Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Information Sciences 158 (2004) 131–147

INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL

www.elsevier.com/locate/ins

Nonlinear system identification using discrete-time recurrent neural networks with stable learning algorithms

Wen Yu *

Departamento de Control Automatico, CINVESTAV-IPN, A.P. 14-740, Av. IPN 2508, México D.F. 07360, Mexico

Abstract

In general, neural networks cannot match nonlinear systems exactly. Neuro identifier has to include robust modification in order to guarantee Lyapunov stability. In this paper input-to-state stability approach is applied to access robust training algorithms of discrete-time recurrent neural networks. We conclude that for nonlinear system identification, the gradient descent law and the backpropagation-like algorithm for the weights adjustment are stable in the sense of L_∞ and robust to any bounded uncertainties.

1. Introduction

Recent results show that neural network technique seems to be very effective to identify a broad category of complex nonlinear systems when complete model information cannot be obtained. Lyapunov approach can be used directly to obtain robust training algorithms for continuous-time neural networks [6,13,19,26]. Discrete-time neural networks are more convenient for real applications. Two types stability for discrete-time neural networks were studied. The stability of neural networks can be found in [5,20]. The stability of

* Fax: +52-55-5747-7089.
E-mail address: yuw@ctrl.cinvestav.mx (W. Yu).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2003.08.002

learning algorithms was discussed in [11,16]. In [16] they assumed neural networks could represent nonlinear systems exactly, and concluded that backpropagation-type algorithm guaranteed exact convergence of identification error. Gersgorin's theorem was used to derive stability conditions for the network learning in [11].

Neural networks can be classified as feedforward and recurrent ones. Most of publications in nonlinear system identification use feedforward networks, for example multilayer perceptrons (MLP), which are implemented for the approximation of nonlinear functions in the right hand side of dynamic model equations. The main drawback of these neural networks is that the weights' updating do not utilize any information on the local data structure and the function approximation is sensitive to the training data [11]. Since recurrent networks incorporate feedback, they have powerful representation capability and can successfully overcome disadvantages of feedforward networks [7].

It is well known that normal identification algorithms are stable for ideal plants [8]. In the presence of disturbances or unmodeled dynamics, these adaptive procedures can go to instability easily. The lack of robustness in parameters identification was demonstrated in [4] and became a hot issue in 1980s. Several robust modification techniques were proposed in [8]. The weight adjusting algorithms of neural networks is a type of parameters identification, the normal gradient algorithm is stable when neural network model can match the nonlinear plant exactly [16]. Generally, some modifications to the normal gradient algorithm or backpropagation should be applied, such that the learning process is stable. For example, in [11] some hard restrictions were added in the learning law, in [20] the dynamic backpropagation was modified with NLq stability constraints. Another generalized method is to use robust modification techniques [8]. Kosmatopoulos et al. [13] applied σ-modification, Jagannathan and Lewis [9] used modified δ-rule, and Song [18] used dead-zone in the weight tuning algorithms. By ISS theory, we prove that the normal gradient law and backpropagation-like algorithm without robust modifications are L_∞ stable for discrete-time feedforward neuro identification [27].

Neuro identification is in sense of black-box approximation. All uncertainties can be considered as parts of the black-box, i.e., unmodeled dynamics are within the black-box model, not as structured uncertainties. Therefore the common used robustifying techniques are not necessary. By passivity theory, gradient descent algorithms for continuous-time recurrent neural networks were stable and robust to any bounded uncertainties without robust modification [24], and for continuous-time recurrent neural identification they were also robustly stable [25]. Nevertheless, do discrete-time recurrent neural networks have the similar characteristics? This paper gives an answer for it. To the best of our knowledge, system identification without robust modification via discrete-time recurrent neural networks has not yet been established in the literature.

Input-to-state stability (ISS) is another elegant approach to analyze stability besides Lyapunov method. It can lead to general conclusions on stability by using input/state characteristics. In this paper, ISS approach is applied to obtain some new learning laws that do not need robust modifications. A simple simulation gives the effectiveness of the suggested algorithm.

2. Preliminaries

The main concern of this section is to understand some concepts of ISS and backpropagation of MLP. Consider following discrete-time state-space nonlinear system

$$x(k+1) = f[x(k), u(k)] \tag{1}$$

where $u(k) \in \Re^m$ is the input vector, $x(k) \in \Re^n$ is a state vector, f is general nonlinear smooth function $f \in C^\infty$. Let us now recall following definitions.

Definition 1

(a) If a function $\gamma(s)$ is continuous and strictly increasing with $\gamma(0) = 0$, $\gamma(s)$ is called as $\mathscr{K}$-function.

(b) For a function $\beta(s,t)$, $\beta(s,\cdot)$ is $\mathscr{K}$-function, $\beta(\cdot,t)$ is decreasing and $\lim_{t\to\infty} \beta(\cdot,t) = 0$, $\beta(s,t)$ is called as $\mathscr{K}\mathscr{L}$-function.

(c) If a function $\alpha(s)$ is $\mathscr{K}$-function and $\lim_{s\to\infty} \alpha(s) = \infty$, $\alpha(s)$ is called as $\mathscr{K}_\infty$-functions.

Definition 2

(a) A system (1) is said to be *input-to-state stability* if there exists a $\mathscr{K}$-function $\gamma(\cdot)$ and $\mathscr{K}\mathscr{L}$-function $\beta(\cdot)$, such that, for each $u \in L_\infty$, i.e., $\sup\{\|u(k)\|\} < \infty$, and each initial state $x^0 \in R^n$, it holds that

$$\|x(k, x^0, u(k))\| \leqslant \beta(\|x^0\|, k) + \gamma(\|u(k)\|)$$

(b) A smooth function $V : \Re^n \to \Re \geqslant 0$ is called a ISS-Lyapunov function for system (1) if there exist $\mathscr{K}_\infty$-functions $\alpha_1(\cdot)$, $\alpha_2(\cdot)$ and $\alpha_3(\cdot)$, and $\mathscr{K}$-function $\alpha_4(\cdot)$ such that for any $s \in \Re^n$, each $x(k) \in \Re^n$, $u(k) \in \Re^m$

$$\alpha_1(s) \leqslant V(s) \leqslant \alpha_2(s)$$

$$V_{k+1} - V_k \leqslant -\alpha_3(\|x(k)\|) + \alpha_4(\|u(k)\|)$$

Theorem 1. *For a discrete-time nonlinear system, the following are equivalent* [12].

