第 20 卷第 2 期
2003 年 4 月
控制理论与应用
Control Theory & Applications
Vol.20 No.2
Apr.2003

Article ID: 1000 - 8152(2003)02 - 0205 - 06

Training of parameters and time delays of universal learning network with switching mechanism

HAN Min[1], XI Jian-hui[1], HIRASAWA Kotaro[2]

(1. School of Electronics and Information Engineering, Dalian University of Technology, Liaoning Dalian 116023, China;
2. Graduate School of Information Science and Electronic Engineering, Kyushu University, Fukuoka 812 - 8581, Japan)

Abstract: A new modeling method using the ULN with switching mechanism is studied, where both parameters and time delays are adjusted to model the nonlinear systems. The simulation results of nonlinear system identification problems show that better performances can be obtained by the proposed method than the conventional method using only the parameter optimization. And how the generalization ability of modeling dynamic systems is influenced by the network size is also studied using the ULN with switching mechanism.

Key words: universal learning network; random search; time delays; neural network; generalization ability

CLC number: TP183 **Document code**: A

带有开关机制的通用学习网络的参数和时间延迟的学习

韩 敏[1], 席剑辉[1], 平泽宏太郎[2]

(1. 大连理工大学 电子与信息工程学院, 辽宁 大连, 116023;
2. 九州大学 大学院系统信息科学与电子工程研究院, 日本 福冈 812 - 8581)

摘要: 研究了一种新的带有开关机制的 ULN 建模方法, 不仅对其中的参数而且对时间延迟都进行了调整以适应非线性系统的建模要求. 对非线性系统识别问题的仿真结果表明, 所提方法具有比仅使用参数优化的传统方法更好的性能. 还用带有开关机制的 ULN 对网络规模是如何影响动态系统泛化能力的问题进行了研究.

关键词: 通用学习网络; 随机搜索; 延迟时间; 神经网络; 泛化能力

1 Introduction

Neural networks have been widely studied in recent years. Based on a natural extension of the well known recurrent neural networks[1], Hirasawa and his coworkers have proposed a universal learning network (ULN)[2,3], which is constructed by arbitrarily connecting nonlinearly operated nodes to each other with multi-branches that may have arbitrary time delays including zero or minus ones, so that it can be used to model and control large scale complicated systems naturally. In our previous papers[4], a switching mechanism has been introduced into the ULN in order to realize an optimal modeling of dynamic systems. It has been shown that a compromised model in terms of model error and model parsimony can be achieved by appropriately training the parameters of the ULN with switching mechanism[4~6]. On the other hand, one of the distinctive features of the ULN is that the ULN can have arbitrary time delays between the nodes. Simulation results show that appropriate time delays can improve the performance of the ULN significantly.

In this paper, we shall study how the generalization ability of the ULN can be improved by taking advantage of the switching mechanism and the ability of having arbitrary time delays between the nodes. A new training method for the ULN is presented, in which both the parameters of the ULN with switching mechanism and the time delays between the nodes are adjusted. The effectiveness of the proposed method is investigated via numerical simulations of nonlinear system identification.

2 Basic structure of the ULN

The structure of the ULN is shown in Fig. 1, where

Received date: 2000 - 06 - 06; Revised date: 2002 - 06 - 03.
Foundation item: supported by the Project 973 of China (G1999043602); the National Natural Science Foundation of China(50139020).

N_i denotes the node i, $D_{ij}(p)$ and $\alpha_{ij}(p)$ are the time delay and switching function on the pth branch between node i and node j. It can be seen that ULN provides a fully recurrent connection between nodes. The number of nodes, the number of branches between nodes and the time delays they represent are flexible.

