Sentiment analysis of Chinese text based on Elmo-RNN model

Min Yang1,a, Junwu Xu2*, Kai Luo3, Yiping Zhang4
College of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China
a21807010024@stu.wit.edu.cn, *4300513@wit.edu.cn

Abstract. Traditional word embedding technologies such as Word2Vec and GloVe can only produce a single semantic vector, and cannot get the specific meaning of polysemous words in the text combined with context. To solve this problem, a sentiment analysis model of Chinese text based on Elmo and recurrent neural network is proposed. The model uses Elmo model to learn the pre training corpus, and the BiLSTM network structure in Elmo model makes the word vector generated by Elmo model have context sensitive characteristics, which is more accurate in the expression of polysemous words, and this is not achieved by the traditional word embedding technology. Then, the model uses the recurrent neural network to extract the deep-seated features of the word vector and fuse the features. Finally, the softmax function is used to realize the sentiment classification of text. Experimental results show that Elmo-RNN can effectively improve the accuracy of text sentiment analysis.

1. Introduction

With the development of the Internet and the popularity of mobile communication devices, more and more people are used to expressing their opinions on the social networking platform such as Weibo. Over time, these words with personal emotion color form the text big data on the Internet today. It is very important to analyze the emotional preferences of these text data and understand what people think. With these data, businesses can better understand the needs of customers; the government can better grasp the trend of social public opinion and conduct appropriate guidance. Therefore, text emotion analysis[1] has become an indispensable direction in the field of Natural Language Processing (NLP).

Traditional sentiment analysis methods can be roughly divided into three categories: methods based on emotion dictionary; methods based on machine learning; methods based on deep learning model. The method based on emotion dictionary is limited by the quality and coverage of emotion dictionary [2]. The most frequently used methods are supervised learning and unsupervised learning. Supervised methods mainly use machine learning technology, such as Support Vector Machine (SVM), maximum entropy method and naive Bayes method[3] to learn text, and then to classify emotion. The unsupervised method mainly analyzes the emotional words, grammar and semantics in the text, and finally realize emotion classification by extracting statistical features of text.

In the process of natural language processing, a very important thing is how to accurately convert words into languages that can be recognized by computers. Traditional word embedding technologies, such as Word2Vec, GloVe, usually store multiple semantics of a word in the same word embedding vector when processing polysemous words, which makes it impossible to accurately obtain the information expressed by the word when processing text tasks. For example, apple can refer to both Apple phone and apple in fruit. If the word is vectorized by traditional embedding technology, the result is a vector containing two semantics.
2. Related work
As a popular direction of artificial intelligence, deep learning is also widely used in the field of natural language processing, including emotional analysis. After Hinton [4] proposed deep learning in 2006, deep learning method was introduced into the field of natural language processing to solve the problem that traditional machine learning methods can't mine deep semantics. Deep learning method is the mainstream method in the field of emotional analysis.

In 2016, Bahdana [5] first applied attention mechanism to natural language processing tasks, and constructed the machine translation model with the best performance at that time. Standard machine learning methods are used to solve emotional analysis problems. Feature Engineering is the core of this kind of method. Liang Jun [6] proposed Long Short-term Memory (LSTM) networks for sentiment analysis of Chinese microblog texts.

The main purpose of text sentiment analysis is to mine the subjective emotion contained in the original text [7]. Text is a kind of unstructured data. Before sentiment analysis, it is necessary to transform it into structured data that can be recognized by computer, that is, word embedding vector. However, the traditional word embedding method can not express the exact meaning of polysemous words, and the effect is not good, which affects the follow-up experiments. In order to solve this difficult problem, this paper uses Elmo (Embedding from Language Model) model to generate word vector. The advantage of Elmo model is that the word vector generated by Elmo model can accurately express different semantics of polysemous words by combining context information, which is suitable for different scenes of natural language processing.

3. Elmo RNN Emotion Analysis Model
This paper proposes an emotive analysis model based on Elmo-RNN to study the emotional tendency of Chinese texts. This model combines Elmo model and Recurrent Neural Network (RNN). Firstly, context related word vectors are generated by pre training language model based on large-scale corpus. Then, deep feature extraction of word vectors is carried out by RNN network. Finally, the experimental results are obtained by using Softmax activation function.

3.1. Elmo model
Elmo model was proposed by Peters [8] in 2018. Unlike most widely used word embeddings, Elmo word representations are functions of the entire input sentence, as described in this section. It can learn the complex features of lexical usage, such as syntactic and semantic features, as well as the changes of these usages in different language environments, that is, modeling polysemy with context [9]. The model is pre-trained on a large text corpus, which can significantly improve the research status in a series of challenging NLP problems, including question answering, text implication and sentiment analysis. The model can be divided into two parts: character level coding layer and BiLSTM (Bi-directional Long Short-term Memory) network layer as shown in Figure 1.

