Bayesian regularized NAR neural network based short-term prediction method of water consumption

Jianyu Liu¹, Linxue Zhao², and Yanlong Mao¹

¹Geological Engineering and Surveying Institute, Chang'an University, 710054 Xi'an, China
²Packaging Engineering and Digital Media Technology, Xi'an University of Technology, 710048 Xi'an, China

Abstract. With the continuous construction of urban water supply infrastructure, it is extremely urgent to change the management mode of water supply from traditional manual experience to modern and efficient means. The water consumption forecast is the premise of water supply scheduling, and its accuracy also directly affects the effectiveness of water supply scheduling. This paper analyzes the regularity of water consumption time series, establishes a short-term water consumption prediction model based on Bayesian regularized NAR neural network, and compares and evaluates the prediction effect of the model. The verification results show that the Bayesian based NAR neural network prediction model has higher adaptability to the water consumption prediction than the standard BP neural network and the Bayesian regularized BP neural network. The prediction accuracy can more accurately reflect the short-term variation of water consumption.

1 Introduction

The ultimate goal of water consumption forecasting is to obtain the expected value of the next hour's water volume, providing basic data support for pipe network pressure optimization and pump station scheduling. The urban water consumption is affected by many random factors, and the change is more complicated. However, it can be concluded that the urban water consumption curve is a small cycle of 24 hours a day, 7 days a week is a medium cycle, and 365 days a year is a major cycle. This paper mainly discusses the short-term water consumption forecast, which is less affected by seasonal factors and population growth factors, the change law has two important characteristics: randomness and periodicity, which conforms to the law of "time series". The commonly used forecasting methods for water consumption include time series method, grey prediction method, artificial neural network, etc. At present, the traditional statistical-based prediction method has been gradually replaced by artificial neural network, and the deep learning model has gradually become the most widely studied research method. A deep neural network model with multiple nonlinear mapping levels can abstract the input signal layer by layer and extract features to dig deeper potential laws. In this paper, Bayesian regularization algorithm and nonlinear autoregressive (NAR) model are added to BP neural network, and the short-term prediction model of water consumption is established. The experimental results are compared and analyzed.

2 Theoretical analysis

2.1 Overview of BP neural network

The BP (back propagation) neural network is a concept proposed by scientists led by Rumelhart and McClelland in 1986. It is a multi-layer feedforward neural network trained according to the error back propagation algorithm. It is the most widely used neural network. The basic formula for BP neural network calculation is shown in Equation 1:

\[
\text{net}_j = \sum_{i} W_{ij} O_i \\
O_j = f(\text{net}_j) \\
f(x) = \frac{1}{1 + e^{-x}}
\]

In the formula: \(\text{net}_j\) - the sum of the inputs of unit j; \(W_{ij}\) - the weight between neurons i, j; \(O_j\) - the input of unit i; \(f\) - function function; \(N\) - the number of processing units; \(P\) - the number of training samples; \(j\) - number of neural network units.

2.2 Bayesian regularization overview

Regularization is a training method that further enhances the generalization ability by adding a penalty term to the training performance function of the neural network to simplify the weight connection structure of the network. The training performance function after regularization of the penalty is shown in Equation 2:
\[ F = \beta E_\gamma + \alpha \gamma \]
\[ E_\gamma = \frac{1}{M} \sum_{i=1}^{M} \alpha_i^2 \]  
(2)

In the formula: \( E_\gamma \) - the sum of the squares of the network weights; \( \alpha_i \) - neural network connection weights; \( M \) - the number of connection parameters; \( \alpha, \beta \) - the parameters of the objective function.

Bayesian regularization method can adaptively adjust the size of super parameters in the process of network training, so as to achieve the optimal value. According to bayesian criterion, the parameter \( \alpha \) and \( \beta \) values at the minimum point of objective function \( E \) are solved by posterior probability, and formula 3 is deduced.

\[ \alpha = \frac{\gamma}{2E_\gamma} \]
\[ \beta = \frac{E - \gamma}{2E_\gamma} \]
\[ \gamma = \frac{n}{2} \alpha \cdot \text{tr}(H)^{-1} \]
\[ H = f^2 E_\gamma + \alpha \gamma^2 E_\gamma \]  
(3)

In the formula: \( \gamma \) - the effective weight of the sample network parameters; \( H \) - the Hessian matrix of the objective function. Bayesian neural network training can adaptively adjust the size of \( \gamma \) and make it optimal.

