Self-Organization and Entropy Decreasing in Neural Networks

Masayoshi INOUE and Masayuki KASHIMA

Department of Physics, Kagoshima University, Kagoshima 890

(Received June 13, 1994)

Dynamics of self-organization of binary patterns in a Hopfield model, a Boltzmann machine and a chaos neural network are investigated with the use of an ensemble average entropy and a short time average entropy. Time dependences of these entropies are calculated by numerical simulations when these models are solving traveling salesman problems. Decreasing of the entropies are observed in consequences of the self-organization at their initial time stages.

§ 1. Introduction

Recently, complex systems have been attracted interest in many fields, and the neural network has been regarded as one of typical examples of the systems. The neural network has a distinguished function of information processing which is carried out by making use of self-organization of patterns. The self-organization is a transient process, therefore it cannot treat with equilibrium approaches. In this paper, we study the phenomenon in a Hopfield model, a Boltzmann machine and a chaos neural network with the use of an ensemble average entropy and a short time average entropy which depend on time. These entropies indicate the variety of the patterns which are included in an ensemble or a certain time interval. Each entropy takes a large (small) value if there are many (few) kinds of patterns, therefore, it decreases while the self-organization of patterns is proceeded. In general, the dynamics of the model depends on its control parameter and also its initial condition. The time development of the ensemble average entropy depends on only model’s control parameter where the average over its ensemble members brings a behavior of a simple decay. On the other hand, the short time average entropy depends on not only its control parameter but its initial condition, and it shows rich behavior; especially for the chaos neural network as will be shown later. We also calculate the time development of an energy, which takes the minimum value when its pattern corresponds to the best solution. The dynamical behavior of the short time entropy is similar to the energy for the Hopfield model and the Boltzmann machine, however the similarity is not always valid for the chaos neural network. Namely, the short time entropy contains information which is different from that of the energy. Thus, we calculate the energy and also the two kinds of entropies to characterize the dynamics of these models in more detail.

The dynamics of each model based on the update rule of the state of its neuron. In the Hopfield model, the state at the discrete time \( n(=0, 1, 2, \ldots) \) of the \( i \)-th neuron \( u_i(n) \) is determined by the value of \( I_i(n) = \sum_j w_{ij} u_j(n) + s_i - \theta \), where \( s_i \) is the external input, \( \theta \) is the threshold value and \( w_{ij}(=w_{ji}) \) is the synaptic weight between \( i \)-th and \( j \)-th neurons. The neuron of this model is asynchronously updated according to the rule which can be written as

\[
\delta u_i(n) = \frac{1}{N} \sum_j w_{ij} u_j(n) \quad \text{for} \quad u_i(n) \neq u_i^{\text{eq}}
\]

where \( u_i^{\text{eq}} \) is the equilibrium state of the neuron.

The dynamics of the Boltzmann machine is similar to that of the Hopfield model, but the update rule of the state is different. In the Boltzmann machine, the state at the discrete time \( n(=0, 1, 2, \ldots) \) of the \( i \)-th neuron \( u_i(n) \) is determined by the value of the probability \( P_i(n) = \frac{1}{1 + e^{-I_i(n)}} \), where \( I_i(n) = \sum_j w_{ij} u_j(n) + s_i - \theta \). The neuron of this model is asynchronously updated according to the rule which can be written as

\[
\delta u_i(n) = \frac{1}{N} \sum_j w_{ij} u_j(n) \quad \text{for} \quad u_i(n) \neq u_i^{\text{eq}}
\]

where \( u_i^{\text{eq}} \) is the equilibrium state of the neuron.

The dynamics of the chaos neural network is similar to that of the Hopfield model, but the update rule of the state is different. In the chaos neural network, the state at the discrete time \( n(=0, 1, 2, \ldots) \) of the \( i \)-th neuron \( u_i(n) \) is determined by the value of the probability \( P_i(n) = \frac{1}{1 + e^{-I_i(n)}} \), where \( I_i(n) = \sum_j w_{ij} u_j(n) + s_i - \theta \). The neuron of this model is asynchronously updated according to the rule which can be written as

\[
\delta u_i(n) = \frac{1}{N} \sum_j w_{ij} u_j(n) \quad \text{for} \quad u_i(n) \neq u_i^{\text{eq}}
\]

where \( u_i^{\text{eq}} \) is the equilibrium state of the neuron.
If \( I_i(n) > 0 \),

\[
u_i(n) = \begin{cases} 
1 & \text{if } I_i(n) > 0, \\
0 & \text{if } I_i(n) < 0.
\end{cases}
\]

Namely, one of the neurons is chosen in a statistical manner from all of the neurons, and its state is changed by the deterministic rule.

