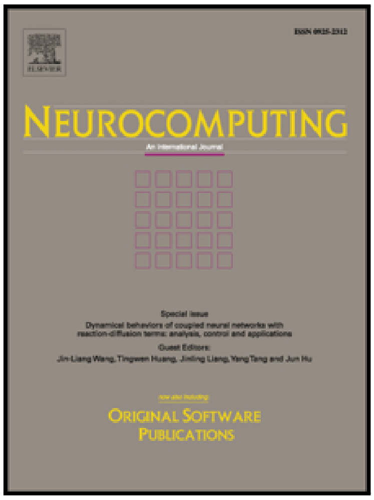
Communicated by Prof. Zidong Wang

Accepted Manuscript



A Switching Delayed PSO Optimized Extreme Learning Machine for Short-Term Load Forecasting

Nianyin Zeng, Hong Zhang, Weibo Liu, Jinling Liang, Fuad E. Alsaadi

PII: S0925-2312(17)30314-4
DOI: 10.1016/j.neucom.2017.01.090
Reference: NEUCOM 18110

To appear in: *Neurocomputing*

Received date: 7 October 2016
Revised date: 20 January 2017
Accepted date: 20 January 2017

Please cite this article as: Nianyin Zeng, Hong Zhang, Weibo Liu, Jinling Liang, Fuad E. Alsaadi, A Switching Delayed PSO Optimized Extreme Learning Machine for Short-Term Load Forecasting, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.01.090

A Switching Delayed PSO Optimized Extreme Learning Machine for Short-Term Load Forecasting

Nianyin Zeng*, Hong Zhang, Weibo Liu, Jinling Liang and Fuad E. Alsaadi

Abstract

In this paper, a hybrid learning approach, which combines the extreme learning machine (ELM) with a new switching delayed PSO (SDPSO) algorithm, is proposed for the problem of the short-term load forecasting (STLF). In particular, the input weights and biases of ELM are optimized by a new developed SDPSO algorithm, where the delayed information of locally best particle and globally best particle are exploited to update the velocity of particle. By testing the proposed SDPSO-ELM in a comprehensive manner on a *tanh* function, this approach obtain better generalization performance and can also avoid adding unnecessary hidden nodes and overtraining problems. Moreover, it has shown outstanding performance than other state-of-the-art ELMs. Finally, the proposed SDPSO-ELM algorithm is successfully applied to the STLF of power system. Experiment results demonstrate that the proposed learning algorithm can get better forecasting results in comparison with the radial basis function neural network (RBFNN) algorithm.

Index Terms

Short-term load forecasting; extreme learning machine; switching delayed particle swarm optimization (SDPSO); neural network; time-delay.

I. INTRODUCTION

Load forecasting, which aims to predict the future load demand with satisfactory accuracy, plays an importance role in the generation scheduling, system reliability and power optimization and economical running of the smart grid. Based on the prediction time, load forecasting can be divided into three types, which are short-term, medium-term and long-term forecasting [27]. In this paper, we focus on the problem of the short-term load forecasting (STLF), which generally refers to the period of its prediction from one hour to one week.

In the past few years, many efforts have been made, and quite a considerable number of methods have been proposed for STLF. Especially, recent research has been going mainly toward two categories: one is the statistical methods, see, e.g., [1], [3], [21], and the other is artificial intelligence methods, see, e.g., [2], [9], [16], [17], [27]. Among various forecasting approaches, artificial neural networks (ANNs) have become the popular ones for the STLF due primarily to its attractive properties such as strong ability

This work was supported in part by the Natural Science Foundation of China under Grants 61403319 and 61673110, in part by the Fujian Natural Science Foundation under Grant 2015J05131, in part by the Six Talent Peaks Project for the High Level Personnel from the Jiangsu Province of China under Grant 2015-DZXX-003, in part by the Fujian Provincial Key Laboratory of Medical Instrumentation and Pharmaceutical Technology, and also Fujian Provincial Key Laboratory of Eco-Industrial Green Technology.

N. Zeng and H. Zhang are with the Department of Instrumental and Electrical Engineering, Xiamen University, Fujian 361005, China. Email address: zny@xmu.edu.cn

W. Liu is with the Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH, United Kingdom.

J. Liang is with the School of Mathematics, Southeast University, Nanjing 210096, China. She is also with the Communication Systems and Networks (CSN) Research Group, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia.

F. E. Alsaadi is with the Communication Systems and Networks (CSN) Research Group, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia.

*Corresponding author.

in pattern recognition, fault tolerance and distributed associative memory [27]. Up to date, a variety of ANN-based algorithms have been under especial intensive investigations for the STLF problem [2], [9], [16], [17], [27]. However, studies have shown that those ANN-based algorithms which are trained by gradient-based methods such as back propagation (BP) or its variants have some limitations when applied to a variety of complex fields like power load forecasting [2], [9], [27].

Recently, a powerful learning algorithm called extreme learning machine (ELM) was proposed by Huang in 2004 [4] for single-hidden-layer feedforward neural networks (SLFNs). The ELM, which can overcome some limitations faced by gradient-based methods, has been gaining particular research attention because of fast learning speed and competitive generalization performance when applied to real-word problems [5], [7], [9], [18], [26], [27]. With rapid development of smart grid, large volumes of data can be acquired and applied to the STLF of power system using ELMs. Notably, unnecessary hidden neurons may be added and ill-condition problems may occur when the input weights and hidden biases are selected randomly in the ELM [9], [27]. Hence, with maintained performance, the evolutionary algorithm, particularly the particle swarm optimization (PSO) algorithm is chosen to select the optimum input weights and hidden biases of the ELM.

Inspired by the swarm behaviours of birds flocking or fish schooling [8], PSO is a swarm intelligence-based optimization algorithm which has been widely used in many complex optimization problems because of easily implementation. However, the basic PSO algorithm suffers from certain shortcomings such as easily trapped in the local optimum and appeared premature convergence [20]. Hence, a great number of strategies have been introduced into the basic PSO algorithm to improve its performance, see e.g. [20], [24], [28]–[30]. In particular, a novel switching delayed PSO (SDPSO) algorithm, which can adjust the model according to an evolutionary factor and a Markov chain has been recently introduced and analyzed in [30], where the delayed information of locally best particle and globally best particle are utilized to update the velocity of particle. It has been proven that the SDPSO can not only enhance the ability of global searching but also improve the possibility of eventually reaching the global optimum solution.

