An Application of a Function of Pattern Separation to an Associative Memory System

Nobuhiko IKEDA* and Toyoshi TORIOKA**
(Received July 9, 1986)

Abstract

There are some studies which applied a function of pattern separation to technical systems. The effect of its function on an associative memory system, however, has not yet been clarified sufficiently.

In the present paper, a new associative memory model having the function of pattern separation is given, and its ability of recall is derived. The ability is compared to that of the traditional model. As the result, it is shown that the model is superior to the traditional model with respect to the recalling ability. Further, it is also cleaned up that the ability of recall in the model is enhanced by the function of pattern separation.

1. Introduction

There are many studies on an associative memory system (Nakano et al[1]., 1970, Uesaka et al[2]., 1972, Kohonen[3], 1972, Kosugi et al[4]., 1975). As the result, two favorable properties of key patterns for the associative memory are pointed out. One is the property that the firing rate of the key pattern is constant. Another is the property that the overlapping rate among the key patterns is small.

Marr (1969)[5] regarded the cerebellar cortex as a pattern separator, and gave a theory of pattern separation on the cerebellar cortex, using the concept of "Codon", viz. the idea of overlapping rate. Wigström (1973)[6] and Nakano et al. (1974)[7] have proposed a model of associative memory which adopted the idea of pattern separation described above. However, they have not discussed about the effect of pattern separation on the recalling ability in the associative memory in detail.

Recently, Torioka (1978, 1979, 1980)[8]~[10], one of authors, have modeled the cerebellar cortex as a two-layer neural network with an inhibitory mechanism and given an extended theory which includes Marr's theory. He also pointed out that the degree of pattern separation has a favorable property for associative memory and pattern recognition.

In the present paper, we propose a model of associative memory using the property of the degree of pattern separation given by Torioka, and derive an ability of recall in the model. We also compare the recalling ability in the model to that in the traditional model.

As the result, it is shown that our model is superior to the traditional model with respect to the recalling ability. It is also cleaned up that the ability of recall in the model is enhanced by the function of pattern separation.

* Tokuyama Technical College
** Technical College, Yamaguchi University

2. Two-layer Random Neural Net and Notations

Figure 1 shows a model of the cerebellar cortex. It is a two-layer neural net consisting of threshold elements with an inhibitory mechanism. In this model, the first and second layers correspond to the layer of mossy fibers and granule cells in the cerebellar cortex, respectively, and are connected randomly. The numbers of elements on the first and second layers, n and N, are large enough. Each element on the second layer recieves R (integer) excitatory connections of weight 1 from the first layer. The i-th element of the second layer has a threshould value h_i. An inhibitory mechanism G corresponds to the Golgi cell. It recieves excitatory inputs from only the layer of mossy fibers and sends an appropriate inhibitory output θ to each element of second layer through an inhibitory connection. At this time, it controls the overall firing rate of second layer at a fixed value and enhances the function of pattern separation on the second layer.

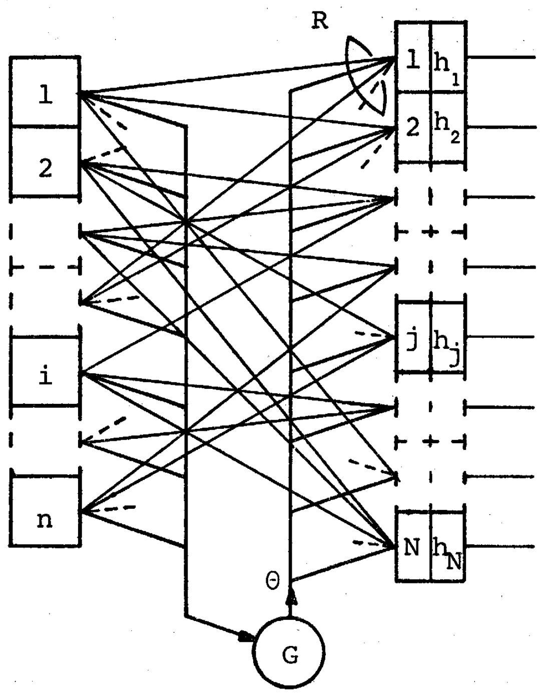

Fig. 1 Two-layer random neural net

We denote two arbitary n-dimensional binary patterns added to the first layer by y_1 and y_2. we give the firing rates of y_1 and y_2, common firing rate between y_1 and y_2 and overlapping rate on y_1 and y_2 by

$$\phi_1 = \frac{1}{n} \sum_{i=1}^{n} y_{1i} , \qquad \phi_2 = \frac{1}{n} \sum_{i=1}^{n} y_{2i} , \tag{1}$$

$$\psi_{12} = \frac{1}{n} \sum_{i=1}^{n} y_{1i} \cdot y_{2i} , \tag{2}$$

and

$$\tau_{12}=\frac{\psi_{12}}{\sqrt{\phi_1\cdot\phi_2}} \tag{3}$$

respectively, where y_{1i} and y_{2i} are i-th elements of y_1 and y_2. We denote two N-dimensional binary patterns corresponding to y_1 and y_2 by Y_1 and Y_2. We also give the firing rates of Y_1 and Y_2, common firing rate between Y_1 and Y_2 and overlapping rate on Y_1 and Y_2 by

$$\Phi_1=\frac{1}{N}\sum_{i=1}^{N}Y_{1i}, \qquad \Phi_2=\frac{1}{N}\sum_{i=1}^{N}Y_{2i}, \tag{4}$$

$$\Psi_{12}=\frac{1}{N}\sum_{i=1}^{N}Y_{1i}\cdot Y_{2i}\ , \tag{5}$$

and

$$T_{12}=\frac{\Psi_{12}}{\sqrt{\Phi_1\cdot\Phi_2}}\ , \tag{6}$$

respectively, where Y_{1i} and Y_{2i} are i-th elements of Y_1 and Y_2.

