Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 3, Issue 3: 129-137, DOI: 10.25236/AJCIS.2020.030314

Text Emotion Detection Based on Bi- LSTM Network

Zihe Wang*

Beijing No.4 High School International Campus, Beijing, China
**Corresponding author e-mail: hzhc_wjdi@163.com*

ABSTRACT. *Emotion detection and opinion mining in social network is a hot research topic nowadays. Most existing studies and research, however, have been using typical binary (positive/negative) or ternary (positive / negative / neutral) classification. Classifying short sequences of text into multi subclasses is relatively rarely reported. Except Bouazizi and Ohtsuki (2017), the accuracy rate of detection in their two studies was only 56.9% and 60.2%. In our study, we proposed a Bi-directional Long Short-Term Memory with Language Model (BiLSTM-LM) to classify text sequences into seven distinct emotional classes. Results showed that the accuracy rate of detection based on our model can reach as high as 64.09% on multi-class classification, which is 4 percentage points higher than the most advanced model in the world to date.*

KEYWORDS: *Emotion detection, opinion mining, LSTM*

1. Introduction

With the popularity of social networking and web reviews, sentiment analysis has quickly become a new research topic and has been paid more and more attention in the aspects of research and commercial application. In addition, it has also produced a lot of value. The sentiment analysis task is a basic and significant task in natural language processing. It is the process of analyzing, processing, concluding and reasoning subjective texts with emotional color, and finds out the tendency expression of users to the goods and the attention to hot events. According to the emotion model theory [1], all emotions can be divided into 3 categories, 6 groups and 22 types. Although there are many emotion theories, the literature research [2] finds that there are six basic emotions in human beings, that is, happiness, sadness, fear, surprise, anger and jealousy. Other emotions are combined with these six basic emotions, which belong to the complex emotion.

The research on emotion all focuses on the generation mechanism of emotion, such as the influencing factors of emotion, emotion classification, the influence of emotion on people and so on, but there are few research on the human emotion

ISSN 2616-5775 Vol. 3, Issue 3: 129-137, DOI: 10.25236/AJCIS.2020.030314

judgement [3]. The current research on emotion judgment lays emphasis on the study of expression, micro-expression and physical behavior, and carries on emotion judgement according to people's expression and behavior [4].

There are few studies on extracting emotional information from text input. The literature [9] develops a semantic network to extract emotions from text contents, but there are few corpus to support this result. For the problem of text emotion analysis, this paper proposes to use Bi-directional Long Short-Term Memory(Bi-LSTM) combined with Language Model(LM), that is BiLSTM-LM. LSTM solves the vanishing gradient of RNN through a gate mechanism, and can effectively learn the long-term dependency. Combined with the language model, the language features of mass data are obtained. The convergence speed of training is accelerated, and the generalization ability of the model is improved. Through the experimental comparative analysis and recognition of 6 basic emotions, the BiLSTM-LM model proposed in this paper is significantly improved compared with other deep learning models.

2. Basic Theory

2.1 Long Short-Term Memory

Long short-term memory(LSTM) is a time recurrent neural network, which is suitable for processing and predicting important events with relatively long interval and delay in time series. LSTM is proposed to solve the problem of "vanishing gradient" in the recurrent neural network (RNN), and it is a special recurrent neural network.

The standard RNN structure has a chained form of repetitive neural network module, which is generally a tanh layer for repetitive learning(figure 1). But in LSTM(figure 2), there are four special structures in the repetitive module. The horizontal line running through the top of the figure is the state of cell (cell), and yellow matrix is the neural network layer through learning. The pink circle represents the operation, and the black arrow represents the vector transport. As a whole, not only h but also the state of cell c flows with the time and the state of cell c represents the long-term memory.

