International Core Journal of Engineering
Volume 7 Issue 8, 2021
ISSN: 2414-1895
DOI: 10.6919/ICJE.202108_7(8).0030

Research on Text Sentiment Analysis based on Attention Mechanism

Zexin Zhai

Key Laboratory of Electronic and Information Engineering (Southwest Minzu University), State Ethnic Affairs Commission, Southwest Minzu University, Chengdu, China.

Abstract

Text classification plays an important role in many natural language processing (NLP) applications, such as sentiment analysis, web search, spam filtering and information retrieval. In these applications, it is necessary to specify one or more predefined categories for a text sequence. For the problem of text classification, traditional classification methods need to design features and or mark parts of speech manually, which is very time-consuming and manual. In the neural network language model, it is difficult to learn the long-term dependence of gradient descent due to the problem of gradient disappearance. In order to overcome the limitations of existing technology, researchers began to increase the depth of network in recent years. However, increasing the depth of the network means increasing the number of network parameters, which makes the network computation expensive. In recent years, convolutional neural network (CNN) and recurrent neural network (RNN) have been applied to language modeling and achieved remarkable results, but they also have their own shortcomings. This paper proposes a model combining CNN, RNN and attention mechanism to overcome the problems of existing deep learning models. An unsupervised neural language model word2vec is used to train the initial word embedding, and then the deep learning network proposed in this paper is used for further training. The convolution layer is used to extract the local features of the text, and then the long-term relevance (global features) of the text learned by the Bi LSTM is input, Finally, through the attention layer, different features are assigned weights to obtain more important text information. Experimental results show that the proposed method is better than other classification methods.

Keywords

CNN; Bi LSTM; Attention Mechanism; Text Sentiment Analysis.

1. Introduction

Natural language processing (NLP) is an interdisciplinary field of computer science, artificial intelligence and linguistics. The goal of NLP is to enable computers to process or understand natural language in order to perform specific tasks (such as booking, purchasing items, answering questions). Text classification is an important part of natural language processing. It is a task to automatically classify a group of documents from a predefined set into multiple categories. It is an important task in many fields of natural language processing. Text classification has been applied in the fields of recommendation system [1] and spam filtering system [2]. Sentiment analysis is a branch of text classification, which is a research field of analyzing users' opinions, emotions, evaluations, attitudes and emotions about the entities and their attributes expressed in the text [3].

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0030

Text sentiment analysis, also known as opinion mining, refers to the analysis of subjective text with emotional color, mining the emotional tendency contained in it, and dividing the emotional attitude. As a research hotspot of natural language processing, text sentiment analysis has great significance in public opinion analysis, user portrait and recommendation system. The process of text sentiment analysis is shown in Figure 1., including the acquisition of original data, data preprocessing, feature extraction, classifier and the output of sentiment categories. The original data is generally obtained through web crawler, such as Sina Weibo content, twitter corpus, comments of major e-commerce websites, etc; Data preprocessing refers to data cleaning and noise removal. The common methods include removing invalid characters and data, unifying data categories (such as Chinese or simplified), using word segmentation tools for word segmentation, stopping word filtering, etc; According the different methods used, feature extraction has different implementation methods. When depending on different tools to obtain the numerical vector representation of text, the common methods are word frequency counting model n-gram and bag of words model TF-IDF, while the feature extraction of deep learning method is generally automatic; The final emotional polarity of the text is obtained from the output of the classifier. The common classifier methods are SVM and softmax.

