Research on the Prediction of Popular Opinion Trend of Web News based on BP neural Network and LSTM

Jiale Wu*, Lulin Zhang and Fulian Yin
Communication University of China, Department of Science and Technology

*E-mail: wujiale0713@163.com

Abstract. Current prediction methods on public opinion trend of network news are often carried out according to human experience or traditional time series prediction, but they sometimes lack scientific principle and immediacy. This paper researches on the prediction of public opinion trend of network news based on deep learning. Firstly, we use the BP neural network and LSTM to predict public sentiment trend, then combine word2vec with LSTM to classify the text sentiment and finally use doc2vec algorithm and k-menas to cluster the text. Experiment result shows that the mean square error with good performance, which reveals the beneficial effect of our prediction method.

Keywords. Network news, Time series predicting, Deep learning, LSTM, Emotional classification.

1. Introduction
The rapid spread of online news in the era of self-media has brought an important influence on the spread of public opinion. However, the progress of existing public opinion monitoring is often not as fast as the speed of Internet information dissemination. However, with the development of deep learning technology, it has become the mainstream research technology in the field of natural language processing, which can capture the more comprehensive and in-depth implicit text features of public sentiment texts of social platforms.

The application of deep learning technology in the field of online news sensation mainly focuses on the sentiment analysis of popular opinion texts. In 2002, the emotion classification method based on the naive Bayes model, maximum entropy model and support vector machine model SVM was first proposed by Pang B at the EMNLP conference[1]. After that, the exploration of text sentiment classification centering on the traditional machine learning model began, and this method was also applied to the sentiment classification of network news and papers. In 2006, the sentiment dictionary method used word matching scores to calculate the sentiment of texts and applied it to the analysis of public opinion in network news[2]. However, this method cannot avoid the flaws out of context. In 2011, Latent Semantic Indexing (LSI), which aims to perform dimensionality reduction[3] on raw text data, but semantics still cannot be extracted by this method. In 2014,taking advantage of the CNN to classify text sentences and using single-layer CNN for text modeling[4] was proposed, which achieved good results under the condition of good quality of word vectors, but if there is a semantic noise in the original network news lyric text word vector, it would make the classification effect worse. In order to solve this problem, Luo Fan et al. proposed a Hierarchical RNN-CNN (H-RNN-CNN) model combining RNN and CNN[5], including double-layer RNN and top-level CNN, which contributes to
the long-text emotional orientation and cross-sentence information judgment[6]. Then, using CNN to capture local information, a network of attention models incorporating LSTM and CNN was proposed by Hu Zhaoju et al. to extract textual context information, and the AT-BL&C model was obtained by replacing the LSTM with the BILSTM that implements the attention mechanism, and then the better classification effect of the online news public opinion text was achieved [7-8].

Based on the above technical achievements, this paper proposes a method for further integrating the application of network news public opinion predicting and deep learning technologies.

2. Research on predicting Trend of Internet News Development

2.1 BP Neural Network and LSTM
Back Propagation Neural Network is a multilayer feed-forward network trained by error back propagation algorithm. The original default identity activation function is shown in Eq. (1):

$$f(x) = x$$  \tag{1}

The Rectified Linear Unit (ReLU), also known as a modified linear unit, is a commonly used activation function in an artificial neural network[9]. The ReLU function is shown in Eq. (2):

$$f(x) = \max(0, x)$$  \tag{2}

The ReLU function converges much faster than the sigmoid function or the tanh function. The disadvantage is that when a large gradient flows through the ReLU neuron and the parameters are updated, the neuron loses any data activation capability. The Leaky-ReLU function after improving ReLU is shown in Eq. (3):

$$f(x) = \begin{cases} x, & x \geq 0 \\ ax, & x < 0 \end{cases}$$  \tag{3}

The LSTM neural network model[10] is based on the RNN model to design the memory unit in the neural network node, which is equivalent to the "processor" for judging the usefulness of information. It is used to control the flow of information in the form of a gate structure. Only the information that meets the algorithm's authentication remain and the non-compliant information was deleted through the Oblivion Gate.

This design solves the problem of information disappearing or exploding when the RNN model transmits information over long distances, which can learn the characteristics of long-distance dependence and is suitable for network news public opinion prediction experiments.

2.2 Experiments and Comparative Analysis
The experiment of this paper takes the case of Jiangge, a network news event, as an example. It collected a total of 63 days of micro-index and Baidu index hot word search volume data before and after the fermentation of the Jiangge case and then obtained an integrated search index after the addition calculation.

