Proceedings of the 11th
KACC, October 1996

Search for Optimal Time Delays in Universal Learning Network

° Min HAN, Kotaro HIRASAWA, Masanao OHBAYASHI and Hirofumi FUJITA
Department of Electrical and Electronic Systems Engineering, Kyushu University
6-10-1 Hakozaki, Higashi-ku, Fukuoka 812, Japan
Tel : +81-92-642-3907, Fax : +81-92-642-3962
E-mail : han @scel.ees.kyushu-u.ac.jp

Abstract - Universal Learning Network(U.L.N.), which can model and control the large scale complicated systems naturally, consists of nonlinearly operated nodes and multi-branches that may have arbitrary time delays including zero or minus ones. Therefore, U.L.N. can be applied to many kinds of systems which are difficult to be expressed by ordinary first order difference equations with one sampling time delay. It has been already reported that learning algorithm of parameter varibles in U.L.N. by forward and backward propagation is useful for modeling, managing and controlling of the large scale complicated systems such as industrial plants, economic, social and life phenomena. But, in the previous learning algorithm of U.L.N., time delays between the nodes were fixed, in other words, criterion function of U.L.N. was improved by adjusting only parameter variables. In this paper, a new learning algorithm is proposed, where not only parameter variables but also time delays between the nodes can be adjusted. Because time delays are integral numbers, adjustment of time delays can be carried out by a kind of random search procedure which executes intensified and diversified search in a single framework.

keywords : Universal Learning Network, Random Search, Learning, Time Delays

1. INTRODUCTION

Neural networks have been widely studied in recent years. By learning algorithms, neural networks can simulate a certain complicated systems. But, general current neural networks are composed of fixed nodes and branches and can not be equipped with arbitrary time delays. So it is difficult to apply these kinds of neural networks to modeling of the large scale complicated systems. In order to solve this problem, Universal Learning Network (U.L.N.) has been proposed[1][2]. Learning algorithm of Universal Learning Network by forward and backward propagation is useful for modeling and controlling large scale complicated systems such as industrial plants, economic, social and life phenomena.

The basic idea of U.L.N. is that most of the large scale complicated systems can be modeled by the networks which consist of nonlinearly operated nodes and multi-branches that may have arbitrary time delays including zero or minus ones. Therefore, U.L.N. can be applied to many kinds of systems which are difficult to be expressed as ordinary first order difference equations with one sampling time delay.

From an application oriented point of view, an excessive dimensionality of the network implies lengthened processing and learning times, even though the processing and learning performance is satisfactory.

In this paper, a new learning algorithm is proposed, where not only parameter variables but also time delays between the nodes can be adjusted considering both the performance and dimensionality of the U.L.N.[3]. Because time delays are integral numbers, adjustment of time delays can be carried out by a kind of random search which executes intensified and diversified search in a single framework.

Nowadays, adaptive time delay neural network which can also adjust time delays by the gradient method has been proposed[5]. The difference between our method and adaptive time delay neural network is that the searching of optimal time delay in our method is based on a random search different from the gradient method. And our proposed method can be applied to not only feed forward networks but also recurrent networks, while adaptive time delay neural network can only be used for feed forward networks.

2. BASIC STRUCTURE OF MULIT-BRANCH UNIVERSAL LEARNING NETWORK

The structure of Universal Learning Network with multi-branches and filtering structures[3] is shown in Figure 1.

In order to make the network compact, every branch has filtering structures, namely, switching functions such as $\alpha_{ij}(p)$ on pth branch from i node to j node. The learning parameter variables in the switching function should be learned so that $\alpha_{ij}(p)$ becomes 0.0 if branch from i node to j node is unnecessary and $\alpha_{ij}(p)$ becomes 1.0 if branch from i node to j node is necessary for the network to have predetermined performance.

Basic equation of U.L.N. is represented by Equation (1)

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

where

$h_j(t)$	:	output value of j node at time t,
$\lambda_m(t)$	:	value of mth parameter variable at time t,
$r_n(t)$	:	value of nth external input variable at time t,
O_j	:	nonlinear function of j node,
$D_{ij}(p)$	:	time delay of pth branch from i node to j node,
$JF(j)$	:	set of node numbers whose outputs are connected to j node,
$JB(j)$	:	set of node numbers whose inputs are connected from j node,
$B(i,j)$	:	set of branches from i node to j node,
$N(j)$	:	set of external input variables that are connected to j node,
N	:	set of external input variables,
$M(j)$	:	set of parameter variable numbers that are included in j node,
M		set of parameter variable numbers,
J	:	set of node numbers,
T	:	set of sampling times.

