Fast Recursive Identification Algorithm for Nonlinear Time Series model based on improved Extreme Learning Machine

[*1] Jie Jia, [1] Yanyan Zhou, [*2] Yong Yang, [1] Xiaona Luo
[1] *Institute of Aerospace Information and Security Technology, Nanchang Hangkong University, Nanchang 330063, China, E-mail: jiajie757@126.com*
[2]*School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China, E-mail:greatyangy@126.com*

Abstract

Extreme Learning Machine for its fast learning speed and generalization performance advantages has been concerned widespreadly by scholars of the system identification and pattern recognition and others fields. However, there are some problems in traditional Extreme Learning Machine. Matrix morbidly and adaptive tracking slow and low accuracy and large-scale network are caused by the randomness in the process of transition fitting and hidden node selection. The development of extreme Learning Machine was restricted by these problems. The Akaike information criterion was introduced as the optimal stopping criterion to select the appropriate number of hidden nodes. And a new method for fast identification nonlinear time-varying system based on the improved Extreme Learning Machine algorithm is proposed. In addition, learned from the ELM a learning ideas and recursive least squares theory, transition fitting and local optimal solution was avoided. The nonlinear time-varying dynamic characteristics of the system can be better tracked, with good generalization, robustness and controllability. Through typical nonlinear system identification simulation examples, compared with the existing common methods, the algorithm's good performance of simple calculation, fast convergence and high accuracy of recognition can be shown.

Keywords*: Extreme Learning Machine, Fast Identification, Recursive Least Squares*

1. Introduction

There are many problems which can not be ignored by time-varying and nonlinear of system in the fields of aerospace and mechanical engineering, for example, the tethered system for small spacecraft flexible manipulator, solar array, large antenna with a flexible multi-body system characteristics, and parametric excitation of nonlinear vibration system [1-2]. It is difficult to accurately describe structural features due to the complexity of the structural design and the usage of a variety of advanced materials, system identification is a major way to model such complex systems [3-4]. Since 1980, Billings published review articles of the nonlinear system identification; the various aspects of the theory of nonlinear identification have made some progress. In recent years, the knowledge of the neural network, support vector machine and extreme learning machine, which are based on the intelligent control theory, have produced many new identification methods to open up a new avenue for the identification of nonlinear systems [5].

Highly nonlinear fitting, the neural network method is in full development and application in the identification of nonlinear systems. But usually neural network method applies only to the identification of non-linear time-invariant systems. Only in recent years, for the time-varying nonlinear system identification algorithm [6-10] was proposed. The main Internet models include the minimum radial basis function neural network model, which yet is less effective for continuous time-varying nonlinear systems. For this reason, Hachino [12] explored continuous time nonlinear system identification based on radial basis function network model and immune algorithm, in which the immune algorithm can effectively track time-varying system parameters. There are some of the latest developments in neural network methods [13, 14].

Support vector machine based on statistical learning theory is similar with neural network in the structure and function. As is known in literature [15], the support vector machine can approach any of a class of non-linear function, which provides a theoretical basis for the use of support vector machines for system identification and control. Meanwhile, support vector machine computing speed, no local minimum point, and will not be the same as the neural network into local solution. In recent years,

support vector machines has made breakthrough progress in the system identification and control. SVM was used in linear and nonlinear system modeling by GS Santos [16] , the linear model was used for describing parameter estimation method, and the nonlinear model was used for describing model structure identification method. Besides, SVM was applied to the black-box system identification by Sariyildiz. E.[17] to perform the identification simulation for the dynamic behavior of the mechanical arm. The least squares support vector machine (LS - SVM) as an extension of the standard support vector machines, only solving linear equations, computational speed, has been widely used in the parameter estimation and other fields [18-19]. These algorithms are differ from one another in aspect of noise's sensitivity, calculation accuracy and efficiency, the research target of the new method is always hope that there are better improvements in these performance indicators than existing algorithms.