This paper used the time series sliding window method to process the comprehensive search index into a form that can be used as a neural network input. For example, suppose that the BP neural network structure is 6-X-1, that is, there are 6 input nodes and 1 output node, which uses the 6-day time series as the prediction basis to predict the 7th day. This paper successively uses 6 consecutive days of data as input data for the prediction model of the development trend of network news events, and the 7th-day data is output as the predicted target data. At the same time, we used the first two-thirds of the collected network news event data as a training sample that constitutes a prediction for the development trend of network news events, and the latter one-third of the data as a test sample. Next, this paper compared the experiments of three layers of BP neural network structure in a single hidden layer and four layers of BP neural network structure in a double hidden layer, and we found that the mean squared error of the four-layer BP neural network structure 6-5-3-1 was the lowest, equal to 0.017396. In the case of consistent data conditions, the experimental results are shown in Table 1.
Table 1. Comparison of different transfer functions in BP three-layer four-layer neural network output layer

| Output Layer            | Identity activation function MSE | Leaky-ReLU MSE |
|-------------------------|----------------------------------|----------------|
| three-layer BP 6-5-1    | 0.019322                         | 0.015920       |
| three-layer BP 6-5-3-1  | 0.017396                         | 0.015938       |

Then we choose LSTM time recursive neural network for predictive analysis of data series and uses Python Keras to construct and train the LSTM model. Based on the LSTM time-recursive neural network algorithm, this experiment compares the prediction effect using different related input layer node time-steps and iterative times epochs. According to the comprehensive experimental results, it can be concluded that the optimal result of the prediction effect is the mean square error 0.0030. Figure 1 shows the trends of the original search data and prediction data of the comprehensive search index in the Jiangge case.

![Figure 1. The developmental raw data and prediction data trends of LSTM experimental jiangge case](image)

After a comprehensive comparison of the BP neural network algorithm and the LSTM deep learning network for the data set of the jiangge case, we can see that LSTM has a better prediction effect. Moreover, the mean square error that can be achieved after training is significantly reduced, and the fitting effect between the original data curve and the predicted data curve is better.

3. Research on the Prediction of Public Sentiment Trend

3.1 Sentiment Classification based on LSTM and sentiment index

The LSTM time recursive neural network model is good at processing sequence data and can learn long-distance data dependencies. Before putting network news lyric text data into the LSTM model, we need to first perform word vectorization of Word2vec. In 2013, using Word2vec to learn the vector representation of words was proposed by Mikolov et al. It is mainly divided into two methods, which are the CBOW (continuous bag of words) that predicts the target word with the context and the Skip-gram [11-12] that predicts the context with the center word. LSTM deep learning classification requires training and learning through training sets. Therefore, the field and quality of training set corpus in text sentiment classification influenced the training of the model to a great extent, and thus affect the classification effect. In order to fit the theme of this paper's sentiment analysis based on social news media in this paper, we introduced the Chinese microblog sentiment analysis of the 2012 CCF Natural Language Processing and Chinese Computing Conference under the e-commerce sentiment training data of various fields with a large amount of test data. We add the parsed emotional text data of social media comments to the original Chinese sentiment analysis corpus, and then iteratively train the LSTM model with the merged corpus to obtain the sentiment classification model. Finally, we divide the overall Weibo lyric corpus data into daily text collections by date and then put it into a model based on deep learning, so we can get the network news sensational daily positive emotional text quantity posnum and negative emotional text quantity negnum. This paper defines the daily emotional index to reflect the proportion of positive sentiment texts and the calculation method is shown in Eq. (3):
3.2 Algorithm for text vectorization

The categorization of web news lyric text topics requires first processing the text as a vector based on Word2vec. In 2014, doc2vec, or paragraph2vec, was proposed by Le, Mikolov or some others. This is an unsupervised framework, which can get text vectors and expand Word2vec. The learned vector can calculate the similarity between texts by calculating the distance. Doc2vec's two models: DBOW (distributed bag of words) and DM (distributed memory) correspond to CBOW and skip-gram in Word2vec respectively. The DBOW model is the probability of predicting a set of random words in a document given a document vector, while the DM model is the probability of predicting words given a context and a document vector.

This paper vectorizes the online news lyrics and then perform K-means clustering. The K-means algorithm is the most classical clustering method. It sets the K points in the space as the cluster center, and classifies the closest objects, and then iteratively calculates and updates the values of each cluster center. Finally, until the center point position becomes stable, the number of custom classifications is Clustering is complete.

In order to determine the appropriate number of clusters for the network news public opinion topic, we set up different clustering numbers for comparison and then perform text processing on the classified network news comment subsets: text segmentation, text preprocessing and word frequency calculation. The specific calculation method is: We select the top 30 words, and then calculate the topic overlapping high-frequency words again and again. After calculating the average of all combinations, the obtained value is the average high-frequency word repetition rate. This paper proposed the average high-frequency word within the suitable range with the lowest repetition rate, where the topic difference is relatively largest, and the clustering value is clustered as the topic classification number.