- *It is input-to-state stability (ISS).*
- *It is robustly stable.*
- *It admit a smooth ISS-Lyapunov function.*

Property. If a nonlinear system is input-to-state stable, the behavior of the system remains bounded when its inputs are bounded.

The learning algorithm used to adjust the synaptic weights of a multilayer perceptron is known as backpropagation. This algorithm provides a compositionally efficient method for the training of multilayer networks. Even if it does not give a solution for all problems, it put to rest the criticism about neuro learning. A part of multilayer perceptron is shown in Fig. 1, where d_j is the desired output, y_i is the ith neuron output, i, j and k indicate neurons, w_{ij} is the weight between neuro i and neuron j. For the ith neuron, the nonlinear active function is defined as $\varphi_i(\cdot)$.

The output of MLP is

$$y_j = \varphi_j(v_j), \quad v_j = \sum_{i=1}^{n} w_{ij} y_i$$

The error at the output of neuron j is given as

$$e_j(k) = y_j(k) - d_j(k)$$

The instantaneous sum of the squared output errors is given by

$$\varepsilon(k) = \frac{1}{2} \sum_{j=1}^{l} e_j^2(k)$$

where l is the number of neurons of the output layer. Using the gradient decent, the weight connecting neuron i to neuron j is updated as

$$w_{ij}(k+1) = w_{ij}(k) - \eta \frac{\partial \varepsilon(k)}{\partial w_{ij}(k)}$$

where η is learning rate. The term $\frac{\partial \varepsilon(k)}{\partial w_{ij}(k)}$ can be calculated as

$$\frac{\partial \varepsilon(k)}{\partial w_{ij}(k)} = \frac{\partial \varepsilon(k)}{\partial e_j(k)} \frac{\partial e_j(k)}{\partial y_j(k)} \frac{\partial y_j(k)}{\partial v_j(k)} \frac{\partial v_j(k)}{\partial w_{ij}(k)}$$

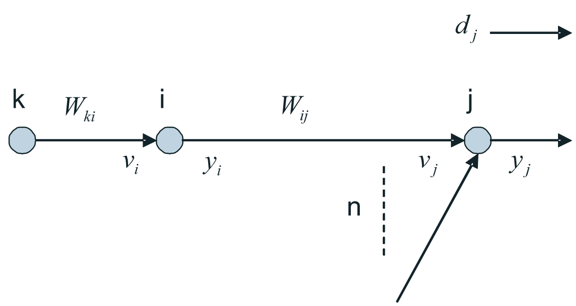

Fig. 1. One part of MLP.

The partial derivatives are given by $\frac{\partial \varepsilon(k)}{\partial e_j(k)} = e_j(k)$, $\frac{\partial e_j(k)}{\partial y_j(k)} = 1$, $\frac{\partial y_j(k)}{\partial v_j(k)} = \varphi'_j[v_j(k)]$, $\frac{\partial v_j(k)}{\partial w_{ji}(k)} = y_i(k)$. Usually, linear function is used for the output layer, so

$$w_{ij}(k+1) = w_{ij}(k) - \eta y_i(k) e_j(k) \tag{2}$$

Similar w_{ki} can be updated as

$$w_{ki}(k+1) = w_{ki}(k) - \eta \frac{\partial \varepsilon(k)}{\partial w_{ki}(k)}$$

and

$$\begin{aligned}\frac{\partial \varepsilon(k)}{\partial w_{ki}(k)} &= \frac{\partial \varepsilon(k)}{\partial e_j(k)} \frac{\partial e_j(k)}{\partial y_j(k)} \frac{\partial y_j(k)}{\partial v_j(k)} \frac{\partial v_j(k)}{\partial y_i(k)} \frac{\partial y_i(k)}{\partial v_i(k)} \frac{\partial v_i(k)}{\partial w_{ki}(k)} \\ &= \{e_j(k)\varphi'_j[v_j(k)]W_{ij}\}\varphi'_i[v_i(k)]y_k(k) = e_i\varphi'_i[v_i(k)]y_k(k)\end{aligned}$$

So

$$w_{ki}(k+1) = w_{ki}(k) - \eta y_k(k) e_i(k) \varphi'_i[v_i(k)] \tag{3}$$

where $e_i(k)\varphi'_i[v_i(k)]$ is error backpropagation.

Backpropagation algorithm has become the most popular one for training of the multilayer perceptron [17]. It is easy to calculate and is able to classify information non-linearly separable. The algorithm is a gradient technique, implementing only one step search in the direction of the minimum, which could be a local one, and not an optimization one. So, it is not possible to demonstrate its convergence to a global optimum.

The research on faster algorithms of MLP falls roughly into two categories. The first category involves the development of heuristic techniques, which arise out of a study of the distinctive performance of standard backpropagation algorithm. These heuristic techniques include such ideas as varying the learning rate [10], using momentum [22] and re-scaling variables [21]. Another category of research has focused on standard numerical optimization techniques. Training feedforward neural networks becomes a simply a numerical optimization problem [1].

3. System identification with single layer recurrent neural networks

Consider discrete-time recurrent neural networks

$$\beta \hat{x}(k+1) = A\hat{x}(k) + \sigma[W_1(k)x(k)] + \phi[W_2(k)x(k)]U(k) \tag{4}$$

where $\hat{x}(k) \in \Re^n$ represents the internal state of the neural network. The matrix $A \in \Re^{n\times n}$ is a stable matrix which will be specified after. The matrices $W_1(k)$, $W_2(k) \in \Re^{n\times n}$ are the weights of the neural network. β is a positive constant $\beta > 1$ which is a design parameter. $U(k) = [u_1, u_2, \ldots, u_m, 0, \ldots, 0]^{\mathrm{T}} \in \Re^n$.

