Available online at www.ijpe-online.com

International Journal of Performability Engineering

vol. 14, no. 6, June 2018, pp. 1171-1180
DOI: 10.23940/ijpe.18.06.p8.11711180

Monitoring and Warning Methods of Tailings Reservoir using BP Neural Network

Tianyong Wu[a], Chunyuan Zhang[b,*], and Yunsheng Zhao[a]

[a]*Faculty of Engineering, China University of Geosciences, Wuhan, 430074, China*
[b]*School of Computer Science, China University of Geosciences, Wuhan, 430074, China*

Abstract

The tailings reservoir is a major hazard source with high potential energy, which may cause artificial debris flow. The stability of the tailings reservoir is extremely important to the normal operation of the mining enterprises and the safety of people's lives and property. In order to reduce the risk of a tailings accident, a multivariate linear regression model, a BP neural network and a regression analysis model optimized by genetic algorithm are established in this article to discuss the monitoring and warning method of the tailings reservoir. It takes the safety monitoring data of the Huangmailing tailings as an example to make a comparison of three forecasting models by taking fitness, simulating capability of initial data and the predicting ability of new data into consideration. The results of the experiment show that the BP neural network forecasting model is better able to predict safety monitoring data over the other two models. The predicting ability of the regression analysis model optimized by genetic algorithm is better than the forecasting capability of the multivariate linear regression model.

Keywords: tailings; multivariate linear regression; BP neural network; genetic algorithm

(Submitted on February 27, 2018; Revised on April 1, 2018; Accepted on May 21, 2018)

1. Introduction

Tailings are constructed by damming intercept troughs or encirclement and are used to store waste residue that is discharged. The safety of the tailings is related to the safety of life and property of the state and the people. Discovery and early warning of security risks is an important part that cannot be neglected in the process of safe production. It is of great significance to real-time monitor the operation status and provide technical support for the actual work of tailings. In recent years, with the popularization and rapid development of computer and information technology, more and more information theory has been proven in the practice of engineering.

In 2012, Y. B. Wang [9] etc. established a prediction model for the mine tailing accident rate by adopting a harmony search algorithm and BP neural network to prevent mine tailing accidents effectively. In the same year, B. Yang [12] combined a neural network theory with a basic theory of genetic algorithms and established the early warning model of tailings reservoir based on a genetic algorithm neural network. In 2013, X. R. Zhang [15] etc. established a prediction model based on a RBF neural network to predict the built-up area in Hefei City, and used a BP neural network, simple linear regression model and multiple linear regression model as a comparison. In the same year, X. Z. Lang [3] proposed an improved genetic algorithm to optimize weights of BP neural networks and built a saturation line forecasting model to provide decision support for safety assessment on tailing ponds. In 2014, L. Shang [6] etc. chose to use a multiple linear regression analysis model and BP neural network model to analyze and forecast the scale of construction land of Xingtai City in Hebei Province. In the same year, Q. L. Zhang [13] etc. established a GA_SVM prediction model to optimize the flocculating sedimentation parameters. In 2015, X. M. Wang [8] etc. established the support vector machine regression model, and the model parameters were optimized through the genetic algorithm to make an optimal prediction. In the same year, Y. B. Wang [10] etc. optimized the radial basis function neural network with genetic algorithm and applied the optimized RBF neural network in the safety prediction of

* Corresponding author.
E-mail address: 2472140040@qq.com

tailings. In 2016, Q. L. Zhang [14] etc. applied a back propagation neural network and genetic algorithm to establish the flocculation sedimentation parameters prediction model of crude tailings to get the optimum parameters of flocculating sedimentation. In the same year, T. L. Xu [11] etc. built the saturation line forecasting model based on a genetic neural network algorithm to predict the complex and non-linear saturation line of the tailings dam. D. C. Huang and S. C. Xie [1] established the tailings dam sedimentation forecast model based on a BP neural network to make scientific and rational predictions against upcoming deformations. Z. K. Pi [5] etc. integrated a multivariate principal component regression analysis algorithm to get the main factors and established a PCR-SPA predictive model for gas emission. In 2017, Y. X. Ke [2] etc. introduced magnetic treatment technique into dewatering and concentrating CTR, and they optimized the SVM model by obtaining a GA-SVM model for optimizing sedimentation parameters of pre-magnetized CTR. In the same year, Z. X. Liu [4] etc. selected a fractional dimension number and a correlation coefficient of the fractional dimension number to characterize the geometric features of whole tailing and constructed a fractal-BP neural network model.

