Growing neural network for acquisition of 2-layer structure

Ryusuke KURINO Masanori SUGISAKA and Katsunari SHIBATA
Dept. of Electronic Engineering, Oita University.
700 Dannoharu Oita 870-1192 Japan. Email: shibata@cc.oita-u.ac.jp

Abstract

Neural network are broadly used to approximate non-linear functions. However, it is difficult to decide an appropriate structure for a given problem. In this paper, “growing neural network” is proposed as an extension of Back Propagation (BP) learning. The propagated error signal is diffused from a target neuron as a substance. The axon of a growing neuron grows according to the concentration gradient of the substance. In a simulation, it was examined that the simplest problems, “AND” and “OR”, could be solved by the neural network and 2-layer structure was properly obtained.

1. Introduction

Artificial neural networks suggested by the natural nerve system of living things are broadly used to approximate non-linear functions because of its advantage of the leaning and generalization ability.

It is said that there are billions of neurons in the brain of human, and they are mutually connecting complicatedly. As opposed to it, when the artificial neural network [NN] is used, a simple 3-layer structure is employed in most cases. That is because the way to decide an appropriate structure for a given problem has not been established yet. Furthermore, it is well known that 3-layer neural network with enough number of hidden neurons can approximate any continuous functions with any precision [1], [2].

However, in the field of intelligent robot hereafter, high order functions will be required. In order to realize such functions, a various levels of abstract state representation and memory must be required. In other words, not only the approximation of the function from sensors to motors, but also utilization of the abstract representation in hidden layers obtained through learning is essential to realize intelligence. However, each hidden neuron in the 3-layer NN represents just a linear and smooth classification of the input space. It is clear that the representation is not enough as the abstract state representation. For the memory, feedback connections are also necessary. These are supported by the complicated structure of neural network in our brain. Accordingly autonomous acquisition of an appropriate structure of the neural network including recurrent structure must be required increasingly for the brain of intelligent robots.

In several methods to decide the structure of the neural network at present, since rough structure is assumed, the role of the designer is large and the degree of freedom to decide the structure is small. Further, it is difficult to decide the structure purposively according to the given problems on-line (see section 2).

It has been considered that in the brain of living things, the number of neurons decreases after its birth, while the number of connections between neurons rapidly increases[3]. This suggests a strategy in the brain as follows. More neurons than necessity are prepared at first, then they grow their axons and learn their synapse weights, and an appropriate network structure is formed flexibly. After that, unnecessary neurons are removed by apoptosis. It has been also reported that chemical substances such as NGF (Nerve Growth Factor) make a role of promoting the growth of axons and maintaining the connections [3](see section 3).

In this paper, “growing neural network” is proposed in which the concept of “growth” is introduced in the conventional artificial neural networks. Then, the growth is formulized as an extension of Back Propagation (BP) learning that is a popular supervised learning. Getting a hint from NGF mentioned above, axons grows according to the concentration gradient of chemical substance which is diffused according to the error signal propagated in BP leaning. Therefore, the degree of freedom to decide the structure is larger than the conventional methods. Furthermore, since the growth is performed as an extension of learning, it is expected that purposive structure can be obtained. The final goal of this research is to build growing neural network that is able to obtain various structures including recurrent structure. In this paper, as the first step of this research, considering the acquisition of two-layer structure, the fundamental algorithm is introduced at first, and then the simulation result when the algorithm is applied to the learning of the simplest logical functions, “AND” and ”OR”, is reported.

2. Conventional methods to decide the structure

There are well-know several methods to decide the structure of the neural network at present: (a) the method using information criterion [4], (b) the method to cut out unnecessary connections by Forgetting [5], (c) the method to append hidden neurons successively [6], and (d) the method using genetic algorithm (GA)[7]. Here, these methods are mentioned, a growing neural network and the conventional methods are compared, and the advantage of the growing neural network is explained.

(a) The method using information criterion

- Information criterion is used to compare some structures, but it cannot find an appropriate structure.
- The main purpose is to avoid over-fitting.
- The structure is evaluated after learning.

(b) The method to cut out unnecessary connections by Forgetting.

- The main purpose is to avoid over-fitting.
- The structure can be obtained by learning.
- All the connections must be made at first.

(c) The method to append hidden neurons successively

- The structure can be obtained together with learning.
- Since a neuron is appended, all the connections of the neuron are appended simultaneously.
- The computational cost is low.

(d) The method using genetic algorithm (GA)

- The degree of freedom to decide the structure is comparatively large.
- The structure is evaluated after learning.