Extreme Learning Machine (ELM) learning algorithm is an effective forward neural network construction algorithm [20]. The algorithm has been verified by a number of benchmark problems and the practical application; the results show that the algorithm has good generalization capability and fast learning speed [21-23]. Extreme Learning Machine algorithm has been developed and improved [24-27], and been widely applied in the system identification [28-29]. However, there are two aspects defects of the algorithm in the process of application. Initially, some of the hidden nodes, selecting randomly in ELM network algorithm, lack of representation. The ELM network algorithm needs more hidden nodes comparing with traditional neural network learning algorithm based on parameter tuning; thereby it increases the complexity of the network. Secondly, it is necessary for learning parameters to determine by trial and error in the practical application, accordingly, it increases the subjectivity in the process of determination for network hidden node.

The Akaike information criterion is brought in this paper as the optimal stopping criterion to choose the appropriate number of hidden layer nodes, in addition, ELM's one-time learning idea and based on recursive least squares theory are motivated, over fitting and the local minimum are averted. What's more, over large-scale networks, matrix morbid caused by randomness in selection process of Network hidden node and the trial and error of learning parameters are all solved. And based on the improved Extreme Learning Machine algorithm, a new fast recursive identification algorithm for nonlinear time series model are proposed.

2. System description and improved ELM

2.1 System description

For general discrete nonlinear time-varying system, the relationship among input, output and parameter vector can be described in the following relationship,

$$y(k) = f(y(k-1),\cdots y(k-n_y),u(k),u(k-1),\cdots u(k-n_u),\theta(k)) + v(k) \tag{1}$$

Where the $y(k) \in R^m$, $u(k) \in R^n$ and $v(k) \in R^n$ are respectively the output vector, input vector and noise vector. Where n_y and n_u are respectively the maximum delay of the output and input. Here v_k is assumed to be Gaussian independently distributed sequence with zero-mean. Where $\theta(k)$ is the parameter vector of time-varying system, and f is unknown nonlinear vector function. The goal is, given the system input and output data, to design learning algorithms, and to simulate the nonlinear time-varying characteristics relationship of output, input and time-varying parameters.

2.2 Improved ELM

Given the training sample set (u_k, y_k) , $k = 1,\cdots,N$, where $u_k \in C^n$ is input vector, $y_k \in C^m$ is the corresponding desired output. The output of single hidden layer forward neural network with $\tilde{N}$ additive hidden nodes can be modeled as

$$f_{\tilde{N}}(x_k) = \sum_{i=1}^{\tilde{N}} \beta_i G(a_i \cdot u_k + b_i) \tag{2}$$

Where $\alpha_i = [\alpha_{1i},\cdots,\alpha_{ni}]^T$ ($i = 1,\cdots,\tilde{N}$) is weight vector connecting the input nodes and the $i-th$ hidden node, β_i is weight vector connecting the $i-th$ hidden node and the output nodes. b_i is

threshold of the $i-th$ hidden node, $a_i \cdot u_k$ represents the inner product of a_i and u_k in R^n [30-31]. Activation function G(x) is S type additive function. This method can be applied to the case of the scalar output and the vector output. In order to simplify the representation and take the scalar output for example here.

The ultimate aim of the single hidden layer feed forward neural network learning algorithm is how to determine the values of β_i , a_i and b_i .

$$f_{\tilde{N}}(u_k) = \sum_{i=1}^{\tilde{N}} \beta_i G(a_i \cdot u_k + b_i) = y_k, \quad k = 1, \cdots N \tag{3}$$

Equation (3) can be viewed as a special case of the following linear regression model:

$$y_k = \sum_{i=1}^{\tilde{N}} \beta_i G(a_i \cdot u_k + b_i) + \varepsilon_k, \quad k = 1, \cdots N \tag{4}$$

