ISSN: 2414-1895 DOI: 10.6919/ICJE.201909_5(10).0019

# Emotional analysis of neural network text combined with attention mechanism

Xueming Zhai [a], Wei Wei [b]

School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China.

[a]15103115775@163.com, [b]1012438684@qq.com

## Abstract

**In order to solve the problems of ignoring key words, insufficient learning of context information and gradient dispersion in the current neural network model, this paper proposes a text emotion analysis model of neural network based on attention mechanism.In this paper, word2vec is combined with term frequency-inverse document frequency (TFIDF) algorithm, and attention mechanism is added into the gated loop unit (GRU) network model to highlight the role of keywords in the text through weighted probability.At the same time, Maxout neuron is introduced at the output end of the model, which can effectively alleviate the gradient dispersion problem.The model in this paper is tested on IMDB English data set. The experimental results show that the model in this paper has a good effect on IMDB data set and can achieve better accuracy and F value compared with other classification models.**

## Keywords

**Emotion analysis, maxout neurons, TFIDF , GRU.**

## 1. Introduction

Emotional analysis is a process of analyzing, processing, concluding and reasoning subjective texts with emotional colors by computer to help people quickly obtain and sort out opinion information on the Internet [1]. At present, emotion analysis has become an important task in industry and academia, and the research of emotion analysis has greatly promoted the development of artificial intelligence and all the research results resulting from it.

The most important technique in the study of emotion analysis is emotion classification technique. At present, the mainstream emotion classification technology mainly falls into the following two types: the method based on emotion dictionary, the method based on artificial extraction of features classification, and the method based on deep learning. The emotional lexicography-based method is mainly used to divide words into parts of speech, find out the words of different parts of speech and calculate their corresponding scores. This method relies too much on the dictionary of emotion and has a serious domain, so the effect is not ideal [2]. The method based on manual feature extraction is a method based on traditional machine learning, which requires a large amount of manual annotation data. Machine learning algorithms such as support vector machine (SVM) and naive bayes (naive bayes) are adopted to classify emotions. In recent years, methods based on deep learning have attracted the attention of scholars. The deep learning model can fully explore the emotional information of the text without manual annotation, thus achieving a good classification effect.

Most existing deep learning methods transform text into word vectors after word segmentation, and do not pay enough attention to the key words and context information in the text. At the same time, the problem of gradient dispersion is often ignored by people. To solve the above problems, this paper

ISSN: 2414-1895 DOI: 10.6919/ICJE.201909_5(10).0019

proposes a wt-ag-m emotion analysis model. Firstly, preprocess the input text, construct a vector for each word using word2vec, and assign different weights to each word vector through TFIDF algorithm to form the input of GRU. Add Attention mechanism on the basis of GRU network model to highlight the role of keywords; In order to solve the gradient dispersion problem in stochastic gradient reduction algorithm, Maxout neuron is introduced at the output end of the network model. Through experiments on IMDB, a public data set, compared with the existing results, it is proved that the model constructed in this paper is helpful to improve the accuracy of emotional analysis.

## 2. Analysis of existing problems

### 2.1 Word vector representation.

Before word2vec, some people used neural networks to train word vectors.Bengio[3] first used multi-layer neural network to build a language model. However, because the vocabulary is generally above the million level, the processing process of the model is very time-consuming, which means that the output probability of each word calculated by the output layer of the model is very large.The biggest feature of the log-bilinear model proposed by Mikolov[4] is that it eliminates the hidden layer of traditional neural network and USES its linear representation ability to calculate the real number of words, which is used to represent word vectors.

Word2vec trains n-element (n-gram) language model through neural network machine learning algorithm, and finds the corresponding vector of words in the training process.The word2vec model generates vectors that are not only low dimensional but also carry semantic information about the context, compared to the vectors generated by previous text representation methods.However, word2vec model itself cannot calculate the word weight in the text, so this paper USES TFIDF algorithm to weight word2vec word vector.

