2005 International Symposium on Nonlinear
Theory and its Applications (NOLTA2005)
Bruges, Belgium, October 18-21, 2005

Entropy based Association Model

Masahiro NAKAGAWA

Nagaoka University of Technology, Kamitomioka 1603-1, Nagaoka,
Niigata 940-2188, JAPAN
e-mail: masanaka@vos.nagaokaut.ac.jp

Abstract In this paper, an entropy based associative memory model will be proposed and applied to memory retrievals with an orthogonal learning model to compare with the conventional model based on the quadratic Lyapunov functional to be minimized during the retrieval process. In the present approach, the updating dynamics will be constructed on the basis of the entropy minimization strategy which may be reduced asymptotically to the above-mentioned conventional dynamics as a special case. From numerical results, it will be found that the presently proposed novel approach realizes twice of the memory capacity in comparison with the quadratic Lyapunov functional approach,e.g. associatron.

§1. Introduction

During the past quarter century, the numerous autoassociative models have been extensively investigated on the basis of the autocorrelation dynamics. Since the proposals of the retrieval models by Anderson,[1] Kohonen, [2] and Nakano, [3] some works related to such an autoassociation model of the inter-connected neurons through an autocorrelation matrix were theoretically analyzed by Amari, [4] Amit *et al* .[5] and Gardner.[6] So far it has been well appreciated that the storage capacity of the autocorrelation model, or the number of stored pattern vectors, L , to be completely associated vs the number of neurons N, which is called the relative storage capacity or loading rate and denoted as $\alpha_c=L/N$, is estimated as $\alpha_c \sim 0.14$ at most for the autocorrelation learning model with the activation function as the signum one (sgn(x) for the abbreviation).[7,8]

In contrast to the above-mentioned models with monotonous activation functions, the neuro-dynamics with a nonmonotonous mapping was recently proposed by Morita,[9] Yanai and Amari,[10] Shiino and Fukai.[11] They reported that the nonmonotonous mapping in a neuro-dynamics possesses a remarkable advantage in the storage capacity, $\alpha_c \sim 0.27$, superior than the conventional association models with monotonous mappings, e.g. the signum or sigmoidal function.

In the present paper, we shall propose a novel approach based on the entropy defined in terms of the overlaps, which are defined by the innerproducts between the state vector and the embedded vectors.

§2. Theory

Let us consider an associative model with the embedded binary vector $e^{(r)}_i=\pm 1$ $(1\le i\le N, 1\le r\le L)$, where N and L are the number of neurons and the number of embedded vectors. The states of the neural network are characterised in terms of the output vector s_i $(1\le i\le N)$ and the internal states σ_i $(1\le i\le N)$ which are related each other in terms of

$$s_i=f\left(\sigma_i\right) \qquad (1\le i\le N) , \tag{1}$$

where $f(\bullet)$ is the activation function of the neuron.

Then we introduce the following entropy which is to be related to the overlaps;

$$I=-\sum_{r=1}^{L}\left|m^{(r)}\right|\log\left|m^{(r)}\right|, \tag{2}$$

where the overlaps $m^{(r)}$ (r=1,2,...,L) are defined by

$$m^{(r)}=\sum_{i=1}^{N}e^{\dagger(r)}_i s_i \quad ; \tag{3}$$

here the covariant vector $e^{\dagger(r)}_i$ is defined in terms of the following orthogonal relation,

$$\sum_{i=1}^{N}e^{\dagger(r)}_i e^{(s)}_i=\delta_{rs} \qquad (1\le r,s\le L) \quad , \tag{4}$$

$$e^{\dagger(r)}_i=\sum_{r'=1}^{L}a_{rr'}e^{(r')}_i \quad , \tag{4a}$$

$$a_{rr'}=(\omega^{-1})_{rr'} \quad , \tag{4b}$$

$$\omega_{rr'}=\left(\sum_{i=1}^{N}e^{(r)}_i e^{(r')}_i\right) \quad . \tag{4c}$$

The entropy defined by eq.(2) can be minimized by the following condition

$$\left|m^{(r)}\right|=\delta_{rs} \qquad (1\le r,s\le L), \tag{5a}$$

and

$$\sum_{r=1}^{L}\left|m^{(r)}\right|=1 . \tag{5b}$$

That is, regarding $\left|m^{(r)}\right|$ ($1\le r\le L$) as the probability distribution in eq.(2), a target pattern may be retrieved by minimizing the entropy I with respect to $m^{(r)}$ or the state vector s_i to achieve the retrieval of a target pattern in which the eqs.(5a) and (5b) are to be satisfied. Therefore the entropy function may be considered to be a functional to be minimized during the retrieval process of the auto-association model instead of the conventional quadratic energy functional, E, i.e.