$\|U(k)\|^2 \leqslant \overline{u}$. $\sigma(\cdot) \in \Re^m$ is sigmoid vector functions, $\phi(\cdot)$ is $\Re^{n\times n}$ diagonal matrix,

$$\sigma[W_1(k)x(k)] = \left[\sigma_1\left(\sum_{j=1}^{n} w_{1j}^1 x_j\right), \sigma_2\left(\sum_{j=1}^{n} w_{2j}^1 x_j\right), \ldots, \sigma_n\left(\sum_{j=1}^{n} w_{nj}^1 x_j\right)\right]^{\mathrm{T}}$$

$$\phi[W_2(k)x(k)]U(k) = \left[\phi_1\left(\sum_{j=1}^{n} w_{1j}^2 x_j\right)u_1, \phi_2\left(\sum_{j=1}^{n} w_{2j}^2 x_j\right)u_2, \ldots, \phi_n\left(\sum_{j=1}^{n} w_{nj}^2 x_j\right)u_n\right]^{\mathrm{T}}$$

The identified nonlinear system is represented as (1), which is bounded-input and bounded-output (BIBO) stable, i.e., $x(k)$ and $u(k)$ are bounded. According to the Stone–Weierstrass theorem [3], this nonlinear system can be written as

$$\beta x(k+1) = Ax(k) + \sigma[W_1^* x(k)] + \phi[W_2^* x(k)]U(k) + \mu(k) \tag{5}$$

where W_1^* and W_2^* are constant weights which can minimize the modeling error $\mu(k)$. Since σ and ϕ are bounded functions, $\mu(k)$ is bounded as $\mu^2(k) \leqslant \bar{\mu}$, $\bar{\mu}$ is an unknown positive constant. $\mu^2(k)$ is vector norm which is defined as $\mu^2(k) = \mu_1^2(k) + \cdots + \mu_n^2(k)$, $\mu(k) = [\mu_1(k), \ldots, \mu_n(k)]^{\mathrm{T}}$. This is corrected in the finite horizon, $k < K$, for which system identification has been carried out. In the infinite time horizon, Stone–Weierstrass theorem cannot be applied. But the plant is assumed BIBO, $\mu^2(k)$ is also bounded. The neuro identification error is defined as

$$e(k) = \hat{x}(k) - x(k)$$

From (5) and (4)

$$\begin{aligned}\beta e(k+1) = {} & Ae(k) + \{\sigma[W_1(k)x(k)] - \sigma[W_1^* x(k)]\} \\ & + \{\phi[W_2(k)x(k)] - \phi[W^* x(k)]\}U(k) - \mu(k)\end{aligned}$$

Using Taylor series around the points of $W_1(k)x(k)$

$$\sigma[W_1(k)x(k)] - \sigma[W_1^* x(k)] = \sigma'[W_1(k)x(k)]\widetilde{W}_1(k)x(k) + \varepsilon_1(k) \tag{6}$$

where $\widetilde{W}_1(k) = W_1(k) - W_1^*$, $\varepsilon_1(k)$ is second order approximation error. σ' is the derivative of nonlinear activation function $\sigma(\cdot)$ at the point of $W_1(k)$. Since ϕ is a sigmoid activation function, $\varepsilon_1(k)$ is bounded as $\|\varepsilon_1(k)\|^2 \leqslant \bar{\varepsilon}_1$, $\bar{\varepsilon}_1$ is an unknown positive constant. Similar

$$\phi[W_2(k)x(k)]U(k) - \phi[W_2^* x(k)]U(k) = u(k)\phi'[W_2(k)x(k)]\widetilde{W}_2(k)x(k) + \varepsilon_2(k) \tag{7}$$

where $u(k) = \operatorname{diag}(u_i)$, $\widetilde{W}_2(k) = W_2(k) - W_2^*$, $\|\varepsilon_2(k)\|^2 \leqslant \bar{\varepsilon}_2$, $\bar{\varepsilon}_2$ is an unknown positive constant. So

$$\beta e(k+1) = Ae(k) + \sigma'\widetilde{W}_1(k)x(k) + u(k)\phi'\widetilde{W}_2(k)x(k) + \zeta(k) \tag{8}$$

where $\zeta(k) = \varepsilon_1(k) + \varepsilon_2(k) - \mu(k)$. The following theorem gives a stable learning algorithm of discrete-time single-layer neural network.

Theorem 2. *If the single-layer neural network* (4) *is used to identify nonlinear plant* (1) *and the eigenvalues of* A *is selected as* $-1 < \lambda(A) < 0$, *the following gradient updating law without robust modification can make the identification error* $e(k)$ *bounded* (*stable in an* L_∞ *sense*)

$$\begin{aligned} W_1(k+1) &= W_1(k) - \eta(k)\sigma' x(k)e^{\mathrm{T}}(k) \\ W_2(k+1) &= W_2(k) - \eta(k)u(k)\phi' x(k)e^{\mathrm{T}}(k) \end{aligned} \tag{9}$$

where $\eta(k)$ *satisfies*

$$\eta(k) = \begin{cases} \dfrac{\eta}{1 + \|\sigma' x(k)\|^2 + \|u(k)\phi' x(k)\|^2} & \text{if } \beta\|e(k+1)\| \geqslant \|e(k)\| \\ 0 & \text{if } \beta\|e(k+1)\| < \|e(k)\| \end{cases}$$

$0 < \eta \leqslant 1$.

Proof. Select Lyapunov function as

$$V(k) = \|\widetilde{W}_1(k)\|^2 + \|\widetilde{W}_2(k)\|^2$$

where $\|\widetilde{W}_1(k)\|^2 = \sum_{i=1}^{n} \tilde{w}_1(k)^2 = \mathrm{tr}\{\widetilde{W}_1^{\mathrm{T}}(k)\widetilde{W}_1(k)\}$. From the updating law (9)

$$\widetilde{W}_1(k+1) = \widetilde{W}_1(k) - \eta(k)\sigma' x(k)e^{\mathrm{T}}(k)$$

So

$$\begin{aligned} \Delta V(k) &= V(k+1) - V(k) \\ &= \|\widetilde{W}_1(k) - \eta(k)\sigma' x(k)e(k)^{\mathrm{T}}\|^2 - \|\widetilde{W}_1(k)\|^2 \\ &\quad + \|\widetilde{W}_2(k) - \eta(k)u(k)\phi' x(k)e^{\mathrm{T}}(k)\|^2 - \|\widetilde{W}_2(k)\|^2 \\ &= \eta^2(k)\|e(k)\|^2\|\sigma' x(k)\|^2 - 2\eta(k)\|\sigma'\widetilde{W}_1(k)x(k)e^{\mathrm{T}}(k)\| \\ &\quad + \eta^2(k)\|e(k)\|^2\|u(k)\phi' x(k)\|^2 - 2\eta(k)\|u(k)\phi'\widetilde{W}_2(k)x(k)e^{\mathrm{T}}(k)\| \end{aligned}$$

