Prediction Model of Soil Nutrients Loss Based on Artificial Neural Network

Wang Zhiliang*[1]*, Zhang Li*[2]*, Liang Chuan*[1] and *Fu Qian

[1]Hydroelectric College, Sichuan University
[2]Water resources College, Hohai University
E-mail: lhlzi@371.net

Abstract: On the basis of Artificial Neural Network theory, a back propagation neural network with one middle layer is building in this paper, and its algorithms is also given. Using this BP network model, study the case of Malian-River basin. The results by calculating show that the solution based on BP algorithms are consistent with those based multiple-variables linear regression model. They also indicate that BP model in this paper is reasonable and BP algorithms are feasible.
Keywords: artificial neural network, multiple-variables linear regression analysis, soil erosion

1 Introduction

Water and soil loss has caused severe damage to ecological environments. With soil erosion, many nutrients in the soil run off and land's productivity decline. This leads crops reduction of output in a large area. The silt caused by soil erosion deposits in the reservoir to decrease the volume of reservoir and deposits in water way to obstructer shipping. Objectively, although many measures have been taken to control soil erosion, the excepted goal doesn't be reach. To strengthen scientific research on water and soil loss and to heighten prediction of water soil loss in focal point regions are very urgent. The main method to predict the loss of soil nutrients is multiple-variables linear regression model. But some hypothesizes which are necessary to building multiple-variables linear regression model are as follows [1]: all independent variables are non-stochastic variables; there is no relationship of multiple-layer common linear between all the independent variables. The representative problem that this method can solve is one dependent variable versus several independent variables. Nevertheless, when researching the problem about loss of soil nutrients, we study the relationship between soil nutrients (assuming it includes three components: N, P and K) and precipitation in the catchments, sediment volume and runoff of floods, and that this relation is obviously multiple-variables versus multiple-variables. At the same time, since correlative relationship exists among three independent variables, multiple-layer common linear phenomenon occurs. In view of this point, to seek new method and new way for this kind of problem is necessary.

Artificial Neural Networks (ANN) is one kind of information processing system, which imitates neuron structure and functions, and it takes on the similar characteristics: parallel property of processing problem, distributed property of knowledge storage, and so on. After being trained, it is able to adapt outer environment, to recognize pattern and to reason synthetically. ANN is a new method of processing data. Compared with traditional methods, it possesses some advantages as follows: (1) Data is fuzzy or non-linear; (2) Pattern character is dim; (3) There is much noise in data. Consequently, ANN technology is widely used in many fields of nature science and social science. Because solving the learning problem of forward feedback networks, presently, back propagation network is the most popular ANN. This paper wants to build a BP network model with three input items and three output items, and to research the predict problem of soil nutrients loss.

2 Structure of BP model design

2.1 Layer number of BP network model

There are two layers, which are input layer and output layer, in ANN at least. This simple model is rarely applied, because it is not useful to solve XOR (eXclusive OR) problem. On the basis of experiences [2], adopting two or more middle layer is no more benefit for solving problem. The more middle layer, the more time is taken for BP model to be excised. So, only one middle layer is designed to establish the BP network model in this paper.

2.2 Middle layer unit number of BP network model [2]

To determinate the number of middle unit is the key to design a successful network. If numbers is very few, the complex problem is difficultly solved. On the contrary, if it is much more, the time excised will drastically increase, and exceeding excise may be caused. In this condition, network will storage verbose information, and then it doesn't recognize true information easily. The geometrical average rule is adopted to determinate the number of middle layer unit in this paper. Let the number of input layer unit is *n*, here *n*=3,and let the number of output layer unit is *m*, here *m*=3.

So, the number of middle layer unit, *l*, is calculated as follows:

$$l = \sqrt{m \cdot n} = \sqrt{3 \times 3} = 3 \tag{1}$$

2.3 Topological structure of BP networks model

Three layers topological structure in this paper appears in Figure 1.