### 2.3 Overview of NAR Model

The NAR model is a special form of the NARMA (Nonlinear auto-regressive Moving Average) model. NARMA model can be defined as formula 4:

\[ y_n = \beta(y_{n-1}, \ldots, y_{n-k}, x_{n-k}, \ldots, x_{n-1}) \]  
(4)

In the formula: \( x \) - model input; \( y \) - model output; \( n \) - discrete time; \( G \) - a nonlinear function.

When the measured data is a time series and the model input is unknown, Equation 3 can be rewritten to the NAR model form shown in Equation 5.

\[ y_n = f(y_{n-1}, \ldots, y_{n-k}) + ke_n \]  
(5)

In the formula, \( F \) - nonlinear function; \( k \) - constant; \( e_n \) - a random variable subject to Gaussian allocation. In the NAR model, the output of each \( y \) is passed to the input of the meta-neural network layer, which acts as an adjustment parameter to influence the next output, thereby realizing the adjustment of the neural network. Therefore, the NAR model can reflect the historical state of the system and is a dynamic, memory-enabled neural network.

### 3 Short-term prediction model of water consumption

#### 3.1 Data acquisition and preprocessing

The data is from water consumption monitoring data of a county town’s water supply company in Yulin, from 2019-04-01 00:00 to 2019-04-16 00:00. A total of 385 training data were divided into 269 training sets, 58 verification sets, and 58 test sets according to the ratio of 70%, 15%, and 15%.

First, the historical data is normalized, and the min-max standardization method is used to linearly transform the water consumption data, so that the result value is mapped between \([0, 1]\), and the conversion function is as the formula 6.

\[ X' = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \]  
(6)

In the formula, \( X' \) is normalized data; \( X \) is raw data; \( X_{\text{min}} \) is the minimum value in all historical data; \( X_{\text{max}} \) is the maximum value in all historical data.

Table 1 below is the partially normalized flow data.

| Tab.1 Normalized water flow data |
|--------------------------------|
| Before normalization | After normalization | Before normalization | After normalization |
|-----------------------|---------------------|----------------------|---------------------|
| 96                    | 0.1518              | 105                  | 0.1920              |
| 96                    | 0.1518              | 206                  | 0.6429              |
| 94                    | 0.1429              | 271                  | 0.9330              |
| 91                    | 0.1295              | 252                  | 0.8482              |
| 92                    | 0.1339              | 212                  | 0.6696              |

#### 3.2 Network parameters

The Bayesian regularized NAR neural network is mainly composed of the input layer, the hidden layer, the output layer and the number of lag variables. Since both the input variable and the output variable are water consumption, the number of input layers and output layers is 1. The output delay of the neural network is set to 10 according to the rate of change and stability of the water consumption. The selection of the number of neurons in the hidden layer is determined by many factors. At present, there is no mature theoretical basis. Therefore, the trial and error method is used to establish the model of the hidden layer neurons from 5 to 40 respectively. The weight and threshold affect the performance change of the neural network, so each model is trained 5 times, and the MSE (mean square error) of the training result is recorded separately. MSE reflects the prediction accuracy throughout the entire prediction period, as shown in Equation 7.

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]  
(7)

In the formula, \( n \) is the predicted point; \( y_i \) is the true value; \( \hat{y}_i \) is the predicted value result.

The MSE performance of the NAR neural network model training under different hidden layer nodes is shown in Figure 1.
It can be seen from the figure that the MSE is lower when the number of neurons in the hidden layer is 20, and the number of neurons in the hidden layer is too large, which increases the number of iterations and training time of the model, resulting in over-fitting of the model. The ability to reduce the model leads to a decline in the predictive power of the model. When the number of neurons in the hidden layer is too small, the model will have problems such as insufficient fitting, poor fault tolerance, and low ability to identify new samples, which is consistent with previous theoretical studies[1].

