

Public Opinion Prediction Model of Food Safety Events Network Based on BP Neural Network

Chen Zheng¹, Yinghua Song¹ and Yaping Ma¹

¹School of Safety Science and Emergency Management, Wuhan University of Technology, Wuhan Hubei 430070, China.

Email: 445396142@qq.com

Abstract. The food safety incident network public opinion has the characteristics of wide audience, complex and changeable, and bad influence. It is of great theoretical and practical significance to study the behavioral and influencing factors of microblog public opinion forwarding in this kind of event. This paper summarizes and enriches the index factors affecting the network's public opinion forwarding volume. Combined with BP neural network algorithm, this paper constructs a network public opinion forwarding behavior prediction model, and applies and verifies it by crawling Sina Weibo food safety event microblog data. The results show that the introduction of fan activity has a certain weight ratio, which has a greater impact on the forwarding of public opinion in food safety events network with identity authentication, hot search, hypertext and other indicators, and the prediction model combined with BP neural network has a better prediction effect.

1. Research Survey

The rapid development of new media information technology provides conditions for the expansion of public opinion dissemination of food safety incidents. The network public opinion of food safety emergencies has the characteristics of wide audience, fast propagation, long duration and great social response. If the guidance is not appropriate, it is easy to cause serious adverse effects on social. At the same time, as a representative of the new media social platform, microblog users and their interaction behaviors, such as forwarding and praising, play an important role in the dissemination and development of public opinion. In recent years, the research on influencing factors of public opinion communication behavior on Twitter and Sina Weibo platform has been widely concerned by scholars at home and abroad. Through the study of twitter data, Suh [1] found that whether the microblog contains links and topic tags has a direct impact on the forwarding rate of microblog, while the number of followers, the number of followers and the length of registration of microblog authors have an indirect impact on the forwarding rate of microblog. Rudat [2] classifies the influencing factors into three aspects: user guidance, Weibo theme and information content, and determines the different influences of various factors. Zhang [3] uses classification algorithm to verify that mentioning other people, identity authentication and other factors in Twitter has a great positive impact on Weibo forwarding. Guan [4] introduces new content factors such as inserting pictures and links. Deng [5] proposed and proved that social influence index and user active time have a high influence on microblog forwarding. Although relevant scholars have carried out a lot of research on the influencing factors of Internet public opinion users' forwarding behavior, and studied the promoting effect of many factors on the public opinion communication, all kinds of emergencies have heterogeneity characteristics, and the influence of various factors on different types of emergencies may be different, which needs further research.



BP neural network and its optimization algorithm are more suitable for dealing with public opinion events with strong outbreak and rapid diffusion because of its excellent approximation effect, faster calculation speed and strong nonlinear fitting ability. Therefore, this paper takes Sina Weibo as the research platform, through analyzing the research results of food safety incidents and public opinion prediction, establishes an index system of influencing factors of online public opinion forwarding for such emergencies as food safety, and combines with BP neural network to build a prediction model to realize the accurate prediction of online public opinion forwarding behavior of food safety.

2. Influence Index Selection

In this paper, we select the index system which has certain influence on microblog forwarding and is easy to obtain from the existing literature, and verify the factor index which affects the online public opinion forwarding volume of food safety events in the new model. User influence index: identity authentication enables users to have different identification advantages in information dissemination. According to the difficulty of authentication, they are represented by 0, 1, 2 and 3. The number of fans, number of concerns, and registration period are more intuitive to reflect the accumulation of users' fans and influence, which can be represented by their real value. High fan activity can maintain a high exposure rate of information in the microblog platform. This paper uses the weighted sum of the average number of forwarding, comments and likes of each microblog, and its calculation is shown in formula (1), where w_1, w_2, w_3 , are weight coefficients, which are represented by the calculated actual value; user activity is directly expressed by the accumulated real number of microblogs every month.

$$F = \frac{\sum (w_1 A + w_2 B + w_3 C)}{D} \quad (1)$$

F is fan activity, A is the number of forwarding, B is the number of comments, C is the number of likes.

