Memory capacity bound and threshold optimization in recurrent neural network with variable hysteresis threshold

Proceedings of 1993 International Joint Conference on Neural Networks

MEMORY CAPACITY BOUND AND THRESHOLD OPTIMIZATION IN RECURRENT NEURAL NETWORK WITH VARIABLE HYSTERESIS THRESHOLD

Kenji NAKAYAMA Katsuaki NISHIMURA Hiroshi KATAYAMA

Dept. of Electrical and Computer Eng., Faculty of Tech., Kanazawa Univ.
2-40-20, Kodatsuno, Kanazawa 920 JAPAN
E-mail: nakayama@haspnn1.ec.t.kanazawa-u.ac.jp

ABSTRACT

Authors have proposed an asymmetrical associative neural network (NN) using variable hysteresis threshold and its learning and association algorithms. It can drastically improve noise performance, that is insensitivity to noise. In this paper, memory capacity bound and threshold optimization in this associative NN are further discussed. Binary random patterns are considered. First, relation between the number of patterns and the number of iterations is investigated. The latter gradually increases until some number of patterns. After that, it suddenly increases. This is a very peculiar phenomenon. This turning point gives the memory capacity bound, that is about 1.56N, where N is the number of units. Next, threshold optimization is discussed. Relation between threshold and noise performance, and effects of connection weight distribution on noise performance are theoretically discussed. Based on these results, a ratio of step-size and the threshold is optimized to be $0.5/(N_P-1)$, where N_P is the number of units on the pattern. Numerically statistical simulation demonstrates efficiency of the proposed methods.

I INTRODUCTION

An associative memory is one of useful applications of artificial neural networks (NNs). Connection weights are adjusted so that the equilibrium states express the patterns to be memorized. Conventional methods include auto-correlation methods and orthogonal methods [1]-[6]. These methods, however, assume symmetrical weights, and are effective only for lineally independent patterns or orthogonal patterns. Therefore, a memory capacity and noise performance, that is insensitivity to noise, are strictly limited.

Authors have proposed an asymmetrical associative neural network using variable hysteresis threshold and its learning and association algorithms [7],[8]. It can drastically improve noise performance. In this paper, the memory capacity bound and threshold optimization are further discussed. Binary random patterns are taken into account.

II ASYMMETRICAL RECURRENT NEURAL NETWORK

The asymmetrical recurrent NN proposed in [7],[8] is briefly described here.

2.1 Network Structure

The network transition is formulated as follows:

$$u_j(n) = \sum_{i=1}^{N} w_{ij} v_i(n), \quad w_{ii}=0 \tag{1}$$

$$v_j(n+1) = f(u_j(n)) = \begin{cases} 1, & u_j(n) \geqq T(n) & \text{(2a)} \\ v_j(n), & |u_j(n)| < T(n) & \text{(2b)} \\ 0, & u_j(n) \leqq -T(n) & \text{(2c)} \end{cases}$$

$u_i(n)$ and $v_i(n)$ are the input and output of the ith unit, respectively. w_{ij} is a connection weight from the ith unit to the jth unit. $T(n)$ is threshold at n.

2.2 Connection Weight Learning

Initial guess of the weights are set to zero. The network state is set to be one of the pattern P(m). The unit input is calculated by Eq.(1), using the ith element of P(m), denoted $p_i(m)$, as $v_i(n)$. The weights are updated as follows:

$$u_j(n) = \sum_{i=1}^{N} w_{ij}(n)p_i(m) \tag{3}$$

$$w_{ij}(n+1) = w_{ij}(n) + \eta(n)\delta_j(n)p_i(m) \tag{4}$$

$$\delta_j(n) = p_j(m)S[T-u_j(n)] + (p_i(m)-1)S[u_j(n)+T] \tag{5}$$

$$S[x] = 1 \text{ for } x>0 \text{ , and } = 0 \text{ for } x\leq 0 \tag{6}$$

$\eta(n)$ is step-size. In order to achieve both fast and stable convergence, $\eta(n)$ is determined to be large value at the beginning, and is gradually decreased. All weights are simultaneously updated taking all patterns into account. This process is counted as one iteration.

2.3 Association from Incomplete Patterns

After the training completed, all unit inputs satisfy

$$\text{If } p_i(m)=1, \text{ then } u_i(n)\geq T \tag{7a}$$

$$\text{If } p_i(m)=0, \text{ then } u_i(n)\leq -T \tag{7b}$$

By adding noise, these conditions may be destroyed. Therefore, based on a noisy pattern, it is difficult to estimate the correct state of each unit. However, if the unit input satisfies $|u_i(n)| >> T$, then the state of this unit can be expected to be 1 or 0 with high probability. In the early stage, it is very important to update the correct units only. Therefore, T(n) is initially set to large, and is gradually decreased toward T used in the training.

