

Neural Network Application Based on GIS and Matlab to Evaluation of Flood Risk

SONG Lirong^{1,2}, ZHAO Shiwei², LIAO Weilin², WANG Zhaoli^{2*}

1. College of Hydrology and Water Resources, Hohai University, Nanjing, China;

2. School of Civil and Transportation Engineering, South China University of Technology, Guangzhou, China
e-mail: wangzhl@scut.edu.cn

Abstract—In order to test the Artificial Neural Networks (ANN) in the applicability of flood risk assessment, this paper applies the traditional BP neural networks (BPNN), radial basis function neural networks (RBFNN) and probabilistic neural networks (PNN) to establish flood risk assessment model using MATLAB combined with GIS technology. It is observed that BPNN is superior among three methods. Taking Beijiang River basin as a case study, the risk assessment map based on BPNN model shows that the dangerous areas are mainly located in these areas: Sihui, Qingyuan city, Fogang, northwest Huaiji, central Yangshan, central Yingde, northeast Nanxiong and so on. Compared with a few historical large floods, above results can better reflect the actual situation of flood risk in Beijiang River basin, which validate the rationality of the presented model and provide a reference for flood control and disaster assessment.

Keywords-flood damage; risk assessment; BP neural network; geographic information system

I. INTRODUCTION

Artificial neural network (ANN) is the rise of the research focus in the field of artificial intelligence since the 1980s. From the perspective of information processing of human brain neurons network, it establishes a simplified and abstract model which composes to different networks according to different connections^[1]. ANN is well-known as a parallel information processing structure which consists of simple processing units (neurons or cells) with a high degree of interconnection between each unit. Each neuron performs a relatively simple job: receive impulses from input cells or neurons, carry out some types of transformation of the inputs and pass on the weighted products to the other cells or other neurons. The function of a neural network replicates that of a biological brain, which is basically composed of a very large number of massively interconnected neurons. Neural networks exhibit mapping capability, learn from examples and possess the capability to generalize. They are robust systems that attempt to mimic the fault-tolerance and can process information in parallel with high speed^[2]. For these outstanding capabilities, ANNs are designed for pattern recognition and classification applications, such as character recognition, voice recognition, fingerprint recognition, remote sensing image recognition, face recognition, handwritten character recognition, industrial fault detection, precision-guided^[3], etc.

Flood is a serious natural disaster with the properties of sudden, high frequency, and wide distribution. Huge losses are caused by flood disasters every year in the world and China has become one of the most serious counties affected by flood disaster^[4]. Flood risk assessment is particularly important to effectively achieve the purpose of flood control, precisely analyze the probability of a flood and evaluate the losses. It is also especially significant for the government to take some measures of flood protection and disaster reduction more effective. However, the flood detriment risk assessment is a comprehensive evaluation and analysis of the flood hazard risk factors, including the dangerousness of disaster-causing factors, the stability of pregnant disaster environment and the vulnerability of hazard-affected body^[5]. Due to flood risk is essentially a three-dimensional concept related to detriment, uncertainty and complexity^[6], it is influenced by many factors including disaster-causing factors, pregnant environment and hazard-affected body. Moreover, there is much multiplicity, complexity, uncertainty and imprecision in the geographic information, remote sensing information, statistical information which is used to express those factors^[7]. Therefore, flood risk assessment has become the worldwide focus of disaster research. Many researchers have conducted studies in this field and achieved acceptable results. The research methods include the analytic hierarchy process (AHP)^[8], fuzzy comprehensive assessment (FAC)^[9], statistical analysis (SA)^[10], etc. This paper attempts to utilize the powerful classification ability of ANN pattern recognition to resolve classification of the flood risk level.

II. METHODS AND PROCEDURES

Artificial neural networks have proven themselves as proficient classifiers and are particularly well suited for addressing non-linear problems. Given the classification of flood risk level is essentially a non-linear and pattern classification problem, artificial neural networks is certainly a good candidate for solving the problem. There are many kinds of neural networks for pattern classification, such as BP neural networks (BPNN), radial basis function networks (RBFNN), generalized regression networks (GRNN), probabilistic neural networks (PNN), self-organizing feature map networks (SOFM), learning vector quantization networks (LVQ), etc. From the perspective of wide application, this paper prepares to respectively select BPNN, RBFNN and PNN to establish a flood risk assessment model.