As opposed to it, in the growing NN, since the growth of neuron progresses according to the error signal, it is thought that the structure can be obtained together with learning, and the obtained structure is expected to be purposive. Thus, it is considered that the degree of freedom to decide the structure is larger than the conventional methods. Furthermore, the position of neurons influences the structure. By this property, it is expected that the symmetry between neurons are broken, and strcturization of network is promoted.

3.Growth of neuron in the living things

In 1950, a NGF (Nerve Growth Factor) was discovered by Levi-Montalcini et al [8], and it is known that the NGF deeply relate to the growth of neuron. After that, except for the NGF, a chemical substance that works as well as the NGF is discovered, and they are called “Nerve Nutrition Factors”. When the target neuron requires some connections, the NGF is made in the target neuron, and is diffused. Then, the growing neuron extends its axon according to the concentration gradient. After making the connection, NGF moves along the axon backward to the pre-synaptic neuron and the neuron and the axon are maintained et al [9]. When the axon is cut, NGF cannot reach the pre-synaptic neuron, and then the axon degenerates.

Furthermore, in the experiment on a plate, a neuron cannot extend its axon within the range with no NGF, and if no part of the neuron and its axon exist in the range with NGF, the neuron is degenerated [10].

4. Growing neural network (NN)

4.1 Fundamental algorithm

In the growing NN, the growth is formulated as an extension of BP learning as mentioned above. Figure.1 shows the flowchart of the algorithm.Figure.2 shows the basic idea of the growing NN. Figure.3 shows the distribution of the concentration gradient around a target neuron. At first, the output and the error are calculated, and the error signal is propagated backward. If the neuron does not have enough connections, the neuron that receives the signal diffuses it as a chemical substance. Then, the concentration gradient is formed by the diffusion around the neuron. The growing neuron extends its axon according to the concentration gradient. After making the connection (synapse), the learning is started from 0 connection weight according to the regular BP algorithm.

1. Decision of each neuron’s position
2. Decision of whether each connection is made or not.
3. Setting of input and training signals
4. Calculation of output and error
5. Back propagation of the error signal
6. Diffusion of the error signal
7. Calculation of the axon growth
8. Judge of whether each connection has been made or not
9. Update of connection weights
   (Only for connected neurons)
10. Return to step 4

Fig.1 Flow chart of the growing neural network

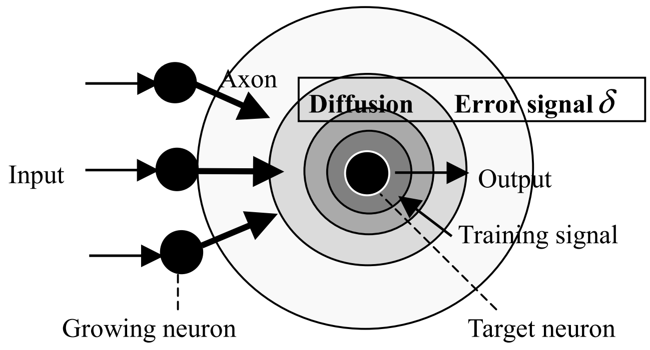

Fig.2 Growing neural network

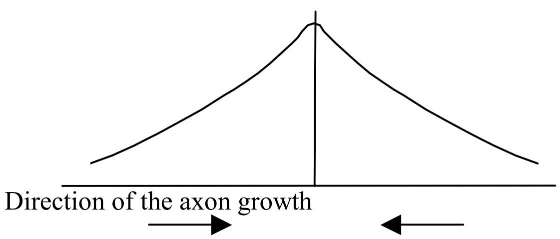

Fig.3 Concentration gradient

4.2 Diffusion of the error signal

Neurons of living things diffuse chemical substance like NGF, and grow its axon according to the concentration gradient that is formed by the diffused substance. To realize such functions in the growing NN, the error signal is diffused as the chemical substance around the target neuron. The error can be positive or negative, but the diffusion of negative is hard to imagine. For this reason, here, it is assumed that there exists the substance for each of negative and positive error, and they diffuse their substance independently. In this paper, since the simplest two-layer structure is assumed, the target neuron diffuses the error signal. The error signal is calculated as

$$\delta_j = -\frac{\partial E}{\partial net_j} = (d_j - o_j) f'(net_j) \qquad (1)$$

where net_j : the internal state of the neuron j , o_j : the output, d_j : the training signal, $f'(net_j)$: the derivative of the output function, j =0,.., $NODE$, $NODE$: the number of target neurons, E : the error function. The diffusion is calculated for each of the positive and negative error signal respectively as

