划词翻译现已支持百度翻译（专业版）和阿里翻译（专业版），覆盖电子科技、水利机械、电商、医药、金融等多个领域。 不再提示

News Text Classification Based on Improved Bi-LSTM-CNN

Chenbin Li, Guohua Zhan, Zhihua Li
Hangzhou Institute of Service Engineering
Hangzhou Normal University
Hangzhou, China
licboct@163.com, ghzhan@hznu.edu.cn, zhihuali_e@163.com

***Abstract*— The traditional text classification methods are based on machine learning. It requires a large amount of artificially labeled training data as well as human participation. However, it is common that ignoring the contextual information and the word order information in such a way, and often exist some problems such as data sparseness and latitudinal explosion. With the development of deep learning, many researchers have also been using deep learning in text classification. This paper investigates the application issue of NLP in text classification by using the Bi-LSTM-CNN method. For the purpose of improving the accuracy of text classification, a kind of comprehensive expression is employed to accurately express semantics. The experiment shows that the model in this paper has great advantages in the classification of news texts.**

Keywords-component; text classification; Bi-LSTM-CNN; word order; text semantics

I. INTRODUCTION

In the past few years, the explosive news information faces the problem that the demand for getting news from a large number of users. The effective classification of news largely solves the mess of news information, enabling users to locate news types more accurately and faster. Text classification is an important application in the NLP field. It refers to the classification of texts into pre-defined categories. In today's Internet age, text data has become one of the most common forms of data, such as user reviews, news, emails, and more. The basic process of text classification generally includes: text preprocessing, feature extraction, text representation and classifier training [1]. The traditional method of feature representation usually ignores the context. The order of information text in the text is still unsatisfactory for the semantics of capturing words. At the same time, there are problems such as data sparseness and dimension explosion in the feature extraction method, which will reduce the training model's generalization ability.

Deep learning once achieved remarkable achievements in the field of Speech Recognition and Image Processing [2], but it is gradually developing in natural language processing. Deep learning technology has gradually replaced traditional machine learning methods and has become a mainstream technology in text classification [3]. Deep learning can express objects more accurately, and it can automatically acquire the features of objects from massive data. The deep learning model based on such functional attributes includes Convolutional Neural Network (CNN) [4], Recurrent Neural Network [5], and Recursive Neural Network [6].

How to efficiently classify these massive text data has caused experts to study. In [7], the support vector machine (SVM) algorithm is used to optimize the parameters of the text classifier, thus improving the classification accuracy of the text classifier. In [8], the traditional machine learning method is used for the classification. The TF-IDF model is used to extract the category keywords, and the cosine similarity calculation is performed by these category keywords and the text keywords to be classified. In [9] proposed text classification based on implicit Dirichlet distribution LDA model and SVM algorithm, but in a large number of short texts, the classification effect is not good due to short text length and excessive noise. In [10], the feature text is extracted from the news text by convolutional neural network to classify the text. However, although such a method can extract features well, the context is usually ignored and the text semantics are not accurate enough.

Based on the above considerations, this paper uses Bi-LSTM-CNN to solve the classification problem of large-scale news text. In order to better extract the characteristics of the text. In this paper, the Bi-LSTM model is used to obtain the representation of two directions, and then the two directions representation is combined into a new expression through the convolutional neural network. Each word expression is added by itself to the left text vector and the right text vector to indicate. For the left and right texts, a loop structure is used, which is a non-linear transformation of the previous word and a text on the left side. This approach preserves contextual information and a wider range of word order better.

The experiments in this paper were conducted using a subset of the THUCNews for training and testing, and were compared using a variety of benchmark models. Experiments show that the model in this paper is the best compared to other models.

II. MODEL ANALYSIS

The model in this paper is mainly composed of a Bi-LSTM layer composed of a word vector and a left and right context, a local feature, a global feature and a softmax layer. At the same time, the word vector is trained by Word2Vec.

