Fifth IFSA World Congress (1993), 1017-1020

A New Effective Learning Algorithm for a Neo Fuzzy Neuron Model

Takeshi Yamakawa*, Hiroaki Kusanagi, Eiji Uchino* and Tsutomu Miki***

*Department of Control Engineering and Science
Kyushu Institute of Technology, Iizuka, Fukuoka 820, Japan
Phone : +81-948-29-7712, Fax : +81-948-29-7742
E-mail : yamakawa@ces.kyutech.ac.jp

**TADANO, LTD.
34 Kou, Shinden-chou, Takamatsu-city 761-01, Japan

Abstract

This paper describes a neo fuzzy neuron which was produced by a fusion of fuzzy logic and neuroscience. Some learning algorithms are presented. The guarantee for the global minimum on the error-weight space is proved by a reduction to absurdity. Enhanced is that the learning speed of the neo fuzzy neuron exceeds 100,000 times of that of conventional multi-layer neural networks.

1. INTRODUCTION

One of the authors presented a model of fuzzy neuron, in which linear synaptic connections are replaced with a nonlinearity characterized by a membership function labeled as "tightly connected", "loosely connected", etc., and excitatory connections and inhibitory connections are represented by fuzzy logic intersections and fuzzy logic complements followed by fuzzy logic intersections, respectively [1, 2]. Sequentially, a neo fuzzy neuron model was presented by the authors [3,4,5], the nonlinear characteristics of which is represented by fuzzy if-then rules with complementary membership functions. One multi-input one-output neo fuzzy neuron model, but not a neural network, can exhibit its good ability to describe a nonlinear relationship between inputs and output as well as its short learning time compared with a conventional neural network.

This paper proves the guarantee of a global minimum in error-weight space of a neo fuzzy neuron by employing the reduction to absurdity. Furthermore, five learning procedures are presented and compared with each other in order to evaluate them. The final error and the learning time of the neo fuzzy neuron are compared with those of a conventional three-layered neural network.

2. ARCHITECTURE OF A NEO FUZZY NEURON[3,4,5]

The structure of the neo fuzzy neuron is shown in Fig.1(a), where the characteristics of each synapse is represented by a nonlinear function f_i and the soma doesn't exhibit a sigmoidal function at all. Aggregation of synaptic signals is achieved by an algebraic sum. Thus the output of this neo fuzzy neuron can be represented by the following equation;

$$\hat{y} = f_1(x_1) + f_2(x_2) + \cdots + f_m(x_m) = \sum_{i=1}^{m} f_i(x_i) \tag{2.1}$$

The structure of the nonlinear synapse is shown in Fig.1(b). The input space x_i is divided into several *fuzzy segments* which are characterized by membership functions $\mu_{i1}, \mu_{i2},\ldots, \mu_{ij},\ldots, \mu_{in}$ within the range between x_{min} and x_{max} as shown in Fig.1(c). 1, 2, ..., j, ...,n are numbers assigned to labels of fuzzy segments. The membership functions are followed by variable weights $w_{i1}, w_{i2}, \ldots, w_{ij}, \ldots, w_{in}$.

Mapping from x_i to $f_i(x_i)$ is determined by fuzzy inferences and a defuzzification. The fuzzy inference adopted here is a singleton consequent, that is, each weight w_{ij} is a deterministic value such as 0.3. It should be emphasized that each membership function in antecedent is triangular and assigned to be *complementary* (so called by the authors) with neighboring ones. In other words, an input signal x_i activates only two membership functions simultaneously and the sum of grades of these two neighboring membership functions labelled by k and k+1 is always equal to 1, that is, $\mu_{ik}(x_i) + \mu_{i,k+1}(x_i) = 1$ So that the defuzzification taking a center of gravity doesn't need a division and the output of the neo fuzzy neuron can be represented by the following simple equation.

$$f_i(x_i) = \frac{\sum_{j=1}^{n} \mu_{ij}(x_i)\cdot w_{ij}}{\sum_{j=1}^{n} \mu_{ij}(x_i)} = \frac{\mu_{ik}(x_i)\cdot w_{ik} + \mu_{i,k+1}(x_i)\cdot w_{i,k+1}}{\mu_{ik}(x_i) + \mu_{i,k+1}(x_i)} = \mu_{ik}(x_i)\cdot w_{ik} + \mu_{i,k+1}(x_i)\cdot w_{i,k+1}\cdot \tag{2.2}$$

This equation can be realized by the architecture shown in Fig.1(b)

The weights w_{ij} are assigned by learning procedures described in the next chapter.

