Prediction of Daily Water Consumption of Residential Communities Based on Correlation Analysis and BP Neural Network

Longlong Liu¹,², Wenzhu Li³ and Xin Liu*¹
¹ School of Information and Electrical Engineering, Hebei University of Engineering, Handan, Hebei, 056038, China
² Handan River Affairs Bureau, Handan, Hebei, 056001, China
³ Institute of Science and Technology, Hebei University of Engineering, Handan, Hebei, 056038, China
*Corresponding author’s e-mail: xin.liu@hebeu.edu.cn

Abstract. In recent years, with the rapid development of cities, water resources are becoming increasingly scarce. Reasonable prediction of community water use is an important basis for regional water resources allocation, effective management of water resources and water conservation. The prediction model of BP neural network based on correlation analysis is established. Firstly, the influence factors of water consumption are analyzed by the correlation analysis theory, the influence size is sorted, and then the partial autocorrelation theory is used to analyze the daily water flow sequence, and the optimal delay time is determined, and then the input variables are determined. Then the BP neural network model is used to predict the water consumption of the community. Finally, the BP neural network model based on the correlation analysis is compared with the other models. The results show that the average error of forecasting model for residential community water consumption is 2.21% and the maximum relative error is 6.87%. Compared with the BP neural network model without correlation analysis and the least squares support vector machine model, the error is the smallest.

1. Introduction
The spatial and temporal distribution of water resources in China is extremely uneven, and the shortage of water resources and the deterioration of water environment seriously affect the sustainable development of the economy and society[1]. Community water use is an important part of water consumption. Reasonable community water use forecast is an important basis for regional water resources allocation, effective water resources management and water conservation.

In recent years, with the rapid development of science and technology, some new technologies and new theories are gradually applied to water demand prediction, and the most commonly used is BP neural network. BP neural network has the characteristics of high-scale parallel processing efficiency, strong learning ability and global approximation, which can be well adapted to the requirements of urban short-term water consumption forecast[2-3]. Yan Xu et al. used BP neural network to predict urban water consumption and achieved good prediction results[4-6]. However, in the modeling process, the selection of input variables is based on the historical water change law, relying on the subjective experience of technicians, lacking of theoretical basis, and affecting the accuracy of prediction results.
Moreover, when the water consumption is predicted, it is strictly in accordance with the time series, and the data are strict and the generalization ability is poor.

In view of these problems, this paper improves on the basis of BP neural network, and establishes BP neural network community water consumption prediction model based on correlation analysis. Firstly, the relevant analysis theory is used to analyze the influencing factors of water consumption, and then sorted according to the influence size. Then, the partial water consumption series is used to analyze the daily water consumption sequence to determine the optimal delay time, and then the input variables are determined. Then, when the data are input, the model randomly disturbs the data sequence and improves the generalization ability. Finally, in order to verify the accuracy of the BP neural network model based on correlation analysis established in this paper, the BP neural network model and the least squares support vector machine model without correlation analysis are compared and compared. The results show that the method has strong generalization ability and high prediction accuracy, which provides a reliable basis for urban water dispatching in the future.

2. Method

In the daily water consumption forecast of the residents' community, the input factors are too small, which will affect the accuracy of the prediction results. The input factors are too many, and may increase the computational complexity and may fall into local optimum. In order to obtain better prediction results, it is important to select the influencing factors reasonably and determine the input variables.

2.1. Correlation analysis of factors affecting daily water consumption

In order to determine the impact of various influencing factors such as weather conditions and holidays on the water consumption of residents, the correlation coefficient test method is used for correlation analysis. To check whether the correlation between daily water consumption and influencing factors of residents is significant, it is to examine the correlation coefficient. According to the relevant analysis theory\(^7\), the calculation formula of the correlation coefficient \(r\) between residents' daily water consumption and influencing factors is as follows:

\[
 r = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i \cdot y_i - \bar{x} \cdot \bar{y})}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2\right) \cdot \left(\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2\right)}}
\]

In the above formula: \(r\) represents the daily water consumption of residents, \(X\) represents the influencing factor; \(x_i\) and \(y_i\) are the influencing factors and the value of the i-day of the resident community water consumption respectively; \(\bar{x}\) and \(\bar{y}\) are the average values of \(X\) and \(Y\), respectively, \(i=1, 2, ..., n\).

