Using the Method Combining PCA with BP Neural Network to Predict Water Demand for Urban Development

Zhanyong Wang
The Research Center for East-West Cooperation in China, The Key Lab of GIScience of the Education Ministry PRC, East China Normal University, Shanghai 200062, China
E-mail: wangzy1026@163.com

Jianhua Xu
The Research Center for East-West Cooperation in China, The Key Lab of GIScience of the Education Ministry PRC, East China Normal University, Shanghai 200062, China
E-mail: jhxu@geo.ecnu.edu.cn

Feng Lu
The Research Center for East-West Cooperation in China, The Key Lab of GIScience of the Education Ministry PRC, East China Normal University, Shanghai 200062, China
E-mail: aprillf@126.com

Yan Zhang
The Research Center for East-West Cooperation in China, The Key Lab of GIScience of the Education Ministry PRC, East China Normal University, Shanghai 200062, China
E-mail: jasminyan@126.com

Abstract

Combining Principal Component Analysis (PCA) with BP Neural Network, the paper has established a model to predict water demand for urban development with a demonstration in Hefei city. The results indicate that the error absolute value of prediction model is less than 0.9 percent with an ideal effect. Viewed from PCA results, the principal factors affecting urban water demand can be summarized up as economic development (first principal component F_1) and population size (second principal component F_2). By model training of BP network with the scores of F_1 and F_2 as inputs and water demand as outputs, we has provided three prediction programs, while we think the medium program is relatively better suitable for guiding Hefei's water resources planning according to a comparative analysis on the balance between water supply and demand.

Key words: *Principal Component Analysis, BP Neural Network, predict, water demand, Hefei*

1. Introduction

As a fundamental and irreplaceable natural resource, water has become the strategic economic resource for urban sustainable development [1-3]. With the rapid population growth and industrial-agricultural productio -n, urban water consumption is increasing drastically year by year resulting into the serious contradiction between water supply and demand [3-4]. So, it's a precondition to correctly understand the impact factors and scientifically predict water demand for urban social-economic sustainable development [2-5].

However, affected by various coupling factors such as natural, economic, social factors and so on [6], the prediction accuracy of water demand has aroused wide attention for a long time, and thereby the nonlinear methods instead of traditional classic methods generate recently just as the representative method of Artificial Neural Networks (ANNs). But, due to the overlapping information among coupling factors, if we use original variables directly to predict water demand, it may cause the unsteady prediction results and it's uneasy to comprehensively understand impact factors [5-7].

In this study, we combine Principal Component Analysis (PCA) with the prevalent BP (Back-Propagation) Neural Network to predict urban water demand, which not only can help to analyze the comprehensive characteristics of impact factors but also can ensure the prediction accuracy and the steady results. With the Research results, the regional social-

978-0-7695-3736-8/09 $25.00 © 2009 IEEE
DOI 10.1109/ICNC.2009.212

IEEE computer society

economic development and water resources planning in Hefei can get a scientific reference.

2. Methods

Based on the representative driving force factors affecting urban water demand, firstly we predict the values of these factors in three standard years (2010, 2020, 2030) using grey prediction, multiple regression, exponential smoothing and so on. Then, we extract a few independent and integrated principal components by PCA. Finally, a BPNN model with the first *m* principal component scores as inputs is established to predict urban water demand in the three standard years.

2.1. Principal component analysis

Principal Component Analysis (PCA), as a method of reducing the complexity of data processing, also can help to fully understand the comprehensive characterist -ics of variables [8]. Suppose there is a geographical sample data matrix with the sample number *n* and the variable number *p*. Firstly, the correlation coefficient $r_{ij}\ (i,j=1,2,...,p)$ between x_i and x_j is calculated. Secondly, the eigenvalue λ_i and the corresponding eigenvector e_i are solved by the equation $|\lambda I - R|=0$. Thirdly, the first *m* ($m\leq p$) principal component F_i corresponding to $\lambda_i\quad i=1,2,...,m)$, whose cumulative contribution rate reaches 85-95%, is selected to meet the analysis demand. Finally, based on the load values of F_i, we get the principal component score Z_i as inputs of BPNN model.