3. LEARNING ALGORITHM

3.1 Incremental Updating (Stepwise Training) : SE.

Let $\mathbf{X}_k=(x_{1k}, x_{2k}, \ldots, x_{ik}, \ldots, x_{mk})$ be the k-th input pattern applied to the m-input neo fuzzy neuron and y_k and te_k the output and its desired value, respectively, corresponding to the k-th input pattern, where k=1, 2, ..., p.

The primitive learning procedure was presented by the authors [3.4,5] as a modification of a steepest descent method [6].

$$\Delta w_{ij} = - \alpha(y_k - te_k)\mu_{ij}(x_{ik}) \tag{3.1}$$

where α represents a learning rate and is empirically obtained. In this learning algorithm, all the initial weights are assigned to be zero.

3.2 Batch Learning (Cumulative Weight Adjustment) : BA.[6]

The learning procedure of 3.1 is the incremental change of weights for each input pattern. The other learning procedure is to change the weights for a set of p input patterns. In this case, the square of the error is cumulative for a set of patterns

$$E = \frac{1}{2}\sum_{k=1}^{p} E_k = \frac{1}{2}\sum_{k=1}^{p} (y_k - te_k)^2 \tag{3.2}$$

The incremental change of weights for minimizing the squared error (Eq.(3.2)) is obtained from Eqs.(2.1) and (2.2) as

$$\Delta w_{ij} = - \alpha\frac{\partial E}{\partial w_{ij}} = - \alpha\sum_{k=1}^{p} (y_k - te_k)\mu_{ij}(x_{ik}) \tag{3.3}$$

In this learning algorithm, all the initial weights are assigned to be zero and the updating of the weights is achieved after calculation of cumulative value in Eq.(3.3).

3.3 Direct Search Method with Second Order Interpolation : D.R.M.

In 3.1 and 3.2, a learning rate is a constant obtained by experiments. In order to realize a learning of high-speed and least error, the learning rate should be optimally adjusted for each case

The optimal value of α to minimize the error E at a g. ,n temporal point t can be obtained by searching α by which the updated weight produces the minimal error in the next step.

In this paper, the searching process is based on interpolation of second order (Lagrange's interpolation) in order to obtain the optimal value of α at higher speed. The error function E is a function of α and unimodal as described in Chap.4. The procedure to obtain the optimum value α_{opt} which gives the approximately minimal error is presented in the following.

Assume the searching for α in the interval $[\alpha^0, \alpha^0+h]$ on the error-α space.

<STEP 1> If $E(\alpha^0)>E(\alpha^0+h)$, then the interval $[\alpha^0, \alpha^0+h]$ is shifted by the increment h to $[\alpha^0+h, \alpha^0+kh]$. If $E(\alpha^0+h)>E(\alpha^0+kh)$, then the interval $[\alpha^0+h, \alpha^0+kh]$ is shifted to $[\alpha^0+kh, \alpha^0+k^2h]$. This procedure is continued until $E(a^0+k^{s-1}h)<E(a^0+k^sh)$. If $E(\alpha^0)<E(\alpha^0+h)$, then the similar shifting of interval to the counter direction is achieved until $E(a^0-k^{s-1}h)>E(a^0-k^sh)$. In this paper k=0.2 and h=0.05.