The way of update of the neuron in the Boltzmann machine is similar to the Hopfield model, however the state is changed by a stochastic rule. In the Boltzmann machine, taking value \( u_i(n) = 1 \) with probability \( P[I_i(n)] \) and value \( u_i(n) = 0 \) with probability \( (1 - P[I_i(n)]) \), where

\[
P[I_i(n)] = \frac{1}{1 + \exp[-I_i(n)/T]}.
\]

Here \( T \) is the "temperature" of the Boltzmann machine. This model reduces to the Hopfield model if \( T \to 0 \).

The chaos neural network has been introduced by one of the authors, where the update rule is expressed by the motion of two chaos oscillators which are coupled with each other. The equation of motion of the coupled oscillator is expressed by

\[
\begin{pmatrix}
x_i(n+1) \\
y_i(n+1)
\end{pmatrix} = \frac{1}{1 + 2D_i(n)} \begin{pmatrix} 1 + D_i(n) & D_i(n) \\ D_i(n) & 1 + D_i(n) \end{pmatrix} \begin{pmatrix} f(x_i(n)) \\ g(y_i(n)) \end{pmatrix},
\]

where \( D_i(n) \) is the coupling coefficient between the two oscillators in the \( i \)-th neuron at time \( n \), and \( x_i(n) (0 \leq x_i(n) < 1) \) and \( y_i(n) (0 \leq y_i(n) < 1) \) are the variables of the first and the second oscillators in the \( i \)-th neuron at time \( n \), respectively. The maps are chosen as \( f(x) = ax(1-x) \) and \( g(y) = by(1-y) \), where \( a (0 < a \leq 4) \) and \( b (0 < b \leq 4) \) are the control parameters.

This coupling system reduces to Yamada-Fujisaka model if \( D_i(n) \) is taken as a constant. In our model \( D_i(n) \) depends on time as \( D_i(n) = I_i(n)(=\sum_j w_{ij} u_j(n) + s_i - \theta) \), but with \( D_i(n) = 0 \) if \( I_i(n) < 0 \). Namely, the coupling coefficient inside \( i \)-th neuron \( D_i(n) \) is regulated by the states of other neurons \( \{u_i(n)\} \) through the medium of the synaptic weights \( \{w_{ij}\} \). It is obvious that the set of states \( \{u_i(n)\} \) itself is determined by individual states of neurons. Thus, it can be explained that the microscopic rule \( (D_i(n)) \) is subjected to the set of states \( \{u_i(n)\} \) by means of a feedback ("feedback slaving principle"), which is the most important mechanism of our model.

When the coupling coefficient \( D_i(n) \) becomes larger than \( D_c = [\exp(\lambda) - 1]/2 \), the two oscillators synchronized with each other if \( a = b \) and nearly synchronized if \( a \approx b \) after a transient time, where \( \lambda \) is the Lyapunov exponent of the map \( f(x) \). This phenomenon can be measured by the difference \( |x_i(n) - y_i(n)| \) which is used to determine the value of \( u_i(n) \) in the following way:

\[
u_i(n) = \begin{cases} 
1 & \text{if } |x_i(n) - y_i(n)| < \epsilon, \\
0 & \text{otherwise},
\end{cases}
\]

where \( \epsilon \) is the criterion parameter of the synchronization.
If we get $u_i$ with the aid of Eq. (4) after \{x_i, y_i\} changing their values $l$ times according to Eq. (3), $u_i$ has no transient time due to the motion of \{x_i, y_i\} when $l \to \infty$. In this case ($l \to \infty$), the chaos neural network behaves similar to the synchronous Hopfield model if $\epsilon \to 0$ with $(a-b) \to 0$.

