9

Forecasting Technology of Spatio-temporal Changes of Water Pollution Public Opinion

Anousavanh Sayavong [1] Zhou Haiwei [2]

Received: October 18, 2021
Revised: December 23, 2021
Accepted: December 24, 2021

Abstract

This article aimed to summarize online public opinion about water pollution from the Weibo website and builds a forecasting framework. Collected from Weibo Website, the data in this study are public opinion in the form of microblog from January 1, 2015, to December 31, 2017, through web crawler technology; for the spatial analysis, Beijing, Nanjing, Nanning, Huludao, Xianyang, and Dehong were chosen as examples to find the online public opinion data. The data were analyzed by using the backpropagation (BP) neural network and the fractal interpolation method. Through the comparison of the two methods, the forecasting accuracy of the number of microblogs

[1] Ph.D. Student from Laos PDR. Business School of Hohai University , Nanjing, Jiangsu, China(nonghaigroup@yahoo.com)

[2] Ph.D. Lecturer of Business School of Hohai University, Nanjing, Jiangsu, China (Huangdechun@hhu.edu.cn)

in 2017 by the BP neural network model is higher than that of the fractal interpolation model; but for the peak forecasting of a sudden increase in the number of microblogs, the fractal interpolation has done better. It can be said that the two models have their advantages in microblog public opinion forecasting due to their different theories. In the case of a spatial forecasting, the fractal interpolation model has less error rate in forecasting and the accuracy at a later point in time is more accurate than predicting the long-term public opinion. Therefore, the fractal interpolation model is suitable for predicting random events, but it needs to take into account the low forecasting results. The implications are included.

Keywords: public opinion; water pollution; forecasting; BP neural network; fractal interpolation

Introduction

At present, many countries are actively promoting the construction of water ecological civilization. Water pollution control is an important task for every country. Under normal circumstances, water pollution refers to water that pollutes the environment caused by harmful chemical substances that reduce or lose the value of water. There are few related studies on online public opinion about water pollution in China. The definition of “Online public opinion” in this study is the collection of cognition, attitude, emotion, and behavioral tendencies generated by the main body of public opinion on the internet.

This article clarifies the definition and constituent elements of online public opinion on water pollution. By transforming and categorizing microblogs on water pollution, it analyzes the temporal and spatial changes of those microblogs, constructs the public opinion monitoring framework, and analyzes those online public opinions about water pollution. The characteristics of microblogs in comparison have laid a theoretical foundation for future research on water pollution microblog public opinion. Public opinion analysis and forecasting are the current research hotspots. A large number of studies have used the backpropagation (BP) neural network to predict the time distribution of public opinion, but the fractal interpolation model has not yet been combined in public opinion forecasting. This research innovatively establishes a fractal interpolation model to predict the time distribution of public opinion. Microblogs about water pollution are predicted, and the BP neural network model and fractal interpolation model are used to analyze and predict the annual changing process of the number of water pollution microblogs and compare the applicability of the two methods. This study also predicted changes in the spatial distribution of those microblogs.

Literature Review

The studies on online public opinion in China are mainly divided into three levels: theory, technology, and application. At the theoretical level, Liu Yi's (2007) Introduction to Public Opinion Research Network is the first monograph in the research theory of Internet public opinion in China. It proposed that online public opinion

is the sum of various emotions, attitudes, and opinions expressed and disseminated through the Internet. Zhao Siyuan (2012) researched network mass events and their governance mechanisms and proposed a "government-led-mass participation" governance model of online public opinion. Lin Min (2013) started from the current status of domestic online public opinion, and explained the conditions for the occurrence of social clustering behavior by using the basic research thinking of empirical communication, thus fully explaining the multi-level factors and internal changes in the process of public opinion derivation.

At the technical level, Mei Zhongling (2007), who studied the public opinion analysis method of Web information mining, and found that this method can find hot spots in network activities in time. Duan Ning (2007) analyzed the space, time of occurrence, and changes in online public opinion; and proposed that changes in online public opinion are hierarchical, profitable, regionally diffuse, and transformative. In recent years, the number of papers on online public opinion technology has gradually increased. Liu Shuting (2017) studied public opinion collecting technology on the WeChat platform, and solved the problem of limited frequency of traditional servers through the combination of python language and Hook technology. Leng Bing (2017) proposed the text classification technology based on the Markov logic network. The classification effect is far better than the single-source domain classification method; the network public opinion technology proposed by Yu Li (2017) was based on improved

k-nearest neighbors. The algorithm proved through experiments to improve the accuracy of a topic classification and solve the problem of excessive dependence on keywords.