Based on the above discussions, a hybrid learning approach, which combines the ELM with a new SDPSO algorithm, is proposed for the STLF of power system in this paper. Especially, the SDPSO algorithm is exploited to optimize the hidden nodes parameters (including the input weights and biases) of ELM. Through comprehensive analysis of the proposed method, the results demonstrate that the performance of the SDPSO-ELM algorithm is superior over other competitive learning algorithms. The main contributions of this paper can be summarized as follows. *(1) A hybrid learning algorithm called SDPSO-ELM is proposed for the problem of STLF. The hidden nodes parameters of ELM algorithm are optimized by the new developed SDPSO algorithm. Hence, the SDPSO-ELM can avoid adding unnecessary hidden nodes and overtraining problems, and also improve the forecasting accuracy. (2) The SDPSO-ELM is tested in a comprehensive and systematic manner on a tanch function and its performance outperforms other state-of-the-art ELMs significantly in terms of generalization performance. (3) The novel SDPSO-ELM algorithm is successfully applied to the STLF of power system. The results demonstrate that the proposed method can get better forecasting results in comparison with the competitive algorithms.*

The rest of this paper is organized as follows. In section II, we presents a detailed introduction on the extreme learning machine and the new developed switching delayed particle swarm optimization algorithm. A hybrid learning method called SPSO-ELM is proposed in Section III. In Section IV, we successfully implement the proposed SDPSO-ELM approach to the STLF for electric power system and a series of experiments are carried out to demonstrate the effectiveness of the approach. Finally, conclusions are summarized in Section V.

II. Preliminaries

In this section, we mainly introduce the extreme learning machine (ELM) and the new developed switching delayed particle swarm optimization (SDPSO) algorithm.

A. Extreme Learning Machine

The single-hidden-layer feedforward neural networks (SLFNs), as shown in Fig. 1, is a three-layer neural network with the ability of approximating complex nonlinearity of the data, which has been extensively investigated and widely applied in the past few decades [23], [26]. Employed in generalized SLFNs, ELM, a novel learning algorithm with highly competitive performance has attracted broad attentions [7].

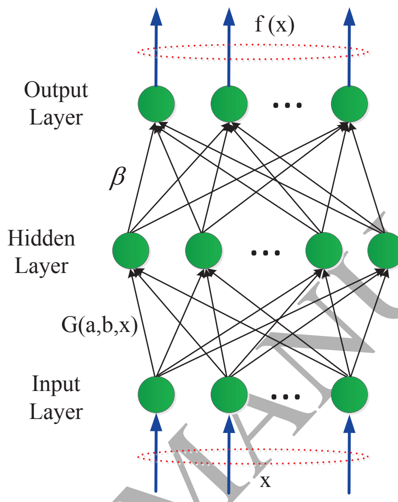


Fig. 1. Schematic diagram of a SLFN.

For a SLFN (d input nodes, L hidden nodes and m output nodes), the output function of the SLFN can be denoted as [26]:

$$f_L(x) = \sum_{i=1}^{L} \beta_i g_i(x) = \sum_{i=1}^{L} \beta_i G(a_i, b_i, x), \quad x \in R^d, \beta_i \in R^m \tag{1}$$

where g_i and $G(a_i, b_i, x)$ represent the activation function of ith hidden node. (a_i, b_i) is the hidden layer parameters which connect the hidden neurons and input neurons, and β_i is the output weight which connects the hidden neurons and output neurons. For additive hidden nodes, the activation function g_i can be described as:

$$g_i = G(a_i, b_i, x) = g(a_i x + b_i), \quad a_i \in R^d, b_i \in R \tag{2}$$

And for RBF hidden nodes, the activation function g_i can be described as:

$$g_i = G(a_i, b_i, x) = g(b_i \|x - a_i\|), \quad a_i \in R^d, b_i \in R^+ \tag{3}$$

Suppose that the training set consists of N samples $\{(x_i, t_i)\}_{i=1}^{N}$, where $x_i = [x_{i1}, x_{i2}, ..., x_{id}]^T \in R^d$ and $t_i = [t_{i1}, t_{i2}, ..., t_{im}]^T \in R^m$, theoretically, the SLFNs can approximate these N samples with zero error, which can be represented as:

$$\sum_{j=1}^{N} \|y_j - t_j\| = 0 \tag{4}$$

where y denotes the actual output of the SLFN. Therefore, there exist parameters (a_i, b_i) and β_i satisfying that:

$$\sum_{i=1}^{L} \beta_i G(a_i, b_i, x_j) = t_j, \quad j = 1, 2, ..., N \tag{5}$$

The matrix form of the above N equations is briefly represented by:

$$H\beta = T \tag{6}$$

where

$$H = \begin{bmatrix} h(x_1) \\ h(x_2) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} G(a_1, b_1, x_1) \cdots G(a_L, b_L, x_1) \\ G(a_1, b_1, x_2) \cdots G(a_L, b_L, x_2) \\ \vdots \\ G(a_1, b_1, x_N) \cdots G(a_L, b_L, x_N) \end{bmatrix}_{N \times L} \tag{7}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \quad and \quad T = \begin{bmatrix} t_1^T \\ t_2^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \tag{8}$$

Unlike the traditional learning algorithms, the highlight of ELM is that the hidden layer parameters (a_i, b_i) are generated randomly and kept fixed instead of tuned by iteration. In ELM, the most important step is to figure out the output wights β_i, which can be solved by minimizing the training error as well as the norm of the output weights. One thing to note here is that the generalization performance of SLFNs tends to get better if the training error and norms of weights are smaller. The objective function can be described as:

$$F = \min_{\beta} \|H\beta - T\|^2 + \lambda \|\beta\|^2 \tag{9}$$

The method of smallest norm least-squares is used to solve the above problem:

$$\beta = H^{\dagger} T \tag{10}$$

where $H^{\dagger}$ represents the Moore-Penrose generalized inverse of matrix H.

Here, various of approaches can be adopted to compute $H^{\dagger}$, such as the orthogonal projection approach, orthogonalization approach, the iterative approach, and the singular value decomposition approach (SVD). Especially, the orthogonal projection method can be efficiently used in two cases: $H^{\dagger} = (H^T H)^{-1} H^T$ when $H^T H$ is nonsingular, or $H^{\dagger} = H^T (H H^T)^{-1}$ when $H H^T$ is nonsingular. Based on the ridge regression theory, the stability and generalization performance of SLFNs can be improved by adding a positive value to the diagonal of $H^T H$ or $H H^T$. In this way, the output wights β can be calculated by:

$$\beta = (\frac{\boldsymbol{I}}{\lambda} + H^T H)^{-1} H^T T \qquad \text{if } H^T H \text{ is nonsingular} \tag{11}$$

or

$$\beta = H^T (\frac{\boldsymbol{I}}{\lambda} + H H^T)^{-1} T \quad \text{if } H H^T \text{ is nonsingular} \tag{12}$$

B. Switching Delayed Particle Swarm Optimization Algorithm

PSO, first proposed by Kennedy and Eberhart in 1995 [8], is a global optimization algorithm motivated by the group behaviors in nature such as fish schooling or birds flocking, etc. The basic PSO and its variants have been successfully applied to various real-word applications due to its effectiveness in performing difficult optimization tasks and its convenience for implementation with fast convergence to a reasonably good solution [20], [24], [28], [30].

