Research on Emotional Tendency Analysis of Weibo Comments Based on Deep Neural Network

Xinmeng Chen

School of Software, Tiangong University, Tianjin 300387, China

xinmengchen121@foxmail.com

Abstract. With the rapid development of China's economy, social media has also developed rapidly, among which Weibo has grown into a more popular social media. Users will comment on hot events on the Weibo platform. Therefore, how to conduct sentiment analysis on user comments has become necessary research for relevant departments. Through natural processing technology, practical information can be extracted from the text of Weibo, which can provide appropriate decision-making for China's network security monitoring, prediction of potential problems, and product analysis. This paper proposes a method based on Word2Vec and a recurrent neural network improved for long and short-term memory networks to conduct sentiment analysis of Chinese Weibo comments through the integration of deep learning. And through experiments, the accuracy of sentiment analysis of Weibo comments in this article has reached 92%. Through the research of this article, we can better guide and supervise the sentiment tendency of public opinion for relevant departments.

Keywords: Weibo, Comment analysis, Deep neural network, Long and short-term memory network, word vector.

1. Introduction

With the development and in-depth application of information technologies such as China's mobile Internet and mobile communications, online users have surged on major social media platforms. This is one of the important ways for online user information interaction. Online comments have emerged as the times require, and the number is enormous. Online reviews are comments made by users on a specific topic on a particular platform. They are the feelings and evaluations formed by users based on their use or experience of products/services (Shan, X, 2018). Online reviews are rich in users' feelings and information, involving their preferences, needs, and evaluations, and have a high value of attention. On the other hand, because Weibo has 160 million daily activities and at the same time maintains a constant growth rate, users' comments on Weibo have attracted more attention (Niu, C. M., G. H. Zhan,2018) Based on this, the sentimental analysis of the comment text of Weibo has great practical significance.

Sentiment analysis of text reviews has received a lot of attention and research results at home and abroad. Some scholars use dual-channel convolutional neural network character embedding to perform sentiment analysis of Chinese Weibo(Chen, S,2018). However, due to the late start in China, there are few research results on Weibo comments， Some Chinese scholars adopt the sentiment analysis of Chinese Weibo with multiple sentiment dictionaries and rule sets(Wu, J, K. Lu,2019). Currently commonly used shallow text analysis methods: a text sentiment analysis based on sentiment dictionary, this effect depends on the quality of sentiment dictionary, building a dictionary often requires a lot of energy, and the tendency is evident in long, intricate sentences and non-emotional words When this method does not work (Zhang, L,2018). The second is based on traditional machine learning methods. This method is better than artificial results, but machine learning methods need to pay attention to the combination of features, which requires a lot of effort(Cho, K,2014). At the same time, the machine has limited processing capabilities for complex functions and cannot be considered in depth. The links within the text further affect the accuracy of the classification. Therefore, this article proposes a method of fusion deep learning to build the model of this article and then study the sentiment classification of Weibo comments.

2. Word2vec word embedding technology principle

In NLP, people use word vectors to represent the feature vectors of words, and Word embedding technology is a technology that maps words to real number domain vectors.

Assuming that the number of different words in a natural language system is N, each word corresponds to an integer between 0 and N-1, which is recorded as the index of the word. To construct a one-hot encoding vector of words, we build an all-zero vector of length N and add the i-th position of the term with index i is 1 (i is an integer between 0 and N-1) so that each word has a one-to-one correspondence with a vector of length N.

The goal of training the word embedding model is to obtain helpful word representations of the surrounding words in the target text. Given a series of training words w1, w2..., wT, the objective function is shown in formula (1).

$$max\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c,j\neq 0}Logp\left(w_{t+J}|Wt,\right)$$

3. Improved recurrent neural network model based on LSTM

A neural network contains at least three layers: input, hidden, and output. The number of features in the data set determines the dimensionality of the number of nodes in the input layer Volume; these nodes are linked through links called "synapses" and the nodes created in the hidden layer. For each input layer node, each link carries some weight. These rights mainly used to determine which inputs or signals may pass or fail, and these weights also reflect the strength or degree of the hidden layer. Each neural network is mainly by adjusting the weight of each synapse to learn.

The long and short-term memory network is a particular neural network with a memory data sequence. It can achieve a logarithm through some gates accompanied by a memory line according to the memory of early trends.

Each LSTM is a set of units or system modules and captures thedata flow. These units are similar to a conveyor line, connected from one module to another. A module that transmits past data and collects current data. Because in each unit

Some gates are used, so the data in each unit can be processed, filter, or add to the next team. Therefore, these gates are based on the neural network layer of the sigmoid function, which enables these units to allow data to pass through selectively or process data.

4. experiments

4.1 Data Sources

This article uses Weibo comment data, which mainly covers 6 different topics. Among them, there are 5080 positive emotions and 4250 negative emotions. In a particular sample, ensure that the number of words is as large as possible. Divide the forecast into a training set and a prediction set, approximately at a ratio of 3:1. Among them, 6620 pieces were selected as a training corpus, and the remaining 3310 pieces were used as a test corpus.