In order to predict the safety monitoring data of tailings timely and accurately, this article establishes a multivariate linear regression predicting model, a regression model optimized by genetic algorithm and a BP neural network forecasting model, as well as a comprehensive and comparative analysis of the three models through the forecasting results of the safety monitoring data of the Huangmailing tailings.

2. Forecasting models and the theories of algorithms

2.1. Multivariate linear regression model

Multivariate linear regression (MLR) analysis is a method that discusses the linear relationship between a dependent variable and multiple independent variables. It returns the parameters determined by the MLR model to the hypothetical equation, and it predicts the trend of the dependent variable through the regression equation. The general equation of the multiple linear regression model is shown as Equation (1).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_j x_j + \cdots + \beta_k x_k + \mu \tag{1}$$

In Equation (1), *k* is the number of explain variables, and β_j is the coefficient of the regression. μ is the random error after removing the influence of independent variables on *y*.

After the parameters of the MLR model are estimated, it is necessary to conduct further statistical examination of the regression function of the sample to determine the reliability of the estimation. The indexes consist of the fitness, the general linear significance of the equation, the significance of variables, and the confidence interval of parameters.

In the process of fitness's examination, the simulating degree of observed values of the samples is measured by the statistical and regressing values of the samples. If total sum of squares (TSS) are $\sum (Y_i - \bar{Y})^2$, the explained sum of squares (ESS) are $\sum (\hat{Y}_i - \bar{Y})^2$, and the residual sum of squares (RSS) are $\sum (Y_i - \hat{Y}_i)^2$, the relationship among them is shown as Equation (2).

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \tag{2}$$

In Equation (2), R2 is the coefficient of determination. The shorter the distance between R^2 and the number 1, the higher the fitness of the model will be.

In the process of testing the general linear significance of the equation, the statistic expression is shown as Equation (3) under the condition that the original hypothesis of H_0 is feasible.

$$F = \frac{ESS / K}{RSS / (n-k-1)} \tag{3}$$

In Equation (3), the freedom degree of distribution *F* is obeyed to $(k, n-k-1)$. The given significance level α often takes 0.05 or 0.1, so the degree of confidence is 95 percent or 90 percent. After calculating the statistic value *F* of the samples and obtaining the critical value F_α in the table, whether the original hypothesis of H_0 is refused or accepted is determined by the size of F and F_α. It is the way to determine whether the general linear relationship of the original equation is significant.

In the process of examination of the significance of variables, the expression of the statistical value t is shown as Equation (4).

$$t = \frac{\hat{\beta}_j - \beta_j}{S_{\beta_j}} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{c_{jj} \dfrac{e'e}{n-k-1}}} \sim t(n-k-1) \tag{4}$$

The original hypothesis for some variable $X_j (j = 1, 2, \cdots, k)$ is $H_0 : \beta_j = 0$, and the alternative hypothesis is $H_1 : \beta_j \neq 0$. If the significance level is given, the critical value can be easily obtained as $t_{\alpha/2}(n-k-1)$. After that, whether the original hypothesis of H_0 is refused or accepted is determined by the size of $|t|$ and $t_{\alpha/2}$. It is the way to determine whether the explanatory variable should be included in the model.

2.2. Regression model optimized by genetic algorithm

- *Theories of genetic algorithm*

Genetic algorithm is an optimized method based on the biological evolution theory of "the survival of the fittest", and it has a strong ability to seeking the optimal solution. The chromosome population is produced to represent the problem to be solved. Different genetic operators among the chromosome population form a cycle evolution process to obtain the optimal solution of the problem to be solved.

In the process of genetic algorithm, *Gen* is the genetic generation, *M* is the number of individuals in the group, P_t is the probability of a cross operation, P_m is the probability of a mutation operation, and *i* is the cumulative number of individuals that have been processed. When the cumulative number *i* is equal to the total number *M*, it means that the generation of individuals have been fully processed, and it is a time to turn into the next generation of the group. Genetic generation *Gen* is not only the number of times that the genetic algorithm is operated repeatedly, but also the number of generations that the population has generated.

- *Optimization of regression model based on genetic algorithm*

When dealing with an optimized problem of multiple linear regression, the ordinary least square theory is usually used to obtain $\beta_0, \beta_1, \cdots, \beta_m$. The noise error of the measured data in many practical problems can greatly affect the outcome of the equations. In comparison, the genetic algorithm can get a more reliable solution.