The degree of pattern separation is defind as

$$\Delta_{12}=\frac{T_{12}}{t_{12}}\ . \tag{7}$$

We can define that the pattern y_1 and y_2 are separated in the sense of overlapping rate, when $\Delta_{12}<1$.

The distribution of the threshold value h_i is given by

$$f\ (h)=\frac{1}{\sqrt{2\pi}\sigma}\ exp\ \left(-\ \frac{(h-\bar{h})^2}{2\sigma^2}\right), \tag{8}$$

where $\bar{h}$ is the mean, and σ is the standard deviation.

3. The Degree of Pattern Separation and Its property

We give the degree of pattern separation. We show that the degree of pattern separation has a property favorable for the association processing. The degree of pattern sparation has the above property when the standard deviation σ is small. Then, we assume that $3\sigma<0.5$. Let us also assume that the firing rate of the second layer is controlled at a fixed value ϕ, even if any pattern is added to the first layer. The control is practically performed only through the shift of the mean value $\bar{h}$ by an inhibitory output θ. At that time, the form of the distribution function f(h) is not changed. For the later explanation, let us denote the means $\bar{h}_1$ and $\bar{h}_2$ of $\bar{h}$ shifted as

$$\bar{h}_1=\bar{h}+\theta\ (\phi_1),$$

and

$$\bar{h}_2 = \bar{h} + \theta\ (\phi_2),$$

where $\theta(\phi_2)$ and $\theta(\phi_2)$ are the inhibitory outputs corresponding to y_1 and y_2 on the first layer, respectively. [For the details see Torioka (1978).]

3.1. The Degree of Pattern Separation

Let y_1 be added to the first layer. The firing rate Φ becomes

$$\Phi = a\ V_i\ (\phi_1)\ + \sum_{p=i+1}^{R} V_p(\phi_1) \tag{9}$$

with

$$a = \int_{-\infty}^{(i-h_1)/\sigma} f\ (h)dh,$$

where V_p is the probability that a neural element on the second layer recieves p excited inputs from y_1. The a is the rate that the elements on the second layer of recieving i excited inputs from the first layer fire, i coincides with the integer threshold which is the nearest to $\bar{h}_1$ and $0 \leqq i \leqq R$. When y_2 is added, the firing rate Φ becomes

$$\Phi = b\ V_j\ (\phi_2)\ + \sum_{q=j+1}^{R} V_q(\phi_2) \tag{10}$$

with

$$b = \int_{-\infty}^{(j-h_2)/\sigma} f\ (h)dh,$$

where V_q is the probability that a neural element on the second layer recieves q excited inputs from y_2. The b is the rate that the elements on the second layer of recieving j excited inputs from the first layer fire, j coincides with the integer threshold which is the nearest to $\bar{h}_2$, and $0 \leqq j \leqq R$. The common firing rate Ψ_{12} is obtained as

$$\Psi_{12} = \mathrm{c}\ W_{ij} + \mathrm{a} \sum_{q=j+1}^{R} W_{iq} + \mathrm{b} \sum_{p=i+1}^{R} W_{pj} + \sum_{p=i+1}^{R} \sum_{q=j+1}^{R} W_{pq}. \tag{11}$$

with

$\mathrm{c} = \min(\mathrm{a}, \mathrm{b})$.

In (11), W_{pq} is the probability that a neural element on the second layer recieves p excited inputs from Y_1 and q excited inputs from Y_2 and is represented by

$$W_{pq} = \sum_{r=max\,(0,\,p+q-R)}^{min\,(p,\,q)} \frac{R!}{(p-r)!\,(p-r)!\,r!\,(R-p-q+r)!} \cdot (\phi_1 - \psi_{12})^{p-r} \cdot (\phi_2 - \psi_{12})^{q-r} \cdot \psi_{12}^{r} \cdot (1 - \phi_1 - \phi_2 - \psi_{12})^{R-p-q+r}, \tag{12}$$

where min (a, b) means the smaller value of a and b, max (a, b) means the larger value of a and b.

Equation (11) is the probability that the elements of recieving i excited inputs from y_1 and j excited inputs from y_2 fire in common for both patterns. In (9), (10) and (11), note that infinitestimal terms are neglected. If we put $a \leqq b$, then the degree of pattern separation becomes

$$\Delta_{12} = \frac{1}{\tau_{12}} \left(A + \frac{B}{\Phi}\right) \tag{13}$$

with

$$A = \frac{\sum_{q=j+1}^{R} W_{iq}}{V_i(\phi_1)} + \frac{\sum_{p=i}^{R} W_{pi}}{V_j(\phi_2)}$$

$$B = \sum_{p=i+1}^{R} \sum_{q=j+1}^{R} W_{pq} - \frac{\left(\sum_{p=i+1}^{R} V_p(\phi_1)\right) \cdot \left(\sum_{q=j+1}^{R} W_{iq}\right)}{V_j(\phi_1)} - \frac{\left(\sum_{q=j+1}^{R} V_q(\phi_2)\right) \cdot \left(\sum_{p=i+1}^{R} W_{pj}\right)}{V_j(\phi_2)} \tag{14}$$

[For the details see Torioka (1978, 1979).]

3.2. The Property Favorable for the Association Processing

There are two properties favorable for the key patterns in the associative memory. One is the property that the overlapping rate among the key patterns is small. Another is the property that the firing rate of the key pattern is constant.