<STEP 2> Let the last three points of α be α_1, α_2 and α_3 $(\alpha_1<\alpha_2<\alpha_3)$, and E be E_1, E_2 and E_3 corresponding to them, respectively. For example $\alpha_1 = \alpha^0 + k^{s-2}h$, $\alpha_2 = \alpha^0 + k^{s-1}h$ and $\alpha_3 = \alpha^0 + k^sh$ in case of $E(\alpha^0)>E(\alpha^0+h)$. From the interpolation of second order [7], the optimum value α_{opt} which gives approximately minimum value of E is obtained by the following

$$\alpha_{opt} = \frac{(\alpha_2^2 - \alpha_3^2)E_1 + (\alpha_3^2 - \alpha_1^2)E_2 + (\alpha_1^2 - \alpha_2^2)E_3}{2(\alpha_2 - \alpha_3)E_1 + (\alpha_3 - \alpha_1)E_2 + (\alpha_1 - \alpha_2)E_3} \tag{3.4}$$

The change of weights is achieved for each set of input pattern foe saving time.

3.4 Incremental Updating with Variable Learning Rate ; V.L.

In the ordinary Incremental updating, the learning rate is constant through the procedure. As the procedure goes on, the weight vector comes up to but does not reach the global minimum and oscillates around the minimum point. The bigger the learning rate is assigned to be , the bigger the oscillation becomes.

In this learning procedure, several learning rates are given and they are all examined at the first learning cycle. The best learning rate which gives the minimum error is adopted in the sequential cycles. When the change rate

$$\frac{|E(W(t+1)) - E(W(t))|}{E(W(t))} \tag{3.5}$$

of error becomes less than a given value (0.1 in this paper), then the learning rate is made to be smaller (1/10 of the value) to continue the procedure.

3.5 Steepest Descent Method with Momentum Term ; MO

In the steepest descent method of learning, there exists a problem when we implement ordinary neural networks or a neo fuzzy neuron. How is the value of learning rate α to be assigned? As might be expected, a large α corresponds to rapid learning but might also result in oscillations. In order to suppress this oscillations a momentum term is added to the learning rule by Rumelhart, Hilton and Williams [6]. Its neo fuzzy neuron version is described as

$$\Delta w_{ij}(t+1) = -\alpha \sum_{k=1}^{p} (y_k - te_k)\mu_{ij}(x_{ik}) + \beta \Delta w_{ij}(t) \tag{3.6}$$

where β represents a momentum parameter. The momentum parameter is a positive value and less than unity, and there is no general criterion how the value is to be chosen as well as the learning rate α . In this paper, α and β are empirically assigned to be 0.3 and 0.2, respectively. This algorithm allows us to expect the less learning cycles to converge.

3.6 Comparison

Learning algorithms described above are examined for a 1-input 1-output neo fuzzy neuron by computer simulation employing a work station. An unpredictable time series of one dimension is previously created and 50 data points are sequentially applied to the input of the delay element followed by the neo fuzzy neuron. This system can achieve the learning of the synapse for predicting one incremental step ahead. A set of 50 data points achieves one learning cycle and thus one batch learning (3.2, 3.3 and 3.5) is accomplished for each set of data. The learning procedure is iterated for 100 machine cycles for each case, that is, the total amount of 5000 data are applied for learning examination. The input range is divided into 12 fuzzy segments. Thus this 1-input 1-output neo fuzzy neuron possesses a nonlinear synapse which is characterized by 12 membership functions and 12 weights. All the initial weights are assigned to be zero which is quite different from an conventional neural networks.

Experimental results from computer simulation is shown in Fig.2, where the horizontal axis represents the learning time mt (machine time) on the work station and the vertical axis the root mean squared error (R.M.S. error), both in logarithmic scale. The root mean squared error is defined [8] by

$$E_{RMS} = \sqrt{\frac{1}{p}\sum_{k=1}^{p} (y_k - te_k)^2} \tag{3.7}$$

Four procedures SE., BA., V.L. and MO. exhibit the convergence within 1 msec, while D.R.M. does about 1 sec. Because in D.R.M. procedure, the optimum value of a is calculated at each learning cycle.

4. GUARANTEE FOR A GLOBAL MINIMUM

In this chapter, it is proved that the neo fuzzy neuron model guarantees the global minimum in the error-weight space under an arbitrary initial condition.

[Theorem]

If the error-weight space E-w possesses a global minimum, it does not do any local minima. That is, E-w necessarily gives the global minimum because of its unimodality.

