A New Modified Elman Neural Network with Stable Learning Algorithms for Identification of Nonlinear Systems

Fatemeh Nejadmorad Moghanloo, Alireza Yazdizadeh and Amir Pouresmael Janbaz Fomani

Abstract In this paper a new dynamic neural network structure based on the Elman Neural Network (ENN), for identification of nonlinear systems is introduced. The proposed structure has feedbacks from the outputs to the inputs and at the same time there are some connections from the hidden layer to the output layer, so that it is called as Output to Input Feedback, Hidden to Output Elman Neural Network (OIFHO ENN). The capability of the proposed structure for representing nonlinear systems is shown analytically. Stability of the learning algorithms is analyzed and shown. Encouraging simulation results reveal that the idea of using the proposed structure for identification of nonlinear systems is feasible and very appealing.

Keywords Elman Neural Network · OIFHO ENN · Nonlinear System Identification

1 Introduction

In broad terms, the ultimate goal of system identification is to obtain a mathematical model whose output matches the output of a dynamic system for a given input. The solution to the exact matching problem, in general, is extremely difficult. Consequently, for practical reasons the original problem is relaxed to development of a model whose output can be made "as close as possible" to the output of the considered dynamic system. Different methods have been developed in recent years for linear/nonlinear system identification. A common characteristic of most of these methods is the use of a parameterized model where parameters are adjusted based on the minimization of a norm of the output identification error. These methods can be

F. Nejadmorad Moghanloo (✉) · A. Yazdizadeh · A. Pouresmael Janbaz Fomani
Department of Electrical Engineering, Abbaspour College of Technology,
Shahid Beheshti University, 16765-1719, Tehran, Iran
e-mail: f.nejadmorad@gmail.com

A. Yazdizadeh
e-mail: Alireza@pwut.ac.ir

A.P.J. Fomani
e-mail: a.pouresmael@gmail.com

R. Lee (ed.), *Computer and Information Science*, Studies in Computational Intelligence 566, DOI 10.1007/978-3-319-10509-3_13

classified into two main categories, namely, conventional and neural network-based methods [1]. Conventional methods are based on well established linear system theory and recently developed nonlinear system techniques. In most of existing methods, under certain conditions, desired characteristics such as convergence of the output error (identification error) to zero and stability of the identifier are shown analytically. The main disadvantage of these methods is that they are generally applicable and extendible to only a special class of nonlinear systems. In order to generalize these results to arbitrary classes of nonlinear systems, restrictive knowledge about the system is required [1].

Fortunately, the characteristics of the Artificial Neural Network (ANN) approach, namely nonlinear transformation, provide effective techniques for system identification, especially for non-linear systems. The ANN approach has a high potential for identification applications because: (1) it can approximate the nonlinear input–output mapping of a dynamic system; (2) it enables to model the complex system's behavior through training, without a priori information about the structures or parameters of systems. Due to these characteristics, there has been a growing interest, in recent years, in the application of neural networks to dynamic system identification and control [2–5].

1.1 Literature Review

Elman neural network (ENN) is a partial recurrent network model proposed by Elman in 1990 [6]. It lies somewhere between a classic feedforward perception and a pure recurrent network. The feedforward connection consists of the input layer, hidden layer, and output layer, in which the weights connecting two neighboring layers are variables. In contrast to the classical feedforward neural networks, the back forward connection employs context layer that is sensitive to the history of input data, therefore, the connections between the context layer and the hidden layer are fixed. Furthermore, since dynamic characteristics of Elman network is provided by internal connections, it does not need to use the state as input or training signals, which makes ENN superior to static feedforward network and is widely used in dynamic system identification [7].

There has been much research interest in Elman Neural Network [2, 8–15]. Elman Neural Network has been applied to dynamic system identification and financial prediction in [8, 10], respectively. A modified Elman Neural Network has been proposed by [4] because it was found that the basic Elman network trained by the standard Backpropagation (BP) algorithm was able to model only first-order dynamic systems. The performance of Elman's RNN has shown by means of two different applications in [14]. Song [15] focuses on the real-time online learning of an extended training algorithm for Elman Neural Network with a new Multiple-Input–Multiple-Output (MIMO) adaptive dead zone scheme and guaranteed weight convergence. Hsu [16] proposes an Elman-based self-organizing RBF Neural Network (ESRNN) for online approximation of the unknown nonlinear system dynamics based on a Lyapunov function and an Adaptive Backstepping Elman-based Neural Control (ABENC) system to eliminate the effect of the approximation error. Pham has described the dynamic BP (DBP) algorithm in [8] which is proper for training the basic Elman Neural Network and shows that the modified Elman Neural Network is an approximation of the Elman Neural Network trained by DBP. Pham has clarified why the modified Elman Neural Network can model higher-order dynamic systems. In [11] OHF and OIF Elman Neural Networks are presented for identification and control of ultrasonic motor. A hybrid Elman- NARX Neural Network is presented by [12] to analyze and predict chaotic time series. A new recurrent Neural Network based on the original Elman Neural Network is introduced in [2] to improve the resolution ratio of Elman Neural Network. Yuan Cheng has presented a new modified Elman Neural Network to improve the dynamic characteristics of the original Elman Neural Network [17]. A novel EMD–ENN approach, a hybrid of Empirical Mode Decomposition (EMD) and Elman neural network (ENN), is presented in [5] to forecast the wind speed. In this study, first, the original wind speed dataset are decomposed into sub-series with EMD and then each sub-series are forecasted using an Elman Neural Network model. The forecasted values of original wind speed are calculated by the sum of the predicted values of every sub-series.

1.2 Contributions

The contributions of this paper are as follows:

- To present a new modified Elman neural network in which covers four classes of nonlinear systems. The feedback information from all layers can improve dynamic characteristics and convergence speed of the new modified Elman neural network. It possesses comparatively higher learning capability and convergence speed.
- To analysis the stability of the learning rates.
- To compare numerical results obtained through the proposed approach of this paper with ones achieved from other modified Elman neural networks reported in the literature.

1.3 Paper Organization

The organization of this paper is as follows. Section 2, introduces the proposed modified Elman neural network, namely OIFHO ENN for identification of general classes of nonlinear systems and develops the dynamic recurrent back-propagation algorithm for the purposed new modified Elman neural network. In Sect. 3, to guarantee the fast convergence, the optimal adaptive learning rates are also derived in the sense of discrete-type Lyapunov stability. Simulation results are presented in Sect. 4. Section 5 uses different norms of error, namely, Mean Square Error (MSE), Root Mean Square Error (RMSE) and Normalized Mean Square Error (NMSE) to evaluate the performance of the new modified Elman Neural Network structure proposed in Sect. 2 in comparison with the OHF and the OIF structures.

2 New Modified Elman Neural Network (OIFHO ENN)

In the original Elman neural network, the hidden layer neurons are fed by the outputs of the context neurons and the input neurons. Context neurons are known as memory units as they store the previous outputs of hidden neurons. Since a typical Elman neural network only employs the hidden context nodes to diverse message, it has low learning speed and convergence precision. On the other hand, in the proposed modified Elman neural network, the feedback of the output layer is taken into account; therefore better learning efficiency can be obtained. Moreover, to make the neurons sensitive to the history of input data, self connections of the context nodes and output feedback node are added. Thus, the proposed modified Elman neural network combines the ability of dealing with nonlinear problems and can effectively improve the convergence precision and reduce learning time.

Figure 1 depicts our proposed new modified Elman neural network, namely, Output to Input Feedback, Hidden to Output Elman Neural Network (OIFHO ENN) that is presented based on the Elman neural network. The OIFHO ENN possesses self-feedback links with fixed coefficient α, β and γ in the context nodes. The feedback information from all layers can improve dynamic characteristics and convergence speed of the new modified Elman neural network. In order to compare the speed of convergence of the proposed method with other method, we try several benchmark examples in Sect. 5.

