Prediction model of dissolved oxygen in ponds based on ELM neural network

Xinfei Li, Jiaoyan Ai[1], Chunhuan Lin and Haibin Guan

College of Electrical Engineering, Guangxi University, Nanning, 530004

[1] shinin@vip.163.com

Abstract. Dissolved oxygen in ponds is affected by many factors, and its distribution is unbalanced. In this paper, in order to improve the imbalance of dissolved oxygen distribution more effectively, the dissolved oxygen prediction model of Extreme Learning Machine (ELM) intelligent algorithm is established, based on the method of improving dissolved oxygen distribution by artificial push flow. Select the Lake Jing of Guangxi University as the experimental area. Using the model to predict the dissolved oxygen concentration of different voltage pumps, the results show that the ELM prediction accuracy is higher than the BP algorithm, and its mean square error is $MSE_{ELM}=0.0394$, the correlation coefficient $R_{ELM}=0.9823$. The prediction results of the 24V voltage pump push flow show that the discrete prediction curve can approximate the measured values well. The model can provide the basis for the artificial improvement of the dissolved oxygen distribution decision.

1. Introduction

Dissolved oxygen is an important index to evaluate the quality of water body. In artificial landscape water bodies and large-scale aquaculture farms, due to physical, chemical, biological and human activities and other factors, the concentration of dissolved oxygen is unevenly distributed in time and space [1-3]. High or low dissolved oxygen in local waters will be harmful to aquatic ecology in a certain time. Low dissolved oxygen will threaten the survival of fish, promote the breeding of anaerobic bacteria, leading to the phenomenon of eutrophication in the region, and affecting the entire body of water. High dissolved oxygen may cause the fish to appear bubble disease [4,5]. The imbalance of dissolved oxygen concentration will endanger the ecological balance of the whole water, so it is important to improve and predict the distribution of dissolved oxygen to guide fishery and aquatic ecological remediation.

Dissolved oxygen prediction model is a complex nonlinear system. In recent years, many scholars have widely used intelligent algorithms to establish predictive modeling. Liu Shuangyin et al, using ant colony algorithm (ACA) to optimize the least squares support vector regression (LSSVR), establish the forecasting model of the dissolved oxygen in crab breeding [6]. Ma Congguo establish the model of dissolved oxygen in aquaculture ponds based on genetic algorithm and RBF neural network [7].

At present, the prediction of dissolved oxygen is mostly the prediction of dissolved oxygen change in the natural environment, and the prediction of the concentration and distribution of dissolved oxygen is less. In this paper, the distribution of dissolved oxygen was improved artificially by using water pump pushing, RBF network is used to analyze the spatial and temporal distributions of

dissolved oxygen. The ELM prediction model is established to provide reference and guidance for manual improvement of dissolved oxygen distribution.

2. Materials and methods

2.1. Experimental area

Taking Lake Jing, Guangxi University as an experimental site, and the area of 7m*5m on the east side of the lake as an experimental area, as shown in figure 1. The water area of Lake Jing is about 3000 m^2, after the artificial ecological restoration, formed by a bottom sediment, aquatic plants, submerged plants, aquatic animal and microorganisms constitute a complete production system, closed to artificial transformation of the concept of water scenery.
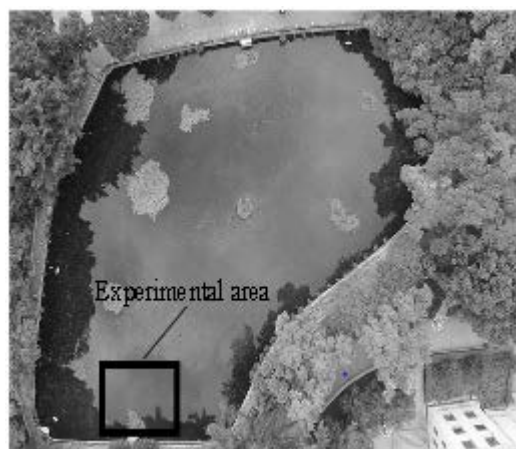
Figure 1. Experimental area.
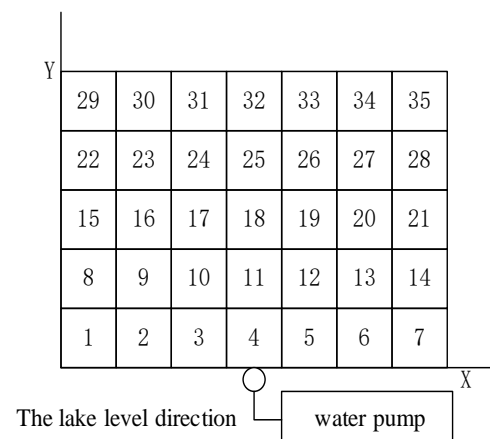
Figure 2. Grid partitioning.

