Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 10, No. 1, 2023

Chinese review sentiment analysis based on deep learning and attention mechanism

Wenjie Hu

Anhui University of Finance and Economics Bengbu, China

Abstract: With the development of the Internet and the continuous growth of the number of netizens, more and more people publish their attitudes and emotions on products, services, etc. on the Internet, accumulating a large number of comments containing personal opinions. The sentiment analysis of text reviews is of great significance to explore the emotional tendencies of users and help merchants adjust product positioning and marketing strategies in a timely manner. The deep learning framework adopts CNN and LSTM models respectively to introduce the attention mechanism for sentiment classification. The effects of CNN model and LSTM model on sentiment classification in Weibo comments were analyzed. From the experimental data, after the same number of iterations, the accuracy of the CNN model reaches 97.94%, and the accuracy of the LSTM model reaches 98.18%. The experimental results show that LSTM is better than CNN in the sentiment classification of Weibo comments.

Keywords: Sentiment classification; Convolutional neural networks; Long short-term memory networks; Attention mechanism.

1. Introduction

Since online shopping, takeaway, ticketing, online education and other websites have user rating functions, consumers can give timely evaluations of purchased products, and merchants can use these feedback data to judge consumers' approval of goods and adjust product positioning and marketing strategies.

The growth in the number of Internet users has led to an explosion of attitudes and perceptions of online reviews, products and online user activities on social media platforms. These review texts contain a wealth of valuable information, have high search value and commercial value, and have applications in almost all areas of business and society. Therefore, extracting emotional information from online feedback data is very important to determine the purpose and emotional tendency of user reviews and to judge the judgment attitude of products, services, etc.

2. Related work

The current research on user online comment sentiment analysis is mainly divided into three categories, using emotional dictionaries, based on traditional machine learning algorithms and based on deep learning algorithms. Yang Shuxin et al. proposed a sentiment analysis model SLP-ELMO that integrates emotional dictionary and contextual language model, which can dynamically adjust the word filling of words according to context, and solve the problem of single semantic representation of word embedding. Xu Minlin proposes a text sentiment analysis model combining sentiment dictionary and parallel neural network. Wang Zhiying completed the construction of a machine learning classification model on the Spark distributed computing platform, and Yan Junchao et al. trained samples using three different algorithms: Naive Bayes, support vector machine (SVM) and recurrent neural network (RNN), and the results showed that Naive Bayes and SVM are good, and SVM is better in comparison. Xu Kangting proposed a sentiment analysis method CLKDL combining language knowledge and deep learning. Xie Caiyun et al. studied CNN-based text sentiment analysis and SVM-based sentiment analysis, and adjusted the parameters of the two models. Experimental results show that the model based on SVM has a good classification effect. Wang Hanqi proposed an ALBERT-LSTM model for user comments, and then combined with the neural network model LSTM to construct the ALBERT-LSTM model. Yinggang Wu constructed the LSTM model of ordered neurons and the LSTM self-attention model of ordered neurons to handle text analysis tasks. Zhang Tingting constructs a Chinese comment text sentiment classification model based on bidirectional time convolutional neural network combined with attention mechanism. Yan Weichen [10] proposed a text sentiment analysis method based on multi-head self-attention mechanism and bidirectional LSTM. This method makes up for the problem of insufficient ability to extract emotional words.

3. Sentiment analysis model

3.1. CNN model

With the deepening of in-depth learning research, convolutional neural networks have also been applied to natural language processing by researchers, and the biggest advantage of CNN is its easy-to-understand neural network structure and very flexible processing of high-dimensional data information. Compared to other time series models, convolutional neural networks do not depend on the output of the last layer of the model, so the training time of the model is very short. The CNN model mainly includes input layer, convolutional layer, pooling layer, fully connected layer and softmax output layer. The role of the convolutional layer is to better represent the characteristic information of the input text. The main function of the pooling layer is to filter the properties obtained by filtering the volume layer according to a certain standard. The whole connection layer is equivalent to the hidden layer of the traditional neural network, which plays the role of fixed dimension, and finally connects the softmax output layer to obtain the final classification output. The CNN structure is shown in Figure 3-1