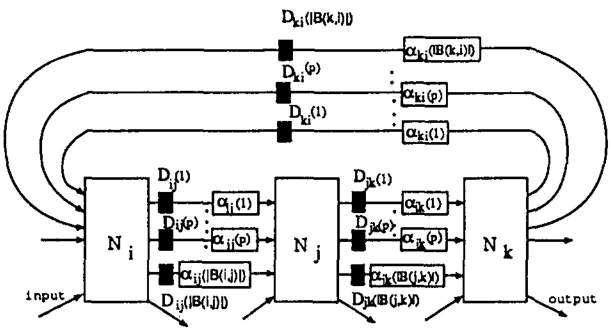


Fig.1 Structure of U.L.N. with multi-branches and switching functions

The switching function $\alpha_{ij}(p)$ is supposed to be

$$\alpha_{ij}(p) = \frac{1}{1+e^{-\varphi\beta_{ij}(p)}} \qquad (2)$$

The extended criterion function including both modeling error E and compactness of the model is given as follows:

$$\mathrm{L} \;=\; E + R_\alpha \sum_i \sum_j \sum_p (\alpha_{ij}(p))^2 \qquad (3)$$

$$E = E(\{h_r(s)\}, \{\lambda_m(s)\})$$

where

E	:	usual criterion function representing general modeling error,
R_α	:	weight coefficient,
$\alpha_{ij}(p)$	:	switching function of pth branch from i node to j node.

Depending on the value of weight coefficient R_α, the balance between the criterion function of E and the compactness of the network may be adjusted.

3. LEARNING OF PARAMETER VARIABLES λ_m and $\beta_{ij}(p)$

Learning of U.L.N.[1] is to adjust λ_m and $\beta_{ij}(p)$ by back propagating $\frac{\partial L}{\partial h_j(t)}$, in the same way as that commonly used in neural networks. The different point of U.L.N. from commonly used neural networks is that U.L.N. can have arbitray time delays between the nodes and has multi-branches between the nodes. From the reference[1], λ_m and $\beta_{ij}(p)$ can be adjusted as follows.

$$\lambda_m \leftarrow \lambda_m - \gamma\frac{\partial^\dagger L}{\partial \lambda_m} \qquad (4)$$

$$\beta_{ij}(p) \leftarrow \beta_{ij}(p) - \gamma\frac{\partial^\dagger L}{\partial \beta_{ij}(p)} \qquad (5)$$

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

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

$$\delta(j,t) = \sum_{k\in JB(j)}\sum_{p\in B(j,k)}\left[\frac{\partial h_k(t+D_{jk}(p))}{\partial h_j(t)} \times\delta(k,t+D_{jk}(p))\right] + \frac{\partial L}{\partial h_j(t)} \qquad (8)$$

$$j\in J,\quad t\in T$$

$\frac{\partial^\dagger L}{\partial \lambda_m}$, $\frac{\partial^\dagger L}{\partial \beta_{ij}(p)}$, in Equation(4) and Equation(5) are ordered derivatire proposed by Werbos[4].

where

$JD(\lambda_m)$	:	set of node numbers including λ_m,
$JB(j)$	:	set of node numbers whose inputs are connected from j node,
$B(j,k)$	:	set of branches from j node to k node.

4.RANDOM SEARCH OF TIME DELAYS

In the previous learning algorithm of U.L.N., time delays between the nodes were fixed, that is, criterion function of U.L.N. was improved by adjusting only parameter variables. In order to minimize extended criterion function a new learning algorithm is proposed to adjust time delays while parameters are also learnt. The basis idea is to search for optimal time delays by a kind of random search procedure, which has intensification and diversification capability.

The feature of the proposed random search is to define lth neighborhood $N^l(x)$ satisfying Equation(9) and to search for the optimal time delays using $N^l(x)$ in such a way as when there is quite a possibility of finding good solutions around the current one, intensified search for the vicinity of the current solution is carried out(l is small), on the other hand, when there is no possibility of finding good solutions, diversified search is executed in order to find good solutions in the region far from the current solutions(l is large).

$$N^l(x) \subset N^{l+1}(x) \qquad (9)$$

5. SIMULATION RESULTS OF NONLINEAR SYSTEM IDENTIFICATION

In simulations, a nonlinear system was modeled by the Universal Learning Network shown in Figure 1, which has 5 nodes, fully recurrent connections with 1 branch between the nodes, one external input and one output.