There exist a constant $\beta > 0$, such that if $\|\beta e(k+1)\| \geqslant \|e(k)\|$, using (8) and $\eta(k) \geqslant 0$,

$$\begin{aligned} & -2\eta(k)\|\sigma'\widetilde{W}_1(k)x(k)e^{\mathrm{T}}(k)\| - 2\eta(k)\|u(k)\phi'\widetilde{W}_2(k)x(k)e^{\mathrm{T}}(k)\| \\ & \leqslant -2\eta(k)\|e^{\mathrm{T}}(k)\|\|\beta e(k+1) - Ae(k) - \zeta(k)\| \\ & = -2\eta(k)\|e^{\mathrm{T}}(k)\beta e(k+1) - e^{\mathrm{T}}(k)Ae(k) - e^{\mathrm{T}}(k)\zeta(k)\| \\ & \leqslant -2\eta(k)\|e^{\mathrm{T}}(k)\beta e(k+1)\| + 2\eta(k)e^{\mathrm{T}}(k)Ae(k) + 2\eta(k)\|e^{\mathrm{T}}(k)\zeta(k)\| \\ & \leqslant -2\eta(k)\|e(k)\|^2 + 2\eta(k)\lambda_{\max}(A)\|e(k)\|^2 + \eta(k)\|e(k)\|^2 + \eta(k)\|\zeta(k)\|^2 \end{aligned}$$

Since $0 < \eta \leqslant 1$,

$$\begin{aligned}\Delta V(k) &\leqslant \eta^2(k)\|e(k)\|^2\|\sigma' x(k)\|^2 + \eta^2(k)\|e(k)\|^2\|u(k)\phi' x(k)\|^2 \\ &\quad - \eta(k)\|e(k)\|^2 + 2\eta(k)\lambda_{\max}(A)\|e(k)\|^2 + \eta(k)\|\zeta(k)\|^2 \\ &= -\eta(k)\left[(1 - 2\lambda_{\max}(A)) - \eta\frac{\|\sigma' x(k)\|^2 + \|u(k)\phi' x(k)\|^2}{1 + \|\sigma' x(k)\|^2 + \|u(k)\phi' x(k)\|^2}\right]e^2(k) \\ &\quad + \eta_k\zeta^2(k) \\ &\leqslant -\pi e^2(k) + \eta\zeta^2(k) \end{aligned} \tag{10}$$

where

$$\pi = \frac{\eta}{1+\kappa}\left[1 - 2\lambda_{\max}(A) - \frac{\kappa}{1+\kappa}\right], \quad \kappa = \max_k(\|\sigma' x(k)\|^2 + \|u(k)\phi' x(k)\|^2)$$

Since $-1 < \lambda(A) < 0$, $\pi > 0$

$$n\min(\tilde{w}_i^2) \leqslant V_k \leqslant n\max(\tilde{w}_i^2)$$

where $n \times \min(\tilde{w}_i^2)$ and $n \times \max(\tilde{w}_i^2)$ are $\mathscr{K}_\infty$-functions, and $\pi e^2(k)$ is an $\mathscr{K}_\infty$-function, $\eta\zeta^2(k)$ is a $\mathscr{K}$-function, so V_k admits the smooth ISS-Lyapunov function as in Definition 2. From Theorem 1, the dynamic of the identification error is input-to-state stable. The "INPUT" is corresponded to the second term of the last line in (10), i.e., the modeling error $\zeta(k) = \varepsilon_1(k) + \varepsilon_2(k) - \mu(k)$, the "STATE" is corresponded to the first term of the last line in (10), i.e., the identification error $e(k)$. Because the "INPUT" $\zeta(k)$ is bounded and the dynamic is ISS, the "STATE" $e(k)$ is bounded.

If $\beta\|e(k+1)\| < \|e(k)\|$, $\Delta V(k) = 0$. $V(k)$ is constant, $W_1(k)$ is constant. Since $\|e(k+1)\| < \frac{1}{\beta}\|e(k)\|$, $\frac{1}{\beta} < 1$, $e(k)$ is bounded. □

Remark 1. The condition $\beta\|e(k+1)\| \geqslant \|e(k)\|$ is dead-zone. If β is selected big enough, the dead-zone becomes small.

Remark 2. The class of networks considered in this paper is non-linear in the weights as in [16,18]. Due to slow learning convergence and high data sets, many practical implementations of neural networks are linear in the weights as in [6,19],

$$\beta\hat{x}(k+1) = A\hat{x}(k) + W_1(k)\sigma[x(k)] \tag{11}$$

In this case identification error dynamic (8) becomes

$$\begin{aligned}\beta e(k+1) &= Ae(k) + W_k\sigma[x(k)] - W^*\sigma[x(k)] + \mu(k) \\ &= Ae(k) + \widetilde{W}_k\sigma[x(k)] + \mu(k)\end{aligned}$$

The following updating law

$$W_{k+1} = W_k - \eta(k)e(k)\sigma^{\mathrm{T}} \tag{12}$$

with η_k satisfying

$$\eta(k) = \begin{cases} \dfrac{\eta}{1+\|\sigma\|^2}, \quad 0 < \eta \leqslant 1 & \text{if } \beta\|e(k+1)\| \geqslant \|e(k)\| \\ 0 & \text{if } \beta\|e(k+1)\| < \|e(k)\| \end{cases}$$

can makes that the identification error $e(k)$ bounded. The proof procedures are the same as Theorem 2.

4. System identification with discrete-time multilayer recurrent neural networks

In order to simplify calculation process, the following discrete-time nonlinear system without control input is discussed

$$x(k+1) = f[x(k)] \tag{13}$$

The discrete-time multilayer recurrent neural networks is represented as

$$\beta\hat{x}(k+1) = A\hat{x}(k) + V_k\sigma[W_k x(k)] \tag{14}$$

where the weights in output layer are $V_k \in R^{1\times m}$, the weights in hidden layer are $W_k \in R^{m\times n}$, σ is m-dimension vector function $\sigma = [\sigma_1 \cdots \sigma_n]^{\mathrm{T}}$. The typical presentation of the element $\sigma_i(\cdot)$ is sigmoid function. The structure of the discrete-time multilayer recurrent neural networks is shown in Fig. 2. MLP is discrete-time multilayer perceptrons.

If the nonlinear plant (13) has a control input as

$$x(k+1) = f[x(k), u(k)]$$

The corresponding discrete-time recurrent neural networks is

$$\beta\hat{x}(k+1) = A\hat{x}(k) + V_{1,k}\sigma[W_{1,k}x(k)] + V_{2,k}\phi[W_{2,k}x(k)]U(k)$$

where $U(k) = [u_1, u_2, \ldots, u_m, 0, \ldots, 0]^{\mathrm{T}}$.