### 2.2 Emotional analysis.

Deep learning can automatically select features to better obtain semantic relations and emotional information of sentences.Kim[5] first proposed to apply CNN to emotion analysis tasks, and achieved good results.Conneau et al. [6] adopted a deep convolutional network method and proposed the VDCNN model to improve the accuracy of emotion classification task.However, CNN can only mine partial information of text, and it lacks the effect of capturing long-distance dependence.Circular neural network (RNN) makes up for this deficiency.Tang et al. [7] built a model of text-level text and proposed a hierarchical RNN model.Although RNN is suitable for context processing, it produces a gradient explosion when dealing with long-distance dependency problems.To solve this problem, Hochreiter et al. [8] proposed LSTM model and optimized the internal structure of RNN.Zhu et al. [9] used LSTM to model the text and divided it into word sequences for emotional classification.Traditional LSTM can only make effective use of the above information and ignore the following information, which affects the accuracy of emotional classification to some extent.

To solve the gradient disappearance problem of text feature selection, the wt-ag-m model proposed in this paper solves the long-term dependence problem in RNN model training.At the same time, this paper adds Attention mechanism on the basis of GRU model, reducing the possibility of ignoring important information or adding unnecessary information in the process of feature vector extraction, and effectively highlighting the role of keywords in the text.

## 3. WT-AG-M model

### 3.1 Model framework.

The existing deep learning methods do not pay enough attention to the key words and contextual information in the text in the task of emotional analysis of text-level text.At the same time, with the increase of the number of iterations of neural network, the influence of overfitting and gradient dispersion on the classification results of emotion gradually expands.To solve the above problems, this paper proposes a wt-ag-m emotion analysis model.The framework of wt-ag-m model is shown in

ISSN: 2414-1895 DOI: 10.6919/ICJE.201909_5(10).0019

figure 1.Firstly, the input comment text is preprocessed. Word2vec is used to construct a vector for each word. Through TFIDF algorithm, different weights are assigned to each word vector to form the input of GRU.Then, the GRU network which adds Attention mechanism is used to train feature vectors, so that the output feature vectors contain semantic features and word sequence features.At last, the feature vector is input into Maxout neurons in the full connection layer, and Softmax regression classifier is used in the output layer to classify emotions.

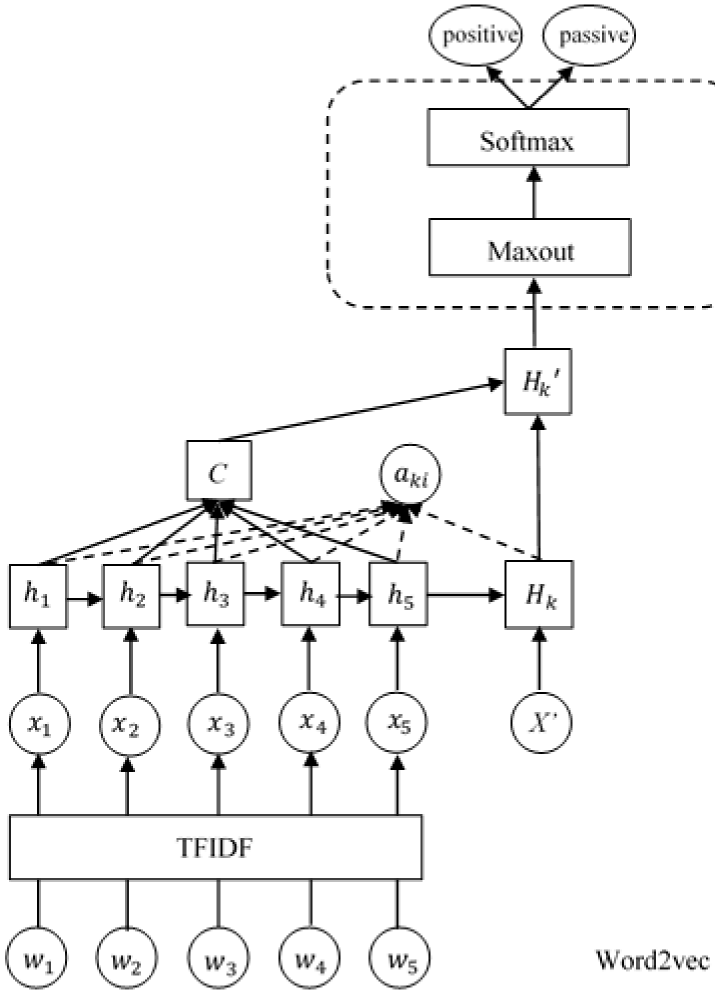


Fig.1 Schematic diagram of Basic Process

### 3.2 Combined with TFIDF algorithm and Word2vec.

The word vectors generated by the word2vec model contain semantic information about the context compared to other methods of text representation. However, the word2vec model cannot recognize which words are relatively important in the text, so the TFIDF algorithm can calculate the weight of the feature items. TFIDF takes into account the probability of each word appearing in a single document *p*, and the weight of that word in the entire document set, *l*. Therefore, this paper uses word2vec combined with TFIDF algorithm to extract word vectors.