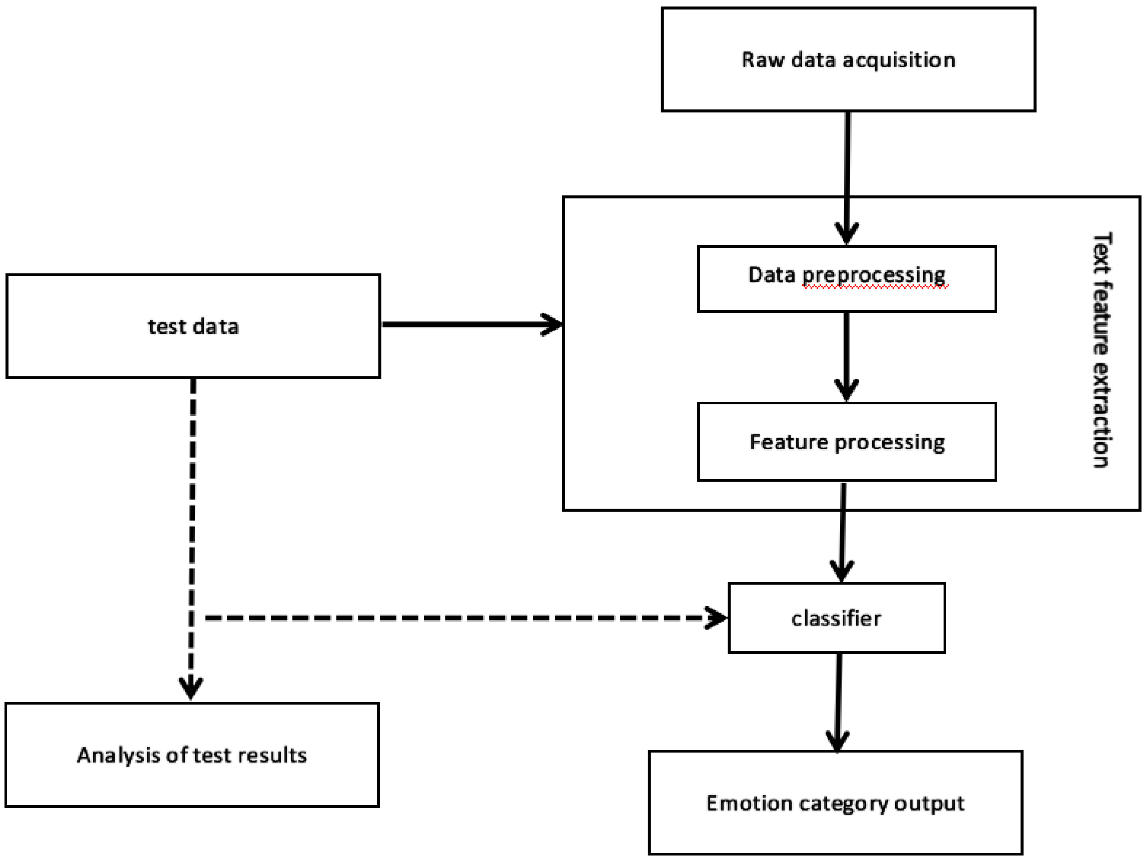

Figure 1. Text sentiment analysis process

2. Review of related technologies

Text classification is a classic topic in natural language processing. In natural language processing, it is necessary to assign predefined categories to free text documents. The traditional method of text classification has changed from manual feature design to machine learning. Almost all of these text classification techniques are based on words, for example, n-grams method has achieved good results. The goal of NLP is to use computer to process text, so as to analyze it, extract information, and finally complete different tasks, such as part of speech tagging, sentiment analysis and machine translation.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0030

2.1 Deep learning for natural language processing

A neural network consists of several neurons, and a single neuron consists of input, activation function and output. Let the input be an n-dimensional vector $x \ni R^n$. The output is calculated by the following functions:

$$a = f(w^T x + b) \tag{1}$$

Where f is the activation function. This function is also called nonlinear function. A common example is sigmoid function:

$$f(x) = sigmoid(x) = \frac{1}{1+e^{-x}} \tag{2}$$

Or hyperbolic tangent function:

$$f(x) = tanh(x) = \frac{e^{x-}e^{-x}}{e^{x}+e^{-x}} \tag{3}$$

The sigmoid activation function maps any real number to the interval [0,1]. In this interval, the activation function can be interpreted as the probability of the neural unit opening with the weight W and the offset parameter B. Tanh function and relu function are also commonly used activation functions. Deep learning has achieved remarkable results in computer vision [4] and speech recognition [5,6]. Deep neural network has become more and more popular in NLP application; A lot of work involves learning word vector representation through neural language model, and then combining the learned word vectors for classification. DNN based methods usually start from an input text and are represented as a sequence of words, in which each sequence is represented as a vector; Then, each word in the sequence is mapped into a continuous vector space. Dense vectors are generated by multiplying them with a weight matrix [7,8]. Then, the sequence is transmitted to DNN, and the DNN processes the sequence in multiple layers to generate classification. RNN can analyze text word by word and store the semantics of all previous texts in a fixed size hidden layer [9], but it also increases the time complexity. The ability to capture advanced, appropriate statistics is valuable for capturing the semantics of long text. However, in the RNN model, the nearest word is more meaningful than the previous word, and the key information of the text can appear anywhere in the document, not just at the end of the document. This can be inefficient when used to capture the semantics of the entire document. Therefore, in order to overcome the difficulty of RNN, long-term and short-term memory networks are introduced.