The selected area is located on the east side of the lake, with a depth of about 0.5m, some of the water above the canopy shade, the bottom of the grass-like plant distribution is uneven. In order to facilitate the measurement and analysis of data, the region is divided into 1m*1m grids, as shown in Figure 2. The x-axis indicates the parallel lakeshore direction, the y-axis indicates the vertical lakeshore direction, the data acquisition points are located at the center of the grid, and each sample can get 35 data. The measurement of experimental data adopts the wireless water quality monitoring equipment of Chengdu Cable Blue Company, which is based on the wireless internet technology, uses the current measuring method to dissolve the oxygen sensor, can obtain the measurement data remotely, the measuring speed is fast and the portable is convenient to carry.

2.2. Interpolation fitting of dissolved oxygen concentration distribution

In order to reflect the distribution of dissolved oxygen concentration in the whole experimental area, the experimental data were interpolated and fitted with 35 data collected from the center of the grid at one time. Dissolved oxygen concentration is affected by many factors, and fitting is a complex surface fitting problem. Radial basis function network has a strong ability of nonlinear fitting, it can approximate any nonlinear function and has good generalization ability and fast learning convergence speed. Therefore, the RBF network is constructed and the experimental data are fitted and interpolated. With the sampling point coordinates (x, y) as input to the neural network, the dissolved oxygen concentration z as output, the RBF network is actually the mapping relation of the function z (x, y).

2.3. ELM prediction model

Elm is developed from a single hidden layer feed-forward neural network learning algorithm [8]. Compared with the traditional BP neural network, the implicit layer weights and thresholds are

randomly generated, the parameters are easier to be set, the learning convergence speed is faster and the generalization ability is stronger. The topology of the Elm network is shown in Figure 3.
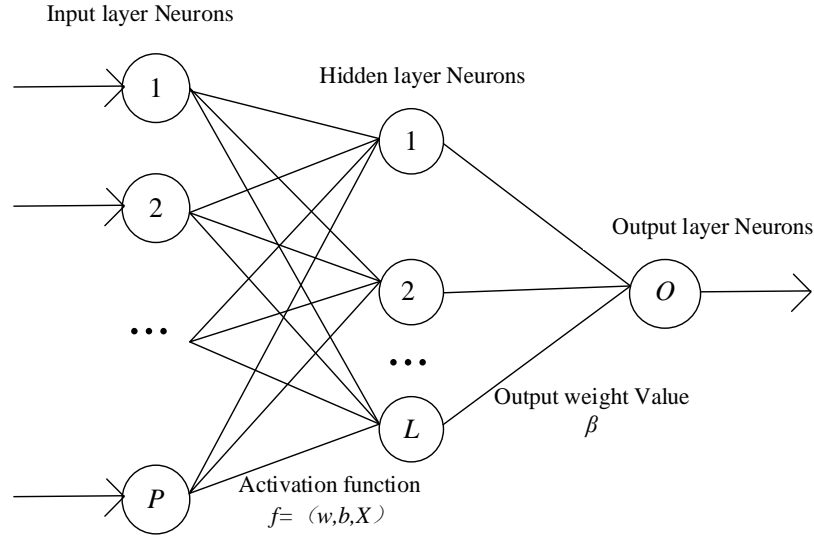
Figure 3. Topological structure of ELM Network.

The network input layer has p-neurons, the output layer neurons have 1, the hidden layer neurons have l, and f is the activation function between the input layer and the hidden layer.

For the total number of N samples $\{x_i, y_i\}$, where $\{x_{1i}, x_{2i}, \ldots, x_{ni}\} \in R^N$, the output of the network can be expressed as:

$$o_j = \sum_{i=1}^{L} \beta_i f\left(w_i, b_i, x_j\right) , j = 1,2,\ldots,N \tag{1}$$

where $W_i = [w_{i1} \ldots w_{in}]^T$ is the weight vector of the i-th hidden neuron and the input; b_i is the threshold of the i-th hidden neuron.