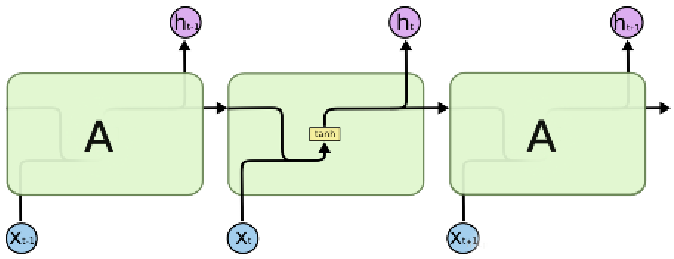

Figure. 1 Basic structure of RNN network

ISSN 2616-5775 Vol. 3, Issue 3: 129-137, DOI: 10.25236/AJCIS.2020.030314

The reason why LSTM can remember long-term information is the designed "gate" structure, which is an approach for the selective pass of information. In LSTM, the first stage is the forgetting gate, which determines the information needed to be forgotten from the cell state. The next stage is the input gate, which determines the new information that can be stored in the cell state. The final stage is the output gate, which determines the output value.

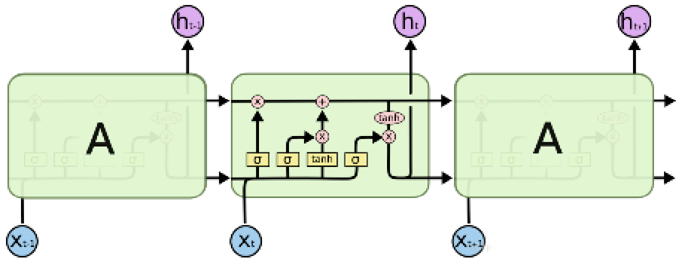

Figure. 2 Basic structure of the LSTM network

The Bi-directional RNN consists of two common RNN, a positive RNN, that uses previous information, and a reversed RNN, that uses future information. Thus at the time of t, the information not only at t-1 but also at t+1 can be used. Generally speaking, because bi-directional LSTM can make use of the previous and future information at the same time, it will be more accurate in final prediction than the mono-directional LSTM. Figure 3 shows the structure of bi-directional LSTM.

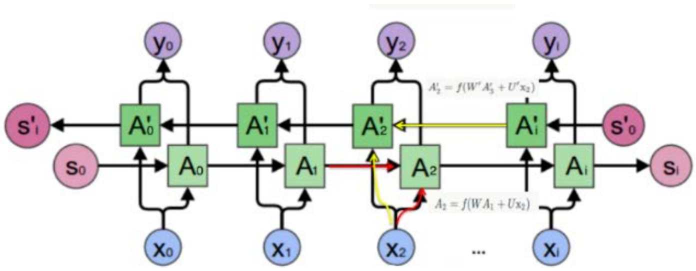

Figure. 3 Bi-directional LSTM network structure

$A_1 \to A_2 \to \cdots \to A_i$ is a positive RNN that participates in the positive calculation. The input at the time of t is the sequential date x_t at the time of t and the output $A_{t\text{-}1}$ at the time of t-1. $A_1 \to A_2 \to \cdots \to A_i$ is a reversed RNN that participates in the inverse calculation. The input at the time of t is the sequential

ISSN 2616-5775 Vol. 3, Issue 3: 129-137, DOI: 10.25236/AJCIS.2020.030314

date x_t at the time of t and the output A'_{t+1} at the time of t+1. The final output value at the time of t depends on A_{t-1} and A'_{t+1}.

2.2 Word Embedding

Word Embedding is an important concept in natural language processing (NLP). WordEmbedding can be used to transform a word into the vector representation with fixed length, which is convenient for mathematical treatment.

