Water demand forecasting based on adaptive extreme learning machine

Jinming Jia
School of Economics and Management
Beijing Jiaotong University
Haidian District, Beijing, China, postcode: 100044
Email:jiajinming@sina.com

Shengyue Hao
School of Economics and Management
Beijing Jiaotong University
Haidian District, Beijing, China, postcode: 100044
Email:haoshyue@bjtu.edu.cn

***Abstract*—Predicting water consumption is of key importance for water supply management, which is also relevant in processes for reviewing prices.In this study, a hybrid method based on extreme learning machine model with the adaptive metrics of inputs is proposed for improving forecasting accuracy. The adaptive metrics of inputs can solve the problems of amplitude changing and trend determination, and reduce the effect of the overfitting of networks. It was found that the proposed model is practical for water demand forecasting and outperforms the auto-regression (AR), artificial neural network (ANN), support vector machine(SVM) and extreme learning machine (ELM) models.**

***Index Terms*—Water consumption, Extreme learning machine, Forecasting, Time series**

I. Introduction

An essential component of sustainable water resources planning and management is water demand forecasting. It provides valuable trigger in determining the time and capacity for new water resources development. Efficiently operating and managing a water supply system requires short-term water demand forecasts; and the estimation of future municipal water demand is central to the planning of a regional water-supply system. Water demand forecasting is becoming an essential tool for the design, operation, and management of water-supply systems.

Recently, many methods have been applied extensively for forecasting. Generally, they are divided into two groups: classical methods based on mathematical and statistical models [1] (e.g. exponential smoothing, regression, Box-Jenkins autoregressive integrated moving average ($ARIMA$), generalized autoregressive conditionally heteroskedastic ($GARCH$) methods), and modern heuristic methods using artificial intelligence techniques including artificial neural networks (ANN) and evolutionary computation [2], [3], [4], [5], [6]. The classical methods employ a linear functional form for modeling, and approximation of linear models to complex real-world problems is not always satisfactory [7]. Recently, artificial neural networks (ANN) have been applied extensively for forecasting. The ANN is a universal function approximator which is capable of mapping any linear or non-linear functions [8]. However, most ANN based forecasting methods use gradient-based learning algorithms, which can not avoids many difficulties such as stopping criteria, learning rate, learning epochs, local minima [9], [10]. A relatively novel learning algorithm for single-hidden-layer feedforward neural networks ($SLFN$) called extreme learning machine (ELM) has been proposed in recently [11], [12]. In learning process, the inputweights and hidden biases are randomly chosen, and the output weights are analytically determined by using the Moore-Penrose generalized inverse. ELM can learn much faster with a higher generalization performance than the traditional gradient-based learning algorithms, and solves the problem of stopping criteria, learning rate, learning epochs, local minima.

Since the BP or ELM lacks a systematic procedure for model-building, the forecasting result is not always accurate when the input data is very different from the training data. Furthermore, like other flexible nonlinear estimation methods [13], the ELM may suffer either under-fitting or over-fitting, especially for water demand data. A network that is not sufficiently complex can fail to fully detect the signal in a complicated data set and lead to under-fitting. A network that is too complex may fit not only the signal but also the noise and lead to over-fitting. The network is said to over-fitting relative to a simpler one if it is more accurate in fitting known data (hindsight) but less accurate in predicting new data (foresight). In order to solve this problem, a novel adaptive extreme learning machine ($AD-ELM$) model is proposed in this study with the adaptive metrics of inputs, and the output data is evolved by a mechanism for admixture. The adaptive metrics of inputs of the model can adapt to local variations of trends and amplitudes. Most inputs of the network are close to the historical data in order to avoid a dramatic increase in the forecasting error due to the big difference between training data and input data.

II. Methodology

In this part, we introduce the adaptive extreme learning machine ($AD-ELM$) algorithm for the water demand forecasting. In this work, the data are divided into training and predicting sets. Based on the input and output weights obtained by training data, the water demand can be predicted directly through the established $AD-ELM$.

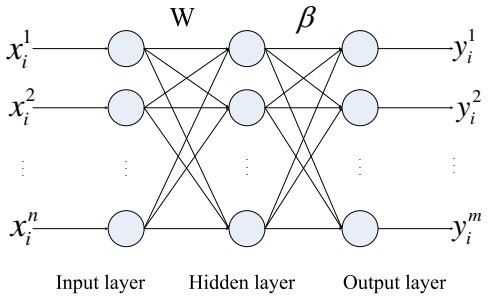

Fig. 1. The structure of ELM model.