$$E=-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}w_{ij}s^{\dagger}_{i}s_{j} , \tag{6a}$$

where $s^{\dagger}_{i}$ is the covariant vector defined by

$$s^{\dagger}_{i}=\sum_{r=1}^{L}\sum_{j=1}^{N} e^{\dagger(r)}_{i}e^{\dagger(r)}_{j}s_{j} , \tag{6b}$$

and the connection matrix w_{ij} is defined in terms of

$$w_{ij}=\sum_{r=1}^{L}e^{(r)}_{i}e^{\dagger(r)}_{j} . \tag{6c}$$

According to the steepest descent approach in the discrete time model, the updating rule of the internal states σ_i ($1\le i\le N$) may be defined by

$$\sigma_i(t+1)=-\eta\frac{\partial I}{\partial s^{\dagger}_{i}} \quad (1\le i\le N), \tag{7}$$

where η is a positive coefficient to realize the entropy descent approach. Substituting eqs.(2) and (3) into eq.(7) and noting the following relation with aid of eq.(6b),

$$m^{(r)}=\sum_{i=1}^{N}e^{\dagger(r)}_{i}s_{i}=\sum_{i=1}^{N}e^{(r)}_{i}s^{\dagger}_{i} \tag{8}$$

one may readily derive the following relation.

$$\begin{aligned}\sigma_i(t+1)&=-\eta\frac{\partial I}{\partial s^{\dagger}_{i}}=+\eta\frac{\partial}{\partial s^{\dagger}_{i}}\sum_{r=1}^{L}\left|m^{(r)}\right|\log\left\{\left|m^{(r)}\right|\right\}\\&=\eta\frac{\partial}{\partial s^{\dagger}_{i}}\sum_{r=1}^{L}\left|\sum_{j=1}^{N}e^{(r)}_{j}s^{\dagger}_{j}(t)\right|\log\left\{\left|\sum_{k=1}^{N}e^{(r)}_{k}s^{\dagger}_{k}(t)\right|\right\}\\&=\eta\sum_{r=1}^{L}e^{(r)}_{i}\mathrm{sgn}\left\{\sum_{j=1}^{N}e^{(r)}_{j}s^{\dagger}_{j}(t)\right\}\left[1+\log\left\{\left|\sum_{k=1}^{N}e^{(r)}_{k}s^{\dagger}_{k}(t)\right|\right\}\right]\\&=\eta\sum_{r=1}^{L}e^{(r)}_{i}\mathrm{sgn}\left\{\sum_{j=1}^{N}e^{\dagger(r)}_{j}s_{j}(t)\right\}\left[1+\log\left\{\left|\sum_{k=1}^{N}e^{\dagger(r)}_{k}s_{k}(t)\right|\right\}\right] .\end{aligned} \tag{9}$$

Generalizing somewhat the above dynamics, we propose the following updating rule for the internal states

$$\begin{aligned}\sigma_i(t+1)&=\eta\sum_{r=1}^{L}e^{(r)}_{i}\mathrm{sgn}\left\{\sum_{j=1}^{N}e^{\dagger(r)}_{j}s_{j}(t)\right\}\frac{1}{\alpha}\left[\alpha+\log\left\{(1-\alpha)+\alpha\left|\sum_{k=1}^{N}e^{\dagger(r)}_{k}s_{k}(t)\right|\right\}\right]\\&=\eta\sum_{r=1}^{L}e^{(r)}_{i}\mathrm{sgn}\left\{m^{(r)}(t)\right\}\frac{1}{\alpha}\left[\alpha+\log\left\{(1-\alpha)+\alpha\left|m^{(r)}(t)\right|\right\}\right] .\end{aligned} \tag{10}$$

In the limit of $\alpha\to 0$, the above dynamics will be reduced to the autocorrelation dynamics.

$$\begin{aligned}\sigma_i(t+1)&=-\eta\lim_{\alpha\to 0}\sum_{r=1}^{L}e^{(r)}_{i}\mathrm{sgn}\left\{m^{(r)}(t)\right\}\frac{1}{\alpha}\left[\alpha+\log\left\{(1-\alpha)+\alpha\left|m^{(r)}(t)\right|\right\}\right]\\&=\eta\sum_{r=1}^{L}e^{(r)}_{i}m^{(r)}(t)=-\eta\sum_{r=1}^{L}e^{(r)}_{i}\sum_{j=1}^{N}e^{\dagger(r)}_{j}s_{j}(t)\\&=\eta\sum_{j=1}^{N}w_{ij}s_{j}(t) .\end{aligned} \tag{11}$$

On the other hand, eq.(10) results in eq.(9) for $\alpha \to 1$. Therefore one may control the dynamics between the autocorrelation ($\alpha \to 0$) and the entropy based approach ($\alpha \to 1$).