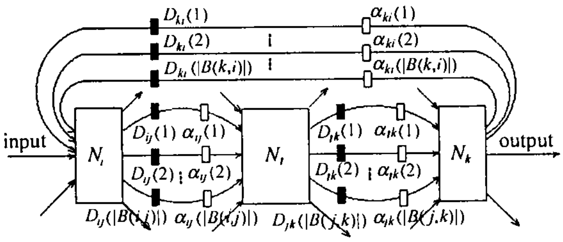


Fig. 1 Structure of the ULN with multi-branches and switching mechanisms

The output value of node j, $h_j(t)$, can be described by

$$h_j(t) = O_j(\{h_i(t - D_{ij}(p)) \mid i \in JF(j), p \in B(i,j)\}, \{r_n(t) \mid n \in N(j)\}, \{\lambda_m(t) \mid m \in M(j)\}), \; j \in J, \; t \in T, \tag{1}$$

where O_j denotes nonlinear function of node j, $\lambda_m(t)$ is value of m-th parameter, $r_n(t)$ is value of n-th external input variable, $JF(j)$ is set of numbers of nodes whose outputs are connected to node j, $B(i,j)$ is set of numbers of branches from node i to node j, $N(j)$ is set of numbers of external input variables of node j, $M(j)$ is set of numbers of parameters, where output of node j can be partially differentiable with respect to these parameters, J is set of numbers of nodes, and T is set of sampling times.

The switching function $\alpha_{ij}(p)$, used for controlling branch deletion, is supposed to be

$$\alpha_{ij}(p) = \frac{1}{1 + e^{-\Psi\beta_{ij}(p)}}, \tag{2}$$

where $\beta_{ij}(p)$ is adjustable parameter and Ψ is a slope factor. $\beta_{ij}(p)$ can be adjusted in the same way as other parameters. Ψ will be set to a small value at the beginning of the training, and scheduled to increase gradually as the learning progresses. In this way, an upper bound $\alpha_{ij}(p) = 1.0$ or a lower bound $\alpha_{ij}(p) = 0.0$, which means the connection and disconnection of the pth branch from node i to node j respectively, can be determined by comparing $\alpha_{ij}(p)$ with a threshold value. If $\alpha_{ij}(p)$ is less than the threshold, then $\alpha_{ij}(p) = 0$ and this branch will be deleted. The switching function is shown in Fig.2.

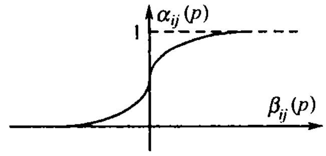


Fig. 2 Switching function

3 Training algorithm for the ULN

Generally, there are two ways to improve the generalization ability of a network. One is to increase the training data. The second is to optimize the structure of the network. We mainly use the second method that is realized by introducing switching mechanism and extended criterion function into the ULN. Furthermore, the generalization ability can also be improved by appropriately training the time delays between the nodes.

3.1 Training of the parameters λ_m and $\beta_{ij}(p)$

The training of the parameters λ_m and $\beta_{ij}(p)$ will be performed by minimizing an extended criterion function L including the parsimony of the model, which are given by

$$L = E + R_\alpha \sum_i \sum_j \sum_p (\alpha_{ij}(p))^2, \quad E = E(\{h_j(t)\}, \{\lambda_m(t)\}), \tag{3}$$

where E is supposed to be usual criterion function, R_α is weighting coefficient that can balance the criterion function E and the parsimony of network.

In references (2) and (7), a unified training algorithm has been proposed for the ULN. The algorithm is distinctive in that it is capable of dealing with arbitrary time delays and multi-branches between the nodes of ULN. Based on the same scheme, the parameters λ_m and $\beta_{ij}(p)$ in the ULN with switching mechanism can be adjusted using the following algorithm[2,7]:

$$\lambda_m(t) = \lambda_m(t) - \gamma \frac{\partial^+ L}{\partial \lambda_m}, \tag{4}$$

$$\beta_{ij}(p) = \beta_{ij}(p) - \gamma \frac{\partial^+ L}{\partial \beta_{ij}(p)}, \tag{5}$$