![Figure 1 Illustration of Elmo model](image_url)
3.2. **BiLSTM model**

BiLSTM network is the basis of Elmo model[10], specifically, given a sequence of N tokens, \(t_1, t_2, ..., t_N\). A forward language model, such as multi-layer LSTM, is used to calculate the probability of the current token given the previous tokens, as shown in formula (1):

\[
p(t_1, t_2, ..., t_N) = \prod_{k=1}^{N} p(t_k|t_1, t_2, ..., t_{k-1})
\]  

(1)

At each position \(k\), each LSTM layer outputs a context-dependent representation \(h_{LM}^{k,j}\) where \(j = 1, ..., L\). The top layer LSTM output, \(h_{LM}^{L,j}\), is used to predict the next token \(t_{k+1}\) with a Softmax layer.

A backward LM is similar to a forward LM, except it runs over the sequence in reverse, predicting the previous token given the future context as shown in formula (2):

\[
p(t_1, t_2, ..., t_N) = \prod_{k=1}^{N} p(t_k|t_{k+1}, t_{k+2}, ..., t_N)
\]  

(2)

At each location \(K\), the reverse LSTM also generates a context sensitive representation \(h_{LM}^{k,M}\) at each level. Bilstm used by Elmo is a language model that combines both forward and reverse directions. The loss function is to maximize the following likelihood values, as shown in formula (3):

\[
\sum_{k=1}^{N} \left( \log p(t_k|t_1, ..., t_{k-1}; \Theta_x, \Theta_{LSTM}, \Theta_s) + \log p(t_k|t_{k+1}, ..., t_N; \Theta_x, \Theta_{LSTM}, \Theta_s) \right)
\]

(3)

Among them, \(\Theta_x, \Theta_s, \Theta_{LSTM}, \Theta_{LSTM}\) are the parameters of word embedding, output layer (before Softmax) and forward and backward LSTM.

From this, we can see that the bilistm is actually equivalent to two LSTM which have been trained in the forward and reverse directions respectively. The network structure is shown in Figure 2.

![Figure 2 BiLSTM network structure](image)

3.3. **RNN (Recurrent Neural Network) structure**

The input and output of traditional neural networks, such as CNN, are independent of each other, while RNN introduces the concept of "memory", that is, the output needs to rely on the previous input sequence and remember the key input. It does not memorize all fixed length sequences rigidly, but stores the information of previous time steps by hiding state, which makes RNN have certain advantages in processing nonlinear features.
4. Experiment

4.1. Experiment data set
The data set used in the experiment is to crawl the comments of users' shampoo on JingDong, and make a two classification according to the emotional color of the comments. The positive emotion is marked as 1 and the negative emotion is 0. There are 10000 data, including 5000 positive comments and 5000 negative comments.

Chinese text needs to be segmented in the process of processing. Here we use the general word segmentation tool pkuseg, and use the space as the separator to separate the sentence from the word, as shown in Table 1.

| Table.1 Model comparison results |
|----------------------------------|
| Word segmentation outputs       |
| Positive comments               |
| I received the goods. The logistics was fast. I didn't have to say anything about it. When I bought the goods, I still worried that the logistics was busy. I was afraid it would not be useful. I opened the things. I like them very much. I sent five travel clothes. Thank you Jingdong. Will continue to pay attention. Look at the pictures. Isn't it great. | 1 |
| Negative comments               |
| The packing is very bad. The tape is stuck askew. The tallyman is very perfunctory. My order clearly says to ask for the invoice and the shopping receipt. As a result, after opening the package, there is nothing. This time, it's OK. Next time, I will return the goods directly. | 0 |

4.2. Experimental environment configuration
The experimental environment configuration is shown in Table 2:

| Table.2 Environment configuration |
|-----------------------------------|
| Experimental environment          |
| Language                          | Python3.6 |
| Tool                              | Pycharm   |
| Frame                             | Pytorch0.4|
| Operating system                  | Ubuntu16.04|

4.3. Valuation index
In this paper, the recognition results are divided into four categories: TP stands for positive, TN for negative, FP for false positive, FN for false negative. The experimental results are judged by accuracy. The higher the accuracy is, the better the classification ability of the model is. The calculation formula is shown in formula (4):

\[
Accuracy = \frac{TP + FP}{TP + FP + TN + FN} \tag{4}
\]

4.4. Contrast experiment setup
In order to verify the effectiveness of the model, three groups of comparative experiments are set up to compare the model with other emotion classification methods.