2474-3828/18/$31.00 ©2018 IEEE
DOI 10.1109/ITME.2018.00199

IEEE computer society

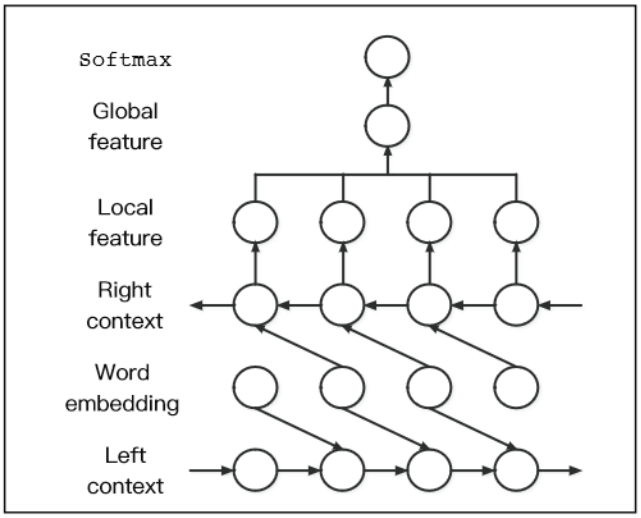

Figure 1. Neural Network Architecture.

A. Word vector training based on Word2Vec

Word2vec uses the n-gram model, which assumes that a word is only related to the surrounding n words, and has nothing to do with other words. Word2vec has two models: Continuous Bag-of-word (CBOW) [11] and Skip-gram model, which can transform text words into word vectors, make assumptions and models of natural languages, and let computers understand natural languages.

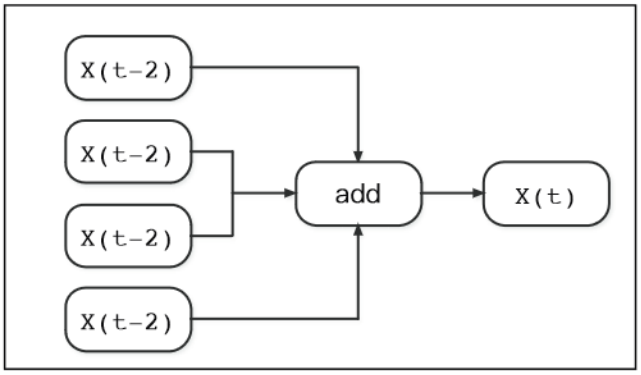

Figure 2. The model of Bag-of-word.

This paper uses the CBOW model (see Figure 1) for the training of word vectors [12], the input layer is 2a (a = 2) word vectors in the context of the word x_i, and the projection layer vector S_x is the 2a word vector's sum. The output layer is a Huffman tree whose weight is the number of occurrences of the corpus. Using the Hierarchical softmax algorithm, each word in the corpus has a unique path from the root node. Each edge in the path has a corresponding weight. The weights of the paths are combined to obtain the encoding of each word. Using the phrase sequence for x_1,x_2,……,x_T, the optimization objective function of CBOW is, as in Equation (1), optimized with gradient descent.

$$\mathrm{p(s)} = \mathrm{p}(x_1, x_2, \ldots, x_i) = \prod_{i=1}^{T} p(x_i|Content_i) \quad (1)$$

B. Context-based Bi-LSTM-CNN

Today, Recurrent Neural Network are used effectively in the field of natural language processing. The Recurrent Neural Network can fully utilize the word order information and semantically synthesize each word to achieve more precise semantic expression [13]. Based on the principle of [13], this paper further optimizes the model to change the unidirectional LSTM layer to a bidirectional LSTM layer as its recurrent neural network. The forward LSTM refers to the sequential reading corpus and the backward LSTM refers to reading corpus in reverse order. Such a structure not only considers the forward semantics, but also considers the reverse order semantics, which greatly improves the expression of text semantics. The word vector at each position has an expression obtained by LSTM in two directions, and the left and right context of each word is expressed as shown in Equations (2) and (3).