When \(r>0\), \(Y\) is said to be positively correlated with \(X\); when \(r<0\), \(Y\) is said to be negatively correlated with \(X\). If the absolute value of the correlation coefficient \(r\) is small (close to 0), it indicates that the correlation between \(Y\) and \(X\) is not significant. When the absolute value of the correlation coefficient \(r\) is large (close to 1), it indicates that the correlation between \(Y\) and \(X\) is significant.

2.2. Daily water consumption partial autocorrelation analysis

Under the condition that other influencing factors are constant, the partial correlation analysis theory is used to study the correlations existing in the daily water consumption time series of residents' communities, which truly reflects the correlation between internal water consumption. The partial autocorrelation coefficient is an indicator of the degree of autocorrelation. The water consumption sample set is \(x(t) = x_1, x_2, ..., x_n\), then the partial autocorrelation coefficient is as follows:
In the above formula: $k$ represents the delay time (time interval), $k = 1, 2, \ldots, m$; $\bar{x}$ is the mean; $r_k$ is the partial autocorrelation coefficient, and the value range is $[-1, 1]$.

Generally, there are two methods for determining the optimal delay time: (1) The value of $k$ is the value corresponding to the first zero crossing of $r_k$; (2) If $r_k$ approaches zero when the delay time is large, the optimal delay time is the value corresponding to the first time when $r_k$ is less than $\frac{1}{e}$.[8-10].

2.3. Daily water consumption prediction model based on BP neural network

BP neural network is a multilayer feedforward network with hidden layers[11-14]. Its basic principle is the gradient steepest descent method. Its central idea is to adjust the weight to minimize the total network error, that is, to use gradient search technology, in order to minimize the error mean value of the actual output value and expected output value of the network. The network learning process is a process of correcting the weight coefficient by the error side backward propagation. A schematic diagram of the topology of the three-layer BP network is shown in Figure 1.

![Schematic diagram of three layer BP neural network topology.](image)

From Fig. 1, the threshold value of the $j$-th neuron in the hidden layer can be $\theta_j$, and the threshold of the neurons in the daily water output layer is $\gamma$, and the input of the $j$-th neuron in the hidden layer is as follows:

$$s_j = \sum_{i=1}^{n} W_{ij} x_i - \theta_j$$  \hspace{1cm} (3)$$

In the above formula: $x_i$ is the input value of the $i$-th neuron when the input variable is input; $n$ is the number of neurons in the input layer. The output is $b_j = f(s_j), j = 1, 2, \ldots, p$, $p$ is the number of neurons in the hidden layer, and $f$ is the excitation function. The specific representation is as follows:

$$f(x) = \frac{1}{1 + e^{-x}}$$  \hspace{1cm} (4)$$

The role of the excitation function $f$ is to simulate the nonlinear properties of biological neurons. The input of neurons for daily water output is as follows:

$$L = \sum_{j=1}^{p} v_j b_j - \gamma$$  \hspace{1cm} (5)$$
The daily water consumption output is \( C = f(L) \). This is a complete one-pass process of the input mode. The weight and threshold are adjusted according to the error size, that is, the inverse propagation process of the error. Network training is repeated by pattern repeat propagation and error inverse propagation until the error meets the requirements.

So far, by using the relevant analysis theory to analyze the influencing factors of daily water consumption and using the partial autorelation theory to analyze the daily water consumption time series, a BP neural network prediction model based on correlation analysis is established, which can effectively predict the daily water consumption of residents' communities.

### 3. Study case

Firstly, the prediction process of the BP neural network model based on correlation analysis in the daily water consumption prediction model of the residents community is determined, as shown in Figure 2. Then, the sample data are analyzed to determine the main influencing factors of the daily water consumption of the community, and the input variables are determined. Finally, the input variables are brought into the BP neural network prediction model for prediction and output.

#### 3.1. Data Sources

The data used in this paper is from the online water quality testing platform of Hebei University of Engineering. The actual community water consumption data of the residents' homes from April 1 to July 9, 2016.