Table 1. Impact indicators and their characterization

impact index		indicator state	characterization value
user influence index	identity authentication	No, orange v, gold v, blue v	0, 1, 2, 3
	number of fans	M	m
	attention number	M	m
	age of registration	M	m
	Weibo volume	M	m
	user activity	M	m
	original proportion	[0,1]	m
	fan activity	M	m
content influence index	emotional tendency	Negative, neutral (no comments), positive	-1, 0, 1
	visual content	Text only, pictures, video	0, 1, 2
	hot search or not	no, yes	0, 1
	super topic or not	no, yes	0, 1
	is there a web link	no, yes	0, 1
	@ others people or not	no, yes	0, 1
	time interval	2: 00-8: 00, 8: 00-14: 00, 14: 00-20: 00, 20: 00-2: 00	0, 1, 2, 3
	Weibo length	M	m

In terms of microblog content, the selected indicators include whether there are pictures, whether

there is visual content in video, whether there is hot search, whether there is super topic, whether there is web link, whether there is @ others. These additional information increases the diversity of microblog content, increases the possibility and way of information acquisition for other users, and improves the amount of information contained in the content. Table 1 shows the indicators of each influencing factor and their specific characterization values, and the indicator characterization values in the table are the real values shown in M.

3. BP Neural Network

BP neural network is a kind of neural network which can learn and store a large number of input-output mode mapping relationships without establishing mathematical equations in advance. It has a strong nonlinear fitting ability. Its structure is shown in Figure 1, including input layer, hidden layer and output layer. The data is imported into the network at the input layer, and then transmitted forward. Each hidden layer will generate connection weights to build the model by itself. The fitting results will be compared with the actual values, and the backward feedback error will be fed back. The hidden layer will continue to adjust the weight self-learning combined with the feedback error, and this process will continue to cycle until the preset fitting effect is achieved. When the fitting effect is achieved, the network will have the minimum feedback error, and the connection weight in the hidden layer can be used to judge the important proportion of each input layer neuron in the network.

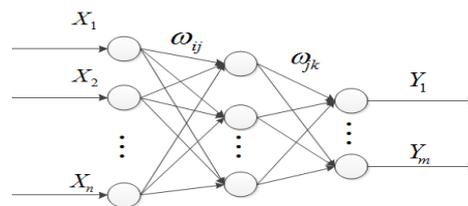


Figure 1. BP neural network structure

The number of neurons in the input layer and the output layer is determined according to the index number. The hidden layer is an important link between the preceding and the following in the BP neural network, but there is no definite criterion for the number of neurons in the hidden layer, so the number can only be adjusted according to the empirical formula and the actual test. The empirical formula is as follows:

$$H = \sqrt{in + out} + a \quad (2)$$

“in” is the number of input layer cells, “out” is the number of output layer cells, “h” is the number of hidden layer cells, and “a” is an integer between 1 and 10.

According to formula (2), the hidden layer is set from 5, and each group of experiments is carried out 20 times. The results show that when the hidden layer neuron is 6, there is the minimum training error mean. Therefore, this paper constructs a three-layer BP neural network of 16-6-1.

4. Experiment and Result Analysis.

4.1. Data Sources and Processing

According to the hot search events and event database provided by Sina Weibo, this paper crawls and collates the Weibo data of 12 typical food safety emergencies in 2019, including "Tongrentang expired honey", "Muji good carcinogenic food". 600 of the 1500 processed valid data are randomly selected as training samples and verification samples, and the remaining data are selected as prediction samples. At the same time, the micro blog forwarding volume is divided into different popularity criteria. In this paper, the forwarding volume less than 100 is set as low popularity, 100-1000 as medium popularity, and 1000 above as high popularity, with corresponding indicator values of 0, 1, 2.

4.2. Analysis of the Impact of Indicators

This paper uses MATLAB 2016 as a simulation tool. Through the learning and training of sample data, the weights of each layer in the network are obtained. The normalized result of the sum of the absolute values of the connection weights from the input index to all hidden layer nodes is taken as the influence weight of the index on the micro blog forwarding volume. The calculation formula is as follows:

$$\omega_i = \frac{\sum_{j=1}^n |\omega_{ij}|}{\sum_{i=1}^m \sum_{j=1}^n |\omega_{ij}|} \quad (3)$$

$$i = 1, 2, \dots, m \quad j = 1, 2, \dots, n$$

m is the number of input indicators; n is the number of nodes in the hidden layer.

