

Distance Prediction of Slope-Foot Landslide in Southwest of China Based on GA-BP Neural Network

Hao Zhou¹, Yong Chen^{1,2} and Ruoying Tian¹

¹School of Resources and Environmental Engineering, Wuhan University of Science and Technology, Wuhan; 430081, China.

²Hubei Key Laboratory for Efficient Utilization and Agglomeration of Metallurgic Mineral Resources, Wuhan; 430081, China.

*Email: Yongchen1968@163.com

Abstract. Landslide is a typical geological hazard in southwest China. In order to reasonably assess the disaster-causing range of earthquake-induced landslides in southwest China, the prediction of landslide movement distance is very important. By making principal component analysis of the factors affecting the distance of landslide movement and combining with the mechanism of landslide movement, the three principal components are named as kinetic energy factor, slope factor and resistance factor, which are used as input neurons of BP neural network. The initial weights and thresholds of BP neural network are optimized by using the global search ability of genetic algorithm, and the prediction models of maximum vertical motion distance (H) and maximum horizontal motion distance (L) of landslide are constructed based on the optimized network. At the same time, on the basis of principal component analysis, a multivariate regression prediction model is constructed, and the results of network model prediction are compared with those of multivariate regression prediction. The optimization effect of genetic algorithm for BP neural network is obvious. The BP neural network model after optimization is accurate and stable in predicting the motion distance of slope foot landslide. The prediction errors of maximum horizontal and vertical motion distances below 10% accounted for 86.67% and 93.33% respectively, which is superior to the original BP neural network and multiple regression prediction model.

1. Introduction

Landslide is one of the major geological disasters in China, which brings huge economic losses to the people every year. For landslides whose horizontal movement is not obviously hindered, the movement distance of landslides is not only affected by the volume and height difference, but also related to the change of terrain gradient. Although there may be more gradient changes in the process of landslide movement, the gradient difference between the moving and accumulating zones of landslides at the foot of the slope is at the foot of the slope, and it is also related to landslide movement. The action is very significant [1]. The average slope (α) of sliding surface, the slope difference between sliding area and accumulation area, that is the angle of slope foot (β) and the average slope (γ) of accumulation area, all have great influence on the movement distance of landslide. Slope foot landslides caused by earthquakes are common in mountainous areas of Southwest China due to topography and geological structure. Landslide movement distance is an important index for Evaluating Landslide disasters, and its reasonable prediction can provide reference and basis for local disaster prevention and mitigation work.

At present, many models and theories have been put forward for the study of landslide movement at home and abroad, such as multi-factor coupling mechanism, air lubrication mechanism and energy



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

transfer mechanism [2]. Through qualitative and quantitative research, Huajun Meng et al. concluded that landslide volume, site conditions and geological structure are the important factors affecting landslide movement [3]. According to the movement distance of landslide, Ran Tang and Qiang Xu design simulation experiments from the dynamic point of view, deduce the theoretical calculation formula of the movement distance of landslide [4]. Based on the statistical idea, through numerical simulation and regression analysis, Jian Guo, Shanshan Hao et al. obtained the prediction methods of different types of landslide movement distances, and established the relationship between the influencing factors and the movement distances of landslides [5-6]. However, due to the complexity of the internal mechanism of landslide, the prediction accuracy of the method and model for the distance of landslide movement needs to be improved.

In this paper, principal component analysis is used to extract the principal component of factors affecting landslide movement, and the mechanism of landslide movement is analysed. Then, a network model is constructed based on BP neural network optimized by genetic algorithm. The network model is trained and tested with 75 slope-foot landslides caused by earthquakes in southwestern China as samples, and the slope-foot type caused by earthquakes in southwestern China is analysed. The prediction of landslide motion distance is expected to improve the prediction accuracy of Landslide Motion distance. At the same time, new ideas and methods are introduced for the study of this kind of problems.