[Proof]

Let us assume that if E-w possesses a global minimum, then it also possesses local minima. Let the weight w_{uv} giving a global minimum E_{min} be $w_{uv,min}$. That is,

$$E = \frac{1}{2}\sum_{k=1}^{p} \left(\sum_{j=1}^{n}\sum_{i=1}^{m} \mu_{ij}(x_{ik})w_{ij,min} - te_k\right)^2 = E_{min} \tag{4.1}$$

$$\left(\frac{\partial E}{\partial w_{uv}}\right)_{w_{uv} = w_{uv,min}} = \sum_{k=1}^{p} \left(\sum_{j=1}^{n}\sum_{i=1}^{m} \mu_{ij}(x_{ik})w_{ij,min} - te_k\right)\mu_{uv}(x_{uk}) = 0 \tag{4.2}$$

Let the weight w_{uv} giving a local minimum E_{loc} be $w_{uv,loc}$. That is ,

$$E = \frac{1}{2}\sum_{k=1}^{p} \left(\sum_{j=1}^{n}\sum_{i=1}^{m} \mu_{ij}(x_{ik})w_{ij,loc} - te_k\right)^2 = E_{loc} \tag{4.3}$$

$$\left(\frac{\partial E}{\partial w_{uv}}\right)_{w_{uv} = w_{uv,loc}} = \sum_{k=1}^{p} \left(\sum_{j=1}^{n}\sum_{i=1}^{m} \mu_{ij}(x_{ik})w_{ij,loc} - te_k\right)\mu_{uv}(x_{uk}) = 0 \tag{4.4}$$

where $E_{min}<E_{loc}$.

An arbitrary real number r satisfies

$$\left(\left(\frac{\partial E}{\partial w_{uv}}\right)_{w_{uv} = w_{uv,min}} + r\left(\frac{\partial E}{\partial w_{uv}}\right)_{w_{uv} = w_{uv,loc}}\right)^2 = 0 \tag{4.5}$$

From Schwarz's inequality and Eqs.(4.1), (4.2), (4.3), (4.4) and (4.5), we can get

$$0=\left(\sum_{k=1}^{p}\left(\left(\sum_{j=1}^{n}\sum_{i=1}^{m}\mu_{ij}(x_{ik})w_{ij,min}-te_k\right)+r\left(\sum_{j=1}^{n}\sum_{i=1}^{m}\mu_{ij}(x_{ik})w_{ij,loc}-te_k\right)\right)\mu_{uv}(x_{uk})\right)^2$$
$$\le\sum_{k=1}^{p}\left(\left(\sum_{j=1}^{n}\sum_{i=1}^{m}\mu_{ij}(x_{ik})w_{ij,min}-te_k\right)+r\left(\sum_{j=1}^{n}\sum_{i=1}^{m}\mu_{ij}(x_{ik})w_{ij,loc}-te_k\right)\right)^2\sum_{k=1}^{p}\mu_{uv}(x_{uk})^2$$
$$\le\left(2E_{min}+2r^2E_{loc}+4r\sqrt{E_{min}E_{loc}}\right)\sum_{k=1}^{p}\mu_{uv}(x_{uk})^2 \qquad (4.6)$$

where the former equality is valid when there exists an arbitrary real number λ satisfying Eqs.(4.7), and the latter equality is valid when there exists an arbitrary real number ε satisfying Eq.(4.8)and Eq.(4.9) is valid.

$$\left(\sum_{j=1}^{n}\sum_{i=1}^{m}\mu_{ij}(x_{ik})w_{ij,min}-te_k\right)+r\left(\sum_{j=1}^{n}\sum_{i=1}^{m}\mu_{ij}(x_{ik})w_{ij,loc}-te_k\right)=\lambda\mu_{uv}(x_{uk}) \qquad (4.7)$$

$$\sum_{j=1}^{n}\sum_{i=1}^{m}\mu_{ij}(x_{ik})w_{ij,min}-te_k=\varepsilon\left(\sum_{j=1}^{n}\sum_{i=1}^{m}\mu_{ij}(x_{ik})w_{ij,loc}-te_k\right) \qquad (4.8)$$

and

$$\left(\sum_{j=1}^{n}\sum_{i=1}^{m}\mu_{ij}(x_{ik})w_{ij,min}-te_k\right)\left(\sum_{j=1}^{n}\sum_{i=1}^{m}\mu_{ij}(x_{ik})w_{ij,loc}-te_k\right)\ge 0 \qquad (4.9)$$