Input value $u(k)$ and output value $y(k)$ of the system to be identified in simulations are shown in Equation(10) and Figure 2.

$$y(k+1) = \begin{cases} 1.34y(k) - 0.277y(k-2) - 0.80y(k-4) \\ +0.01, \qquad u(k) \geq 0 \\ 1.34y(k) - 0.277y(k-2) - 0.80y(k-4) \\ -0.01, \qquad u(k) < 0 \end{cases} \qquad (10)$$

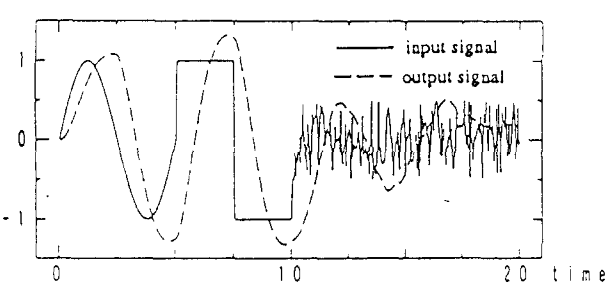


Fig.2 Input and output values of the system to be identified

Input value $u(k)$ to the nonlinear system was assumed to be

$$u(k) = \begin{cases} \sin(\frac{\pi}{50}k), & 0 \leq k < 100 \\ 1.0, & 100 \leq k < 150 \\ -1.0, & 150 \leq k < 200 \\ \text{uniform random,} & 200 \leq k < 400 \\ \text{numbers in}(-0.5, 0.5), & \end{cases}$$

Simulation condition is shown in Table 1..

Table 1. Simulation condition

number of nodes	J=5
number of branches between nodes	1
nonlinear function	$f(x) = A\frac{1-e^{-x}}{1+e^{-x}}$ A=1.5
initial value of parameter λ_m $\beta_{ij}(p)$	random numbers in (-1.0,1.0) 0.3
learning coefficient of λ_m $\beta_{ij}(p)$	γ=0.00002 γ=0.0002
criterion function identification error E coefficient R_α	root square error 0.1, 0.5
number of learning	500000
number of time delay search	50, 500, 5000
φ in switching function	increase from 20 to 5000

Simulation cases are as follows

· only learning without search of time delay (all time delay are one sampling time,case1,6)

· only learning without search of time delay(time delay was assumed to be random numbers in [1,10], case 2,7)

· learning and search of time delay combined (total learning is 500000 times, case 3,4,5,8,9,10)

Table 2,3, Figure 3,4,5,6, and Figure 7 show identification results including learning and searching curves, identified and teaching signals, identification errors, numbers of residual branches,$\beta_{ij}(p)$, and $\alpha_{ij}(p)$.

From theses results, it has been shown that adjusting of not only parameter variables but also time delays is useful for identification of a nonlinear dynamic system by using Universal Learning Network with extented criterion function.

Table 2. Identification results
(weight coefficient R_α=0.1)

case	1	2	3	4	5
learning number after search of delay	500000*	500000	10000	1000	100
search number	0	0	50	500	5000
residual branches	13	8	9	8	9
error $\times$ $(10)^{-3}$	5.50	3.43	3.15	3.52	3.70

Table 3. Identification results
(weight coefficient R_α=0.5)

case	6	7	8	9	10
learning number after search of delay	500000*	500000	10000	1000	100
search number	0	0	50	500	5000
residual branches	7	7	7	7	5
error $\times$ $(10)^{-3}$	12.55	23.15	9.78	7.23	6.20

*: time delays of all branches are assumed to be 1 sampling time.