* Corresponding author. Tel.: +86-20-87111030; fax: +86-20-87114460.

A. Introduction of BPNN

BP neural network [11-13] is a multilayer feedforward network based on reverse error propagation algorithm. In order to approximate the target, it continuously adjusts the network weights and thresholds by making the sum of squared errors minimum through back-propagation network, belonging to the global approximation network. The basic idea of the BP algorithm is this: their learning process includes the two process signal forward propagation and forward error back propagation. Forward propagation, the input signal from the input layer, through the hidden layer units, is transmitted to the output layer to produce an output signal at the output. If the desired output in the output layer will not be gotten, the learning process transfers to the error signal back-propagation process; error signal back-propagation process, the error signal from the output, through the hidden layer, propagates to the input layer by layer, while network weights is adjusted by the error feedback, and constantly revised by the weights and thresholds so that the actual output of the network is closer to the desired output. According to the Kolmogorov theorem, a three-layer BP neural network with one hidden layer (if the nodes of the hidden layer are sufficient) can approximate arbitrary nonlinear continuous function with arbitrary precision on a closed set.

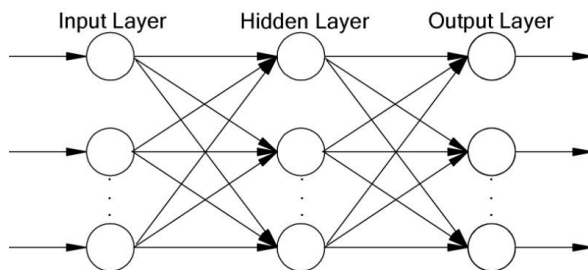


Figure 1. Structure of BP neural network

B. Introduction of RBFNN

Radial basis function neural network (RBFNN) [14] proposed by Moody and Darken in the late 1980s is a type of multi-layer feedforward neural network that learns using a supervised training technique and is known as a multi-nonlinear dynamic system. It is self-adaptive, self-systematic and has the capability of learning, association and compatibility, which make it an appropriate approach for constructing models of complicated and unknown systems. A typical structure of RBFNN, which consists of three layers, namely, input, radial basis, and output layers. The input layer consists of the source nodes. The radial basis layer generally uses Gaussian functions as the kernel function. The output layer uses a linear transfer function. The hidden layer of RBFNN transforms the low dimensional input vector to a high dimensional space, so that the problem not linear separable in the low-dimensional space becomes linear separable in the high dimensional space. It is a local approximation network with the powerful learning speed, convergence and real time, and for each training sample, it is only a small number of weights and thresholds to be amended.

C. Introduction of PNN

The probabilistic neural network (PNN) [15] firstly proposed by Dr. DF Specht in 1989 is commonly used in pattern classification, which combines statistical methods and feedforward neural network, and uses an exponential function instead of the S-type function. PNN can be done any nonlinear transformation, whose curved surface of judgments is closer to the curved surface under Bayesian optimal criteria, belonging to a local approximation network. A typical structure of PNN, which consists of four layers, namely, input, pattern (first hidden layer), summation (second hidden layer), and output layers. The input layer, which is fully connected with the pattern layer, does not perform any computation and only distributes input signals to the neurons in the pattern layer. Each unit in the pattern layer implements a RBF by computing the Gaussian kernel of the Euclidean distance between the existing input vector and the training pattern. The summation layer that accumulates the outputs of the pattern layer consists of one neuron for each class. The output layer classifies a given input set to one of the specified classes based on the output in the summation layer. PNN has a simple learning process (not requiring an iterative training procedure), fault tolerance, pattern classification ability and expansion of the performance advantages.

D. Steps of ANN Application to Flood Risk Assessment

Matlab is a large-scale interactive software researched and developed by MathWorks, Inc., specifically for scientific and engineering computing. Moreover, the neural network toolbox involves almost all categories of neural network model in Matlab7.8. This paper uses the Neural Network Toolbox to achieve the establishment of the flood risk assessment model.

