Random Attractors for Initial Memory Codes†

Akira Date

Division of Applied Mathematics
Brown University
Box F, Providence, RI 02912, USA
date@dam.brown.edu

Department of Biophysical Engineering
Faculty of Engineering Science, Osaka University
1-3 Machikaneyama, Toyonaka, Osaka 560, Japan

Abstract

The problem of how the temporal relations among events are converted to the spatial structure in neural connections are considered. Random coding is one hypothesis to this problem, in which items to be stored are associated with randomly chosen network attractors, independent of spatial relationships between the items. We demonstrate the idea with a simple network model, and discuss the relevant issues.

1. Introduction

How are the temporal relations among events represented in neural connections? No direct evidence for the problem has so far been described, but there is a some clue. Single-unit recording studies from the monkey inferior temporal cortex have shown that neuronal short-term representations of long-term memory for a set of pictorial stimuli did not depend solely on geometric similarity among visual stimuli but depend on temporal sequences of the set of stimuli exposed to the monkeys in the training session to some extent [13,14]. The results suggest that the strategies used by single neurons to categorize objects utilize not only visual shapes themselves but also temporal relationship among objects presented to the monkeys [3,9,11,13]. Földiák [8] and Griniasty et al. [9] independently proposed network models with modified Hebbian learning rules in which the learning process involves contiguous stimuli in a sequence. With these models, several researchers have tried to reproduce the physiological ob-

†Research supported in part by the Ministry of Education, Science and Culture of Japan under Grant-in-Aid for Scientific Research. This work was done during the author was with the Division of Electronic and Information Engineering, Tokyo University of Agriculture and Technology. He is currently supported by JSPS Research Fellowships for Young Scientists.

servation of Miyashita (1988) [4,9]. The mechanisms they proposed are thought to play an important role in order to recognize three-dimensional (3-D) objects invariant to the changes in retinal image projection [17]. Here we show another possibility to consider random coding strategy in which items to be stored are associated with randomly chosen network attractors, independent of spatial relationships between the items. In other words, the idea is that a some part of networks spontaneously generates random patterns of activity as internal representations for events or objects. It will be shown with a simple network model.

The network we treat here is the randomly and symmetrically connected neural network (random symmetric networks). A large number of equilibrium states or fixed points is in the network. Mathematical properties of the network including the average number of equilibrium states, the neuronal model dependencies have already been reported (see [6] and references therein). Morita (1988) proposed an associative memory system in which input patterns were associated with attractors or stable states of the random symmetric network. The essence of his idea is that Hamming distance between a pair of input patterns does not affect the distance between the corresponding pair of attractors [15].

First, we describes an architecture and properties of the random symmetric networks. Following that, we introduce the two ways generating activity patterns or codes (actually these codes are attractors for the network). And properties of the attractors including distances between them are analyzed. In particular, we will show that a set of attractors in which contiguously generated ones share high similarity in their activity patterns can be created with natural perturbation process. Finally the functional roles of the networks are discussed. The present work was greatly influenced by the consideration of Amari et al. [3].
2. Random symmetric networks and its properties

Here we consider a fully interconnected network consisting of $n$ neural-like elements in which the $i$th output $x_i$, $i = 1, \ldots, n$ takes binary states ($x_i = 0, 1$) and $w_{ij}, 1 \leq i, j \leq n$, the connection weight from the $j$th to the $i$th neuron, is i.i.d. random variable under the symmetric condition $w_{ij} = w_{ji}$. And suppose that connection weights \( \{w_{ij}\}, i > j \), are chosen independently from a normal distribution with mean $\mu_w$ and variance $\sigma_w^2$ and $w_{ii} = 0$. The restriction is not essential to the analysis to follow (see [6]).

The neural state of the system changes in time under the following equation,

$$x_i = I \left( \sum_{j=1}^{n} w_{ij} x_j - \theta \right),$$

where $\theta$ is a variable threshold value which is uniform throughout the network, and $I$ is the step function that $I(u) = 0$ if $u \leq 0$, otherwise $I(u)$ takes 1. Each neuron readjusts its state randomly and asynchronously. The network as a whole evolves in the state space \( \{0, 1\}^n \).

For the network with any symmetric connection matrix \( \{w_{ij}\} \), starting from any arbitrary initial states, the system reaches a stable states and cease to evolve [10].

The network properties can be described with following macroscopic parameters [6],

$$W = \frac{\sqrt{n} \mu_w}{\sigma_w}, \quad \Theta = \frac{\theta}{\sqrt{n} \sigma_w}. \quad (2)$$

There are many fixed points in the network. When the network parameters are designed to maximize the average number of fixed points, it becomes to be $e^{0.1408n}$ [6, 7]. Depending on $W, \Theta$, the networks are classified into three categories (mono-stable, bi-stable, oscillatory) according to the dynamics of activity level, i.e., rate of excited neurons (see [2]). In the following computer simulation, for the simplicity we treat mono-stable networks, in which starting from any activity level configuration (state of neurons), network state is going to converge to a specific activity level for sufficiently large $n$. 