$$\frac{\partial^+ L}{\partial \lambda_m} = \sum_{t' \in T} \sum_{d \in JD(\lambda_m)} \left[\frac{\partial h_d(t')}{\partial \lambda_m}\delta(d,t')\right] + \frac{\partial L}{\partial \lambda_m}, \tag{6}$$

$$\frac{\partial^+ L}{\partial \beta_{ij}(p)} = \sum_{t' \in T}\left[\frac{\partial h_j(t')}{\partial \beta_{ij}(p)}\delta(j,t')\right] + \frac{\partial L}{\partial \beta_{ij}(p)}, \tag{7}$$

$$\delta(j,t) = \sum_{k\in JB(j)}\sum_{p\in B(j,k)} [\frac{\partial h_k(t+D_{jk}(p))}{\partial h_j(t)} \cdot \delta(k,t+D_{jk}(p))] + \frac{\partial L}{\partial h_j(t)},$$
$$j \in J,\ t \in T, \tag{8}$$

where $\frac{\partial^+ L}{\partial \lambda_m}, \frac{\partial^+ L}{\partial \beta_{ij}(p)}$ are the ordered derivative proposed by Werbos[8], $JD(\lambda_m)$ is set of numbers of nodes whose output can be partially differentiable with respect to λ_m. When the nonlinear function of the node in the ULN is chosen to be bipolar sigmoidal functions shown in (9), further expressions are given by

$$h_j(t) = A\frac{1-e^{-\alpha_j}}{1+e^{-\alpha_j}}, \tag{9}$$

$$\alpha_j = \sum_{i\in JF(j)}\sum_{p\in B(i,j)} h_{ij}(p)\alpha_{ij}(p) \times h_i(t-D_{ij}(p)) + \theta_j, \tag{10}$$

$$\frac{\partial h_j(t)}{\partial h_{ij}(p)} = \frac{A}{2}\{1-(\frac{h_j(t)}{A})^2\}\alpha_{ij}(p) \times h_i(t-D_{ij}(p)), \tag{11}$$

$$\frac{\partial h_j(t)}{\partial \beta_{ij}(p)} = \frac{A}{2}\{1-(\frac{h_j(t)}{A})^2\}h_{ij}(p) \cdot h_i(t-D_{ij}(p))\ \varphi \times \alpha_{ij}(p)(1-\alpha_{ij}(p)), \tag{12}$$

$$\frac{\partial L}{\partial \beta_{ij}(p)} = 2R_a\varphi\,\alpha_{ij}(p)^2(1-\alpha_{ij}(p)), \tag{13}$$

$$\frac{\partial h_k(t+D_{jk}(p))}{\partial h_j(t)} = \frac{A}{2}\{1-(\frac{h_k(t+D_{jk}(p))}{A})^2\} \cdot h_{jk}(p)\alpha_{jk}(p), \tag{14}$$

where λ_m is denoted as $h_{ij}(p)$, the weight on pth branch from node i to node j, and θ_j is the external input.

3.2 Searching of the time delays

In the previous training algorithm, the time delays between the nodes of the ULN are fixed because it is difficult to train the time delays in the same way as training of the parameters since the time delays can only take integer values. In order to minimize extended criterion function, the time delays should also be adjusted[7,9]. In this paper, such training is performed by a combination of the following two searching procedures.

First, search the optimal time delays using a kind of random search procedure. It can be described as follows: select one branch randomly out of the branches, and calculate the criterion function L with the development of the time delay of this branch, for example, from 1 to 10. Find optimal time delay of the selected branch with the time delays of other branches being fixed. Here, the criterion function L takes the minimum value for the optimal time delay.

Then, the parameters of the ULN is adjusted using gradient method.

A combination of the above two training procedures for the time delays and for the parameters is repeated until the stop condition is satisfied.