$$\frac{\partial u^p_{x,y}}{\partial t} = \rho \delta_i + D \nabla^2 u^p \qquad (2)$$

$$\frac{\partial u^n_{x,y}}{\partial t} = -\rho \delta_i + D \nabla^2 u^n \qquad (3)$$

where $u^{n,p}_{x,y}$: a concentration, D : a diffusion constant, p indicates positive, n indicates negative, ρ : a divergence constant. The quantity of the diffused substance is proportional to the error signal. At the place where no neuron exists, *Eq* (2) and (3) is calculated with δ = 0.0. By dealing with the error signal as the diffused substance, the growth of neuron according to the necessity can be realized.

4.3 Extension of the axon

The axon extends according to the state of growing neuron and the concentration gradient at the tip of the axon. Two types of neurons are prepared. One of them makes a positive connection, while the other makes a negative connection. The former grows its axon to the direction of the gradient of the positive error signal, while the latter grows its axon to that of the negative error signal. The reason is as follows. (1) If there is only one type of neurons, the neuron makes only one of negative or positive connection. (2) In the neuron of living things, it is known that which of the positive or negative connection the neuron makes is decided by the neurotransmitter that is generated at the synapse. The extension of the axon is calculated as,

$$\frac{d\mathbf{a}_i}{dt} = \xi \cdot \nabla (u^p - u^n) \cdot S_i \cdot flag_i \qquad (4)$$

where ξ : a growing constant, i =0,…, $NODE$, $NODE$:the number of growing neurons, and the state S of the growing neuron is defined in the section 4.5. If the growing neuron is for the positive connection, $flag$ =1, while for the negative connection, $flag$ = -1.

4.4 Update of the connection weight

The connection weight is always 0 before the connection is formed. When the axon extends according to the concentration gradient and the connection is formed, the synapse starts the learning from 0 connection weight. The update of connection weight is calculated as well as the conventional BP leaning,

$$\frac{d\omega_{ji}}{dt} = -\eta \frac{\partial E}{\partial \omega_{ji}} = \eta \delta_j o_i \qquad (5)$$

4.5 Adjustment for diffusion delay

It takes some time that the diffused error signal reaches the growing neuron. If the input patterns change frequently, the input signal has changed when the error signal arrives at the growing neuron. Then, the first-order delay is introduced to the output of the growing neuron to adjust the gap of the timing. The state of neuron S is defined as the first-order delay of the output of the growing neuron as

$$\tau \frac{dS_i}{dt} = -S_i + o_i \qquad (6)$$

If a time constant τ is too large, the state of the neuron does not change so much. While, if the time constant is too small, the state of the neuron is not different from the original output value. Thus, the appropriate time constant is required.

4.6 Boundary condition

Since the diffusion is calculated in the limited area, some boundary condition is necessary. By Dirichlet condition or by Neumann condition, natural diffusion cannot be realized. In this paper, extrapolation using exponential curve is done as *Eq*.(7) at the boundary as shown in Fig.4, and it is used to calculate diffusion.

$$u_{n+1} = u_n^2 / u_{n-1} \qquad (7).$$

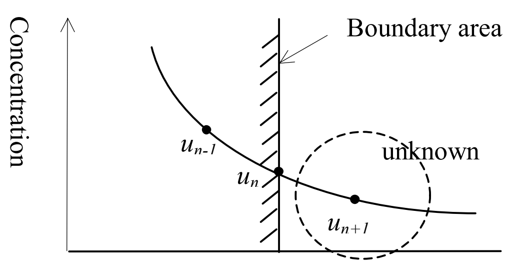

Fig.4 Extrapolation using exponential curve at the boundary

5. Simulation

5.1 Set up

A simple two-layer structure is assumed to verify the fundamental functions of the growing NN, and two simple logical function "AND","OR" were learned. The simulation is done in 0.5(mm)$\times$0.5(mm) of area. This area is divided into 50$\times$50 small regions, and the concentration is calculated for each region. Figure.5 shows the position of each neuron. The left output neuron learns to output "AND" of the input 'a' and 'c', and right one learned to output "OR" of the input 'b' and 'd'. The input 'b' and 'c' have to connect to the father output neuron. The parameters used in the simulation as follows. Leaning constant η =0.1(1/sec), growing constant ξ=0.2(mm/sec), diffusion constant D=0.2(mm^2/sec), divergence constant ρ =2.0(1/sec). Figure.6 shows the training pattern example according to time in the case of AND output. Each of input signals changes randomly from 0 to 1at every 1 (sec). 0.1 or 0.9 was used as the training signal on behalf of 0 or 1. The output function is sigmoid with the range from 0 to 1. The inertia term was not used.