There is only one article about public opinion on water pollution (Wenhui, 2014). The research showed that public opinion on water pollution mainly concentrated in the following four levels: firstly, the early warning mechanism of water pollution incidents mainly strengthened water pollution legislation, government supervision, and the establishment of a coordination system for various departments. Secondly, the emergency mechanisms for water pollution incidents mainly included water pollution treatment methods and risk emergency strategies. This level showed water pollution on a technical level. The third was to put forward scientific countermeasures and responses to water pollution. The fourth was to study the response strategies of water pollution from a professional level, such as analyzing water pollution incidents from the environmental laws and regulations, information disclosure system construction, and human health. Previously, the focus of public opinion research on water pollution was mainly on early warning, emergency response, prevention, and countermeasures.

Forecasting is to find the essence of an object's development and changes, if any, after investigation. It is also to analyze the past, present, and future of any selected object to find the truth about it (Glynn, 1986). Predictive analysis is the investigation and study of the regularity of the development and changes of things based on

scientific analysis of things in the future (Hua, 2014). In 1992, Teruaki Ohnishi proposed a method to predict the change of trend of the public acceptance (PA) of nuclear energy through quantified variables. The parameters that affected the public acceptance are determined, and the predicted value of the future public acceptance is obtained through the mathematical model. Eren Bas (2015) proposed that the linear autoregressive model can produce more accurate forecasts than the fuzzy time series method, so he proposed a new hybrid forecasting method. The method is called the fuzzy time series network forecasting method (FTS-N) and an example analysis proves that the forecasting result is more accurate. Leemann and Wasserfallen (2017) explained that multi-level additive analysis and post-layering (MRP) are relatively mature methods in the evaluation system of online public opinion. On this basis, Leemann proposed a synthetic multi-level additive analysis and posterior stratification (MRSP). Compared with MRP, MRSP expands the range of data edge distribution and increases the forecasting accuracy.

Due to language differences, the rules of online public opinio n in different countries are also different. Therefore, the current research on public opinion forecasting should not only analyze foreign documents but also focus on the current research status of online public opinion forecasting technology in China. The opinion of public opinion forecasting in China first appeared in 2006. Ren Xinxin (2006) put forward the opinion of public opinion forecasting in the process of studying the early warning mechanism of public opinion. At present,

forecasting models such as differential equation model, time-series model, gray theory model, and BP neural network model are more common; Zhang Hua (2014) analyzed and compared the mentioned four models, and summarized the applicability, advantages, and disadvantages of each model.

Research Methodology

This research adopts the quantitative method in collecting data and uses the qualitative method in explaining the results. The data in this study are public opinion in the form of a microblog on the Weibo website. All the online public opinions about water pollution were collected from January 1, 2015, to December 31, 2017, using the crawler technology program. In data analysis, this study employs the backpropagation (BP) neural network model and fractal interpolation method as the main tools to predict the microblogs. This study focused on two aspects of predicting microblogs about water pollution, i.e., time and space.

Several microblogs on Weibo from January 1, 2015, to December 31, 2016, were used as the trial sample of the BP neural network; while the microblogs from January 1 to December 31, 2017, are the actual value. In the trial period, the BP neural network model can better simulate the changing habit of the number of microblogs from 2015 to 2016, but the simulation accuracy of checking any increasing number of microblogs is poor.

Based on the trial use of the BP neural network model from 2015 to 2016, the daily Weibo volume from January 1 to December

31, 2017, can be predicted. From the comparison between the predicted value and the actual number of microblogs, it can be seen that the BP neural network model can predict the time trend of the number of microblogs in 2017; the change of the number of daily microblogs from less than 100 to 300 can be forecasted with a one-day delay. The sudden increase of daily microblogs from about 100 to more than 400 is unpredictable. Taking the absolute value of the difference between the predicted value of the BP neural network and the actual value as the error, and the cumulative statistics of the error, the cumulative average error of the BP neural network model is 49.0.