The structure of the network can be described by the following matrix formula,

$$Y = H\beta + E \tag{5}$$

Where,

$$Y = [y_1, ..., y_{N}]^T$$

$$H = \begin{bmatrix} G(a_1, b_1, u_1) & \cdots & G(a_{\tilde{N}}, b_{\tilde{N}}, u_1) \\ \vdots & \cdots & \vdots \\ G(a_1, b_1, u_N) & \cdots & G(a_{\tilde{N}}, b_{\tilde{N}}, u_N) \end{bmatrix}_{N \times \tilde{N}}$$

$$\beta = [\beta_1, \cdots, \beta_{\tilde{N}}]^T \quad , \quad E = [\varepsilon_1, \cdots, \varepsilon_N]^T$$

Let $\theta(k)$ represent $\tilde{N} \times 1$ order vector of all the network connection weights and get a network structure,

$$y(k) = \sum_{i=1}^{\tilde{N}} \beta_i G(a_i \cdot u_k + b_i) = \overline{f}(u(k), \theta(k)) \tag{6}$$

$\overline{f}$ notes the approximation function of network , the following problems are how to determine the output of network $y(k)$ when input of the nonlinear time-varying is given.

The Akaike information criterion [30] is a criterion, which has been widely used for the statistical model selection in the field of statistical analysis. It also is widely used in the selection of the neural network mode l [30-32], and has been successfully applied [30-33]. Application of the guidelines can reduce network complexity on the the premise of guaranteeing the accuracy, which can guarantee the dialectical accuracy while optimizing generalization performance. Akaike criteria expressed as follows,

$$AIC = N \log\left(\frac{RSS}{N}\right) + 2K \tag{7}$$

Where RSS is residual sum of squares, N means the number of samples; K indicates the number of network independent parameters. By this combination of the information criterion to select the method of random hidden nodes, the smallest AIC value corresponds to the structure of the network can be selected as the optimal network structure.

3. Learning algorithm

Nonlinear time-varying system identification method based on improved ELM, is the connection weights and learning parameters of the network as a system of time-varying parameters to reflect the system's output with the input changes.The estimation methods of network weights usually include Recursive Least Squares (RLS), extended Kalman filter (EKF)and so on. RLS method is usually assumed that the connection weights obey the law of random walks, that is,

$$\theta(k) = \theta(k-1) + v(k) \tag{8}$$

Generated by the network model error $e(k)$ can be expressed as $e(k) = d(k) - y(k)$, then

$$d(k) = y(k) + e(k) = \overline{f}(u(k), \theta(k)) + e(k) \tag{9}$$

Where $d(k)$, $y(k)$ denotes the desired output and the actual output of the network. Let $\overline{\theta}(k)$ become the estimated value of $\theta(k)$. Then the nonlinear function $\overline{f}(u(k), \theta(k))$ is performed the Taylor expansion

at $\bar{\theta}(k-1)$ (at time $k-1$, the estimated value of the state vector), it only retains the linear term and obtains the state equations of system as following,

$$\begin{cases} \theta(k) = \theta(k-1) + v(k) \\ y(k) = F^T(k)\theta(k) + T(k) + e(k) \\ T(k) = \bar{f}(u, \bar{\theta}(k-1)) - F^T(k)\bar{\theta}(k-1) + \eta(k) \end{cases} \tag{10}$$

where $F(k)$ is a $\tilde{N} \times m$ order matrix , $\eta(k)$ is higher-order term of Taylor expansion and it can be neglected in the actual calculation, the minimum variance estimation $\bar{\theta}(k)$ of $\theta(k)$ can directly gain by identification formula of recursive least squares algorithm,

$$\begin{cases} \bar{\theta}(k+1) = \bar{\theta}(k) + K(k+1)(d(k+1) - F^T(k+1)\bar{\theta}(k)) \\ K(k+1) = P(k)F(k+1)(1 + F^T(k+1)P(k)F(k+1))^{-1} \\ P(k+1) = P(k) - K(k+1)F^T(k+1)P(k) \end{cases} \tag{11}$$

The above process only given initial value can be recursive calculation continues, and has been proved to be convergent. These steps are slow convergence and low recognition accuracy, and therefore we hope that these can be improved.