2. Research Method

2.1. GA-BP Neural Network Model

Artificial neural network (ANN) is a mathematical model inspired by the principle of biological neural network perceiving the world. It has strong ability to deal with non-linear problems. It is very suitable for problems with complex internal mechanism such as landslide movement and difficult to be solved by traditional statistical methods. BP neural network is a multi-layer feedforward neural network based on error back propagation algorithm. It has the characteristics of learning and self-adapting. The whole training process includes two processes: data forward transmission and error feedback correction. In this paper, the influencing factors of landslide motion distance are studied, and the prediction model of GA-BP neural network is constructed to predict the motion distance of slope foot landslide caused by earthquake.

2.1.1. BP Basic Structure. In BP neural network, the data is transferred to the hidden layer by the input layer according to the weight, and then to the output layer through the non-linear processing. The output layer adjusts the connection weights and thresholds in the process of data transmission through error feedback. A typical BP neural network consists of an input layer, a hidden layer and an output layer. Its topological structure is shown in figure 1.

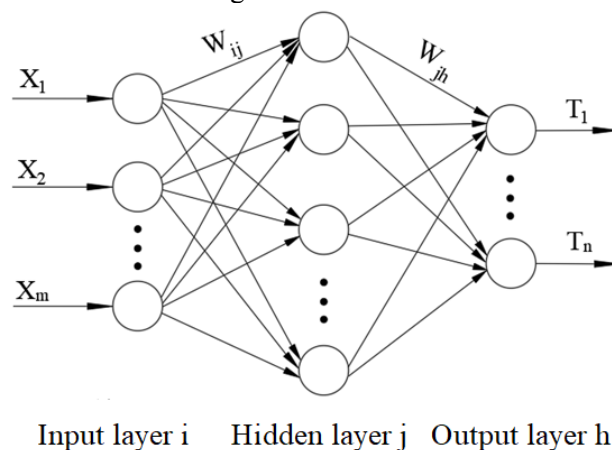


Figure 1. 3-Layer BP Neural Network

2.1.2. Setting of Network Parameters. The number of hidden layer neurons and the number of single layer neurons have a significant impact on the performance of the whole network. The method of determining the number of hidden layer neurons is not unique. The following formula (1) is selected to determine the number of hidden layer neurons preliminarily and adjust it appropriately in the actual training process. The selection of transfer function, learning algorithm, efficiency, initial threshold and weight may also have a great impact on the performance of the whole network. In the training process, the appropriate learning algorithm and efficiency are determined according to the results. The initial threshold and weight are optimized by genetic algorithm.

$$N = \sqrt{m + n} + t \quad (1)$$

Where: N is the number of neurons in the hidden layer, m is the number of neurons in the input layer, n is the number of neurons in the output layer, and it is an integer between 0 and 10.

2.1.3. Optimizing BP Neural Network by Genetic Algorithms. BP neural network has strong ability of data fitting and prediction, but the determination of initial weight and threshold has great influence on the performance of the neural network. With the help of the optimization ability of genetic algorithm, the initial weight and threshold of BP neural network are optimized, so that the neural network model can approach the most excellent solution quickly and accurately in the training process, avoiding local convergence or divergence. The optimization process of genetic algorithm includes population initialization, fitness function selection, selection operation, crossover operation and mutation operation.

2.2. Principal Component Analysis

The intrinsic mechanism of landslide is complex. In order to analyse the mechanism of landslide movement, and to reflect the intrinsic relationship among various factors more fully in the construction of network model, and to improve the accuracy of prediction model, the principal component analysis of factors affecting landslide movement is made by using MATLAB 2016, and the principal component obtained is used as the input of neural network neurons.

2.2.1. Analysis of the correlation of various factors. The original data is standardized by Min-max method (formula 2) to eliminate the influence of each factor dimension. Then the correlation coefficient matrix of each factor is obtained by the function Corrccoef. Then KMO and Bartlett tests are performed on each factor to determine whether the selected factor is suitable for principal component analysis.

$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (2)$$

Where: x_i is the first initial value of index X, x'_i is the value of x_i after standardizing, $\max(x)$ and $\min(x)$ are the maximum and minimum values of index X respectively.