The standard RNN only predicts the past words of the text. This method is suitable for predicting the next word in the context. However, in the text sentiment analysis task, context information is equally important. We hope that we can use past and future words as part of speech markers at the same time. Bidirectional long-term and short-term memory network [10] is a further improvement of LSTM. Bi LSTM extracts features from the front and back directions to obtain more comprehensive and accurate semantic features. Therefore, Bi LSTM can better solve the task of sequence modeling than LSTM.

2.2 Attention mechanism

In the conventional encoder encoder model, researchers found that the context vector C learned by the model is too generalized, and the input of different output (YT) is the same context vector, which is different from people's intuitive understanding. In the case of translation task, for a word in the target language, we should focus on one or several words in the source language instead of the whole sentence. Therefore, people have proposed whether the machine can focus on one or some parts of the source input when doing these tasks. From the model, that is, for different outputs (YT), different "context vectors" need to be calculated so that the model can better learn the alignment between input and output, that is, attention mechanism. The calculation process of context vector is the calculation process of attention. On the surface, attention is just the recalculation of context vector. In essence, attention is a weighted sum of input. The combination of LSTM and attention mechanism can achieve good results, especially for sequence problems. Attention mechanism has been successfully applied to text classification, reading comprehension and other tasks.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202108_7(8).0030

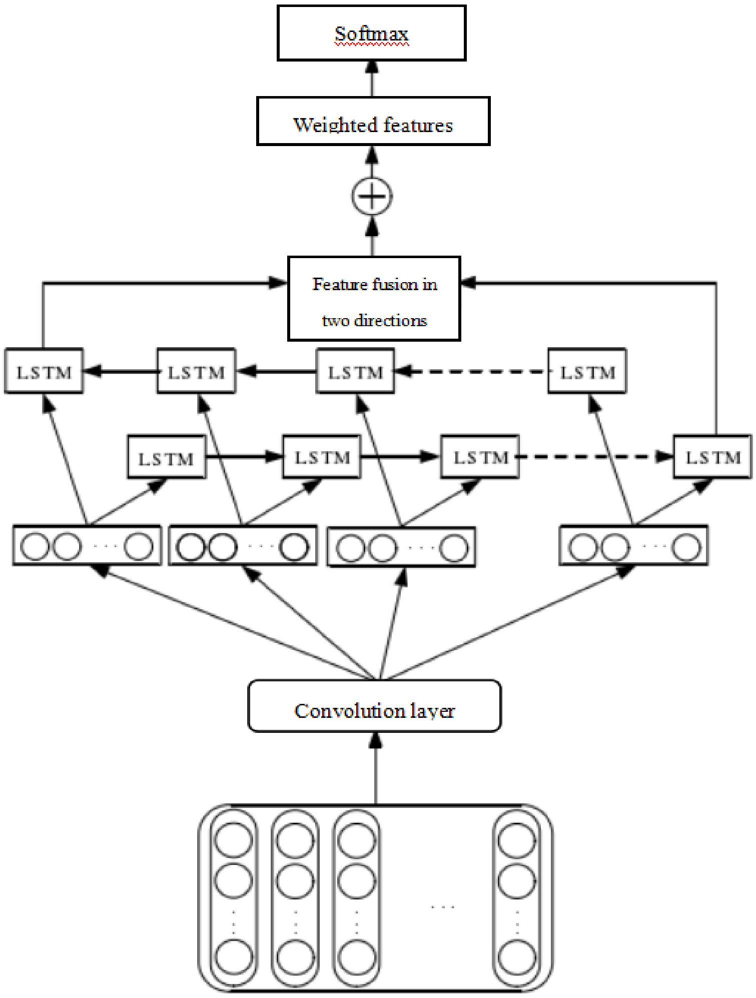

Figure 2. The proposed CNN-Bi LSTM-ATT architecture

3. Text sentiment analysis model

The model consists of convolutional neural network, recurrent neural network and attention layer. The model structure uses word2vec word embedding as input, and inputs them into convolutional neural network to learn and extract local features, and then inputs them into bidirectional LSTM layer. Bidirectional LSTM layer can capture long-term dependency from left and right directions to extract global features of text, and strengthen the weight of important information through attention layer. Finally, it inputs them into classification layer for classification. The model includes embedding layer, convolution layer, bilstm layer, attention layer and classification layer.