#### 3.2. Determination of predictive model input variables

There are many factors affecting the water consumption of residents' communities. This paper mainly considers the maximum temperature, weather conditions, holidays and the amount of water used in the previous day. According to the relevant analysis theory, the correlation between each influencing factor and the water consumption of residents' communities is calculated by using Equation 1. The coefficients are calculated and the results are shown in Table 1.

| Maximum temperature | weather conditions | holiday conditions | water consumption the previous day |
|---------------------|--------------------|-------------------|-----------------------------------|
| 0.200               | -0.196             | 0.182             | 0.891                             |

It can be seen from Table 1 that the influence of each influencing factor on the water consumption of residents is: water consumption the previous day > Maximum temperature > weather condition > holiday conditions, in which the previous day's water consumption, maximum temperature and holiday situation are positively related to the water demand. The weather conditions have a negative
correlation with the demand for water demand. Calculate the daily water consumption partial autocorrelation coefficient using Equation 2. The results are shown in Figure 3.

It can be seen from Fig. 3 that m is less than 0 at n, and there is no problem that m approaches zero when the delay time is large. This paper chooses the first method to determine the optimal delay time of daily water consumption, and the daily water consumption is the most. The optimal delay time is one day. Therefore, the daily water consumption is highly correlated with the previous day's water consumption, and the correlation with other days is weak.

Through the correlation analysis results presented in Table 1 and Figure 3, it is determined that the daily water consumption of the residential community is predicted by the previous day's water consumption, and the BP neural network prediction model is established by using the previous day's water consumption as a model input.

3.3. BP neural network daily water consumption forecast

According to the input variables determined above, in order to make the model more practical, break the strict time series data, divide the obtained 99-day data into two parts, and use any 80-day data as the training set to build a model. The remaining 19 days of data were test sets and the model was tested. Normalize the data before modeling, eliminate the influence of different dimensions and units, and prevent the problem of “decorating large numbers”. Then use MATLAB software to carry out modeling. Firstly, the parameters such as error precision, training times and learning rate of BP neural network model are determined by trial calculation. The evaluation criteria of model fitting and prediction results are relative error ($\delta$) and average relative error (MAPE). Calculated as follows:

$$\delta = \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$$  \hspace{1cm} (6)

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$$  \hspace{1cm} (7)

In the above formula: n is the number of samples in the test set, and $y_i$ and $\hat{y}_i$ are measured values and predicted values.

After repeated adjustment of the parameter calculation, the final optimal model parameters are obtained: the error precision is 0.004, the training frequency is 100, and the learning rate is 0.1. The predicted results are shown in Figure 4. The relative error ($\delta$) is 6.87% and the average relative error (MAPE) is 2.21%.
3.4. Analysis of prediction results of different models

In this paper, BP neural network models (two input BP neural network models) with input variables as the previous day's water consumption and maximum temperature, BP neural network models (three input BP neural network models) with input variables as the previous day's water consumption, maximum temperature and weather conditions, BP neural network model (four input BP neural network models) with input variables as the previous day's water consumption, maximum temperature, weather conditions and holidays, and the least squares support vector machine model with input variables for the previous day's water consumption (LSSVM models) were established. The sample data are trained and predicted while maintaining other conditions. The prediction results are compared with the BP neural network model (one input BP neural network model) whose input variables are the previous day's water consumption, as shown in Figure 5 and Figure 6. The prediction results are evaluated as shown in Figure 7 and Table 2.

![Figure 6. Comparison of BP neural network model and LSSVM model prediction result.](image)

![Figure 7. Comparison of relative error results of different models.](image)

| Evaluation index         | Predictive model                          |
|--------------------------|-------------------------------------------|
|                          | one input BP neural network model          |
| Average relative error (%)| 2.21                                      |
| Maximum relative error (%)| 6.87                                      |
|                          | two input BP neural network models         |
| Average relative error (%)| 4.72                                      |
| Maximum relative error (%)| 12.77                                     |
|                          | three input BP neural network models       |
| Average relative error (%)| 4.76                                      |
| Maximum relative error (%)| 10.78                                     |
|                          | four input BP neural network models        |
| Average relative error (%)| 3.40                                      |
| Maximum relative error (%)| 9.44                                      |
|                          | LSSVM models                              |
| Average relative error (%)| 18.7                                      |
| Maximum relative error (%)| 45.92                                     |

It can be seen from Fig. 4, Fig. 5, Fig. 6, Fig. 7 and Table 2 that the errors obtained by different input variables of the same prediction method are different, and the errors obtained by different prediction methods of the same input variables are also different. In this paper, the input variables are determined based on the correlation analysis theory, and the trend of the predicted values predicted by the BP neural network model is basically the same as the measured value, and the values are very close, the error is small, and the accuracy is high.