Thus we can obtain the following equation with equality under the conditions of Eqs.(4.7), (4.8) and (4.9).

$$0\le E_{min}+r^2E_{loc}+2r\sqrt{E_{min}E_{loc}} \qquad (4.10)$$

When Eqs.(4.7), (4.8) or (4.9) is not valid, the equality in Eq.(4.10) is not valid. That is,

$$E_{min}+r^2E_{loc}+2r\sqrt{E_{min}E_{loc}}>0 \qquad (4.11)$$

is valid for an arbitrary real number r. However,

$$r=-\sqrt{\frac{E_{min}}{E_{loc}}} \qquad (4.12)$$

can make the left hand side of Eq.(4.11) to be zero. Thus Eq.(4.11) includes absurdity in itself.

On the other hand, when Eqs.(4.7), (4.8) and (4.9) are simultaneously valid, the equality in Eq.(4.10) should be valid. That is,

$$E_{min}+r^2E_{loc}+2r\sqrt{E_{min}E_{loc}}=0 \qquad (4.13)$$

for an arbitrary real number r. However, in this case, we can get the following equation from Eqs.(4.1), (4.2), (4.3), (4.4), (4.5), (4.7) and (4.8).

$$\left(\left(\frac{\partial E}{\partial w_{uv}}\right)_{w_{uv}=w_{uv,min}}+r\left(\frac{\partial E}{\partial w_{uv}}\right)_{w_{uv}=w_{uv,loc}}\right)^2=\frac{2E_{loc}(\varepsilon+r)^2}{\lambda} \qquad (4.14)$$

Eq.(4.14) is not necessarily equal to zero for any real numbers ε, λ and r and thus it exhibits absurdity with respect to Eq.(4.5).

Consequently, the assumption , " if the error-weight space E-w possesses a global minimum, then it also possesses local minima, " is not valid, and if the error-weight space E-w possesses a global minimum, it does not do any local minima.

Q.E.D.

5. COMPARISON WITH A CONVENTIONAL THREE-LAYERED NEURAL NETWORK

The ability of a neo fuzzy neuron (NFN) to describe a system is examined by comparison with a conventional three-layered neural network (NN). One dimensional nonlinear dynamical system employing a logistic map is adopted to create the time series which is used for training. The time series of 50 time points is sequentially applied to the delay element to feed the data $x_k(t)$ at time t to the NFN or NN, then the output is the calculated value $y_k(t+1)$ at $t+1$ and the input to the delay element is $te_k(t+1)$. The NN achieves the learning by the error back propagation algorithm [9] and the NFN does by D.R.M. in the same manner to 3.3. The examination was implemented for 100,000 training cycles for each.

Fig.3 shows the decay of R.M.S. errors for various neural networks and neo fuzzy neurons. NN4 and NN10 represent conventional neural networks which possesses 4 hidden layer neurons and 10 ones, respectively. NFN8, NFN12 and NFN16 represents neo fuzzy neurons, the nonlinear synapse of which is described with 8, 12 and 16 pairs of membership functions and weights. The figure shows drastic difference in learning speed between the neo fuzzy neurons and conventional neural networks. The former one can reduce the error by two decades within 1 sec. , while the other one cannot reduce the error less than one hundredth within 100,000 sec. The figure also shows that the neo fuzzy neuron with more segmentation (i.e. more membership functions and weights) exhibits the smaller error at the sacrifice of learning speed.

Fig.4 (a) and (b) shows the identified nonlinear functions in the neural network and the synapse of the neo fuzzy neuron, which correspond to a logistic function to create chaotic time series used for learning. The NFN12 exhibits much higher accuracy rather than the NN10.

6. CONCLUSIONS

Some learning procedures are examined for a neo fuzzy neuron model. Its learning characteristics is compared with that of a conventional three layered neural network. A neo fuzzy neuron model shows a drastic improvement in learning speed and accuracy. Furthermore, the guarantee of convergence to a global minimum in a neo fuzzy neuron is proved. This aspect implies the quite difference from traditional neural networks.

