Scientific
Research
Publishing

Research on Network Group Event Recognition Based on Neural Network

Haiying Ma

Department of Management Science and Engineering, East China University of Science & Technology, Shanghai, China
Email: c_mhy@163.com

Abstract

Sina Weibo is one of China's most popular SNS platforms, which has a large number of user groups. Nowadays, the public opinion on Sina Weibo has great influence and significantly affects the social reality. This paper defines the network group events from aspects of the nature of the incident and influence and establishes a neural network of pattern recognition model that is used to fit the model. The model can recognize the network group events through continuous three days' various topics growth rate. By programing a crawler, we get the data of 23 hot events on Sina Weibo from 2011 to 2015 for neural network training. The training results of pattern recognition model show that the accuracy rate of the judgment of the Internet Group events and the non-network group events is 72.7% and 75% respectively. The overall accuracy was 73.9%.

Keywords

Sina Weibo, Network Group Events, Hot Event, Pattern Recognition Model

1. Research Background

Micro-blog is a platform for information sharing, communication and access based on user relationship. Micro-blog as a sharing and exchange platform, it emphasizes more on timeliness and randomness. Micro-blog can express every thought and the latest developments [1]. According to the thirty-sixth "Development of Chinese Internet Statistics Report", as of September 2015, China's netizens reached 668 million, the number of monthly active micro-blog has reached 222 million and active users reached 100 million, with continuous improvement of the micro-blog platform function, micro-blog users have a gradually steady and sustained growth.

Such as the 2011 Meimei Guo event, as Meimei Guo showed her rich in the micro-blog, her micro-blog identity is a general manager of Chinese Red Cross commerce.

This information had a lot of forwards, causing human flesh search of Meimei Guo and distrust of the Chinese Red Cross. According to statistics, since the Meimei Guo incident, the amount of social donations fell, in July the number of social donations is 500 million yuan, compared with June, a decline of more than 50% [2].

It is because of the huge user groups and the rapid spread of public opinion and the strong resonance of a public opinion. It can sometimes produce an effect of the effect of micro-blog. Public opinion effect has positive and negative aspects. Negative effects will have a very bad impact on society. With the rapid increase of China's Internet users, similar events will be more likely to occur, and the impact will be greater. Therefore, the micro-blog Sina's public opinion monitoring has great practical significance. In this paper, we use artificial intelligence technology to predict the network group events. It can help the government to control public opinion for the first time, to prevent the network mass incidents of social harm.

2. Research Method

At present, micro-blog's research is divided into emotional analysis and public opinion analysis. In the analysis of public opinion, most of the literature is the use of infectious disease model to predict the amount of micro-blog forward, or micro-blog's public opinion level classification. Using the model of infectious disease problem is the difference to the network group events topic (this will cause the transmission coefficient and the coefficient of change and cure) and the impact of derivative public opinion on the duration of the incident. In fact, this article also really found that the use of infectious disease models for micro-blog research is focused on an event, or focus on a post [3]. In recent years, some scholars began to use the neural network for micro-blog research. This is mainly because in recent years, micro-blog has produced more data can be used for neural network training. And the neural network itself has the advantage: strong nonlinear fitting ability, can map any complex nonlinear relationship, and learning rules are simple, easy to achieve computer [4]. Neural network has strong robustness, memory ability, nonlinear mapping ability and strong self-learning ability. But in terms of prediction, these studies are focused on a single event or post, so this paper attempts to train a neural network that can predict a variety of network group events.

So far, there are few literatures about the analysis of the network group event recognition. In this paper, the classification of public opinion on the network group events mainly used micro-blog Sina advanced search. By defining the time nodes and key words we can get popular micro-blog Sina's total number of topics, forwarding the number of topics, the number of the certified number topics, the number of including links and other data [5]. These data are easy to obtain, and this paper believes that these data well contains the characteristics of the network group events information, so it will be based on these indicators to establish the network group events prediction model.

