AN ADAPTIVE HYSTERESIS THRESHOLD METHOD FOR AN ASSOCIATIVE MEMORY USING MUTUALLY CONNECTED NEURAL NETWORK

Kenji NAKAYAMA Naoki MITSUTANI
Dept. of Electrical and Computer Eng., Faculty of Tech., Kanazawa Univ.
2-40-20, Kodatsuno, Kanazawa, 920 JAPAN

ABSTRACT

A new adaptive threshold method is proposed for an associative memory using mutually connected neural network. In a learning process, the network state, that is the unit state $u_i(m)$, is fixed to a pattern P(m) to be memorized. Connection weights are iteratively adjusted so that the unit input satisfies $v_i(k) \geqq \theta$ for $u_i(m)=1$, and $v_i(k) \leqq -\theta$ for $u_i(m)=0$. In an association process, P(m) is recollected from its degraded version Q(m). The network state is initially set to be Q(m). At the nth state transition step, if $v_i(n) \geqq \phi(n)$, then $u_i(n+1)=1$. If $v_i(n) \leqq -\phi(n)$, then $u_i(n+1)=0$. Furthermore, if $-\phi(n) < v_i(n) < \phi(n)$, then $u_i(n+1)=u_i(n)$. $\phi(n)$ is initially chosen to be $\phi(0) > \theta$, and is gradually decreased as $\phi(n)=\phi(0)-\alpha n$, where α is constant. Computer simulation was carried out, using 51 and 153 patterns, which appear on a key board. A neural network has 16x16=256 units and full connections. The results demonstrate that drastic improvements in a memory capacity and association rates can be achieved. For example, an association rate for 51 patterns with 40 noises has been increased from 12.2% to 97.7%, compared with a single threshold method.

I INTRODUCTION

An associative memory is one of useful applications of artificial neural networks. One approach is to memorize patterns on equilibrium states of a mutually connected neural network. Connection weights are adjusted so that equilibrium states express the patterns. Conventional methods for adjusting connection weights include auto-correlation methods and orthogonal methods [1]-[6]. These methods, however, assume symmetrical connection weights, and are effect only for lineally independent patterns and orthogonal patterns. Therefore, a memory capacity for arbitrary patterns is strictly limited. By assuming asymmetrical connection weights, and by adjusting connection weights through a Hebb rule, for instance, a memory capacity can be improved. However, when the patterns are degraded by noise or obstacles, association rates are significantly decreased.

This paper focuses on improvements in a memory capacity for arbitrary patterns and an association rate for degraded patterns. For this purpose, new learning and association methods, employing hysteresis threshold levels are proposed. In a learning process, a fixed hysteresis threshold is employed in order to obtain noise margin [7]. In an association process, an adaptive hysteresis threshold is employed, in order to suppress error propagation. Discussions on a memory capacity and an association rate are provided. In order to confirm efficiency of the proposed method, computer simulation is demonstrated. Two sets of memory patterns, including 51 and 153 patterns, are employed. A mutually connected neural network, having 16x16=256 units and full connections, is used.

II A LEARNING METHOD

The connection weights are adjusted so as to store the patterns on the equilibrium states. A connection weight from the ith unit to the jth unit is denoted by w_{ij}. A self-loop w_{ii} is not used. A unit takes 2-levels states, that is, 1 and 0. The state of a unit is also called 'output' in this paper.

(1)Initial guess:

All connection weights are initially set to be zero.

$$w_{ij}(0)=0, \quad 1 \leqq i,j \leqq N \tag{1}$$

(2)Network state:

The network state is fixed to a pattern P(m) to be memorized. In other words, the unit states are not changed regardless of the inputs.

(3)Adjusting connection weights:

Let the state of the jth unit for P(m) be $u_j(m)$, and the input at the kth adjusting step be $v_j(k)$. In order to obtain enough noise margin, the input is required to satisfy $v_j(k) \geqq \theta$ for $u_j(m)=1$, and $v_j(k) \leqq -\theta$ for $u_j(m)=0$. Figure 1 shows this relation. The input $v_j(k)$ should locate in the shaded portion, according as $u_j(m)$. This hysteresis threshold can provide not only a noise margin but also ability of detecting the correct state of units in degraded patterns. The latter property will be discussed in Sec. IV.