Recently, a novel switching delayed PSO (SDPSO) algorithm is developed and analyzed in [30], in which the update formula of the velocity adaptively changes based on an evolutionary factor and the Markov chain. Furthermore, the delayed information of $pbest$ and $gbest$ are employed to the update formula of the velocity in accordance with different evolutionary states. The SDPSO algorithm can effectively overcome the problem of local optimum and premature convergence, and it is especially proficient in handling with the high-dimensional and multi-modal problems.

The velocity and position equations of the new developed SDPSO algorithm are given as follows:

$$\begin{aligned} v_{i,j}(k+1) = \quad & w(k)v_{i,j}(k) + c_1(\xi(k))r_1(p_{i,j}(k-\tau_1(\xi(k))) - x_{i,j} \\ & + c_2(\xi(k))r_2(p_{g,j}(k-\tau_2(\xi(k)) - x_{i,j}(k)), \\ x_{i,j}(k+1) = \quad & x_{i,j}(k) + v_{i,j}(k+1), \end{aligned} \tag{13}$$

where $c_1(\xi(k))$ and $c_2(\xi(k))$ are the acceleration coefficients, $\tau_1(\xi(k))$ and $\tau_2(\xi(k))$ represent the time-delay, respectively. Especially, these four parameters are decided by a non-homogeneous Markov chain $\xi(k)$ $(k \geq 0)$, which takes values in a finite state space $\mathcal{S} = \{1, 2, \cdots, N\}$ according to the probability transition matrix $\Pi^{(k)} = (\pi_{ij}^{(k)})_{N\times N}$, where $\pi_{ij}^{(k)} \geq 0$ $(i, j \in \mathcal{S})$ and $\sum_{j=1}^{N} \pi_{ij}^{(k)} = 1$.

The searching process towards to the global optimum can be divided into four states: convergence, exploration, exploitation and jumping out, which are denoted by $\xi(k) = 1, \xi(k) = 2, \xi(k) = 3$ and $\xi(k) = 4$, respectively [20], [30]. Particularly, these four states can be discriminatively defined by an evolutionary factor (EF) which can be defined as follows [24], [30]:

$$E_f = \frac{d_g - d_{\min}}{d_{\max} - d_{\min}} \tag{14}$$

where d_g represents the globally best particle among the mean distance d_i, which represents the distance between each particle and the other particles. $d_{\max}$ and $d_{\min}$ represent the maximum and minimum distances in d_i, respectively.

According to the value of E_f, the state of Markov chain can be determined by:

$$\xi(k) = \begin{cases} 1, & 0 \leq E_f < 0.25, \\ 2, & 0.25 \leq E_f < 0.5, \\ 3, & 0.5 \leq E_f < 0.75, \\ 4, & 0.75 \leq E_f < 1, \end{cases} \tag{15}$$

where the probability transition matrix is obtained as follows:

$$\Pi = \begin{pmatrix} \chi & 1-\chi & 0 & 0 \\ \frac{1-\chi}{2} & \chi & \frac{1-\chi}{2} & 0 \\ 0 & \frac{1-\chi}{2} & 0 & \frac{1-\chi}{2} \\ 0 & 0 & 1-\chi & \chi \end{pmatrix} \tag{16}$$

Therefore, the Markov chain may change from one state to another or remain the same state during the iteration, which is depended on the value of probability transition matrix Π.

TABLE I
STRATEGIES FOR SELECTING c_1 , c_2 AND DELAYED INFORMATION

State	Mode	c_1	c_2	*pbest*	*gbest*	$\tau_1(\xi(k))$	$\tau_2(\xi(k))$
Convergence	$\xi(k) = 1$	2	2	$p_i(k)$	$p_g(k)$	0	0
Exploitation	$\xi(k) = 2$	2.1	1.9	$p_i(k - \tau_1(\xi(k)))$	$p_g(k)$	$\lfloor k \cdot rand_1 \rfloor$	0
Exploration	$\xi(k) = 3$	2.2	1.8	$p_i(k)$	$p_g(k - \tau_2(\xi(k)))$	0	$\lfloor k \cdot rand_2 \rfloor$
Jumping-out	$\xi(k) = 4$	1.8	2.2	$p_i(k - \tau_1(\xi(k)))$	$p_g(k - \tau_2(\xi(k)))$	$\lfloor k \cdot rand_1 \rfloor$	$\lfloor k \cdot rand_2 \rfloor$

In addition, the delayed information of *pbest* and *gbest* are exploited to update the velocity equation according to the evolutionary state. The methods for selecting the acceleration coefficients (c_1 and c_2) and delayed information are given in the Table I. Furthermore, the inertia weight w has similar tendency as the evolutionary factor E_f during the iteration process, which can be expressed as follows [20], [30]:

$$w(E_f) = 0.5E_f + 0.4 \in [0.4, 0.9], \forall E_f \in [0, 1]. \tag{17}$$

III. SWITCHING DELAYED PSO OPTIMIZED EXTREME LEARNING MACHINE (SDPSO-ELM)

In ELM, the parameters of input weights and hidden biases are randomly generated instead of using the conventional tuning-based approaches, which saves much training time. Nevertheless, as the output weights are calculated on the basis of input weights and hidden biases, there inevitably exists some nonoptimal or unnecessary input weights and hidden biases. Generally, the approach of randomly selecting parameters may cause two main issues. On one hand, the ELM may need more hidden neurons than the traditional tuning-based neural networks, which can lead to a slow response to the unknown testing data. On the other hand, it tends to appear an ill-conditioned hidden output matrix H, especially in the case that a great number of hidden neurons are used, which may result in a worse generalization performance.

A condition number is introduced to qualitatively describe the condition of a matrix [25]. In general, a small condition number is expected because it stands for good conditioning. On the contrary, a large one may indicate that the matrix is close to ill-conditioned. Based on the literatures, the 2-norm condition number of the matrix H can be defined by [25]:

$$K_2(H) = \sqrt{\frac{\lambda_{max}(H^T H)}{\lambda_{min}(H^T H)}} \tag{18}$$

where $\lambda_{max}(H^T H)$ and $\lambda_{min}(H^T H)$ represent the the largest and smallest eigenvalues of the matrix $H^T H$, respectively.