The two-layer neural net has a function which transforms the patterns on the first layer into the patterns having the above property. The above fact is cleaned up by giving an example. We consider the case that $\phi_1 = \phi_2 = \phi$ and R=2. In this case, Δ_{12} becomes as

$$\Delta_{12} = \begin{cases} \tau_{12} & (0 < \Phi \leq \phi^2), \\ \dfrac{1}{\tau_{12}} \cdot \left[\dfrac{\phi - 2\phi\,\psi_{12} + \psi_{12}}{(1-\phi)\phi} - \dfrac{1}{\Phi} \cdot \dfrac{(\phi - \psi_{12})\{\phi\,(\phi - \psi_{12}) + \psi_{12}\}}{(1-\phi)} \right] & (\phi^2 \leqq \Phi \leqq 2\phi - \phi^2), \\ \dfrac{1}{\tau_{12}} \cdot \left[\dfrac{(1-\psi_{12})^2 - 2(\phi - \psi_{12})^2}{(1-\phi)} + \dfrac{1}{\Phi} \cdot \dfrac{(\phi - \psi_{12}) \cdot (2 - 3\phi + \psi_{12})}{(1-\phi)^2} \right] & (2\phi - \phi^2 \leqq \Phi < 1.0). \end{cases} \tag{15}$$

Figure 2 shows the relation between Φ and Δ_{12} plotted for τ_{12} as parameter, where $\phi = 0.2$. The figure shows that Δ_{12} is constant under τ_{12}=const. and $\Delta_{12} < 1$ when Φ lies in the range $(0, \phi^2]$. The figure also shows that the less τ_{12} becomes, the larger Δ_{12} becomes and vice versa. Figure 3 shows the relation between τ_{12} and T_{12}. As known from the

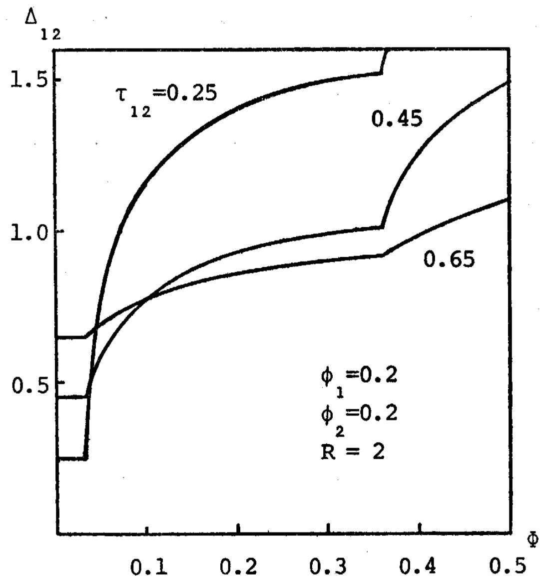

Fig. 2 Relation between Φ and Δ_{12} plotted for τ_{12} as parameter, where $\phi_1=0.2$, R=2 and $\sigma=0.1$

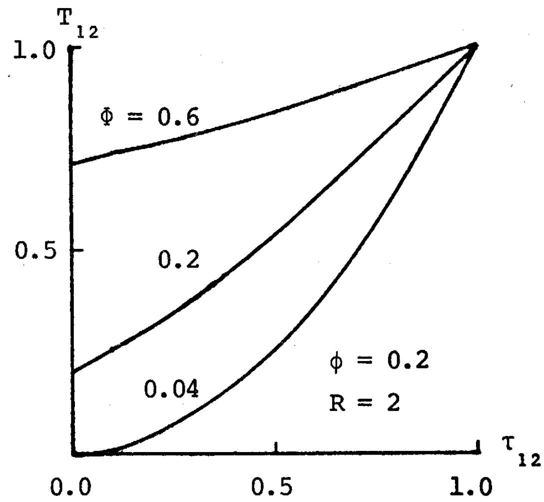

Fig. 3 Relation between τ_{12} and T_{12} plotted for Φ as parameter, where $\phi=0.2$, R =2 and $\sigma=0.1$

figure, in the sense of mean, the patterns y_1 and y_2 having the overlapping rate τ_{12} are transformed into the patterns Y_1 and Y_2 having the smaller overlapping rate by the neural net, in the range $(0, \phi^2]$. The firing rate of the transformed patterns is also controlled at a fixed value Φ. Consequently, the transformed patterns Y_1 and Y_2 on the second layer have two properties favorable for the key patterns in the associative memory. If Φ exceeds ϕ^2, $\Delta_{12}>1$ and the relation between large and small Δ_{12} reverses itself at a certain firing rate, thus the useful properties are lost. These facts are also obtained in a general case that the connection between the first and second layers is R.

4. Traditional Associative Memory System and Its Recalling Ability

For the later consideration, we obtain a recalling ability of the traditional associative memory system.

4.1. Memorization and Recalling Process

Figure 4 shows the traditional associative memory system. Let X_k and y_k $(k=1,\dots,K)$ be the m-dimensional memorized pattern and the n-dimensional key pattern, respectively. Let us assume that the number of memorized patterns, K, is relatively large. The pattern X_k is associated with the key y_k, where the subscript k is the label of a particular pattern. These patterns are given by

$$X_k = (X_{k1}, X_{k2}, \cdots, X_{kj}, \cdots, X_{km}),$$
$$y_k = (y_{k1}, y_{k2}, \cdots, y_{kj}, \cdots, y_{kn}), \qquad (16)$$

where each element takes on value 1 or 0.

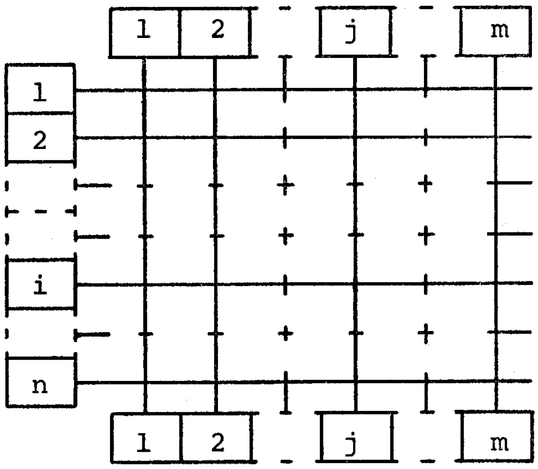

Fig. 4 Traditional associative memory system

The elements X_{kj} and y_{kj} are independent each other stochastically and random variables. These elements take on value 1 with probability,

$$P_r\ (X_{ki}=1)=\phi_x \tag{17}$$

and

$$P_r\ (y_{kj}=1)=\phi_y, \tag{18}$$

respectively.