A. Extreme learning machine (ELM)

The ELM is a flexible computing framework for a broad range of nonlinear problems [15]. A single hidden-layer feed-forward network ($SLFN$) is the most widely used model for forecasting modeling[16]. As shown in Fig.1, the model is characterized by a network of three layers of simple processing units connected by acyclic links. The hidden layers can capture the nonlinear relationship among variables. Each layer consists of multiple neurons that are connected to neurons in adjacent layers. Suppose there are N distinct samples (X_i, T_i), where $X_i = [\ x_i^1, x_i^2, \cdots, x_i^n\ ]^T \in \mathbb{R}^n$ and $T_i = [\ t_i^1, t_i^2, \cdots, x_i^m\ ]^T \in \mathbb{R}^m$. The $SLFN$ with k hidden neurons and an activation function vector $g(x) = (\ g_1(x), g_1(x), \cdots, g_k(x)\ )$ are described as

$$\sum_{i=1}^{k} \beta_i g_i(W_i \cdot X_j + b_i) = Y_j, \quad j = 1, 2, \ldots, N, \tag{1}$$

where $W_i = [w_{i1}, w_{i2}, \cdots, w_{in}]^T$ is the weight vector connection the ith hidden neuron and the input neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{im}]^T$ is the weight vector connecting the ith hidden neuron and the output neurons, and b_i is the threshold of the ith hidden neuron. The operation $W_i \cdot X_j$ is the inner product of W_i and X_j. If the SLFNs can approximate these N samples with a zero error viz. $\sum_{i=1}^{N} \parallel Y_i - T_i \parallel = 0$. Thus, there also exist parameters W_i, β_i, and b_i such that

$$\sum_{i=1}^{k} \beta_i g_i(W_i \cdot X_j + b_i) = T_j, \quad j = 1, 2, \ldots, N, \tag{2}$$

Thus, above equations can be compactly described as

$$\mathbf{H}\beta = T, \tag{3}$$

where $\mathbf{H}$ is the hidden-layer output matrix, the ith column of $\mathbf{H}$ denotes the ith hidden neuron output with respect to inputs $X_1, X_2, \cdots, X_N$

$$\mathbf{H} = \begin{bmatrix} g_1(W_1 \cdot X_1 + b_1) & \ldots & g_k(W_k \cdot X_1 + b_k) \\ \vdots & \ldots & \vdots \\ g_1(W_1 \cdot X_N + b_1) & \ldots & g_k(W_k \cdot X_N + b_k) \end{bmatrix} \tag{4}$$

Unlike the traditional function approximation theories which require to adjust input weights and hidden layer biases, the input weights and hidden biases are randomly generated. Thus, training an $SLFN$ is simply equivalent to finding a least squares solution $\hat{\beta}$ of the linear function $\mathbf{H}\beta = T$:

$$\parallel \mathbf{H}\hat{\beta} - T \parallel = \min_{\beta} \parallel \mathbf{H}\beta - T \parallel, \tag{5}$$

The smallest norm leastsquares solution of the above linear system is

$$\hat{\beta} = \mathbf{H}^{\dagger} T, \tag{6}$$

where $\mathbf{H}^{\dagger}$ is the Moore-Penrose generalized inverse of matrix $\mathbf{H}$. Owing to the Moore-Penrose generalized inverse, the learning speed are dramatically increased for the network.

B. Adaptive extreme learning machine model for forecasting ($AD-ELM$)

As ELM described above, it can avoid many difficulties faced by gradient-based learning methods. However, like other flexible nonlinear estimation methods, the ELM may suffer either under-fitting or over-fitting [13]. For these two problems, the over-fitting is more important when the signal data is sufficient and the network is sufficiently complex. Over-fitting is especially misleading because it can easily lead to wild prediction far beyond the range of the training data even with the noise-free data. A model which has been over-fitting will generally have poor predictive performance, as it can exaggerate minor fluctuations in the data. In this work, the water demand history data are large and the network is also complex. Thus, in this paper we emphasize on the problem of over-fitting for the ELM. In order to avoid over-fitting, the adaptive extreme learning machine model($AD-ELM$) is proposed, and this algorithm can reduce the chance of over-fitting. In this model, the hindsight data is used to modify the inputs of the ELM in the prediction processing making the inputs approach to the learning data. In this work, the output of the network is only one value, that is the predicted water demand. Thus, we discuss our model only for one output.