In order to overcome the above mentioned limitations, a hybrid method, named SDPSO-ELM, which combines the SDPSO algorithm with an ELM is proposed in this paper. In particular, the SDPSO algorithm is used to optimize the parameters of input weights and hidden biases of ELM.

Firstly, initialize particles of the swarm and the parameters of the SDPSO. Each particle formed by a set of inputs weights and hidden biases is represented by $\theta_i = [w_{11}, w_{12}, ..., w_{1K}, w_{21}, w_{22}, ..., w_{2K}, ..., w_{n1}, w_{n2}, ..., w_{nK}, ..., b_1, b_2, ..., b_K]$, where w_{ij} and b_j are randomly initialized from [-1,1].

Secondly, calculate the corresponding output weights of particles according to the Eqn. (10) and evaluate the fitness value of each particle. In order to avoid over-fitting of the SLFN, the fitness value of each particle is calculated by the root mean square error (RMSE) on the validation set rather than the whole training set. The fitness function is defined as follows:

$$f() = \sqrt{\frac{\sum_{j=1}^{n_\nu} \| \sum_{i=1}^{K} \beta_i g(w_i \cdot x_j + b_i) - t_j \|_2^2}{n_\nu}} \tag{19}$$

where n_ν is the number of the validation samples.

Thirdly, determine the the local best $pbest$ for all particles and the globally best $gbest$ for the swarm. With the mean distance value of each particle, the evolutionary factor can be figured out by the Eqn. (14), and then the state in the next generation is updated based on the current state and the probability transition matrix. Subsequently, the inertia weight is calculated by the Eqn. (17), as well as the acceleration coefficients and the $pbest$, $gbest$ are selected according to the Table I. Especially, neural networks are apt to have better generalization performance with smaller norm of weights [23]. Hence, for the sake of further enhancing the generalization performance of SLFN, the norm of output weights as well as the RMSE are both taken into account to determine the local best $pbest$ for all particles and the global best $gbest$ for the swarm. The corresponding equations are denoted as follows [23]:

$$p_{ib} = \begin{cases} p_i & (f(p_{ib}) - f(p_i) \geq \varepsilon f(p_{ib})) \text{ or } (|f(p_{ib}) - f(p_i)| < \varepsilon f(p_{ib}) \text{ and } \|\beta_{p_i}\| < \|\beta_{p_{ib}}\|) \\ p_{ib} & \text{else} \end{cases} \tag{20}$$

$$p_{g} = \begin{cases} p_i & (f(p_{g}) - f(p_i) \geq \varepsilon f(p_{g})) \text{ or } (|f(p_{g}) - f(p_i)| < \varepsilon f(p_{g}) \text{ and } \|\beta_{p_i}\| < \|\beta_{p_{g}}\|) \\ p_{g} & \text{else} \end{cases} \tag{21}$$

where $f(p_i)$, $f(p_{ib})$ and p_g are the corresponding fitness values for the ith particle, the best position of ith particle and global best position of the swarm, respectively. β_{p_i}, $\beta_{p_{ib}}$ and β_{p_g} are the corresponding output wights obtained by Eqn. (10) when the parameters of input weights and hidden biases are set as the ith particle, the best position of ith particle and global best position of the swarm, respectively. $\varepsilon > 0$ is the tolerance rate.

Fourthly, each particle updates the velocity and position according to Eqn. (13), and the next generation is produced.

Finally, the above iteration process continues until the optimal solution or the maximum iteration is reached. Therefore, the optimal parameters of input weights and hidden biases are achieved, and eventually the corresponding ELM with the optimal parameters can be utilized to the testing data.

IV. Experimental Results and Discussions

In this section, a series of experiments are carried out to demonstrate the effectiveness of the proposed SDPSO-ELM approach. To illustrate the superiority of the SDPSO-ELM, three well-known ELM algorithms are compared with the SDPSO-ELM algorithm which consists of the IPSO-ELM algorithm [6], the E-ELM algorithm [23] and the basic ELM [4]. At first, the four algorithms are used to reconstruct a given nonlinear function, in order to discuss their properties in all respects. And then, we try to apply the proposed SDPSO-ELM approach to the STLF for electric power system.

The parameters of four algorithms in the experiments are set as: for ELM, the hidden nodes $L = 30$; for the other three algorithms, the hidden nodes $L = 10$, the particle number $N = 20$, the tolerance rate for fitness function $\varepsilon = 0.1$, and the maximum iteration number is set to 200. To obtain the mean results, the experiment is repeated 50 trails independently for each algorithm in order to avoid the influence of random factor.

TABLE II
THE APPROXIMATION PERFORMANCE OF FUNCTION *tanh* WITH FOUR ELMS.

Algorithms	Training RMSE	Testing RMSE	Testing Dev	Training time	Hidden nodes	Norm	Condition
ELM	0.0938	0.0157	0.0031	0.0031	30	3.8330e+09	3.6579e+16
E-ELM	0.0949	0.0144	0.0075	4.8541	10	1.0245e+03	4.16604e+31
IPSO-ELM	0.0944	0.0121	0.0043	5.5040	10	4.4097e+03	7.56568e+15
SDPSO-ELM	0.0943	0.0118	0.0029	8.7688	10	1.3820e+03	4.0069e+05

A. *Function Approximation*

Here, the *tanh* function is selected for the four algorithms to approximate, which is represented as:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{22}$$

The training set and testing set are formed by 1000 data, respectively, where the inputs are uniformly distributed in [-5,5]. In addition, a large uniform noise distributed in [-0.2,0.2] is added to the training data while the testing data remain clean. The corresponding results are shown in Table II and Fig. 2.