2.2. BP neural network model

BP (Back-Propagation) Neural Network, a more representative and prevalent model among ANNs, is a multi-layer and feed-forward network with the network training based on error back-propagation and weight or threshold adjustment [9]. In this study, we adopt a fully conjunctive three-layer network with input layer, hidden layer and output layer to predict water demand according to the following principal.

(1) Propagation process

Suppose $X_k=(x_1,x_2,...,x_n), k=1,2,...,m$ is the sample vector for inputs with the training sample number *m* and the neuron number of input layer *n*, and $Y_k=(y_1,y_2,...,y_q)$ is the desired output vector corresponding to the input sample while *q* is the neuron number, so the inputs or outputs of nodes in hidden layer can be obtained respectively according to the formula (1) and (2) as follows:

$$S_j=\sum_{i=1}^{n} w_{ij}x_j-\theta_j \qquad (1)$$

$$b_j=f(S_j) \qquad (j=1,2,...,p) \qquad (2)$$

in which, *p* is the neuron number of hidden layer, w_{ij} is the weight linking i^{th} neuron in input layer to j^{th} neuron in hidden layer, θ_j is the threshold of j^{th} neuron in hidden layer, *f* is the excitation function.

While the inputs of nodes in output layer is calculated as the following formula (3):

$$L_t=\sum_{j=1}^{p} v_{jt}b_j-\gamma_t \qquad (3)$$

where L_t is the inputs of t^{th} neuron in output layer, v_{jt} is the weight linking j^{th} neuron in hidden layer to t^{th} neuron in output layer, γ_t is the threshold of t^{th} neuron in output layer. Meanwhile, $C_t=f(L_t)$ is used to calculate the output of t^{th} neuron in output layer.

At the moment, an input mode has completed a positive-sequence propagation process.

In order to reduce the chance of falling into local minimum value to raise the speed of convergence of network model, an improved momentum BP algorithm is adopted to train network, and the training process can be described as the following equation (4),

$$w(k+1)=w(k)+\alpha[(1-\eta)D(k)+\eta D(k-1)] \qquad (4)$$

where $w(k)$ is the single weight or the weight vector, $D(k)=\frac{-\partial E}{\partial w(k)}$ means negative gradient of time, $\alpha(\alpha>0)$ is learning rate, $\eta(0\leq\eta<1)$ is momentum factor.

In fact, momentum item, equivalent to damping item, has reduced the oscillation trend of training process to improve the convergence. Besides, the initial w_{ij}, v_{jt}, θ_j and γ_t are all random values and α can be adjusted following the equality (5)-(7):

$$w(k+1)=w(k)+\alpha(k)D(k) \qquad (5)$$

$$\alpha(k)=2\lambda\alpha(k-1) \qquad (6)$$

$$\lambda=sign[D(k)D(k-1)] \qquad (7)$$

If the gradient change remains the same direction at two consecutive iteration moments, it shows the convergence rate declines too slowly and we should double the step. Conversely, the step should be halved.

(2) Excitation function

Excitation function, including linear, sigmoid, tangent and other types, can transform the input of a neuron in hidden layer or output layer into the output of this neuron with the purpose of simulating the characteristics of biological neurons.

(3) Error function

Suppose the actual output is $y_j(t)$ and the desired output is $d_j(t)$ respectively corresponding to j^{th} neuron at t^{th} moment, then we define error function of network at t^{th} moment as $E(t)$ calculated as follows:

$$E(t)=\frac{1}{2}\sum_{j=1}^{q}(y_j(t)-d_j(t))^2 \tag{8}$$

When $E(t)\le\varepsilon$ (ε is a given error in advance), BP network will stop training and the network model at this time is just what we need.