§3. Numerical Results

The embedded vectors are set to the binary random vectors as follows.

$$e^{(r)}_i = \mathrm{sgn}(z^{(r)}_i) \qquad (1 \le r \le L) \quad , \tag{12}$$

where $z^{(r)}_i$ $(1 \le i \le N,\ 1 \le r \le L)$ are the zero-mean pseudo-random numbers between -1 and +1. For simplicity, the activation function , eq.(1), is set to

$$s_i = f(\sigma_i) = \mathrm{sgn}(\sigma_i) \ , \tag{13}$$

where $\mathrm{sgn}(\bullet)$ denotes the signum function defined by

$$\mathrm{sgn}(x) = \begin{cases} -1 & (x<0) \\ 0 & (x=0) \\ +1 & (x>0) \end{cases} \quad . \tag{14}$$

The initial vector $s_i(0)$ $(1 \le i \le N)$ is set to

$$s_i(0) = \begin{cases} -e^{(s)}_i & (1 \le i \le H_d) \\ +e^{(s)}_i & (H_d + 1 \le i \le N) \end{cases} \ , \tag{15}$$

where $e^{(s)}_i$ is a target pattern to be retrieved and H_d is the Hamming distance between the initial vector $s_i(0)$ and the target vector $e^{(s)}_i$. The retrieval is succeeded if an overlap defined by eq.(3) results in 1 for $t \ge 1$, in which the system may be in a steady state such that

$$s_i(t+1) = s_i(t) \quad , \tag{17a}$$

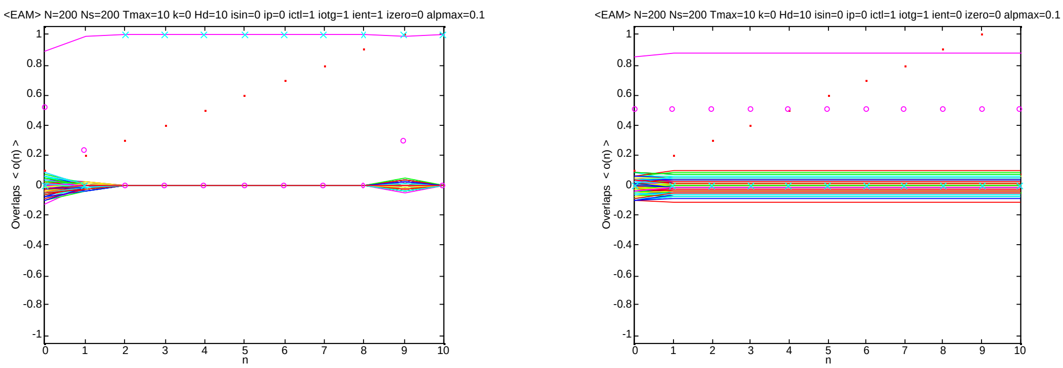
$$\sigma_i(t+1) = \sigma_i(t) \quad . \tag{17b}$$

To see the retrieval ability of the present model, the success rate S_r is defined as the rate of the success for 1000 trials with the different embedded vector sets $e^{(r)}_i$ $(1 \le i \le N,\ 1 \le r \le L)$. To control from the autocorrelation dynamics after the initial state $(t \sim 1)$ to the entropy based dynamics $(t \sim T_{max})$, the parameter α in eq.(10) was simply controlled by

$$\alpha = \frac{t}{T_{max}} \alpha_{max} \qquad (0 \le t \le T_{max}) \quad , \tag{18}$$

where T_{max} and α_{max} are the maximum values of the iterations of the updating according to eq.(10) and α, respectively.

Choosing N=200, η=1, H_d=10, T_{max}=10, L/N=0.5 and α_{max}=0.1, we first present an example of the dynamics of the overlaps in Fig.1(a) (Entropy based approach) and (b) (Associatron). Therein the cross symbols(×) and the open circles(o) represent the success of retrievals, in which eqs.(5a) and (5b) are satisfied, and the entropy defined by eq.(2), respectively, for a retrieval process. In addition the time dependence of the parameter α/α_{max} defined by eq.(18) are depicted as dots (.). In Fig.1(a) after a transient state, it is confirmed that the complete association corresponding to eqs.(5a) and (5b) can be achieved. On the other hand, in Fig.1(b), a trapping at a local minimum is found to be inevitable for L/N=0.5, in which eqs.(5a) and (5b) can not be achieved. From this results one may apparently confirm the advantage of our approach.



(a) The time dependence of overlaps of the present entropy based model defined by eq.(10).

(b) The time dependence of overlaps of the associatron defined by eq.(11).