Firstly, a strategy is used to initialize the input data $Q_i = [\ q_i^1, q_i^2, \cdots, q_i^n\ ]$. The strategy adopts the adaptive metrics which are similar to the $adaptive\ k-nearest\ neighbor$ method [14]. The data set $Q_i = [\ q_i^1, q_i^2, \cdots, q_i^n\ ]$ is compared with the learned input patterns $X_i = [\ x_i^1, x_i^2, \cdots, x_i^n], i = 1, 2, \ldots, N$. The determination of the closeness measure is the major factor in prediction accuracy. Closeness is usually defined in terms of metric distance on the Euclidean space. The most common choices are the Minkowski metrics:

$$\begin{aligned} L_M(Q_i, X_i) \quad = \quad & \big(\mid q_i^1 - x_i^1 \mid^d + \mid q_i^2 - x_i^2 \mid^d \\ & + \ldots + \mid q_i^n - x_i^n \mid^d \big)^{\frac{1}{d}} \end{aligned} \tag{7}$$

This equation gives the value difference between Q_i and X_i, but the differences of trends and amplitudes are not presented. In time-series forecasting, the information on trends and amplitudes is the crucial factor. In this study, adaptive metrics are introduced to solve this problem and the arithmetic is presented as:

$$L_A(Q_i, X_i) = min_{\lambda_r, u_r} f_r(\lambda_r, u_r), \tag{8}$$

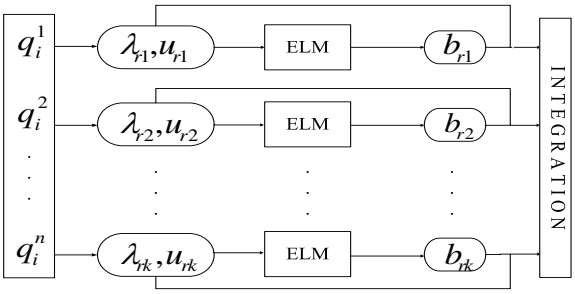

Fig. 2. The forecasting scheme for adaptive extreme learning machine

$$f_r(\lambda_r, u_r) = \big(\mid q_i^1 - \lambda_r x_i^1 - u_r \mid^d + \mid q_i^2 - \lambda_r x_i^2 - u_r \mid^d + \ldots + \mid q_i^n - \lambda_r x_i^n - u_r \mid^d \big)^{\frac{1}{d}} \quad (9)$$

where h_r and l_r are the largest and smallest elements of vector correspondingly, $\lambda_r \in [1, \frac{h_r}{l_r}]$, $u_r \in [0, h_r - l_r]$. The parameter of minimization λ_r equilibrates the amplitude difference between Q_i and X_i.

The parameter u_r is responsible for the trend of data. The optimization problem (8) can be solved by the algorithm of Levenberg-Marquardt optimization or other gradient methods for $d \geq 1$. In this study, d is assumed to be 2 and gives the widely used Euclidean metrics.

$$f_r(\lambda_r, u_r) = \big(\mid q_i^1 - \lambda_r x_i^1 - u_r \mid^2 + \mid q_i^2 - \lambda_r x_i^2 - u_r \mid^2 + \ldots + \mid q_i^n - \lambda_r x_i^n - u_r \mid^2 \big)^{\frac{1}{2}} \quad (10)$$

For $d = 2$, two equations are considered:

$$\begin{cases} \frac{\partial f_r(\lambda_r, u_r)}{\partial \lambda_r} = 0 \\ \frac{\partial f_r(\lambda_r, u_r)}{\partial u_r} = 0 \end{cases}, \quad (11)$$

When the corresponding linear system is solved, the solution of the minimization problem can be obtained analytically:

$$u_r = \frac{z_1 z_2 - z_3 z_4}{n z_2 - z_3^2}, \lambda_r = \frac{n z_4 - z_1 z_3}{n z_2 - z_3^2},$$

where $z_1 = \sum_{i=1}^{n} q_i$, $z_2 = \sum_{i=1}^{n} q_i^2$, $z_3 = \sum_{i=1}^{n} x_i$, $z_4 = \sum_{i=1}^{n} q_i x_i$.

Based on this strategy, the adaptive k-nearest neighbors are chosen. And the input vector of the first network (known as the main network) can be defined as:

$$input_v = \big(\hat{q}_i^{1v}, \hat{q}_i^{2v}, \ldots, \hat{q}_i^{nv}\big) = \left(\frac{q_i^1 - u_{r_v}}{\lambda_{r_v}}, \frac{q_i^2 - u_{r_v}}{\lambda_{r_v}}, \ldots, \frac{q_i^n - u_{r_v}}{\lambda_{r_v}} \right).$$