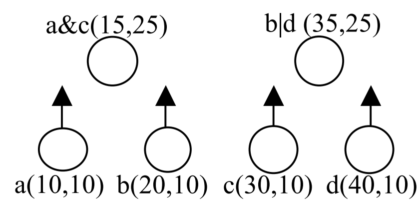

Fig.5 Position of the each neuron

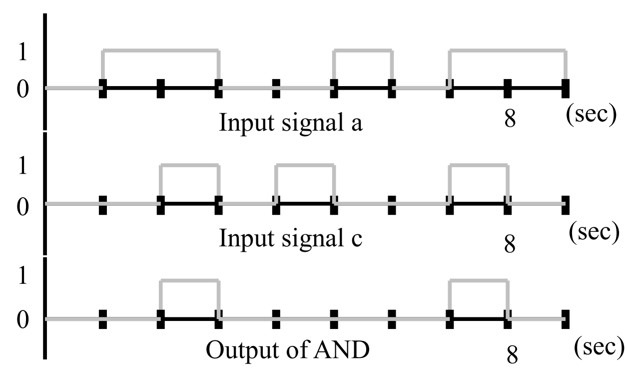

Fig.6 Training pattern for AND output

5.2 Adjustment for the diffusion delay

At first, the simulation to adjust the diffusion delay is done. The position of a target neuron is (15,25), the error of positive or negative generates by turns and the concentration gradient at (10,10) was observed. Figure.7 shows the correlation of two patterns between the concentration gradient and the state of the growing neuron. The simulation is done for two cases. In the first case, input interval is 1(sec), in the second case, it is 2(sec). The product of the concentration gradient and the state of the neuron is calculated, and then it is integrated along time axis. In this figure, when a time constant is 0.1, the correlation is the largest. Moreover, in the case of 0.1, Fig.8 and Fig.9 show the change of the state of the neuron and the concentration gradient. Input interval is 1.0 (sec) in

Fig.8, while it is 2.0 (sec) in Fig.9. In each of Fig.8 and Fig.9, the timing gap between the concentration gradient and the state of the neuron is small. Thus, 0.1 is employed as a time constant.

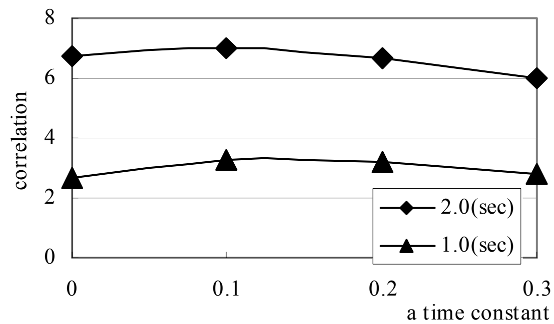

Fig.7 The correlations between the concentration gradient and the state of the neuron.

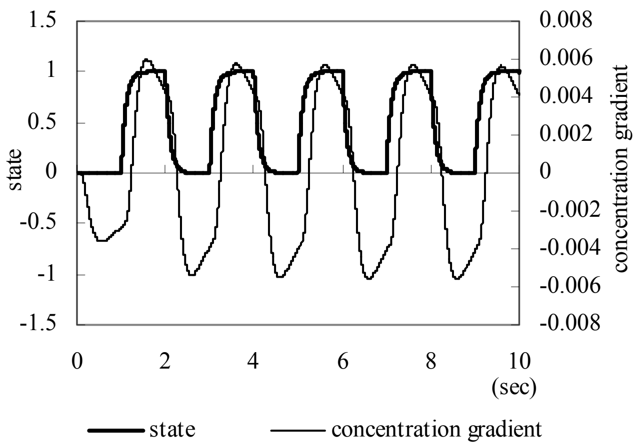

Fig. 8 The change of the concentration gradient and state of the neuron (the input interval is 1.0(sec))

