Junfei CHEN[1,2], Qiongji JIN[2], Huimin WANG[1,2] ,Shufang ZHAO[2]

State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University (1), Business School, Hohai University (2)

A Risk Assessment Model of Flood Based on Information Diffusion Method and BP Neural Network

***Abstract**. Climate change has caused more frequent floods in China which have already resulted in huge losses. Thus flood risk assessment and management is an important research topic. In this paper, a new model of flood risk assessment is proposed based on the information diffusion theory and the back propagation (BP) neural network. Due to the fact that flood statistics data are relatively short and often insufficient for flood risk assessment, the information diffusion method can transform imperfect flood historical data from a point in a traditional data sample to a fuzzy data set and obtain optimized data sample. Then, the optimized data are used to train neural networks with back propagation and can improve neural network adaptive ability. The flood data of Dongting Lake's different encirclement dikes are used to assess the flood risk of Dongting Lake with the proposed model in this research. The results are consistent with the actual situation of Dongting Lake area, which thus verifies the model's effectiveness for flood risk management. This method can be easily applied to effectively resolve problems of insufficient samples in flood risk assessment.*

***Streszczenie.** W artykule zaprezentowano nowy model oceny ryzyka powodzi bazujący na teorii dyfuzji informacji I wykorzystujący sieci neuronowe. Dane statystyczne o powodziach są relatywnie krótkie i często niewystarczające do oceny ryzyka. W pierwszym etapie przetwarza się dane historyczne do otrzymania bardziej kompletnych danych. Te dane wykorzystane są do trenowania sieci neuronowych. (**Model oceny ryzyka powodzi bazujący na metodzie dyfuzji informacji i wykorzystujący sieci neuronowe**)*

Keywords: Information diffusion; Back propagation (BP) neural network; Flood disaster; Risk Assessment
Słowa kluczowe: dyfuzja inofmacji, sieci neuronowe,przewidywanie ryzyka powodzi

Introduction

Large property losses and casualties have occurred frequently in China because of floods, such as the flood of Yangtze River in 1998 and the flood of Huaihe River in 2009. According to flood disaster statistical data from 1950 to 2010 in China, the disaster-affected area is 98.23 million hectares on average every year, and the disaster-damaged area is about 54.46 million hectares, and the rate of disaster-damaged is 55.44%. Floods bring tremendous losses and seriously hamper the development of Chinese society and economy. Therefore, it is very important to evaluate the flood risk level for carrying out flood risk management effectively.

There are many researchers conducting flood risk assessment both at home and abroad. Huang chose four first-grade factors and fourteen second-grade factors based on Delphi method and used analytical hierarchy process (AHP) method to evaluate the flood risk of 67 counties in Fujian province[1]. Wang established flood risk evaluation model based on a fuzzy comprehensive assessment method, and used it to assess the flood risk of Huzhou City, China[2]. Jiang et al., discussed how to assess the regional flood risk by AHP, fuzzy comprehensive evaluation method and spatial analysis in theory[3]. Luo evaluated the Chongqing flood risk based on fuzzy evaluation method and obtained the region risk maps according to the flood risk level[4]. Wei et al. , set up a neural network-based predictive method for flood disaster problems and used an example of flood disaster-damaged in China from 1949 to 1994 for demonstration[5]. Huang et al., established a comprehensive evaluation model of flood disaster loss based on support vector machine (SVM) which is fit for evaluating the grades of flood disaster[6]. According to the fuzzy uncertainty of drought and flood disasters, Zhou established the risk assessment model based on the variable fuzzy set theory and took the drought and flood disasters assessment of Shandong province as an example for risk assessment[7]. Chen et al., used a grey clustering analysis method to analyze the disaster losses of every region in Xinjiang "96.7" floods, and the results show that the grey clustering analysis is an effective method to grade the flood disaster[8]. In most situations, the historical disaster data are scarce for flood risk evaluation. Feng established a risk analysis method based on the information diffusion theory and applied it to the risk assessment of the Jinhuajiang and Qujiang drainage basins in Zhejiang province, China. The study indicates that the information diffusion model exhibits fairly stable analytical results and can improve the evaluation accuracy[9].