2.2.2. Get the principal component expression. Using the PCA function of MATLAB 2016 to do correlation analysis, the eigenvalue, the contribution rate of principal component and the score matrix are obtained, and the principal component expression is as follows.

$$Y_i = X * W_i \quad (3)$$

Where: Y_i is the first principal component, X is the vector composed of all factors affecting landslide motion, and W_i is the score vector of the second principal component.

3. Results and Analysis

3.1. Influencing Factors of Landslide

The main factors affecting the movement of slope foot landslide are geological structure, lithology, seismic grade, water content of landslide body, volume of landslide body, slope, topographic

fluctuation, elevation, normalized vegetation index and so on. The landslide samples selected in this study are slope-foot landslides caused by Wenchuan earthquake. The geological structure, lithology and water content of the landslide body in this area are relatively close. Earthquake is the starting factor of landslide, so magnitude will not affect the subsequent movement of the landslide body. These four factors are not taken into account in the construction of the model. Normalized vegetation index has a low correlation with landslide development [7], and has little influence on the distance of landslide movement. In order to improve the efficiency of model training, this factor is discarded when building the model. According to the relevant research [8-10], the following six factors are finally determined as research indicators: the length of slip surface above the foot of slope (L), the average slope degree (α), the angle of slope foot (β), the average slope gradient (γ), the volume of the landslide body (V), and the height difference of the front and rear edges of the landslide body (H).

3.2. Principal Component Analysis Results

Making principal component analysis of the determined research indicators, the correlation coefficient matrix can be obtained (table 1), it can be seen that most of the selected factors are highly correlated. KMO and Bartlett tests for each factor showed that the KMO test coefficient was 0.606, bigger than 0.5, and the Sig value of Bartlett test was 0.001, less than 0.05. In summary, we can see that there is a strong correlation between the selected factors, which is suitable for principal component analysis. Table 2 shows that the common factor variance of each factor is close to 1.0, which indicates that the extracted principal components can better reflect the information of the original data.

Eigenvalues and principal component contribution rate is shown in table 3. The cumulative variance contribution rate of the first three principal components reaches 92.87%, and the eigenvalues of principal components 1, 2 and 3 are 3.17, 1.35 and 1.03 respectively, which are all bigger than 1. These three principal components can fully reflect the influence of each factor on the movement distance of landslide. The principal component score matrix is shown in table 4, from which the expressions of the three principal components are as follows:

$$Y_1 = 0.4x'_V + 0.51x'_L + 0.5x'_H + 0.09x'_\alpha + 0.41x'_\beta - 0.39x'_\gamma \quad (4)$$

$$Y_2 = -0.34x'_V - 0.18x'_L + 0.01x'_H + 0.8x'_\alpha + 0.45x'_\beta + 0.1x'_\gamma \quad (5)$$

$$Y_3 = 0.27x'_V + 0.26x'_L + 0.34x'_H + 0.31x'_\alpha - 0.4x'_\beta + 0.69x'_\gamma \quad (6)$$

Where: x'_V , x'_L , x'_H , x'_α , x'_β , x'_γ respectively indicate the volume of landslide body, the length of sliding surface above the foot of slope, the height difference of landslide body, the average gradient of sliding surface, the angle of foot of slope and the average gradient of accumulation area after standardizing.

It can be seen that the volume of landslide body (V), the length above the foot of the slope (L) and the height difference of the landslide (H) all have higher loads on the first principal component Y_1 , which reflects the energy accumulation ability of the landslide in the acceleration stage, and can be called the kinetic energy factor; the average slope of the sliding surface (α) and the angle of the slope foot (β) both have high load on the second principal component Y_2 , which reflects the acceleration ability of the landslide body in the acceleration stage, and can be called acceleration factor; the angle of the slope foot (β) and the average slope of the accumulation area (γ) both have high load on the third principal component Y_3 , which reflects the energy loss of the landslide body from excessive to deceleration stage, and can be called resistance factor.