The connection weights are up-dated by

$$w_{ij}(k+1) = w_{ij}(k) + \Delta w_{ij}(k) \tag{2}$$

Based on the above idea, the change $\Delta w_{ij}(k)$ is determined as follows:

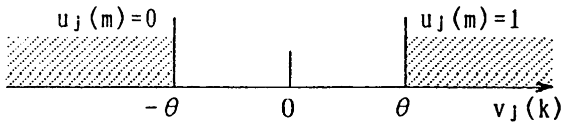


Fig. 1 Regions of the input $v_j(k)$ for $u_j(m)$, in a learning process.

$$v_j(k) = \sum_{i=1}^{N} w_{ij}(k) u_i(m), \quad w_{jj}=0 \tag{3}$$

i) For $u_j(m)=1$, if $v_j(k) \geqq \theta$ then $\Delta w_{ij}(k)=0$ (4a)

if $v_j(k) < \theta$ then $\Delta w_{ij}(k)=\Delta w$ (4b)

ii) For $u_j(m)=0$, if $v_j(k) \leqq -\theta$ then $\Delta w_{ij}(k)=0$ (5a)

if $v_j(k) > -\theta$ then $\Delta w_{ij}(k)=-\Delta w$ (5b)

provided that $\Delta w>0$ and $u_i(m)=1$. Δw_{ij} is always zero for $u_i(m)=0$, regardless of $u_j(m)$. All connection weights are changed simultaneously for one pattern P(m).

(4)The above process (3) is repeated for all patterns P(1)~P(M). This process is counted as one learning step.

(5)The above processes (3) and (4) are repeated until $\Delta w_{ij}(k)$ becomes zero for all patterns and for all units. Δw is gradually decreased during a learning process, in order to avoid any vibration.

III MEMORY CAPACITY

A memory capacity of an associative memory is usually evaluated by a ratio of the numbers of memory patterns (M) and of units (N), that is, M/N. By employing asymmetrical connection weights, a memory capacity can be increased. It is, however, very difficult to derive a general formula for a memory capacity. Because, it is highly dependent on cross-correlation among the patterns. Therefore, in this section, the critical patterns are provided. Proofs are omitted due to a page limit. P(m) is used as a set of units, whose state is 1, in the following.

(1)If $P(1) \cup P(2) - P(1) \cap P(2) = 1$ unit, (6)
then P(1) and P(2) cannot be memorized simultaneously.

(2)If $P(1) \cup P(2) - P(1) \cap P(2) \geqq 2$ units, (7)
then P(1) and P(2) can be memorized simultaneously.

(3)Let u_1 and u_2 express the 1st and 2nd units.

If $u_1 = P(1) \cap \{P(1) \cup P(2) - P(1) \cap P(2)\}$ (8a)

$u_2 = P(2) \cap \{P(1) \cup P(2) - P(1) \cap P(2)\}$ (8b)

$P(3) = \{u_1, u_2\}$ (8c)

then P(1), P(2) and P(3) cannot be memorized simultaneously.

(4)Under the condition (3), if P(3) further includes another unit u_3, besides u_1 and u_2, then P(1), P(2) and P(3) can be memorized simultaneously.

(5)By combining the patterns, satisfying the conditions (1) and (3), more complicated prohibitory sets of patterns can be generated. In other words, prohibitory patterns can be detected using the conditions (1) and (3).

Examples of patterns corresponding to (1)~(3) are shown in Fig.2.

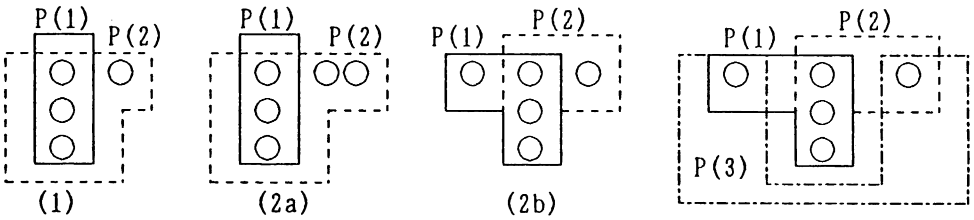


Fig. 2 Examples of patterns satisfying conditions (1), (2) and (3).

In a real world, however, patterns satisfying the conditions (1) and (3), seem to be limited. So, basically speaking, a mutually connected neural network, with asymmetrical connection weights, can memorize a large number of patterns. Another problem for a large memory capacity is a probability of recollecting the memorized pattern from its degraded pattern. This will be discussed in in the next section.

IV AN ASSOCIATION METHOD

A. Adaptive Hysteresis Threshold

In an association process, the memorized pattern P(m) is recollected from its incomplete pattern Q(m), which is degraded by random noise and obstacles.

$$Q(m) = P(m) + E \qquad (9)$$

where E represents errors. The network state transits following

$$v_j(n) = \sum_{i=1}^{N} w_{ij} u_j(n), \quad w_{jj}=0 \qquad (10)$$

$$u_j(n+1) = f(v_j(n)) \qquad (11)$$

where $v_j(n)$ and $u_j(n)$ are the input and output of the jth unit at the nth state transition step. f() is a nonlinear function.