When applying the genetic algorithm to deal with multiple linear regression problems, it usually regards the residual sum of squares as the fitness function, and the range of the coefficients is $[-a, a]$. During the process of the genetic algorithm, *n* is set as the population size, and *m* is set as the number of parents. The condition to end the iteration is that the number of iterations is more than 50000 times, or the difference of the fitness value between the worst individual and the optimal individual in the population is less than 10^{-4}.

In order to better apply the genetic algorithm to the potential problem in the multivariate linear analysis, this article presents two methods to conduct the crossover process of the multi-parent. The expression of the first method is shown as Equation (5).

$$X = \sum_{i=1}^{m} \alpha_i X_i ' \tag{5}$$

In Equation (5), *m* is the number of parents, and α_i should meet the conditions $\sum_{i=1}^{m}\alpha_i = 1$, $-0.5 \le \alpha_i \le 1.5$.

However, the individuals generated in the new crossover method will usually exceed the specified range. It will take a lot of the execution time of the algorithm to regenerate new individuals aimed at replacing the invalid individuals.

In the second crossover method, all the values of α_i are equal, and the ranges of α_i should meet the condition that $0 \le \alpha_1 = \alpha_2 = \cdots = \alpha_m \le (1/m + 0.15)$. The expression of the second method is shown as Equation (5) as well, and the individuals generated in the second method have a higher chance to meet the boundary conditions. The expression easily falls into partial convergence, owing to the lack of changes of α_i and the lack of diversity of new individuals.

Taking the two methods above into consideration, the new crossover method is where the two methods are performed by a certain number of iterations.

2.3. BP neural network model

The BP neural network algorithm is a supervised learning process. The input and output samples are used to train the network, and weights have been adjusted during the training process. It makes sure that the adjustment of the weights is proportional to the gradient of the errors. In this way, errors can be reduced to meet the needs of the application [7]. The BP neural network consists of an input layer, hidden layer and output layer, and its network structure is shown in Figure 1.

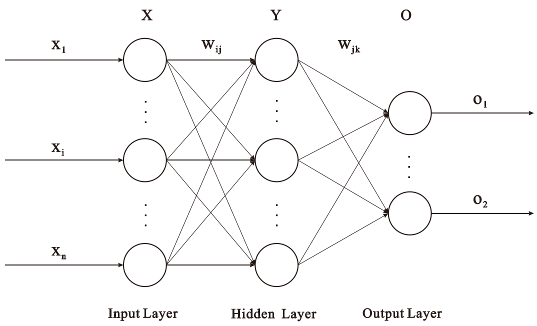

Figure 1. BP neural network structure

In the BP neural network, the unipolar sigmoid function is usually used as the transfer function, and the expression is shown as Equation (6).

$$f(x) = \frac{1}{1+e^{-x}} \tag{6}$$

In practice, bipolar sigmoid functions can also be used to meet the needs, and the expression is shown as Equation (7).

$$f(x) = \frac{1-e^{-x}}{1+e^{-x}} \tag{7}$$

In order to realize the predicting work, the BP neural network should be trained, and the training steps are as follow:

- **Step 1** Calculating errors of the network

If the values of the network output are difference than the values of the expected output, the expression of output errors is shown as Equation (8).

$$E_i = \frac{1}{2(d_i - o_i)} \tag{8}$$

The expression of total network errors is shown as Equation (9).

$$E = \frac{1}{2}\sum_{i=1}^{n} E_i \tag{9}$$

In Equation (8) and Equation (9), the variables meet the condition that $i = 1, 2, \cdots, n$.

- **Step 2** Adjusting the weights of each layer

In order to obtain the error signal δ^0 and adjust the connection weight ω, the expected outputs of the output layer d are compared to the actual outputs, and the errors signal is reverted to the hidden layer to obtain δ_j^y. The adjusting expression of the connection weight ω is shown in Equation (10) to Equation (12).

$$\delta_j^y = \delta^0 z_j y_j (1 - y_j) \tag{10}$$

$$z_j(t) = z_j(t-1) + \eta\delta^0 y_j + \mu\Delta z_j(t-1) \tag{11}$$

$$v_{ij}(t) = v_{ij}(t-1) + \eta\delta_j^v x_j + \mu\Delta v_{ij}(t-1) \tag{12}$$

In Equation (11) and Equation (12), η is the rate of learning, and μ is the momentum. The variables should meet the condition that $\eta \in (0,1)$ and $\mu \in (0,1)$.

- **Step 3** Entering the next sample and returning to step 1 to continue training.
- **Step 4** When all samples are trained, the whole training will end on the condition that $E < E_{\min}$. Otherwise, E will be set to 0 and the training will be restarted.