Due to the fact that flood statistics data are often insufficient for flood risk assessment. Therefore, how to utilize the incomplete information effectively to assess flood risk and obtain effective risk evaluation results becomes quite important for flood risk management. In this paper, the information diffusion theory and the BP neural network are introduced, and a new risk assessment model of floods based on the information diffusion theory and the BP (Back Propagation) neural network is established. Then, we use this model to assess the flood risk of different encirclement dikes in Dongting Lake of Hunan province in China. The research can provide guidance for flood risk management by reasonable flood risk assessment.

Flood risk assessment model

Information Diffusion Principle

Information diffusion is the process by which fuzzy mathematics handles samples using the set-valued method. The method can transform an observed sample into a fuzzy set, that is, turn a single point sample into a set-valued sample[10,11]. Its basic idea is that when there are not many sample points, the knowledge provided by the samples is incomplete, fuzzy and uncertain. At this time, we should not regard a sample point as exact information or exact observation value, but as a set of values. The Information diffusion method can turn a sample observed value into a fuzzy set to achieve a better effect for risk analysis[12]. Huang proved different diffusion functions, the simple normal distribution is better than exponential distribution and logarithmic normal distribution under the condition of a small sample size[13]. Therefore, the normal diffusion method is adopted in this paper. This method transforms a single value into set-valued samples according to the following formula:

$$(1)\qquad q\left(v,V_j\right)=\frac{1}{h\sqrt{2\pi}}\exp\left[-\left(v-V_j\right)^2\Big/\left(2h^2\right)\right]$$

where: $v(v \in V)$ is the information absorption point. h is window width to control the scope of diffusion. It is related to the capacity n of the flood samples X .

h can be computed as follows:

$$(2) \qquad h = \begin{cases} 1.6987(b-a)/(n-1) & 1 \le n \le 5 \\ 1.4456(b-a)/(n-1) & 6 \le n \le 7 \\ 1.4230(b-a)/(n-1) & 8 \le n \le 9 \\ 1.4208(b-a)/(n-1) & n \ge 10 \end{cases}$$

Here a and b are the minimum value and maximum value of the sample data respectively and n is the sample number.

To ensure that all the samples have same status, it is necessary to conduct normalization to the information distribution results which are obtained from the same sample. Normalization is done by following formula.

$$(3) \qquad q'(v,V_j) = q(v,V_j) \Big/ \sum_{1 \le j \le m} q(v,V_j)$$

Here, m is the point number of discrete V .

BP Neural Network

In general, artificial neural networks(ANNs) possess attributes of learning, generalizing, parallel processing and error endurance[14]. ANNs try to discover the inherent relationship between parameters by learning process. Theoretically, neural networks can simulate any kind of data patterns given sufficient training. There are many different kinds of ANNs, among which the BP neural network is the most widely used. In our study, the BP neural network with multi-input and multi-output is employed to build the risk assessment model[15]. BP neural networks consist of three or more layers that include the input, hidden and output layers. Any input information is transmitted forward to nodes of the hidden layer, and calculated by the activation function of each element. Output information is transmitted from the hidden nodes to the output nodes. A BP neural network of three-layer topological structure is shown in Figure 1.

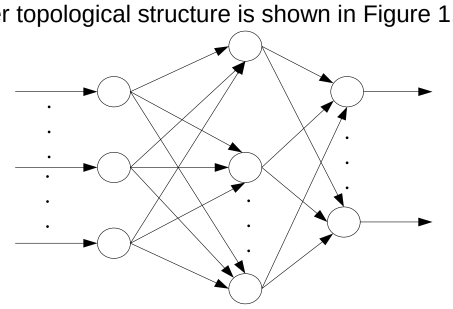

Fig.1. The three-layer BP neural network

BP neural network is trained by repeatedly presenting a series of input/output pattern sets to the network and the network gradually "learns" the input/output relationship by adjusting the weights to minimize the error between the actual and predicted output patterns of the training set. The status of the nodes in every layer only affects the nodes in the next layer. The difference between actual and expected output is calculated recursively and step by step and the error signal is returned through the original link access. The processes of information forward propagation and error back propagation are the processes for adjusting weights in each layer constantly, and they are also the neural network learning and training process. These processes will stop when the output error of training process has reached an acceptable level or a predetermined number of learning times[16].