The premise of using the fractal interpolation model to predict is that the time change of public opinion has fractal characteristics. The basic principle of fractal interpolation model forecasting is to use the state continuity of the time series to predict by fractal interpolation extension. According to the forecasting theory of fractal interpolation extension, the fractal interpolation model can better predict the change of the number of microblogs in 2017 overtime, but the predicted peak value is too small. There is also a delay on the upper side and there is a phase difference. Taking the absolute value of the difference between the predicted value and the actual value as the error, the cumulative average error of the predicted value of the fractal interpolation model is 52.

In the analysis of public opinion in different spaces, this study took Beijing, Nanjing, Nanning, Huludao, Xianyang, and Dehong as examples of cities from different regions, scales, and economic levels.

Analyze the spatial distribution of public opinion on microblogs of water pollution in the year. To predict the spatial change of general public opinion on water pollution, the microblogs from each city from January 1, 2015, to December 31, 2016, were selected. The water pollution in the Jialing River is chosen as an example and the microblogs about its water pollution were obtained through crawler software. Due to a large amount of data in searching for the keyword "water pollution" from May 5th to May 7th, the researcher extracted the microblogs through a combination of multiple keywords such as "Jialing River".

Since each microblog contains multiple field attribute information, the research had to extract the data to more detail, i.e., field user name, text information, user origin, and manually filter unrelated information. A total of 254 Weibo data from 18:00 on May 5th to 23:59 on May 7th, 2017 were obtained. For the forecasting of the spatial distribution of general public opinion, the BP neural network model with high accuracy for long-term public opinion forecasting is adopted. For the trial period, the input data is the number of daily microblogs from 2015 to 2016. First, all input data is normalized, and then the neural network is trialed and predicted.

Findings

The entire development process of microblogs that increased from the beginning-reaching the peak and then decrease is also a matter of great concern. From the changes in the number of

microblogs in 2017, it is found that there were 4 times that microblogs exceed 500 in a single day, namely January 6, 2017, June 3, August 15, and September 7, respectively. In the first two public opinion processes with a high number of microblogs, the forecasting results of the two models are compared. The methods and parameters used in the forecasting are the same as those of the one-year microblog public opinion process. The two events having many unprecedented opinions are described below.

(1) The first event: From January 1st to 13th, 2017, a water pollution topic suddenly increased due to the Chai Jing survey. There were 2,077 microblogs in 13 days, with an average of about 159.8 per day. The minimum number of Weibo in a single day is 4, and the maximum number of Weibo in a single day is 538 (January 6). The total number of microblogs predicted by the network model and the fractal interpolation model are both 10% and 20% smaller than the actual value.

(2) The second event: From June 1st to 9th, 2017, there was a sudden increase in microblogs about water pollution caused by children's drinking water. There were 1,550 microblogs in 9 days, with an average of about 172 a day, a single day The minimum number of Weibo is 48, and the maximum is 687 in a single day (June 3). The total number of microblogs predicted by the BP neural network model and the fractal interpolation model is 36.8% and 32% smaller than the actual value, respectively.

The number of microblogs regarding water pollution in

Nanning, Xianyang, Huludao, and Dehong were very few in 2015-2016. Therefore, the spatial distribution of water pollution microblogs forecasts results in Beijing and Nanjing in 2017 were analyzed. The neural network model can reflect the changing trend of water pollution microblogs in Beijing and Nanjing in 2017, but for the sudden increase or decrease, the forecasting results have a large deviation. For regional emergencies, it covers the influencing factors related to the event and the connection with the city and requires special research.

Sudden water pollution events (special events) often show the characteristics of small scale, high frequency, and rapidity, and are low-level behaviors. Special events of water pollution have the characteristics of unknown and randomness. Therefore, the spatial distribution of public opinion on special events cannot be predicted before the event, but the trend of the spatial distribution of public opinion after the event can be predicted. According to the experimental results, the forecasting results of special events through the fractal interpolation model are more accurate.

Through the analysis of the changes of online public opinion in various places, the number of microblogs in the incubation period is 0. So, the spatial change of public opinion is predicted from the point where the number of microblogs is not 0, that is, the time point t0+1 after the incubation period t0 ends. The number of water pollution microblogs in the five cities of Guangyuan, Nanchong, Chongqing, Chengdu, and Beijing were predicted according to the fractal interpolation model.