In memorizing process, the pattern X_k is superimposed on the matrix memory [M] by the key y_k associated with it.

$$[M]=\sum_{k=1}^{K} X_k \cdot y_k^t, \tag{19}$$

where the superscript t means the transposition of the vector.

In recalling process, we operate $\tilde{y}_r$ as the input pattern to the matrix memory [M]. The pattern $\tilde{y}_r$ means a incomplete version of y_r associated with X_r. The incomplete version is given by adding a noise to y_r. The method for generating the noise is explained later on. The pattern $\tilde{y}_r$ is denoted by

$$\tilde{y}_r=(\tilde{y}_{r1}, \tilde{y}_{r2}, \cdots, \tilde{y}_{rj}, \cdots, \tilde{y}_{rn}). \tag{20}$$

Then, we obtain the output pattern $\hat{X}_r$ similar to X_r. If $\hat{X}_r=X_r$, the recall is perfect for X_r. The recall of particular X_r is made by the following operation

$$\hat{X}_r=u_h[[M]\cdot\tilde{y}_r], \tag{21}$$

where

$$u_h(a)=\begin{cases}1 & (a>h)\\ 0 & (a\leq h)\end{cases}$$

The j-th element $\widehat{X}_{rj}$ of $\widehat{X}_r$ is given by

$$\begin{aligned} \hat{X}_{rj} &= u_h[R_{rj}] \\ &= u_h[(\sum_{k=1}^{K} X_{kj}\, y_k^t)\ \tilde{y}_r] \\ &= u_h[X_{rj}\, y_r^t\ \tilde{y}_r + \sum_{\substack{k=1 \\ k\neq r}}^{k} X_{kj}\, y_k^t\ \tilde{y}_r]. \end{aligned} \tag{22}$$

We examine the probabity distribution of R_{rj} to derive the recalling ability. R_{rj} of (22) takes two different distributions according as the value of X_{rj} is 1 or 0. The threshold value λ is determined from the above distributions. [For the details see Shimura (1976).]

4.2. Case of Fixed Firing Rate

Let us assume that the firing rate of the key patterns in the noiseless case is ϕ_0. We derive the probability density function of R_{rj} under the condition $X_{rj}=1$. Let us also assume that $\tilde{y}_r$ in (21) is the pattern containing a noise s. The pattern is generated by means of changing the value of each element of y_r from 1(0) to 0(1) with probability s. At this time, the distribution of s becomes as

$$s \doteq N(E(s),\ V(s)) \tag{23}$$

with

$$E(s)=\bar{s}, \quad V(s)=\bar{s}\ (1-\bar{s}), \tag{24}$$

approximately, where N (μ, σ^2) means the density function of normal distribution with mean value μ and variance σ^2. Then, the distribution of the first term in (22) becomes as

$$X_{rj}\, y_r\, \tilde{y}_r \doteq N(\mu_\alpha, a_\alpha^2) \tag{25}$$

with

$$\mu_\alpha = n\phi_0\ (1-\bar{s}),$$

$$\sigma_\alpha^2 = n\phi_0\bar{s}\ (1-\bar{s}),$$

approximately. The distribution of the second term in (17) becomes as

$$\sum_{\substack{k=1 \\ k\neq r}}^{K} X_{kj}\, y_k^t\, \tilde{y}_r \doteq N(\mu_\beta, \sigma_\beta^2) \tag{26}$$

with

$$\mu_\beta = (w-1)L\ \{\phi_0 + \bar{s}\ (1-2\phi_0)\},$$

$$\sigma_\beta = (w-1)L\ [(1-\phi_0)\{\phi_0+\bar{s}\ (1-2\phi_0)\} + (W-1)\ \phi_0\ (2\phi_0^2-2\phi_0+1)\ \bar{s}\ (1-\bar{s})],$$

approximately, where $w=\phi_x K$ and $L=\phi_0 n$. From (25) and (26), the denisity function of R_{rj} becomes as

$$f_1(x) \simeq N(\mu_1, \sigma_1^2) \tag{27}$$

with

$$\mu_1 = L(1-\bar{s}) + (w-1)L\{\phi_0 + \bar{s}(1-2\phi_0)\},$$

$$\sigma_1^2 = L\bar{s}(1-\bar{s}) + (w-1)L[(1-\phi_0)\{\phi_0+\bar{s}(1-\phi_0)\} + (w-1)\phi_0(2\phi_0^2-2\phi_0+1)\bar{s}(1-\bar{s})],$$

approximately. Similarly, the density function of R_{rj} under $X_{rj}=0$ becomes as

$$f_0(x) \simeq N(\mu_0, \sigma_0^2) \tag{28}$$

with

$$\mu_0 = wL\{\phi_0+\bar{s}(1-\phi_0)\},$$

$$\sigma_0^2 = wL[(1-\phi_0)\{\phi_0+\bar{s}(1-2\phi_0)\}+W\phi_0(2\phi_0^2-2\phi_0+1)\bar{s}(1-\bar{s})],$$

approximately.

The threshold value λ is set so that the loss is minimum.[11] The threshold value is obtained from (27) and (28) which $\bar{s}=0$.

$$\lambda = \frac{\mu_1^2-\mu_0^2-2\sigma^2 \ln T_x}{2(\mu_1-\mu_0)}, \tag{29}$$

with

$$T_x = \phi_x/(1-\phi_x).$$

In (29), σ^2_1 becomes nearly equal to σ^2_0 when s is small and w is large. Then we put $\sigma^2_1=\sigma^2_0=\sigma^2$.