III MEMORY CAPACITY BOUND FOR RANDOM PATTERNS

The memory capacity is highly dependent on correlation among the patterns. Combinations of patterns, which cannot be simultaneously memorized, have been discussed [],[]. In this paper, random patterns are taken into account. Half units, takes 1. The memory capacity is investigated based on relation between the number of patterns and the number of training iterations.

Figure 1 shows the simulation results. The number of units N is 64. The threshold T is chosen to be 10, 20 and 30. The result using T=20 is shown here. The other thresholds provide the same results. The number of iterations gradually increases up to 80 patterns. After that, it suddenly increases. This is a very peculiar phenomenon. In the simulation, the training converges for 100 patterns, with 7619 iterations. The number of the patterns could be increased a little more. However, from the very sharp slope in Fig.1, it is almost limited

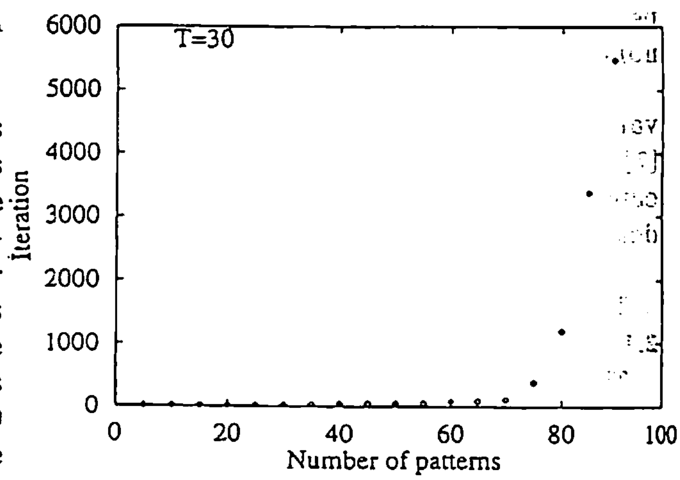

Fig.1 Relation between the number of iterations and the number of patterns to be memorized.

near by 100 patterns.
Thus, the memory capacity is about $100/64 \fallingdotseq 1.56$ times as large as N. This bound is independent on the threshold.

IV HYSTERESIS THRESHOLD OPTIMIZATION

4.1 Threshold and Noise Performance

If connection weights are increased in proportional to the threshold, then the noise performance is not improved. Because the error transfer gain is also increased. Suppose the training converges using the threshold T, and the weights w_{ij} are obtained, which satisfy

$$u_i(n) = \Sigma w_{ij} p_i(m) \tag{8}$$

$$| u_i(n) | \geq T \tag{9}$$

$$\text{Noise margin} = | u_i(n) | - T \tag{10}$$

For the scaled threshold αT, αw_{ij} could be one solution. The noise margin is also scaled by α. However, at the same time, the error transfer gain is scaled by α, then noise performance is not improved any more.

4.2 Variance of Connection Weights and Noise Performance

Solutions for the connection weights are not limited to only one set. Noise performance of the NN, using each solution, can be evaluated based on the distribution of the weights, which is defined as follows:

$$M_{ABS} = \frac{1}{N_W} \sum_{i=1}^{N} | w_{ij} | \tag{11}$$

$$\underline{w}_{ij} = | w_{ij} | / M_{ABS} \tag{12}$$

$$\mathrm{Var}[\underline{w}_{ij}] = \frac{1}{N_W} \sum_{i=1}^{N} \sum_{j=1}^{N} (\underline{w}_{ij} - w_D)^2 \tag{13}$$

N_W is the number of the weights, and w_D is mean of $\underline{w}_{ij}$. Since large weights mean high transfer gains, noise performance is inversely proportional to $\mathrm{Var}[\underline{w}_{ij}]$. Therefore, the threshold T should be optimized by minimizing the variance.

4.3 Ratio of Step-size and Threshold

The variance $\mathrm{Var}[\underline{w}_{ij}]$ is highly dependent on a ratio of $\eta(n)$ and T.

$$r(n) = \eta(n)/T \tag{14}$$

In the training process, if the input does not satisfy Eq.(7), then the weights come from the other units on the pattern are increased by $\eta(n)$. Thus, letting the number of the units on the pattern be N_P, the unit input is increased by $\eta(n)(N_P-1)$. If this change is large compared with T, then the input easily exceeds T. This over-adjusting causes a large variance. Therefore, the condition required for r(n) becomes as follows:

$$\eta(n)(N_P-1) < T, \tag{15}$$

$$r(n) = \eta(n)/T < 1/(N_P-1) \tag{16}$$

4.4 Simulation Results

In order to justify the above discussions, simulation was carried out. $\eta(n)$ is chosen to 0.5. r(n) is controlled by changing T. The following variance is evaluated.

$$r_{wij}(n) = w_{ij}(n)/w_{ij}(n-1) \tag{17}$$

$$\mathrm{Var}[r_{wij}] = \frac{1}{N_W} \sum_{i=1}^{N} \sum_{j=1}^{N} (r_{wij}(n) - r_D)^2 \tag{18}$$