The neuro identification discussed in this paper is a kind of on-line identification, i.e., the weights are always updated by the identification error $e(k)$,

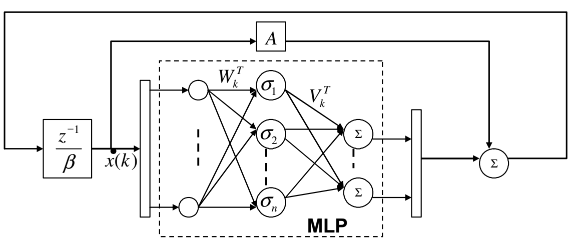

Fig. 2. Discrete-time multilayer recurrent neural networks.

without $e(k)$ the neural networks model cannot be used separately. The identified nonlinear system (13) can be represented by

$$\beta x(k+1) = Ax(k) + V^* \sigma[W^* x(k)] - \mu(k) \tag{15}$$

where V^* and W^* are set of unknown weights which may minimize the modeling error $\mu(k)$. In the case of two independent variables, a smooth function f has Taylor formula as

$$f(x_1, x_2) = \sum_{k=0}^{l-1} \frac{1}{k!} \left[(x_1 - x_1^0) \frac{\partial}{\partial x_1} + (x_2 - x_2^0) \frac{\partial}{\partial x_2} \right]_0^k f + R_l$$

where R_l is the remainder of the Taylor formula. Let x_1 and x_2 correspond V^* and W^*, x_1^0 and x_2^0 correspond V_k and W_k. Using Taylor series around the point of $W_k X(k)$ and V_k, the identification error $e(k) = \hat{x}(k) - x(k)$ can be represented as

$$\beta e(k+1) = Ae(k) + \widetilde{V}_k \sigma[W_k x(k)] + V_k \sigma' \widetilde{W}_k x(k) + \zeta(k) \tag{16}$$

where σ' is the derivative of nonlinear activation function $\sigma(\cdot)$ at the point of $W_k X(k)$, $\widetilde{W}_k = W_k - W^*$, $\widetilde{V}_k = V_k - V^*$, $\zeta(k) = R_1 + \mu(k)$, here R_1 is second order approximation error of the Taylor series.

This paper only discusses open-loop identification, so the plant (13) can be assumed to be bounded stable, i.e., $x(k)$ in (13) is bounded. By the boundness of the sigmoid function σ, $\mu(k)$ in (15) is bounded, also R_1 is bounded. So $\zeta(k)$ in (16) is bounded. Compared to normal backpropagation algorithm (2) and (3), the following theorem gives a stable backpropagation-like algorithm for discrete-time multilayer neural network.

Theorem 3. *The multilayer neural network* (14) *is used to identify nonlinear plant* (13) *and A is selected as $-1 < \lambda(A) < 0$, the following gradient updating law without robust modification can make identification error $e(k)$ bounded*

$$\begin{aligned} W_{k+1} &= W_k - \eta_k e(k) \sigma' V_k^{\mathrm{T}} x(k) \\ V_{k+1} &= V_k - \eta_k e(k) \sigma^{\mathrm{T}} [W_k(k) x(k)] \end{aligned} \tag{17}$$

where η_k satisfies

$$\eta_k = \begin{cases} \dfrac{\eta}{1 + \|\sigma' V_k^{\mathrm{T}} x(k)\|^2 + \|\sigma\|^2} & \text{if } \beta\|e(k+1)\| \geqslant \|e(k)\| \\ 0 & \text{if } \beta\|e(k+1)\| < \|e(k)\| \end{cases}$$

$0 < \eta \leqslant 1$. *The average of the identification error satisfies*

$$J = \limsup_{T \to \infty} \frac{1}{T} \sum_{k=1}^{\mathrm{T}} e^2(k) \leqslant \frac{\eta}{\pi} \bar{\zeta} \tag{18}$$

where

$$\pi = \frac{\eta}{1+\kappa}\left[1 - \frac{\kappa}{1+\kappa}\right] > 0, \quad \kappa = \max_k(\|\sigma' V_k^{\mathrm{T}} x(k)\|^2 + \|\sigma\|^2),$$
$$\bar{\zeta} = \max_k[\zeta^2(k)]$$

$T > 0$ *is identification time.*

Proof. Selected a positive scalar L_k as

$$L_k = \|\widetilde{W}_k\|^2 + \|\widetilde{V}_k\|^2 \tag{19}$$

From the updating law (17),

$$\widetilde{W}_{k+1} = \widetilde{W}_k - \eta_k e(k)\sigma' V_k^{\mathrm{T}} x^{\mathrm{T}}(k), \quad \widetilde{V}_{k+1} = \widetilde{V}_k - \eta_k e(k)\sigma^{\mathrm{T}}$$

Since ϕ' is diagonal matrix, and by using (16)

$$\begin{aligned}\Delta L_k &= \|\widetilde{W}_k - \eta_k e(k)\sigma' V_k^{\mathrm{T}} x^{\mathrm{T}}(k)\|^2 + \|\widetilde{V}_k - \eta_k e(k)\sigma^{\mathrm{T}}\|^2 - \|\widetilde{W}_k\|^2 - \|\widetilde{V}_k\|^2 \\ &= \eta_k^2 e^2(k)(\|\sigma' V_k^{\mathrm{T}} x^{\mathrm{T}}(k)\|^2 + \|\sigma\|^2) - 2\eta_k\|e(k)\|\|V_k^{\mathrm{T}}\sigma'\widetilde{W}_k X(k) + \widetilde{V}_k\sigma\| \\ &= \eta_k^2 e^2(k)(\|\sigma' V_k^{\mathrm{T}} x^{\mathrm{T}}(k)\|^2 + \|\sigma\|^2) - 2\eta_k\|e(k)[\beta e(k+1) - Ae(k) - \zeta(k)]\|\end{aligned} \tag{20}$$

Using (16) and $\eta(k) \geqslant 0$, there exist a constant $\beta > 0$, such that

If $\|\beta e(k+1)\| \geqslant \|e(k)\|$

$$\Delta L_k \leqslant -\eta_k e^2(k)[1 - \eta_k(\|\sigma' V_k^{\mathrm{T}} x^{\mathrm{T}}(k)\|^2 + \|\sigma\|^2)] + \eta\zeta^2(k) \leqslant -\pi e^2(k) + \eta\zeta^2(k)$$

where π is defined in (18). Because

$$n[\min(\tilde{w}_i^2) + \min(\tilde{v}_i^2)] \leqslant L_k \leqslant n[\max(\tilde{w}_i^2) + \max(\tilde{v}_i^2)]$$

where $n[\min(\tilde{w}_i^2) + \min(\tilde{v}_i^2)]$ and $n[\max(\tilde{w}_i^2) + \max(\tilde{v}_i^2)]$ are $\mathscr{K}_\infty$-functions, and $\pi e^2(k)$ is an $\mathscr{K}_\infty$-function, $\eta\zeta^2(k)$ is a $\mathscr{K}$-function. From (16) and (19), V_k is the function of $e(k)$ and $\zeta(k)$, so L_k admits a smooth ISS-Lyapunov function as in Definition 2. From Theorem 1, the dynamic of the identification error is input-to-state stable. Because the "INPUT" $\zeta(k)$ is bounded and the dynamic is ISS, the "STATE" $e(k)$ is bounded.