3. Data Preparation

According to the above, this paper will use the data of every day for a period of time

Sina micro-blog hot events after the outbreak, including the total number of topics, the original topic number, including picture topic number, etc. this paper write a crawler to obtain the required data.

According to this program, the web crawler can choose to access the web page of the worldwide web, and automatically get the information in the web page. One of the most basic web crawler mainly contains three modules: access to the source code of the target web page, matching the desired content, saved the content for the data analysis and use.

This article uses the python3.5 to compile the reptile procedure. The target page is http://weibo.cn/search. The target page is a search page for micro-blog mobile. In this paper, the search interface can choose to enter the key words, types, user types, nickname, event, sort, etc.

First through the Baidu, we search micro-blog event, micro-blog popular events, etc.. Then we selected some of the hot topics of micro-blog, the time span is from 2011 to 2015. For example, saving the children begging, Meimei Guo incident, 723 motor car accident, China's tongue, such as events. Then through the Baidu encyclopedia and related reports we determined the time of the whole hot. Then used the developed crawler we crawled the hot time data. In climbing, we use the keyword method to search. At the same time, because some events records have been removed by micro-blog Sina, so before determining to climb to take an event, we search for all records between the occurrence of the event and the end time. If the number of topics is too small, the lifting of the event is canceled. As the keyword search method is used, so the inevitable data may produce some unnecessary noise. After grabbing some data of the events, we found that most of the micro-blog hot spot event duration is less than 30 days. Therefore, we think that 30 days is the time span of micro-blog hot events. Taking this as a basis, we climbed the hot spot events data from the beginning of the events in 30 days. Finally, we climbed up 23 hot events of the micro-blog, as well as 30 days data of the event before the start of the event.

4. Data Preanalysis

We save the data from the crawl to the excel file. Then we select a number of hot events to draw the total number of topics of the graph. As shown in Figure 1, we can find that the number of topics in different micro-blog hot events is different. Xinjiang nut cake event, after the outbreak of the event, the topic number quickly rose to the top, then the number of topics related reduced rapidly. Compared with the Anti Japanese demonstrations, the number of topics in the climax of the two events is almost the same, the rate of rise is almost, but the decline of the Anti Japanese demonstrations is volatile. In addition, the total number topics of different hot event is in the 0 ~ 60 million, a number topics of hot events is large (Meimei Guo, Malaysia Airlines, Hui Tang, Rescue beggars children, Zaike), a number topics of hot events is small (Anti-Japanese march, Nut cake, Ice bucket challenge, Di Yao, Money willful).

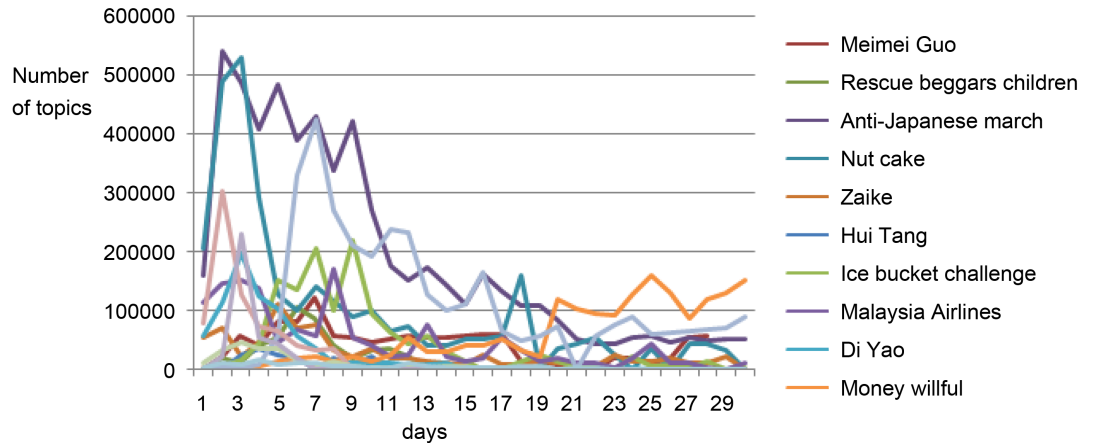This is due to three consecutive days of various topics to predict the growth rate of