As can be seen from the previous algorithm recursive formula, a main process of algorithm is the variation of $P(k)$, we convert their patterns of variation to make it simple, and then it is applied to the learning of the network as a learning algorithm. The estimation procedures of changed network weights are as follows,

$$\begin{cases} \bar{\theta}(k+1) = \bar{\theta}(k) + K(k+1)(d(k+1) - F^T(k+1)\bar{\theta}(k)) \\ K_1(k+1) = F(k+1)(1 + F^T(k+1)P(k)F(k+1))^{-1} \\ K(k+1) = P(k)K_1(k+1) \\ P(k+1) = P(k) - K_1(k+1)F^T(k+1) \end{cases} \tag{12}$$

The significance of each variable in the equations is not changed, an intermediate variable is just introduced, but the calculation of $P(k)$ in new algorithm obtains a certain degree of simplification.

4. Algorithm performance experiments

Considering a general nonlinear Wiener ARARMAX model, simulation object can be described as the following,

$$\begin{aligned} & A(z)\bar{y}(k) = B(z)u(k) + \frac{D(z)}{C(z)}v(k) \\ & A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} = 1 - 1.60 z^{-1} + 0.80^{-2} \\ & B(z) = b_1 z^{-1} + b_2 z^{-2} = z^{-1} + 0.65 z^{-2} \\ & C(z) = 1 + c_1 z^{-1} = 1 + 0.8 z^{-1} \\ & D(z) = 1 + d_1 z^{-1} = 1 - 0.64 z^{-1} \\ & \bar{y}(k) = g(y(k)) = k_1 y(k) + k_2 \sin(y(k)) + k_3 \cos(y(k)) \\ & \quad = -0.40 y(k) + 0.20 \sin(y(k)) - 0.80 \cos(y(k)) \end{aligned}$$

The performance of ELM is generally better than many commonly used algorithms, such as BP algorithm, support vector machines (SVM) algorithm [20], thus it only compares improved parameter identification algorithm of ElM with the traditional in this paper. For convenience sake, it takes advantage of two comparison algorithms to approximate the training set (x_i, y_i) and the test set (x_i, y_i), respectively. Data length $L = 5000$, define process parameter estimation error $\delta_s = \|\bar{\theta} - \theta\| / \|\theta\|$, where x_i are randomly distributed on the interval [-10, 10]. In order to make this parameter identification problem close to the practical problems, the standard random noise whose amplitude in [-0.2, 0.2] can be joined in each data of training set, yet it does not add in noise in data of test set.

The activation function of improved ELM algorithm in this experiment is additive S function $G(x)=1/(1+\exp(-(a\cdot x+b)))$, initial input weight a randomly selects from [-1,1] while the bias b randomly selects from [0,1]. The number of hidden nodes is determined by trial and error in the practical application of the conventional ELM. In this experiment, it can meet the requirement to solve the problem because the initial number of hidden nodes are 500, so the number of hidden nodes of the algorithm set [10,500].The result shows that the value of ACI was minimum when the hidden units number (HUN) equals 14 , the value of AIC is decided through calculating the hidden nodes, 14 hidden nodes should be chosen in the experiment to construct a network of ELM. The experimental results are the average of 50 independent computing in the experiment of improved algorithm.

The curve for Parameter estimation error versus k can be shown in Figure 1 and Table 1.