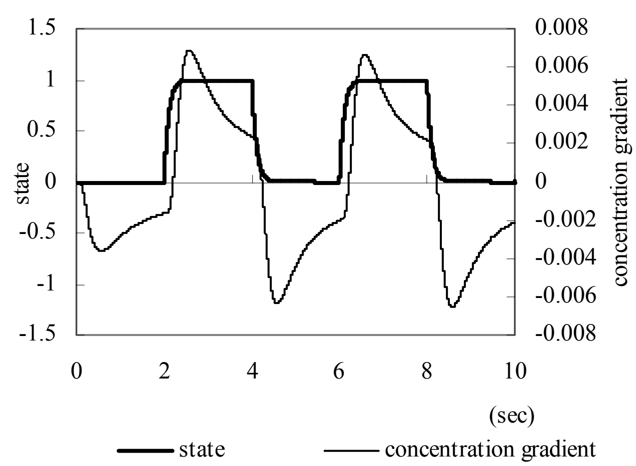

Fig. 9 The change of the concentration gradient and state of the neuron (the input interval is 2.0(sec))

5.3 Result

Figure.10 shows how the axon of each input neuron grows. Although, there are two types of neurons, for positive connection and for negative connection, only the neurons for positive are shown in the figure. The neuron grows with repetition of extension and degeneration and the input ‘a’ forms the connection at 115 (sec) and the input ‘d’ forms at 97(sec). Since the input ‘b’ and ‘c’ are located closer to the target neuron that should not be connected, the loci prowl by the influence of the closer target neuron. However, there is no correlation between the input and the output. Finally, the both loci crossed and arrived at the correct target neuron at 240(sec) and 310(sec) respectively.

Figure.11, 12 show the change of the error for each output. The arrows in these figures indicate the timing when each connection was made. After making the first connection, the error for some input patterns are gradually decreased according to the change of the connection weights. Though it is difficult to notice from this figures. After making the second connection, the error decreases quickly. Moreover, it is seen that a small error remains after convergence. That is because the output for AND is stuck at 0.0 when the input is (0,0), and the output for OR is stuck at 1.0 when the input is (1,1) even though the training signal is 0.1 and 0.9 respectively.

Figure.13, 14 show the change of the connection weights for the “AND” and ”OR” output respectively. Since the neuron ‘a’ and ‘d’ makes their connection, the learning of the connection weight for the neuron ‘a’, ‘d’ starts at first. The connection weight of the neuron ‘a’ fluctuates more than that of the neuron ‘d’. That is because when the input signal ‘a’ is 1, the training signal is 0.1or 0.9, but when the input signal ‘d’ is 1, the training signal is always 0.9.