When $\epsilon=0$ with $a=b$, the excitation $u_i(n)=1$ signifies that the two oscillators completely synchronize with each other. Once the complete synchronization occurs, the neuron is always in the same state even when the coupling coefficient $D_i(n)$ becomes small. This phenomenon can be easily found if we rewrite Eq. (3) as $x_i(n+1)=f[x_i(n)]+D_i(n)[y_i(n+1)-x_i(n+1)]$ whose coupling term takes 0 at the complete synchronized state. We choose $a \neq b$ to avoid the complete synchronization.

In the Hopfield model, if we set $w_{ij}=0$ and $s_i=0$, the state $u_i(n)$ becomes the boundary between $u_i(n)=1$ and $u_i(n)=0$ when the threshold takes $\theta=0$. On the other hand, in the chaos neural network the same situation occurs when $\theta=-D_c$ ($=-[\exp(\lambda)-1]/2$). Namely, $\theta=0$ in the Hopfield model corresponds to $\theta=-D_c$ in the chaos neural network.

Motion of a synchronous Hopfield model falls into a fix point or a two-state cycle. In the chaos neural network, all neurons are updated synchronously, however it runs properly owing to transient motion of the coupled-oscillator in the neurons. For example, it has abilities of a self association, solving a traveling salesman problem (TSP), and learning: (1) Boltzmann machine scheme high speed learning without simulated annealing, (2) stochastic back propagation learning with the use of the binary neuron.

§ 2. Time dependent entropies of ensembles

We analyze the self-organization of binary patterns when the models are solving a traveling salesman problem (TSP). The TSP for $L$ cities ($C_1 - C_L$) is solved with the method of Hopfield and Tank where $L^2$ neurons are arranged on an $L \times L$ square lattice. The row and the column numbers represent the city and the position of the path, respectively. In this method the synaptic weight between the neuron with (row $i$, column $k$) and the neuron with (row $j$, column $l$) and the external input of the neuron (row $i$, column $k$) are chosen as

\begin{align*}
  w_{ikl} &= -A[\delta_{ij}(1-\delta_{kl})+\delta_{kl}(1-\delta_{ij})] - Bd_i(\delta_{ik+1}+\delta_{ik-1}), \\
  s_{ik} &= A,
\end{align*}

where $A$ and $B$ are positive parameters, $d_{ij}$ is the distance between cities $C_i$ and $C_j$, and $\delta_{ij}=1$ if $i=j$ and $\delta_{ij}=0$ otherwise.

The most important macroscopic quantity is the "energy" $E(n)$ which has been introduced by Hopfield. The energy $E(n)$, whose form is the same with that of an Ising model, is defined by

\begin{equation}
  E(n) = -\frac{1}{2} \sum_{ik} \sum_{jl} w_{ijkl}u_{ik}(n)u_{jl}(n) - \sum_{ik} (s_{ik} - \theta) u_{ik}(n).
\end{equation}

The threshold value $\theta=0$ of the Hopfield model corresponds to $\theta=-D_c$ of the
M. Inoue and M. Kashima

chaos neural network, thus the second term on the r.h.s. of the above equation must be replaced by $\sum_{ik}(s_{ik} - \theta - D_e)$ for the chaos neural network. If we substitute Eqs. (5) and (6) into the above equation, the energy can be written as $E(n) = E_1(n) + E_2(n) + E_3(n)$, where $E_1(n) = (A/2)(\sum_i(\sum_k u_{ik}(n) - 1)^2 + \sum_k(\sum_j u_{jk}(n) - 1)^2)$, $E_2(n) = (B/2)\sum_{ijk}d_{ijk}(n)(u_{i,k+1}(n) + u_{i,k-1}(n))$ and $E_3(n) = -\sum_{ik}(-\theta)u_{ik}(n)$. In the Hopfield model and the Boltzmann machine, $\theta$ is taken as $\theta = 0$ for the TSP case, thus $E_3(n) = 0$. The additional energy takes also $E_3(n) = 0$ if the threshold value of the chaos neural network is chosen as $\theta = -D_e$, however $E_3(n) = -\sum_{ik}(-\theta - D_e)u_{ik}(n) = 0$ if a biased $\theta(\neq -D_e)$ is chosen.