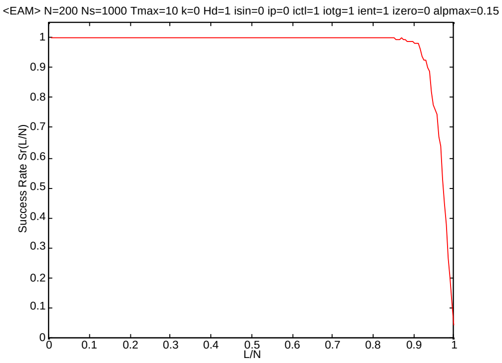
Fig.1 The time dependence of the overlaps defined by eq.(3).

Then we present the dependence of the success rate S_r on the loading rate L/N are depicted in Fig.2(a) and (b) for

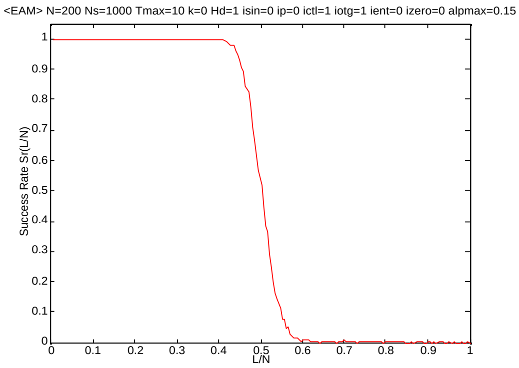
H_d/N=1/200. From Fig.2(a), one may confirm the larger memory capacity of the presently proposed model defined by eq.(10) (Entropy based approach) in comparison with the conventional model defined by eq.(11) (quadratic Lyapunov functional approach). Therefore the presently proposed nonlinear dynamics based on the entropy functional to be minimized has a great advantage beyond the conventional one based on eqs.(6a) and (11). The depression of the success rate at L/N~1 in Fig.2(a) may be considered to result from the fact such that

$$w_{ij}=\sum_{r=1}^{L} e^{(r)}_i e^{\dagger(r)}_j=\delta_{ij} \quad (L=N) . \qquad (19)$$

For comparison, the corresponding result of the autocorrelation model with , i.e. eq.(11), is shown in Fig.2(b). Comparing Figs. 2(a) and 2(b), it is found that the present approach may achieve twice of the memory capacity beyond the conventional strategy based on the quadratic Lyapunov functional to be minimized.



(a) The dependence of the success rate on the loading rate α=L/N of the present entropy based model defined by eq.(10).



(b) The dependence of the success rate on the loading rate α=L/N of the associatron defined by eq.(11).

Fig.2 The characteristics of the memory retrievals of autoassociation models.

§4 Concluding Remarks

In the present paper, we have proposed an entropy based association model instead of the conventional autocorrelation dynamics. From numerical results, it was found that the large memory capacity may be achieved on the basis of the entropy approach.

As a future problem, it seems to be worthwhile to involve a chaotic dynamics in the present model introducing a periodic activation function such as sinusoidal one and to extend the autocorrelation model replacing $e^{\dagger(r)}_i$ by $e^{(r)}_i/N$ in the present approach, in which the connection matrix w_{ij} and the overlaps $m^{(r)}$ read

$$w_{ij}=\frac{1}{N}\sum_{r=1}^{L} e^{(r)}_i e^{(r)}_j , \qquad (20)$$

and

$$m^{(r)}(t)=\frac{1}{N}\sum_{i=1}^{N} e^{(r)}_i s_i(t) , \qquad (21)$$

corresponding to eq.(6c) and eq.(3), respectively. The entropy based approach with eq.(20), i.e. autocorrelation dynamics, is now in progress in the relation with chaos dynamics[12] and will be reported in the near future.

References

[1] J. A.Anderson (1972):Math. Biosci.**14**,197.
[2] T.Kohonen (1972): *Correlation matrix memories* , IEEE Trans. **C-21**,353.
[3] K.Nakano: IEEE Trans. **SMC-2** (1972)380.
[4] S .Amari : Biol. Cybern.**26**(1977) 175.
[5] D. J. Amit , H. Gutfreund, and H. Sompolinsky : Phys. Rev. Lett.**55** (1985) 1530.
[6] E. Gardner : J. Phys. **A19**(1986)L1047.
[7] R. J. McEliece , R.J. Posner, E.R. Rodemich and S.S.Venkatesh : IEEE Trans. on Information Theory, **IT-33**(1987)461.
[8] S.Amari and K.Maginu : Neural Networks **1**(1988)63.
[9] M.Morita : Neural Networks, **6**(1993) 115.
[10] Hiro-F. Yanai and S.Amari : Proc. of ICNN'93, San Francisco (1993)1385.
[11] M.Shiino and T.Fukai : Phys. Rev. **E48** (1993) 867.
[12] M.Nakagawa : *Chaos and Fractals in Engineering*, World Scientific Inc.(1999) 960pp.