Sentiment analysis of film reviews based on CNN-BLSTM-Attention

Lu Shang, Liying Sui, Shusong Wang, Dexue Zhang*

College of Electronic Information Engineering, Shandong University of Science and Technology, Qingdao, 266590, China

gxc1228@yeah.net

Abstract. In order to accurately analyse the emotional tendency of film reviews, help investors make decisions and improve the quality of works, an optimized CNN-BLSTM-Attention sentiment analysis model was designed. The CNN model has a strong ability to capture the local correlation of spatial or temporal structures. The RNN model can either process sequences of any length or capture long-range dependencies, but it is easy to cause the problem of gradient disappearance. The CNN-BLSTM-Attention sentiment analysis model designed in this paper, which combines the advantages of CNN and RNN, is more accurate when being used to analyze the sentiment characteristics of texts. The experimental results show that the accuracy of the CNN-BLSTM-Attention after optimization model is better than that of CNN and RNN models in the experiment, which proves the effectiveness of the analysis method in this paper and can provide some significance for the optimization of related sentiment analysis models.

1. Introduction
With the rapid development of Internet technology and the rise of social networks, more and more people choose to express their opinions on film and television works through the Internet, which makes it easier for filmmakers to understand the public’s opinions and evaluations of movies. Under this kind of network environment, a large number of movie reviews with personal emotions are generated. Analyzing these texts with personal emotions is a beneficial work for the film industry and consumers.

The main task of sentiment analysis is to complete the classification of sentimental text. The text sentiment analysis mainly includes text classification, information extraction and text generation [1]. For sentiment analysis, a large number of related sentiment classification methods have emerged. For example, Kim proposed the CNN (Convolutional Neural Network) model in 2014 [2], which used word vectors to classify text and achieved impressive results. Mikolov [3] and others proposed RNN (Recurrent Neural Network) model in 2010. Later on the basis of RNN, some scholars proposed the LSTM model, which is a variant of RNN. Convolutional neural network (CNN) and recurrent neural network (RNN) are network models that are currently used in sentiment analysis.

In this paper, a new model of sentiment analysis, CNN-BLSTM-Attention, is proposed through the investigation and research of various deep learning models, which is superior to CNN and RNN models.

2. Sentiment analysis model

2.1 CNN model
CNN is a multi-layer neural network. Its basic structure includes an input layer, a convolution layer, a pooling layer, a fully connected layer, and an output layer. In this paper, CNN is first used to train the data to obtain a training model. The training process is shown in Figure 1.

![Training process diagram](image)

**Figure 1 CNN model training sample process**

To build a CNN model, multiple one-dimensional convolution kernels need to be defined firstly, then these convolution kernels should be used to perform convolution calculations on the inputs. In this process, convolution kernels with different widths may capture the correlation of different numbers of neighboring words. Secondly, timing maximum pooling should be performed to all output channels, and then the pooled output values of these channels would be linked into a vector. Finally, the connected vectors would be transformed into the output of each class by the fully connected layer. This step generally uses discarded layers to deal with overfitting.

### 2.2 LSTM model

Recurrent neural network (RNN) has a wide range of applications in processing serialized inputs. It is a typical model in deep learning frameworks. Long short-term memory (LSTM) network [4] is a variant of recurrent neural network. LSTM introduces 3 gates, namely input gate, forget gate and output gate, and memory cells with the same shape as the hidden state of recurrent neural network. LSTM introduces 3 gates, namely input gate, forget gate and output gate, and memory cells with the same shape as the hidden state.

The input gate $I_t \in R^{n \times h}$, forget gate $F_t \in R^{n \times h}$, and output gate $O_t \in R^{n \times h}$ of the time step are calculated as follows:

$$I_t = \sigma(X_tW_{ix} + H_{t-1}W_{hi} + b_i)$$  \hspace{1cm} (1)

$$F_t = \sigma(X_tW_{xf} + H_{t-1}W_{hf} + b_f)$$  \hspace{1cm} (2)

$$O_t = \sigma(X_tW_{io} + H_{t-1}W_{ho} + b_o)$$  \hspace{1cm} (3)

Where $W_{xi}, W_{xf}, W_{xo} \in R^{d \times h}$ and $W_{hi}, W_{hf}, W_{ho} \in R^{h \times h}$ are weight parameters and $b_i, b_f, b_o \in R^{1 \times h}$ are bias parameters.