If the traveler visits all $L$ cities on his list once and once only, the energy $E_1(n)$ takes the minimum value zero. There are $L!$ patterns which represent the paths whose $E_1(n) = 0$. The energy $E_2(n)$ of each pattern depends on the total length of its path, and it takes minimum if the path is the shortest one (best solution). The values of $A$ and $B$ are determined by empirically, usually $A/B = 10$ is used for a TSP of $10$ cities, which are distributed in a unit square, in the Boltzmann machine. The ratio $A/B \sim 1$ is good for the chaos neural network. It becomes evident that the TSP can be solved to search for the minimum of the energy $E(n)$, and the best solution corresponds to the global minimum if the parameters $\{A, B\}$ are properly chosen.

In the Hopfield model, the energy always decreases until it is trapped by one of the local minimum states. As a result of this property the Hopfield model has a function of a self association, however it can scarcely find the best solution of the TSP. On the contrary, the energy stochastically decreases in the initial time stage in the Boltzmann machine and the chaos neural network. If the control parameters are chosen properly, an initial pattern of the Boltzmann machine or the chaos neural network converges to a pattern which represents one of the paths. We statistically analyze this self-organization of the pattern with the use of a time dependent information entropy of an ensemble. In the ensemble, all possible different kinds of patterns have been prepared uniformly at the initial time $n = 0$, and each ensemble member develops owing to its dynamical rule. In the chaos neural network, the initial value $x_i(0)$ is chosen as $x_i(0) = r_i/2$, and $y_i(0)$ is determined using $x_i(0)$ as $y_i(0) = x_i(0) + 0.1\epsilon$ if $u_i(0) = 1$ or $y_i(0) = x_i(0) + 10\epsilon$ if $u_i(0) = 0$, where $r_i$ is a uniform random number in the interval $[0, 1]$. The ensemble average entropy $S_{en}(n)$ is defined by

$$S_{en}(n) = -\sum_p f(n, p)\ln[f(n, p)],$$

where $f(n, p)$ is the probability of the pattern $p$ at time $n$, and the ensemble average has been used to calculate it.

The network of $L$ cities TSP has $N = L \times L$ neurons and there are $2L = 2\sqrt{N}$ patterns which give the best solution, because the traveler can choose any city as the first and travels clockwise or anticlockwise. The energy of the best solution takes minimum value $E_0$, which is the global minimum. Namely the degeneracy of the ground state is $W_\theta = 2\sqrt{N}$, thus the minimum entropy is given by $\min[S_{en}] = \ln[2\sqrt{N}]$, where we assume that there is no accidental degeneracy and the parameters $\{A, B\}$ are properly chosen. On the other hand, the maximum entropy is obtained as $\max[S_{en}] = \ln[2^N]$. 
Self-Organization and Entropy Decreasing in Neural Networks

TSP of 4 cities

Fig. 1. The time development of the ensemble average entropies $S_{en}(n)$ for the Hopfield model with \{A=10, B=1\}, the Boltzmann machine with \{1/T=0.7, A=10, B=1\} and the chaos neural network with \{c=0.00001, a =3.82825\(+2/2\), \(b=a-18c\), \(\theta=-2\), \(A=1.5\), \(B=1\}\}, where these models are solving the TSP of 4 cities.

We numerically calculate $S_{en}(n)$ for the three models which are solving the TSP of 4 random cities \(N=4\times4\) neurons) where these cities are distributed in a unit square. Typical examples of the chaos neural network and the Boltzmann machine are compared with the Hopfield model in Fig. 1. There are $2^4=5367$ different kinds of patterns at \(n=0\), thus $S_{en}(0)=-\ln[1/5367]=11.09\ldots\max[S_{en}]$, and it decreases according to the self-organization process in each ensemble of the model. Namely, variety of patterns in each ensemble becomes poor due to the developing of the self-organization. There is no control parameter other than \{A, B\} in the Hopfield model whose saturation value $S_{en}^H(\infty)\approx 4.67$ is larger than those of the other two models with properly chosen control parameters. Because there are many local minimum states in which ensemble members are trapped. The number of local minimum states $W_0$ can be estimated as $W_0=\exp[S_{en}^H(\infty)]\approx\exp[4.67](=\exp[0.29\times16])\approx 107$. Edwards and Tanaka\(^{10}\) have calculated the number of "ground" states of a spin glass with \(N\) spins as $W_{g-\tau}=\exp[0.2\times N]$ where the "ground" state is defined that there appears to be a distribution of levels from which $\delta E$ is zero and $\delta^2 E$ is positive with a distribution function near $E=E_0$. The number $W_0\approx 107$ is larger than the number $W_{g-\tau}$ with 16 spins $W_{g-\tau}=\exp[0.2\times16]\approx 25$. However, the number of ground state of the neural network $W_0=\sqrt{N}=8$ is smaller than $W_{g-\tau}=25$.