In the last place, the BP neural network should be examined by test samples.

3. Analysis and comparison of the case

In this article, the safety monitoring data of the Huangmailing tailings are taken as an example to compare three forecasting models. There are 10 monitoring points in the Huangmailing tailings, and one of the monitoring points is the switching center of the safety monitoring data. In order to ensure the safe operation of the tailings reservoir, it is necessary to monitor the network status of the monitoring points in the monitoring network of the Huangmailing tailings and diagnose whether the monitoring network has failed by predicting the transmission amount of safety monitoring data.

The security monitoring system database of the Huangmailing tailings is called for the monitoring data transmission of each monitoring point in the period T = 5min. There are 610 groups of data, and a part of them is shown in Table 1. After that, three forecasting models are used to predict the data transmission rate of each monitoring point.

Table 1. Monitoring data transmission amount of each monitoring point in the Huangmailing tailings

	data transmission amount									data reception amount
	①	②	③	④	⑤	⑥	⑦	⑧	⑨	M
T_1	4573.8	9783.7	13009.4	4399.2	12970.8	4487.7	4408.4	13732.2	19732.9	87132.1
T_2	4860.6	9378.4	13402.3	4603.7	13049.9	4209.4	4378.6	13213.1	19083.6	87047.7
T_3	4798.3	9583.7	13145.2	4478.8	12864.7	4389.2	4579.2	13579.4	19537.1	87219.3
T_4	4687.6	9473.6	13347.8	4037.7	13156.2	4109.3	4281.5	13689.7	19934.9	87166.8
T_5	4978.9	9564.6	13564.5	4654.6	13365.4	4234.6	4321.5	13879.9	19564.6	87891.4
T_6	4352.1	9126.1	13565.2	4234.2	12875.3	4552.1	4210.8	13215.4	19423.1	88533.3
T_7	4895.6	9642.8	13246.7	4531.2	12964.3	4359.7	4119.4	13147.8	19827.8	87369.8
T_8	4811.3	9246.5	13289.1	4154.8	13276.4	4167.8	4701.3	13642.4	19830.7	86841.8
T_9	4714.6	9478.1	13765.4	4312.6	13531.7	4296.7	4639.8	13479.2	19688.8	88772.5
T_{10}	4434.9	9603.5	13673.7	4598.7	13481.2	4516.3	4267.9	13197.6	19381.7	87276.8

3.1. Analysis of multivariate linear regression model

Variables need to be selected according to the composition of the data switching center and the data transmission amount of each monitoring point. The dependent variable y is the data receiving amount of the data switching center. The independent variable is the data transmission amount of each monitoring point, and all of them are recorded separately as x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9.

The multivariate linear regression model is established by making use of the T_1 to T_{600} groups of data in a stepwise regression method. The probability of the independent variable being close to the dependent variable and satisfying the judgment conditional argument is less than 0.05. If the probability is greater than or equal to 0.1, the independent variable will be removed. In this model, each independent variable has a strong correlation with the dependent variable. According to the output of SPSS software, the expression of the multivariate linear regression model of the data receiving amount of the data switching center is shown as Equation (13).

$$y = 0.653x_1 + 0.808x_2 + 0.870x_3 + 0.694x_4 + 0.784x_5 + 0.856x_6 + 0.766x_7 + 0.821x_8 + 0.951x_9 + 14465.186 \tag{13}$$

The new multivariate linear regression model needs to be tested for fitness, general linear significance of the equation and significance of independent variables. The testing steps are as follow:

- **Step 1** Testing the fitness of the model

The multivariate linear regression model is summarized in Table 2. According to Table 2, R=0.915, R^2=0.837, and fitness is close to 1. It indicates that the equation is well-fitted and the independent variable can explain 83.7 percent of the change of the dependent variable.

Table 2. Summary sheet of multivariate linear regression model

Model	R	R2	Adjustment of R2	Error of standard estimation
9	0.915	0.837	0.835	343.52245

- **Step 2** Testing the general linear significance of the equation

According to the Anova table of the multivariate linear regression model, the model can better explain the sum of squares. The statistic F is 331.923, and its probability is 0.00. The significance probability is far less than 0.01 with the introduction of independent variables. Therefore, the original hypothesis that the overall regression coefficient is 0 can be significantly rejected.

- **Step 3** Testing the significance of the variables

The coefficients of the multivariate linear regression are shown in Table 3, and the tolerances of each variable are different and greater than 0. Variance expansion factors are different and less than 10. Therefore, there is no collinearity between the various variables. In model 9, the t values of each variable are 0.00 and less than 0.05, so all variables are introduced into the regression mode.