Let P_{f_1} and P_{f_0} be the probabilities that the elements of X_r, 1 and 0, are recalled correctly by the key $\tilde{y}_r$, respectively. From (27), (28) and (29), these become as

$$P_{f_1} = \frac{1}{\sqrt{2\pi}\sigma}\int_{\lambda}^{\infty} exp\left\{-\frac{(x-\mu_1)^2}{2\sigma^2}\right\}dx = F(\alpha_1) \tag{30}$$

and

$$P_{f_0} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\lambda} exp\left\{-\frac{(x-\mu_0)^2}{2\sigma^2}\right\}dx = F(\alpha_0) \tag{31}$$

with

$$\alpha_1 = \frac{\mu_1-\lambda}{\sigma}, \qquad \alpha_0 = \frac{\lambda-\mu_0}{\sigma},$$

$$F(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} exp\left(-\frac{t^2}{2}\right)dt,$$

respectively.

The recalling ability P_f is generally defined as the probability mean of P_{f1} and P_{f0} as

$$P_f=\phi_x\cdot F(\alpha_1)+(1-\phi_x)\cdot F(\alpha_0). \tag{32}$$

As we direct our attention to the property of the key pattern, we put $\phi_x=0.5$. P_f becomes as

$$P_f=F\left(\sqrt{\frac{n(1-\phi_0)}{2K}}\right) \tag{33}$$

when $s=0$.

4.3. Case of Variable Firing Rate

In this case, the distribution of the firing rate of key patterns becomes as

$$\phi_y\doteq N(E(\phi_y),\ V(\phi_y)) \tag{34}$$

with

$$E(\phi_y)=\phi_0,$$
$$V(\phi_y)=\phi_0(1-\phi_0),$$

approximately.

The denisity function of R_{rj} is derived in the same manner as the case of fixed firing rate. The mean values and variances corresponding to (27) and (28) are given by

$$\begin{aligned}
\mu_1&=L(1-\bar{s})+(w-1)\{\phi_0+\bar{s}(1-2\phi_0)\}\\
\sigma_1^2&=L(w-1)[(1-\phi_0)\{\phi_0+\bar{s}(1-2\phi_0)\}\\
&\quad+(w-1)\phi_0\{(2\phi_0^2-2\phi_0+1)\bar{s}(1-\bar{s})+(2\bar{s}^2-2\bar{s}+1)\phi_0(1-\phi_0)\}),\\
\mu_0&=wL\{\phi_0+\bar{s}(1-2\phi_0)\},\\
\sigma_0^2&=wL[(1-\phi_0)\{\phi_0+\bar{s}(1-2\phi_0)\}\\
&\quad+w\phi_0\{(2\phi_0^2-2\phi_0+1)\bar{s}(1-\bar{s})+(2\bar{s}^2-2\bar{s}+1)\phi_0(1-\phi_0)\}],
\end{aligned} \tag{35}$$

approximately. The threshold value is determined from (29) and (35) which $\bar{s}=0$. The recalling ability is determined from (30), (31) and (32). We denote it by P_v. The recalling ability P_v becomes

$$P_v=F\left(\frac{1}{K}\cdot\sqrt{\frac{n(1-\phi_0)}{\phi_0}}\right), \tag{36}$$

when $\phi_x=0.5$ and $s=0$.

As known from (33) and (36), the smaller ϕ_0 becomes, the higher the recalling ability becomes. From the comparison between (33) and (36), we can know that the model has the higher recalling ability in the case of fixed firing rate than that of variable firing rate when the number of the memorized patterns increases (Fig. 5).

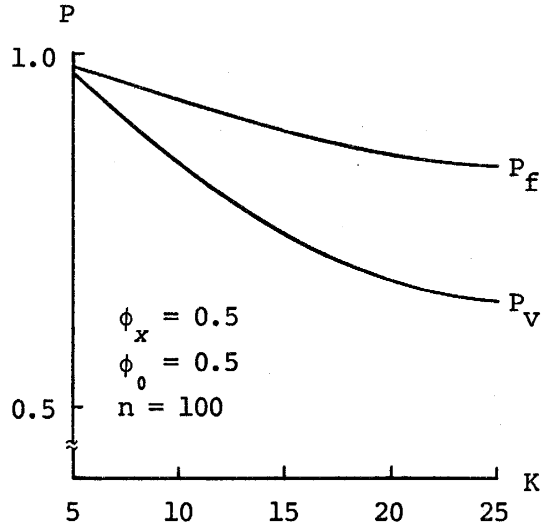

Fig. 5 Relations between K and P_f and between K and P_V, where $\phi_X=0.5$, $\phi_0=0.5$ and $n=100$

From the above consideration, we can conclude as follows. The two properties that the firing rate of the key patterns is constant and the overlapping rate among them is small are favorable for the associative memory. Therefore, we can surely apply the function of pattern separation to producing of the key pattern having the above properties.

5. Associative Memory having a Function of Pattern Separation

Figure 6 shows the model of associative memory having a function of pattern separation. In the model, the pattern separation system, viz. two-layer neural net, is directly coupled as preprocessing device of the traditional associative memory system. Then, the key pattern y_k on the first layer is transformed to the pattern Y_k on the second layer

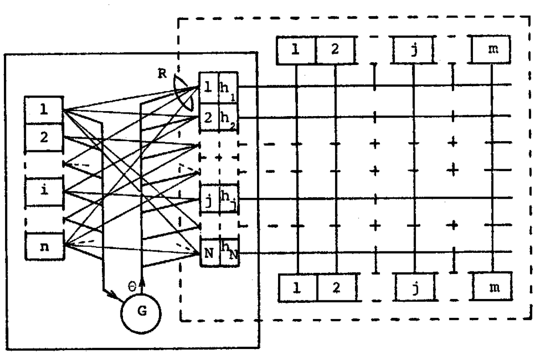

Fig. 6 Associative memory system with a function of pattern separation

through the neural net, where the firing rate of Y_k is controlled at Φ. The pattern Y_k is expressed by

$$Y_k = (Y_{k1}, Y_{k2}, \cdots, Y_{kj}, \cdots, Y_{kn}), \tag{37}$$

where $k=1,\ldots,K$ and each element takes on value 1 or 0. In this system, the pattern Y_k substantially works as the key pattern which is associated with X_k. The memorizing and the recalling processes are performed in the same manner as the traditional model. Those are described as

$$[M] = \sum_{k=1}^{K} X_k \cdot Y_k^t, \tag{38}$$

$$\hat{X}_r = u_h\,[[M] \cdot \tilde{Y}_r], \tag{39}$$

$$\hat{X}_{rj} = u_h\,[R_{rj}] = u_h\,[X_{rj} Y_r^t \tilde{Y}_r + \sum_{\substack{k=1 \\ k \neq r}}^{K} X_{kj} Y_k^t \cdot \tilde{Y}_r]. \tag{40}$$