The control parameters \{\(\epsilon, a, b, \theta, A, B\)\} fairly affect the self-organization in the chaos neural network, and the results of $S_{en}(n)$ for three typical cases are shown in Fig. 2. The chaos neural network has various features; for example, the case \{\(\epsilon=0.005, a=4, b=a-\epsilon, \theta=-0.5, A=10, B=1\)\} behaves similar to the Boltzmann...
machine with a high temperature. On the contrary, a low temperature of the Boltzmann machine is similar to the case \( \{ \varepsilon = 0.0001, \ a = 4, \ b = a - \varepsilon, \ \theta = -0.5, \ A = 10, \ B = 1 \} \). Peculiar behavior is observed in the case \( \{ \varepsilon = 0.00001, \ a = 3.82825 \approx 1 + 2\sqrt{2}, \ b = a - 18\varepsilon, \ \theta = -2, \ A = 1.5, \ B = 1 \} \), which is an efficient parameter set to find the best solution of the TSP, where the value \( a \) is slightly smaller than the critical value of the window three \((1 + 2\sqrt{2})\). In this case, the Lyapunov exponent \( \lambda \) of the map \( f(x) = ax(1-x) \) with \( a = 3.82825 \) is obtained as \( \lambda = 0.24214 \cdots \) which gives \( D_c = 0.1369 \cdots \), however we use the biased value \( \theta = -2 \) to improve the efficiency. We sometimes use a biased threshold value instead of \( \theta = -D_c \).

In the Boltzmann machine, the rate of decrease and the saturation value of \( S_{en}(n) \) depends on its temperature \( T \), and the ensemble always becomes equilibrium as \( n \to \infty \) because the size of the system is finite \((N = 16)\). The saturation entropy \( S_{en}(\infty) \), which is given by the equilibrium ensemble, decreases as the temperature falls and \( \lim_{T \to 0} S_{en}(\infty) = \min S_{en} \). However the system stays its metastable state for the long time interval \( n_{re} \gg n \gg 1 \), where \( n_{re} \) is the relaxation time which becomes infinitely large as \( T \to 0 \). Temperature dependences of \( S_{en}(n) \) for several \( n \) are shown in Fig. 3 which shows that \( S_{en}(n) \) becomes large as the temperature \( T \) rises at a high temperature region and the reverse relation is observed at a low temperature region. The entropy \( S_{en}(n) \) of the metastable state \((n_{re} \gg n \gg 1)\) at the very low temperature region takes \( S_{en}(n) \approx S_{en}^n \). It is obvious that \( S_{en}(n) \approx \max[S_{en}] \) for the very high temperature region. The entropy \( S_{en}(n) \) takes the minimum value at a certain temperature \( T_{m}(n) \), and the minimum value together with \( T_{m}(n) \) are decreased as the time \( n \) is increased (Fig. 3). It may be noted that it spends much time to simulate for large \( n \). The results of Fig. 3 show that the selection \( T = T_{m}(n) \) is the best to obtain a small entropy under the condition of limited simulation time \( n \) where a simulated annealing is not considered. The temperature \( T_{m}(n) \) is determined by the structure of the energy levels spacing near \( E = E_0 \). Generally it is difficult to calculate \( T_{m}(n) \) theoretically.

§ 3. Short time average entropies of typical cases

In the preceding section we analyze the self-organization using the ensemble average entropy \( S_{en}(n) \). The dynamical behavior of the Boltzmann machine is different from that of the chaos neural network, however the difference is not clearly observed in \( S_{en}(n) \). An alternative quantity is necessary to characterize typical
Self-Organization and Entropy Decreasing in Neural Networks

We introduce a short time average entropy $S_{sh}(n, m)$ of a dynamical system, where $m$ is the interval of the average, which is thought to be one of the candidates of the quantity. The short time average entropy is defined by

$$S_{sh}(n, m) = -\sum_p f(n, m, p) \ln[f(n, m, p)],$$

(9)

where $f(n, m, p)$ is the probability of pattern $p$ in the time interval $((n-m) \sim n)$. This definition gives the maximum value of the entropy as $\max[S_{sh}(n, m)] = \ln[m]$. The short time average energy $E_{sh}(n, m)$ can be defined by the same way where $E(n)$ is averaged over the interval $((n-m) \sim n)$.