Therefore, this study selects the number of neurons in the hidden layer as 20, and the empirical formula for selecting the number of neurons in the hidden layer is basically consistent.[2]

\[ n_1 = \sqrt{n + m + a} \]  

(8)

\( n_1 \) is the number of neurons in the hidden layer; \( n \) is the initial sample data; \( m \) is the number of output neurons; \( a \) is a constant between [1, 10].

### 3.3 Network training

The Bayesian regularization algorithm is used to train the NAR neural network, which can avoid the problem of the local optimality caused by the random acquisition of the initial weight of the BP neural network[3]. The weighting of the network weight mean square and the error mean square is taken as the performance function. Under the premise of ensuring that the network mean square error is as small as possible, the complexity of the network is effectively controlled, so that better generalization performance is improved and the network promotion capability is improved. The modeling sample is divided into 269 training samples, 58 verification samples, 58 test samples, the transfer function is set to the tansig function, the training function is set to the trainbr function, the weight adaptive learning function is set to the leannmd function, and the performance function uses MSE. The network was trained 434 times and the mean square error was 0.00147. After the training, a short-term prediction model of water consumption time series based on NAR neural network was established, the structure of network shown in Fig. 2.

### 3.4 Analysis of results

This paper uses MSE and R Squared to evaluate the training effect of neural networks.

MSE refers to the square root of the average of the sum of the squares of the errors of the measurements. R Squared represents the degree of fit to the training sample, ranging from 0-1, as shown in Equation 9:

\[ R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \]

(9)

In the formula: \( y_i \) - observations; \( \bar{y} \) - the average of the observations; \( \hat{y} \) - Predictive value.

It can be seen from Fig. 3 and Table 2 that compared with the standard BP neural network and the Bayesian regularized BP neural network, the Bayesian regularized NAR neural network has better MSE and R Squared performance, and has a higher fitness for the data.

### Tab.2 MSE and R Squared performance of different neural networks

|                 | BP neural network | BR neural network | BR NAR neural network |
|-----------------|------------------|------------------|----------------------|
| MSE             | 0.0276           | 0.0187           | 0.000196             |
| R Squared       | 0.79718          | 0.85973          | 0.95618              |
| Number of training | 1000            | 597              | 434                  |

Fig.1 MSE performance of model training under different number of hidden layer nodes

Fig.2 Schematic diagram of short-term prediction of time series of water consumption

Fig.3 Comparison of different neural networks R Squared
4 Model evaluation

To verify the prediction effect of model, we use three well-trained networks to predict the water consumption from 2019-04-16 00:00 to 2019-04-16 06:00.

It can be seen from Table 3 that the Bayesian regularized based NAR neural network has smaller relative error and more accurate prediction than the standard BP neural network and the Bayesian regularized neural network.

The predicted value and the actual value of the three methods were drawn into a broken line graph, where the straight line was the actual value and the "o" line was the predicted value of the Bayesian regularized NAR neural network. It can be seen that the data fitting degree was the highest and the closest to the actual value, which was better than the other methods, as shown in Figure 4.

5 Conclusion

The Bayesian regularized based NAR network can effectively simplify the network complexity under the premise of ensuring the smallest square error of the network error, thereby significantly improving the generalization ability of the neural network without being restricted by the amount of sample data.

The verification results show that when the data set is the same, compared with the standard BP neural network and the Bayesian regularized BP neural network, the Bayesian regularized based NAR network prediction model MSE is only 28.7153, R Squared is 0.94949, the performance is better, so it has higher adaptability and prediction accuracy for short-term prediction of water consumption, and can more accurately reflect the short-term variation of water consumption.

References

1. H.F. Zhang, R. L.Yang, D.S.Zhang.Journal of Jiangxi normal university 34,290-294(2010)
2. X.H.Zhao,Y.Zhang.Journal of Xi’an technological university 37,544-549(2017)
3. Z.F.Zhang,R.N.Yang,Y.H.Fang.Fire control and command control 3,25-29(2018)