Table 1. Correlation matrix

Index	V	L	H	α	β	γ
V	1.000	0.715	0.672	0.055	0.492	-0.509
L	0.715	1.000	0.941	0.066	0.442	-0.447
H	0.672	0.941	1.000	0.332	0.522	-0.330
α	0.055	0.066	0.332	1.000	0.501	0.216
β	0.492	0.442	0.522	0.501	1.000	-0.736
γ	-0.509	-0.447	-0.330	0.216	-0.736	1.000

Table 2. Common factor variance

Index	Initial	Extraction	Index	Initial	Extraction	Index	Initial	Extraction
V	1.000	0.731	H	1.000	0.930	β	1.000	0.992
L	1.000	0.935	α	1.000	0.985	γ	1.000	0.998

Table 3. Eigenvalues and principal component contribution rate

Component	Initial principal component		
	Eigenvalue	Contribution of variance/(%)	Cumulative/(%)
1	3.17	53.11	53.11
2	1.35	22.56	75.67
3	1.03	17.20	92.87
4	0.41	6.88	99.75
5	0.01	0.24	99.99
6	0.01	0.01	100.00

Table 4. Principal component scoring matrix

Index	Principal component score		
	Principal component one	Principal component two	Principal component three
V	0.40	-0.34	0.27
L	0.51	-0.18	0.26
H	0.50	0.01	0.34
α	0.09	0.80	0.31
β	0.41	0.45	-0.40
γ	-0.39	0.10	0.69

3.3. Network Model Prediction and Analysis

Finally, the genetic population size is 20, the number of evolutions is 50, the crossover probability is 0.4, the mutation probability is 0.2, the network learning efficiency is 0.1, and the target error is 0.01. Based on this, the prediction model of slope foot landslide motion distance caused by earthquake is constructed. During the experiment, 60 samples were randomly selected from the sample population as the training set of the network model, and the remaining 15 samples were used as the test set. Absolute error (AE), absolute percentage error (APE) and average absolute percentage error (MAPE) are used to evaluate the accuracy of prediction results. The calculation formulas are as follows:

$$E_{ae} = \text{abs}(L_p - L_{\max}) \quad (7)$$

$$E_{ape} = \frac{\text{abs}(L_p - L_{\max})}{L_{\max}} \times 100\% \quad (8)$$

$$E_{mape} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\text{abs}(L_p - L_{\max})}{L_{\max}} \times 100\% \right) \quad (9)$$

Where: E_{ae} is the absolute error of prediction, E_{ape} is the relative error of prediction; E_{mape} is the average relative error of prediction; L_p is the predicted value of the model, L_{max} is the expected value of the model, n is the capacity of the test set.

In order to verify the superiority of the neural network compared with the traditional mathematical statistics method and the optimization ability of genetic algorithm to BP neural network in this kind of problems, the least squares method is used to do multiple regression and BP neural network is used to predict the distance of Landslide Motion on the basis of principal component analysis, and the maximum vertical motion distance H The regression equation of the maximum horizontal motion distance L can be found in formula (10) and (11). The prediction results of each method can be seen in figure 2 and 3 below. The absolute error can be seen in figure 4 and 5 below.