It is found that for the cumulative microblog process in each city, the fractal interpolation model forecast can reflect the growth process of microblogs, but the predicted value of the change in the number of microblogs in the five cities is smaller than the actual value. The number is low and the forecast deviation is small.

Combining the forecasting results of microblogs regarding water pollution with the GIS platform can display the predicted spatial distribution more intuitively, and show the density of microblogs in a certain area. The GIS platform displays the geographical distribution of public opinion on water pollution at each point in time. For the geographic information of public opinion, it can be standardized and converted into the latitude and longitude coordinates of the prefecture-level city to determine the location of the point coordinates. To prevent confusion between the collected past regional distribution of public opinion and the predicted regional distribution of public opinion, different file names can be used to distinguish them, such as “20170101 collections” and “20170101 forecastings”. By searching for relevant file names, you can check the real-time or predicted distribution of public opinion on that day. Select the China map and the prefecture-level city coordinate points to display on the ArcGIS platform, create a map of China with prefecture-level cities as the boundary, and import the excel form with the microblog public opinion regional divisions, and the GI platform can visually display the city where the dense public opinion is located at.

Results

1. Characteristics of fractal interpolation model in forecasting and analysis of its causes.

This paper uses a fractal interpolation model to predict the spatial distribution of public opinion about water pollution incidents in the Jialing River. According to the forecasting results of the fractal interpolation model, the characteristics and reasons are as follows:

1.1 The forecasting error rate is small. Because of the randomness of the parameters, the fractal interpolation model shows certain advantages in predicting uncertain events. Thus, the forecasting accuracy is high.

1.2 The test result is low. Although the experimental results show that the fractal interpolation model is more accurate in predicting the microblogs regarding sudden water pollution, the fractal interpolation model hardly predicts the increase of water pollution microblogs. The average number of public opinions is slightly lower than the actual measured value. Analyzing this phenomenon, it is believed that the reason is that step size $\sum$=0.01 is used to search for the forecasting point yN+1 step by step from 0 to 1 to find the yB corresponding to the minimum mean square error. Therefore, when predicting the cumulative number of microblogs, it will be lower than the actual value.

(1.3) The accuracy of predicting the value at a later point in time is more accurate than predicting the long-term public opinion. Due to the extension characteristics of the fractal

interpolation model, it is suitable for predicting the value at an unknown time point after the forecasting, but the accuracy of the results at several time points after the forecasting is low. This study uses an affine fractal interpolation model with a relatively simple structure and predicts the long-term public opinion for the whole year of 2017 using the same iterated function systems (IFS) to fit the forecasting results. If a long public opinion sequence is divided into multiple parts for forecasting, the parameters of the fractal interpolation model will be changed, which will effectively improve the accuracy of the forecasting results.

2. Analysis of the difference between the forecasting results of the BP neural network model and fractal interpolation model and its causes.

In this study, the BP neural network model and the fractal interpolation model are applied to the forecasting process of the temporal and spatial changes of water pollution public opinion, and the forecasting results are different, which are specifically reflected as follows:

2.1 The BP neural network model is more accurate in predicting public opinion for a long time, and the fractal interpolation model is more accurate in predicting a sudden increase in Weibo.

2.2 The BP neural network model is relatively inaccurate in predicting the spatial distribution of public opinion during the sudden increase process.

2.3 The fractal interpolation model is more accurate

in predicting the spatial distribution of public opinion in the process of a sudden increase in water pollution public opinion.

2.4 The fundamental reason for the difference in the above forecasting results lies in the different principles of the BP neural network model and the fractal interpolation model.

The BP neural network model is a forecasting model based on machine learning while the fractal interpolation model forecasting is a forecasting model based on the unique randomness and uncertainty of fractal theory.

Discussions

Contrary to the singularity of public opinion on daily policy and natural disasters, public opinion on water pollution contains long-term general public opinion with a certain regularity and public opinion with many attentions and has a complex composition. In this section, the public opinion on water pollution microblog is taken as an example to discuss the public opinion forecasting method.

1. Microblogs regarding water pollution have a complicated composition. Therefore, when predicting public opinion, general public opinion and emergency public opinion should be classified and discussed. The general public opinion of Weibo is predicted by the BP neural network model, and the fractal interpolation model is used to predict the emergency microblogs. Through the combination of the two public opinion forecasting models, one can maximize the strengths and avoid the weaknesses and obtain more accurate results.