In (39) and (40), $\tilde{Y}_r$ is the transformed pattern corresponding to $\tilde{y}_r$. It is expressed by

$$\tilde{Y}_r = (\tilde{Y}_{r1}, \tilde{Y}_{r2}, \cdots, \tilde{Y}_{rj} \cdots, \tilde{Y}_{rn}). \tag{41}$$

6. Effect of Pattern Separation on Recalling Ability

The density functions of R_{rj} in this case corresponding to (27) and (28) can be derived by using the overlapping rate among the key patterns. Let τ_r and Δ_r be the mean of overlapping rate between y_r and $\tilde{y}_r$ and the degree of pattern separation under those patterns, respectively. At this time, from (7), the mean of overlapping rate T_r between Y_r and $\tilde{Y}_r$ becomes as

$$T_r = \tau_r \cdot \Delta_Y. \tag{42}$$

Similarly, the mean of overlapping rate T_k between Y_k and $\tilde{Y}_r$ becomes as

$$T_k = \tau_k \cdot \Delta_k, \tag{43}$$

where τ_k is the mean of overlapping rate between y_k and $\tilde{y}_k$, and Δ_k is the degree of pattern sparation under the patterns y_k and $\tilde{y}_r$. We derive the equations of τ_r and τ_k concretely. In the sense of mean value, $\tilde{y}_{rj}$ takes on value 1 with probability $1-\bar{s}$ when $y_{rj}=1$. On the other hand, it takes on value 1 with probability $\bar{s}$ when $y_{rj}=0$. Therefore, the mean of firing rate ϕ of $\tilde{y}_r$ becomes

$$\tilde{\phi} = \phi_0 + \bar{s}\,(1-2\phi_0). \tag{44}$$

The mean of common firing rate ψ_r between y_r and $\tilde{y}_r$ becomes

$$\psi_r = \phi_0\,(1-\bar{s}). \tag{45}$$

From (3), the mean of overlapping rate τ_r becomes as

$$\tau_r = \frac{\psi_r}{\sqrt{\phi_0 \cdot \tilde{\phi}}} = \sqrt{\frac{\phi_0}{\phi_0 + \bar{s}\ (1-2\phi_0)}}\ (1-\bar{s})\,. \tag{46}$$

τ_k is also derived similarly. y_k and $\tilde{y}_r$ are independent one another in this case. Therefore, the mean of common firing rate ψ_k and the mean of overlapping rate τ_k under those patterns become as

$$\psi_k = \phi_0 \{\phi_0 + \bar{s}\ (1-2\phi_0)\}, \tag{47}$$

$$\tau_k = \sqrt{\frac{\psi_k}{\phi_0 \cdot \tilde{\phi}}} = \sqrt{\phi_0 \{\phi_0 + \bar{s}\ (1-2\phi_0)\}}\,. \tag{48}$$

T_r and T_k can be also regarded as the probabilities that $Y_{rj} \cdot \tilde{Y}_{rj}$ and $Y_{kj} \cdot \tilde{Y}_{rj}\,(k \neq r)$ take on value 1 under the condition $\tilde{Y}_{rj}=1$, respectively. By using τ_r and τ_k, the density function of R_{rj} under $X_{rj}=1$ is obtained as follows,

$$f_1(x) \simeq N(\mu_1, \sigma_1^2) \tag{49}$$

with

$$\mu_1 = M \cdot T_r + (w-1) M \cdot T_k,$$

$$\sigma_1^2 = M \cdot T_r\ (1-T_r) + (w-1) M \cdot T_k\ (1-T_k),$$

approximately, where $M=\Phi \cdot N$. Similarly, the density function of R_{rj} under $X_{rj}=0$ is given by

$$f_0(x) = N(\mu_0, \sigma_0^2) \tag{50}$$

with

$$\mu_0 = w \cdot M \cdot T_k,$$

$$\sigma_0^2 = w \cdot M \cdot T_k\ (1-T_k),$$

approximately. The threshold value is determined from (29), (49) and (50), in which $\bar{s}=0$. When $\bar{s}=0$, the mean of overlapping rates τ_r and τ_k become as

$$\tau_r = 1 \tag{51}$$

and

$$\tau_k = \phi_0. \tag{52}$$

The recalling ability is derived from (30), (31) and (32). We denote it by P_s. The recalling ability P_s becomes

$$P_s = F\left(\sqrt{\frac{M\,(1-\Delta\phi_0)}{2K\Delta\phi_0}}\right), \tag{53}$$

when $\phi_x = 0.5$ and $s = 0$. In (53), Δ is the degree of pattern separation under the patterns y_k and y_r.

Figure 7 shows the relation between the degree of pattern separation and the recalling ability of the present model. The values of parameters are as follows.

$\phi_x = 0.5$, $\phi_0 = 0.5$, $m = 10$, $n = 100$, $N = 300$, $R = 2$, $\bar{s} = 0$.