REFERENCES

[1]Takeshi Yamakawa, Japanese Patent Application, No.TOKU-GAN-HEI 1-133690 (May 1989).

[2]Takeshi Yamakawa and Shin-ya Tomoda, "A Fuzzy Neuron and Its Application to Pattern Recognition", Proc. the Third IFSA Congress, Seattle, Washington , August 6-11 (1989).

[3]Takeshi Yamakawa, Japanese Patent Application, No.TOKU-GAN-HEI 4-132897 (May 1992).

[4]Takeshi Yamakawa, Eiji Uchino, Tsutomu Miki and Hiroaki Kusanagi, "Identification of Nonlinear Dynamical Systems by a Neo Fuzzy Neuron and Prediction of Their Behavior," Proc. 8th Fuzzy System Symposium, Hiroshima, May 26-28, pp.249-252 (1992).

[5]Takeshi Yamakawa, Eiji Uchino, Tsutomu Miki and Hiroaki Kusanagi, "A Neo Fuzzy Neuron and Its Applications to System Identification and Prediction of the System Behavior," Proc. the 2nd International Conference on Fuzzy Logic & Neural Networks, Iizuka, Japan, July 17-22, pp.477-483 (1992).

[6]Jacek M. Zurada, "Introduction to Artificial Neural Systems," West Publishing Company , (1992).

[7]David E. Rumelhart, James J. McClelland and the PDP Research Group, Parallel Distributed Processing, Vol.1: Foundations, The MIT Press, Cambridge, Massachusetts, pp.322-330 (1986).

[8]Heisuke Hironaka, Encyclopedia of Mathematical Sciences, Osaka Shoseki, p.483 (1991) (in Japanese).

[9]David E. Rumelhart, James J. McClelland and the PDP Research Group, Parallel Distributed Processing, Vol.1 : Foundations, The MIT Press, Cambridge, Massachusetts (1986).

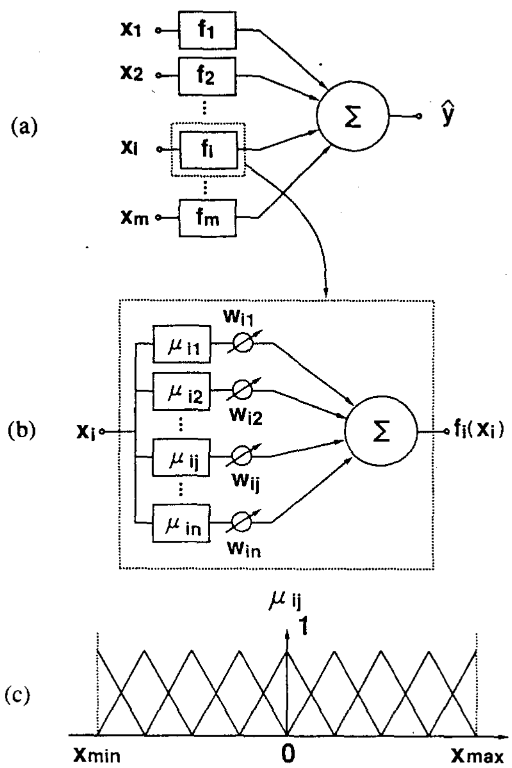

Fig.1 (a)Structure of the neo fuzzy neuron, each synaptic characteristics of which is represented by a nonlinear function f_i. (b)Structure of the nonlinear synapse which is described with a set of if-then rules including singletons in consequents. (c)Triangular and complementary membership functions assigned for the fuzzy segments in the input space.