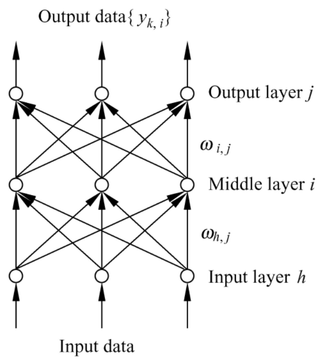

Fig. 1 The topological structure schematic diagram of BP network

3 Algorithms on BP network model [3]

The concrete stage is as follows:

Step 1：data initialization.。First, input data and output data are processed into unitary data.

Assuming input and output unitary data are respectively x_{kh} and d_{kj}, here k=1, 2, …, *nk*, and *nk* is the sample volume; h=1,2,3, and h represents the number of input units; j=1, 2, 3, and j represents the number of input units. Moreover, assign initial values to conjunction weights and threshold value $\{w_{hi}\}$、$\{w_{ij}\}$、$\{\theta_i\}$、$\{\theta_j\}$,which are limited in (–1,1), here i=1,2,3 ,and it represents the number of middle units. At the same time, assign initial value to learning rate η, here $\eta \in (0, 1)$.

Step 2: k=1. Start excising network.

Step 3: compute input and output of each middle unit.

$$x_i = \sum_{k=1}^{3} w_{hi} \cdot x_{kh} + \theta_i \tag{2}$$

$$y_i = 1/(1+\mathrm{e}^{-x_i}) \qquad i = 1,2,3 \tag{3}$$

Step 4: compute input and output of each middle unit

$$x_j = \sum_{i=1}^{3} w_{ij} \cdot y_i + \theta_j \tag{4}$$

$$y_j = 1/(1+e^{-x_j}) \qquad j = 1,2,3 \tag{5}$$

Step 5: compute the partial different of the single sample error E_k to each output unit.

$$E_k = \frac{1}{2}\sum_{j=1}^{3}(y_j - d_{kj})^2 \tag{6}$$

$$\frac{\partial E_k}{\partial x_j} = \frac{\partial E_k}{\partial y_j}\frac{\partial y_j}{\partial x_j} = (y_j - d_{kj})y_j(1-y_j) \qquad j = 1,2,3 \tag{7}$$

Step 6: compute the partial different of the single sample error E_k to each output unit.

$$\frac{\partial E_k}{\partial x_i} = \frac{\partial E_k}{\partial y_i}\frac{\partial y_i}{\partial x_i} = y_i(1-y_i)\sum_{j=1}^{3}\left(\frac{\partial E_k}{\partial x_j}\frac{\partial x_j}{\partial x_i}\right) = y_i(1-y_i)\sum_{j=1}^{3}\left(\frac{\partial E_k}{\partial x_j}\cdot w_{i,j}\right),\ i = 1,2,3 \tag{8}$$

Step 7: improve value of each conjunction weight and threshold. In the four equations: h,i,j=1,2,3, y_h=x_{kh}, and t represent modifying times.

$$w_{ij}^{t+1} = w_{ij}^{t} - \eta\frac{\partial E_k}{\partial x_j} \tag{9}$$

$$\theta_j^{t+1} = \theta_j^{t} - \eta\frac{\partial E_k}{\partial x_j} \tag{10}$$

$$w_{hi}^{t+1} = w_{hi}^{t} - \eta\frac{\partial E_k}{\partial x_i} \cdot y_h \tag{11}$$

$$\theta_i^{t+1} = \theta_i^{t} - \eta\frac{\partial E_k}{\partial x_i} \tag{12}$$

Step 8：k=k+1，turn to Step3，until all samples are finished.
Step 9：repeat Step2—Step8 until：

$$E \leqslant \varepsilon \tag{13}$$

Thus, the train is over. In the above formula, ε is a given error limit.

$$E = \sum_{k=1}^{nk} E_k = \frac{1}{2}\cdot\sum_{k=1}^{nk}\sum_{j=1}^{3}(y_j - d_{kj})^2 \tag{14}$$

Step 10: provide the initialized data to the BP network, and then y_1, y_2 and y_3 can be obtained by computing Step 1 to Step 4.The final results will be figured out after y_1, y_2 and y_3 being return to original values.

4 Application of the model

Adopting the data of a hydrological station in 19 years (Table 1), Guo Rui had ever studied the loss of a basin in western north of China [4], and he found that water and soil loss was serious from the first year to the 9th year, by fuzzy classification analysis and gray system theory computing analysis, and that the reason of serious water and soil loss is the unreasonable action of human being.