If the "sigmoid" function of the formula (2) is the activation function f, the N samples are approximated by zero errors means that $\sum_{j=1}^{N}\|o_j - y_j\| = 0$, i.e., then there is β, W, and b which makes the formula (3) established.

$$f(u) = \frac{1}{1+e^{-u}} \tag{2}$$

$$\sum_{i=0}^{L} \beta_i f\left(w_i, b_i, x_j\right) = y_j \quad j = 1,2,\ldots,N \tag{3}$$

The formula (3) can be abbreviated as:

$$H \cdot \beta = Y \tag{4}$$

where

$$H(w_1, w_2, \ldots, w_L, b_1, b_2, \ldots, b_L, x_1, x_2, \ldots, x_N) = \begin{bmatrix} f(w_1, b_1, x_1) & \cdots & f(w_L, b_L, x_1) \\ \vdots & \cdots & \vdots \\ f(w_1, b_1, x_N) & \cdots & f(w_L, b_L, x_N) \end{bmatrix}_{N \times L} \tag{5}$$

$$\beta = [\beta_1^T \ldots \beta_L^T]_L \tag{6}$$

$$Y = [Y_1^T \ldots Y_N^T]_N \tag{7}$$

H is the implicit layer output matrix of the network.

$$\beta = H^+ Y \tag{8}$$

where H^+ is the inverse matrix of H.

The main steps of the ELM algorithm are as follows:

1) Determines the number of neurons in the hidden layer L, randomly generates the output weights of the hidden layer W_t and bias $b_t, t = 1,2, ... , L$;

2) Selects the activation function of the hidden layer and calculates the output matrix H.

3) Computes the weight of the output layer β: $\beta = H^+Y$. Where H, β and Y are defined as formula (5) and (6) [9].

3. Results and discussion

3.1. Experimental data processing and analysis

In the experiment, the pumps with the operating voltage of 12V, 24V and 48V three kinds of voltages were artificially pushed in the natural environment, and a group of no push flow experiments were used as control. Each experiment starts at 8:30 a.m., ends at 11:00 a.m., and measures data at intervals of 30 minutes. Each experiment can be measured 35*6 group data, a total of 4 groups of experiments. The partial data obtained by the no push flow group is shown in table 1.

Table 1. The partial data obtained by the no push flow group.

| (x, y)	8:30	9:00	9:30	10:00	10:30	11:00
(1,1)	1.312	1.710	3.121	4.221	5.322	6.572
(1,2)	1.309	1.891	3.685	5.232	6.105	7.522
(1,3)	1.490	1.956	4.132	5.365	6.621	8.321
...	...	...	...	...	...	...
(7,5)	3.341	3.632	5.122	6.384	7.294	8.743

3.2. Fitting results of oxygen concentration distribution
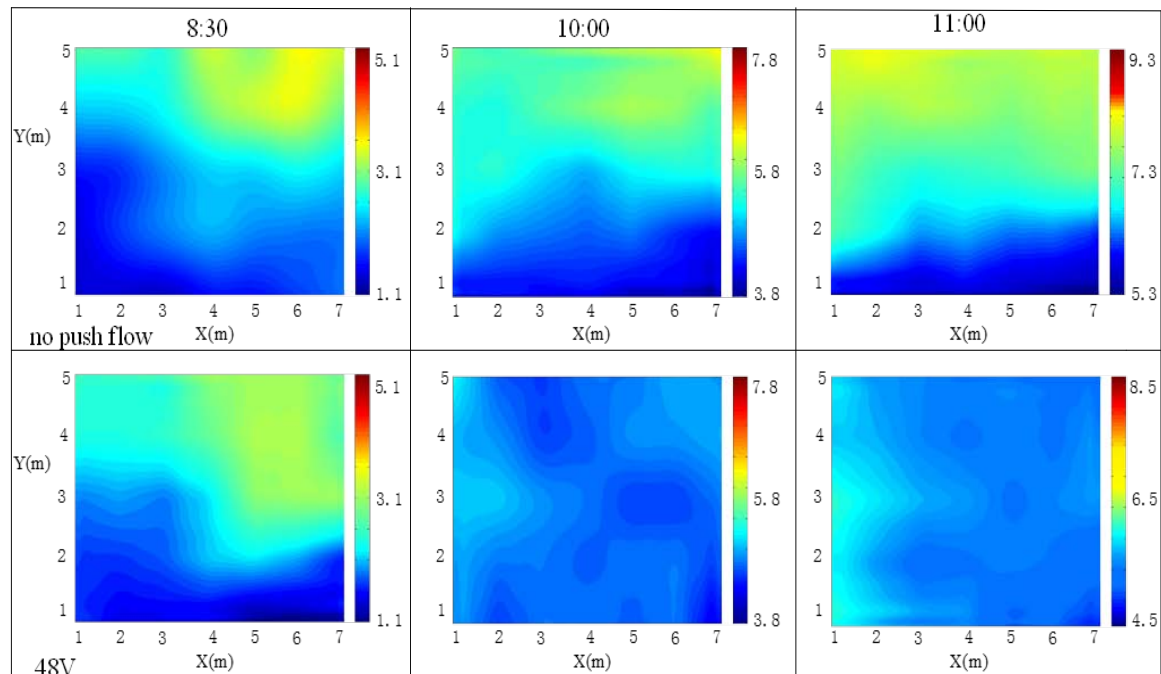
Figure 4. Pseudo-colour map of dissolved oxygen concentration.