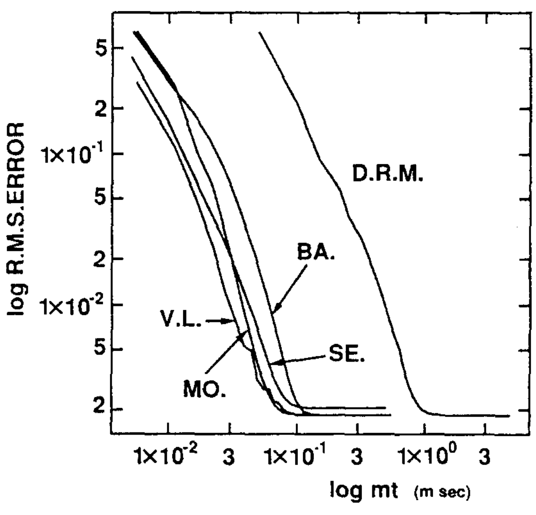

Fig.2 Comparison of learning rates between different learning procedures (root mean squared error vs. machine time for learning). The input space is divided into 12 fuzzy segments in the neo fuzzy neuron under test.. **SE.**: Incremental updating (α=0.3). **BA.**: Batch learning (α=0.3). **D.R.M.**:: Direct search method with second order interpolation (h=0.05, k=0.2). V.L.: Incremental updating with variable learning rate (α_0=0.05, 0.1, 0.2, 0.5, 1.0, 2.00; permissible change rate of error = 0.1). **MO.**: Steepest descent method with momentum term (α=0.3, β=0.2).

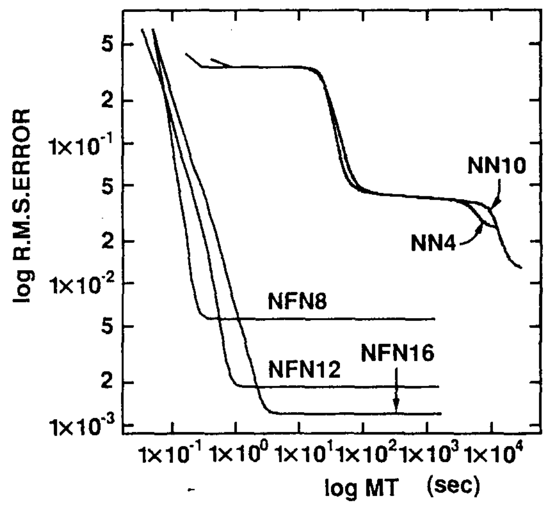

Fig.3 Comparison of learning rates between neo fuzzy neurons (NFN) and conventional three-layered neural networks (NN). NN4, NN10: 4 and 10 neurons in a hidden layer. NFN8, NFN12, NFN16: 8, 12 and 16 fuzzy segments in input space. A learning speed of a neo fuzzy neuron exceeds 100,000 times of conventional neural networks.

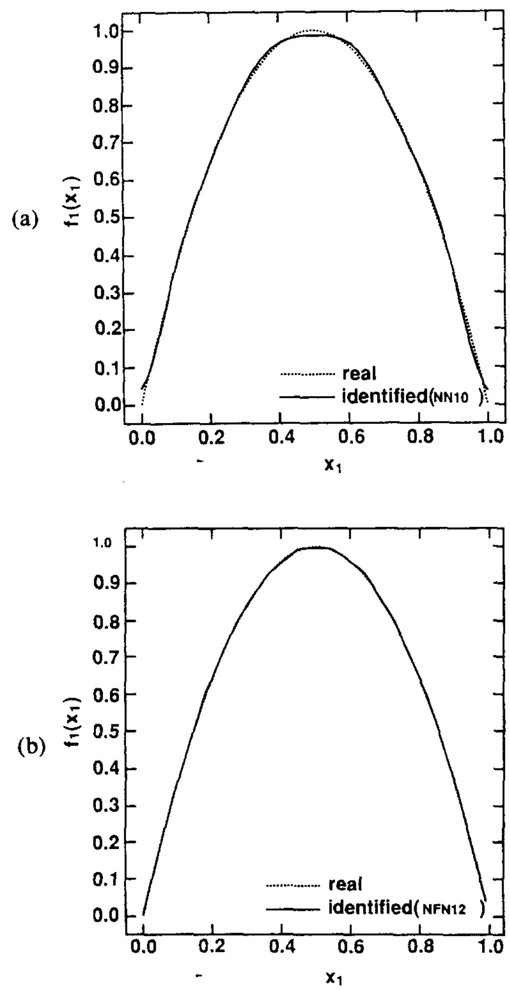

Fig.4 Nonlinear functions identified by (a) a neural network and by (b) the synapse of a neo fuzzy neuron.