The Input–output equation of OIFHO ENN is:

$$\begin{aligned} Y(k) &= g\left(W^4(k)\,X_c(k) + W^5(k)\,X(k)\right) \\ &= W^4(k)\left(\gamma X_c(k-1) + X(k-1)\right) + W^5(k)\,X(k) \\ &= N_f\left[U(k), U(k-1), \ldots, Y(k-1), Y(k-2), \ldots\right] \end{aligned} \tag{1}$$

$$y_{c,l}(k) = \beta y_{c,l}(k-1) + y_l(k-1), \quad l = 1, \ldots, n \tag{2}$$

$$x_{c,k}(k) = \gamma x_{c,k}(k-1) + x_k(k-1), \quad k = 1, \ldots, n \tag{3}$$

where $y_{c,l}$ is the output of the lth output context unit, $x_{c,k}$is the output of the kth hidden context unit and β $\left(0 < \beta \leq 1\right)$ and γ $\left(0 < \gamma \leq 1\right)$are the self-feedback coefficients. It is a type of recurrent neural networks with different layers of neurons, namely: input nodes, hidden nodes, output nodes and context nodes. The input and output nodes interact with the outside environment, whereas the hidden and context nodes do not. The context nodes are used only to memorize previous activations of the hidden nodes and the output nodes. The feed–forward connections are modifiable, whereas the recurrent connections are fixed.

If we assume that there are r nodes in the input layer, n nodes in the hidden layer and the hidden context layer and m nodes in the output layer and the output context layer, then the input u is an r dimensional vector and the output x of the hidden layer and the output $x_{c,k}$ of the hidden context nodes are n dimensional vectors, where the output y of the output layer and the output $y_{c,l}$ of the output context nodes are m dimensional vectors, and the weights W^1, W^2, W^3, W^4, and W^5 are the weights between hidden layer and input layer, input layer and output context layer, hidden layer and output context layer, output layer and hidden context layer and output layer and hidden layer and are $n \times r$, $r \times m$, $n \times m$, $m \times n$, and $m \times n$ dimensional matrices, respectively.

The mathematical model of the new modified Elman Neural Network can be described as follows:

$$y\,(k) = g\left(W^4 x_c\,(k) + W^5 x\,(k)\right) \tag{4}$$

$$x_c\,(k) = \gamma x_c\,(k-1) + x\,(k-1) \tag{5}$$

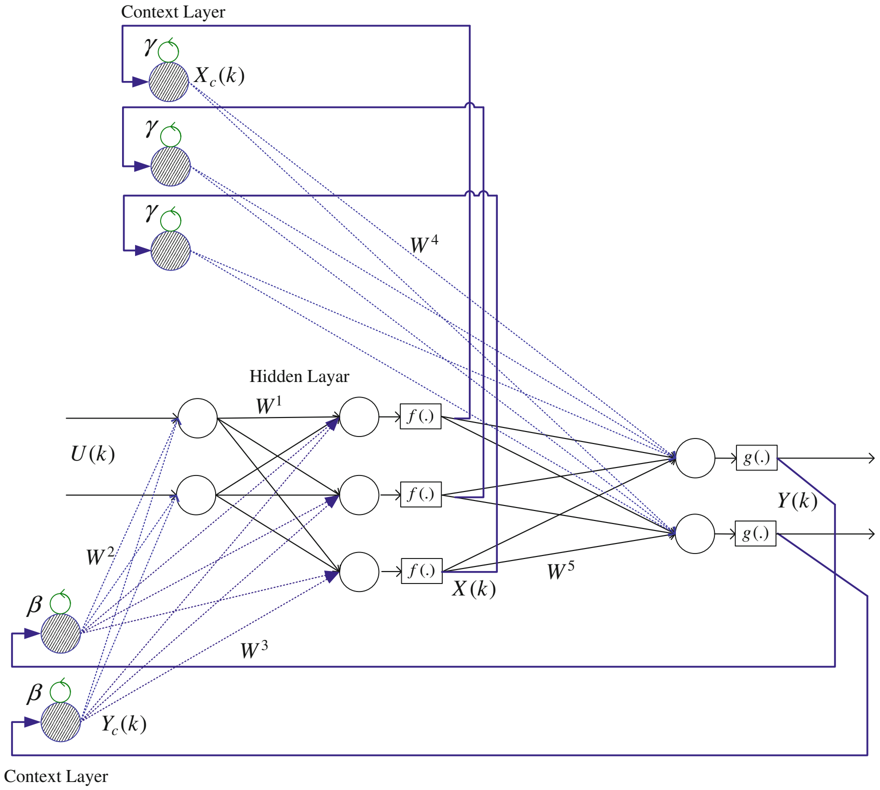

Fig. 1 Architecture of the new modified Elman neural network (OIFHO ENN)

$$x\left(k\right)=f\left(W^{1}\left(W^{2}y_{c}\left(k\right)+u\left(k\right)\right)+W^{3}y_{c}\left(k\right)\right) \tag{6}$$

$$y_{c}\left(k\right)=\beta y_{c}\left(k-1\right)+y\left(k-1\right) \tag{7}$$

$f\left(x\right)$ is often taken as the sigmoid function:

$$f(x)=\frac{1}{1+e^{-x}} \tag{8}$$

and $g\left(x\right)$ is often taken as a linear function, that is:

$$y(k)=W^{4}x_{c}(k)+W^{5}x(k) \tag{9}$$

We may define a norm of error as:

$$E(k)=\frac{1}{2}(y_{d}(k)-y(k))^{T}(y_{d}(k)-y(k)) \tag{10}$$

By differentiating E with respect to W^1, W^2, W^3, W^4, and W^5, according to the gradient descent method, we obtain the following equations:

$$W(k+1)=\Delta W(k)+W(k)$$

$$\Delta w_{ij}^{5}(k)=-\eta_{5}\frac{\partial(e_{i}(k))}{\partial w_{ij}^{5}(k)}=\eta_{5}.\delta_{i}^{o}(k)(x_{j}(k)+w_{ij}^{5}(k)\frac{\partial x_{j}(k)}{\partial w_{ij}^{5}(k)})$$
$$(i=1,\ldots,m),(j=1,\ldots,n) \tag{11}$$

$$\frac{\partial x_{j}(k)}{\partial w_{ij}^{5}(k)}=f_{j}^{\prime}(.)\left(\sum_{q=1}^{r}w_{jq}^{1}(k).A_{5}+\sum_{l=1}^{m}w_{jl}^{3}(k).B_{5}\right) \tag{12}$$

where

$$A_{5}=\sum_{l=1}^{m}w_{ql}^{2}(k)(\gamma\frac{\partial y_{c,l}(k-1)}{\partial w_{ij}^{5}(k)}+(x_{j}(k-1)+w_{ij}^{5}(k)\frac{\partial x_{j}(k-1)}{\partial w_{ij}^{5}(k)}))$$
$$B_{5}=\left(\gamma\frac{\partial y_{c,l}(k-1)}{\partial w_{ij}^{5}(k)}+(x_{j}(k-1)+w_{ij}^{5}(k)\frac{\partial x_{j}(k-1)}{\partial w_{ij}^{5}(k)})\right)$$
$$(i=1,\ldots,m),(j=1,\ldots,n)$$

$$\Delta w_{ih}^{4}(k) = -\eta_4 \frac{\partial(e_i(k))}{\partial w_{ih}^{4}(k)} = \eta_4 \delta_i^{o}(k)\left(x_h(k-1) + \gamma \frac{\partial y_i(k-1)}{\partial w_{ih}^{4}(k)}\right) \quad (h = 1,\ldots,n), (i = 1,\ldots,m) \tag{13}$$