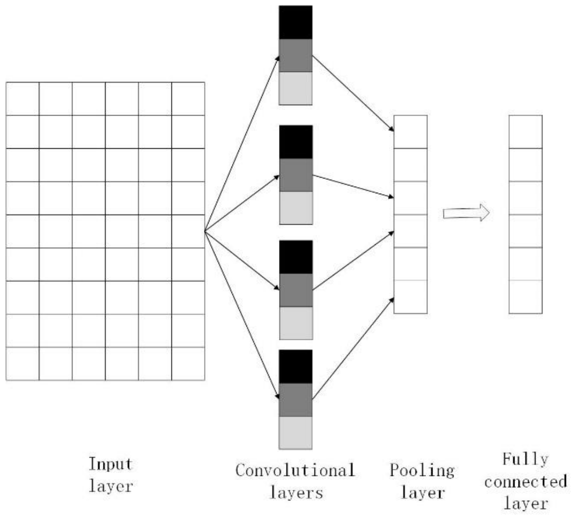

Figure 1. CNN structure diagram

(1) Input layer: The dimension of the input vector is usually two-dimensional.

(2) Convolutional layer: According to the convolution operation, the higher hierarchy of data feature information is extracted, and the formula is shown in (1).

$$c_i = f(w^T H_{i:i+h-1} + b) \quad (1)$$

$$c = [c_1, c_1, \cdots, c_{1n-h+1}] \epsilon R^{n-h+1} \quad (2)$$

(3) Pooling layer: not only reduces the amount of data processing, but also retains the effective feature information of text data to the greatest extent, the specific formula is shown in (3).

$$\hat{c} = max\{c\} \quad (3)$$

$$z = [\hat{c}_1, \cdots, \hat{c}_m] \quad (4)$$

(4) Fully connected layer: Connects the high-level text information features extracted by the previous layer.

(5) Output layer: Predict feature information through activation function.

3.2. Long short-term memory networks

Recurrent Neural Network (RNN) can process sequence information well, so RNN can be selected to process comment data with sequence characteristics. RNNs solve some of the limitations of feedforward neural networks in synchronization problems, but they also bring self-imposed problems, such as gradient vanishing and explosion during RNN network training etc. In RNN, as a node moves backwards, the perception of the previous node gradually decreases, causing the gradient to disappear. Long Short-Term Memory networks (LSTMs) are special RNN networks designed to solve long-dependency problems. LSTM uses the gate mechanism to store information from previous moments, which can solve the gradient vanishing problem in RNNs well. LSTMs generally treat sentences as sequences in text sentiment classification. Within these words, each word can be thought of as a node. Each word is arranged in order by the LSTM unit. In this way, each word can learn something from the previous words. Finally, the context information of the entire sentence is output as a word vector. Therefore, the sentiment of the text can be classified.

The internal cell structure of the LSTM is shown in Figure 3-2:

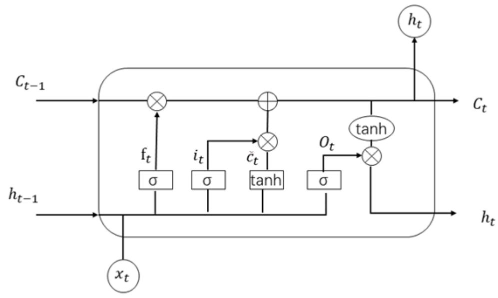

Figure 2. LSTM block diagram

In an LSTM model, include an input gate, i_ta forget gate, $f_t$an output gate, $O_t$and a cell state$\tilde{C}_t$.

At the heart of the LSTM is the cellular state, represented by horizontal lines running through the cell. Like a conveyor belt, the state of the cell has a very small number of branches throughout the cell. LSTM networks can remove or add cell state information through structures called gates. The forget gate $f_t$takes the output of the previous moment $h_{t-1}$and the word vector at that moment x_t as input, and the LSTM network is calculating the value of the Sigmoid unit Then decide which information to forget and which to keep. After passing the Sigmoid cell, the value is reflected to the [0,1] interval, where the information is retained when the value is equal to 1. When it is equal to 0, the information is forgotten. The input gate $i_t$determines the value of the conveyor belt to update. The output gate $O_t$determines which values flow out.

3.3. Attention mechanisms

The common thing in the attention mechanism is the self-attention mechanism, and three different vectors are required for the input of the self-attention mechanism, namely: Query(Q), Key(K), Value (). V）。 Then the result of attention is obtained by calculating the formula, which is shown in (5).

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

Among them, it d_k represents the dimension, and the reason why it is divided by the $\sqrt{d_k}$penalty factor is to avoid QK^T excessive inner product. The Softmax $\frac{QK^T}{\sqrt{d_k}}$ function normalizes the operation and then obtains the attention coefficient of each word for other words.