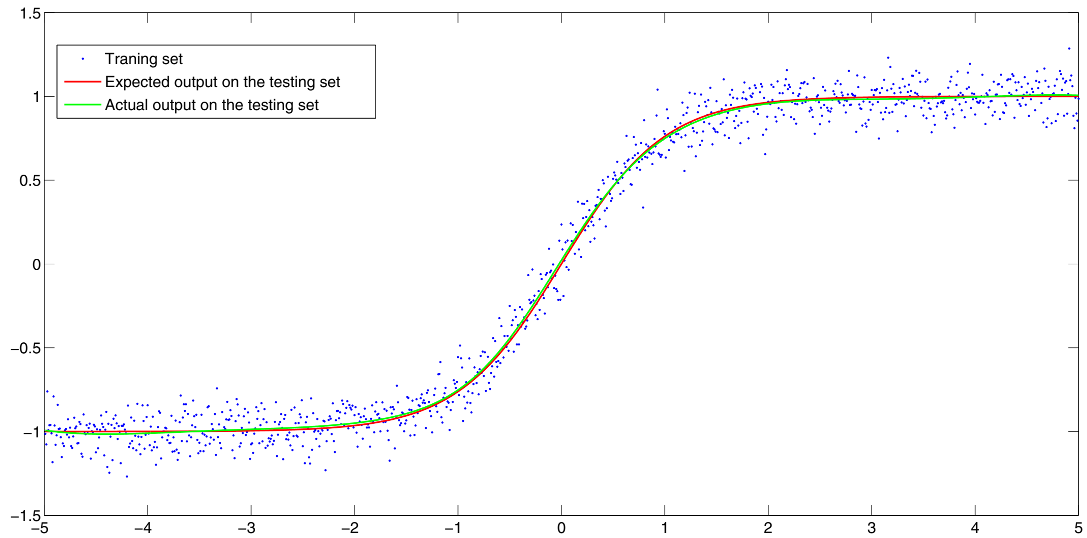
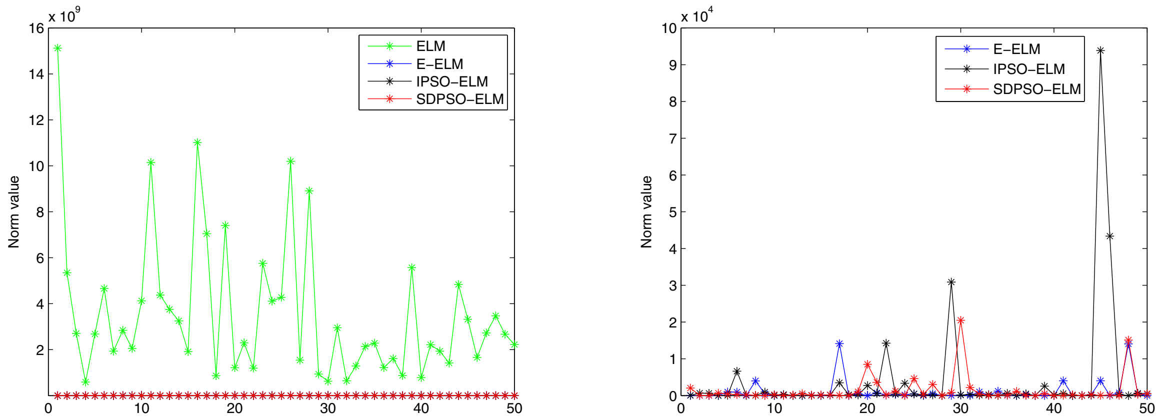


Fig. 2. The output of the SDPSO-ELM algorithm.

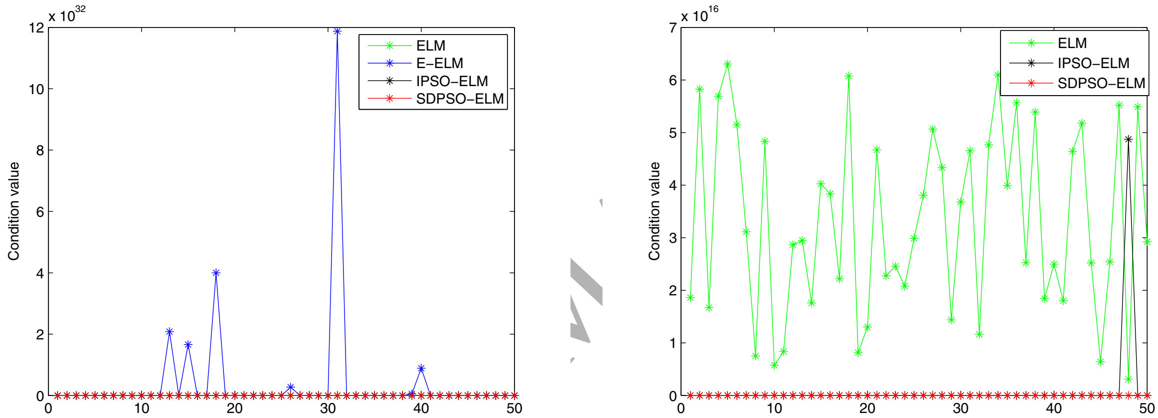
As shown in Fig. 2, the output of the SDPSO-ELM algorithm for "tanh" function approximation is extremely close to the expected value, which demonstrates the effectiveness of the proposed SDPSO-ELM approach.

From the Table II, we firstly observe that all three improved ELMs of SDPSO-ELM, IPSO-ELM and E-ELM achieve smaller testing RMSE than the ELM, where the SDPSO-ELM achieves the best performance. Meanwhile, only 10 hidden neurons are assigned for the three improved ELMs, while 30 hidden neurons are used in ELM. Therefore, a more compact SLFN with good generalization performance can be obtained by the three improved ELMs. Moreover, the norm values of the output weights in the E-ELM, IPSO-ELM and SDPSO-ELM are much smaller than the ELM, which further confirms that these



(a) The norm value of the output weights obtained by four ELMs (b) The norm value of the output weights obtained by three ELMs

Fig. 3. The norm value of the output weights obtained by ELMs



(a) The condition value of SLFN obtained by four ELMs (b) The condition value of SLFN obtained by three ELMs

Fig. 4. The condition value of SLFN obtained by ELMs

three improved ELMs enhance the SLFN generalization performance. However, the condition values of IPSO-ELM and E-ELM are not so satisfactory, for the reason that a few sharp spikes are contained in the total 50 trials. In spite of this, the SDPSO-ELM also achieves a stable and low condition value of SLFN. In addition, the standard deviation (Dev) of SDPSO-ELM and ELM are much better than the Dev of E-ELM and IPSO-ELM, possibly because of the volatility of 50 trials in E-ELM and IPSO-ELM. Note that the three improved ELMs sacrifice more times for selecting the parameters of input weights and basis. In conclusion, the proposed SDPSO-ELM algorithm outperforms the other three algorithms of IPSO-ELM, E-ELM and ELM, in terms of the RMSE, Dev, norm and condition value.

Furthermore, as shown in Fig. 3 and Fig. 4, the norm of the output weights and condition values of four ELMs are analyzed respectively. From Fig. 3(a), we notice that the algorithms of SDPSO-ELM, IPSO-ELM and E-ELM achieve much lower value of the norm of the output weights than the ELM in each trial. Similarly, as observed in Fig. 4(a), the condition values of SLFN obtained by the SDPSO-ELM, IPSO-ELM and ELM are much smaller and more steady than those of the E-ELM.