Most input values can be close to the historical data using this method. The forecasting error increases dramatically due to the big difference between training data and input data. In order to get more accurate results for time series $q_i^1, q_i^2, \ldots, q_i^n$, k sets of inputs are used and the output vector are $output_v = b_v$, $v = 1, 2, \ldots, k$. The mechanism for admixture of outputs is presented as follows:

$$Result = \frac{1}{k-1} \sum_{v=1}^{k} (\lambda_{r_v} b_v + u_{r_v})(1 - \frac{d_v}{U}), \quad (12)$$

$$U = \sum_{v=1}^{k} d_v, \quad (13)$$

where d_v is the distance between Q_i's vth nearest pattern and Q_i. From Equation (12), the forecasting result is calculated from b_v $v = 1, 2, \ldots, k$ with different weighing coefficients, the larger coefficient is given for more closer pattern of Q_i. Based on the methodology proposed above, the forecasting scheme can be formulated as shown in Figure 2.

In the water consumption forecasting, we focus on one-step-ahead point forecasting. Let $y_1, y_2, y_3, \ldots, y_t$ be a time series. At time t for $t \geq 1$, the next value y_{t+1} will be predicted based on the observed realizations of $y_t, y_{t-1}, y_{t-2}, \ldots, y_1$. For ELM model, its solution is usually different from time to time because the input weights and hidden biases are randomly chosen. It is well known that the mean value the forecasting is more reliable. So, a regression integration method is proposed in this paper to obtain higher prediction accuracy. And the final predicted water demand data is only the mean of the s predicted water consumption series $\bar{y}_{t+1} = \frac{1}{s} \sum_{i=1}^{s} \tilde{y}_{t+1}^s$.

III. Numerical simulations studies

In order to compare the $AD - ELM$ method with other methods on all the water demand data series, an adequate error measure method must be selected. The mean squared error ($NMSE$) is used as the error criterion, which is the ratio of the mean squared error to the variance of the time series. It defined, for a time series y_i, by

$$NMSE = \frac{\sum_{i=1}^{M} (y_i - \widetilde{y}_i)^2}{\sum_{i=1}^{M} (y_i - \hat{y}_i)^2} = \frac{\sum_{i=1}^{M} (y_i - \widetilde{y}_i)^2}{M\sigma^2},$$

$$\hat{y}_i = \frac{1}{M} \sum_{i=1}^{M} y_i,$$

where y_i is the source point, $\widetilde{y}_i$ is the predicted point and M is the number of predicted points. σ^2 is the mean value and variance estimated from the source data. A value of $NMSE = 1$ means simply predicting the average. Another measure is the Mean Absolute Percentage Error ($MAPE$). The $MAPE$ is regarded as one of the standard statistical performance measures and takes the following form

$$MAPE = \frac{1}{M} \sum_{i=1}^{M} \mid \frac{y_i - \widetilde{y}_i}{y_i} \mid \cdot 100\%$$

There are 84 observations of the water demand data of a Chinese city from January 2006 to December 2012. Fig. 3

TABLE I
COMPARISONS SUMMARY FOR MONTHLY FORECASTING

	$NMSE$	$MAPE$
ANN	0.5643	26.64%
AR	1.0232	32.42%
ELM	0.6242	13.65%
$AD-ELM$	0.0743	8.51%

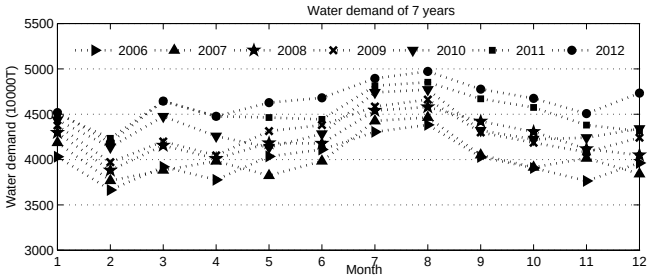

Fig. 3. The structure of ELM model.

shows the time series for each year. The number of input neurons is selected as the cycle of the learning data if possible. In Fig. 3, it can be found that the cycle for the water demand data is 12 months. Thus, the number of input nodes for ELM, BP, and AD-ELM should be set as 12, and the number of hidden is set to be 8. Here, we choose two neighbor for the model that is $k = 2$ in Eq. (12). In our work, the AR method using AR(m) where m is the number of input nodes of AD-ELM. In this paper, we forecasting the water demand data from January 2012 to December 2012. The water demand of last 12 months are forecasted in Fig. 4. In the simulation, the NMSE are 0.5643, 1.0232, 0.6242, and 0.0743 for ANN, AR, ELM, AD-ELM respectively, and the MAPE are 26.64%, 32.42%, 13.65%, 8.51% for ANN, AR, ELM, AD-ELM respectively. It is obvious that the proposed AD-ELM method outperforms the three other models as the proposed method has the advantage of adapting to local variations of trends and amplitudes.