Table 3. Coefficients of multivariate linear regression model

Model	Non-standardized coefficient		Standardized coefficient	t	Sig.	Collinear statistics	
	B	Standardized errors	Trial version			Tolerance	VIF
Constant	14465.186	1579.890		9.156	0.000		
X_9	0.951	0.024	0.671	39.801	0.000	0.986	1.014
X_8	0.821	0.046	0.303	17.799	0.000	0.966	1.035
X_7	0.870	0.050	0.297	17.516	0.000	0.976	1.024
X_6	0.856	0.051	0.286	16.856	0.000	0.971	1.029
X_5	0.784	0.056	0.237	14.102	0.000	0.991	1.009
X_4	0.808	0.062	0.222	13.049	0.000	0.964	1.037
X_3	0.766	0.060	0.213	12.668	0.000	0.987	1.013
X_2	0.694	0.062	0.190	11.261	0.000	0.988	1.012
X_1	0.653	0.061	0.179	10.640	0.000	0.990	1.010

3.2. Analysis of the regression model optimized by genetic algorithm

During the analyzing process of the regression model optimized by the genetic algorithm, the residual sum of square is chosen as the fitness, and the range of $\beta_0, \beta_1, \cdots, \beta_9$ is the interval $[-50, 50]$. In this model, the population size is set to 150, and the number of parents is set to 7. The end condition of the iteration is that the number of iteration times is greater than 50000, or that the value of the difference between the worst individual and the optimal individual in the population is less than 10^{-4}. The expression of the regression model optimized by genetic algorithm is shown as Equation (14).

$$y = 0.660x_1 + 0.812x_2 + 0.870x_3 + 0.702x_4 + 0.792x_5 + 0.858x_6 + 0.770x_7 + 0.829x_8 + 0.952x_9 + 14067.071 \quad (14)$$

The minimum of the residual sum of square of the regression data and the samples data is 6.88×10^7, and the value of fitness is 0.917. Therefore, the normal probability of errors is 84.1 percent, and the value of *F* is 352.969. Compared with the ordinary multivariate linear regression model, the regression results of the two models are nearly the same, and the regression model optimized by the genetic algorithm is slightly better.

3.3. Analysis of BP neural network model

The samples selected in the BP neural network model need to be normalized to eliminate the different dimensions between the initial data. In this model, the T_1 to T_{590} groups of data are treated as the training samples, and the T_{591} to T_{600} groups of data are regarded as the testing samples. The input variable is the data transmission amount of each monitoring point, and the output variable is the data receiving amount of the data switching center.

The BP neural network model needs to be established. A three-layer BP neural network can easily meet the mapping requirements of general functions and approximate to any number of variable functions through any precision requirements. Therefore, a three-layer BP neural network is built in this article, and there are 9 nodes in the input layer, 3 nodes in the hidden layer, and 1 node in the output layer in the initial structure of the model. The number of nodes in the hidden layer is determined by the training results and is finally set to 6.

The BP neural network model needs to be trained. The goal of errors is set to 0.0000000001, and the learning rate is set to 0.01. The maximum number of iterations is set to 500, and the performance function is *MSE*. After that, the BP neural network model is trained to obtain the intended target. According to the results of the tests, the BP neural network gains the fastest converge rate and meets the requirements at step 6. The training process is shown in Figure 2.

At this time, the value of the gradient is 0.00318, and the value of MU is 0.0001. The value of validation checks is 6, and the total *R* is 0.92171. Therefore, the model has a good performance in fitness, and the normal probability of errors is 84.95 percent. The regression analysis process of the BP neural network is shown as Figure 3.

Simulating function is applied to test the trained BP neural network by making use of the test samples, and the test results can nearly meet the needs of the intended setting.

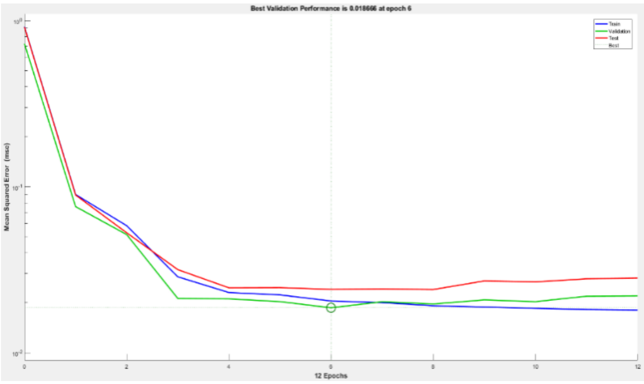

Figure 2. Training process of BP neural network model

3.4. Comparison and analysis of the results of models

- *Comparison of three models*

The fitness of the ordinary multivariate linear regression model is 0.915, and the normal probability of errors is 83.7 percent. The fitness of the regression model optimized by genetic algorithm is 0.917, and the normal probability of errors is 84.1 percent. The fitness of the BP neural network model is 0.92171, and the normal probability of errors is 84.95 percent.