Figure 1. The total number of topics of the graph.

the network group events, so it can effectively avoid the influence due to total different trends and topics of different magnitude of different hot events. In addition, from the structural properties of the data obtained, and some of the conclusions of the literature are the same, and we can find the cause of this structural characteristics. To a certain extent, it shows that the data contains the characteristics of micro-blog hot events information. It can be used as a data sample for micro-blog prediction research.

5. Empirical Analysis

The neural network pattern recognition toolbox of matlab 2015 is used in this paper. As shown in **Figure 2**, we use three layers BP neural network (input layer, hidden layer and output layer), the transfer function is the sigmoid nonlinear transfer function, the output layer is a function of soft max multi classification function, training algorithm is the Scaled conjugate gradient algorithm. In this paper, the number of hidden layer nodes is set to 6.

The training process is shown in **Figure 3**, the blue line is the error of neural network training, the red line is the error of test samples, the green line is the error of validation samples. When the neural network is iterated to 15 times, the error of the sample data is not decreased, and the error of the sample data is not decreased for 6 times. The training results are best for the ninth iteration, the error of the sample is the smallest.

Figure 4 is the discrimination diagram. The upper left of the figure is the training result matrix, there are 7 pieces of network group events is right, there are 8 non-network group events is right, there are 3 pieces of network group events was mistaken for non-network group events, there are 3 non-network group events were misclassified as groups of network event. The upper right corner of the graph is the result matrix. The lower left corner of the graph is the test result matrix. From the graph, we can see that they are all right to judge the network group events and non-network group events. Overall, the accuracy of the training results in the first class of events (the network group events) was 72.7%, the accuracy rate of the second categories of events (non-network group events) was 75%, the overall accuracy rate was 73.9%.
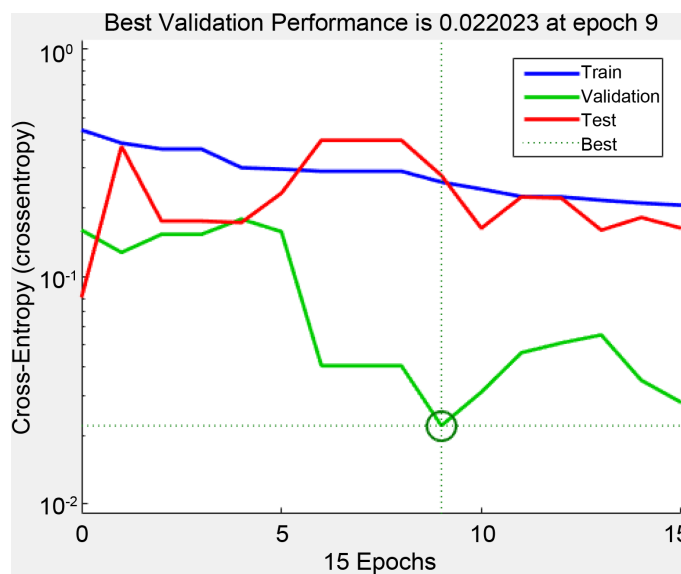
Figure 2. Neural network settings.
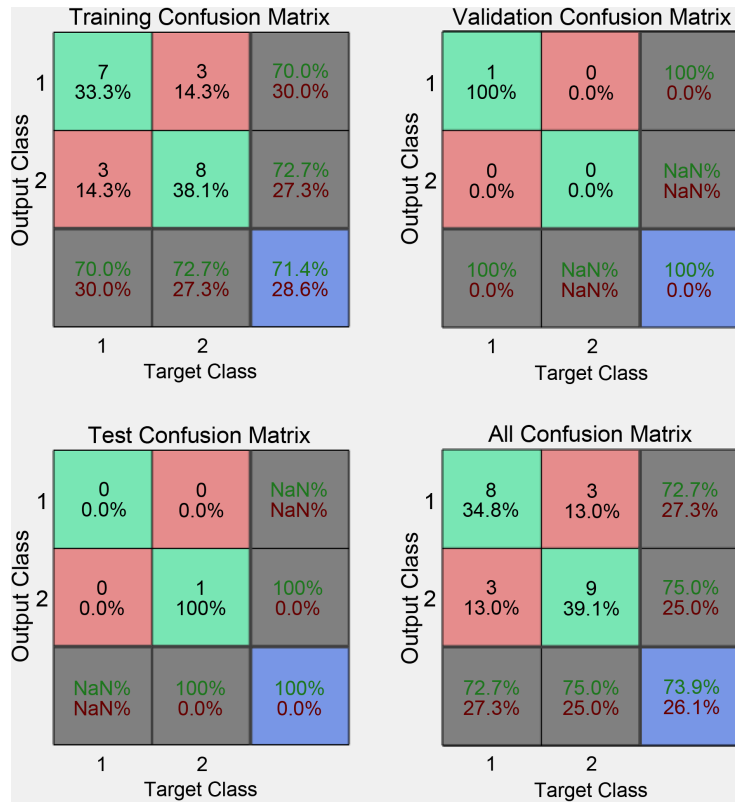