4 Simulation of nonlinear system identification

To illustrate the investigation of the training and the generalization ability, a nonlinear system described by (15) was modeled by the ULN shown in Fig.1, which has 5 nodes denoted as N_1 to N_5 and has fully recurrent connections with three branches between every two nodes. So the initial branches is $5\times5\times3=75$. N_1 has an external input and the output of the ULN is obtained from the N_5. The teaching signal described in Fig.3 at the training stage was obtained by exciting the nonlinear system with the input signal obtained by (16).

$$y(k+1) = \begin{cases} 1.34y(k) - 0.277y(k-2) - \\ 0.80y(k-4) + 0.01,\ u(k) \geqslant 0, \\ 1.34y(k) - 0.277y(k-2) - \\ 0.80y(k-4) - 0.01,\ u(k) < 0, \end{cases} \tag{15}$$

$$u(k) = \begin{cases} \sin(\frac{\pi}{50}k), & 0 \leqslant k < 100, \\ 1.0, & 100 \leqslant k < 150, \\ -1.0, & 150 \leqslant k < 200, \\ \text{uniform radom number in}(-0.5,0.5), & 200 \leqslant k < 400. \end{cases} \tag{16}$$

Figure 4 and Fig.5 show the input-output data for investigating the generalization ability, which are obtained by changing the order of sine function, rectangular function and random number for input $u(k)$ in (16). The other simulation conditions are shown in Table 1.

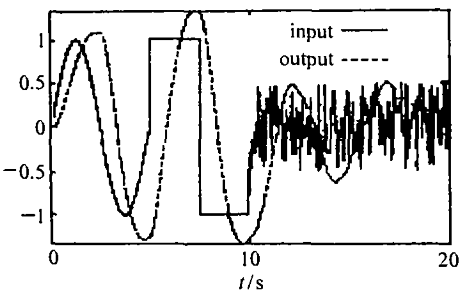


Fig. 3 Input and output data for training (case 0)

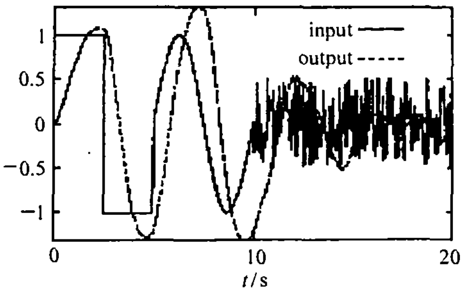


Fig. 4 Input and output data for investigating the generalization ability (case 1)

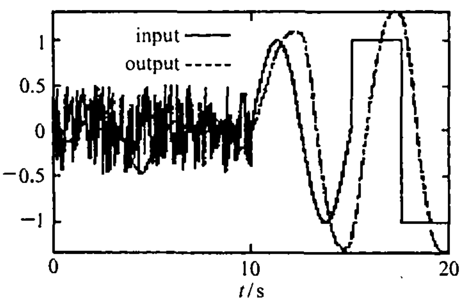


Fig. 5 Input and output data for investigating the generalization ability (case 2)

Table 1 Simulation conditions

A	$A = 1.5$
Initial value of parameter	
$h_{ij}(p)$	Random numbers in $(-1.0, 1.0)$
$D_{ij}(p)$	Positive discrete value in $[1, 10]$
$\beta_{ij}(p)$	0.3(fully connected)
Learning coefficient of $h_{ij}(p), \beta_{ij}(p)$	$\lambda_h = 0.00002, \lambda_\beta = 0.0002$
Identification error E	Root square error of the teaching signal
φ in switching function	Increase from 20 to 5000

Table 2 is the identification results for the training data. For example, cases (a) and (b) mean that only the parameter training of 500 000 times were carried out with the time delays being fixed to 1 sampling time and initial random sampling time respectively. Case (c) means that the parameter training of 10000 times and the time delay searching of 50 times. From Table 2, the average identification errors for the training data were calculated by averaging the 7 cases out of 3 initial parameters $\times$ 3 initial time delays (the best and the worst errors were omitted). So as apparently indicated by the results, adjusting not only parameters but also time delays is effective to improve the identification errors. And, it is also seen that the identification errors and parsimony of the network are appropriately determined corresponding to the R_α.