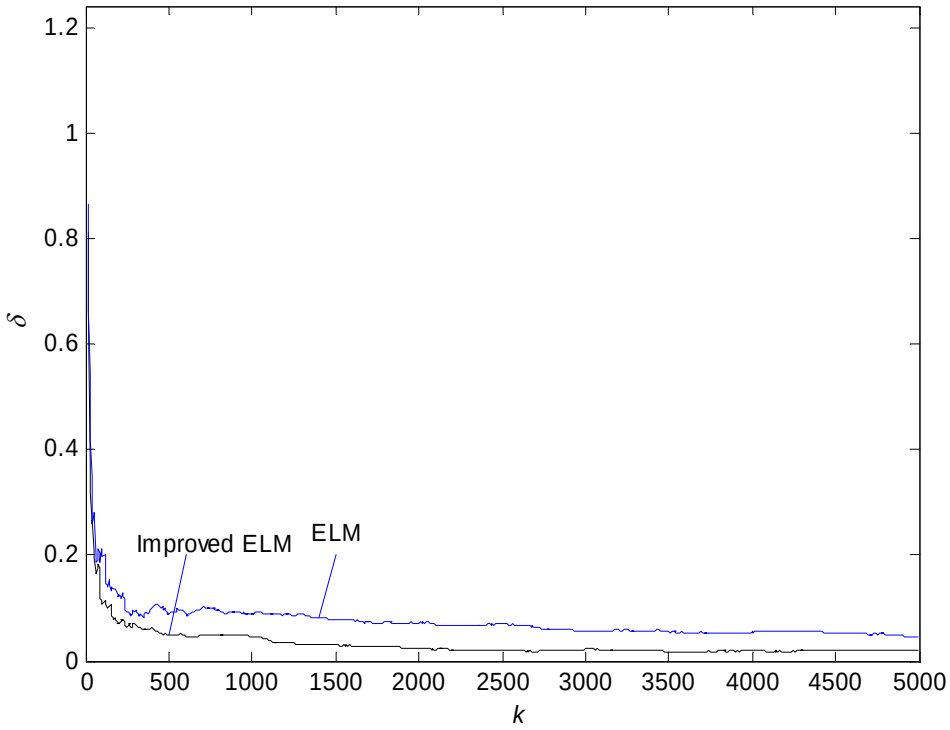

Figure 1. The parameter estimation errors versus k for the algorithms (σ2=0.402)

Table 1. The Improved ELM estimates and errors $\boldsymbol{\sigma_s}$

As can be seen from the data, AIC criterion as the cut-off criteria for online learning, the improved Extreme Learning Machine compared with the traditional Extreme Learning Machine at a much faster speed computing has been considerable estimation accuracy.

The performance of ELM and improved ELM can be shown by comparison in the Table 2. As it reveals in the table, improved ELM reaches to considerable estimation accuracy with much less computation time. In addition, the network size of improved ELM is also smaller than the traditional.

Table 2. Comparison of generalization performance

Algorithm	Hidden node number	Test results		Training time
		RMS	Dev	
Improved ELM	14	0.0072	0.0012	24.4047
ELM	20	0.0074	1.6163e-004	160.1541

5. Conclusions

There are some problems in the Traditional Extreme Learning Machine, Matrix morbidly and adaptive tracking slow and low accuracy and large-scale network caused by the randomness in the process of transition fitting and hidden node selection. In order to solve those problems, the Akaike information criterion was introduced as the optimal stopping criterion to select the appropriate number of hidden nodes. What's more, over large-scale networks, matrix morbid caused by randomness in selection process of Network hidden node and the trial and error of learning parameters are all solved. In addition, learned from the ELM a learning ideas and recursive least squares theory, transition fitting and local optimal solution was avoided. The nonlinear time-varying dynamic characteristics of the system can be better tracked, with good generalization, robustness and controllability. Through the simulation example of typical nonlinear system identification, compared with the traditional ELM, Streamline the structure of the neural network is constructed. Moreover, the considerable computation accuracy is obtained with the breakneck speed of operation, and can be called an efficient neural network learning algorithm.