The recalling ability of the traditional model is also shown to compare with that of the present model. In the figure, note that the dimension of the key pattern in the traditional model is fitted with that of the pattern on the second layer in the present model. As known from the figure, the present model has the higher recalling ability than that of the traditional model when Φ is controlled in the range indicated by a heavy line, and has the highest recalling ability when Φ is controlled at ϕ_0^2. Then, we try to derive the recalling ability in the general case that the number of connections is R, where Φ is controlled at ϕ_0^R. From (14), the degree of pattern separation Δ in (53) becomes

$$\Delta = \phi_0^{R-1} \tag{54}$$

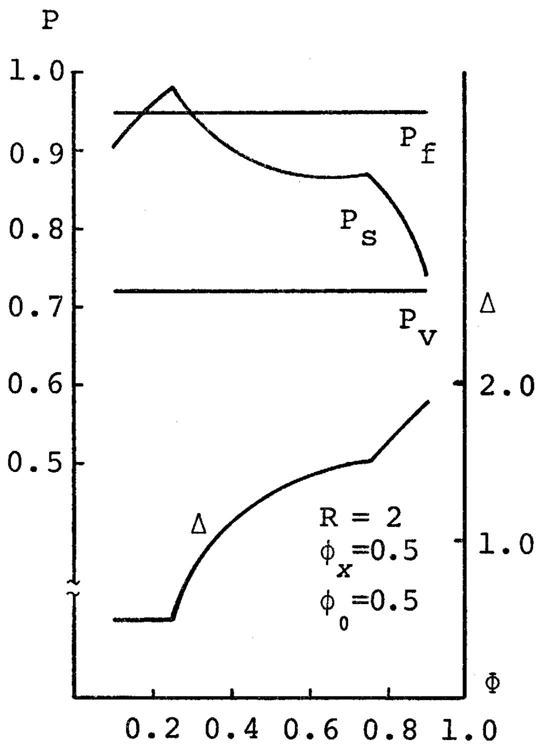

Fig. 7 Relations between Φ and Δ and between Φ and P_s, where $\phi_x = 0.5$, $\phi_0 = 0.5$, $R = 2$, $\bar{s} = 0$ and $N = 300$

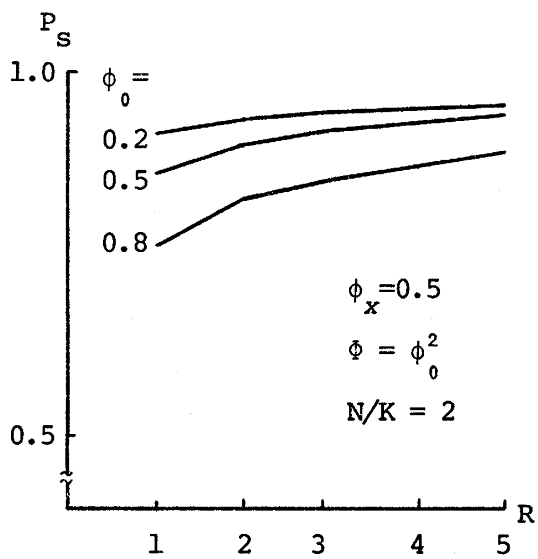

Fig. 8 Relation between R and P_s plotted for ϕ_0 as parameter, where $\phi_x = 0.5$, $\Phi = \phi^2_0$, $\bar{s} = 0$ and $N/K = 2$

The recalling ability P_s becomes

$$P_s = F\left(\sqrt{\frac{M(1-\phi_0^R)}{2K\phi_0^R}}\right), \tag{55}$$

where $\phi_x = 0.5$ and $s = 0$

Figure 8 shows the relation between R and P_s, where $\phi_x = 0.5$, $N/K = 2$, $\bar{s} = 0$ and $\Phi = \phi_0^R$. As know from the figure, the higher recalling ability is obtained by increasing R. The reason for this is as follows. The function of pattern separation is enhanced by the increment of R. Therefore, the overlapping rate among the key patterns on the second layer becomes small. However, the recalling ability is not very improved even if the number of connections becomes larger than about four. The value of ϕ_0^R becomes very small as the number of connections increases even if ϕ_0 is large.

For this reason, the control of the firing rate Φ is difficult. In addition, there is some fear that the number of the firing elements on the second layer, $N \cdot \phi_0^R$, becomes smaller than one under a fixed N. Therefore, an appropriate R should be adopted according to N, with considering the mean of firing rate ϕ_0 of the patterns on the first layer.

7. Computer-simulated Experiments

On computer simulation, the patterns, X_k and y_k, are produced by pseudo-random numbers. The simulation is repeated several times by changing the initial value of pseudo-random numbers. Figure 9 shows the recalling abilities P_s, P_f and P_v against the noise. The value of parameters are as follows.

$\phi_x = 0.5$, $\phi_0 = 0.5$, $m = 10$, $n = 100$, $N = 300$, $K = 30$, $R = 2$, $\Phi = 0.25$.

The solid lines are the theoretical recalling ability. The chain lines are the mean value of the recalling abilities obtained from computer-simulated experiments, where the recalling ability is the mean value of the rates of the elements recalled correctly to all elements of X_r. The figure shows that the recalling abilities, P_s, P_f and P_v, get worse according to the increment of the noise s. The present model, however, holds the high recalling ability over the wide range, and the recalling ability is higher than that of the traditional model. This means that the present model is tolerant of the noise.

Figure 10 is the relation between the mean of firing rate ϕ_0 and the recalling abilities, P_s, P_f and P_v, under the condition $\bar{s} = 0.05$. The values of other parameters are same as those in the figure 9. As known from the figure, the recalling ability P_s of the present model holds the high value, while P_f and P_v of the traditional model get worse drasticlly when the firing rate ϕ_0 becoms large. This means that the property of pattern separation in the present model is very effective for the associative memory system when the overlapping rate among the key patterns is large.

To sum up the above theoretical and experimental examination, the excellent points of the present model are as follows.

(1) The model has the higher recalling ability than that of the traditional model when Φ is controlled at near to ϕ_0^R.

