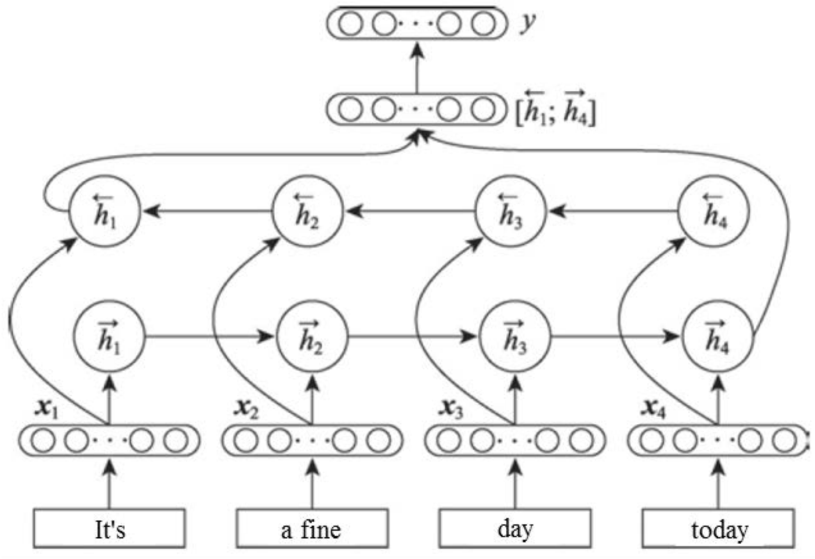

Fig.5 Sentiment analysis network based on BGRU

For a given n dimension input $(x_1, x_2, \cdots, x_n)$, the hidden layer output h of the BGRU at the time t is calculated as follows:

$$\vec{h}_t = \sigma\left(W_{x\vec{h}} x_t + W_{\vec{h}\vec{h}} \vec{h}_{t-1} + b_{\vec{h}}\right) \quad (1)$$

$$\overleftarrow{h}_t = \sigma\left(W_{x\overleftarrow{h}} x_t + W_{\overleftarrow{h}\overleftarrow{h}} \overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}\right) \quad (2)$$

Where, W is the weight matrix connecting the two layers, b is the bias embedding, σ is the activation function, $\vec{h}_t$ and $\overleftarrow{h}_t$, respectively, the positive and negative GRU outputs.

In this model, the output of the last node of the forward and reverse GRU is spliced as the emotional feature of the text, and then connected to the output layer to obtain the output y, as shown in Equation (2):

$$y = W_{hy}\left[\vec{h}_n; \overleftarrow{h}_1\right] + b_y \quad (2)$$

3.5 Model Training

Text affective analysis is actually a classification problem. In this thesis, the text emotion is divided into positive and negative poles. Therefore, in the process of training BGRU, the activation function of this thesis selects sigmoid function, as shown in Equation (3):

$$S(x) = \frac{1}{1+\mathrm{e}^{-x}} \quad (3)$$

The output of the model takes the value $\{0,1\}$, and the objective function selects the logarithmic loss function, as shown in Equation (4):

$$L(Y, P(Y \mid X)) = -\ln P(Y \mid X) \quad (4)$$

In the formula, X is the true emotion distribution of the text and Y is the predicted emotion distribution of the model. The objective of model training is to minimize logarithmic loss function. For model optimizer selection, Adam (adaptive moment estimation) algorithm can calculate the adaptive learning rate of each parameter, and has the characteristics of fast implementation of good results, so Adam is used as optimizer. At the same time, in order to balance the update speed and update times of network weights, Mini-batch gradient descent strategy is adopted and batch-size is set to 128, which improves the learning speed and ensures the convergence of the model. In order to prevent over-fitting in the training process of deep learning networks, a Dropout strategy is added between the BGRU layer and the output layer, which improves the generalization ability of the model by abandoning some modifications of weights, and the dropping rate of Dropout is set to 0.5.

4. Conclusion

In this thesis, an effective method for affective analysis of English texts based on BGRU depth neural network is proposed. After the text is converted into a word embedding sequence, the contextual affective features of the text are extracted automatically by BGRU network, and then the affective tendencies of the text are classified. Experimental results show the effectiveness of the proposed method. Compared with other models, the proposed method can capture affective features better and has a better effect. Moreover, the training speed is 1.36 times of that of BLSTM. In the next step, attention model and linguistic knowledge can be combined into the model, which makes the network model more targeted to the emotional characteristics and more effective in learning the emotional characteristics of the text.

Acknowledgement

Scientific Research Projects of Hunan Education Department

References

[1] Li R, Lin Z, Lin H L, et al. Text emotion analysis: a survey[J]. Journal of Computer Research and Development, 2018, 55(1): 30-52.

[2] Zhao YY, Qin B, Shi Q H, et al. Large-scale sentiment lexiconcollection and its application in sentiment classification[J].Journal of Chinese Information Processing, 2017, 31(2):187-193.

[3] Wang C. Research on the analysis of weibo comments tendency based on machine learning methods [D]. Harbin: Harbin In-stitute of Technology, 2016.