Table 1 The loss of soil nutrients and precipitation, sand transferred volume and runoff volume of floods

Year No.	The lost efficient nutrients			The efficient precipitation (mm)	Runoff volume of floods (10^4m^3)	Sand transferred volume (10^4t)
	N_2 (10^4t)	P_2O_5 (10^4t)	K_2O (10^4t)			
1	23.91	0.07	593.7	461.5	32,007	18,400
2	27.18	0.11	601.0	414.7	25,151	18,333
3	12.09	0.04	450.0	345.6	21,473	8,969
4	0.15	0.01	116.2	346.0	17,034	8,479
5	0.25	0.01	204.3	308.9	14,458	7,560
6	46.11	0.27	703.9	470.0	51,009	31,597
7	10.77	0.15	130.3	208.1	6,417	2,260
8	44.80	0.26	617.8	547.1	47,967	26,558
9	10.87	0.04	156.5	357.6	10,879	5,655
10	11.05	0.04	435.6	404.6	23,316	9,243
11	7.70	0.01	367.5	354.5	16,466	6,628
12	34.30	0.19	697.5	498.2	49,189	30,560
13	11.89	0.03	417.5	348.1	29,068	11,321
14	4.75	0.01	115.7	302.8	14,977	7,318
15	3.32	0.01	311.6	285.7	14,729	6,589
16	3.50	0.01	274.7	387.0	17,992	8,717
17	1.15	0.00	187.9	237.5	9,540	5,044
18	0.22	0.01	189.0	282.6	10,911	3,167
19	2.76	0.05	460.5	410.0	29,788	14,070

He also found that the loss of water and soil from the first year to the 9th year is lower-grade than that from the 10th year to the 19th year, because of the human being treatment; loss of water and soil is relative with precipitation, sand transferred volume and runoff volume of floods. The multivariate linear regression model of the relation between the three kinds of soil nutrients and precipitation, sand transferred volume and runoff volume of floods is built in literature[4], and the influence to the loss of water and soil caused by the action of human being is computed, and the results is set in Table 2.

Using the standardization method in this paper:

$$x_{lm}^{*}=(x_{lm}-Ex_m)/s_m \tag{15}$$

here x_{lm}^{*} is the standardized value; x_{lm} is the original value; Ex_m is the mean of the m-th index series; s_m is deviation, initiate the data from the first year to the 9th year. Regarding the standardized data as the train data, train the BP network, and provide precipitation, sand transferred volume and runoff volume of floods from the 10th year to the 19th year to the model. Finally, the influence to the loss of water and soil caused by the action of human being is figured out, and the results are set in Table 2.

Table 2 The results of the multivariate linear regression mode and BP network model

No.	Items	Loss volume of soil nutrients (10^4t)		
		N	P	K
1	Real loss volume	80.6	0.3609	3457.5
2	Loss computed by multivariate regression	163.2	0.8697	4245.9
3	Change rate (%)(Responding 2)	50.6	58.5	19.50
4	Loss computed by BP network model	124.3	0.5880	3488.9
5	Change rate (%) (Responding 2)	35.1	38.6	11.9

5 Conclusion

From the above table, we can see that the computed result of BP network is identical to that of multivariate linear regression model. It demonstrates that the BP network model in this paper is reasonable and algorithms are feasible. This paper gives a new way to study the loss of water and soil.

References

[1] Ren Ruo, an, Wang Huiwen. 1997 Multivariate Statistic Data Analysis——Theory, method, examples. Beijing: National Defence Industry Press.,18-32.

[2] Zhang Jixian, Mi Xia. 1996 Neural Network and Its Application in Engineering. Beijing: Mechanism Industry Press., 1-19,30-37.

[3] Jin Juliang, Ding Jing. 2000 Genetic Algorithm and Its Applications to Water Science. Chengdu: Sichuan University Press., 68-83.

[4] Upriver and middle-river management bureau of Yellow River water conversancy commission edit. Practice and research of water and soil preserve in yellow soil plateau. 1995 Zhengzhou: Yellow River Water Conversancy Press., 44-49.