3. A demonstration

3.1. Study area

Hefei, the capital of Anhui, is located in the central China between Yangtze and Huaihe Rivers and near Chaohu Lake, where enjoys the subtropical humid monsoon climate featuring mild climate. Due to the uneven temporal-spatial distribution of the rainfall, the availability of surface water is limited. Besides, Hefei is a severely depleted district of the groundwater with the average annual water-output of only about 516 millions m^3. Together with the serious pollution in west half Lake of Chaohu, Nanfeihe River, Dianbu River etc, water pressure becomes severe. So, it is an urgent task for scientifically understanding impact factors and designedly predicting water demand in Hefei with the rapid social-economic development.

3.2. Driving force analysis of water demand

In view of the selected impact factors of water demand as follows mutually coupling together, we adopt PCA to eliminate the overlapping and redundant information among the variables, which is helpful to understand their comprehensive characteristics.

According to the actual situation in Hefei, 14 representative driving forces are selected as follows: agricultural population (x_1, ten thousand), non-agricultural population (x_2, ten thousand), per capita GDP (x_3, yuan), primary industrial outputs (x_4, billion yuan), secondary industrial outputs (x_5, billion yuan), tertiary industrial outputs (x_6, billion yuan), fixed investment (x_7, billion yuan), urban per capita annual disposable income (x_8, yuan), rural per capita annual net income (x_9, yuan), common cultivated area (x_{10}, hectare), effective irrigation farmland (x_{11}, hectare), irrigation quota of farmland (x_{12}, m^3 per hectare), water demand of ten thousand industrial added value (x_{13}, m^3 per ten thousand), per capita water demand of urban living (x_{14}, m^3 per capita per day).

Through PCA, cumulative contribution rate of F_1 and F_2 has reached 93.94%, which indicates F_1 and F_2 can well represent the original variable information. Viewed from Table 1, F_1 has strong correlation with x_3, x_4, x_5, x_6, x_8, x_9, x_{10}, x_{11}, x_{12} etc, most of which are related to economic development, while there is great relativity between F_2 and x_1, x_2, x_{14}, which are related to population size. In brief, the impact factors of water demand can be summarized as economic development and population size in Hefei.

Table 1. Load values of F_1 or F_2

	F_1	F_2
x_1	-0.1792	-0.9795
x_2	0.0641	0.9938
x_3	0.9895	-0.0526
x_4	0.9936	0.0601
x_5	0.9924	0.0943
x_6	0.9905	0.1018
x_7	0.7544	0.3216
x_8	0.9921	0.0765
x_9	0.9904	0.0936
x_{10}	-0.9314	0.2561
x_{11}	-0.9537	0.2295
x_{12}	-0.9261	0.2294
x_{13}	-0.7203	0.6698
x_{14}	-0.3730	-0.8035

(1) Economic development

In Hefei, GDP increased from 1.415 billion yuan in 1980 to 107.376 billion yuan in 2006 with the main growth in secondary and tertiary industries, and the average annual growth rate ranked first among the capital cities in China. Also, fixed investment grew from 8.137 billion yuan in 1997 to 82.48 billion yuan in 2006 with a nearly 10 times increase. In recent 30 years, per capita GDP, urban per capita annual disposable income as well as rural per capita annual net income increased with the average annual growth rate of 199.4, 98.5 or 102.1 percent respectively. Just because of this, a large amount of water is demanded.

(2) Population size

The total population in Hefei grew from 3,261,100 in 1980 to 4,698,500 in 2006, while non-agricultural population grew continuously from 633,300 in 1980 to 1,961,600 in 2006 with the average annual growth rate of 8.07 percent and the growth trend still can last for a long time in future. Meanwhile, more and more floatin -g population also will be attracted by the strengthened capital status under the rising strategy in central region of China.