The characteristics of ordinary LSTM model is that the current time step is determined by the earlier sequence, and the information is passed from front to back through the hidden state. However, the current time step sometimes may also be determined by subsequent time steps. BLSTM handles this kind of information more conveniently by adding hidden layers passed from back to front. The BLSTM model is shown in Figure 5.

### 2.3 CNN-BLSTM-Attention model

Aiming at resolving the weakness of the CNN model which is sensitive to local features, and the RNN model which is prone to generate the problem of gradient disappearance, we fused and transformed the CNN and RNN model, and proposed a new model combining the advantages of the two models. The workflow of the model contains the following process. Firstly, we should convert the texts into the corresponding vocabularies. Secondly, we put these vocabularies through the CNN's convolution layer, pooling layer, and Dropout operation, and then through the BLSTM layer and Attention layer. Finally, after full connection layer and Softmax layer, we will get the final output. This model not only solves the weakness that CNN model is sensitive to local features but also uses a chain neural network to store and propagate information by accessing the BLSTM network. Figure 2 is a schematic diagram of the design model of this paper.
Among them, Dropout specifically refers to the inverted dropout method, and its calculation expression is
\[ h_i = \phi(x_1w_{1i} + x_2w_{2i} + \cdots + x_nw_{ni} + b_i) \] (4)

Here \( \phi \) is the activation function, \( x_1, ..., x_n \) are inputs, the weight parameter of the hidden unit \( w \) is \( w_{1i}, ..., w_{ni} \), and the deviation parameter is \( b_i \). Let the probability of the random variable \( \xi_i \) equals 0 and 1 is \( p \) and \( 1-p \), respectively. Calculate the new hidden unit \( h'_i \) when using the discard method
\[ h'_i = \frac{\xi_i}{1-p} h_i \] (5)

Since \( E(\xi_i) = 1-p \), therefore
\[ E(h'_i) = \frac{E(\xi_i)}{1-p} h_i = h_i \] (6)

That is to say, the discard method does not change the expected value of its input. Since the discarding of hidden layer neurons is random during training, it plays a role of regularization when training the model, and can be used to deal with overfitting.

In addition, this paper also adds attention mechanism on the basis of CNN and RNN. The key point of using attention mechanism is to calculate the background variable. The reset gate, update gate, and candidate hidden states are
\[ r_{t'} = \sigma(W_{yr}y_{t'-1} + W_{sr}s_{t'-1} + W_{cr}c_{t'} + b_r) \] (7)
\[ z_{t'} = \sigma(W_{yz}y_{t'-1} + W_{sz}s_{t'-1} + W_{cz}c_{t'} + b_z) \] (8)
\[ \tilde{s}_{t'} = \text{tanh}(W_{ys}y_{t'-1} + W_{ss}(s_{t'-1} \odot r_{t'}) + W_{cs}c_{t'} + b_s) \] (9)

Among them, \( W \) and \( b \) with subscripts are the weight parameter and deviation parameter of the gating cycle unit, respectively.

3. Experiment and analysis

3.1 Data set
This paper uses Stanford's IMDb dataset (Stanford's Large Movie Review Dataset) as the dataset for text sentiment classification [5]. This data set is divided into two data sets for training and testing, each containing 25,000 reviews of movies downloaded from IMDb. In each dataset, the number of reviews labeled "positive" and "negative" is equal.