Taking the results of the three models into comparison, the BP neural network model has more advantages than the regression model optimized by the genetic algorithm in fitness and modeling effects, and the ordinary multivariate linear regression model has the worst performance in the modeling process.

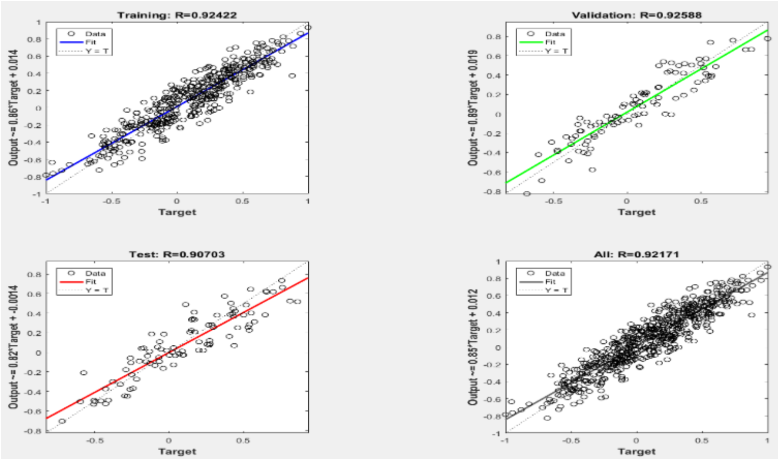

Figure 3. Regression analysis diagram of BP neural network model

- *Comparison of the simulation capabilities of the initial data of three models*

In order to find out the simulation capabilities of the initial data of the three models, the T_{11} to T_{20} groups of the initial data are taken into comparison as the simulating values, and the comparison process is shown as Table 4.

Table 4. The comparison between the real value and the simulating value of three models

	Real value	MLR model	Relative errors (%)	GA-MLR model	Relative errors (%)	BP network model	Relative errors (%)
T_{11}	86143.4	86165.9	-0.0261	86127.1	0.0189	86155.8	-0.0143
T_{12}	86365.4	86326.6	0.0449	86375.6	-0.0118	86366.4	-0.0011
T_{13}	87010.6	87366.6	-0.4091	87336.9	-0.3750	87276.9	-0.3060
T_{14}	87741.6	87506.3	0.2682	87533.3	0.2374	87634.8	0.1217
T_{15}	88253.0	88008.5	0.2770	88044.4	0.2363	88123.2	0.1471
T_{16}	87200.0	87011.0	0.2167	87041.6	0.1816	87103.1	0.1111
T_{17}	87322.2	87503.4	-0.2075	87471.7	-0.1712	87441.7	-0.1368
T_{18}	88181.2	88398.1	-0.2459	88365.1	-0.2085	88080.1	0.1146
T_{19}	87929.5	87192.9	0.8377	87226.2	0.7998	87461.4	0.5323
T_{20}	86568.6	87010.4	-0.5104	86977.5	-0.4726	86800.8	-0.2682

The curve chart of the comparison between the real value and the simulating value is shown as Figure 4. The real value is indicated by a black solid line, and the simulating value of the ordinary multivariate linear regression model is indicated by a red dotted line. The blue dotted line indicates the simulating value of the regression model optimized by the genetic algorithm, and the green dotted line indicates the simulating value of the BP neural network model. According to Figure 4, the simulating value of BP neural network model is closest to the real value, and the simulating value of the ordinary multivariate linear regression model is furthest from the real value.

The comparison chart of absolute errors is shown in Figure 5. The red dotted line indicates the absolute errors of the ordinary multivariate linear regression model, and the blue dotted line indicates the absolute errors of the regression model optimized by the genetic algorithm. The absolute errors of the BP neural network model are indicated by the green dotted line. According to Figure 5, the BP neural network model has the smallest absolute errors fluctuation of the simulating value, and the biggest absolute errors fluctuation of the simulating value are in the ordinary multivariate linear regression model. To sum up, the BP neural network model has a better performance in the initial data simulation and has a higher stability compared to the other two models.