The first step of using the mathematical model to process text corpus is to transform the text into the mathematical representation, which has two methods. The first method is that a word can be represented through one-hot matrix. One-hot matrix refers to the matrix that only has an element of 1 in each row while other elements of 0. Each word in the dictionary is assigned a number. When encoding a sentence, you can only transform each word into the one-hot matrix with the corresponding position of 1 of this word in the dictionary. For example, if you intend to express "the cat sat on the mat", you can use the matrix shown in figure 4.

$$\begin{pmatrix} the \\ cat \\ sat \\ on \\ the \\ mat \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Figure. 4 One-dot coding method

The one-hot representation is intuitive, but has two disadvantages. First, each dimensional length of the matrix is the length of the dictionary. For example, the dictionary contains 10000 words, so the one-hot vector corresponding to each word is the vector of 1X10000, and this vector has only one position of 1, while others of 0, which wastes the space and is not conductive to the calculation. Second, the one-hot matrix simply numbers each word, but the relationship between words is completely invisible. For example, the correlation between "cat" and "mouse" is higher than that of "cat" and "cellphone", which is not reflected in one-hot representation.

The Word Embedding matrix assigns a fixed-length vector to each word, which can be set by itself, such as 300. In fact, it is much lower than the dictionary length(such as 10000). The included angle value between the two word vectors can be used as a measure of their relationship, as shown in figure 5.

ISSN 2616-5775 Vol. 3, Issue 3: 129-137, DOI: 10.25236/AJCIS.2020.030314

anarchism	0.5	0.1	−0.1
originated	−0.5	0.3	0.9
as	0.3	−0.5	−0.3
a	0.7	0.2	−0.3
term	0.8	0.1	−0.1
of	0.4	−0.6	−0.1
abuse	0.7	0.1	−0.4

Figure. 5 WordEmbedding coding scheme

By means of a cosine function, the correlation between the two words can be calculated, which is simple and efficient:

$$similarity = \cos(\theta) = \frac{A \cdot B}{\| A \|_2 \| B \|_2} \tag{1}$$

Due to the characteristics of saving space and calculation convenience of Word Embedding, it is widely used in NLP field.

3. Recognition Algorithm

3.1 Network Model

In order to generate words for emotional detection from the context, this paper creates a lexical library containing 65620 words. All words in lexical vocabulary are divided into two categories: content words and EFWs. If the word type is an emotional keyword, it contains six emotional state tags (happiness sadness, anger, surprise, hatred and fear) and corresponding weights. For modifiers and metaphoric words, they are more related to exaggerated emotion or introverted expression. The whole dictionary is shown in Figure 6.

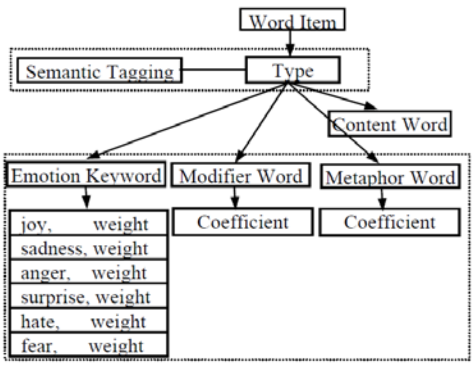

Figure. 6 Text emotion recognition model

ISSN 2616-5775 Vol. 3, Issue 3: 129-137, DOI: 10.25236/AJCIS.2020.030314

In order to make full use of the word features and word feature information of the text, a two-channel LSTM neural network model is proposed in this paper. The model combines word feature information and word feature information of the text, and can learn the deeper feature representation of the text. The structure of the model is shown in Figure 7.

The words are encoded by using WordEmbedding, and the text is represented as the word feature representation $\mathbf{T}_w$ and character feature representation $\mathbf{T}_c$. Then bi-directional LSTM is used to learn the hidden layer features of higher dimensions from two sets of feature representations.

$$\begin{aligned} \mathbf{h}_w &= LSTM_w(\mathbf{T}_w) \\ \mathbf{h}_c &= LSTM_c(\mathbf{T}_c) \end{aligned} \tag{2}$$

Where $\mathbf{h}_w$ and $\mathbf{h}_c$ represent word hidden layer features and character hidden layer features learned by LSTM