6. Acknowledgments

This work was supported by the National Natural Science Foundation of China (61263012, 60963012, 61262034), by the Postdoctoral Foundation of China (2012M510593), by the Key Project of Chinese Ministry of Education (No. 211087), and by the Natural Science Foundation of Jiangxi Province (No. 2010GZS0052, No. 20114BAB211020).

7. References

[1] Gert H. K, Heirman, Frank Naets, Wim Desmet, "A system-level model reduction technique for the efficient simulation of flexible multibody systems", Numerical Methods in Engineering, vol. 85, no. 3, pp.330-354, 2011.
[2] Qiang Tian, Cheng Liu, Margarida Machado and Paulo Flores, "A new model for dry and lubricated cylindrical joints with clearance in spatial flexible multibody systems", Nonlinear Dynamics, vol.64, no.1, pp.25-47, 2011.
[3] Bidyadhar Subudhia, Debashisha Jenab, "A differential evolution based neural network approach to nonlinear system identification", Applied Soft Computing, vol.11, no.1, pp.861-871, 2011.
[4] Feng Ding, "Hierarchical multi-innovation stochastic gradient algorithm for Hammerstein nonlinear system modeling", Applied Mathematical Modelling, Available online 30 April 2012.
[5] J. Ward MacArthur, "A new approach for nonlinear process identification using orthonormal bases and ordinal splines", Journal of Process Control, vol. 22, no.2, pp. 375-389, 2012.
[6] R.K. Al Seyab, Y. Cao, "Nonlinear system identification for predictive control using continuous time recurrent neural networks and automatic differentiation", Journal of Process Control, vol. 18, no.6, pp. 568–581, 2008.