According to the 11th five-year planning for social-economic development and Urban Master Planning

(2006-2020) in Hefei, the agricultural and industrial development is still the focus of economic growth in future. Data shows that the repetition rate of integrated water in industry only reaches about 0.45 and water consumption of GDP per ten thousand yuan is higher than the average value of 42 national major cities. Furthermore, a low water utilization coefficient of agricultural irrigation between 0.4 and 0.5 is only about half of the developed countries (0.7-0.8), which is caused by water leakage of the aging or out-dated equipments. The present situation in Hefei explains the water-saving capacity is suboptimal but has a great potential. However, per capita water demand of urban living increased from 0.113 m^3 per capita per day in 1980 to 0.214 m^3 per capita per day in 2003 with the average annual growth rate of 3.89 percent beyond the assessment criteria of water-saving cities in China. In addition, the booming service trades also drive the rapid rise in water demand of public facilities.

3.3. BP network modeling and accuracy test

Based on experimenting repeatedly, if we use 14 impact factors directly as inputs of BP network model without correlation eliminating, training model is easy to fall into local minimum and the prediction results also isn't ideal, while it reaches ideal effect with Z_1 and Z_2 by PCA as inputs and the convergence rate of model speeds up.

So, with the inputs of Z_1 and Z_2 and outputs of water demand corresponding to 1980,1985,1990,1995 and from 1997 to 2006, the main model parameters are: initial weight between -0.3 to 0.3, initial value of learning coefficient and momentum factor being 0.01 and 0.02 respectively, 4 nodes in a single hidden layer. In order to eliminate the magnitude influence and escape local minimum in the training process, the normalized standardization is carried out firstly among inputs or outputs. Then, BP convergence model is obtained by training. Meanwhile, it's to test the reliability of prediction model that we used sample data again to predict water demand compared with the historical water demand listed in Table 2.

Table 2. Accuracy test of BP network model

	Historical value (billion m^3)	Simulation value (billion m^3)	Relative error (%)
1980	1.070	1.065	-0.4673
1985	1.270	1.280	0.7874
1990	1.408	1.398	-0.7102
1995	1.668	1.673	0.2998
1997	1.498	1.485	-0.8678
1998	1.523	1.522	-0.0657
1999	1.559	1.552	-0.4490
2000	1.529	1.542	0.8502
2001	1.716	1.724	0.4662
2002	1.497	1.503	0.4008
2003	1.480	1.478	-0.1351
2004	1.567	1.561	-0.3829
2005	1.733	1.745	0.6924
2006	1.800	1.803	0.1667

Seen from Table 2, the fitting average error is little about ±0.5234 percent, while the error absolute value is less than 0.9 percent, which proves that using BP neural network technology to establish water demand prediction model is practicable.

3.4. Results and analysis

Referring to the 11th five-year planning for social-economic development, Urban Master Planning (2006-2020), Water Planning Achievement in Hefei and so on, the paper firstly predicted the driving force factors in 2010, 2020 or 2030 under the high, medium or low program. Then, Z_1 and Z_2 of three standard years acted as inputs of prediction model and we obtained three programs of water demand in three standard years in Hefei showed in Table 3.

Table 3. Water supply and demand in Hefei[1]

	Program	2010	2020	2030	
				Local water supply	Water transfer project
Water supply		2.160	2.497	2.879	3.484
Water demand	High	2.393	3.164	3.389	3.389
	Medium	2.139	2.743	2.968	2.968
	Low	2.056	2.528	2.651	2.651
Water deficit	High	0.233	0.667	0.510	-0.095
	Medium	-0.021	0.246	0.089	-0.516
	Low	-0.104	0.031	-0.228	-0.833

Viewed from water supply and demand in Table3, we conclude a summary of the results. Water demand under the high, medium or low program shows a minor change between 2010 and 2020 but a great change between 2020 and 2030, which is due to the gradually strengthened water saving considered in the prediction. Besides, water demand under three programs is greater than water supply respectively in 2020, explaining the contradiction between water supply and demand is

1. ①The unit of water supply and demand is billion m^3. ②Water supply is derived according to the water integrated planning results provided by Hefei Water Authority. Referring to other references, we considered two scenarios just as local water supply and water transfer project from Yangtze River in 2030.

more serious in this year. In particular, water demand under the low program has a minor change over time and performs less than water supply in 2010 or 2030 with little water pressure. If we consider an inter-basin water transfer project in 2030, water demand under the low program is about 0.833 billion m^3 less than water supply, which tells the carrying potential of water reso-urces is enormous for social-economic development in Hefei under this program in 2030. In contrast, the contradiction between water supply and demand under the high or medium program performs obviously in three standard years, while the target of the balance between water supply and demand can be realized only thinking of inter-basin water transfer project in 2030.