The attention mechanism is to weight the output value of the hidden layer, and the output value of the hidden layer at different moments is assigned different weights by the attention mechanism, obtaining information that is more important to the result, and ignoring secondary information, and at the same time because the encoding-decoding model works well. Therefore, when establishing the encoding-decoding structure model, the attention mechanism can be combined with the model to improve the effect of the model.

3.4. Model Framework

The CNN model has five layers, input layer, CNN layer, attention layer, full connection and classification layer, and output layer. The model structure diagram is shown in the following figure:

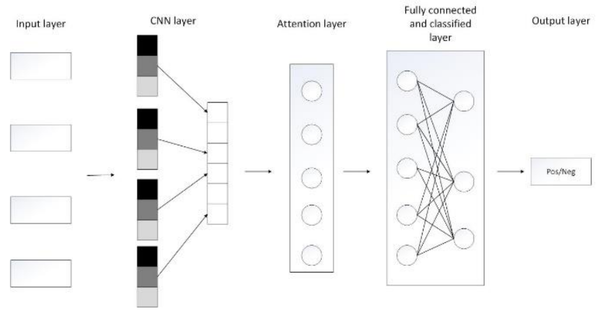

Figure 3. CNN model framework structure diagram

The input layer uses Word2vec technology to vectorize the text, and changes the comment data to the word vector form input model. It is then input to the CNN model, and the input is convolved with multiple convolutional layers. The pooling layer reduces data dimensions and preserves useful information. The attention layer integrates local features, global features and the weight of importance between sentences, and a higher level of semantic expression can be obtained. Use the classifier to get the final emotional tendency.

The LSTM model has five layers, the input layer, the LSTM layer, the attention layer, the fully connected and classified layer, and the output layer. The model structure diagram is shown in the following figure:

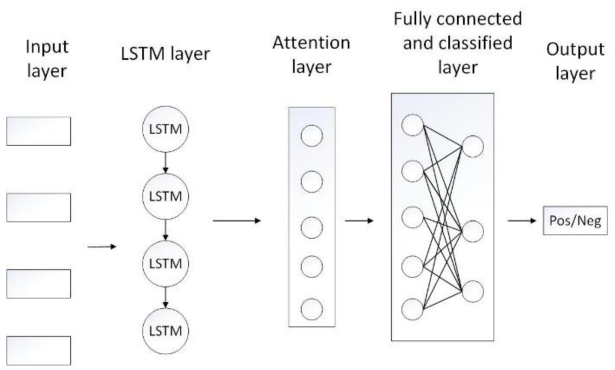

Figure 4. LSTM model frame structure diagram

The input layer of the model is processed in the same way as the CNN model, and then the LSTM layer is used to extract the global features of the review data, and the attention layer is also used to synthesize the local features, global features, and importance weights between sentences. Use the classifier to get the final emotional tendency.

4. Experimental analysis

4.1. Datasets

The dataset used in this article is the Weibo comment dataset published on the Internet. Weibo commented on a total of 1,19,988, of which 5,9993 were positive and 5,9995 were negative Strip.

4.2. Data Preprocessing

The data preprocessing stage has three main tasks: one is to analyze and filter unwanted data to correct irregular data. The second is to perform word segmentation processing on Chinese comment data, and this paper uses the nodule word segmentation tool to give word segmentation to the Chinese comment data, and uses the collection of four glossaries of "Chinese stop word list", "Baidu stop word list", "Harbin Institute of Technology stop word list", and "Sichuan University machine intelligence laboratory stop thesaurus" as reference use thesaurus Remove stop words from comment data. Third, unify the length of the comment data, supplement the part that is insufficient the standard length, and truncate the part that exceeds the standard length.

4.3. Text Representation

The input to a neural network model usually represents data in the form of a two-dimensional vector, and this model represents the review data as a matrix of word vectors. The CBOW of the Word2vec model is used to train the batch data with data preprocessing to obtain distributed lexical representation. By using this method to train word vectors, it is not only possible to express the different characteristics of words in detail, but also to reflect the similarity and correlation between words.