$$\begin{aligned} \Delta w_{jl}^{3} &= \eta_3 \frac{\partial(x_j(k))}{\partial w_{jl}^{3}}.\delta_j^{h}(k) = \eta_3 \sum_{j=1}^{n} w_{ij}^{5}\delta_i^{o}(k)\frac{\partial x_j(k)}{\partial w_{jl}^{3}} \\ &= \eta_3 \sum_{j=1}^{n} w_{ij}^{5}\delta_i^{o}(k)\left(f_j'(.)y_l(k-1) + \beta \frac{\partial x_j(k-1)}{\partial w_{jl}^{3}}\right) \\ &\quad (j = 1,\ldots,n), (q = 1,\ldots,r), (p = 1,\ldots,m) \end{aligned} \tag{14}$$

$$\begin{aligned} \Delta w_{ql}^{2}(k) &= \eta_2 \frac{\partial(x_j(k))}{\partial w_{ql}^{2}(k)}.\delta_j^{h}(k) \\ &= \eta_2 \sum_{q=1}^{r}\sum_{j=1}^{n} w_{jq}^{1}(k) w_{ij}^{5}(k) f_j'(.) \\ &\quad \times \left(\begin{array}{l} \gamma \frac{\partial y_{c,l}(k-1)}{\partial w_{ql}^{2}(k)} \\ +(\sum_{h=1}^{n} w_{ih}^{4}(k)\frac{\partial x_{c,h}(k-1)}{\partial w_{ql}^{2}(k)} + \sum_{j=1}^{n} w_{ij}^{5}(k)\frac{\partial x_j(k-1)}{\partial w_{ql}^{2}(k)}) \end{array}\right).\delta_i^{o}(k) \\ &\quad (l = 1,\ldots,m) \end{aligned} \tag{15}$$

$$\begin{aligned} \Delta w_{jq}^{1}(k) &= \eta_1 \frac{\partial(x_j(k))}{\partial w_{jq}^{1}(k)}.\delta_j^{h}(k) = \eta_1 \sum_{j=1}^{n} w_{ij}^{5}(k) f_j'(.)\frac{\partial(x_j(k))}{\partial w_{jq}^{1}(k)}.\delta_i^{o}(k) \\ \frac{\partial x_j(k)}{\partial w_{jq}^{1}(k)} &= (u_q(k) + \sum_{l=1}^{m} w_{ql}^{2}(k) y_{c,l}(k)) + (\sum_{q=1}^{r} w_{jq}^{1}(k).A_1 + \sum_{l=1}^{m} w_{jl}^{3}(k).B_1) \\ &\quad (q = 1,\ldots,r) \end{aligned} \tag{16}$$

where

$$\begin{aligned} A_1 &= \sum_{l=1}^{m} w_{ql}^{2}(k)(\gamma \frac{\partial y_{c,l}(k-1)}{\partial w_{jq}^{1}(k)} + (\sum_{h=1}^{n} w_{ih}^{4}(k)\frac{\partial x_{c,h}(k-1)}{\partial w_{jq}^{1}(k)} + \sum_{j=1}^{n} w_{ij}^{5}(k)\frac{\partial x_j(k-1)}{\partial w_{jq}^{1}(k)}))) \\ B_1 &= (\gamma \frac{\partial y_{c,l}(k-1)}{\partial w_{jq}^{1}(k)} + (\sum_{h=1}^{n} w_{ih}^{4}(k)\frac{\partial x_{c,h}(k-1)}{\partial w_{jq}^{1}(k)} + \sum_{j=1}^{n} w_{ij}^{5}(k)\frac{\partial x_j(k-1)}{\partial w_{jq}^{1}(k)})) \end{aligned}$$

which form the learning algorithm for the OIFHO ENN, where η_1, η_2, η_3, η_4, and η_5 are learning rates of W^1, W^2, W^3, W^4, and W^5, respectively, and

$$\delta_i^{o}(k) = (y_{d,i}(k) - y_i(k))g_i'(.) \tag{17}$$

$$\delta_j^h(k) = \sum_{i=1}^{m} w_{ji}^5(k)\delta_i^o(k) f_j'(.) \tag{18}$$

$$\delta_q^i(k) = \sum_{j=1}^{n}\sum_{i=1}^{m} w_{jq}^1(k) w_{ji}^5(k)\delta_i^o(k) f_j'(.) \tag{19}$$

if $g(x)$is taken as a linear function, then $g'(.) = 1$.

3 Convergence of Output to Input Feedback, Hidden to Output Elman Neural Network (OIFHO ENN)

The update rules in Eqs. (12–16) need appropriate choice of the learning rates. For the learning rate with a small value, the convergence can be guaranteed, but the speed of convergence is very slow. On the other hand, if the value of the learning rate is too large, the algorithm will become unstable [2]. In order to train neural networks efficiently, we propose five criterions of selecting proper learning rates for the dynamic back propagation algorithm adaptively based on the discrete-type Lyapunov stability analysis. The following theorems give sufficient conditions for the convergence of OIFHO ENN. Suppose that the modification of the weights of by Eqs. (12–16). For the convergence of OIFHO ENN we have the following theorems.

Theorem 1 *The stable convergence of the update rule on* W^1 *is guaranteed if the learning rate* $\eta_1(k)$ *satisfies the following condition:*

$$0 < \eta_1(k) < \frac{8}{nr\left|\max_{ij}(w_{ij}^5(k))\right|\left|\left(\max_q \left|u_q(k)\right| + \max_q\left|\sum_{p=1}^{m} w_{qp}^2 y_{c,p}(k)\right|\right)\right|} \tag{20}$$

Proof Define the Lyapunov energy function as follows:

$$E(k) = \frac{1}{2}\sum_{i=1}^{m} e_i^2(k) \tag{21}$$

where

$$e_i(k) = y_{d,i}(k) - y_i(k) \tag{22}$$

and consequently, we can obtain the modification of the Lyapunov energy function

$$\Delta E(k) = E(k+1) - E(k) = \frac{1}{2}\sum_{i=1}^{m}[e_i^2(k+1) - e_i^2(k)] \tag{23}$$

the error during the learning process can be represented as

$$e_i(k+1) = e_i(k) + \sum_{j=1}^{n}\sum_{q=1}^{m} \frac{\partial e_i(k)}{\partial w_{jq}^1}\Delta w_{jq}^1 = e_i(k) - \sum_{j=1}^{n}\sum_{q=1}^{m} \frac{\partial y_i(k)}{\partial w_{jq}^1}\Delta w_{jq}^1 \quad (24)$$

therefore

$$\begin{aligned}\Delta E(k) &= \frac{1}{2}\sum_{i=1}^{m} e_i^2(k)\left[\left(1 - \eta_1(k)[\frac{\partial y_i(k)}{\partial W^1}]^T[\frac{\partial y_i(k)}{\partial W^1}]^2\right)^2 - 1\right] \\ &= \frac{1}{2}\sum_{i=1}^{m} e_i^2(k)\left[\left(1 - \eta_1(k)\left\|\frac{\partial y_i(k)}{\partial W^1}\right\|^2\right)^2 - 1\right] = -\sum_{i=1}^{m} e_i^2(k)\beta_i^1(k)\end{aligned} \quad (25)$$

where

$$\begin{aligned}\beta_i^1(k) &= \tfrac{1}{2}\left[1 - \left(1 - \eta_1(k)\left\|\tfrac{\partial y_i(k)}{\partial W^1}\right\|^2\right)^2\right] \\ &= \tfrac{1}{2}\eta_1(k)\left\|\tfrac{\partial y_i(k)}{\partial W^1}\right\|^2\left(2 - \eta_1(k)\left\|\tfrac{\partial y_i(k)}{\partial W^1}\right\|^2\right)\end{aligned} \quad (26)$$