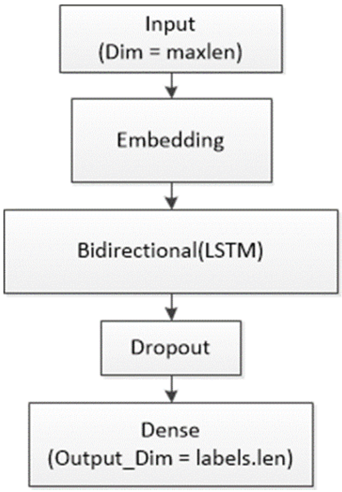

Figure. 7 Text emotion recognition model

In order to obtain the word-character integration feature of the text, the splicing integration of $\mathbf{h}_w$ and $\mathbf{h}_c$ is carried out.

$$\mathbf{h}_{wc} = \mathbf{h}_w \oplus \mathbf{h}_c \tag{3}$$

Where $\mathbf{h}_{wc}$ is the word-character integration feature of the text, and the symbol $\oplus$ represents the splicing of vectors.

Then there is a fully connected layer, which is used to learn the deeper representation of word-character integration feature. In order to reduce the

ISSN 2616-5775 Vol. 3, Issue 3: 129-137, DOI: 10.25236/AJCIS.2020.030314

complexity of the model and prevent the network from overfitting the training samples, the Dropout layer is added above the fully connected layer. The Dropout layer can randomly make some hidden layer nodes in the network not work when the model is trained.

Finally, the model completes the emotion recognition task of text through sigmoid layer. The Sigmoid layer accepts the output of the previous layer as the input, and the output length is 1. The value of the vector is mapped to 0 - 1 by Sigmoid function, which is used as the prediction probability of whether the text contains some emotion or not.

In the training process, the back propagation algorithm (BP) is used to update the weights.

4. Experimental Result Analysis

The CPU of the experimental platform used in this paper is Intel Core i7 / 8700, and the graphics card model is NVIDIA GeForce GTX 1080Ti. The video memory capacity is 8GB and the memory is 64GB. The Tensorflow deep learning framework is used for development and test, and CUDA is used for parallel acceleration of the GPU training process. Other specific configuration parameters are shown in Table 1.

Table 1 Parameter configuration of experimental platform

Configuration	Parameter
GPU	NVIDIA GeForce GTX 1080Ti 8GB GDDR5X
CPU	Intel(R) Core (TM) i7-8700 CPU@3.2-4.6GHz
RAM	64GB DDR4 2133MHz
Operating system	Linux Ubuntu 14.0
Development platform	Anaconda3
Development language	Python 3.5
Framework	Tensorflow 10.0

The network model shown in figure 7 is used to train 37617 texts and recognize the six basic emotions of happiness, sadness, fear, surprise, anger, and jealousy. The changing curve of loss and recognition accuracy in the training process with the epoch are shown in Figure 8 and 9.