4.4. Experimental parameters

The model parameters are shown in the following table:

Table 1. Model parameter table

Batch Size	128
Epochs	3
Word vector dimension	128
Size of convolution kernel	(3,4,5)
Dropout rate	0.5
Activation function	Rlife
Theptimizer	Adam

4.5. Experimental evaluation and comparative experiments

4.5.1. Evaluation Methods

The evaluation parameters used in this model are accuracy (Accuracy, ACC) and cross-entropy loss function (C ross Entropy Loss,CEL）。

The equation for ACC is shown in (6).

$$ACC(i) = \frac{T_i}{N} \tag{6}$$

Among them, it represents the number of correctly classified Weibo comment $T_i$data, and N represents the total number of Weibo comments.

C EL represents the distance between the actual output and the desired output, and when the value of CEL is smaller, the better the model fits. The formula for calculating CEL is shown in (7).

$$CEL = -\frac{1}{N}\sum[y \ln a + (1-y)\ln(1-a)] \tag{7}$$

where a represents the true output, y represents the desired output, and N represents the total number of text.

4.5.2. Comparative experiments

Epoch is the number of times the algorithm designed in training is used on the dataset, and multiple iterations of the dataset are required in training to ensure the accuracy of the model. The accuracy of CNN and LSTM models under different epoch in this experiment is as follows:

Table 2. CNN and LSTM results

Epoch	CNN		TheSTM	
	Time-consuming (s).	Accuracy (%).	Time-consuming (s).	Accuracy (%)
1	661	0.9814	1427	0.9820
2	659	0.9813	1405	0.9809
3	651	0.9794	1371	0.9818

4.6. Results and analysis

The experimental results show that with the increase of the number of iterations, the accuracy of the CNN model decreases slightly, and the accuracy of LSTM is higher than that of CNN. From the experimental data, after the same number of iterations, the accuracy of the CNN model reaches 97.94%, and the accuracy of the LSTM model reaches 98.18%. The experimental results show that LSTM is better than CNN in the sentiment classification of Weibo comments. After replacing the traditional pooling method with the self-attention mechanism and improving the extraction and retention ability of important local vectors, the accuracy of both models was improved.

5. Concluding remarks

Aiming at the sentiment analysis of Chinese comments, this paper uses the CNN model and the L STM model to classify the sentiment of Weibo comment data, and the experimental results show that the LSTM model is better than the sentiment analysis of Chinese comments The CNN model has a higher accuracy rate. Because the self-attention mechanism can highlight the features related to sentiment classification, the text sentiment classification can be better realized. Therefore, this paper conducts experiments after introducing the self-attention mechanism in the original model, and the experimental results show that the accuracy of the model has been improved.

Acknowledgements

This article belongs to the research results of the innovation and entrepreneurship training project for college students of Anhui University of Finance and Economics.

Project Title: Chinese Review Sentiment Analysis Based on Deep Learning and Attention Mechanism.

Item number: 202210378303.

References

[1] Yang Shuxin, Zhang Nan Text sentiment analysis combining sentiment dictionary and contextual language model[J]. Journal of Computer Applications, 2021, 41(10): 2829-2834.

[2] XU Minlin. Research on text sentiment analysis combining sentiment dictionary and neural network[D] Jiangxi University of Science and Technology, 2020.

[3] WANG Zhiying. Research on text sentiment analysis based on Spark and machine learning[D] Central China Normal University, 2021.

[4] Yan Junchao, Zhao Zhihao, Zhao Rui Research on Social Media Text Sentiment Analysis Based on Machine Learning[J]. Information and Computer (Theoretical Edition), 2019, 31(20): 44-47.

[5] Xu Kangting, Song Wei A Chinese-text sentiment analysis method combining language knowledge and deep learning[J/OL] Big Data: 1-16[2022-05-04].

[6] XIE Caiyun,HE Mingzhi,ZHOU Panfeng,PENG Lin,LI Xuanying Research on Text Sentiment Analysis Based on Deep Learning[J]. Information and Computer (Theoretical Edition), 2022, 34(03): 84-86.

[7] WANG Hanqi. E-commerce platform review text sentiment analysis based on deep learning[D] Guilin University of Electronic Technology, 2021.

[8] WU Yinggang. Research on text sentiment analysis based on deep learning[D] Guangdong University of Technology, 2021

[9] ZHANG Tingting. Research on Chinese Review Text Sentiment Analysis Based on Deep Learning and Attention Mechanism[D]. Anhui Jianzhu University, 2021.

[10] YAN Weichen. Research on text sentiment analysis based on deep learning[D] Harbin University of Science and Technology, 2021.