If $\beta\|e(k+1)\| < \|e(k)\|$, $\Delta V(k) = 0$. $L(k)$ is constant, the constants of W_k and V_k means $e(k)$ is bounded.

Eq. (20) can be rewritten as

$$\Delta L_k \leqslant -\pi e^2(k) + \eta\zeta^2(k) \leqslant \pi e^2(k) + \eta\bar{\zeta} \tag{21}$$

Summarizing (21) from 1 up to T, and by using $L_T > 0$ and L_1 is a constant

$$L_T - L_1 \leqslant -\pi\sum_{K=1}^{\mathrm{T}} e^2(k) + T\eta\bar{\zeta}$$

$$\pi\sum_{K=1}^{\mathrm{T}} e^2(k) \leqslant L_1 - L_T + T\eta\bar{\zeta} \leqslant L_1 + T\eta\bar{\zeta}$$

Eq. (18) is established. □

Remark 3. The normalizing learning rates $\eta(k)$ in (9) and (17) are time-varying in order to assure the identification processes are stable. These learning gains are easier to be decided, no any prior information is required, for example $\eta = 1$. This learning law is simpler to use, because the user does not need to care about how to select a better learning rate to assure both fast convergence and stability. No any previous information is required. The contradiction in fast convergence and stable learning may be avoided.

Remark 4. Because $\|\sigma' V_k^{\mathrm{T}} x(k)\|^2 + \|\sigma\|^2 \geqslant 0$, $\|\sigma' x(k)\|^2 + \|u(k)\phi' x(k)\|^2 \geqslant 0$, $1 \geqslant \eta > 0$, the upper and lower bounds for the adaptive learning rate $\eta(k)$ in (9) and (17) is

$$1 \geqslant \eta(k) \geqslant 0$$

If η is selected as dead-zone function:

$$\begin{cases} \eta = 0 & \text{if } |e(k)| \leqslant \bar{\zeta} \\ \eta = \eta_0 & \text{if } |e(k)| > \bar{\zeta} \end{cases}$$

Eq. (9) is the same as [18,26]. If a σ-modification term or modified δ-rule term are added in (9), it becomes that of [9] or that of [14]. But all of them need the upper bound of modeling error $\bar{\zeta}$. And the identification error is enlarged by the robust modifications [8].

Remark 5. Since neural networks cannot match nonlinear systems exactly, the parameters (weights) cannot converge. It can only be assured the output of neural networks follows the output of the plant, i.e. the identification error is stable. Although the weights cannot converge to their optimal values, Eq. (18) shows that the identification error will converge to the ball radius $\frac{\eta}{\pi}\bar{\zeta}$. Even if the input is persistent exciting, the modeling error $\zeta(k)$ will not make the weights convergent to their optimal values. It is possible that the output error is convergent, but the weight errors are very high when the networks structure are not fine defined. The relations of the output error and the weight errors are shown in (16). Simpler case is that the weights is linear with the networks and the neural networks can match the nonlinear plant exactly

plant: $y = W^*\sigma[x(k)]$

neural networks: $\hat{y} = W_k\sigma[x(k)]$

output error: $(y - \hat{y}) = (W^* - W_k)\sigma[x(k)]$

If $\sigma[x(k)]$ is large, small output error $(y - \hat{y})$ does not mean good convergence of the weight error $(W^* - W_t)$.

Remark 6. Noise (or disturbance) is an important issue in the system identification. There are two types of disturbances: external and internal. Internal disturbance can be regarded as unmodeled dynamic $\mu(k)$ in (15). A bounded internal disturbance does not effect the theory results in this paper, but can enlarge the identification error if the internal disturbance becomes bigger. External disturbance can be regarded as measurement noise, input noise, etc. In the point of structure, input noises are increased feedforward through each layer [2]. For example, a noise $\varsigma(k)$ is multiple by $V_k\sigma[W_k\varsigma(k)]$ and achieves the output. Measurement noise is enlarged due to backpropagation of identification error (17), therefore the weights of neural networks are influenced by output noise. On the other hand small external disturbance can accelerate convergent rate according to the persistent exciting theory [16], small disturbances in the input $u(t)$ or in output $y(t)$ can enrich frequency of the signal $X(t)$, this is good for parameters convergence. In the following simulation one can see this point.

5. Simulation

In this section one typical chaotic system, Lorenz model, is chosen to demonstrate the abilities of neuro-identifier. Lorenz model is used for the fluid conviction description especially for some feature of atmospheric dynamic [23]. The uncontrolled model is given by

$$\begin{aligned} \dot{x}_1 &= -\beta x_1 + x_2 x_3 \\ \dot{x}_2 &= \omega(x_3 - x_2) \\ \dot{x}_3 &= -x_1 x_2 + \rho x_2 - x_3 \end{aligned} \tag{22}$$

where x_1, x_2 and x_3 represent measures of vertical and horizontal temperatures, fluid velocity, correspondingly. ω, ρ and β are positive parameters representing the Prandtl number, Rayleigh number and geometric factor. If $\rho < 1$, the origin is a stable equilibrium. If

$$1 < \rho \leqslant \rho^*(\sigma, \beta) := \omega(\omega + \beta + 3)/(\omega - \beta - 1), \tag{23}$$

the system has two stable equilibrium points $(\pm\sqrt{\beta(\rho - 1)}, \pm\sqrt{\beta(\rho - 1)}, (\rho - 1))$ and one unstable equilibrium (the origin). If $\rho^*(\omega, \beta) < \rho$, all three equilibrium points become unstable. As in the commonly studied case, in this paper $\omega = 10$ and $\beta = 8/3$, that leads to $\rho^*(\omega, \beta) = 24.74$. In this example the system is considered as $\omega = 10$, $\beta = 8/3$ and $\rho = 28$. The dynamic of this chaotic system is shown in Fig. 3.