We have

$$\left|\tfrac{\partial y_i(k)}{\partial w_{jq}^1}\right| = \left|\tfrac{\partial y_i(k)}{\partial x_j(k)}.\tfrac{\partial x_j(k)}{\partial w_{jq}^1}\right| = \left|w_{ij}^5(k).f_j'(.).(u_q(k) + \sum_{p=1}^{m} w_{qp}^2 y_{c,p}(k))\right| \quad (27)$$
$$(i = 1, \ldots, m: \; j = 1, \ldots, n: \; q = 1, \ldots, r)$$

then

$$\left|\tfrac{\partial y_i(k)}{\partial w_{jq}^1}\right| \le \tfrac{1}{4}\left|\max_{ij}(w_{ij}^5(k))\right|\left\|\left(\max_q |u_q(k)| + \max_q \left|\sum_{p=1}^{m} w_{qp}^2 y_{c,p}(k)\right|\right)\right| \quad (28)$$
$$(i = 1, \ldots, m: \; j = 1, \ldots, n: \; k = 1, \ldots, n: \; q = 1, \ldots, r)$$

then

$$\left\|\tfrac{\partial y_i(k)}{\partial W^1}\right\| \le \sqrt{\tfrac{nr}{4}\left|\max_{ij}(w_{ij}^5(k))\right|\left\|\left(\max_q |u_q(k)| + \max_q \left|\sum_{p=1}^{m} w_{qp}^2 y_{c,p}(k)\right|\right)\right|}$$
$$(i = 1, \ldots, m: \; j = 1, \ldots, n: \; k = 1, \ldots, n: \; q = 1, \ldots, r) \quad (29)$$

and we have

$$0 < \eta_1(k) < \frac{8}{nr\left|\max\limits_{ij}(w^5_{ij}(k))\right|\left\|\left(\max\limits_{q}\left|u_q(k)\right| + \max\limits_{q}\left|\sum_{p=1}^{m} w^2_{qp}y_{c,p}(k)\right|\right)\right\|} \tag{30}$$

We have $\beta_i^1(k) > 0$, then from Eq. (25) we obtain $\Delta E(k) < 0$. According to the Lyapunov stability theory, this shows that the training error will converge to zero as $t \rightarrow \infty$. This completes the proof.

Theorem 2 *The stable convergence of the update rule (15) on* W^2 *is guaranteed if the learning rate* $\eta_2\,(k)$ *satisfies the following condition:*

$$0 < \eta_2(k) < \frac{8}{mn\left|\max\limits_{ij}(w^5_{ij}(k))\right\|\max\limits_{jq}(w^3_{jq}(k))\left\|\max\limits_{p} y_{cp}(k)\right|} \tag{31}$$

Proof The error during the learning process can be expressed as

$$e_i(k+1) = e_i(k) + \sum_{j=1}^{n}\sum_{q=1}^{r}\frac{\partial e_i(k)}{\partial w^2_{jq}}\Delta w^2_{jq} \;= e_i(k) - \sum_{j=1}^{n}\sum_{q=1}^{r}\frac{\partial y_i(k)}{\partial w^2_{jq}}\Delta w^2_{jq} \tag{32}$$

therefore

$$\begin{aligned}\Delta E(k) &= \frac{1}{2}\sum_{i=1}^{m} e_i^2(k)\left[\left(1 - \eta_2(k)[\tfrac{\partial y_i(k)}{\partial W^2}]^T[\tfrac{\partial y_i(k)}{\partial W^2}]\right)^2 - 1\right] \\ &= \frac{1}{2}\sum_{i=1}^{m} e_i^2(k)\left[\left(1 - \eta_2(k)\|\tfrac{\partial y_i(k)}{\partial W^2}\|^2\right)^2 - 1\right] = -\sum_{i=1}^{m} e_i^2(k)\beta_i^2(k)\end{aligned} \tag{33}$$

where

$$\begin{aligned}\beta_i^2(k) &= \frac{1}{2}\left[1 - \left(1 - \eta_2(k)\left\|\frac{\partial y_i(k)}{\partial W^2}\right\|^2\right)^2\right] \\ &= \frac{1}{2}\eta_2(k)\left\|\frac{\partial y_i(k)}{\partial W^2}\right\|^2\left(2 - \eta_2(k)\left\|\frac{\partial y_i(k)}{\partial W^2}\right\|^2\right)\end{aligned} \tag{34}$$

Notice that the activation function of the hidden neurons in the modified Elman neural network is the sigmoidal type, we have

$$\left|\frac{\partial y_i(k)}{\partial w_{jq}^2}\right| = \left|w_{ij}^5(k)w_{jq}^3(k)f_j'(.)y_{cp}(k)\right|$$
$$\leq \frac{1}{4}\left|\max_{ij}(w_{ij}^5(k))\right|\left|\max_{jq}(w_{jq}^3(k))\right|\left|\max_{p} y_{cp}(k)\right|$$
$$(i = 1, \ldots, m : \; j = 1, \ldots, n : \; q = 1, \ldots, r) \tag{35}$$

According to the definition of the Euclidean norm, we have

$$\left\|\frac{\partial y_i(k)}{\partial W^2}\right\| \leq \sqrt{\frac{mn}{4}\left|\max_{ij}(w_{ij}^5(k))\right|\left|\max_{jq}(w_{jq}^3(k))\right|\left|\max_{p} y_{cp}(k)\right|} \tag{36}$$

then

$$0 < \eta_2(k) < \frac{8}{mn\left|\max_{ij}(w_{ij}^5(k))\right|\left|\max_{jq}(w_{jq}^3(k))\right|\left|\max_{p} y_{cp}(k)\right|} \tag{37}$$

We have $\beta_i^2(k) > 0$, then from Eq. (33) we obtain $\Delta E(k) < 0$. According to the Lyapunov stability theory, this shows that the training error will converges to zero as $t \to \infty$. This completes the proof.

Theorem 3 *The stable convergence of the update rule (14) on W^3 is guaranteed if the learning rate $\eta_3(k)$ satisfies the following condition:*

$$0 < \eta_3(k) < \frac{32}{mn\left|\max_{ij}(w_{ij}^5(k))\right|\left|\max_{p} y_{c,p}(k)\right|^2} \tag{38}$$

Proof The error during the learning process can be expressed as

$$e_i(k+1) = e_i(k) + \sum_{j=1}^{n}\sum_{p=1}^{m}\frac{\partial e_i(k)}{\partial w_{jp}^3}\Delta w_{jp}^3 = e_i(k) - \sum_{j=1}^{n}\sum_{p=1}^{m}\frac{\partial y_i(k)}{\partial w_{jp}^3}\Delta w_{jp}^3 \tag{39}$$

therefore

$$\Delta E(k) = \frac{1}{2}\sum_{i=1}^{m} e_i^2(k)\left[\left(1 - \eta_3(k)[\tfrac{\partial y_i(k)}{\partial w_{jp}^3}]^T[\tfrac{\partial y_i(k)}{\partial w_{jp}^3}]^2\right) - 1\right]$$
$$= \frac{1}{2}\sum_{i=1}^{m} e_i^2(k)\left[\left(1 - \eta_3(k)\left\|\frac{\partial y_i(k)}{\partial w_{jp}^3}\right\|^2\right)^2 - 1\right] = -\sum_{i=1}^{m} e_i^2(k)\beta_i^3(k) \tag{40}$$

where

$$\beta_i^3(k) = \frac{1}{2}\left[1 - \left(1 - \eta_3(k)\left\|\frac{\partial y_i(k)}{\partial w_{jp}^3}\right\|^2\right)^2\right] \tag{41}$$

therefore

$$\left|\frac{\partial y_i(k)}{\partial w_{jp}^3}\right| = \left|\frac{\partial y_i(k)}{\partial x_j(k)} \cdot \frac{\partial x_j(k)}{\partial w_{jp}^3}\right| = \left|w_{ij}^5 f_j'(.) y_{c,p}(k)\right| \quad (i = 1, \ldots, m: \ j = 1, \ldots, n: p = 1, \ldots, m) \tag{42}$$

We have

$$\left\|\frac{\partial y_i(k)}{\partial W^3}\right\| < \frac{\sqrt{mn}\left|\max_{ij}(w_{ij}^5(k))\right|\left|\max_p y_{c,p}(k)\right|}{4} \tag{43}$$

then

$$0 < \eta_3(k) < \frac{32}{mn\left|\max_{ij}(w_{ij}^5(k))\right|\left|\max_p y_{c,p}(k)\right|^2} \tag{44}$$

therefore, $\eta_3(k)$ is chosen as above, then we have $\beta_i^3(k) > 0$ and $\Delta E(k) < 0$, According to the Lyapunov stability theory, this shows that the training error will converges to zero as $t \to \infty$.