3.3. Experiments and results

Set the number of nodes in each hidden layer unchanged to change the number of input layer nodes, which is the number of days related to the prediction. Finally, we find that the mean square error of the four-layer BP neural network structure 6-5-3-1 is the lowest, equal to 0.026184, under the condition that other data conditions are consistent. Comparing the experimental results of the different nodes in the hidden layer, this paper can conclude that the 6-2-1 mean-squared error of the three-layer BP neural network structure is relatively lowest, which is approximately equal to 0.024151.

By comprehensively comparing the BP neural network algorithm and LSTM deep learning network experiment results, we can conclude that when the set value time-steps is 5 in the LSTM network experiment, which means the LSTM set prediction output data is related to the data of the first 5 days, and the mean square error is the lowest, which is 0.0200.

When the number of network news media clustering topics is specified as 2, 3, 4, and 5 respectively, text data vectorization and clustering experiments are performed and the average high-frequency word repetition rate is calculated. According to the experimental results, this paper selected the results with a relatively large degree of thematic differences, which means the number of cluster classifications defined at this time is 2. The experimental results are shown in Table 2.

| Table 2. Different cluster classification number average high-frequency word repetition rate comparison results |
|---------------------------------------------------------------|
| Number of clusters  | 2     | 3     | 4     | 5     |
| The average repetition rate of high-frequency words | 0.60   | 0.69   | 0.63   | 0.62   |

4. Conclusion

This paper introduces deep learning technology into the research of network news dissemination effect
predicting, and analyzes and studies the development trend of network news events and the prediction model of online newsgroup sentiment trends. This paper takes jiangge case, a popular online news incident, as an example. Firstly, we use the BP neural network and LSTM to predict public sentiment trend, then combine word2vec with LSTM to classify the text sentiment and finally use doc2vec algorithm and k-means to cluster the text. Based on the Python language, we conduct prediction experiments on various modules of network news events. Among them, the best result of the network news event trend predicting experiment is the mean square error of 0.0030, and the best result of the network news public opinion sentiment trend predicting experiment is the mean square error of 0.02, which has achieved a good experimental effect.

ACKNOWLEDGMENT
This work is supported by the National Natural Science Foundation of China (No. 61801440), the Fundamental Research Funds for the Central Universities and the Communication University of China’s state-of-the-art training research project (No. CUC18A015-1).

References
[1] Pang B, Lee L, Vaithyanathan S. 2002 Sentiment Classification Using Machine Learning Techniques[C] Proc of EMNLP 79-86.
[2] Kim SM, Hovy E. 2006 Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text[C] Proc. of Wksp. on Sentiment and Subjectivity in Text at Proc. of the Int'l Conf. on Computational Linguistics the, Ann. Meeting of the Assoc. for Computational Linguistics:1-8.
[3] Bader B W, Kegelmeyer WP, Chew P A. 2011 Multilingual Sentiment Analysis Using Latent Semantic Indexing and Machine Learning[C] IEEE, International Conference on Data Mining Workshops. IEEE Computer Society,45-52.
[4] Kim Y. 2014 Convolutional Neural Networks for Sentence Classification[J]. Eprint Arxiv, 31(1):132-164.
[5] Luo Fan, Wang Houfeng. 2018 Chinese Text Sentiment Classification Based on RNN and CNN Hierarchical Networks[J]. Journal of Peking University(Science and Technology), (4):1-6.
[6] Hu Zhaoju, Zhao Xiaowei. 2018 Sentiment Analysis Based on Word Vector Technology and Hybrid Neural Network[J]. Research on Application of Computer, (12):1-6.
[7] Huang Jianguo, Luo Hang, Wang Houjun, et al. 2009 Application of GA-BP Neural Network to Study Time Series Prediction [J]. Journal of University of Electronic Science and Technology, 38(5):687-692.
[8] Ding S, Jia W, Su C, et al. 2010 An Improved BP Neural Network Algorithm Based on Factor Analysis[J]. Journal of Convergence Information Technology, 5(4):103-108.
[9] Jiang Anbo, Wang Weiwei. 2018 Research on Optimization of ReLU Activation Function[J].Transducer and Microsystem,37(02):50-52.
[10] Hochreiter S, Schmidhuber J. 1997 Long Short-Term Memory[J]. Neural Computation, 9(8):1735-1780.
[11] Mikolov T, Chen K, Corrado G, et al. 2013 Efficient Estimation of Word Representations in Vector Space[J]. Computer Science, 31(1):132-164.
[12] Pan Bo, Zhang Qingchuan, Yu Zhongzhong, Cao Shuai. 2018 The Application of Doc2vec in Salary Forecast[J].Application Research of Computers, 35(01):155-157.