Following difference technique is used to get the discrete-time states of the system (22). (22) can be written as

$$\dot{x} = Ax$$

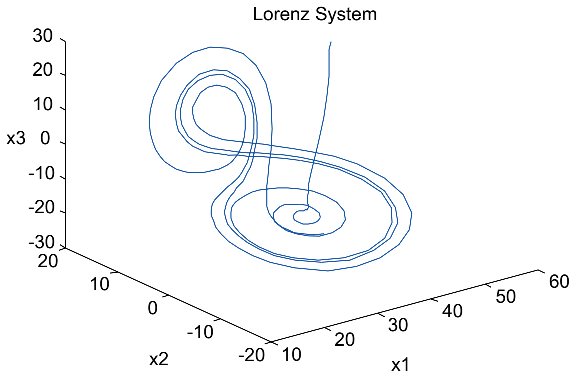

Fig. 3. Chaotic dynamic of Lorenz system.

with

$$x = [x_1, x_2, x_3]^{\mathrm{T}}, \quad A = \begin{bmatrix} -\beta & 0 & x_2 \\ 0 & -\sigma & \sigma \\ -x_2 & \rho & -1 \end{bmatrix}$$

Let us define $s_1 = Ax_k$, $s_2 = A(x_k + s_1)$, $s_3 = A\left(x_k + \frac{s_1+s_2}{4}\right)$. If

$$\left|\frac{s_1 - 2 \times s_3 + s_2}{3}\right| \leqslant \frac{\|x_k\|}{1000} \quad \text{or} \quad \left|\frac{s_1 - 2 \times s_3 + s_2}{3}\right| < 1$$

then

$$x_{k+1} = x_k + \frac{s_1 + 4s_3 + s_2}{6}, \quad k = 0, 1, 2, \ldots, \quad x_0 = [50, 10, 20]^{\mathrm{T}}$$

First single layer neural network (14) is used to identify (22),

$$\beta\hat{x}(k+1) = A\hat{x}(k) + W_1(k)\sigma[x(k)]$$

where $\beta = 4$, $\sigma(\cdot) = \tanh(\cdot)$, $\hat{x} = [\hat{x}_1, \hat{x}_2, \hat{x}_3]^{\mathrm{T}}$, $A = \mathrm{diag}[0.8, 0.8, 0.8]$, $W_1(k) \in \Re^{3\times 3}$. The elements of $W_1(0)$ are random number between [0,1]. The updating law for the weight is (9)

$$W_1(k+1) = W_1(k) - \frac{1}{1 + \|\sigma\|^2} e(k)\sigma^{\mathrm{T}}$$

The on-line identification result is shown in Fig. 4. Theorem 2 gives a sufficient condition of η for stable learning, $\eta \leqslant 1$. In this example, the simulation shows that if $\eta \geqslant 2.5$, the learning process becomes unstable. The neuro identification discussed in this paper is on-line, the convergence of the weight is not studied. The weights do not converge to some constants or optimal values. The total simulation time is 600, there are 14 times $\beta\|e(k+1)\| < \|e(k)\|$.

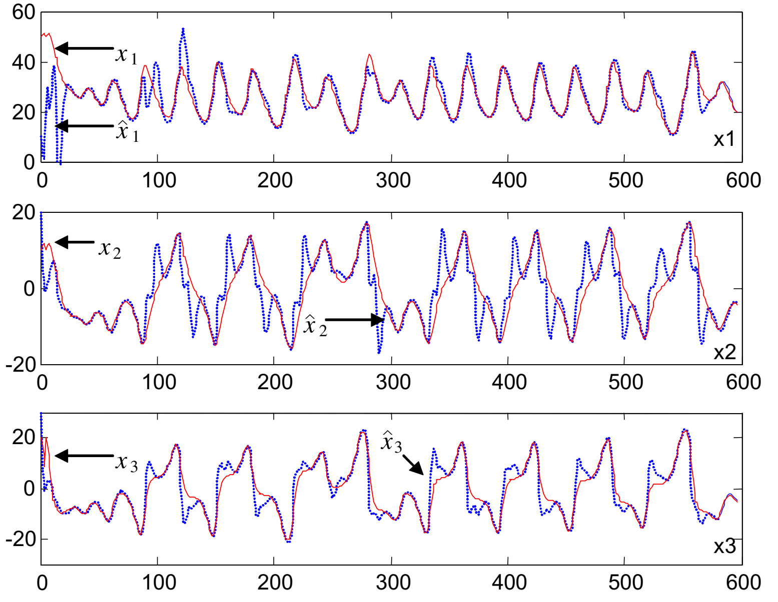

Fig. 4. Single layer neural network identification.

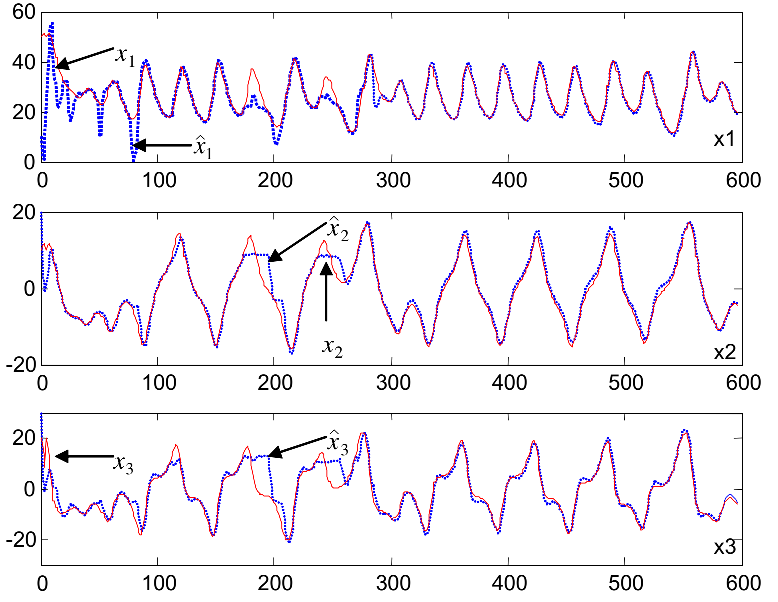

Fig. 5. Multilayer neural network identification.