The X axis in Figure 4 represents the parallel shore direction, and the Y axis represents the vertical lakeshore direction. The color represents the concentration of dissolved oxygen. At 8:30, the concentration distribution of dissolved oxygen from the shore to the middle of the lake is significantly

stepped. Because the grass-like plant distribution is uneven, plant lush regional photosynthesis, rapid recovery, therefore, under the condition of no push-flow, the difference of dissolved oxygen distribution increases with time, while in the 48V pump flow, the water flow makes the oxygen and oxygen region mixed, the dissolved oxygen difference is more and more small, and the dissolved oxygen concentration in the experimental area is balanced at 11:00. In order to quantify the distribution equilibrium of dissolved oxygen concentration, the dispersion coefficient Vx is chosen as an index. The smaller the value of Vx, the more balanced the distribution of dissolved oxygen.
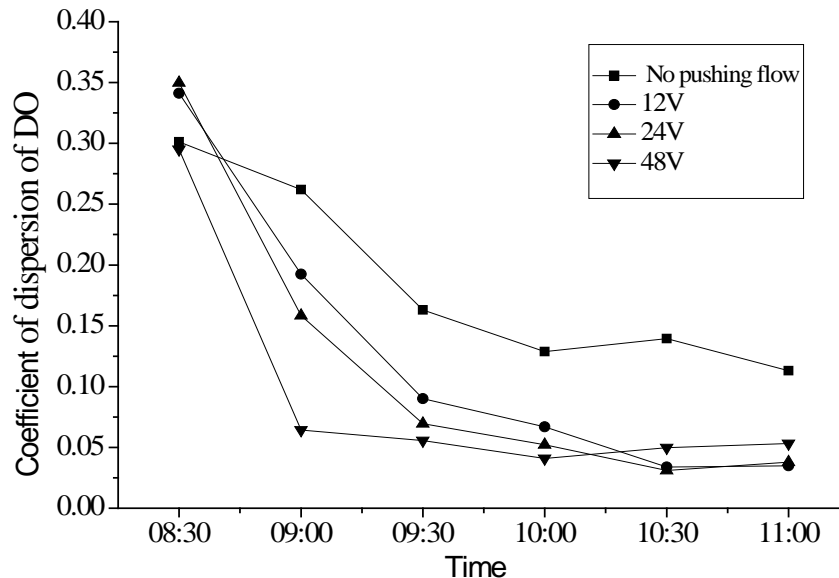
Figure 5. Variation curve of dissolved oxygen concentration dispersion coefficient.

As shown in Figure 5, at the end of the experiment, the water pump push-flow group can eventually maintain the dissolved oxygen dispersion in the vicinity of 0.05, and the 48V water pump push down the dispersion coefficient first reached 0.05.

3.3. ELM prediction model
In order to provide reference and basis for selecting water pump for artificial push flow and determining the length of push-flow, a predictive model of dissolved oxygen concentration based on limit learning machine is established. The modeling steps are as follows:

1) Processing sample data. The new sample set obtained by the fitting is normalized.