(2) The recalling ability is the highest when $\Phi = \phi_0^R$.

(3) The recalling ability is improved by increasing the number of connections between two layers.

(4) The model is tolerant of the noise.

(5) The model maintains the high recalling ability even if the overlapping rate a-

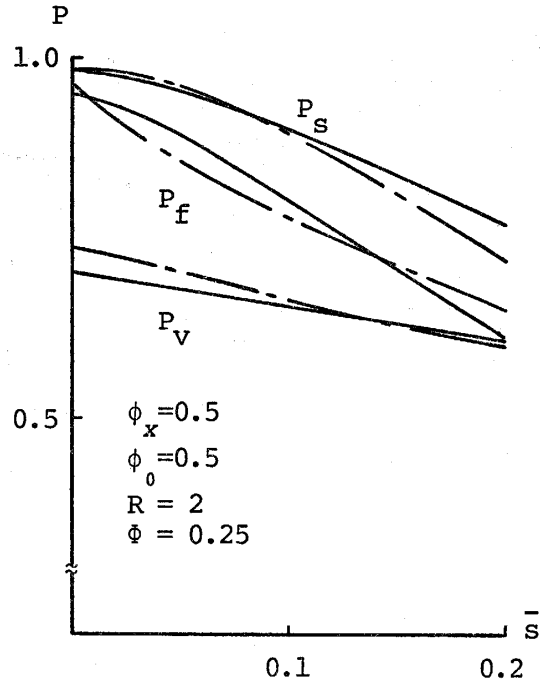

Fig. 9 Recalling abilities, Ps, Pf and Pv, agaist the noise $\bar{s}$, where $\phi_x=0.5$, $\phi_0=0.5$, N=300, K=30, R=2 and $\Phi=0.25$

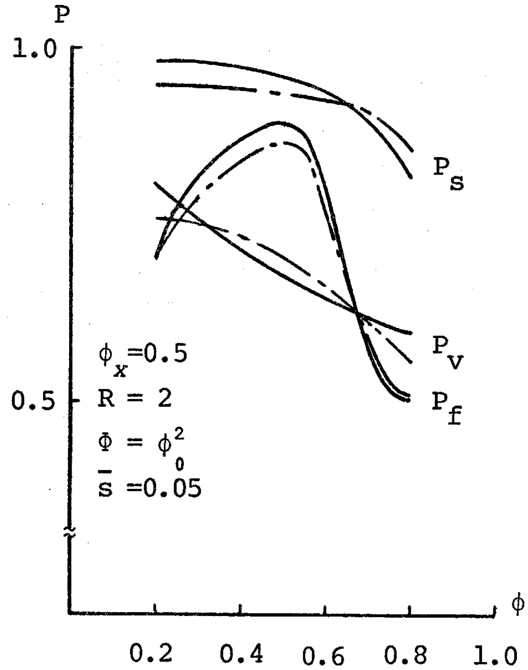

Fig. 10 Recalling abilities, P_s, P_f and P_v, agaist the mean of firing rate ϕ_o, where $\phi_x=0.5$, $\bar{s}=0.05$, N=0.05, N =300, K=30, R=2 and $\Phi=\phi^R{}_0$

mong the key patterns becomes large.

The excellent points of the present model are attributed to the function of pattern separatıon of a two-layer random neural net described before, which produces the patterns favorable for the associative memory.

8. Conclusion

We have clarified that the function of pattern separation of a two-layer random neural net has the property favorble for the associative memory. Then, we have proposed a model of associative memory having the function of pattern separation. The recalling ability in the case including the noise has been derived, and has been compared with that of the traditional model. In addition, the recalling ability has been examined through computer-simulated experiments. As the result, as expected, it has been proved that the model has been superior to the traditional model with respect to the recalling abiity. Further, it has been cleaned up that the recalling ability has been enhanced by the function of patteın separation.

In conclusion, in the comparison of the recalling abilities, we have fitted the dimension of the pattern Y_k on the second layer in the present model with that of the key pattern y_k in the traditional model. However, the total number of the elements in the present model is larger than that in the traditional model by the elements of the first layer. The comparison considering these points should in the future.

References

1) Nakano, K. and Nagumo, J.: "Studies of associative memory using a model of neuron network." Inst. Elertron. Commun. Eng. Jpn. MBE-71-6, 1-18 (1970).
2) Uesaka, Y. and Ozeki, K.: "On some properties of associative memory." Electron. Commun. Eng. 55-D, 323-330 (1972).
3) Kohonen, T.: "Correlation matrix memories." IEEE Trans. Comput. C-21, 353-359 (1972).
4) Kosugi, Y., Kanosue, K and Naito, Y. "Investigation on new types of assocaitive memories." Inst. Electron. Commun. Eng. Jpn. MBE-75-21, 1-12 (1975).
5) Marr, D.: "Atheory of cerebellar cortex." J. Physiol. (London) 208, 437-470 (1969).
6) Wigstrom, H.: "A neuron model with learning capacity and its relation to mechanisms of association." Kybenetik 12, 204-215 (1973).
7) Nakano, K. and Nagumo, J.: "A new development on "associatron". "Inst. Electron. Eng. Jpn. AL-74-7, 1-12 (1974).
8) Torioka, T: "Pattern separability and the effect of the number of connections in a random neural net with inhibitory connections." Biol. Cybernetics 31, 27-35 (1978)
9) Torioka, T.: "Pattern separability in a random neural net with inhibitory connections." Biol. Cybernetics 34, 53-62 (1979).
10) Torioka, T.: "Further consideration on pattern separability in a random neural net with inhibitory connections." Biol. Cybernetics 35, 203-212 (1980).
11) Shimura, M.: "Pattern recognition and learning machine." (Shokodo, Tokyo) p. 63-71 (1972).
12) Ikeda, N. and Torioka, T.: "A model of associative memory with a function of pattern separation." Electron. Commun. Eng. Jpn. 63-A, 943-944 (1980).