Table 2. Weight of each indicator

index	connection weight						influence weight
identity	-0.301	-0.202	0.736	-0.838	0.405	-0.277	0.071
number of fans	-0.326	0.665	0.56	-0.118	-1	-0.348	0.0776
attention number	0.026	-0.748	-0.105	0.779	0.057	-0.133	0.0476
age of registration	0.5	0.301	0.426	0.54	-0.001	-0.449	0.057
Weibo volume	-0.11	0.308	-0.153	0.301	0.376	-0.283	0.0394
user activity	-0.414	-0.52	-0.965	-0.229	0.862	0.408	0.0874
original proportion	-0.165	-0.23	-0.333	0.804	0.004	0.043	0.0406
fan activity	0.39	-0.535	-1	0.112	-0.995	-0.154	0.082
emotional tendency	0.112	-0.511	0.848	0.063	0.209	-0.389	0.0549
visual content	0.634	-0.486	-0.303	0.478	-0.2	-0.091	0.0564
hot search or not	0.175	0.859	-0.532	-0.579	-0.284	-0.32	0.0707
super topic or not	0.579	-0.647	-1	-0.117	-0.659	-0.814	0.0982
is there a web link	-0.049	-0.419	-0.679	0.038	0.69	-0.162	0.0524
@ others people or not	-0.518	-0.229	-0.723	-0.45	-0.228	0.15	0.0591
time interval	-0.864	0.353	-0.153	0.25	-0.133	-0.285	0.0524
Weibo length	0.092	0.634	-0.783	0.053	-0.333	-0.179	0.0533

Table 2 shows the connection weight and influence weight of each index. The average influence weight of each index is 0.0625, among which, the six indexes such as user's identity authentication, number of fans, monthly average activity, fan activity, whether microblog has hot search, whether microblog has hypertext and so on all exceed the average, indicating that the high activity of users and their fans has a greater role in promoting the dissemination of public opinion information. Whether there is a super topic in the content of microblog has the highest weight of 0.0982, while the super topic students have the functions of settling related high-quality topics and gathering fans who are interested in such topics, etc., and have the conditions to create high popularity public opinion, so once such microblog appears, it needs to focus on monitoring. In addition, this paper introduces a fan activity weight of 0.082, which shows that high-quality fans will greatly promote the spread of public

opinion, which is also an important factor in public opinion monitoring. However, the weight of the indicators such as blog volume, number of concerns and proportion of original works is low, which shows that the impact on the development of microblog public opinion is limited, which is consistent with the results of some literatures. Therefore, we can consider not as an impact factor as appropriate to simplify the model structure.

4.3. Prediction Result Analysis

In order to compare the influence of fans' activity on prediction accuracy, the index system including and excluding this index is tested. 30 sample data are randomly selected from the processed prediction data, and the popularity of its forwarding volume is simulated and predicted, and compared with the actual popularity of the forwarding volume. After repeated experiments, randomly select three test results, and make interval statistics for errors, the results are shown in Table 3. It can be found that in the popularity prediction experiment which is divided into three levels, the model with fan activity in the index system has better prediction accuracy.

Table 3. Statistical for the prediction of the error

error group model	Prediction model without fan activity				Prediction model including fans activity			
	1	2	3	probability	1	2	3	probability
[0, 0.1]	15	16	15	51.1%	16	20	18	60.0%
[0, 0.2]	21	22	22	72.2%	23	24	25	80.0%
(0.2, 1]	9	8	8	27.8%	7	6	5	20.0%

It can be seen from the above statistical results that, compared with the original index system, the index system with fan activity improves the accuracy of the prediction model to a certain extent and has a better matching effect.

5. Conclusion

This paper constructs a prediction model of online public opinion of food safety events, verifies the prediction index system of 16 indexes including identity authentication, user activity and fan activity, and confirms the high influence of identity authentication, number of fans, hot search and super talk and other factors on micro blog forwarding in the existing system. At the same time, it verifies the importance of the proposed fan activity in such micro blog forwarding, enriches and rationalizes the existing prediction index system to a certain extent, and proves that the appropriate index can improve the accuracy of BP neural network public opinion prediction.

6. Reference

- [1] Suh B, Hong L, Pirolli P, et al. Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network[C]//2010 IEEE 2nd International Conference on Social Computing, Minneapolis, USA:IEEE, 2010
- [2] Rudat A, Buder J, Hesse F W. Audience design in Twittter: Retweting behavior between informational value and fowllower' interests[J].Computers in Human Behavior, 2014, 35:132-139.
- [3] Zhang Yang, Lu Rong, Yang Qing. A study on the prediction of forwarding behavior in Weibo customers [J]. Journal of Chinese Information, 2012, 26 (4): 109 -114.
- [4] GUAN Wanqiu, GAO Haoyu, YANG Mingmin. Analyzing user behavior of the micro-blogging website Sina Weibo during hot social events[J].Physica A, 2014, 395: 340-3.
- [5] Deng Qing, Ma Yefeng, Liu Yi. Prediction of Weibo forwarding volume based on BP neural network [J]. Journal of Tsinghua University (Natural Science Edition), 2015, 55 (12): 1342-1347.

- [6] Zhang Yuliang. Research on the Construction of risk Evaluation Index system of Public opinion Information flow in Emergency Network based on UML method [J]. Books and Information, 2016, (3): 1: 10, 18.