1. CNN: Single CNN, network input is word2vec training word vector.
2. BiLSTM: Word2vec is used to train the word vector, and the Bi-LSTM model is used to learn the emotional features of the text.
3. Baseline: This is the baseline model proposed by nlpcc2018 multi language emotion analysis and evaluation task. The model uses unigram features and SVM classifier.
4. CNN+BiGRU: Network input is word vector trained by Word2Vec. First, a CNN network is used, and then a BiGRU is added to it.
4.5. Analysis of experimental results
In this paper, a pre-trained Elmo model is used to obtain context sensitive word vectors. The pre-training corpus includes the news captured from Wikipedia and WMT 2008-2012, and the weights and options data of pre-training are obtained. The input of the model can be character or vector. In this paper, Chinese characters are used as the input of the model. The dimension of character vector in Elmo is set to 128 dimensions, the context window is associated with 20 words, and the high-speed network is set to 2048 dimensions. The experimental results are shown in Table 3:

| Experiment result | Accuracy |
|-------------------|----------|
| CNN               | 0.8285   |
| BiLSTM            | 0.8614   |
| Baseline          | 0.8375   |
| CNN+BiGRU         | 0.8756   |
| Elmo+RNN          | 0.8891   |

It can be seen from the table that the accuracy rate of using a single CNN network is only 0.8285, because the network structure of a single CNN network is relatively simple, with the advantages of fast speed and the shortcoming is relatively low accuracy. BiLSTM is Bi-directional Long Short-term memory neural network, which has complex structure, high accuracy and takes longer than CNN. The accuracy of baseline model is greatly improved compared with CNN, but the effect is not as good as BiLSTM. CNN + BiGRU is a combination of CNN network and BiGRU network, which is better than the single network structure. The accuracy rate is 0.8756. Compared with the three models, the model in this paper has the best effect and the correct rate reaches 0.8891.

5. Conclusion
Different from the traditional sentiment analysis methods, this paper proposes a method of using Elmo model to obtain the word vector as the input of the network, which can effectively improve the accuracy of emotion classification. Elmo output word vector combines its own and context features, and has context relevance. The classifier uses RNN network, which can well fuse phrase features, which is conducive to subsequent classification. Experimental results show that the accuracy of the proposed method is better than the existing methods. However, due to the complexity of the model, the time cost of the model is increased to a certain extent. In the future, it is the goal of the next step to study how to improve the classification accuracy and reduce the time cost of the model.

Acknowledgments
This work was supported by the Wuhan Institute of Technology Software engineering laboratory and the eleventh Graduate Innovation Fund of Wuhan Institute of Technology (CX2019235, CX2019236).

References
[1] Wei W, Xiang Y, Survey on Chinese text sentiment analysis[J]. Journal of Computer Applications, 2011, 31(12):3321-3323.
[2] Nasukawa T, Yi J. Sentiment analysis: capturing favorability using natural language processing [C] / Proc of Int Conf on Knowledge Capture, 2003:70-77.
[3] PANG B, LEE L. Opinion mining and sentiment analysis[J]. Foundations and Trends in Information Retrieval, 2008, 2(1-2): 1-135.
[4] Geoffrey E Hinton, Simon Osindero, Yee-Whye Teh. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554.
[5] BAHDDANA U.D, CHO K, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. arXiv preprint arXiv: 1409.0473, 2016: 1-15.
[6] Liang J, Cai Y.M, Yuan H.B, et al. Polarity Shifting and LSTM Based Recursive Networks for
Sentiment Analysis [J]. Journal of Chinese Information Processing, 2015, 29 (5): 152-159.

[7] Wang L.Y, Liu C.H, et al. Text Sentiment Analysis Based on CNN-BiLSTM Network and Attention Mode[J]. Journal of Wuhan Institute of Technology, 2019, 41(04):386-391.

[8] Peters M, Neumann M, Iyyer M, et al. Deep Contextualized Word Representations[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). Stroudsburg: Association for Computational Linguistics, 2018: 2227-2237.

[9] Liu Y.J, Che W.X, et al. Deep Contextualized Word Embeddings for Universal Dependency Parsing[J]. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 2019, 19(1).

[10] Zhao Y.O, Zhang J.Z, et al. Sentiment analysis using embedding from language model and multi-scale convolutional neural network[J]. Journal of Computer Applications, 2020, 40(03): 651-657.