Figure 3. Training chart.

Figure 4. Discrimination diagram.

6. Conclusions

In this paper, through the preparation of a crawler program, we climbed up 23 hot events of the micro-blog, as well as 30 days data of the event before the start of the event. Then we took the 23 pieces of micro-blog hot events into the network group events and non-network group events. We use pattern recognition toolbox of neural networks. Then we input data of three consecutive days, the total number of topics, the number of original topic, including pictures and the link topics number, the number of users of the authentication, to determine whether the event is the network group events. The accuracy of the training results in judging the first class events (network group events) was 72.7%, the accuracy rate of the second categories of events (non-network group events) was 75%, and the overall accuracy was 73.9%.

In this model, the input sample is three days data before the highest growth rate, so in the actual use, the government can monitor three consecutive days data of an event to determine whether the event will be the network group events, and get the prediction results in the day before to get the highest.

But the neural network model is still inadequate, and the accuracy of the judgment is not particularly high. This may be because the number of topics from the network group events and non-network group events may be relatively vague. In addition, more indicators should be added such as emotional indicators, event types, etc., as well as increased training samples to improve the accuracy of the model's discrimination.

Acknowledgements

References

[1] Guan, W., Gao, H., Yang, M., *et al.* (2014) Analyzing User Behavior of the Micro-Blogging Website Sina Weibo during Hot Social Events. *Physica A Statistical Mechanics & Its Applications*, **395**, 340-351. https://doi.org/10.1016/j.physa.2013.09.059

[2] Yu, L., Asur, S. and Huberman, B.A. (2012) Artificial Inflation: The True Story of Trends in Sina Weibo. *Computer Science*, **3**, 35-40.

[3] Wang, X., Zhang, Z., Xiang, Y.U., *et al.* (2015) Finding the Hidden Hands: A Case Study of Detecting Organized Posters and Promoters in SINA Weibo. *Wireless Communication of China*, **12**, 1-13.

[4] Cui, X., Shi, H. and Yi, X. (2014) Application of Association Rule Mining Theory in Sina Weibo. *Journal of Computer & Communications*, **2**, 19-26. https://doi.org/10.4236/jcc.2014.21004

[5] Su, Q., Huang, J. and Zhao, X. (2015) An Information Propagation Model Considering Incomplete Reading Behavior in Microblog. *MpraPaper*, **419**, 55-63. https://doi.org/10.1016/j.physa.2014.10.042