Theorem 4 *The stable convergence of the update rule (13) on W^4 is guaranteed if the learning rate $\eta_4(k)$ satisfies the following condition:*

$$0 < \eta_4(k) < \frac{2}{n} \tag{45}$$

Proof The error during the learning process can be expressed as

$$e_i(k+1) = e_i(k) + \sum_{j=1}^{n} \frac{\partial e_i(k)}{\partial w_{ij}^4}\Delta w_{ij}^4 \ = e_i(k) - \sum_{j=1}^{n} \frac{\partial y_i(k)}{\partial w_{ij}^4}\Delta w_{ij}^4 \tag{46}$$

therefore

$$\begin{aligned}\Delta E(k) &= \frac{1}{2}\sum_{i=1}^{m} e_i^2(k)\left[\left(1 - \eta_4(k)[\frac{\partial y_i(k)}{\partial W^4}]^T[\frac{\partial y_i(k)}{\partial W^4}]^2\right) - 1\right] \\ &= \frac{1}{2}\sum_{i=1}^{m} e_i^2(k)\left[\left(1 - \eta_4(k)\left\|\frac{\partial y_i(k)}{\partial W^4}\right\|^2\right)^2 - 1\right] = -\sum_{i=1}^{m} e_i^2(k)\beta_i^4(k)\end{aligned} \tag{47}$$

where

$$\beta_i^4(k) = \frac{1}{2}\left[1 - \left(1 - \eta_4(k)\left\|\frac{\partial y_i(k)}{\partial W^4}\right\|^2\right)^2\right] \quad (48)$$

W^4 represents an n dimensional vector and $\|.\|$ denotes the Euclidean norm. Noticing that the activation function of the hidden neurons in the modified Elman Neural Network is the sigmoidal type, we have

$$\left|\frac{\partial y_i(k)}{\partial w_{il}^4}\right| = \left|x_{c,k}(k)\right| \qquad (i = 1, \ldots, m : l = 1, \ldots, m) \quad (49)$$

According to the definition of the Euclidean norm, we have

$$\left\|\frac{\partial y_i(k)}{\partial W^4}\right\| < \sqrt{n}\left|x_{c,k}(k)\right| \quad (50)$$

then

$$0 < \eta_4(k) < \frac{2}{n\left|x_{c,k}(k)\right|^2} \quad (51)$$

therefore, $\eta_4\,(k)$ is chosen as:

$$0 < \eta_4(k) < \frac{2}{n\left|x_{c,k}(k)\right|^2} \quad (52)$$

then we have $\beta_i^4\,(k) > 0$ and $\Delta E(k) < 0$, according to the Lyapunov stability theory, this shows that the training error will converges to zero as $t \to \infty$.

Theorem 5 *The stable convergence of the update rule (12) on* W^5 *is guaranteed if the learning rate* $\eta_5\,(k)$ *satisfies the following condition:*

$$0 < \eta_5(k) < \frac{2}{mn} \quad (53)$$

Proof

$$e_i(k+1) = e_i(k) + \sum_{j=1}^{n}\frac{\partial e_i(k)}{\partial w_{ij}^5}\Delta w_{ij}^5 = e_i(k) - \sum_{j=1}^{n}\frac{\partial y_i(k)}{\partial w_{ij}^5}\Delta w_{ij}^5 \quad (54)$$

therefore

$$\Delta E(k) = \frac{1}{2}\sum_{i=1}^{m} e_i^2(k)\left[\left(1-\eta_5(k)[\frac{\partial y_i(k)}{\partial W^5}]^T[\frac{\partial y_i(k)}{\partial W^5}]^2\right)-1\right]$$
$$= \frac{1}{2}\sum_{i=1}^{m} e_i^2(k)\left[\left(1-\eta_5(k)\left\|\frac{\partial y_i(k)}{\partial W^5}\right\|^2\right)^2-1\right] = -\sum_{i=1}^{m} e_i^2(k)\beta_i^5(k) \quad (55)$$

where

$$\beta_i^5(k) = \frac{1}{2}\left[1-\left(1-\eta_5(k)\left\|\frac{\partial y_i(k)}{\partial W^5}\right\|^2\right)^2\right] \quad (56)$$

W^5 represents an n dimensional vector and $\|.\|$ denotes the Euclidean norm. Noticing that the activation function of the hidden neurons in the modified Elman neural network is the sigmoidal type, we have

$$\left|\frac{\partial y_i(k)}{\partial w_{ij}^5}\right| = \left|x_j(k)\right| < 1 \quad (i = 1, \ldots, m : \ l = 1, \ldots, m) \quad (57)$$

According to the definition of the Euclidean norm, we have

$$\left\|\frac{\partial y_i(k)}{\partial W^5}\right\| < \sqrt{mn} \quad (58)$$

then

$$0 < \eta_4(k) < \frac{2}{mn} \quad (59)$$

therefore, $\eta_5\,(k)$ is chosen as $0 < \eta_5(k) < (2/mn)$, then we have $\beta_i^5\,(k) > 0$ and $\Delta E(k) < 0$, According to the Lyapunov stability theory, this shows that the training error will converges to zero as $t \to \infty$.

4 Simulation Results

The objective of this section is to illustrate the performance and capabilities of the proposed structure shown in Fig. 1 for identification of four classes of nonlinear systems considered in [18]. The reference input $u\,(t)$ to all the identifiers must be selected to be "persistently exciting". For identification of linear systems the persistent excitation of the input guarantees the convergence of the identifier parameters to their true values [1]. The following results are compared with OHF and OIF Elman Neural Network [11] and the amplitude and the frequency of the reference inputs are selected experimentally as recommended in [18].

4.1 Application to Model I

Example 1 The governing equation of the system is given by

$$y\,(t) = 0.3y\,(t-1) + 0.6y\,(t-2) + \frac{0.6}{1 + u(t-1)^2} \tag{60}$$

where the output at time t is a linear function of past output at times $t-1$ and $t-2$ plus a nonlinear function of the input at time $t-1$. The reference input $u\,(t-1)$ to the system is selected as $u\,(t-1) = \sin(2\pi(t-1)/100))$.

To show the robustness of the proposed structure to the variations in the amplitude and frequency of the input, an input with 50 % reduction in the frequency (within the 400–600 time steps) and 100 % increase in the frequency (within the 600–800 time steps) is applied to the system. Figures 2 and 3 depict the simulation results using the OIFHO, OHF and OIF Elman NN.

As can be seen from Fig. 3 the performance of the OIFHO Elman NN structure is more robust to the variations in the amplitude as well as the frequency of the input than two other Networks.

Figure 4 shows the variation of learning rates during the simulation for Example 1. For each step the learning rates are chosen according to Eqs. (30), (37), (44), (52) and (59), in which *max (w)* is chosen from available information for the same step. Selecting learning rates in the determined bounds assures the stability of OIFHO ENN.