The structure of Weibo's public opinion is non-uniform. In addition to the general microblogs with the characteristics of "fluctuations around a certain value" and the emergency microblogs with the characteristics of "rapid rise and rapid decline", there are also public opinions that mix two kinds of public opinion changes. Therefore, before predicting a certain kind of public opinion, it is necessary to specifically analyze the characteristics of public opinion changes and conduct classification discussions.

2. The geographical factor of the water pollution incident is an essential element. Microblogs are based on prefecture-level cities. When predicting the spatial distribution of microblogs about water pollution, the interaction and influence between cities must be considered.

The essence of forecasting microblogs regarding water pollution is to predict the density on the timeline of each city and summarize the topic distribution. Attention should be paid to the influence of spatial dependence on the true characteristics of the distribution pattern. The spatial effect produced by spatial dependence may be a large-scale trend or a local effect. The former is generally referred to as a first-order effect, which describes the overall variability of the mean value of a parameter. That is the global trend. It is a second order effect, which is produced by spatial dependence and expresses the tendency of neighboring values to converge with each other. It is obtained by calculating the deviation of the mean value. For example, the spatial process of infectious diseases requires a second-order

effect description. The study of spatial point patterns from a global perspective is mainly based on first-order velocities, which can be quantitatively described according to the density of the process, that is, the average amount of time per unit area. The influence of spatial dependence on the point pattern can be measured by the second-order property, that is, the relationship between points and the distance between points is used to describe.

Public opinion is spatially dependent on large-scale and high-level behaviors with a contagious nature, such as air pollution and infectious diseases. Changes in the spatial distribution of public opinion will center on the initial city and gradually expand outward, and cities will influence each other. Because the geographic location of the main body of the microblog public opinion is based on the prefecture-level city, the spatial distribution of this kind of public opinion during a certain period is predicted, that is, the spatial change of the prefecture-level city in a certain time is predicted, and finally the integration is performed.

From the previous research and the comparison results of the BP neural network model and the fractal interpolation model in this article, it is shown that choosing different model methods to predict the same thing, due to the different principles of the forecasting model itself, the predicted results cannot be the same. There are also differences in the error rate. The accuracy of forecasting results should be the primary basis for model selection. Taking the public opinion forecasting of water pollution emergencies in this study as

an example, the fractal interpolation model can only predict the results by a one-time unit in advance, and the period that can be predicted in the future is relatively short, but the results of the fractal interpolation model are more accurate. The fractal interpolation model is preferable when sending public opinion on the event. Since the significance of the results of public opinion forecasting lies in the early warning of public opinion, the most accurate forecasting result is fitted by choosing a suitable method. The length of time that can be predicted by the method is also an important reference when choosing a public opinion forecasting method. Performance differences such as the calculation time of the forecasting method are secondary considerations. In addition, in the process of public opinion forecasting, the most suitable model method should be used, or multiple methods should be combined to avoid errors caused by various model methods in the forecasting process as much as possible.

Conclusion

Based on the microblogging platform, this research analyzes the public opinion of water pollution and designs an experimental plan for predicting the public opinion of water pollution to predict the public opinion of water pollution on the microblogs by using the BP neural network model and fractal interpolation model to predict the time distribution of water pollution microblog public opinion. Based on the time distribution of water pollution microblog public

opinion from January 1, 2015, to December 31, 2016, the above two models were established to predict the time distribution of the whole year of 2017 and compared with the actual distribution of microblog public opinion in 2017. Through the comparison of the two methods, it can be seen that for the public opinion of the whole year, the forecasting accuracy of the number of microblogs in 2017 by the BP neural network model is higher than that of the fractal interpolation model; for the peak forecasting of a sudden increase in the number of microblogs, the fractal interpolation has done better. The forecasting result of the model is closer to the actual peak value than the BP neural network model. The daily minimum number of microblogs, the maximum number of microblogs, and the average error predicted by the fractal interpolation model are closer to the actual value than the BP neural network model. However, for the daily average number of microblogs, the BP neural network model outplayed the fractal interpolation method. The neural network model is close to the actual value. In addition, due to the characteristics of fractal interpolation theory, the theory is more suitable for the forecasting of continuous nonlinear curves. Therefore, for a large number of continuous-time intervals, the forecasting method of fractal interpolation theory can obtain more consistent forecasting results. The forecasting experiment of the time distribution of microblogs regarding water pollution through the BP neural network model and fractal interpolation model shows that the two models have their advantages in microblog public opinion forecasting due

to their different theories.