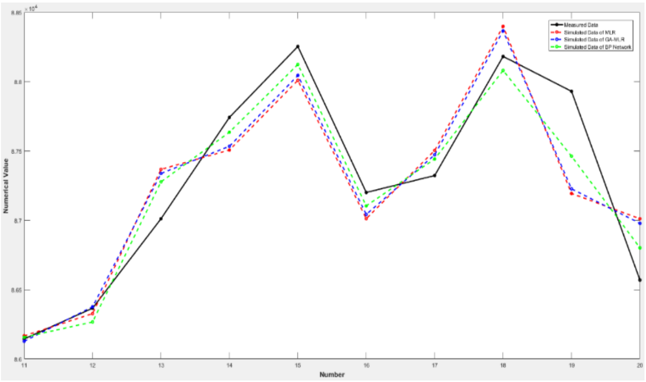

Figure 4. The comparison curve between the real value and the simulating value of three models

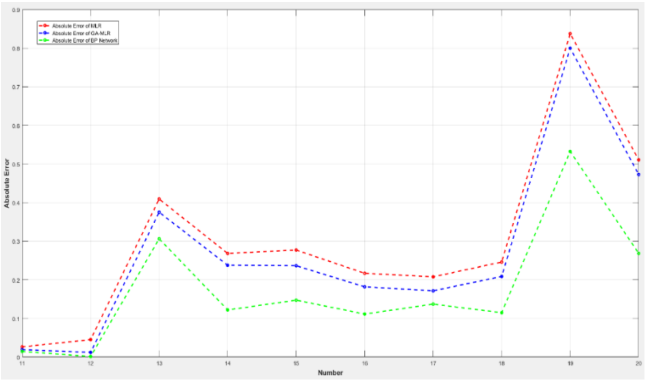

Figure 5. The absolute errors comparison of three models

- *Comparison of the predicting capabilities of the new data of three models*

In order to learn the predicting capabilities of the new data of the three models, the T_{601} to T_{610} groups of new data are taken into comparison as the predicting values, and the comparison process is shown in Table 5.

Table 5. Comparison between real values and predicting values

	Real value	MLR model	Relative errors (%)	GA-MLR model	Relative errors (%)	BP network model	Relative errors (%)
T_{601}	88780.5	88923.9	-0.1615	88695.8	0.0954	88728.0	0.0591
T_{602}	87995.3	87871.5	0.1406	87841.5	0.1747	88026.2	-0.0351
T_{603}	86678.2	86945.9	-0.3088	86905.4	-0.2621	86874.6	-0.2266
T_{604}	88599.3	88365.0	0.2644	88389.5	0.2644	88443.6	0.1757
T_{605}	86468.4	86792.4	-0.3747	86751.6	-0.2819	86712.2	-0.2819
T_{606}	86801.0	86938.7	-0.1586	86902.4	-0.1168	86840.5	-0.0455
T_{607}	89107.4	88940.8	0.1869	88953.3	0.1729	88981.4	0.1414
T_{608}	87695.7	87678.7	0.0193	87646.1	0.0565	87671.1	0.0281
T_{609}	87217.0	87452.3	-0.2697	87332.7	-0.1326	87297.5	-0.0923
T_{610}	87003.4	86879.8	0.1421	87070.2	-0.0768	87036.3	-0.0378

According to Table 5, the BP neural network model has smaller predicting errors of the T_{601} to T_{610} groups of data than the regression model optimized by the genetic algorithm. The highest predicting errors of the T_{601} to T_{610} groups of data are in the ordinary multivariate linear regression model. Taking the three models into comparison, the BP neural network model has more advantages than the other two models in forecasting accuracy and modeling stability.

4. Conclusions

In this article, the multivariate linear regression model, the regression model optimized by a genetic algorithm and the BP neural network model are established to simulate and predict the safety monitoring data of Huangmailing tailings. According to the results, it can be proven that the BP neural network model has a better fitness, stronger simulating abilities of initial data and more powerful predicting capabilities of new data than the regression model optimized by a genetic algorithm. The ordinary multivariate linear regression model has the worst predicting ability among the three forecasting models. However, the learning efficiency of the BP neural network is fixed, and the convergence speed of the network algorithm is slow. In the future, our work will focus on the optimization of the BP neural network algorithm to improve the convergence rate.

Acknowledgements

The authors are grateful to the area editor and the anonymous reviewers for their valuable comments and suggestions on this paper. This work was a provincial and ministerial level Vertical Project under Grant No. 2012056068, and it was conducted by the China University of Geosciences.