4.2 Building the LSTM framework

Chung, J used a deep neural network for verification(Chung, J，2014).This article also uses the Keras extension library to build the LSTM structure. Keras is a Python deep learning framework that can easily define almost all deep learning models, has a user-friendly API, and can quickly develop prototypes of deep learning models. The primary model of Keras is Sequential. Users can easily stack the network layers they need to form an overall model.

The specific structure is as follows:

1). Use word2vec technology to map words into 128-dimensional vectors for feature extraction.

2). LSTM module. The exact steps are: first, put the extracted features into the input unit, then send the data stream from the input unit to the hidden term, and at the same time send another data stream from the hidden team to the output unit. The memory term of the neural network is the hidden term. For a hidden term, use X1 to represent the input of the t-the step, then the current activation value of the unit s=f(UM+wt-1). Among them, f represents the activation function, and Relu is used in this experiment. The output of step t is obtained through the softmax layer.

The model training process and prediction process of this article can be summarized as:

(1) Extract the feature vector of related data to obtain training data.

(2) Import the data in training set into the set model in batches, continuously update the connection weights in each node of the network organization, and minimize the value of the loss function to minimize the error between the actual value and the fitted value, thereby Learn the nonlinear characteristics between input variables and output variables.

(3) Save the well-trained model parameters, input the data in the test set into the trained model to get the test value, and compare it with the actual marked result in the test set to get the accuracy rate. The specific parameter table is shown in Table 1.

5. Experimental results

This paper builds the LSAM model of Weibo comment text according to the above process and iterates 600 times. The experimental results are shown in the following table.

Table 1. Parameters of LSTM sentiment analysis model for Weibo comment text

parameter name/ Parameter value	Learning rate	Training batch size	Number of hidden layer units	Hidden layers	Word vector size
	120	68	120	6	230

The experimental results are given in the table and the LSTM model loss value, training accuracy rate, and test accuracy rate under specific iteration steps. It can be seen that the highest accuracy rate of the LSTM model within 600 iterations is 92%. And the training loss and the training accuracy rate are kept within the average value, verifying that the model can avoid overfitting and gradient disappearance.

Table 2. The experimental results

Number of iterations	Training loss	Training accuracy	Test accuracy
1	0.7034	51.36	51.53
2	0.7103	52.56	52.64
3	0.7001	53.45	54.43
4	0.7006	54.56	56.49
6	0.7005	58.67	57,42
98	0.4005	81.34	82.34
100	0.3867	82.54	84,21
102	0.3968	83.41	84,34
199	0.2506	84.26	86.70
200	0.2307	85.69	87.70
201	0.2106	87.98	88.69
599	0.0234	95.69	85,36
600	0.0362	96.83	85.64

6. Conclusion

With the interaction of network platforms and the diffusion effect of information, online reviews have exploded, and the difficulty of identifying information in online reviews has increased significantly. It has attracted widespread attention from the industry and academia. This article conducts an in-depth study on the emotional tendency of Weibo comments.

This paper integrates the idea of deep learning, extracts features based on Word2Vec, and implements sentiment analysis of Chinese microblogs using the LSTM-improved recurrent neural network method. And it has obtained a better test accuracy rate than previous methods, which proves that the model proposed in this paper can solve the short text sentiment analysis problem very well and has specific feasibility. Gal Y, Ghahramani Z thinks that we should not ignore its shortcomings (Gal Y, Ghahramani Z,2016).

References

[1] Shan, X., et al. "Identification Method Research on the Usefulness of Online Product Review." Journal of Beijing University of Technology(Social Sciences Edition) (2018).

[2] Niu, C. M., G. H. Zhan, and L. I. Zhi-Hua. "Chinese Weibo Sentiment Analysis Based on Deep Neural Network." Computer Systems & Applications (2018).

[3] Chen, S., et al. "Chinese Weibo sentiment analysis based on character embedding with dual-channel convolutional neural network." 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA) IEEE, 2018.

[4] Wu, J., and K. Lu. "CHINESE WEIBO SENTIMENT ANALYSIS BASED ON MULTIPLE SENTIMENT LEXICONS AND RULE SETS." Computer Applications and Software (2019).

[5] Zhang, L., et al. "A Recommendation Model Based on Deep Neural Network." IEEE Access 6.99(2018):9454-9463.

[6] Cho, K., et al. "On the Properties of Neural Machine Translation: Encoder-Decoder Approaches." Computer Science (2014).

[7] Chung, J., et al. "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling." Eprint Arxiv (2014).

[8] Gal Y, Ghahramani Z. Dropout as a bayesian approxima**-**tion: Representing model uncertainty in deep learning[C]//international conference on machine learning. PMLR,2016: 1050-1059.