There are at most only $m$ different kind of patterns, therefore we can easily calculate $S_{sh}(n, m)$ for a large network. The essential properties of $S_{sh}(n, m)$ are not so related to the number of the neurons $N$, so that we treat a $10 \times 10$ (10 random cities) system instead of the $4 \times 4$ system which has been studied in the preceding section. The 10 random cities of Hopfield-Tank \(^{11}\) is chosen in the present study.

In the following calculations, a typical parameter set is selected for each case and its initial condition is randomly chosen. The entropy $S_{sh}(n, m)$ of the Hopfield model simply decreases in time, and the saturation value always takes zero where its pattern is frozen. However, the saturation values of $E(n)$ and $E_{sh}(n, m)$ depend on their initial condition. The time development of $E(n)$, $E_{sh}(n, m)$ and $S_{sh}(n, m)$ of the Hopfield model are very simple, namely they simply decrease until the system is trapped by one of the local minimum. Our simulation shows that it takes $n \approx 5$ to reach their saturation values.

The Boltzmann machine has the control parameter $T$, thus $E(n)$, $E_{sh}(n, m)$ and $S_{sh}(n, m)$ depend on not only their initial condition but $T$. As the temperature $T$ rises, the fluctuation of $E(n)$, $E_{sh}(n, m)$ and $S_{sh}(n, m)$ become large. A typical example of $E(n)$, $E_{sh}(n, m)$ and $S_{sh}(n, m)$ is shown in Fig. 4. A fluctuation of $S_{sh}(n, m)$, which appears after the self-organization at its initial time stage, becomes large as the temperature rises.

There are many typical cases for the

\[\text{Fig. 4. The time development of the energy } E(n), \text{ the short time average energy } E_{sh}(n, m) \text{ and the short time average entropy } S_{sh}(n, m) \text{ for the Boltzmann machine with } (1/T = 0.7, A = 10, B = 1) \text{ where the model is solving the TSP of 10 cities. The interval of the average } m \text{ is chosen as } m = 10. \text{ The energies } E(n) \text{ and } E_{sh}(n) \text{ take very large values } (\sim 10000) \text{ at } n = 0 \text{ which depends on } (A, B) \text{ and the initial condition.} \text{ The values of a few initial time steps are scaled out in the figure.}\]
chaos neural network which can be characterized by \( \{ \epsilon, a, b, \theta, A, B \} \), where \( \{ \epsilon, a, b, \theta \} \) control the neuron and \( \{ A, B \} \) determined the strength of the synaptic weights and the inputs. In the following, we focus our attention on the fully developed chaos case \( a=4 \) and the below the window three case \( a \leq 1+2/2 \). A similar behavior to the Boltzmann machine of Fig. 4 is obtained if \( \{ \epsilon=0.00003, a=4, b=3.9999, \theta=-0.5, A=10, B=1 \} \). The case \( a=4 \) gives also another type of \( \{ E(n), E_{sn}(n, m), S_{sn}(n, m) \} \) whose results are shown in Fig. 5 where the parameters are chosen as \( \{ \epsilon=0.00001, a=4, b=a-\epsilon, \theta=-2, A=1.3, B=1.3 \} \). The Lyapunov exponent takes \( \lambda=\ln 2 \) for the case \( a=4 \), and the \( \lambda \) gives \( D_c=0.5 \). However, we have used the biased threshold value \( \theta=-2(\neq-D_c) \) in the present case. Figure 5 shows that after the initial time stage, the entropy \( S_{sn}(n, m) \) fluctuates while \( E(n) \) and \( E_{sn}(n, m) \) are nearly constant. This phenomenon is attributable to the reason that the state (pattern) of this case fluctuates around one of local minimum states.