$$L = -184.15 - 0.85Y_1 + 2.42Y_2 + 4.26Y_3 \quad (10)$$

$$H = 58.83 + 0.38Y_1 - 0.39Y_2 + 1.44Y_3 \quad (11)$$

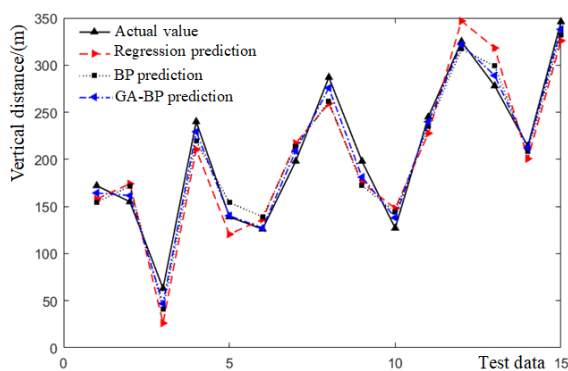


Figure 2. Prediction of maximum horizontal motion distance

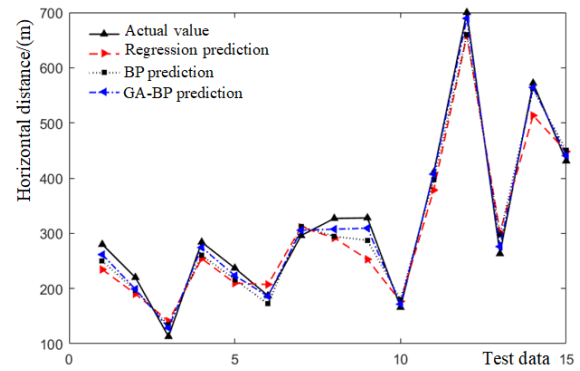


Figure 3. Prediction of maximum vertical motion distance

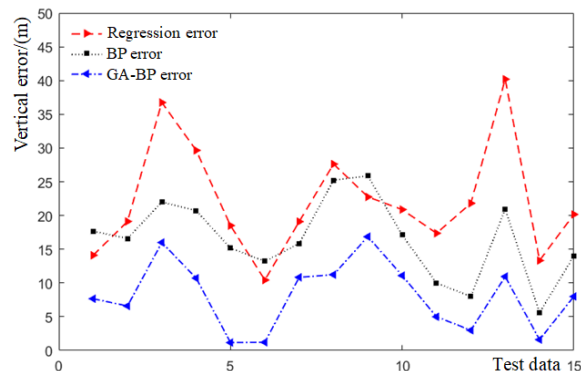


Figure 4. Absolute error of vertical motion distance

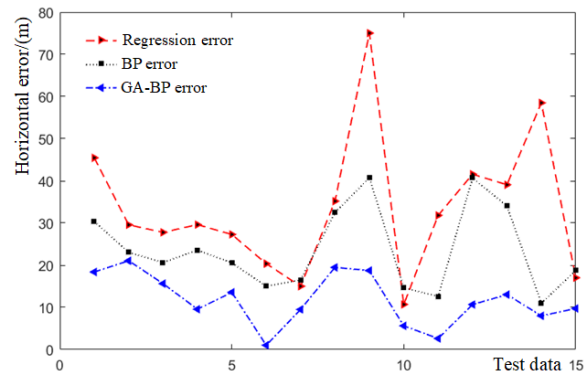


Figure 5. Absolute error of horizontal motion distance

Through the analysis of the predicted data, GA-BP neural network has the best prediction effect on the landslide movement distance. The minimum absolute error of vertical and horizontal movement distance is 1.16 meters and 1.01 meters, the minimum relative error is 0.74% and 0.54%, and the relative error is less than 10% which accounts for 93.33% and 86.67% of the test samples respectively. The standard deviations of errors are 4.87 and 5.91 respectively. The accuracy and stability of GA-BP neural network prediction are better than those of multiple regression and BP neural network prediction. The specific prediction errors are shown in table 5 and 6 below.

Table 5. Absolute error distribution

Prediction model	AE Absolute error/(m)			AE analysis Standard deviation	APE Average percentage error/(%)			APE analysis	
	Max	Min	AVG		Max	Min	AVG	<=10%	Proportion/(%)
Regression	40.19	10.43	22.11	8.02	58.31	5.82	13.35	8	53.33
BPNN	25.86	5.59	16.52	5.69	34.91	2.45	9.99	9	60.00
GA-BPNN	16.86	1.16	8.12	4.87	25.38	0.74	5.12	14	93.33

Table 6. Relative error distribution

Prediction model	AE Absolute error/(m)			AE analysis Standard deviation	APE Average percentage error/(%)			APE analysis	
	Max	Min	AVG		Max	Min	AVG	<=10%	Proportion/(%)
Regression	75.04	10.69	33.60	16.33	24.59	3.94	11.66	8	53.33
BPNN	40.72	10.98	23.63	9.46	18.26	1.92	8.62	10	66.67
GA-BPNN	21.04	1.01	11.77	5.91	13.83	0.54	4.57	13	86.67