Figure 2. is the network model proposed in this paper, which combines the advantages of convolutional neural network and recurrent neural network, akes word embedding as the output of the model, learns the local features of the text through the convolution layer, and takes the learned features as the input of the bidirectional long-term and short-term memory layer to obtain the global features of the text, and then gives the weight value to the obtained features through the attention mechanism, The important information is highlighted to increase the accuracy of classification.

4. Experiment and result analysis

In order to test the effectiveness of the proposed model, we compare it with CNN, LSTM, Bi LSTM and cnn-lstm on the English IMDB movie review dataset and the Chinese Hotel Review dataset..

IMDB data set analysis:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0030

Table 1. Comparison of experimental results

Model	Accuracy/%	Precision/%	Recall/%	F1
CNN	83.44	84.61	84.10	0.8434
LSTM	83.66	84.22	84.34	0.8439
Bi LSTM	85.33	86.09	85.46	0.8577
CNN-Bi LSTM	86.87	86.96	85.93	0.8644
The model proposed in this paper	88.24	89.52	88.81	0.8920

Experimental results show that the accuracy of the proposed model is significantly improved, and compared with CNN Bi LSTM network, the accuracy is also improved by 2%. The model proposed in this paper combines the two kinds of networks, not only reduces the depth of the network, but also can learn the local and global features of the text at the same time. Then the attention mechanism strengthens the weight of important information, so as to improve the classification effect. The comparison of experimental results is shown in Table 1.

Hotel review data set:

Table 2. Comparison of experimental results

Model	Accuracy/%	Precision/%	Recall/%	F1
CNN	85.65	87.57	84.25	0.8588
LSTM	82.75	84.31	84.16	0.8423
Bi LSTM	86.51	88.89	85.75	0.8729
CNN-Bi LSTM	87.35	86.89	86.74	0.8698
The model proposed in this paper	89.23	89.74	88.36	0.8807

The experimental results show that adding bilstm layer and attention layer to the model can better learn the complex semantics in the text. Compared with CNN Bi LSTM model, the accuracy of the proposed model is also improved, which proves the effectiveness of attention mechanism.

5. Conclusion

In text sentiment analysis task, convolutional neural network learns text information by extracting high-level features of text. However, due to the locality of convolution layer, it needs many convolution layers to capture long-term dependencies. With the increase of the length of the input sequence, this situation becomes more serious. Finally, a network with many convolution layers is needed. In this paper, we propose a new framework to solve this problem. The goal is to capture text information and reduce the number of parameters in the architecture. In this framework, CNN and RNN are combined on unsupervised and pretrained word vectors, local features of text are learned through convolution layer, and Bi LSTM layer is used as the replacement of pooling layer to reduce the loss of details in local information and capture long-term dependency more effectively. Finally, attention mechanism is used to further improve the classification effect. The method presented in this paper has good performance on both datasets and is superior to other methods. The experimental results show that the same level of classification performance can be achieved by using smaller architecture.

References

[1] Watanabe A, Sasano R, Takamura H, et al. Generating Personalized Snippets for Web Page Recommender Systems [C]. 2014 IEEE/ WIC/ ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). ACM, 2014,218-225.

[2] Almeida T A, Silva T P, Santos I, et al. Text Normalization and Semantic Indexing to Enhance Instant Messaging and SMS Spam Filtering [J]. Knowledge-Based Systems, 2016, 108 :25-32.

[3] Liu, Bing. Sentiment Analysis (Mining Opinions, Sentiments, and Emotions) || Opinion Summarization and Search[J]. Cambridge University Press,2015:303-308.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0030

[4] Krizhevsky A, Sutskever I, HintonG E. Imagenet classification with deep convolutional neural networks [C]. Advances in neural information processing systems. 2012: 1097-1105.

[5] Graves A. Generating sequences with recurrent neural networks[J]. ar Xiv preprint ar Xiv:1308.0850, 2013.

[6] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks [C]. 2013 IEEE international conference on acoustics, speech and signal processing. IEEE, 2013: 6645-6649.

[7] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. ar Xiv preprint ar Xiv:1409.0473, 2014.

[8] Chorowski J K, Bahdanau D, Serdyuk D, et al. Attention-based models for speech recognition [C]. Advances in neural information processing systems. 2015: 577-585.

[9] Elman J L. Finding structure in time[J]. Cognitive science, 1990, 14(2): 179-211.

[10] Graves A, Jürgen Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Neural Networks, 2005, 18(5-6):602-610.