In Figs. 6~9, four typical cases of the results are shown where \( a \) is chosen as \( a \leq 1+2/2 \). A typical self-organization of a pattern is observed in the case \( \{ \epsilon=0.00001, a=3.828427\pm1+2/2, b=a-18\epsilon, \theta=-1.5, A=1.5, B=1 \} \) (Fig. 6) where there are three time stages as: (1) the first time stage \( (n=0\sim40) \), the entropy \( S_{sn}(n, m) \) holds the minimum value \( S_{sn}(n, m)=\ln[m] \), (2) the second time stage \( (n=40\sim60) \), it simply decreases until \( S_{sn}(n, m)=0 \), (3) the third time stage \( (n>60) \), the entropy takes

\[ E(n), E_{sn}(n, m), S_{sn}(n, m) \]
zero almost all times. Namely, preparations needed for the self-organization is made in the first time stage and the self-organization is carried out in the second time stage. The parameter set gives a good example of dynamics of the self-organization which is similar to the self association of the pattern. It may be noted that the self-organized pattern does not always correspond to the best solution of the TSP and the function of search of the best solution is not good because there is almost no fluctuation after the self-organization is finished. The Lyapunov exponent of this case \(a = 3.828427\) is obtained as \(\lambda = 0.0118\ldots\) which gives \(D_c = 0.0059\ldots\), however the biased threshold \(\theta = -1.5\) is used.

A pathological phenomenon can be recognized in Fig. 7 where the parameters are chosen as the same as Fig. 6 except for the threshold value \(\theta = -2.5\) which is smaller than the case of Fig. 6. This case shows "epileptic fits" after a transient time, where a strongly excited state (many neurons are excited) and a weakly excited state (almost all neurons are inhibited) appear alternatively. The epileptic motion is somewhat similar to a two-state cycle motion which can be observed in the synchronous Hopfield model. It is interesting that premonitory symptoms of the epilepsy is usually noticed in \(S_{sh}(n, m)\) which becomes small just before the fit (Fig. 7). The small value of \(S_{sh}(n, m)\) indicates that the pattern is nearly frozen where the effect of the transient motion of \(\{x_i(n), y_i(n)\}\) becomes weaker, and then the dynamics of \(\{u_i(n)\}\) may be shown similar to that of the synchronous Hopfield model. However, we cannot clearly interpret the correct causal relation of the symptoms and the epilepsy.

If we simulate a very long time \(n = 0 \sim 10000\) with \(\theta \simeq -2.2\), where other parameters are the same with Fig. 7, we can observe that the "epilepsy" settles down for a while and then it starts again. This highly intermittent behavior (Fig. 8) is similar to a "chaotic itineracy" whose concept has been proposed by Kaneko, Ikeda, and Tsuda. No analogous phenomenon appears in the Boltzmann machine for any temperature.

One of efficient parameter sets to find the best solution of the TSP is chosen as the last typical case \(\{e = 0.00001, a = 3.82825, b = a - 18e, \theta = -2, A = 1.1, B = 1.4\}\). This parameter set is similar to the cases of Figs. 6 and 7, where the threshold value is taken as a middle value of the two cases and \(e \) and \(\{a, b\}\) are selected to generate a
Fig. 8. The long time behavior of the energy $E(n)$ for the chaos neural network with the set of parameters \{\( \varepsilon = 0.000001, a = 3.828427 \approx 1 + \frac{2}{2}, b = a - 18\varepsilon, \theta = -2.2, A = 1.5, B = 1 \)\} which is the same with Figs. 6 and 7 other than $\theta$.

Proper "random" motion which is stronger than the two cases. The results of the last case are shown in Fig. 9. The time dependence of $E(n)$ shows very noisy and the noisy behavior is smoothed in the short time average energy $E_{sh}(n, m)$ which decreases in the initial time stage ($n = 0 \sim 200$). This case has an efficient function to find the best solution, however the first initial time stage is long and the entropy suddenly takes a small value after the stage where the network frequently finds the best or better solutions. Sometimes the path of the travel is rearranged owing to a large fluctuation, which is generated by chaos. The pattern is not frozen even when the pattern corresponds to the best solution, and the network runs to search for another good solution. If we replace \{\( A = 1.1, B = 1.4 \)\} with \{\( A = 3, B = 3 \)\}, noisy behaviors of $E(n)$, $E_{sh}(n, m)$ and $S_{sh}(n, m)$ calm down, where the system stays in longer one of the solutions than the case of Fig. 9, but the efficiency to find the best solution is not better than Fig. 9.