Especially, in order to provide a proper performance comparison among the algorithms, we delete the worst performance of four ELMs. Then, the results are shown in Fig. 3(b) and Fig. 4(b), from which we can obviously observe the improvements on the condition and the norm of the output weights of the SLFN. As Fig. 3(b) shows, the curves of norm value for E-ELM and SDPSO-ELM are more stable, while the curve of IPSO-ELM contains several sharp spikes. From Fig. 4(b), the SDPSO-ELM also achieves the most stable performance of condition value among the three algorithms of SDPSO-ELM, IPSO-ELM and ELM.

B. Short-term Load Forecasting for Electric Power System

As a basic but indispensable task of power system, STLF generally predicts the load for a lead time ranging from one hour to several days, which plays an important role in the generation scheduling and reserving activities. In this section, the proposed SDPSO-ELM algorithm is applied to the short-term load forecasting of power system, where the 24 hourly loads of the forecasting day are required to make prediction.

1) Architecture: The architecture of the proposed SDPSO-ELM based STLF is shown in Fig. 5. As shown, the SDPSO algorithm is introduced to optimize the parameters of input weights and hidden biases of ELM, where the norm of output weights and the RMSE on the validation set are both considered as the criterion. The combination of the SDPSO and ELM makes the selection of input weights and hidden biases in ELM more reasonable and effective, instead of the conventional randomly generating. With the optimized parameters, the SDPSO-ELM algorithm achieves stronger generalization performance as well as better condition of the SLFN. Therefore, the forecasting results with high accuracy are obtained using the SDPSO-ELM when the testing data inputs.

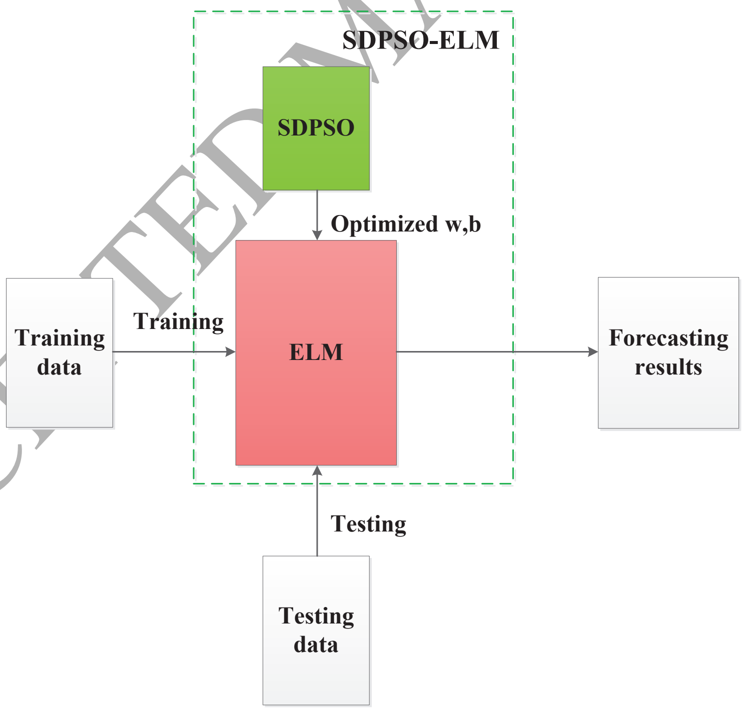


Fig. 5. Architecture of proposed SDPSO-ELM based STLF model.

In particular, in order to obtain more precise forecasting data, the STLF model is designed as the combination of the prediction results of 24 hours, which means the load of each hour in the day is

TABLE III
SELECTION OF INPUT AND OUTPUT VARIABLES

Input neurons	Input variables
1	the load of hour t in the day $k-7$
2	the load of hour t in the day $k-2$
3	the load of hour t in the day $k-1$
4	the load of hour $t-1$ in the day $k-1$
5	the load of hour $t-2$ in the day $k-1$
6	the day type of the day $k-1$
7	the day type of the day k
8	the weather condition of the day $k-1$
9	the weather condition of the day k
10	the temperature of the day $k-1$
11	the temperature in the day k
Output neuron	Output variable
1	the load of hour t in the day k

predicted by the SDPSO-ELM, independently. For the load of hour t in the day k, the input variables are chosen according to the correlation studies and literatures, as shown in Table III.

As for the performance indicator, mean absolute percentage error (MAPE) and mean absolute error (MAE) are introduced, which are computed as follows:

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|Y_k(i) - \hat{Y_k}(i)|}{Y_k(i)} \times 100 \tag{23}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|Y_k(i) - \hat{Y_k}(i)| \tag{24}$$

where $Y_k(i)$ and $\hat{Y_k}(i)$ are the real and prediction load value of hour i, respectively.

2) Test results: The historical load data from January to June in 2010 of the Ningde City in Fujian Province of China are analyzed using the proposed SDPSO-ELM approach. The data from January to May are used as the training set, and then we forecast the hourly load on any one day in June. Furthermore, a comparison between the SDPSO-ELM and the state-of-the-art ANN method, radial basis function neural network (RBFNN), is conducted in order to demonstrate the performance of the proposed method. It should be pointed out that under the same circumstance, we can easily confirm the parameters of SDPSO-ELM while requiring much more time to tune the parameters of the RBFNN.

The forecasting results are shown in the Fig. 6 and the generalization performances are listed in the Table IV.

Fig. 6 describes that the output of the SDPSO-ELM algorithm for STLF is obviously more approximate to the real load than that of the RBFNN, which indicates the proposed SDPSO-ELM can efficiently forecast the hourly load with remarkable accuracy.

Meanwhile, from Table IV, the MAPE obtained by the SDPSO-ELM decreases 0.72% compared with the MAPE of RBFNN. It is of great significance to improve the prediction accuracy for the power system generation scheduling and reserving activities. Similarly, the SDPSO-ELM algorithm achieves better MAE as much as 7.566MW than the MAE of RBFNN.

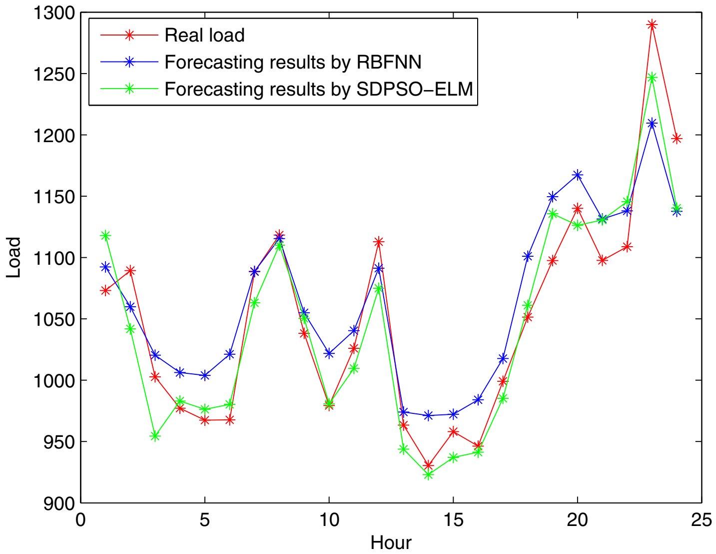


Fig. 6. Forecasting results of two methods.