1) Data preparation

- ① Select R indexes that can reflect the impact factors (disaster-causing factors, pregnant environment and hazard-affected body).
- ② Collect the data of the indexes previously and convert them into raster layer based on GIS.
- ③ Classify the index value with the method of natural breaks (Jenks) based on GIS and the interval value of every index can be obtained.
- ④ Set the number of risk grade and map each column value including R indexes to only one risk grade.
- ⑤ Normalize the data of each index and divide it into two groups, one used for training and the other for testing.

2) The network creating and training

- ① Determine the structure of the network, the number of layers, the number of each layer's neurons, the transfer function and the training algorithm according to the input sample data and application requirements.
- ② Initialize the weights and thresholds of the network layers (usually randomly selected) according to different requirements.
- ③ Provide the sample data (input matrix P with the desired output matrix T) to the network and train it.

3) The network testing

Provide the test data to the trained network and the prediction results will be obtained. Compared with the desired output we get the test accuracy of the previous network. Identify

the network parameters through several tests in order to obtain more satisfactory test accuracy.

4) The network application

Provide the index raster data of the whole basin to the network selected at step (3) and the risk grade of each grid will be obtained. Acquire the flood risk assessment map based on GIS.

III. CASE STUDY

A. Beijiing River Basin Overview

Beijiing River is the second largest river of Pearl River. Its upstream originates Shijie in Xinfeng county, Jiangxi province. It meets Xijiang River in Sixianku in Sanshui and flows into the river network of Pearl River Delta. The mainstream flows into the sea through Humen. The area of Beijiing River basin is about 46649km².The first grade tributary includes Wengjiang River, Pajiang River, Jinjiang River, Wujiang River, Nanshui River, Lianjiang River, Binjiang River and Suijiang River and so on.

B. Data Preparation s

Select 10 indexes as follow: 3 days maximum of rainfall, typhoon frequency, elevation, slope, river buffer, distance to road, depth of runoff, percentage of agricultural land, GDP density, population density. Rainfall is a direct factor in the formation of a flood. This paper chooses the 3 days maximum of rainfall (mm) as a rainfall index because of its great influence to the flood. The typhoon frequency index is acquired according to the average typhoon frequency (times/year) based on units of the county in Guangdong province. The river buffers are divided into different levels due to different discharge of each river. National roads and highways are chosen to consider the distance to road (m). The shorter the distance to road is, the faster the lives, property and relief materials would be transferred, then the lower risk attacked by floods would be, otherwise the higher. The depth of runoff (mm) can reflect some condition of underlying, such as vegetation, soil, urbanization and other characteristics. The units of population density and GDP density are person/km² and 10 thousand yuan/km².The percentage of agricultural land (%) means the proportion of agricultural land. Of the 10 indexes, positive indexes include 3 days maximum of rainfall, typhoon frequency, river buffer, distance to road, depth of runoff, percentage of agricultural land, GDP density, population density. Negative indexes include elevation and slope.

C. Build ANN flood risk assessment models

The flood risk assessment of BP neural network model is a three-layer network including an input layer, a hidden layer and an output layer. The input dimension is 10, the output 5. The transfer functions of the hidden layer and the output layer are both the function *logsig*. The LM algorithm is used to optimize network computing. Based on Matlab, create the network with the function *patternnet*, and set the ratio of training, validation, testing data at 15:4:1. Make the risk indexes as the training input, the risk grades as the training output. After several tests, the number of hidden neurons is set to 23. At the same time, the accuracy of 1000 test samples reaches to 98.8%.

Based on Matlab, create the RBF network including 10 input and 5 output with the function *newrb*. Hidden layer neurons will be added automatically after each epoch. The kernel function is Gaussian function, and the output transfer function is the pure linear function *purelin*. The mean squared error goal *goal* and the spread of radial basis functions *spread* can be calibrated through several tests. Make the risk indexes as the training input, the risk grades as the training output. Transfer the output into the standard output type (e.g. (0.988, 0.002, 0.006, 0.003, 0.001) to (1, 0, 0, 0, 0)) with the function *compet*. Calculate the accuracy between the actual output and the desired output with the function *confusion*. After several tests, *goal* is set to 0.005, *spread* 75, and the optimum test accuracy reaches to 95.7%.