Table 1 summarizes the prediction performance of different prediction models on the water demand data mentioned above. It is obvious that the proposed AD-ELM method outperforms the three other models.

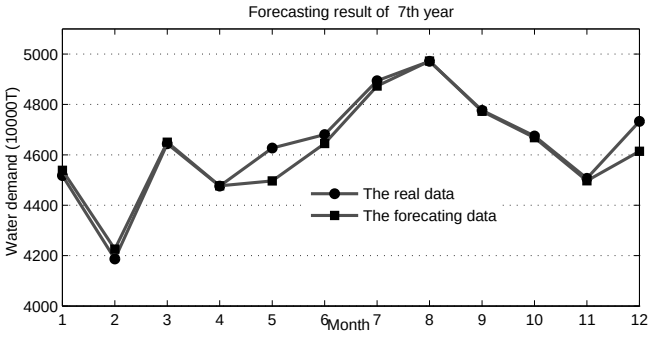

Fig. 4. The structure of ELM model.

IV. CONCLUSION

In this work we have addressed the task of trying to predict the future water demand in a city. The efficiency of supply management optimization relies on the forecast accuracy of the water demand. How to develop more accurate forecasting methods becomes an important research topic. It is known that ELM not only has a higher generalization performance than the traditional gradient-based learning algorithms but it also avoids many difficulties faced by gradient-based learning methods. But like other flexible nonlinear estimation methods, the ELM may suffer either underfitting or overfitting. In order to reduce the effect of this problem, this study presents a novel adaptive approach to extending the extreme learning machine in which the adaptive metrics of inputs is proposed for water demand prediction. The experimental results generated by a set of consistent performance measures with different metrics (MAPE, NMSE) have successfully demonstrated that the proposed model can be employed in water demand forecasting and can produce smaller predicting errors than some other forecasting methods. Thus, the AD-ELM is a better choice for the practical forecasting of water demand.

REFERENCES

[1] S.C. Huang, Online option price forecasting by using unscented Kalman filters and support vector machines, Expert Systems with Applications, 34(4)(2008)2819-2825.

[2] P.S.A. Freitas, A.J.L. Rodrigues, Model combination in neural-based forecasting, European Journal of Operational Research, 173(2006)801-814.

[3] Y. Bodyanskiy, S. Popov, Neural network approach to forecasting of quasiperiodic financial time series, European Journal of Operational Research, 175(2006)1357-1366.

[4] C.T. Lin, I.F. Lee, Artificial intelligence diagnosis algorithm for expanding a precision expert forecasting system, Expert Systems with Applications, 36(4)(2009)8385-8390.

[5] R. A. Araujo, A class of hybrid morphological perceptrons with application in time series forecasting, Knowledge-Based Systems, 24(4)(2011)513-529.

[6] M. Xia, Y. Zhang, L. Weng, and X. Ye, Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs, Knowledge-Based Systems, 36(2012)253-259.

[7] S. Makridakis, S. Wheelwright, R. Hyndman, Forecasting: Methods and Applications, New York: Wiley, (1998).

[8] K. Funahashi, On the approximate realization of continuous mappings by neural networks, Neural Networks, 2(1989)183-192.

[9] G.B. Huang, Learning capability and storage capacity of two-hidden-layer feedforward networks, IEEE Transactions on Neural Networks, 14(2003)274-281.

[10] G.B. Huang, L. Chen, C.K. Siew, Universal approximation using incremental constructive feedforward networks with random hidden nodes, IEEE Transactions on Neural Networks, 17(2006)879-892.

[11] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: theory and applications, Neurocomputing, 70(2006)489-501.

[12] Q.Y. Zhu, A.K. Qin, P.N. Suganthan, G.B. Huang, Evolutionary extreme learning machine, Pattern Recognition, 38(2006)1759-1763.

[13] Z. Guo, J. Wu, H. Lu, J. Wang, A case study on a hybrid wind speed forecasting method using BP neural network, Knowledge-Based Systems,24(7)(2011)1048-1056.

[14] M. Kulesh, M. Holschneider, K. Kurennaya, Adaptive metrics in the nearest neighbours method, Physics D, 237(2008)283-291.

[15] G.B. Huang, L. Chen, Enhanced random search based incremental extreme learning machine, Neurocomputing, 71(2008)3460-3468.

[16] Z.L. Sun, T.M. Choi, K.F. Au, Y. Yu, Sales forecasting using extreme learning machine with applications in fashion retailing, Decision Support Systems, 46(2008)411-419