Table 2 Identification results (initial branches = 75)

case		(a)	(b)	(c)	(d)	(e)
Number of learning of parameters		500000	500000	10000	1000	100
Number of time delay search		0	0	50	500	5000
Residual branches	$R_\alpha = 0.0$	25	25	25	25	25
	$R_\alpha = 0.1$	13	8	9	8	9
	$R_\alpha = 0.5$	7	7	7	7	5
Average error ($\times 10^{-3}$)	$R_\alpha = 0.0$	4.42	2.78	2.97	3.21	3.09
	$R_\alpha = 0.1$	6.33	5.25	5.18	3.68	3.78
	$R_\alpha = 0.5$	10.4	9.0	7.85	7.65	5.5

In Fig.6 ~ Fig.8, the average identification errors for case 0,1,2 are obtained by averaging the same cases as Table 2. We can see that the results of case (a) and (b) without optimizing time delays are worse than that of case (d) and (e) with enough searching of optimal time delays for various R_α. Teaching signal and simulated signal obtained by the ULN after training are also shown in Fig.9 ~ Fig.11, in which case 0 denotes the simulated and teaching signals for the training data, while cases 1 and 2 show the simulated and teaching signals for the generalization investigating data. From the simulation results, the following can be concluded:

1) A new training algorithm that can adjust the time delays as well as the parameters of the ULN is presented. The simulation results indicate that the proposed algorithm is effective. Especially, the identification errors of a nonlinear system by the network that has less searching for time delays become worse compared with the identification errors by the network whose time delays are sufficiently adjusted.

2) From the results of the simulations for investigating the generalization ability, it has become quantitatively clear that ① the larger the size of the network is (in other words, the larger the number of the branches is), the smaller the identification errors are for the training data; ② the errors for the generalization investigating data are

bigger than the errors for the training data; ③ the smaller the size of the network is, the smaller the identification errors are for the generalization investigating data; ④ in case that the size of the network is too small, the identification errors increase (See Fig.6 ~ Fig.8).

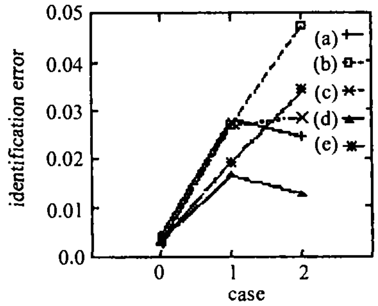


Fig. 6 Average identification error (R_a= 0.0)

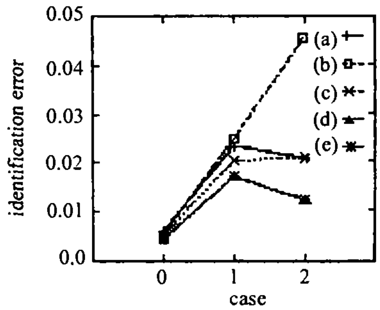


Fig. 7 Average identification error (R_a= 0.1)

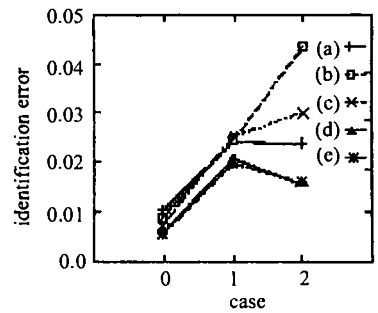
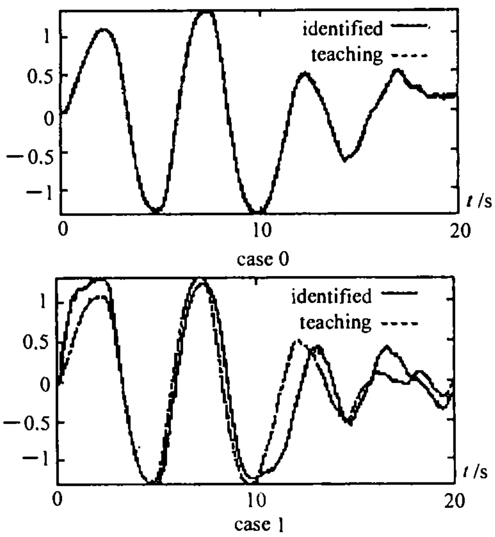