4. Conclusion and Discussion

Firstly, the principal components affecting the movement distance of slope-foot landslide are extracted by principal component analysis, and the intrinsic mechanism of landslide movement is studied. Then, the prediction model of the movement distance of slope-foot landslide is constructed based on BP neural network optimized by genetic algorithm, and the movement distance of slope-foot landslide is predicted, and the results of the network model were compared with those of the multiple regression prediction. The conclusions are as follows:

(1) The GA-BP neural network model is accurate and stable in predicting the distance of landslide motion, and its strong ability to deal with non-linear problems is suitable for solving Landslide Motion Problems with complex internal mechanism.

(2) GA-BP neural network model is superior to the original BP neural network model in predicting accuracy and stability. It shows that in this problem, genetic algorithm has a great optimization on BP neural network from the training process of the network to the output of the results.

(3) For the prediction of landslide movement distance, GA-BP neural network is the best, followed by BP neural network. The deviation based on multiple regression prediction is large and unstable, and the effect is the worst. It shows that the neural network is more suitable for the study of this kind of problem than the traditional mathematical statistics method.

Using GA-BP neural network to construct the prediction model of landslide motion distance is an attempt to combine artificial intelligence with practical engineering problems. It introduces new methods and new ideas for this kind of research. At the same time, it has practical significance to predict the impact range and disaster intensity of slope foot landslides caused by earthquakes in southwestern China. Because of the complexity of influencing factors and intrinsic mechanism of landslide movement, this paper only selected the foot-slope landslide movement caused by earthquake in southwestern China as the research object. It is found that the prediction model of landslide motion distance is also affected by the scale interval of landslide, and further study is needed for this problem.

5. References

- [1] FAN Xiaoyi, LENG Xiaoyu et al. The movement distance and topographic factors of slope foot type and transition earthquake landslide [J]. Geotechnical Mechanics, 2015, 36(5): 1380-1388.
- [2] LI Ruikuan, LIANG Qingguo et al. Research achievements and prospects of landslide mechanism and slope stability [J]. Soil and Water Conservation in China, 2019(3): 40-42.

- [3] MENG Huajun, JIANG Yuanjun, et al. Variation analysis of influencing factors of landslide movement distance in Dujiangyan mountain area before and after Wenchuan earthquake [J]. Journal of Geomechanics, 2017, 23(6): 904-913.
- [4] TANG Ran, XU Qiang, et al. Calculating model of moving distance of push-type landslide [J]. Geotechnical Mechanics, 2018, 39(3): 1009-1020.
- [5] ZHENG Guang, XU Qiang, et al. Study on the Formula for Calculating the Movement Distance of Rock Landslide-Clastic Flow [J]. Geotechnical Mechanics, 2019, 40(12): 01-10.
- [6] HAO Shanshan, LI Menghua, et al. Significance analysis of soil erosion factors in loess hilly gully region [J]. Science of Soil and Water Conservation, 2019, 17(2): 77-86.
- [7] ZHANG Feng. Application of PSO-RBF Neural Network in Landslide Risk Prediction in Historic Landslide Developing Areas [J]. Geographic and Geographic Information Science, 2018, 34(4): 118-124.
- [8] LIAO Jun, DONG Qian, et al. Influencing factors of motion parameters of high-level landslides based on UDEC [J]. Science, Technology and Engineering, 2018, 18(34): 29-36.
- [9] WEI Changli, CHEN Liang, et al. Prediction of Causes and Stability of Landslide in Ozhai Village, Upper Reaches of Minjiang River [J]. South-to-North Water Transfer and Water Science and Technology, 2019, 17(3): 185-192.
- [10] WANG Wenpo, HAN Aiguo, et al. Sensitivity Analysis of Environmental Factors for Landslide Disaster Pregnancy in Puge County, Sichuan Province [J]. Journal of Yangtze Academy of Sciences. 2018, 38 (9): 63-67.