2) Building and training networks. Grid coordinate value x, y, initial concentration value DO_0, time interval ∆t, pump voltage v as network input value, dissolved oxygen concentration at each coordinate point as output value. The training set and test set are allocated according to the proportion of 7:3. We adopt the "trial-and-scrape" method to select the appropriate number of hidden layer nodes, and obtain the input weights of the ELM model *w*, hidden neurons threshold *b*, output weights *β*.

3) Testing network performance and prediction. The predictive output is compared to the real value, and if the error is within the required range, the model compound requirement is described.

After many training tests, when the number of neurons in the hidden layer of the network is set to 50, the mean square error of the prediction model is $MSE_{ELM}=0.0394$, and the correlation coefficient is $R_{ELM}=0.9823$. The same training sample is brought into the BP network for training, and the mean square error of the BP model prediction results is $MSE_{BP}=0.1221$, the correlation coefficient $R_{BP}=0.9681$. As shown in Figure 6, the predicted curve of ELM model can approximate the measured value curve well, and the prediction accuracy is better than BP network model.
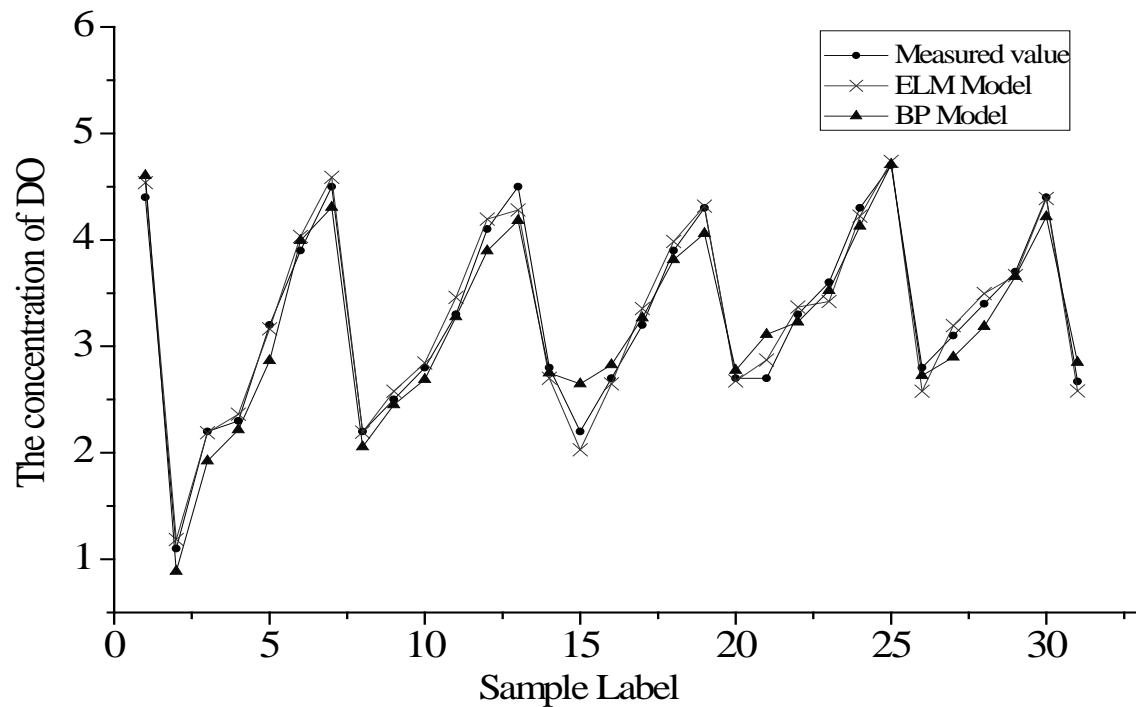
Figure 6. Model prediction results.