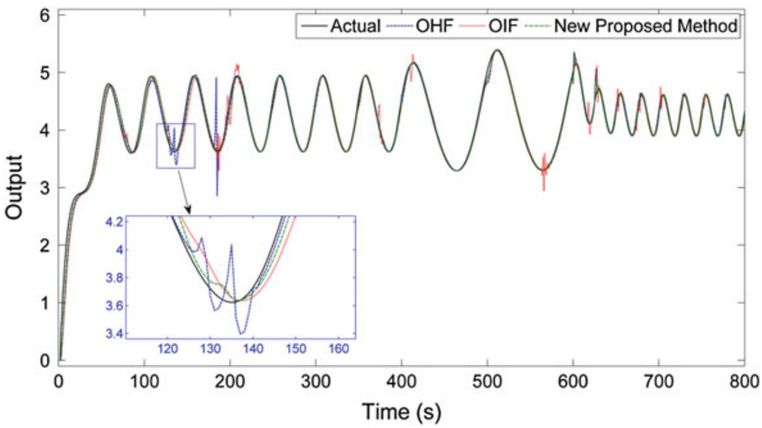

Fig. 2 Responses of the OIFHO, OHF and OIF Elman neural network applied to Example 1 for changing input characteristics

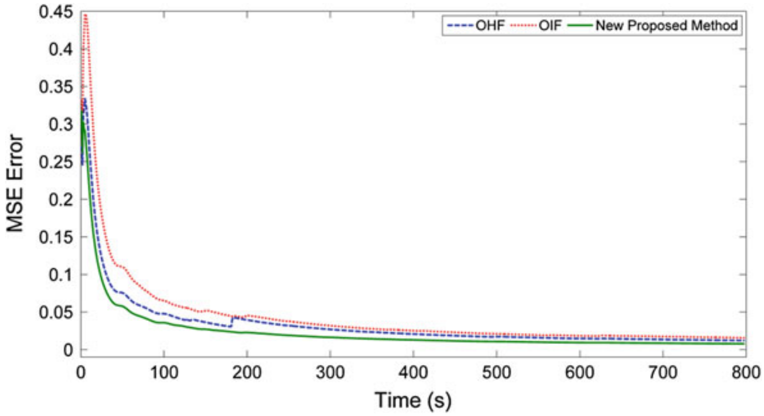

Fig. 3 Comparison of the MSE error of the OIFHO, OHF and OIF Elman neural network applied to Example 1 for changing input characteristics

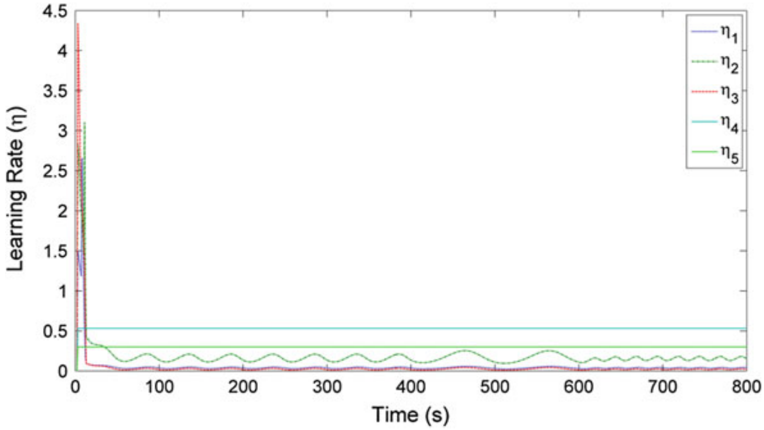

Fig. 4 Variations of learning rates for Example 1

4.2 Application to Model II

Example 2 The governing equation of the system is given by

$$y(t) = \frac{y(t-1)\,y(t-2) + (y(t-2) + 2.5)}{1 + y(t-1)^2 + y(t-2)^2} + u(t-1) \tag{61}$$

where the output at time t is a nonlinear function of the outputs at times $t-1$ and $t-2$ plus a linear function of the input at time $t-1$. The reference input $u(t-1)$ is selected as $u(t-1) = \sin(2\pi(t-1)/25)$).

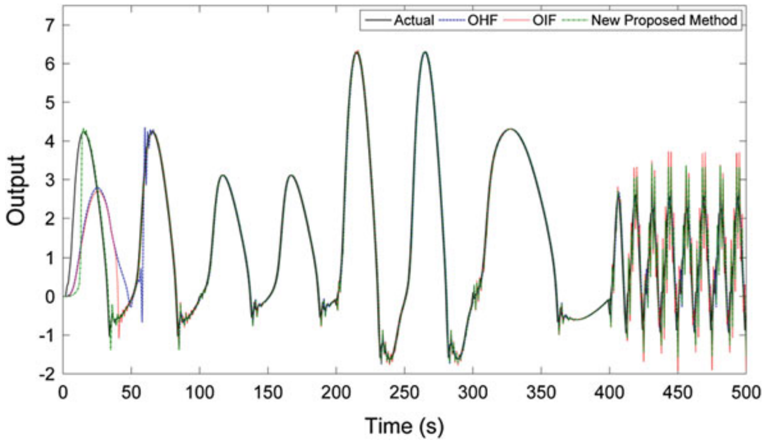

Fig. 5 Responses of the OIFHO, OHF and OIF Elman neural network applied to Example 2 for changing input characteristics

To show the robustness of the proposed structure to variations in the input amplitude and frequency, an input with 50 % reduction in the amplitude (within the 100–200 time steps), 100 % increase in the amplitude (within the 200–300 time steps), 50 % reduction in the frequency (within the 300–400 time steps), and 100 % increase in the frequency (within the 400–500 time steps) is applied to the system. Figures 5 and 6 depict the simulation results using the OIFHO, OHF and OIF Elman NN.

As can be seen from Fig. 6 the performance of the OIFHO Elman NN structure is more robust to the variations in the amplitude as well as the frequency of the input than two other Networks.

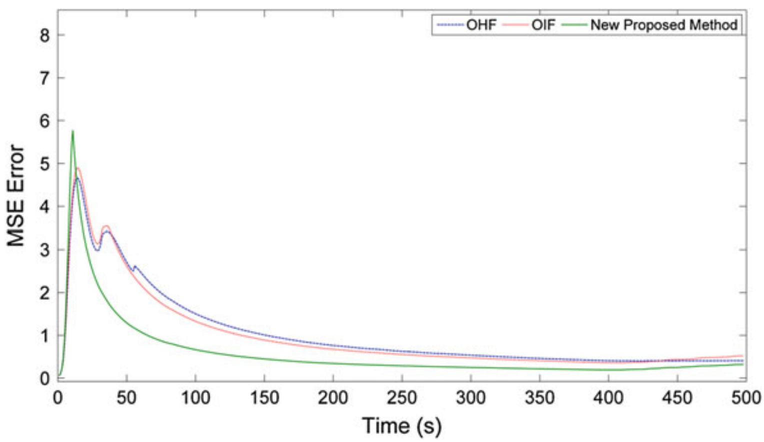

Fig. 6 Comparison of the MSE error of the OIFHO, OHF and OIF Elman neural network applied to Example 2 for changing

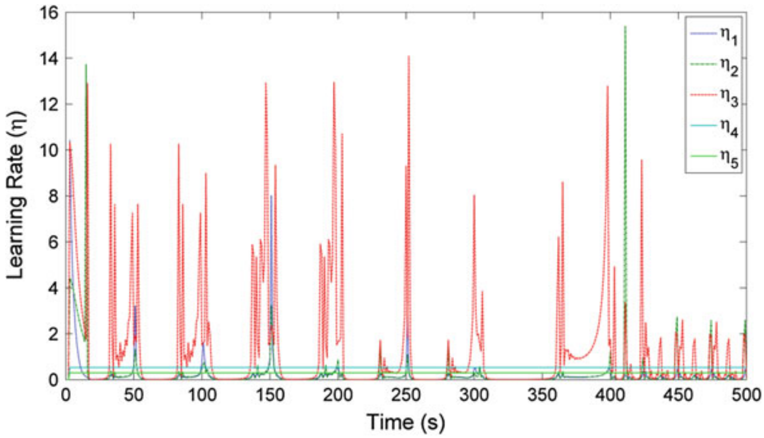

Fig. 7 Variations of learning rates for Example 2

Figure 7 shows the variation of learning rates during the simulation for Example 2. For each step the learning rates are chosen according to Eqs. (30), (37), (44), (52) and (59).