According to the comparative analysis above, we think that the high program belongs to an extensive way of social-economic development, not suitable for Hefei with so severe water shortage, while the low program provides an ideal and efficient water-saving mode and only can be as a long-term development goal in Hefei. However, considering the planning goal adequately and referring to the development stages of advanced water-saving examples at home or abroad as well as the assessment criteria of water-saving cities in China, the practicable values of impact factors of water demand has been determined in the medium program, the prediction results of water demand in which is relatively more suitable as a scientific basis of water planning in Hefei.

4. Conclusions

(1) It's effective to comprehensively understand the driving forces of urban water demand by PCA, which also plays a good role in noise filtering for the inputs of BP network model and can speed up the model convergence. Also, Experiment has proved that this has overcome the negative influence of coupling varia-bles on convergence rate and model accuracy, and the nonlinear behavior simulation of water demand is ideal with an error absolute value less than 0.9 percent.

(2) In the empirical study, the impact factors of water demand can be summarized up as economic development and population size, and especially, the rapid economic growth acts as the uppermost driving force. According to the analysis of supply-demand balance in water resources, the contradiction between water supply and demand is relatively prominent, which can be relieved gradually through the way of efficient water saving or inter-basin water transfer. Based on the prediction programs of water demand and combining with urban development trend in Hefei, we think that the medium program is relatively better suitable for guiding its regional socioeconomic development and water resources planning.

Acknowledgement

The study is supported by Shanghai Key Laboratory of Urbanization and Ecological Restoration, and Natio-nal Natural Science Foundation of China (J0730534).

5. References

[1] Keeney R.L., Wood E.F., An illustrative example of the use of multiattribute utility theory for water resources planning, Water Resources, Nterperiodica, Birmingham, 1997, vol. 13, pp. 705-712.

[2] Falkenmark M, Lundqvist J, Towards water security: political determination and human adaptation crucial, Natural Resources Forum, Blackwell, Oxon, 1998, vol. 21, pp. 37-51.

[3] Feng Qi, Cheng Guodong, et al., Water resources in China: Problems and counter measures, AMBIO, Royal Swe-dish Acad Sciences, Stockholm, 1999, vol. 28, pp. 202-203.

[4] Cheng Guodong, Evolution of the Concept of Carrying Capacity and the Analysis Framework of Water Resources Carrying Capacity in Northwest of China, Journal of Glaciology and Geocryology, Science Press, Beijing, 2002, vol. 24, pp. 361-367.

[5] Shi Yafeng, Qu Yaoguang, Water resources carrying capacity and rational development and utilization of Urumqi River, Science Press, Beijing, 1992.

[6] Maier H.R., Dandy G.C., Neural network for the prediction and forecasting of water resources variables: a review of modeling issues and applications, Environmental Modeling & Software, Elsevier, Amsterdam, 1999, vol. 15, pp. 101-123.

[7] Xu Zhongmin, Cheng Guodong, The Predicted Demand of Water Resources in the Middle Reaches of Heihe River from 1995 to 2050, Journal of Glaciology and Geocryology, Science Press, Beijing, 2000, vol. 22, pp. 139-146.

[8] Xu Jianhua, Mathematical Methods in Contemporary Geography, Higher Education Press, Beijing, 2002.

[9] Wen Xin, Zhou Lu, et al., Application of neural network with MATLAB, Science Press, Beijing, 2001.