3.3 GloVe model
The GloVe model is a new word embedding model proposed after word2vec. The GloVe model uses a squared loss and changes the hop model based on the squared loss [6]. Let the probabilistic distribution
variables \( p'_{ij} \) equals \( x_{ij} (p'_{ij} = x_{ij}) \) and \( q'_{ij} \) equals \( \exp(u^T_j v_i) (q'_{ij} = \exp (u^T_j v_i)) \) and calculate the logarithm. Therefore, the squared loss term is

\[
(loxp'_{ij} - logq'_{ij})^2 = (u^T_j v_i - logx_{ij})^2
\]

The target of the GloVe model is to minimize the loss function, which is

\[
\sum_{i \in V} \sum_{j \in V} h(x_{ij})(u^T_j v_i + b_i + c_j - logx_{ij})^2
\]

The weight function \( h(x) = (x/c)^\alpha \) when \( x < c \), otherwise, \( h(x) = 1 \). Because \( h(0) = 0 \), the flat loss term for \( x_{ij} = 0 \) can be ignored directly.

### 3.3 Parameter setting

The experimental environment of this article is python 3.6.9, using Mxnet deep learning open source framework. In the experiments in this paper, the hyperparameter settings are shown in Table 1:

| Experimental parameters          | Value  |
|----------------------------------|--------|
| Word vector dimension            | 500    |
| Convolution kernel size          | 3, 4, 5|
| Convolution kernels              | 100, 200, 200 |
| Number of LSTM units             | 100    |
| Dropout rate                     | 0.5    |
| Batch size                       | 64     |
| Number of iterations             | 20     |

### 3.4 Comparative test

In this paper, a total of three sets of comparative experiments are designed to compare the impact of using hybrid neural networks and attention models on classification results.

- Model 1: CNN, a text sentiment analysis model based on convolutional neural network.
- Model 2: BLSTM, a text sentiment analysis model based on BLSTM.
- Model 3: CNN-BLSTM-Attention, a network model combining a convolutional neural network and a BLSTM unit.

### 3.5 Experimental results and analysis

The experimental results are shown in Table 2. The accuracy rate of the CNN model is the lowest. The BLSTM model is one percentage point higher than the CNN model. The CNN-BLSTM-Attention optimized by this paper has a significantly higher accuracy rate, reaching 0.99. It shows that our optimization model has better accuracy than the original CNN and RNN model.

| Algorithm               | Accuracy |
|-------------------------|----------|
| CNN                     | 0.95     |
| BLSTM                   | 0.96     |
| CNN-BLSTM-Attention     | 0.99     |

Figures 3, 4 and 5 are the CNN, BLSTM, optimized CNN-BLSTM-Attention model accuracy, and Loss change diagrams, respectively.
Figure 3 CNN model accuracy and Loss change

Figure 4 BLSTM model accuracy and Loss change

Figure 5 CNN-BLSTM- Attention model accuracy and Loss change
4. Conclusions and prospects
The sentiment analysis of movie reviews has strong practical significance. This paper makes full use of the advantages of CNN and LSTM models, and proposes a new sentiment analysis model based on CNN and BLSTM models which adds attention mechanism. Experimental results also show that the optimization algorithm in this paper is superior to CNN and BLSTM models. However, due to the increased complexity, the training time of the model is greatly increased, and there is still room for improvement.

Acknowledgments
Special thanks to the careful teaching of my teachers and the help of classmates.

References
[1] Liu, B. . (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies* 5.1 (2012): 1-167.. Morgan & Claypool.
[2] Kim, Y. . (2014). Convolutional neural networks for sentence classification. *Eprint Arxiv*.
[3] Mikolov, T. , Sutskever, I. , Chen, K. , Corrado, G. , & Dean, J. . (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26, 3111-3119.
[4] Li, H. , & Yu, Q. . (2017). The wire beltline diameter ACO-KF-PID contro research. *Prognostics & System Health Management Conference*. IEEE.
[5] Maas, Andrew & Daly, Raymond & Pham, Peter & Huang, Dan & Ng, Andrew & Potts, Christopher. (2011). Learning Word Vectors for Sentiment Analysis. 142-150.
[6] Pennington, Jeffrey & Socher, Richard & Manning, Christopher. (2014). Glove: Global Vectors for Word Representation. *EMNLP*. 14. 1532-1543. 10.3115/v1/D14-1162.