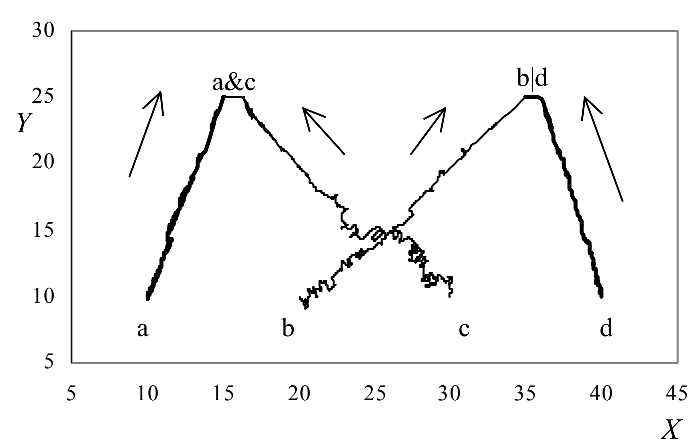

Fig.10 Growth of each growing neuron

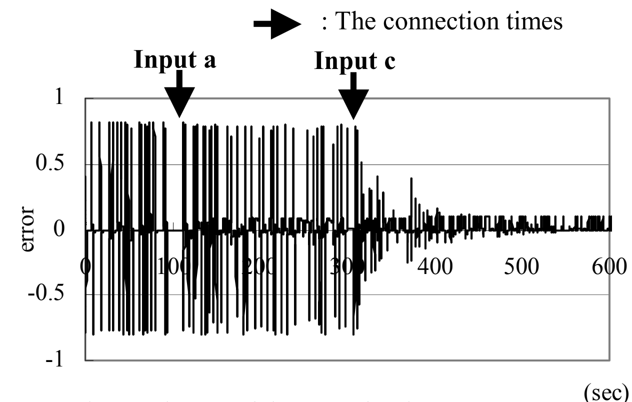

Fig.11 Change of the error for the AND output

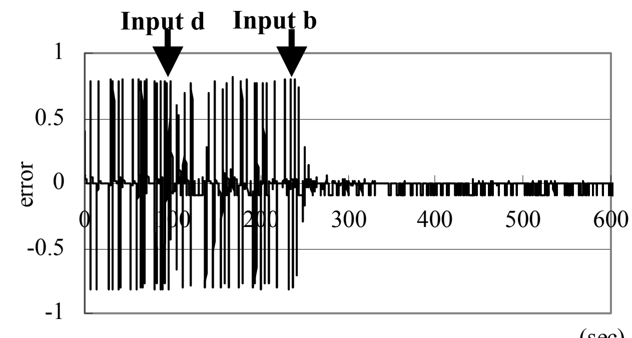

Fig.12 Change of the error for the OR output

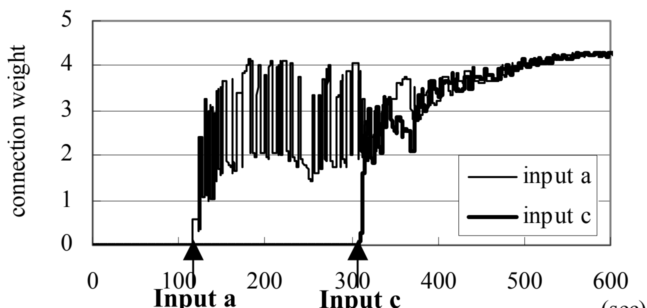

Fig.13 Change of the connection weights for the AND output

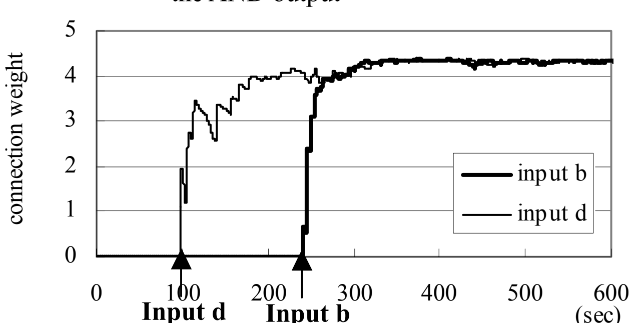

Fig.14 Change of the connection weights for the OR output

6. Conclusion

In this paper, the growing NN that is formulized as an extension of the conventional BP learning was proposed. In the simulation with two target neurons, it was confirmed that the growing neurons grow their axons to the appropriate target neuron even if the target neuron is farther than the other target neuron.

7. Future work

The growing NN is extended so as to solve the problems those needs hidden neurons, such as EXOR. There are two difficult points to be solves. One is that there are sometimes no correlation between each input and each output. The other is that the hidden neuron cannot receive either of input signals and error signals at first.

Acknowledgement

A part of this research was supported by the Sci. Res. Foundation of the Ministry of Edu., Sci., Sports and Culture of Japan (#13780295).

Bibliography

[1]Kurt HORNIK, Multiayer Feedforward Networks are Universal Approximators, Neural Networks, Vol.2, pp. 359-366(1989)

[2]KEN-ICHI FUNAHASHI, on the Approximate Realizaton of Continuous Mappings by Neural Networks, Neural Networks, Vol.2, pp. 183-192(1989).

[3]Tadaharu TUMOTO. Brain and Development, Asakura Publishing Co.,Ltd(1986). (in Japanese)

[4]D.B. Fogel, An information criterion for optimal neural network, IEEE,Vol.2,No.5, pp.490-497(1991)

[5]Masumi. Ishikawa, A Structural Connectionist Learning Algorithm with Forgetting, Artificial Intelligence, Vol.5, No.5, pp.595-603(1990)

[6]S. E. Fahlman and C. Lebiere, The cascade– correlation learning architecture, Advances in Neural information Processing System, Vol.2, pp.524-532 (1990)

[7]Naoki SHIBA, Mnabu KOTANI and Kenzo AKAZAWA, designing Multi-layered Neural Networks Using Genetic Algorithm, Trans. of SICE, Vol.34,No.8,pp.1080-1087(1998). (in Japanese)

[8]Levi-Montalcini, R. and Hamburger, V., Selective growth-stimulating effects of mouse sarcoma on the sensory and sympathetic nervous system of the chick embryo. J. Exp. Zool.,Vol.116, pp.316-362(1951)

[9]Brunso-Bechtold, J.K. and Hamburger, V., Retrograde transport of nerve growth factor in chicken embryo. Proc. Natl. Acad. Sci. USA., No.76,pp.1494-1496(1979)

[10]Campenot, R.B., Local control of neurite development by nerve growth factor, Proc. Natl. Acad. Sci. U.S.A., Vol.74, pp.4516-4519(1977).