ISSN 2616-5775 Vol. 3, Issue 3: 129-137, DOI: 10.25236/AJCIS.2020.030314

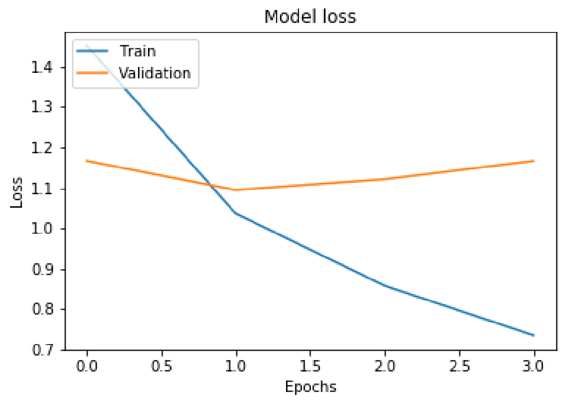

Figure. 8 Training loss curve

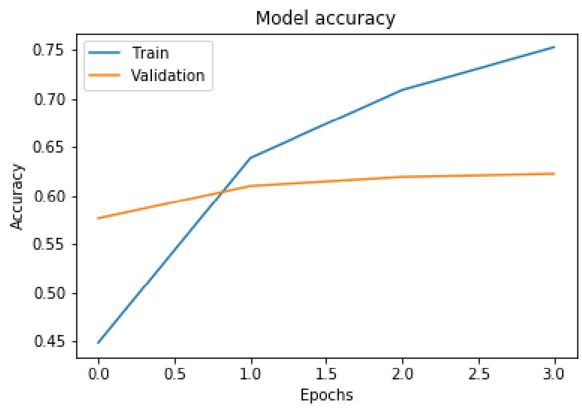

Figure. 9 Training accuracy curve

As you can see from the figure, the loss can be stable by using few epoch for the verification set. The changing curve of recognition rate with epoch is consistent with loss and becomes stable after 2 epoch. The recognition rate can reach 64%, which has a certain improvement compared with the highest 60% recognition rate of text emotion recognition, and it has rapid convergence.

5. Conclusion

This paper proposes a bi-directional LSTM language model for English short text classification task based on text emotion recognition. This paper uses the WordEmbedding to encode words, express text as word feature representation and

ISSN 2616-5775 Vol. 3, Issue 3: 129-137, DOI: 10.25236/AJCIS.2020.030314

character feature representation, and combine content words with emotion functional words to estimate the final emotional output. The experimental results show that the multi-classification detection accuracy of this model is as high as 64.09%, and the convergence speed is fast. The English text can be effectively classified according to six basic emotional features.

References

[1] J. Cahn, “Generating Expression in Synthesized Speech,” Master’s thesis, MIT, 1989.

[2] Keikichi Hirose, Nobuaki Minematsu, etc, “Analytical and perceptual study on the role of acoustic features in realizing emotional speech”, ICSLP2000.

[3] Donna Erickson, 1 Arthur Abramson, “Articulatory characteristics of emotional utterances in spoken English”, ICSLP2000

[4] Schröder M1, Cowie R2, Douglas-Cowie E2, “Acoustic Correlates of Emotion Dimensions in View of Speech Synthesis”, Eurospeech2001

[5] K. Nurzynska and B. Smolka, "Recognition between smiling and neutral facial display with power LBP operator," IEEE EUROCON 2015 - International Conference on Computer as a Tool (EUROCON), Salamanca, 2015, pp. 1-6.

[6] T. Kiran and T. Kushal, "Facial expression classification using Support Vector Machine based on bidirectional Local Binary Pattern Histogram feature descriptor," 2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Shanghai, 2016, pp. 115-120.

[7] C. Cheng, X. Wei and Z. Jian, "Emotion recognition algorithm based on convolution neural network," 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nanjing, 2017, pp. 1-5.

[8] S. Basu, J. Chakraborty and M. Aftabuddin, "Emotion recognition from speech using convolutional neural network with recurrent neural network architecture," 2017 2nd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, 2017, pp. 333-336.

[9] Ze-Jing Chuang and Chung-Hsien Wu, “Emotion recognition from textual input using an emotional semantic network”, ICSLP2002, Denver.

[10] Feng Xiang, Qiu Longhui, Guo Xiaoran. A Method of Student Feedback Text Recognition Based on LSTM Model [J]. Open Education Research, 2019, 25 (02): 114-120.

[11] Taboada M, Brooke J, Tofiloski M, et al. Lexicon-Based Methods for Sentiment Analysis [J]. Computational Linguistics, 2011, 37 (2): 267-307.

[12] Cheng Bo, Li Weihong, Tong Tong. Chinese hierarchical address segmentation based on BiLSTM-CRF [J/OL]. Journal of Geo-Information Science, 2019 (08): 1-9.