4.3 *Application to Model III*

Example 3 The governing equation of the system is given by

$$y\,(t) = \frac{0.2y\,(t-1) + 0.6y\,(t-2)}{1 + y\,(t-1)^2} + \sin\left(u(t-1)\right) \tag{62}$$

where the output at time t is a nonlinear function of the output at time $t-1$ and $t-2$ plus a nonlinear function of the input at time $t-1$. The reference input applied to the system is $u\,(t-1) = \sin(2\pi(t-1)/10)) + \sin(2\pi(t-1)/25))$. To show the robustness of the proposed structure to variations in the input amplitude and frequency, an input with 50 % reduction in the amplitude (within the 400–600 time steps), and 100 % increase in the frequency (within the 600–800 time steps) is applied to the system. Figures 8 and 9 depict the simulation results using the OIFHO, OHF and OIF Elman NN.

As can be seen from Fig. 9 the performance of the OIFHO Elman NN structure is more robust to the variations in the amplitude as well as the frequency of the input than two other Networks. Figure 10 shows the variation of learning rates during the simulation for Example 3. For each step the learning rates are chosen according to Eqs. (30), (37), (44), (52) and (59).

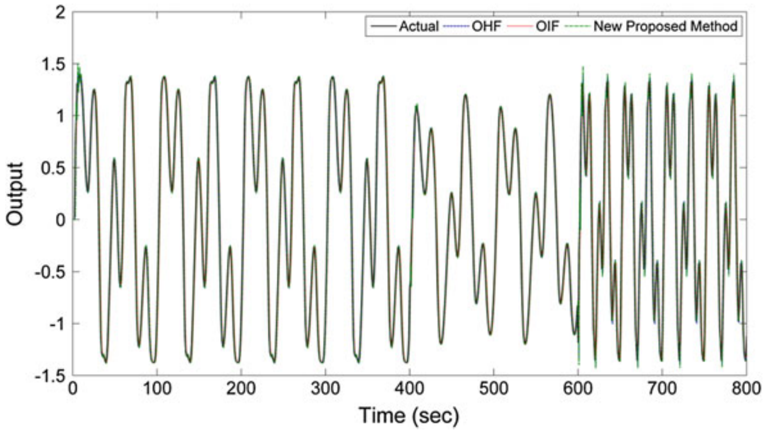

Fig. 8 Responses of the OIFHO, OHF and OIF Elman neural network applied to Example 3 for changing input characteristics

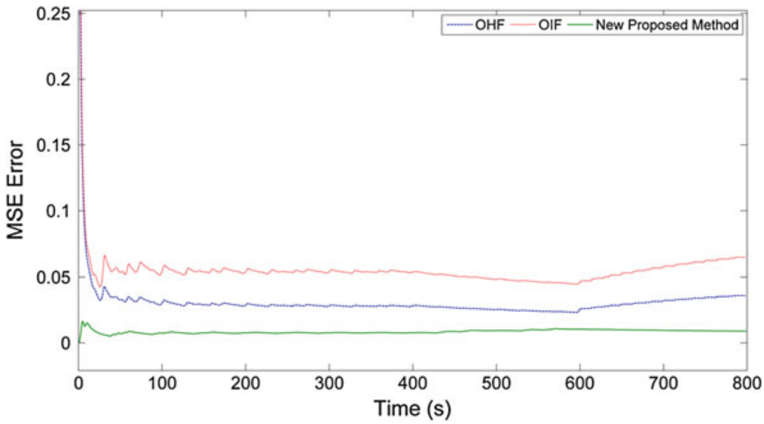

Fig. 9 Comparison of the MSE error of the OIFHO, OHF and OIF Elman neural network applied to Example 3 for changing input characteristics

4.4 Application to Model IV

Example 4 The governing equation of the system is given by

$$y\,(t) = \frac{y\,(t-1) + u\,(t-1)}{1 + y\,(t-1)^2} \tag{63}$$

where the output at time t is a nonlinear function of the outputs at times $t-1$ and the inputs at times $t-1$. The reference input is $u\,(t-1) = \sin(2\pi(t-1)/50))$.

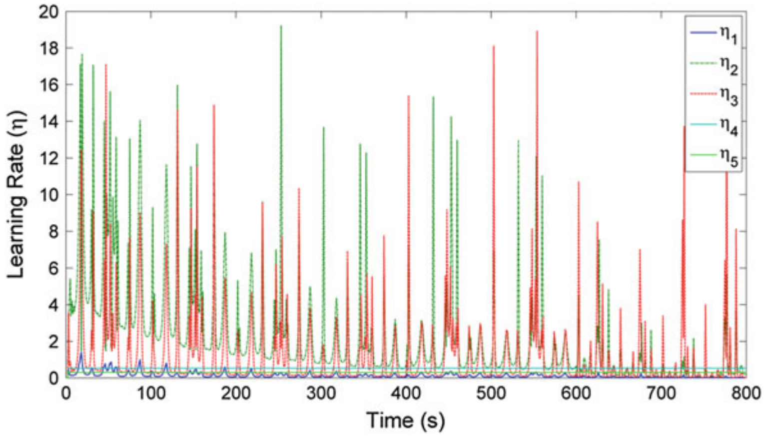

Fig. 10 Variations of learning rates for Example 3

To show the robustness of the proposed structure to variations in the input amplitude and frequency, an input with 50 % reduction in the amplitude (within the 250–500 time steps), 100 % increase in the amplitude (within the 500–750 time steps), 50 % reduction in the frequency (within the 750–1,000 time steps), and 100 % increase in the frequency (within the 1,000–1250 time steps) is applied to the neuro-dynamic structure. Figures 11 and 12 depict the simulation results using the OIFHO, OHF and OIF Elman NN.

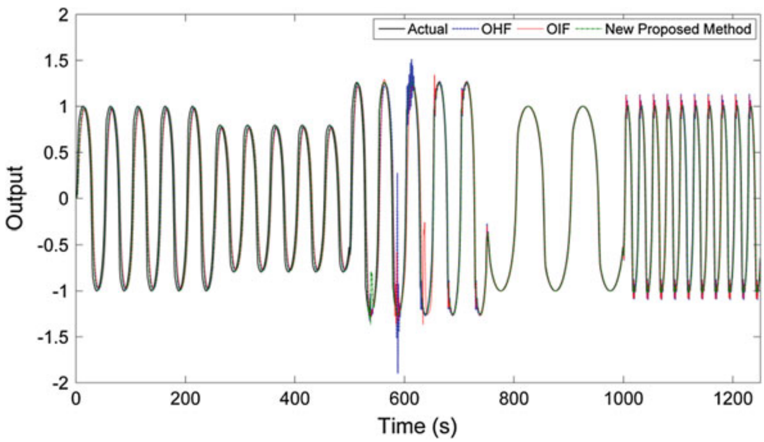

Fig. 11 Responses of the OIFHO, OHF and OIF Elman neural network applied to Example 4 for changing input characteristics

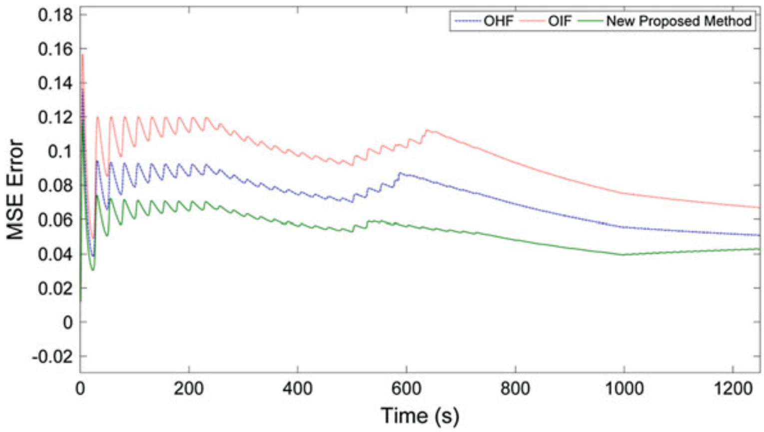

Fig. 12 Comparison of the MSE error of the OIFHO, OHF and OIF Elman neural network applied to Example 4 for changing input characteristics

As can be seen from Fig. 12 the performance of the OIFHO Elman NN structure is more robust to the variations of the amplitude as well as the frequency of the input than two other Networks.