4. Conclusion

Resident community water consumption forecast is an important basis for regional water resources allocation, effective management of water resources and water conservation. Aiming at the problem that the input variables are not well determined in the current daily water consumption forecasting process, relying on subjective experience and lack of theoretical basis, this paper establishes a BP neural network prediction model based on correlation analysis. Taking the family colleges of Hebei University of Engineering as an example, the relevant factors were analyzed to determine the water consumption of the residents in the previous day as the model input, and the daily water consumption...
was the model output. The data of the day are modeled and predicted, and the predicted results are compared with other models.

The results show that the input variables based on the correlation analysis theory are determined, and the trend of the predicted values and the measured values are basically consistent with the BP neural network model, and the values are very close. The average error is 2.21%, and the maximum relative error is 6.87%. The error is minimal compared to other models.

In general, the BP neural network model based on correlation analysis proposed in this paper is suitable for the analysis and prediction of water consumption in residential communities, and the prediction effect is good. It provides a new idea for accurately predicting the water consumption of residents' communities, and provides a theoretical basis for the urban water resource dispatching.

Acknowledgments
Thanks to the relevant data provided by the Hebei University of Engineering Water Saving Platform. Thanks to the instructor’s careful guidance and thanks to the brothers and sisters of the same family for their struggle. I am grateful to the Education Department of Hebei Province for supporting the graduate students' innovative ability training projects.

References
[1] Wang H, You J J. (2016) China's water resources allocation for 30 years. Journal of Hydraulic Engineering, 47(3): 265-271.
[2] Cheng G P, Qi X H. (2011) Evaluation of Private Equity Investment Project Based on BP Neural Network. Journal of Tsinghua University (Science and Technology), 51(11): 1917-1920.
[3] Qi Cheng, Chang N B. (2011) System dynamics modeling for municipal water demand estimation in an urban region uncertain economic impacts. Journal of Environmental Management, 92(6): 1628-1641.
[4] Yan X, Li S Y, Zhang Z. (2016) Application of BP Neural Network Based on Genetic Algorithm in Urban Water Consumption Forecasting. Computer Science, 43(11A): 547-550.
[5] Chu C S, Zhang H W, Guo J. (2006) Water Consumption Forecast Based on Genetic Algorithm and BP Neural Network. China Rural Water and Hydropower, 36-38.
[6] Zhou Y C, Li S P, Zhao Z W, et al. (2015) Short-term water consumption forecast based on BP neural network toolbox. Water and Wastewater, 41: 375-377.
[7] Hu X, Sun H X, Wang L X. (2011) Probability theory. Mathematical statistics. Stochastic Process. Beijing University of Posts and Telecommunications Press, Beijing.
[8] Holzfluss J, Mayer-Kres G. (1986) An approach to error estimation in application of dimensional girths. Dimensions and Entropies in Chaotic Systems, 114-122.
[9] Ding J, Wang W S, Zhao Y L. (2003) Study on Chaos Variation Characteristics of Daily Flow in the Yangtze River--Identification of I-phase Space Embedding Delay. Advances in Water Science, 14(4): 407-411.
[10] Bai Y. (2014) Research and application of water supply quantity prediction method driven by time series characteristics. Chongqing University.
[11] Dong C H. (2003) Matlab neural network and application [M]. National Defense Industry Press, Beijing.
[12] Zhang H W, Niu Z G. (2003) Research on Establishing Macro Model of Urban Water Supply Pipeline Network by Neural Network Method[J]. Systems Engineering Theory and Practice, 23(10): 121-126.
[13] Huang Y H. (2013) Application of intelligent algorithm based on neural network in coke quality prediction. Shenyang University of Technology.
[14] He C H, Huang Q Y, Shen S F, et al. (2011) Estimation of forest combustible load based on BP neural network. Journal of Tsinghua University (Natural Science), 51(2): 230-233.