Fig. 8 Average identification error (R_a= 0.5)



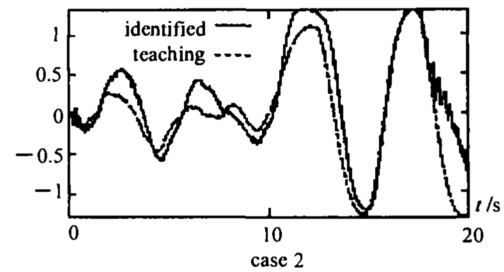


Fig. 9 Result of identified signal and teaching signal (R_a=0.0, the number of search is 5000.)

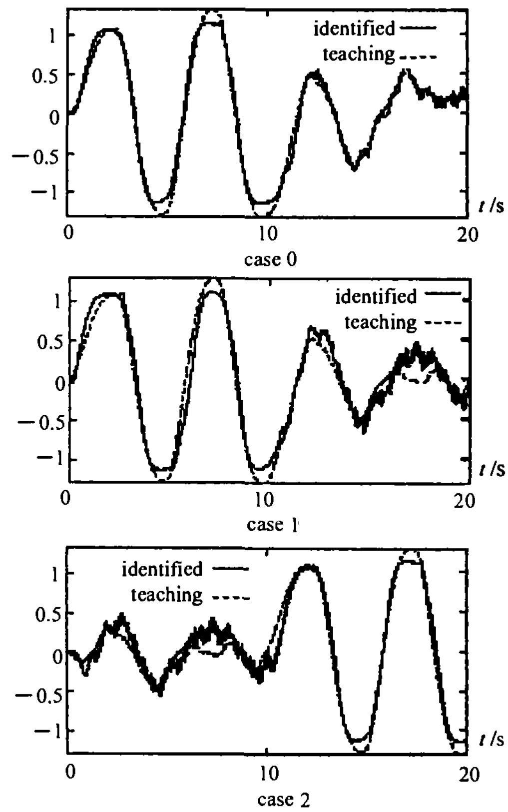
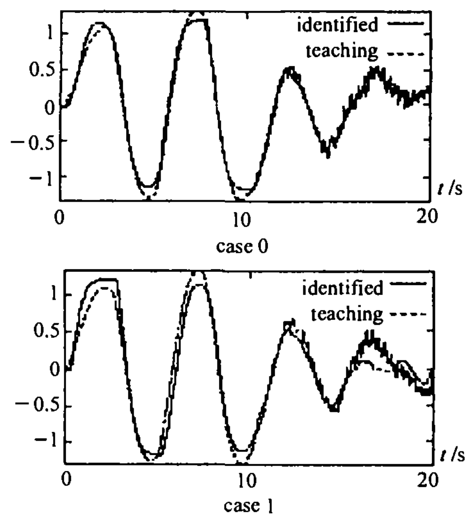


Fig. 10 Result of identified signal and teaching signal (R_a=0.1, the number of search is 5000.)



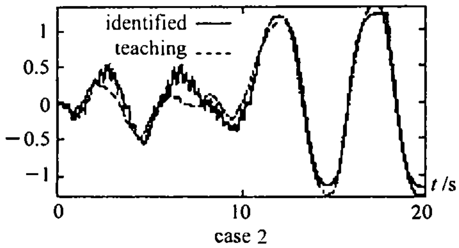


Fig. 11 Result of identified signal and teaching signal (R_α=0.5, the number of search is 5000.)