Second multilayer neural network (14) is used to identify (22),

$$\beta\hat{x}(k+1) = A\hat{x}(k) + V_k\sigma[W_k(k)x(k)]$$

where $\beta = 4$, $\sigma(\cdot) = \tanh(\cdot)$, $A = \mathrm{diag}[0.8, 0.8, 0.8]$. Model complexity is important in the context of system identification, which is corresponded to the number of hidden units of the neuro model. In this simulation different numbers of hidden nodes is tested, the simulation shows that after the hidden nodes is more than 20, the identification accuracy will not be improved a lot. In [15], they also used 20 hidden nodes for the first hidden layer. So $W_k \in R^{20\times 3}$, $V_k \in R^{3\times 20}$. The learning algorithm for W_k and V_k are (17) with $\eta_k = \frac{1}{1+\|\sigma' V_k^{\mathrm{T}} x(k)\|^2+\|\sigma\|^2}$, $\sigma'(\cdot) = \sec h^2(\cdot)$. The initial conditions for W_k and V_k are random number, i.e., $W_0 = \mathrm{rand}(\cdot)$, $V_0 = \mathrm{rand}(\cdot)$. The on-line identification result is shown in Fig. 5. The total simulation time is 600, there are only 1 time when $\beta\|e(k+1)\| < \|e(k)\|$.

Compared to recurrent backpropagation [11], the time-varying learning rate η_k in (17) is easier to be realized.

6. Conclusion

In this paper, nonlinear system identification via discrete-time recurrent single layer and multilayer neural networks are studied. By using ISS approach, this paper gives conclusions that the commonly used robustifying techniques for system identification, such as dead-zone and σ-modification, are not necessary for the gradient descent law and the backpropagation-like algorithm when discrete-time recurrent neural networks are used.

References

[1] E. Barn, Optimization for training neural nets, IEEE Trans. Neural Networks 3 (2) (1992) 232–240.

[2] M. Brown, C.J. Harris, Neurofuzzy Adaptive Modelling and Control, Prentice-Hall, Upper Saddle River, NJ, 1994.

[3] G. Cybenko, Approximation by superposition of sigmoidal activation function, Math.Control, Sig. Syst. 2 (1989) 303–314.

[4] B. Egardt, Stability of adaptive controllers, in: Lecture Notes in Control and Information Sciences 20, Springer-Verlag, Berlin, 1979.

[5] Z. Feng, A.N. Michel, Robustness analysis of a class of discrete-time systems with applications to neural networks, in: Proceedings of American Control Conference, San Deigo, 1999, pp. 3479–3483.

[6] S.S. Ge, C.C. Hang, T.H. Lee, T. Zhang, Stable Adaptive Neural Network Control, Kluwer Academic, Boston, 2001.

[7] J.J. Hopfield, Neurons with grade response have collective computational properties like those of a two-state neurons, Proc. Natl. Acad. Sci., USA 81 (1984) 3088–3092.

[8] P.A. Ioannou, J. Sun, Robust Adaptive Control, Prentice-Hall, Upper Saddle River, NJ, 1996.

[9] S. Jagannathan, F.L. Lewis, Identification of nonlinear dynamical systems using multilayered neural networks, Automatica 32 (12) (1996) 1707–1712.
[10] R.A. Jacobos, Increased rates of convergence through learning rate adaptation, Neural Networks 1 (4) (1988) 295–308.
[11] L. Jin, M.M. Gupta, Stable dynamic backpropagation learning in recurrent neural networks, IEEE Trans. Neural Networks 10 (6) (1999) 1321–1334.
[12] Z.P. Jiang, Y. Wang, Input-to-state stability for discrete-time nonlinear systems, Automatica 37 (2) (2001) 857–869.
[13] E.B. Kosmatopoulos, M.M. Polycarpou, M.A. Christodoulou, P.A. Ioannou, High-order neural network structures for identification of dynamical systems, IEEE Trans. Neural Networks 6 (2) (1995) 431–442.
[14] F.L. Lewis, A. Yesildirek, K. Liu, Multilayer neural-net robot controller with guaranteed tracking performance, IEEE Trans. Neural Networks 7 (2) (1996) 388–399.
[15] K.S. Narendra, K. Parthasarathy, Identification and control of dynamical systems using neural networks, IEEE Trans. Neural Networks 1 (1) (1990) 4–27.
[16] M.M. Polycarpou, P.A. Ioannou, Learning and convergence analysis of neural-type structured networks, IEEE Trans. Neural Networks 3 (1) (1992) 39–50.
[17] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (1986) 533–536.
[18] Q. Song, Robust training algorithm of multilayered neural networks for identification of nonlinear dynamic systems, IEE Proc.—Control Theory Appl. 145 (1) (1998) 41–46.
[19] J.A.K. Suykens, J. Vandewalle, B. De Moor, Lur'e systems with multilayer perceptron and recurrent neural networks; absolute stability and dissipativity, IEEE Trans. Automatic Control 44 (1999) 770–774.
[20] J.A.K. Suykens, J. Vandewalle, B. De Moor, NLq theory: checking and imposing stability of recurrent neural networks for nonlinear modelling, IEEE Trans. Signal Process (special issue on neural networks for signal processing) 45 (11) (1997) 2682–2691.
[21] T. Tollenaere, SuperSAB: fast adaptive back progration with good scaling properties, Neural Networks 3 (5) (1990) 561–573.
[22] T.P. Vogl, J.K. Mangis, A.Z. Zigler, W.T. Zink, D.L. Alkon, Accelerating the convergence of the backprogration method, Biol. Cybernetics 59 (1988) 256–264.
[23] Wen Yu, Passive equivalence of chaos in Lorenz System, IEEE Trans. Circuits Syst., Part I 46 (7) (1999) 876–879.
[24] W. Yu, X. Li, Some stability properties of dynamic neural, networks, IEEE Trans. Circuits Syst., Part I 48 (1) (2001) 256–259.
[25] W. Yu, X. Li, Some new results on system identification with dynamic neural networks, IEEE Trans. Neural Networks 12 (2) (2001) 412–417.
[26] W. Yu, A.S. Poznyak, X. Li, Multilayer dynamic neural networks for nonlinear system on-line identification, Int. J. Control 74 (18) (2001) 1858–1864.
[27] Wen Yu, Xiaoou Li, Discrete-time neuro identification without robust modification, IEE Proc.—Control Theory Appl. 150 (3) (2003) 311–316.