In order to further test the prediction effect of the model, using 24V voltage pump to push flow. The initial value and voltage of dissolved oxygen are brought into the model, and the dispersion coefficient is calculated according to the predicted dissolved oxygen concentration values, and compared with the measured value, as shown in Figure 7.
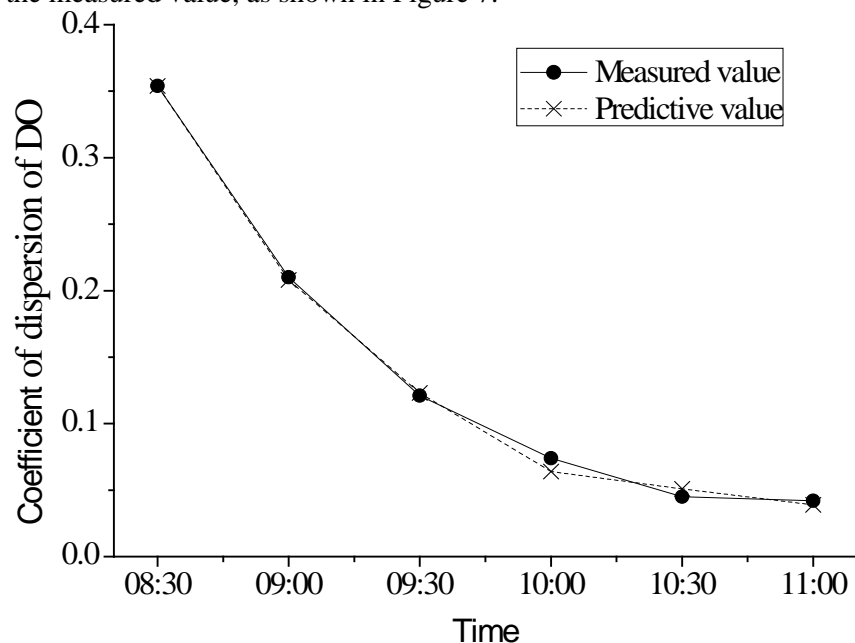
Figure 7. Variation curve of dissolved oxygen concentration dispersion coefficient.

As can be seen from Figure 7, the dispersion coefficients predicted by the model can well follow the measured values, and further verify the accuracy of the prediction model. According to the

predicted results, it takes only about 2 hours to reduce the dispersion coefficient of DO to 0.05 by using a 24V pump to push the flow.

4. Conclusions

The equilibrium of dissolved oxygen concentration is of great significance to the ecological stability of water bodies. In this paper, the prediction model of dissolved oxygen under artificial flow is established based on ELM algorithm, the initial concentration of dissolved oxygen, pump voltage and push time are input variables of the model. Using this model to predict the dissolved oxygen in the Lake Jing experimental area, the mean square error of the predicted results in the experimental time period is MES_{ELM}=0.0394, which is lower than the BP network. The model has good predictive accuracy. By using the prediction model, we can get the variation curve of dissolved oxygen concentration in different voltage pumps, and by analysing the discrete curve, it can provide reference for pump selection and push time, reduce the manual maintenance strength and improve the efficiency.

Acknowledgement

References

[1]   Cao Y J 2011 Temporal and spatial distribution characteristics of dissolved oxygen in the lake reservoir of Zhejiang Province and analysis of influencing factors *J. Journal of Jiangxi Agriculture*. **23 (10)** 151-153

[2]   Dai H, Li C, Ma X, et al 2013 Study on the distribution of dissolved oxygen in the ponds of Eriocheir sinensis *J. Journal of Shanghai Ocean University*. **22 (1)** 66-73

[3]   Hung W W, Zheng X C, Liao F F, et al 2008 Characteristics of eutrophication of reclaimed water landscape water *J. China Water Supply and Drainage*. **24 (1)** 65-68

[4]   Zhu Y L 2004 Discussion on dissolved oxygen in aquaculture water *J. Guide to the enrichment of Fisheries*. **(24)** 23-24

[5]   Gu J, Meng T, Liu X G, et al 2011 Energy-saving technology of pond aerobic energy based on oxygen mass transfer *J.Chinese Journal of Agricultural Engineering*. **27 (11)** 120-125

[6]   Liu S Y et al 2012 Estimation model of dissolved oxygen in crab breeding based on ant colony optimization least squares support vector regression *J. Journal of Agricultural Engineering*. **28 (23)** 167-175

[7]   Ma C G and Zhao D A 2011 Dissolved Oxygen Model of Pond in Pond Based on Genetic Algorithm and RBF Network *J. China Rural Water and Hydropower*. **(2)** 14-16

[8]   Huang G B, Chen L, Siew C K 2006 Universal approximation using incremental constructive feed forward networks with random hidden nodes *J. IEEE Transactions on Neural Networks*. **17(4)** 879   −892

[9]   Huang G B, Zhu Q Y, Siew C K 2004 Extreme learning machine: a new learning scheme of feed forward neural networks *J. Proc.int.joint Conf.neural Netw*. 2:985-990 vol.2