Steps of Flood Risk Assessment Model

The basic idea of model based on the information diffusion principle and the BP neural network is that the new samples data are generated according to the original flood samples data by the information diffusion method. Then, the new samples data are used as the training samples of the BP neural networks. Based on these new samples, the BP network is designed and trained.

The main steps of model based on the information diffusion principle and the BP neural network as follows.

Step 1. Samples analysis and the selection of evaluation factors for the problem of flood risk.

Step 2. Carry out information diffusion for the original sample data and normalization by formula (1) and (3) respectively.

Step 3. Initialize the connection weights of the network and select a sample from the training set data as the input of the network.

Step 4. Calculate the output vector of the network.

Step 5. Calculate the errors of the network and adjust the connection weights from the output layer back to the input layer.

Step 6. Repeat steps 3-5 until the errors are acceptable.

Steop 7. Select another sample of the training set data and repeat the upper steps until convergence.

Risk Assessment of Dongting Lake area

The application of risk assessment methods is illustrated by the Dongting Lake area in the northern Hunan Province, China. Dongting Lake is the fourth great lakes of China and it is the most important storage lakes for the Yangtze River. The flood occurs frequently in Dongting Lake area, and there is more and more serious trend. In the past 50 years, there are more than 40 years occurring floods. The 33 thousand people were killed in the 1954's flood and the disaster-affected area is 256.7 thousand km^2. The catastrophic floods happened in 1998 and caused direct economic loss of 32.9 billions. In this paper, because of the importance of encirclement dikes of Dongting Lake area for preventing floods, the encirclement dikes of the Dongting Lake area are selected as the evaluated samples[17]. Five risk evaluation indexes are chosen, that is, density of population (M), density of industrial and agricultural products (N), density of one line embankment (W), average difference between control points embankment top elevation and every year's highest water level (S) and proportion of danger embankment length to the total embankment length (T).

Mao used the projection pursuit method to assess the flood vulnerability level[17]. He summed up from the literature on disaster science that vulnerability mainly emphasized the resistance ability, coping ability and recovery ability of human social-economic system suffering disasters. Vulnerability can be regarded as another aspect of safety. A higher vulnerability leads to a low level of safety. When the vulnerability is much higher, the resistance ability and the recovery ability are much lower. In other words, under the same environment, if the vulnerability of an area is much stronger, then the flood disaster risk level of the area is much higher. Therefore, the data and flood risk grades of Dongting Lake area are shown in Table 1 by the projection pursuit method. In Table 1, there are three different risk levels called level I, level II and level III, representing high risk level area, middle risk level area and low risk level area respectively.

Information Diffusion of the Original Data Samples

The calculating process is as follows.

Step1. There are 30 samples, that is n=30. Then, the diffusion coefficient h is determined according to formula (2) .

$$h_M = 1.4208(1050-346)/(30-1) = 34.3911,$$
$$h_N = 1.4208(930.65-72.09)/(30-1) = 42.0635,$$
$$h_W = 1.4208(0.86-0.1)/(30-1) = 0.0372,$$
$$h_S = 1.4208(2.48-0.8)/(30-1) = 0.0823,$$
$$h_T = 1.4208(86.77-4.71)/(30-1) = 4.0204.$$