The weights of probabilistic neural network do not need training, and the input layer and pattern layer weights are set for all types of training samples. All weights will be set automatically when the network is created. Make the risk indexes as the training input, the risk grades as the training output. Based on Matlab, create the PNN with the function *newpnn*. Try to create several PNNs with different spread of radial basis functions *spread* in order to obtain the optimal network accuracy. After numerous tests, *spread* is set to 2.8, and the network accuracy of the 1000 test samples reaches 93.8%.

TABLE I. THE OPTIMAL TRAINING RESULTS FOR NETWORKS

Network Types	Training samples	Testing samples	Training accuracy	Testing accuracy	epochs
BP	2088	1000	99.50%	98.80%	82
RBF	2088	1000	99.20%	95.70%	39
PNN	2088	1000	100%	93.80%	—

According to Table 1, we obtain: BP, RBF and PNN all can reach satisfactory accuracy, and can be applied to flood risk assessment; Among them, PNN does not need training, networking quickly, and the BPNN training is slightly slower than the RBFNN; Only from the perspective of accuracy is BP network more suitable for the Beijiing basin flood risk assessment.

Provide the 10-index raster data of the whole Beijiing basin to the previous BP network and the risk grade of each grid will be obtained. Based on GIS, export the Beijiing basin flood risk assessment diagram shown in Fig. 2.

D. Result analysis

Based on the function of spatial analysis in GIS, count the areas and percentage of every risk zone as Tab.2. According to Fig.4 and Tab.3, the flood risk in south areas is higher than the north areas, the downstream areas are higher than the upstream areas, and the urban areas are higher than the rural or mountain areas. The detailed distribution is as follows:(1)The percentage of highest risk zone is 12.35%, mainly in Sihui, Qingyuan city, Qingxin, northwest Huanji, northeast Nanxiong and so on;(2)The percentage of higher risk zone is 26.64%, mainly in Yingde, Yangshan and so on;(3)The percentage of the medium risk zone is 15.45%, mainly in Guangning, Lianshan, Wengyuan, Shixing and so on;(4)The percentage of the lower risk zone is 25.24%,

mainly in Lianzhou, Ruyuan, Renhua, Qujiang and so on;(5)The percentage of lowest risk zone is 20.32%, mainly in the rural or mountain areas in the north of Beijiang River basin, such as Lechang, Linwu in Hunan province and so on.

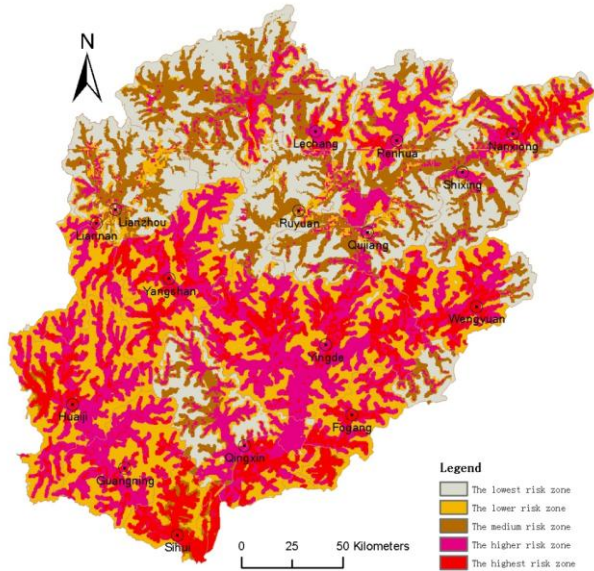


Figure 2. Flood risk assessment map in Beijiang River basin

TABLE II. FLOOD RISK ASSESSMENT MAP IN BEIJIANG RIVER BASIN

Grade	Lowest risk	Lower risk	Medium risk	Higher risk	Highest risk
Areas/km ²	9478.05	11773.92	7207.11	12427.97	5762.23
Percentage/%	20.32	25.24	15.45	26.64	12.35