TABLE IV
GENERALIZATION PERFORMANCE

Hour	Real load (MW)	RBFNN		SDPSO-ELM	
		Forecasting load (MW)	Error (%)	Forecasting load (MW)	Error (%)
00:00	1073.183	1092.301	1.782	1117.925	4.170
01:00	1089.324	1059.841	2.706	1041.834	4.360
02:00	1002.713	1020.360	1.760	954.403	4.781
03:00	977.133	1006.232	2.978	983.083	0.609
04:00	967.507	1003.989	3.770	976.336	0.913
05:00	967.713	1021.168	5.524	980.339	1.305
06:00	1088.405	1088.747	0.032	1063.158	2.320
07:00	1118.365	1115.456	0.261	1109.988	0.749
08:00	1038.160	1054.954	1.618	1050.364	1.108
09:00	979.343	1021.839	4.340	980.662	0.135
10:00	1026.024	1040.448	1.406	1009.597	1.601
11:00	1112.799	1091.302	1.932	1074.986	3.398
12:00	963.442	974.023	1.099	943.827	2.036
13:00	930.390	971.187	4.385	923.009	0.793
14:00	958.153	972.280	1.475	937.127	2.194
15:00	946.315	984.055	3.988	941.411	0.518
16:00	999.088	1017.660	1.859	985.235	1.387
17:00	1051.502	1101.045	4.712	1061.113	0.914
18:00	1097.499	1149.633	4.750	1135.738	3.484
19:00	1140.066	1167.394	2.397	1126.263	1.211
20:00	1097.639	1131.422	3.078	1130.619	3.005
21:00	1108.865	1137.994	2.626	1145.627	3.315
22:00	1289.926	1209.635	6.225	1246.893	3.336
23:00	1196.944	1137.627	4.955	1140.290	4.733
MAPE (%)		2.902		2.182	
MAE (MW)		30.496		22.930	

V. Conclusions

In this paper, we have presented a hybrid learning approach for the problem of the STLF based on an improved ELM which is optimized by a new developed switching delayed PSO algorithm. That is, the input weights and biases of ELM are optimized by the SDPSO algorithm. The proposed SDPSO-ELM is firstly verified on a *tanh* function, which shows that the performance of the proposed method is superior to other popular ELMs. Finally, the SDPSO-ELM is successfully applied to the STLF of power system. Experiment results show that the proposed method can significantly improve the forecasting accuracy in comparison with the RBFNN algorithm. In the near feature, some latest adaptively control strategies (e.g. [11], [13]–[15], [19], [22], [32], [33]) will be exploited for further improving the performance of the ELM algorithm, and also some advanced computational intelligent methods (e.g. [10], [12], [31]) will be applied to the problem of the STLF.

References

[1] N. Amjady, Short-term hourly load forecasting using time-series modeling with peak load estimation capability, *IEEE Transactions on Power Systems*, vol. 16, no. 3, pp. 498-505, 2001.

[2] D. Chaturvedi, A. Sinha, O. MalikShort term load forecast using fuzzy logic and wavelet transform integrated generalized neural network, *Electrical Power and Energy Systems*, vol. 67, pp. 230-237, 2015.

[3] S. Huang, K. Shih, Short-term load forecasting via ARMA model identification including non-Gaussian process considerations, *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 673-679, 2003.

[4] G. B. Huang, Q. Zhu, C. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, *2004 International Joint Conference on Neural Networks*, vol. 2, pp. 985-990, 2004.

[5] G. B. Huang, L. Chen and C. Siew, Universal approximation using incremental constructive feedforward networks with random hidden nodes, *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879-892, 2006.

[6] F. Han, H. Yao, Q. Ling, An improved evolutionary extreme learning machine based on particle swarm optimization, *Neurocomputing*, vol. 116, pp. 87-93, 2013.

[7] G. B. Huang, An Insight into Extreme learning machines: random neurons, random features and kernels, *Cognitive Computation*, vol. 6, no. 3, pp. 376-390, 2014.

[8] J. Kennedy and R. Eberhart, Particle swarm optimization, *Proc. IEEE International Conference On Neural Network*, pp. 1942-1948, 1995.

[9] S. Li, P. Wang, L. Goel, A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection, *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1788-1798, 2016.

[10] S. Liu, G. Wei, Y. Song, and Y. Liu, Extended Kalman filtering for stochastic nonlinear systems with randomly occurring cyber attacks, *Neurocomputing*, vol. 207, pp. 708–716, 2016.

[11] Y. Liu, W. Liu, M. A. Obaid, and I. A. Abbas, Exponential stability of Markovian jumping Cohen-Grossberg neural networks with mixed mode-dependent time-delays, *Neurocomputing*, vol. 177, pp. 409–415, 2016.

[12] W. Li, G. Wei, F. Han, and Y. Liu, Weighted average consensus-based unscented Kalman filtering, *IEEE Transactions on Cybernetics*, vol. 46, no. 2, pp. 558–567, 2016.

[13] S. Liu, G. Wei, Y. Song, and Y. Liu, Error-constrained reliable tracking control for discrete time-varying systems subject to quantization effects, *Neurocomputing*, vol. 174, pp. 897–905, 2016.

[14] Q. Li, B. Shen, Y. Liu, and F. E. Alsaadi, Event-triggered H_∞ state estimation for discrete-time stochastic genetic regulatory networks with Markovian jumping parameters and time-varying delays, *Neurocomputing*, vol. 174, pp. 912–920, 2016.

[15] D. Liu, Y. Liu, and F. E. Alsaadi, A new framework for output feedback controller design for a class of discrete-time stochastic nonlinear system with quantization and missing measurement, *International Journal of General Systems*, vol. 45, no. 5, pp. 517–531, 2016.

[16] J. Nowotarski, B. Liu, R. Weron, T. Hong, Improving short term load forecast accuracy via combining sister forecasts, *Energy*, vol. 98, pp. 40-49, 2016.

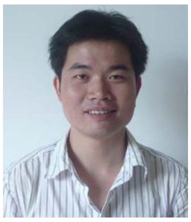
[17] T. Senjyu, H. Takara, K. Uezato, T. Funabashi, One-hour-ahead load forecasting using neural network, *IEEE Transactions on Power Systems*, vol. 17, no. 1, pp. 113-118, 2002.