The text set *M* contains *K* text, assuming that the text $M_i(i = 1,2, \dots, K)$ word segmentation has been done, the word segmentation text is trained through word2vec model, and the corresponding n-dimensional word vector $w = [v_1, v_2, \dots, v_n]$ is generated. TFIDF algorithm is used to calculate the word t in the text $M_i(i = 1,2, \dots, K)$ and $T(\mathrm{t},\ M_i)$. The formula for calculating the probability *p* and weight *l* of the word t is as follows:

$$p(\mathrm{t}, M_i) = \frac{n_t}{n_{M_i}} \tag{1}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.201909_5(10).0019

$$l(t) = log\left(\frac{K}{n_t} + 0.01\right) \tag{2}$$

In the formula: $n_t$ represents the word *t* in the text $M_i(i = 1,2, \ldots, K)$, and $n_{M_i}$ represents the total number of words in the text $M_i$. *K* is the total number of training texts, and $n_t$ is the number of texts in which the word *t* appears in the training text set. Weight value

$$T(\mathrm{t}, M_i) = \frac{p(\mathrm{t}, M_i) * l(t)}{\sqrt{\sum_{t \epsilon M_i}(p(\mathrm{t}, M_i) * l(t))^2}} \tag{3}$$

In the formula: $p(\mathrm{t}, M_i)$ is the word frequency of the word *t* in the text. For each text $M_i$, its text vector

$$x_i = \sum_{t \epsilon M_i} w_t * T(\mathrm{t}, M_i) \tag{4}$$

In the formula: $w_t$ represents the word vector of the word *t*.

### 3.3 GRU based on Attention.

Compared with LSTM, GRU model can not only solve the problem of different time dependence, but also has simpler structure, fewer parameters and better convergence. The loop network unit of the GRU consists of two gate components, the reset gate r and the update gate z. The calculation method of reset gate is as follows:

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \tag{5}$$

The larger the update threshold, the greater the impact. The calculation method is as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{6}$$

GRU helped $h_t$ cell activation values at the same time by time *t-1* activation values $h_{t-1}$, the candidate activation value $\tilde{h}_t$ and *z* control and update the door. The calculation methods of the following, ⊙ expressed as elements multiplication.

$$\tilde{h}_t = \tanh(\mathrm{W} x_t + \mathrm{U}(r_t \odot h_{t-1})) \tag{7}$$

$$h_t^j = (1 - z_t^j) h_{t-1}^j + z_t^j \tilde{h}_t^j \tag{8}$$

The Attention mechanism reflects the relationship between encoder input Attention position and decoder output Attention weight. The Attention mechanism-based GRU model retains the intermediate output results of the input sequence of the GRU encoder, and selectively learns these intermediate output results through training, and correlates the output sequence with the model output.

## 4. Experiment and analysis

### 4.1 Experimental data set.

In order to verify the validity of the joint model, open IMDB is used as the data set in the experiment. This data set is mainly about product reviews, which are all short and suitable for the task of emotion

ISSN: 2414-1895 DOI: 10.6919/ICJE.201909_5(10).0019

analysis at the level of text. The training set and the test set were divided according to the ratio of 8:2, which were used to train the emotional classification model and evaluate the effect respectively. The performance of the model was evaluated by repeated experiments and the average value of the experiment was selected as the final result.

### 4.2 The evaluation index.

In this paper, accuracy rate A, recall rate R and F value are used as evaluation indexes of the experiment. Accuracy A represents the percentage of test sets classified correctly in A certain category after classification. The recall rate R represents the proportion of the correct number predicted as a certain category after classification to all the real Numbers of that category. F value is the weighted harmonic average of accuracy A and recall rate R. The specific calculation method is as follows:

$$\text{Accuracy A} = \frac{m}{m+l} \tag{9}$$

$$\text{Recall rate R} = \frac{m}{m+n} \tag{10}$$

$$\text{F} = \frac{m}{m+l} \tag{11}$$

Table 1 is the discriminant confusion matrix established according to the classification results, which introduces the meanings represented by each letter in the above evaluation indexes.