§ 4. Discussion and some remarks

Ensemble average smoothes over accidental initial conditions and it gives universal properties of the system, however some essential aspects are lost by the average processing. These essential aspects are important for biological systems and
Self-Organization and Entropy Decreasing in Neural Networks

machines. We have introduced the short time average entropy $S_{sh}(n, m)$ to describe the aspects. The $S_{sh}(n, m)$ for typical examples are calculated, while universal properties are analyzed by the ensemble average antropy $S_{en}(n)$. The short time interval $m$ cannot be determined uniquely, so that we empirically select the value as $m=10$ for the $N=100$ system.

In the present study, we treat the self-organization of binary patterns in the neural networks which are solving the TSP. The results do not essentially depend on the form of the distribution of the cities, therefore we choose 4 and 10 random cities. A similar self-organization process is also found in a self association in a neural network, however the process is completed very rapidly and almost no fluctuation after the initial time stage. The analyses of this process using $S_{en}(n)$ and $S_{sh}(n, m)$ are not interesting.

In the neural network the synaptic weight $\{w_{i,j}\}$ is determined by the coordinates of the cities, namely the particular value of $\{w_{i,j}\}$ has its own meaning. Furthermore, the number of neurons $N$ is determined by the number of cities $L$ as $N=L \times L$. On the other hand, a probability of distribution of the exchange constant $\{J_{i,j}\}$ is chosen from a viewpoint of easy calculation in a spin glass theory. A particular distribution $\{J_{i,j}\}$ has no mean, and usually the number of spins $N$ is taken as $N \rightarrow \infty$. The spin glass is a thermodynamic system where an average over the distribution $\{J_{i,j}\}$ is performed, however our neural network is a finite system and the particular value of $\{w_{i,j}\}$ has an essential meaning.

In our simulation for the calculation of $S_{en}(n)$, all possible different kinds of patterns have been prepared uniformly at the initial time $n=0$, which corresponds to the state of $T=\infty$ for the Boltzmann machine. It can be considered that the temperature $T=\infty$ falls to $T=T$ during the self-organization is carrying out in the Boltzmann machine with the temperature $T$. However, the temperature of the Boltzmann machine is not the thermodynamic temperature, and our entropy $S_{en}(n)$ and $S_{sh}(n, m)$ are information entropies which are irrelevant to the second law of thermodynamics.

The update rule of the neuron of the Boltzmann machine reduces to that of the Hopfield model if $T \rightarrow 0$. However, the ensemble of the Hopfield model at $n=\infty$ is not the same with the ensemble of the Boltzmann machine with a successful simulated annealing ($T \rightarrow 0$ at $n=\infty$) whose entropy takes $S_{en}(\infty)=\min S_{en}$. The metastable state of a very low temperature of the Boltzmann machine is similar to the state of the Hopfield model at $n=\infty$.

The Hopfield model has only one typical case, because it has no control parameter. There are three typical cases in the Boltzmann machine which can be characterized as; (1) a high temperature case, (2) a middle temperature case, (3) a low temperature case. In contrast to these two models, we cannot narrow down to a few typical cases for the chaos neural network, which is diversified in nature due to the control parameters $\{\epsilon, a, b\}$. Several typical cases are investigated in the present study. Stochastic motion of $\{u_i(n)\}$ is generated by chaotic motion of $\{x_i(n), y_i(n)\}$, and $\{u_i(n)\}$ sometimes shows an intermittent random motion which cannot observed in the Boltzmann machine. Character of randomness is different from each other, thus there is no simple relation between $\{\epsilon, a, b\}$ and $T$.

Generally, a simulated annealing is indispensable for the Boltzmann machine to
find the best solution, however the process has not been treated in the present study to avoid a complex situation. Relaxation process is closely related to $T_m(n)$, however we cannot obtain a simple relation of $T_m(n)$ in the present study. Problems of the simulated annealing and a detail analysis of the $n$ dependences of $T_m(n)$ are left for the future study.

Acknowledgements

This work is partially supported by the Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture (No. 05836030).

References

14) K. Kaneko, Physica D41 (1990), 137.