Table 1. Data and flood risk grade

No.	M	N	W	S	T	Flood Risk Grade
1	921	201.45	0.14	1.21	32.71	II
2	857	100.68	0.39	2.15	39.18	II
3	512	85.88	0.29	2.48	33.72	III
4	636	266.69	0.11	2.03	42.70	II
5	526	106.37	0.10	2.20	44.03	III
6	1050	557.19	0.18	0.80	21.54	I
7	529	122.29	0.14	2.04	41.36	III
8	742	236.47	0.35	1.28	10.42	II
9	692	204.08	0.15	1.82	41.39	II
10	627	107.95	0.41	2.09	49.70	II
11	716	136.99	0.15	2.13	20.60	III
12	507	84.57	0.59	1.02	8.98	II
13	489	84.77	0.84	1.57	27.34	II
14	544	95.33	0.57	1.88	63.9	I
15	568	72.09	0.73	1.56	17.26	II
16	535	169.45	0.42	2.21	38.23	II
17	562	140.41	0.39	1.63	25.42	II
18	560	122.83	0.19	1.46	38.71	II
19	633	171.7	0.49	1.09	59.41	I
20	532	307.66	0.21	1.86	44.83	II
21	549	425.43	0.43	1.46	33.50	II
22	504	267.57	0.10	1.24	27.16	II
23	721	86.90	0.18	1.65	39.91	II
24	444	199.34	0.47	1.89	15.40	III
25	555	307.92	0.38	1.30	83.62	I
26	346	930.65	0.23	1.25	86.77	I
27	660	93.03	0.54	1.56	4.71	II
28	484	83.81	0.86	1.09	10.43	II
29	568	76.45	0.48	0.97	4.88	II
30	678	154.32	0.69	0.94	40.34	I

Step2. The number of selected discrete points is 6, and the step length is equal. Then the discrete points of these indexes are as follows:

$$M = \{m_1, m_2, m_3, m_4, m_5, m_6\} = \{300, 460, 620, 780, 1040, 1100\},$$
$$N = \{n_1, n_2, n_3, n_4, n_5, n_6\} = \{50, 230, 410, 590, 770, 950\},$$
$$W = \{w_1, w_2, w_3, w_4, w_5, w_6\} = \{0, 0.2, 0.4, 0.6, 0.8, 1\},$$
$$S = \{s_1, s_2, s_3, s_4, s_5, s_6\} = \{0.5, 1, 1.5, 2, 2.5, 3\},$$
$$T = \{t_1, t_2, t_3, t_4, t_5, t_6\} = \{0, 18, 36, 54, 72, 90\}.$$

Step3. Calculate $\mu(m_i, M_j)$, $\mu(n_i, N_j)$, $\mu(w_i, W_j)$, $\mu(s_i, S_j)$ and $\mu(t_i, T_j)$ Using formula (1), then using formula (3) to do the normalization processing, and the input vector X can be obtained. The flood risk level is divided into three kinds: high risk level, middle risk level and low risk level. As the output, the corresponding output vectors of the three kinds of risk are represented by (1,0,0), (0,1,0) and (0,0,1). The BP neural network sample data which is processed by the information diffusion method are obtained.

Results and Analysis

The structure of the BP neural network is 30-6-3 in this paper. No.1 to No.27 encirclement dikes data are as the training samples. The number of input layer nodes is 30, the hidden layer nodes are 6, and the output layer nodes are 3. The goal of training the networks is obtained stable and precision networks. By the number of 922 times training, the permissible error reached 0.001. The learning results of the BP neural network are shown in Table 2. The error curve of training process is shown in Figure 2. As it is shown in Table 2, the output values of BP neural network fit very well with the original value. The results shows that the model based on based on the information diffusion principle and the BP neural network is effective. Using the neural network which is trained to predict the samples 28, 29 and 30, and the prediction results are (-0.1878,0.6972,0.2643), (0.1591,0.8032,0.3245), and (1.0175,-0.0180,-0.014). So, the risk of samples 28, 29 and 30 are middle risk, middle risk, and high risk respectively. The risk assessment results of Dongting Lake in this paper are consistent with the results of projection pursuit model.