In the above state transition, the errors propagate and spread over the whole network. As a result, the network state easily falls into undesired equilibrium state. So, it is very important to suppress the error propagation.

The idea behind adaptive hysteresis threshold is to use only the correct information about the state of units. Accuracy of the states can be evaluated using the unit inputs. Because they are adjusted to satisfy ether $v_j(k) \geqq \theta$ or $v_j(k) \leqq -\theta$, in the previous learning process. Thus, the state of units, whose input satisfies ether $v_j(n) \geqq \phi(n)$ or $v_j(n) \leqq -\phi(n)$, could be ether $u_j(n)=1$ or $u_j(n)=0$, respectively. $\phi(n)$ is a threshold level at the nth network state transition step. It is determined as follows: Since errors quickly spread over the network, accuracy of the states, which are used in the early stage, has strong effect on the convergence route, namely, the final destination. Therefore, the initial threshold $\phi(0)$ is chosen to be larger than θ. In the resulting equilibrium state, however, the requirements for the inputs are ether $v_j(k) \geqq \theta$ or $v_j(k) \leqq -\theta$. Therefore, $\phi(n)$ is decreased from $\phi(0)$ to around θ step by step during the association process.

B. Association Process

Based on the idea, mentioned above, the following association process is proposed.

(1)Initial State:

The network state is initially set to the degraded pattern Q(m).

(2)Network State Transition:

The network state transits following Eqs.(10) and (11). A nonlinear function f() has the following hysteresis threshold levels.

$$\text{If } v_j(n) \geqq \phi(n) \text{ then } u_j(n+1)=1 \qquad (12a)$$

$$\text{If } v_j(n) \leqq -\phi(n) \text{ then } u_j(n+1)=0 \qquad (12b)$$

$$\text{If } -\phi(n) < v_j(n) < \phi(n) \text{ then } u_j(n+1)=u_j(n) \qquad (12c)$$

The above unit state transition is illustrated in Fig.3. $\phi(n)$ is gradually decreased during the association process as follows:

$$\phi(n)=\phi(0)-\alpha n \qquad (13)$$

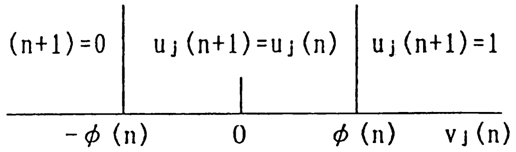


Fig. 3 Unit state transition according as $v_j(n)$, in an association process.

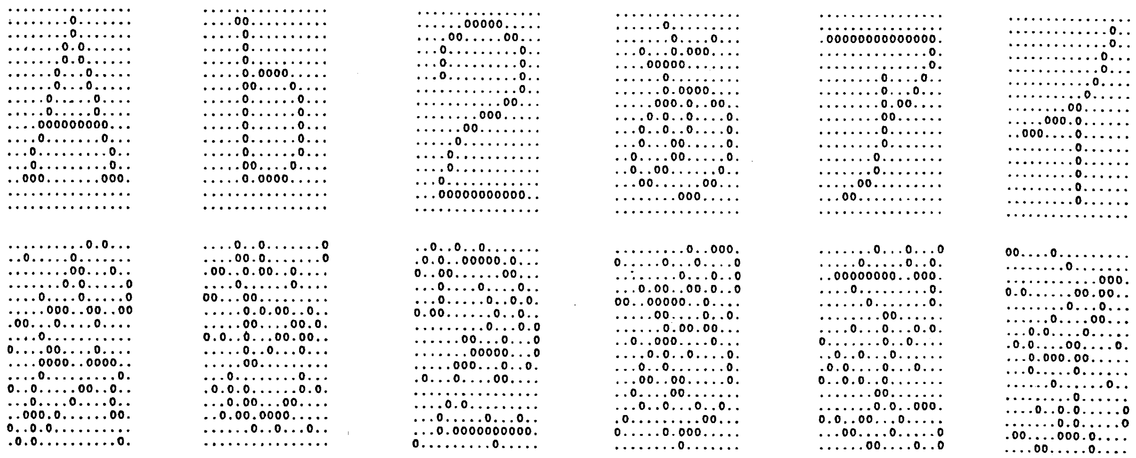
$\phi(0)$ is chosen to be larger than θ. α is constant.

All units simultaneously change their state following Eqs.(12a)-(12c). This process is repeated until the network state does not change any more, namely, it arrives at an equilibrium state. This state represents the associated pattern.

V SIMULATION

A mutually connected neural network, having 16x16=256 units and full connections, is used. Two sets of patterns, including 52 and 154 patterns, are evaluated. One of them includes alphabet(52). The other set includes numbers(10) and the Japanese syllabary (92) in addition to the former set.