Figure 1. a, Distance and $b$, similarity among stable states as a function of the average activity level $\bar{p}$ of the states for random symmetric networks. a, The average distances were obtained by computer simulations for $\Theta = 0.1, 0.0, -0.2, -0.4, -0.5, -0.6, -0.8, -1.0, -2.0, -3.0, -4.0, -5.0, -6.0$. Dotted line shows the average distance between i.i.d. randomly generated patterns in which each component takes 0,1 with probability of $1 - \overline{p}, \overline{p}$ respectively. b, Same for $a$ but on average similarity (see text in detail).
2.1. Attractors from random initial configuration

So far it seems no exact theory on stability of equilibrium states except that a set of those has ultrametricity [16]. To examine the distance or similarity between stable states, computer simulations were carried out on systems of \( n = 1000 \) and \( W = -5 \) (\( \mu_w = -5.0, \sigma_w = 30.0 \)). The dynamics algorithm was initiated from randomly chosen initial starting configurations of activity level \( p = 0.5 \), in which each component takes 0, 1 with probability 0.5 respectively. This trial was repeated for \( m = 1000 \) times.

Following two indices, \( d \) and \( s \), are defined as distance and similarity between stable states,

\[
d(x^\alpha, x^\beta) = \frac{1}{n} \sum_{i=1}^{n} |x_i^\alpha - x_i^\beta| \quad (3)
\]

\[
s(x^\alpha, x^\beta) = \frac{1}{np} \sum_{i=1}^{n} x_i^\alpha x_i^\beta \quad (4)
\]

where \( p \) is the activity averaged over all stable states, \( \bar{p} = 1/(nm) \sum_{\alpha=1}^{m} \sum_{i=1}^{n} x_i^\alpha \). The normalized Hamming distance \( d \) between two binary states \( x^\alpha \) and \( x^\beta \) is defined as the rate of the number of places in which the digit are different, whereas the index \( s \) of similarity or overlap between stable states is also defined as how two compared states share the component of 1. Notice that distance \( d \) treats the two binary 0, 1 as equivalent while the similarity \( s \) ignores the share of the component of 0.

Simulation results were obtained for various values of \( \Theta \) (see legend of Fig.1), while a set of the connections \( \{w_{ij}\} \) was fixed. In Fig.1a, the distances \( d \) (ordinate) are represented as a function of the average activity \( \bar{p} \) of stable states (abscissa). Since the networks we used were mono-stable, depending on \( W, \Theta \), any initial states converged to a specific activity level \( \bar{p} \) independent of the initial activity levels. A dotted line shows the distances between i.i.d. random patterns in which each component takes binary values 0 or 1 with probability 1 - \( p \), \( p \) respectively. The same data but for similarity is presented in Fig.1b. For the i.i.d. random patterns, the expected value of distance is \( 2p(1 - p) \), that of similarity is \( p \). In both cases, when \( n \) is sufficiently large, variances are going to be 0.

The result presented in Fig.1a indicates that the distances between stable states are smaller than that for i.i.d. random patterns. And the distance is maximized when the network parameters are adjusted to make stable states at activity level \( p \approx 0.25 \). It seems interesting, because the average distance between i.i.d. random patterns is maximized at \( p = 0.5 \). From another point of view, the network designed to have stable states at lower activity level can generate pseudo i.i.d. random activity patterns.

From the simulation study for various network parameters \( W, \Theta \) (not shown here), it seems that the distance between stable states depends on the activity level \( p \) of the stable states that the network actually converge, although it is not theoretically demonstrated (the theoretical base should be constructed).

2.2. Attractors generated by threshold modulation

Various schemes can be used to generate patterns of activity. One example has been presented in the previous section, in which patterns of activity (stable states) were obtained by evolving the networks from randomly chosen initial configurations. The process can be viewed as the patterns were generated by i.i.d. random noise driven to individual neurons. Here we show a different way generating activity patterns by modulating the neuronal threshold \( \Theta \), which is equivalent to the external signal \( -\Theta \) for all elements.

First, suppose that the state 00000 ... for the network \( W = -5 \) (same one to the previous section). The 0000 state is stable for \( \Theta \geq 0 \). When \( \Theta \) is decreased from a particular positive value, and at \( \Theta < 0 \) the non-0000 state or activity pattern will appear [10, 15]. And \( \Theta \) is decreased until a particular value \( \Theta_0(<0) \). Then, to make a pattern having lower activity level, \( \Theta \) is increased until \( \Theta = 0 \). The average activity level for \( \Theta = 0 \) was approximately 0.13. In such a manner, one pattern of activity can be generated. Notice that the generated pattern is an attractor or a stable state for the network. With the process of modulating \( \Theta : 0 \to \Theta_0 \to 0 \to \Theta_0 \cdots \), the network can generate a extremely large number of different patterns of activity. This modulation was repeated for \( m = 1000 \) times in our simulation. The generating method utilizes the behavior of each neuron which act asynchronously and update at random in time.