Table 2. The training results of BP neural network

No.	Risk level	Training results
1	(0,1,0)	(-0.0139,0.9673,-0.1238)
2	(0,1,0)	(-0.0155,1.0129,0.0242)
3	(0,0,1)	(-0.0166,0.0098,1.0111)
4	(0,1,0)	(-0.0215,1.0142,-0.0296)
5	(0,0,1)	(0.0550,-0.0448,0.9918)
6	(1,0,0)	(1.0132,0.0145,0.0257)
7	(0,0,1)	(-0.0243,0.0442,0.9895)
8	(0,1,0)	(-0.0015,0.9971,-0.0039)
9	(0,1,0)	(-0.0144,0.9815,0.0253)
10	(0,1,0)	(0.0354,0.9820,-0.0569)
11	(0,0,1)	(-0.0349,0.0049,0.9820)
12	(0,1,0)	(0.0009,0.9939,0.0083)
13	(0,1,0)	(-0.0834,1.0719,0.0459)
14	(1,0,0)	(0.9956,0.0077,0.0048)
15	(0,1,0)	(0.0023,0.9823,0.0046)
16	(0,1,0)	(0.0450,0.9772,0.0123)
17	(0,1,0)	(0.0470,0.9607,0.0242)
18	(0,1,0)	(0.0192,1.0373,-0.0072)
19	(1,0,0)	(0.9796,-0.0067,0.0022)
20	(0,1,0)	(-0.0438,0.9711,0.0813)
21	(0,1,0)	(0.0027,1.0274,0.0077)
22	(0,1,0)	(0.0061,1.0208,0.0062)
23	(0,1,0)	(0.0552,0.9778,0.0124)
24	(0,0,1)	(0.0315,-0.0088,0.9698)
25	(1,0,0)	(0.9610,0.0332,0.0129)
26	(1,0,0)	(1.0192,-0.0311,-0.0118)
27	(0,1,0)	(-0.0066,0.9972,-0.0144)

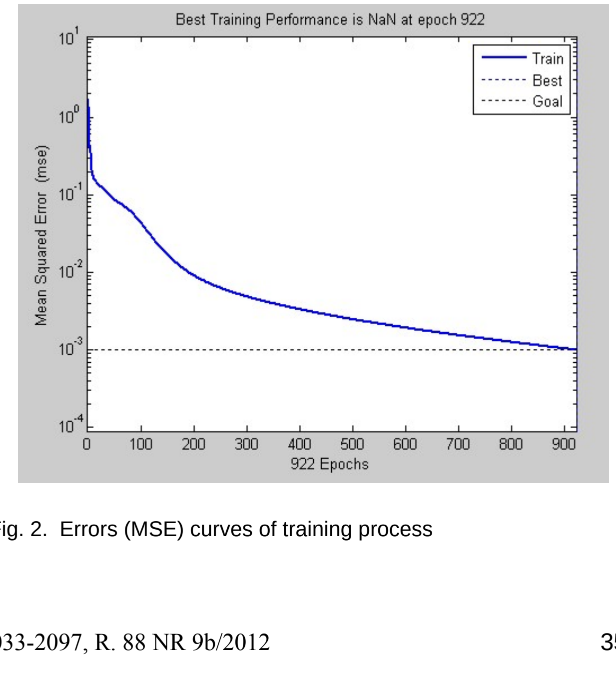

Fig. 2. Errors (MSE) curves of training process

Conclusions

With the global climate change, floods become frequently in China. In the process of flood risk assessment, it is difficult to get abundant flood samples and data. In this study, a flood risk assessment method is proposed based on based on information diffusion method and BP neural network. The model was applied to evaluate the flood risk of Dongting Lake in Hunan province, China. The results show that the model is effective for flood risk assessment. The information diffusion method is capable of extracting useful information and therefore improves the accuracy of risk assessment. The flood risk assessment of Dongting Lake is an important part for flood control decision-making, and it has important significance for flood risk management of Dongting Lake. Due to the complexity of flood risk evaluation coupled with the climate change and the complex geographical environment of area, the evaluation index system is complicated. The methods used in this paper can be improved by using multiple flood risk index. In the meantime, flood risk evaluation is only one part of flood risk management, which still needs further researching.

Acknowledgments

This work was supported by the National Nature Science Foundation of China (No.90924027), the National Society Science Fund of China (No.09CJY020), Public-interest Industry Project of Ministry of Water Resources of China (No.200801027), the Special Fund of State Key Laboratory of China (No.2011585312), the Science and Technology Projects of Yunnan Province (No.2010CA013), 2010 Jiangsu Province Qing Lan Project.