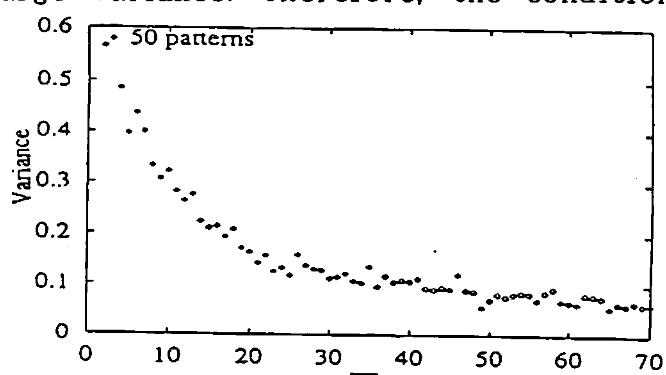

Fig.2 Relation between the variance defined by Eq.(18) and threshold T.

r_D is mean of $r_{wij}(n)$. For large r(n), Var[r_{wij}] is also large. It can be decreased as r(n) decreases. After r(n) reaches to some value, Var[r_{wij}] will saturate. This means the obtained weights are scaled versions of the previous.

Figure 2 shows simulation results using 50 patterns. From the starting point to T=30, that is r(n)=0.0167, the variance can decrease. Therefore, r(n) is not optimum in this interval. After this point, the variance is almost saturated. Therefore, the obtained weights are scaled versions of the weights obtained at the turning point r(n)=0.0167. Noise performance is not improved any more.

Although a smaller r(n) than 0.0167 can guarantee a little smaller variance, it requires a large number of computations. Taking both noise performance and computational load into account, the turning point gives the best selection.

In this simulation, N_P=32, then change of the unit input at one iteration is $\eta(n)(N-1)$=15.5 for T=30. Actually, the change is smaller than 15.5 due to interference by the other patterns. This relation was held for the different number of patterns. Therefore, the following condition on the step-size $\eta(n)$ and the threshold T can be obtained.

$$\eta(n)(N_P-1) \fallingdotseq 0.5T \quad \text{or} \quad \eta(n)/T \fallingdotseq 0.5/(N_P-1) \qquad (19)$$

V NOISE SENSITIVITY AND η (n)/T

Noise performance is investigated based on the ratio η (n)/T. Conditions of the simulation are the same as in Sec. IV. Random noise is added. The state of 5～20% units, randomly selected, are changed. Figure 3 shows the simulation results. As expected in the previous section, association rates increase until about T=30, that is r(n)=0.0167. After that, they tend to saturate. Thus, noise performance also saturate after this turning point.

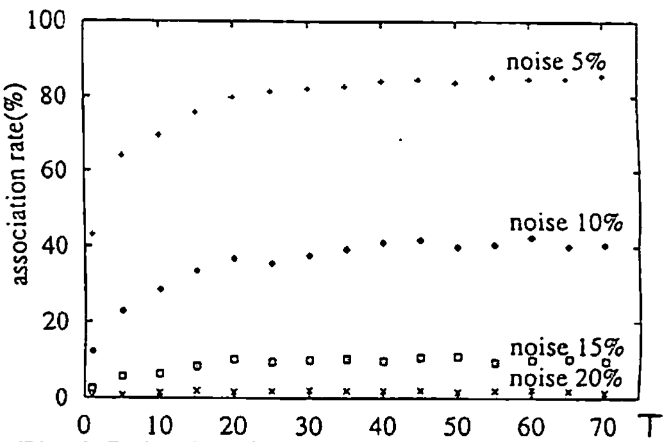

Fig.3 Relation between association rates for noisy patterns and threshold T. The number of patterns is 50.

VI CONCLUSIONS

The memory capacity bound for random patterns and the threshold optimization method have been proposed for the asymmetrical recurrent NN with variable hysteresis threshold. Efficiency of the proposed has been justified through numerically statistical simulation.

REFERENCES

[1]T.Kohonen, Self-Organization and Associative Memory, 3rd Ed., Springer-Verlag 1989.

[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 system ～",Proc.Natl.Sci.USA,vol.79,pp.2554-2558, 1982.

[5]D.Amit et al,"Storing infinite number of patterns～",Phys.Rev.Lett.,pp.1530-1533,1985.

[6]S.Amari and K.Mginu,"Statistical neurodynamics of～",Neural Networks,vol.1,pp.63-73,1988.

[7]N.Mitsutani and K,Nakayama, IEICE Japan Rep. Tech. Meeting, vol.NC90-89, pp.125-130, March 1991.

[8]K.Nakayama and N.Mitsutani,"An adaptive hysteresis～", Proc. IJCNN'91 Seattle, p.Ⅱ A-914, 1991.