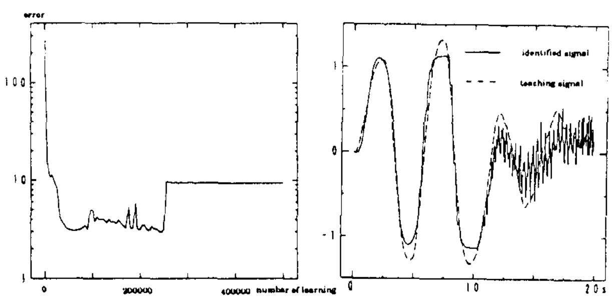


Fig.3.Learning and searching curves, identified and teaching signal (case 7)

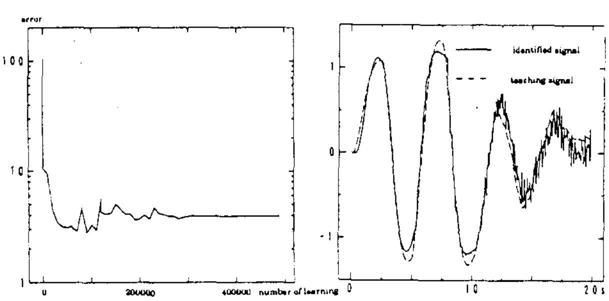


Fig.4.Learning and searching curves, identified and teaching signal (case 8)

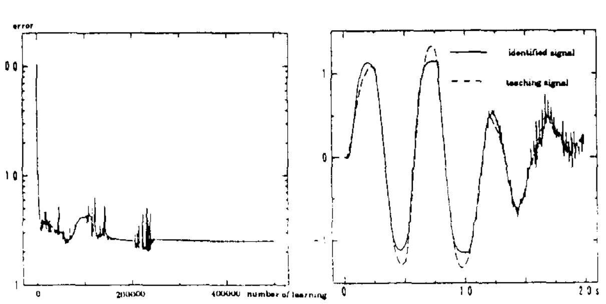


Fig.5.Learning and searching curves, identified and teaching signal (case 10)

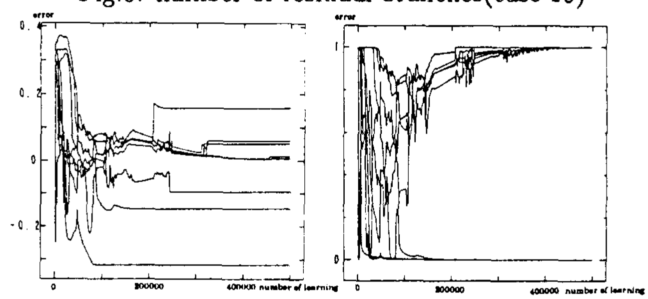
i\j	1	2	3	4	5
1	1	7	4	4	8
2	9	10	9	7	10
3	8	10	4	5	4
4	3	6	5	1	1
5	1	1	6	10	8

initial

i\j	1	2	3	4	5
1				1	7
2					
3					
4	1				
5	9				5

final value

Fig.6. number of residual branches(case 10)



$\beta_{ij}(p)$ $\alpha_{ij}(p)$

Fig.7. curves of $\beta_{ij}(p)$,$\alpha_{ij}(p)$(case 10)

6.CONCLUSION

The Universal Learning Network is proposed for modeling and controlling large scale complicated systems. One of the important things in U.L.N. is that U.L.N. can optimize the structure of large scale systems considering both modeling error and compactness of the network structure. In this paper a new learning algorithm which adjusts parameter variables as well as time delays at the same time is presented. The simulation results indicate that the proposed algorithm is effective. Especially, identification error of a nonlinear system by the network which has less searching for time delays become worse compared with the identification error by the network whose time delays are sufficiently adjusted.

REFERENCES

(1)K. Hirasawa, M. Ohbayashi,J. Murata :**Universal Learning Network and Computation of its Higher Order Derivatives**, Proc. of 1995 IEEE International Conference on Neural Networks,1995

(2)K. Hirasawa,M. Ohbayashi,M. Koga and M. Harada :**Forward Propagation Universal Learning Network**, Proc. of 1996 IEEE International Conference on Neural Networks, 1996

(3)M.Han K.Hirasawa,M.Ohbayashi,H.Fujita : **Modeling Dynamic Systems using Universal Learning Network**, Proc. of 1996 IEEE International Conference on Systems, Man and Cybernetics, 1996

(4)P.Werbos :**Beyond regression: New Tools for Prediction and Analysis in the Behavior Science**, Ph.D. dissertation, Harvard University, 1974

(5) Daw-Tong Lin, Judith E. Dayhoff, Panos A. Ligomenides:**Trajectory Production with the Adaptive Time-Dalay Neural Network**, Neural Networks, Vol.8, No3, pp447-461,1995