REFERENCES

[1] Huang M.S., Huang C.C., Research on Grade Model of Flood Risk Assessment, *Journal of Catastrophology*, 22(2007), No. 1,1-5 (In Chinese)

[2] Wang J. H., Risk Evaluation on Flood Disaster Based on the Fuzzy Comprehensive Assessment Method, *Water Conservancy Science and Technology and Economy*, 15(2009), No. 4,338-340 (In Chinese)

[3] Jiang W.G., Li J., Chen Y.H., Risk Assessment System for Regional Flood Disaster: Principle and Method, *Journal of Natural Disasters*, 17(2008), No. 6, 53-59 (In Chinese)

[4] Luo P., Zhang T.R., Du J., GIS and Fuzzy Evaluation Method Based on Hazard Evaluation for Flood Disaster: A Case Study of Chongqing,China, *Journal of Xihua Normal-University*, 28(2007),No. 2, 165 -171(2007). (In Chinese)

[5] Wei Y.M., Xu W.X., Fan Y., Artificial Neural Network Based Predictive Method for Flood Disaster, *Computers & Industrial Engineering*, 42(2002), 383-390.

[6] Huang Z.W., Zhou J.Z., Song L.X., Flood Disaster Loss Comprehensive Evaluation Model Based on Optimization Support Vector Machine, *Expert Systems with Applications*,37(2010), 3810-3814.

[7] Zhou H.C., Zhang D., Assessment Model of Drought and Flood Disasters with Variable Fuzzy Set Theory, *Transactions of the CSAE*, 25 (2009), No. 9, 56-61 (In Chinese)

[8] Chen Y. N., Yang S. Q., Application of Grey Clustering Analysis in the Classification of Flood Disaster Grade, *Arid Land Geography* , 22(1999), No. 3,37-41 (In Chinese)

[9] Feng L. H., Luo G.Y., Flood risk analysis based on information diffusion theory, *Human & Ecological Risk Assessment* ,14(2008),No. 6,1330-1337

[10] Chatman E. A., Diffusion Theory: a Review and Test of a Conceptual Models in Information Diffusion, *Journal of the American Society for Information Science*, 37(1986), No. 6, 377-386.

[11] Huang C. F. , Principle of Information Diffusion, *Fuzzy Sets and Systems*, 91(1997),No.1, 69-90

[12] He S., Wang J. D., Wang H., Han X. M., The Evaluation of Loess Slope Stability Based on Combination of Information Diffusion Theory and BP Neural Network, *Journal of Northwest University*, 38(2008), No. 6, 983-988 (In Chinese)

[13] Huang C.F., Wang J.D., Fuzzy Information Optimization Technology and Its Application, University of Aeronautics & Astronautics Press, Beijing (1995). (In Chinese)

[14] Kermanshahi B., Iwamiya H., Up to Year 2020 Load Forecasting Using Neural Nets. *Electrical Power and Energy Systems*, 24(2002), 789–797.

[15] Enke D., Thawornwong, S., The Use of Data Mining and Neural Networks for Forecasting Stock Market Returns, *Expert Systems with Applications*, 29(2005), No. 4, 927-940.

[16] Azadeh A., Ghaderi S. F., Sohrabkhani S., Annual Electricity Consumption Forecasting by Neural Network in High Energy Consuming Industrial Sectors, *Energy Conversion and Management*, 49(2008), 2272-2278.

[17] Mao D.H., Research on Theory and Method and Application of Synthetic Risk Analysis on Flood Disaster, China Water Conservancy and Hydropower Press,Beijing(2009). (In Chinese)

***Authors**: Associate Prof. Dr. Junfei Chen, State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China, E-mail:chenjunfei@yahoo.com.cn; Qiongji JIN, Business School, Hohai University, Nanjing 210098, China, E-mail:jinqiongji0431@yahoo.com.cn; Prof. Dr. Huimin Wang, State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China, E-mail: hmwang@hhu.edu.cn; Shufang Zhao, Business School, Hohai University, Nanjing 210098, China, E-mail:shufangzhao@yahoo.com.cn;*

The correspondence address is:
e-mail: chenjunfei@yahoo.com.cn