The patterns are degraded by random noises. Noises are generated by reversing the state of units, which are randomly selected. The number of noises is 40. One hundred sets of random noises are used to evaluate statistic association rates for all patterns. Examples of the memorized patterns and their degraded version are shown in Fig.4.



「A」 「b」 「2」 「あ」 「ア」 「イ」

Fig. 4 Examples for patterns to be memorized, and their noisy patterns, with 40 noises.

The threshold level θ in the learning process is chosen to be 30. Change of connection weights Δw is gradually decreased as 1, 0.5, 0.2, 0.1, for every 50 learning steps. The threshold in an association process is varied as $\phi(n)=\phi(0)-n$. The initial value $\phi(0)$ is chosen to be 100. Furthermore, a single threshold method, using $\theta=0$ and $\phi(n)=0$, and a semi-hysteresis threshold method, using $\theta=30$ and $\phi(n)=0$, are also evaluated for comparison.

Association rates are listed in Table 1. In two pattern sets, since alphabet patterns 「I」 and 「l」 correspond to the prohibitory patterns given by (1) in Sec.III, 「I」 cannot be memorized. So, it is omitted in the following. Method I is the single threshold method. Method II is the semi-hysteresis threshold method. Method III is the proposed one. In this table, the numbers indicated by *2, *3 and *4 represent association rates(%), that is, probabilities of recollecting the original pattern, another memorized pattern and non-memorized patterns, respectively. Let association rates for them from Q(m) be p(m), p'(m) and r(m), respectively. They are calculated using one hundred sets of random noises. For instance, letting the number of the equilibrium states, which represent P(m), be $N_{P(m)}$, p(m) is given by $N_{P(m)}/N_R$ (x100%), where N_R is the number of the random noise sets. Furthermore, data listed in Table 1 are calculated as follows:

$$\text{Upper}(*2):\quad p = \frac{1}{M}\sum_{m=1}^{M} p(m) \qquad (14a)$$

$$\text{Middle}(*3):\quad p' = \frac{1}{M}\sum_{m=1}^{M} p'(m) \qquad (14b)$$

$$\text{Bottom}(*4):\quad r = \frac{1}{M}\sum_{m=1}^{M} r(m) \qquad (14c)$$

Table 1 Association rates(%) for noisy patterns.

	Patterns *1	51 Patterns	153 Patterns
	Noises	40	40
I	$\theta=0$ $\phi(n)=0$	12.2 *2 13.7 *3 74.1 *4	0.2 0.4 99.4
II	$\theta=30$ $\phi(n)=0$	82.5 8.9 8.6	7.5 1.8 90.7
III	$\theta=30$ $\phi(n)=\phi(0)$ $-n$	97.7 1.3 1.0	64.5 0.6 34.9

*1 An alphabet pattern 「I」 is omitted.
*2 Original pattern, *3 Another memorized pattern, *4 Non-memorized pattern.

Although Method I can memorize all patterns, association rates for noisy pattens are very poor. On the contrary, by employing the hysteresis threshold only in the learning process, association rates (*2) for the original patterns can be significantly improved from 12.2% to 82.5% for 51 patterns. However, in the case of 153 patterns, the proposed method (Method III) is very effective compared with Method II. Thus, it can be confirmed that the proposed method is very useful both in association rates for noisy patterns and a memory capacity.

VI CONCLUSION

A new adaptive hysteresis threshold method has been proposed for an associative memory using a mutually connected neural network. Computer simulation has demonstrated that the proposed method can provide drastic improvements in association rates and a memory capacity. For example, an association rate for 51 patterns with 40 random noises can be increased from 12.2% to 97.7%.

REFERENCES

[1]K.Kohonen, Self-Organization and associative Memory, 2nd Ed., Springer-Verlag, 1987.
[2]K.Nakano,"Associatron-A model of associative memory", IEEE Trans. vol.SMC-2, pp.380-388, 1972.
[3]S.Amari,"Neural theory of association and concept-formation", Biol. Cybern., vol.26, pp.175-185, 1977.
[4]J.J.Hopfield,"Neural networks and physical systems with emergent collective computational abilities", Proc. Natl. Sci. USA, vol.79, pp.2554-2558, 1982.
[5]D.Amit et al,"Storing infinite number of patterns in a spin-glass model of neural networks", Phys. Rev. Lett., vol.55, pp.1530-1533, 1985.
[6]S.Amari and K.Maginu,"Statistical neurodynamics of associative memory", Neural Networks, vol.1, pp.63-73, 1988.
[7]N.Mitsutani and K.Nakayama,"A learning method for memorizing arbitrary patterns in mutually connected neural network (in Japanese)", IEICE Japan, Report of Technical Meeting on Nero Computing, vol.NC89-60, pp.67-72, Feb. 1990.