IV. SUMMARY AND CONCLUSIONS

In this study, an efficient approach for flood risk assessment based on artificial neural networks (ANN) and GIS was proposed. Three types of ANNs, namely, back-propagation neural network (BPNN), radial basis function neural network (RBFNN), probabilistic neural network (PNN), which all can reach satisfactory accuracy, and can be applied to flood risk assessment. Among them, PNN does not need a separate training phase, networking quickly, and the BPNN training is slightly slower than the RBFNN. Beijiang River basin is taken as a study case to evaluate the flood risk and the flood risk assessment map is produced.

The test results demonstrated that the proposed method could achieve 98%, 95% and 93% accuracy, respectively, by BPNN, RBFNN and PNN models. Upon the excellent results, ANN has been shown as an effective tool for detecting the complex nonlinear nature of flood risk level classification problem. Of course, due to the complexity of flood formation and influence of many indexes, it's pretty difficult to analyze the flood risk quantitatively. The choice of other indexes, for example, house, reservoir, dike etc, the determination of indexes weights, the index classification and other problems remain to be further studied.

ACKNOWLEDGMENT

We gratefully thank the funding for this study from the Open Research Fund Program of State Key Laboratory of Water Resources and Hydropower Engineering Science (No. 2010B065) and Natural Science Foundation of China (No. 51209095).

REFERENCES

- [1] Han L Q. Artificial Neural Network. Beijing: Beijing University of Posts and Telecommunications Press, 2006.
- [2] Y S Saman, Tsang H H. A new site classification approach based on neural networks. *Soil Dynamics and Earthquake Engineering*, 2011, 31(7): 974-981
- [3] Oz C, Leu MC. American sign language word recognition with a sensory glove using artificial neural networks. *Engineering Applications of Artificial Intelligence*, 2011, 24(7): 1204-1213.
- [4] Du J, He F, Shi P J. Integrated flood risk assessment of Xiangjiang River Basin in China. *Journal of Natural Disasters*, 2006, 15(6):38-44.
- [5] Jiang W G, Li J, Chen Y H, et al. Risk assessment system for regional flood disaster (I):principle and method. *Journal of Natural Disasters*, 2008,17(6):53-59.
- [6] Huang C F. Risk assessment of natural disaster theory & practice [M]. Beijing: Science Press, 2004.
- [7] Shi W Z, Wang X Z, Wang S L. Fuzzy Spatial Information processing [M]. Wuhan: Wuhan University Press, 2003.
- [8] Liu J F, Li J, Liu J, et al. Integrated GIS/AHP-based flood risk assessment: a case study of Huaihe River Basin in China. *Journal of Natural Disasters*,2008, 17(6):110-114.
- [9] Jiang W G, Deng L, Chen L Y. Risk assessment and validation of flood disaster based on fuzzy mathematics .*Progress in Natural Science*, 2009, 19(10):1419-1425.
- [10] Tawatchai T, Mohammed F K. Flood hazard and risk analysis in the southwest region of Bangladesh . *Hydrological Process*, 2005, 19: 2055-2069.
- [11] Wang T, Yang K L. Ice forecast by artificial neural networks in the Middle Route of the South-to-North Water Diversion Project. *Journal of Hydraulic Engineering*, 2009.40(11):1403-1406.
- [12] Zhao L N, Song S B. Application of BP neural network in the prediction of Dali River runoff. *Yellow River*, 2010.32(8):30-31,33
- [13] Zhou H R, Zhen P E, Niu B, et al. HGA-BP-Based Pattern Classification Method .*Journal of System Simulation*, 2009,21(8): 2243-2247.
- [14] Li L, Sun H, Di Q M, et al. Application of RBF Neural Network in the Surface Water Assessment in Pingdingshan City. *Journal of Anhui Agricultural Sciences*, 2008,36(26): 11514- 11516.
- [15] Dong Y H, Zhou W B, Lai K R, et al. The Application of Probabilistic Neural Network Model in Evaluating the Groundwater Quality of Xi an Area. *Journal of Natural Resources*, 2009,24(4): 737-742.