Figure 13 shows the variation of learning rates during the simulation for Example 4. For each step the learning rates are chosen according to Eqs. (30), (37), (44), (52) and (59).

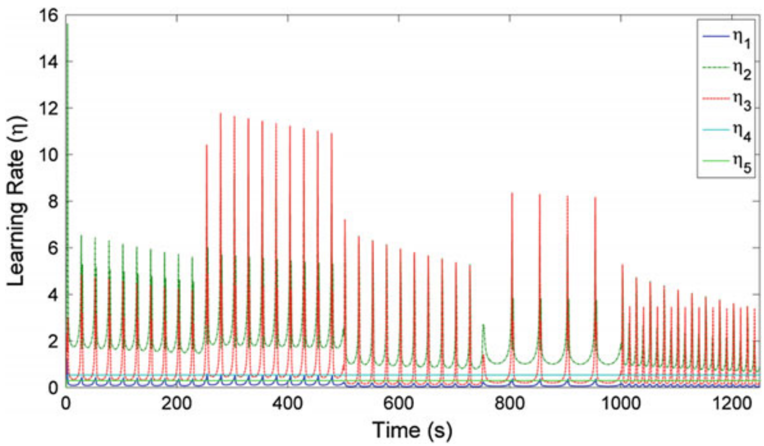

Fig. 13 Variations of learning rates for Example 4

Table 1 Comparison of errors for OIFHO, OHF and OIF structures

	Error	OIFHO	OHF	OIF
Example 1	MSE	0.0078	0.0121	0.0157
	RMSE	0.0882	0.1102	0.1254
	NMSE	0.018	0.0281	0.0365
Example 2	MSE	0.322	0.4073	0.5264
	RMSE	0.5675	0.6382	0.7256
	NMSE	0.0854	0.108	0.1396
Example 3	MSE	0.0312	0.0359	0.0649
	RMSE	0.1765	0.1895	0.2548
	NMSE	0.0405	0.0466	0.0843
Example 4	MSE	0.0427	0.0508	0.0667
	RMSE	0.2067	0.2254	0.2583
	NMSE	0.0603	0.0717	0.0941

5 The Identification Error

Depending on the nature and desired specifications of an application, different error norms may be used to evaluate the performance of an algorithm. We use the Mean Square Error (MSE), Root Mean Square Error (RMSE) and Normalized Mean Square Error (NMSE) to evaluate the performance of OIFHO Elman Neural Network structure proposed in Sect. 2 in comparison with OHF and the OIF structure. The results are given in Table 1.

According to the above tables, we can draw the conclusion that OIFHO structure provided a better performance than other structures, which is due to the excellent nonlinear function approximation capability of OIFHO structure.

6 Conclusion

This paper proposes an improved Elman Neural Network with better performance in comparison with other improved Elman Neural Network by employing three context layers. Subsequently, the dynamic recurrent Backpropagation algorithm for OIFHO is developed according to the gradient descent method. To guarantee the fast convergence, the optimal adaptive learning rates are also derived in the sense of discrete-type Lyapunov stability. Furthermore, capabilities of the proposed structures for identification of four classes of nonlinear systems are shown analytically. Simulation results indicate that the proposed structure is very effective identifying the input–output maps of different classes of nonlinear systems.

References

1. Yazdizadeh, A., Khorasani, K.: Adaptive time delay neural network structure for nonlinear system identification. Neurocomputing **47**, 207–240 (2001)
2. Ge, H.W., Du, W.L., Qian, F., Liang, Y.C.: Identification and control of nonlinear systems by a time-delay recurrent neural network. Neurocomputing **72**, 2857–2864 (2009)
3. Yazdizadeh, A., Khorasani, K.: Identification of a class of nonlinear systems using dynamic neural networks. In: Proceedings of the International Conference on Neural Networks, pp. 161–166. Houston, Texas (1997)
4. Pham, D.T., Liu, X.: Identification of linear and nonlinear dynamic systems using recurrent neural networks. Artif. Intell. Eng. **8**, 67–75 (1993)
5. Wang, J., Zhang, W., Li, Y., Wang, J., Dang, Z.: Forecasting wind speed using empirical mode decomposition and Elman neural network. In: Original Research Article, Applied Soft Computing, Uncorrected Proof, Available online 27 June (2014 in press)
6. Qi. W.M., Chu. C.Y., Ling, J.Q., You, C.W.: A new Elman neural network and its application in hydro-turbine governing system. In: Power and Energy Engineering Conference, APPEEC (2009)
7. Lin, W.M., Hong, C.M.: A new Elman neural network-based control algorithm for adjustable-pitch variable-speed wind-energy conversion systems. IEEE Trans. Power Electron. **26**, 473–481 (2011)
8. Pham, D.T., Liu, X.: Dynamic system modeling using partially recurrent neural networks. J. Syst. Eng. **2**, 90–97 (1992)
9. Hertz, J., Korgh, A., Palmer, R.G.: Recurrent neural networks. Introduction to the Theory of Neural Computing. Addison- wisely, California (1991). Chapter 7
10. Kamijo, K., Tanigawa, T.: Stock price pattern recognition-a recurrent neural network approach. In: Proceedings of the International Joint Conference on Neural Networks, pp. 215–221 (1990)
11. Shi, X.H., Liang, Y.C., Lee, H.P., Lin, W.Z., Xu, X., Lim, S.P.: Improved Elman networks and applications for controlling ultrasonic motors. Appl Artif Intell **18**(7), 603–629 (2004)
12. Ardalani-Farsa, M., Zolfaghari, S.: Chaotic time series prediction with residual analysis method using hybrid Elman-NARX neural networks. Neurocomputing **73**, 2540–2553 (2010)
13. Toqeer, R.S., Bayindir, N.S.: Speed estimation of an induction motor using Elman neural network. Neurocomputing **55**, 727–730 (2003)
14. Seker, S., Ayaz, E., Urkcan, E.T.: Elman's recurrent neural network applications to condition monitoring in nuclear power plant and rotating machinery. Eng. Appl. Artif. Intell. **16**, 647–656 (2003)
15. Song, Q.: On the weight convergence of Elman networks. IEEE Trans. Neural Netw. **21**, 463–480 (2010)
16. Hsu, Chun-Fei: Adaptive backstepping Elman-based neural control for unknown nonlinear systems. Original Research Article Neurocomputing **136**, 170–179 (2014)
17. Cheng, Y.C., Qi, W.M., Zhao, J.: A new Elman neural network and its dynamic properties.In: Cybernetics and Intelligent Systems IEEE Conference, pp. 971–975 (2008)
18. Narendra, K.S., Parthasarathy, K.: Identification and control of dynamical systems using neural networks. IEEE Trans Neural Netw. **1**, 4–27 (1990)