This paper also uses two models to predict the spatial distribution of general microblogs on water pollution and emergency microblogs on water pollution. The study found that the forecasting results of the spatial distribution of microblog public opinion in various cities through the BP neural network model are basically within the forecasting range, but the sudden increase in the peak is difficult to predict. For the emergency public opinion on water pollution, there is a connection between cities and cities. Cities are divided into infection source cities, infected cities, and media cities. The results show that the forecasting result of the fractal interpolation model has the characteristics of a small forecasting error rate, a low forecasting result, and the accuracy of predicting a value at a later point in time is more accurate than predicting the long-term public opinion. This is due to the randomness with parameters, the step-by-step search method with a fixed step size for the forecasting points, and the simple structure. Therefore, the fractal interpolation model is suitable for predicting random events, but it needs to take into account the low forecasting results.

At present, there are few studies on online public opinion on water pollution in China, and the development trend of online public opinion research should be integrated with various industries. Therefore, there is more room for the development of public opinion research on water pollution, which can focus on public opinion monitoring or the spread of public opinion. In addition,

the research of online public opinion can be combined with other industries besides water pollution, which promotes the development of various industries and the network environment. In addition, online public opinion can predict distribution outside of time and space, such as the microblog forwarding that brings connections to the social space, and there is still a lot of room for development in the research of microblog public opinion forecasting.

Bibliography

Bas, E., Egrioglu, E., Aladag, C. H., Yolcu, U. (2015). Fuzzy-time-series network used to forecast linear and no nlinear time series[J]. *Applied Intelligence,* 43(2);343-355.

Bing, L. (2017). *Research and implementation of network public opinion text classification technology based on Markov logic network* [D]. Nanjing: Nanjing University of Posts and Telecommunications.

Glynn, C. J. (1986). The Spiral of Silence: Public Opinion--Our Social Skin. By Elisabeth Noelle-Neumann. University of Chicago Press[J]. Social Forces, 64(3): 828-829

Hua, Z. (2014). *Research on Weibo Public Opinion Forecasting Model Based on Optimized BP Neural Network* [D]. Wuhan: Central China Normal University.

_____ (2014). *Weibo Public Opinion Forecasting Model Based on Optimized BP Neural Network* [D]. Wuhan: Central China Normal University.

Leemann L, Wasserfallen F. (2017) Extending the Use and Forecasting Precision of Subnational Public Opinion Estimation[J]. *American Journal of Political Science,* 61(4).

Li, Y. (2017). Online Public Opinion Technology Based on Improved k-Nearest Neighbor Algorithm [J]. *Automation and Instrumentation,* (06): 64-68

Min, L. (2013). *Internet Public Opinion: Research on Influencing Factors and Their Mechanisms* [D]. Hangzhou: Zhejiang University.

Ning, D. (2007). *The method for information mining in online public opinion* [D]. Harbin: Harbin Institute of Technology.

Ohnishi, T. (1992). A method to forecast quantitative variables re lating to nuclear public acceptance[J]. *Annals of Nuclear Energy,* 19(7):381-392.

Shutingm L. (2017). *Research on key technologies of public opinion based on social networks* [D]. Changchun: Jilin University.

Siyuan, Z. (2012). *Network mass events and their governance mechanisms* [D]. Changchun: Jilin University.

Wenhui, L. (2014). Network Public Opinion Report on Water Pollution Incidents in 2013 [J]. *New Media and Society,* (03): 196-213

Xinxin, R. (2006). Establish an all-round social early warning mechanism [J]. Reference for Middle School Political Teaching, (2);6-7.

Yi, L. (2007). *Introduction to Public Opinion Research Network.* Tianjin: Tianjin People's Publishing House.

Zhongling, M. (2007). Network Public Opinion Analysis Technology Based on Web Information Mining[J]. Journal of Chinese People's Public Security University (Natural Science Edition), (04); 85-88.