Table 1 Classification discrimination confusion matrix

| Real results | Predict results | |
|---|---|---|
| | Belongs to category L | Not belongs to category L |
| Belongs to category L | m | n |
| Not belongs to category L | l | p |

### 4.3 Experimental parameters.

The experimental environment in this paper is Win10 system with 8G of memory. JetBrains PyCharm software and Keras deep learning framework are used to build the network model, and the TensorFlow architecture is at the bottom. The activation function used in the network model was Tanh, Adam was adopted as the gradient update method, and the learning rate was set as 0.05.The dimension of word vector is a main parameter to be adjusted. If the dimension is too high, the model is prone to overfitting. The dimensions are too low to contain all the information needed. In the case of a single variable, F value reaches the optimal value when the word vector dimension is 80. Therefore, 80 is selected as the set value of the vector dimension in the training of neural network.

### 4.4 Experimental results and comparative analysis.

In order to verify the classification performance of the proposed emotional analysis model based on neural network and conditional random field association, four groups of experiments were designed to compare the models.

Group 1: compare and analyze the model in this paper with the bi-gru single model. It is proved that the effect of emotion analysis of the model in this paper is better on the same data set.

Group 2: the model in this paper is compared with the GRU model with attention mechanism added. The same data set is input into the model, and the emotional probability of the whole sentence is selected according to the emotional probability of each word in the sentence, which proves that the emotional analysis effect of this model is better.

Group 3: the GRU classification model combined with W2C and TFIDF was compared and analyzed. Compared with the model in this paper, this model USES Softmax to classify directly instead of

ISSN: 2414-1895 DOI: 10.6919/ICJE.201909_5(10).0019

adding Maxout classifier, which proves that the model classification method proposed in this paper is more effective.

The test results of the comparison experiment are shown in table 2.Compared with GRU single model, GRU model combined with attention mechanism, and GRU classification model combined with W2C and TFIDF, the model in this paper shows good advantages in accuracy, recall rate, F value and other indicators.Experiments show that this model does have better effect on emotion analysis tasks than single model.

Table 2 Test results of different models

| Model | index A | index R | index F |
|---|---|---|---|
| BiGRU | 0.8936 | 0.8839 | 0.8874 |
| Att-BiGRU | 0.9055 | 0.9172 | 0.9038 |
| WT-GRU | 0.9241 | 0.9127 | 0.9268 |
| WT-AG-M | 0.9389 | 0.9349 | 0.9326 |

## 5. Conclusion

This paper analyzes the problems existing in the traditional neural network model and puts forward a kind of neural network text emotion analysis model combined with multilevel attention mechanism. Word2vec is combined with the term frequency inverse document frequency (TFIDF) algorithm to extract more semantic information and structural features from sentences. At the same time, the gated loop unit (GRU), which introduces attention mechanism, is used as the input of Maxout neuron classifier to realize the task of emotion classification and to predict the emotional category of chapters. The model in this paper enriches the features learned. Training and testing on IMDB English data set, generally speaking, the method in this paper achieves good results.

## References

[1] Chen Long, Guan Yu, He Jinhong, et al. Progress in sentiment classification research [J]. Journal of Computer Research and Development, 2017, 54 (06): 1150-1170.

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

[3] Bengio Y, Ducharme R, Vincent P. A neural probabilistic language model [J]. Journal of Machine Learning Research, 2003, 3(2):1137-1155.

[4] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space [DB/OL]. https://arxiv.org/pdf/1301.3871.pdf,2019-05-07.

[5] Kim Y, Convolutional neural networks for sentence classification [C] //Proceedings of EMNLP, 2014.

[6] Conneau, et al. Very deep convolutional networks for text classification [C] //Proceedings of the 15 th Conference of the European Chapter of the Association for Computational Linguistics, 2017.

[7] Tang D, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification [C] //Proceedings of EMNLP, 2015:1422-1432.

[8] Hochreiter S, Schmidhuber J. Long Short-term Memory [J]. Neural Computation, 1997, 9(8): 1735-1780.

[9] Zhu Xiaodan, Parinaz S, Guo Hongyu. Long short-term memory over recursive structures [C]// Proc of Int Conf on Machine Learning, New York: ACM Press, 2015: 1604-1612.