$$c_l(x_i) = g(w_l c_l(x_{i-1}) + w_{sl} E(x_{i-1})) \quad (2)$$

$$c_r(x_i) = g(w_r c_r(x_{i+1}) + w_{sr} E(x_{i+1})) \quad (3)$$

$c_l(x_i)$ is x_i word left context vector, w_l is a matrix that converts the hidden layer to the next hidden layer, w_{sl} is a matrix that combines the semantics of the current word with the left context of the next word, $c_l(x_{i-1})$ is left context of the previous word, $E(w_{i-1})$ is word vector of the previous word. Similarly, the vector $c_r(x_i)$ of the right context of the x_i word can be concluded.

Through this Bi-LSTM layer, using convolution to combine the left and right contexts of the current location to form a new expression of the fusion context word. As shown in the equation (4), there is a concatenation of the left context vector $c_l(x_i)$, the current word's word vector $\mathrm{E}(x_i)$, and the right context vector $c_r(x_i)$. With such an expression, the semantics of the text can be more accurately represented, and the ambiguity of the x_i word can be better eliminated.

$$x_i = (c_l(x_i), \mathrm{E}(x_i), c_r(x_i)) \quad (4)$$

After getting the new expression of the x_i word, applying the linear transformation to the tanh activation function, as in equation (5), to x_i and send the result to the max-pooling layer. The pool layer converts text of different lengths into fixed-length vectors, and can capture information throughout the text.

$$y_i = \tanh(w_i x_i + b_i) \quad (5)$$

Finally, after the softmax function is processed, the output number is converted into a probability, which is converted into a corresponding probability problem, as in Equation (6).

$$p_i = \frac{\exp(y_i)}{\sum_{k=1}^{n} \exp(y_k)} \quad (6)$$

III. EXPERIMENT AND ANALYSIS

A. Data Set

The data set used in this experiment is a subset of THUCNews for training and testing. It selects news from the ten categories of sports, finance, real estate, home, education, technology, fashion, politics, games and entertainment as experimental data. The corpus has a total of 65,000 corpora,

of which the training set includes 50,000 corpora, the verification set includes 5000 corpora, and the test set includes 10,000.

B. *Optimization Method And Parameter Setting*

In the experiment, we did some experiments on the hidden state dimension of the bidirectional LSTM layer, and concluded that the best effect was set to 100 dimensions. In this paper, using the random gradient descent algorithm to update parameters. Each time every experiment uses 128 samples, the initial learning rate is set to 0.8, and the Dropout ratio is 0.5. The experiment is based on the TensorFlow framework.

C. *Comparison of experimental results*

In this paper, the classification methods of TF-IDF, SVM, CNN and LSTM are mainly used for comparison experiments. The classification results are shown in Table 1. The main indicators for evaluating text classification are the following: Accuracy Rate, Loss Rate, and F1 score.

TABLE I. THE COMPARISON EXPERIMENTS RESULTS

Model	The Comparison experiments results		
	Accuracy	*Loss*	*F1*
TF-IDF	90.27	0.31	0.86
SVM	93.49	0.25	0.92
LSTM	94.26	0.21	0.94
CNN	95.61	0.15	0.96
Bi-LSTM-CNN	96.45	0.11	0.99

In order to avoid over-fitting in the experimental process, Cross-validation is performed with different test objects and training objects. The accuracy and loss rate of the loaded pre-training model during the training process are shown in Figure 2 and Figure 3. It is concluded from the figure that the accuracy of the model in this paper is gradually increasing, and the loss rate is steadily decreasing. When the training data reaches 1,800, the accuracy and loss rate tend to be stable.