[18] Z. Sun, T. Choi, K. Au, Y. Yu, Sales forecasting using extreme learning machine with applications in fashion retailing, *Decision Support Systems*, vol. 46, no. 1, pp. 411-419, 2008.

[19] H. Shu, S. Zhang, B. Shen, and Y. Liu, Unknown input and state estimation for linear discrete-time systems with missing measurements and correlated noises, *International Journal of General Systems*, vol. 45, no. 5, pp. 648–661, 2016.

[20] Y. Tang, Z. Wang, J. Fang, Parameters identification of unknown delayed genetic regulatory networks by a switching particle swarm optimization algorithm, *Expert Systems with Applications*, vol. 38, no. 3, pp. 2523-2535, 2011.

[21] S. Villalba, C. Bel, Hybrid demand model for load estimation and short-term load forecasting in distribution electrical systems. *IEEE Transactions on Power Delivery*, vol. 15, no. 2, pp. 764-769, 2000.
[22] C. Wen, Y. Cai, Y. Liu, and C. Wen, A reduced-order approach to filtering for systems with linear equality constraints, *Neurocomputing*, vol. 193, pp. 219–226, 2016.
[23] Q. Zhu, A. Qin, P. Suganthan, G. B. Huang, Evolutionary extreme learning machine, *Pattern Recognition*, vol. 38, no. 10, pp. 1759-1763, 2005.
[24] Z. Zhan, J. Zhang, Y. Li and H. Chung, Adaptive particle swarm optimization, *IEEE Transactions on System, Man and Cybernetics-B*, vol. 39, no. 6, pp. 1362-1381, 2009.
[25] G. Zhao, Z. Shen, C. Miao, Z. Man, On improving the conditioning of extreme learning machine: a linear aase, *7th International Conference on Information, Communications and Signal Processing, 2009*, pp. 1-5, 2009.
[26] G. Zhu, D. Wang, Y. Lan, Extreme learning machines: a survey, *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 2, pp. 107-122, 2011.
[27] R. Zhang, Z. Dong, Y. Xu, K. Meng, K. Wong, Short-term load forecasting of Australian National Electricity Market by an ensemble model of extreme learning machine, *IET Generation, Transmission & Distribution*, vol. 7, no. 4, pp. 391-397, 2013.
[28] N. Zeng, Z. Wang, B. Zineddin, Y. Li, M. Du, L. Xiao, X. Liu, and T. Young, Image-based quantitative analysis of gold immunochromatographic strip via cellular neural network approach, *IEEE Transactions on Medical Imaging*, vol. 33, no. 5, pp. 1129-1136, 2014.
[29] N. Zeng, Z. Wang, H. Zhang, Inferring nonlinear lateral flow immunoassay state-space models via an unscented Kalman filter, *SCIENCE CHINA Information Sciences*, vol. 59, no. 11, 112204, 2016.
[30] N. Zeng, Z. Wang, H. Zhang, F. Alsaadi, A novel switching delayed PSO algorithm for estimating unknown parameters of lateral flow immunoassay, *Cognitive Computation*, vol. 8, no. 2, pp. 143-152, 2016.
[31] N. Zeng, Z. Wang, H. Zhang, W. Liu, F. Alsaadi, Deep belief networks for quantitative analysis of gold immunochromatographic strip, *Cognitive Computation*, vol. 8, no. 4, pp. 684-692, 2016.
[32] W. Zhang, Z. Wang, Y. Liu, D. Ding, and F. E. Alsaadi, Event-based state estimation for a class of complex networks with time-varying delays: a comparison principle approach, *Physics Letters A*, vol. 381, no. 1, pp. 10–18, 2017.
[33] J. Zhang, L. Ma, and Y. Liu, Passivity analysis for discrete-time neural networks with mixed time-delays and randomly occurring quantization effects, *Neurocomputing*, vol. 216, pp. 657–665, 2016.



Nianyin Zeng was born in Fujian Province, China, in 1986. He received the B.Eng. degree in electrical engineering and automation in 2008 and the Ph. D. degree in electrical engineering in 2013, both from Fuzhou University. From October 2012 to March 2013, he was a RA in the Department of Electrical and Electronic Engineering, the University of Hong Kong.

Currently, he is an Assistant Professor with the Department of Instrumental & Electrical Engineering of Xiamen University. His current research interests include intelligent data analysis, computational intelligent, time-series modeling and applications. He is the author or co-author of several technical papers and also a very active reviewer for many international journals and conferences.

Dr. Zeng is currently serving as an Editorial board member for Biomedical Engineering Online (Springer), Journal of Advances in Biomedical Engineering and Technology, and Smart Healthcare.

Hong Zhang received her Bachelor's degree in electrical engineering and automation from the Department of Mechanical & Electrical Engineering, Xiamen University, Xiamen, China, in 2015. She is currently pursuing the Master's degree in Electrical Testing Technology and Instruments at Xiamen University, Xiamen, China. Her research interests include image processing and deep learning techniques.

Weibo Liu received his B. S. degree in electrical engineering from the Department of Electrical Engineering & Electronics, University of Liverpool, Liverpool, UK, in 2015. He is currently pursuing the Ph. D. degree in Computer Science at Brunel University London, London, UK. His research interests include big data analysis and deep learning techniques.

Jinling Liang received the B.Sc. and M.Sc. degrees in mathematics from Northwest University, Xian, China, in 1997 and 1999, respectively, and the Ph.D. degree in applied mathematics from Southeast University, Nanjing, China, in 2006. She is currently a Professor in the Department of Mathematics, Southeast University.

Prof. Liang has published around 80 papers in refereed international journals. She serves as an associate editor for several international journals such as IEEE Transactions on Neural Networks and Learning Systems, IET Control Theory & Applications, and International Journal of Computer Mathematics. She is also a member of the program committees of several international conferences and serves as a very active reviewer for many international journals. Her current research interests include stochastic systems, complex networks, robust filtering and bioinformatics.

Fuad E. Alsaadi received the B.S. and M.Sc. degrees in electronic and communication from King AbdulAziz University, Jeddah, Saudi Arabia, in 1996 and 2002. He then received the Ph.D. degree in Optical Wireless Communication Systems from the University of Leeds, Leeds, UK, in 2011. Between 1996 and 2005, he worked in Jeddah as a communication instructor in the College of Electronics and Communication. He is currently an associate professor of the Electrical and Computer Engineering Department within the Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia. He published widely in the top IEEE communications conferences and journals and has received the Carter award, University of Leeds for the best PhD. He has research interests in optical systems and networks, signal processing, synchronization and systems design.