5 Conclusions

A new method for optimal modeling of the dynamic systems using the ULN is presented. The word 'optimal' means that both the modeling error and the parsimony of the model are taken into consideration, and the time delays are optimized for modeling the dynamic systems.

The important features of the proposed method are the introduction of the switching mechanism and the searching the optimal time delays.

The identification errors of a nonlinear system by the network that has less searching for the time delays become worse compared with the identification errors by the network whose time delays are sufficiently adjusted.

Especially, in this paper, the generalization ability of the proposed method has been studied using the simulations of nonlinear system identifications. And quantitative relations between the size of the network and the modeling errors and the effect of time delays search have been revealed.

References:

[1] GILES G L. Dynamic recurrent neural networks, theory and applications [J]. *IEEE Trans on Neural Networks*, 1994, 5(2): 153 - 155.

[2] HIRASAWA K, OHBAYASHI M, MURATA J. Universal learning network and computation of its higher order derivatives [A]. *Proc of 1995 IEEE Int Conf on Neural Networks* [C]. Perth, WA: [s.n.], 1995: 1273 - 1277.

[3] HIRASAWA K, OHBAYASHI M, KOGA M, et al. Forward propagation universal learning network [A]. *Proc of 1996 IEEE Int Conf on Neural Networks* [C]. Washington, DC: Microfiche, 1996: 353 - 358.

[4] HAN M, HIRASAWA K, FUJITA H, et al. Modeling nonlinear dynamic systems using universal learning network with filtering mechanism [J]. *Trans IEE of Japan*, 1997, 117(9): 1259 - 1266.

[5] HAN M, HIRASAWA K, OHBAYASHI M, et al. Modeling dynamic systems using universal learning network [A]. *Proc of IEEE Int Conf on Systems, Man, and Cybernetics* [C]. Beijing: Tsinghua University, 1996: 1172 - 1177.

[6] HAN M, HIRASAWA K, OHBAYASHI M, et al. Generalization ability of modeling dynamic systems using universal learning network [A]. 11*th IFAC Symposium Identification* [C]. Fukuoka: Nichigen Co, 1997: 1243 - 1248.

[7] HIRASAWA K, OHBAYASHI M, FUJITA H, et al. Universal learning network theory [J]. *Trans IEE of Japan*, 1996, 116(7): 794 - 801.

[8] WERBOS P. Back propagation through time what it does and how to do it [J]. *Proc of IEEE*, 1990, 78(10): 1550 - 1560.

[9] LIN D T, DAYHOFF J E, LIGOMENIDES P A. Trajectory production with the adaptive time-delay neural network [J]. *Neural Networks*, 1995, 8(3): 447 - 461.

作者简介:

韩 敏 (1959 —),女,工学博士研究生,大连理工大学电子与信息工程学院副教授,1982 年毕业于大连理工大学电子系,1992 年至 1999 年留学日本,获日本国立九州大学工学博士学位,目前承担国家自然科学基金重点项目中 3S 信息系统部分,参加国家重点基础研究发展计划(973)项目研究专题及教育部科学技术重点项目等 5 个课题,研究方向为:神经网络,专家系统,3S 系统及混沌序列分析,在国际国内会议及杂志上,共发表论文 45 篇,其中被 EI, ISTP 检索论文 11 篇次,E-mail: minhan@dlut.edu.cn;

席剑辉 (1975 —),女,1996 年获北京航空航天大学自动控制系学士学位,2002 年 3 月在大连理工大学电信学院获硕士学位并继续攻读博士学位,研究方向为神经网络系统辨识与混沌序列分析;

平泽宏太郎 (1942 —),男,1966 年获日本国立九州大学电子工程硕士学位,1966 年至 1992 年就职于日本日立公司,1992 年 12 月以来,任日本国立九州大学教授,现就职于该校信息科学和电子工程研究院,日本电子工程学会会员和 IEEE 会员.