References

1. D. C. Huang and S. C. Xie, "A Way to Predict the Settlement of Tailings Dam based on BP Neural Network," *Engineering of Surveying and Mapping*, vol. 25, no. 8, pp. 53-56, 2016
2. Y. X. Ke, X. M. Wang, and Q. L. Zhang, "Optimal Prediction Model on Sedimentation Parameters of Pre-Magnetized Crude Tailings Slurry," *Journal of Chongqing University. Natural Science Edition*, vol. 40, no. 1, pp. 48-56, 2017
3. X. Z. Lang, "Study on Tailings Dam Saturation Line Prediction based on Genetic Algorithm and Neural Network," *Shandong: Shangdong University of Technology*, 2013
4. Z. X. Liu, Y. C. Gong, and X. B. Li, "Study on the Backfilling Material Properties based on Fractal Theory and BP Neural Network," *Gold Science and Technology*, vol. 25, no. 2, pp. 38-44, 2017
5. Z. K. Pi, B. S. Jia, and Z. X. Li, "Study on Prediction Model of Gas Emission in Coal Working Face based on Spss," *Computer Engineering and Application*, vol. 52, no. 15, pp. 228-232, 2016
6. L. Shang, Y. P. Da, and P. T. Zhang, "Study on Prediction Model of Urban Construction Land Scale a Case Study of Xingtai in Hebei Province," *Chinese Agricultural Science Bulletin*, vol. 30, no. 32, pp. 154-158, 2014
7. T. Y. Wu, C. Y. Zhang, and Y. S. Zhao, "A Study on Faults Diagnosis and Early-Warning Method of Tailings Reservoir Monitoring Points based on Intelligent Discovery," *International Journal of Performability Engineering*, vol. 13, no. 5, pp. 697-710, 2017
8. X. M. Wang, J. W. Zhao, and D. M. Zhang, "Optimal Prediction Model of Flocculating Sedimentation Velocity of Unclassified Tailings," *The Chinese Journal of Nonferrous Metals*, vol. 25, no. 3, pp. 793-798, 2015
9. Y. B. Wang, L. Wang, and Z. X. Li, "Safety Evaluation of Mine Tailings Facilities based on HS-BP Algorithm," *Systems Engineering-Theory & Practice*, vol. 32, no. 11, pp. 2585-2590, 2012
10. Y. B. Wang, J. F. Yan, and X. Zhong, "Tailings Dam Safety Prediction based on Optimizing RBF Neural Networks by Genetic Algorithm," *Computer Applications and Software*, vol. 32, no. 3, pp. 182-184, 2015
11. T. L. Xu, Y. B. Wang, and K. Chen, "Tailings Saturation Line Prediction based on Genetic Algorithm and BP Neural Network," *Journal of Intelligent & Fuzzy Systems*, vol. 30, no. 4, pp. 1947-1955, 2016
12. B. Yang, "Research on the Key Technologies of Tailings Pond Online Safety Monitoring and Pre-Warning," *Wuhan: China University of Geosciences*, 2012
13. Q. L. Zhang, Q. S. Chen, X. M. Wang, and C. C. Xiao, "Study on GA_SVM Optimal Prediction Model on Flocculating Sedimentation Parameter of Unclassified Tailings," *Journal of Safety Science and Technology*, vol. 10, no. 5, pp. 24-30, 2014
14. Q. L. Zhang, Q. Liu, and J. W. Zhao, "Study on the Parameters Prediction Model of Flocculating Sedimentation of Crude Tailings," *Journal of Northeastern University. Natural Science*, vol. 37, no. 6, pp. 875-879, 2016
15. X. R. Zhang, C. L. Fang, Z. B. Wang, and H. T. Ma, "Prediction of Urban Built-up Area based on RBF Neural Network," *Resources and Environment in the Yangtze Basin*, vol. 22, no. 6, pp. 691-697, 2013

Tianyong Wu is a Ph. D. student from the Faculty of Engineering, China University of Geosciences. His research interests include machine learning and neural network algorithm.

Chunyuan Zhang is a Master's student from the School of Computer Science, China University of Geosciences. Her research interests include neural network algorithm, remote sensing classification and recognition.

Yunsheng Zhao graduated from the Faculty of Engineering, China University of Geosciences, with a Bachelor's, Master's and Ph. D degree. Now, he is a professor of the China University of Geosciences, and the Director of the Department of Safety Engineering, China University of Geosciences, Wuhan, China. His current research interests include system security theory and application, safety information engineering, and occupational safety and health.