[7] Thomas B. Schön, Adrian Wills, Brett Ninness, “System identification of nonlinear state-space models”, Automatica, vol. 47, no.1, pp. 39-49, 2011.
[8] Chunxiao Bao, Hong Hao,Zhong-Xian Li, Xinqun Zhu, “Time-varying system identification using a newly improved HHT algorithm”, Neurocomputing, vol. 87, no.23, pp. 1611-1623, 2009.
[9] Xueli Wu, Jianhua Zhang, Quanmin Zhu, “A generalized procedure in designing recurrent neural network identification and control of time-varying-delayed nonlinear dynamic systems”, Neurocomputing, vol. 73, no.7, pp. 1376-1383, 2010.
[10] Abiyev, R. H, “Type 2 Fuzzy Neural Structure for Identification and Control of Time-Varying Plants”, IEEE Transactions on Industrial Electronics , vol. 57, no.12, pp. 4147-4159, 2010.
[11] F. Bonanno,G. Capizzi,G. Graditi, C. Napoli, G.M. Tina, “A radial basis function neural network based approach for the electrical characteristics estimation of a photovoltaic module”, Applied Energy, vol. 97, no.11, pp. 956-961, 2012.
[12] Hachino, Tomohiro Takata, Hitoshi, “On-line identification of continuous-time nonlinear systems using radial basis function networks and iuunune algorithm”, In Proceedings of the 5th International Conference on Control and Automation, ICCA'05, pp.587-592, 2005.
[13] Zhang Youmin, “A fast and numerically robust neural network training algorithm”, Artificial Intelligence and Soft Computing ICAISC 2006-8th International Conference, pp.160-169, 2006.
[14] Alessio Titti, Stefano Squartini and Francesco Piazza, “A new Time-Variant Neural based approach for nonstationary and non-linear System Identification Circuits and Systems”, IEEE International Symposium, vol.5,no23, pp. 5134-5137, 2005.
[15] Chia-Nan Ko, “Integration of support vector regression and annealing dynamical learning algorithm for MIMO system identification”, Expert Systems with Applications, vol.38,no.12, pp. 15224-15233, 2011.
[16] Glauber Souto dos Santos, Luiz Guilherme Justi Luvizotto, Viviana Cocco Mariani, Leandro dos Santos Coelho, “Least squares support vector machines with tuning based on chaotic differential evolution approach applied to the identification of a thermal process”, Expert Systems with Applications, vol.39,no.5, pp. 4805-4812, 2012.
[17] Sariyildiz, E., Ucak, K., Oke, G., Temeltas, H.,Ohnishi, K, “Support Vector Regression based inverse kinematic modeling for a 7-DOF redundant robot arm”, Innovations in Intelligent Systems and Applications (INISTA), 2012 International Symposium on, pp. 1-5, 2012.
[18] Ginés Rubio, Héctor Pomares, Ignacio Rojas, Luis Javier Herrera, “A heuristic method for parameter selection in LS-SVM: Application to time series prediction”, International Journal of Forecasting, vol.27, no.3, pp. 725-739, 2011.
[19] S. Deng, Tsung-Han Yeh, “Using least squares support vector machines for the airframe structures manufacturing cost estimation”, International Journal of Production Economics, vol.131, no.2, pp. 701-708, 2011.
[20] Qiang Lu, Zhiguang Wang, "A Semantic based P2P Personal Knowledge Sharing Model", IJACT, Vol. 4, No. 1, pp. 33 ~ 41, 2012.
[21] Liang N Y, Huang G B, “A fast and accurate online sequential learning algo-rithm for feedforward networks. IEEE Transactions on Neural Networks”, vol.17, no.6, pp.1411-1423, 2006.
[22] Han F, Huang D S. Improved extreme learning machine for function approximation by encoding a priori information. Neurocomputing, vol.69, no.16, pp. 2369-2373, 2006.
[23] Zhu Q Y, Qin A K, Suganthan P N, “Evolu-tionary extreme learning machine Pattern Recogni-tion”, vol.38, no.10, pp. 1759-1763, 2005.
[24] Wang G, Zhao Y, Wang D, “A protein secondary structure prediction framework based on the Extreme Learning Machine”, Neurocomputing, vol.72, no.1, pp. 262-268, 2008.
[25] Lan Y, Soh Y C, Huang G B, “Ensemble of online sequential extreme learning machine”, Neurocomputing, vol.72, no.13, pp. 3391-3395, 2009.
[26] Colin Lemmon, Siu Man Lui, Ickjai Lee, "Review of Location-Aware Routing Protocols", AISS, Vol. 2, No. 2, pp. 132 ~ 143, 2010.
[27] Suresh S, Babu R V, Kim H J, “No-reference image quality assessment using modified extreme learning ma-chine classifier”, Applied Soft Computing, vol.9, no.2, pp. 541-552, 2009.
[28] Ming-Bin Li, Meng Joo Er, “Nonlinear system identification using extreme learning machine”, the9th International Conference on Control, Automation, Robotics and Vision, ICARCV '06 , pp. 1-4, 2006.

[29] Jing Deng, Kang Li, George W. Irwin, "Fast automatic two-stage nonlinear model identification based on the extreme learning machine", Neurocomputing, vol.74, no.16 pp. 2422-2429, 2011.
[30] Panchal G, Ganatra A, Kosta Y P, "Searching most efficient neural network architecture using Akaike's information criterion(AIC)", International Journal of Computer Applications, vol.1, no.5, pp.41-44,2010.
[31] Yin Jianchuan, Zou Zaojian, XU Feng, "An improved extreme learning machine based on Akaike criterion. Journal of Shandong University (Engineering Science)", vol.41, no.6, pp.7-11, 2011.
[32] Bjorck A, "Numerics of Gram-Schmidt orthogonalization", Linear Algebra Application, vol.198, pp.297-316, 1994.
[33] Zhao Z Y, Zhang Y, Liao H J, "Design of ensemble neural network using the Akaike information criterion", Engineering Applications of Artificial Intelligence, 2008, vol.21, no.8, pp.1182-1188,2008.