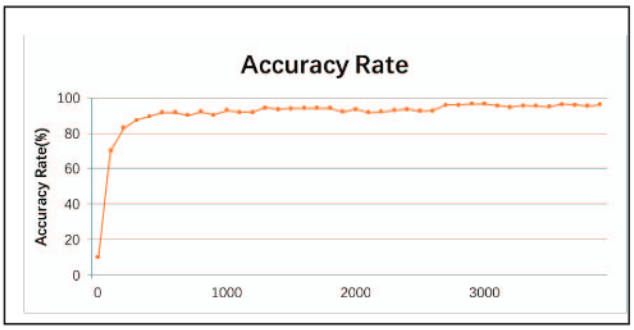

Figure 3. Model training accuracy rate line chart.

Figure 4. Model training loss rate line chart.

From the above data analysis show that the deep learning model outperform the traditional methods. The model of this paper is superior to the results of several other comparative experiments in terms of Accuracy Rate, Loss Rate and F1 score. By using this model which combining Bi-LSTM and CNN, and the efficiency of classification is improved by 0.84%. It can be explained that the Bi-LSTM-CNN-based model can capture features more accurately and introduces less noise. Moreover, this model can preserve more contextual information, and the text semantics can be expressed more accurately, so that the accuracy of news text classification can be improved, and generalization ability is better.

IV. CONCLUSION

In this paper, the Bi-LSTM-CNN model utilizes the loop structure to obtain the context information, and constructs the left and right contexts of each word through the Convolutional Neural Network (CNN) to construct the textual expression of the word, which is more accurately expressed the semantics of the text.

At the same time, there are still many guesses and areas for improvement in the model of this paper. First, for example, the literature [8] will increase the effect of the model after adding the attention mechanism (Attention mechanism). Second, there are many other words that interfere with the subject in the news corpus. The next step is to build a dictionary to filter these noise words.

REFERENCES

[1] *Xiaobo Jin, "A Survey On Text Categorization," AutomationPanorama,* vol. *23(S1), 2006,* pp.*24-29.*

[2] Xuefeng Xi and Guodong Zhou, "A Survey on Deep Learning for Natural Language Processing," ACTA AUTOMATICA SINICA, vol. 42(10), 2016, pp.1445-1465.

[3] Tang D, Qin B and Liu T, "Deep learning for sentiment analysis: successful approaches and future challenges," John Wiley & Sons Inc, 2015.

[4] Wenfei Lan, Wei Xu and Tao Wang, "Text Classification of Chinese News Based on Convolutional Neural Network," Journal of South-Central University for Nationalities(Nat. Sci. Edition), 2018(1), pp.138-143.

[5] Lei Huang and ChangShun Du, "Application of recurrent neural network in text classification," Journal of Beijing University of Chemical Technology(Natural Science), vol. 44(1), 2017, pp.98-104.

[6] Qianjian Gong, "A text classification based on the Recurrent Neural Networks," Huazhong University, 2016.

[7] Jianming Cui, Jianming Liu and Zhouyu Liao, "Research of Text Categorization Based on Support Vector Mechine," Computer Simulation, vol. 30(2), 2013, pp.299-302,

[8] Yongliang Wu, Shuliang Zhao and Changjing Li, "Text Classification Method Based on TF-IDF and Cosine Similarity," Journal of Chinese Information Processing, vol. 31(5), 2017, pp.138-145.

[9] Quanzhu Yao, Zhili Song and Cheng Peng, "Research on text categorization based on LDA," Computer Engineering and Applications, vol. 47(13), 2011, pp.150-153.

[10] Chongling Xia, Tao Qian and Donghong Ji, "Event Convolutional feature based news documents classification," Application Research of Computers, 2017(4), pp.991-994.

[11] Ming Tang, Lei Zhu and Xianchun Zou, "Document Vector Representation Based on Word2Vec," Computer Science, vol. 43(6), 2016, pp.214-217.

[12] Wu L, Hoi S C and Yu N, " Semantics-preserving bag-of-words models and applications," IEEE Transactions on Image Processing, vol. 19(7), 2010, pp.1908-1920.

[13] Wei Jiang and Zhong Jin, "Text Classification Based on Phrase Attention Mechanism," Journal of Chinese Information Processing, vol. 32(2), 2018.