We examined the relationships between the random patterns thus generated for \( \Theta_0 = -0.1, -0.3, -0.6, -1.0, -2.0, -5.0, -6.0 \). In our computer simulations, the system was started at the 0000 state at \( \Theta = 0.0 \). \( \Theta \) was then varied to \( \Theta_0 \). And the state of the network was allowed to evolve until stationary. Parts of the generated patterns (100 patterns for the first 100th neurons) for \( \Theta_0 = -1.0 \) are exemplified in Fig.2a, where the ordinate is the arbitrary number attached to each neuron, and the abscissa is the number sequentially attached to the patterns according to the generated order. These patterns were stable states for \( \Theta = 0.0 \). For the generated patterns, single neurons tended to be either active or silent. Indeed, 122 neurons (12.2%)
Figure 2. Property of generated patterns. a, Part of the patterns generated by the network $W = -5.0, \Theta = 0.0$. The patterns were generated by modulating neuronal threshold $\Theta : 0 \rightarrow \Theta_0 \rightarrow 0 \rightarrow \Theta_0 \cdots$ ($\Theta_0 = -1.0$). b, Similarity among random patterns. c, Temporal correlation between random patterns according to the generated order. Each arrowhead represents the average overlaps among all possible pairs.
were in silent to all of the 1000 generated patterns, and 10 neurons were in active more than the 990 patterns.

The average similarity between generated patterns is shown in Fig.2b where A,B,C,D,E,F,G represents for the case of $\Theta_0 = -0.1, -0.3, -0.6, -1.0, -2.0, -5.0, -6.0$ respectively. The overlaps between the patterns depended on $\Theta_0$, whereas we used the same network and no difference between the activity level of stable states. They were higher than those between stable states presented in the previous section, except F and G.

We then examined whether the similarity between the patterns depended on the temporal order that the patterns were generated. The results are shown in Fig.2c where the abscissa represents the difference between the number attached to each pattern, and average similarity between the patterns are plotted in ordinate. The combinatorial number required for obtaining the average similarity was different among the value of the abscissa (neighbor). For example, 999 pairs were required for neighbor 1, and 990 for neighbor 10. The arrow on the right side of Fig.2c shows the overlap between patterns averaged over all possible pairs. The overlaps between the contiguously generated patterns had high on the average, again except F and G. The extent of the similarity depended on $\Theta_0$. For $\Theta_0 = -1.0$, the overlaps were higher than the average until neighbor 8, and for $\Theta_0 = -0.6$, the overlaps decreasing rapidly until neighbor 21. Notice that the constant overlaps between the pattars, for example, 0.47 for $\Theta_0 = -1.0$, can be eliminated by successive stages. For $\Theta_0 = -5.0, -6.0$, the averaged overlaps did not depend on the neighbor. This indicates that once $\Theta$ is decreased and the activity level is high (average activity level becomes over 95% at $\Theta = -5.0$), patterns to be generated are distributed within a particular region of the state space independent of the previous state.

3. Discussion

To show the idea of what the brain really does might be to create spontaneously internal representation for a novel event or object, we used a simple network model. The statistical properties of the network attractors were analyzed. In particular, it was shown that the extent of similarity between patterns of activity generated sequentially can be controlled by the value of a external signal (−$\theta$ in our case). If there are various modules which have different scales of $\Theta_0$, representations behaving like traveling-waves for the various time scale, such as second, minute, hour and so on, can be created. This might be used as the initial internal representation for the concept of time.

![Figure 3. Schematic representation of the random coding.](image)

To fix the idea of the random coding presented so far, a schematic picture is depicted in Fig.3. Each dot in representation I,II (R1,R2) is a state of activity pattern. The dots in R2 correspond to stable states (recall that a random symmetric network has a extremely large number of fixed points) in which large dots represent the states selected randomly as shown in section 2.2. Patterns of activity in R1 are input signals or items to be memorized. These items are memorized using stable states in R2. Therefore each item has two representations in both fields, R1 and R2. States between the two representations are bound by conventional Hebbian learning. If attractors in R2 are selected as mentioned in section 2.2, two items presented sequentially to R1 are represented as similar activity patterns in R2, whatever spatial relationships these signals have in R1 (refer to item 7 and 8 in Fig.3). Notice that similarity between two spatial patterns in R1 is not necessarily maintained in R2 (see relationship between 1 and 7 in R1 and R2).

We did not attempt to describe the system in full detail. Actually we have not demonstrated that these systems really work. For example, we have not considered here the situation that the inputs or stimuli are presented periodically with fixed order. Suppose there are 100 stimuli or patterns to be stored, and they are used repeatedly in a fixed order. Many representations are then created for a stimulus in R2. How the network knows these different states in R2 represent the only one stimulus? In the real brain, the problem might be solved by different levels of processing in which the data structure for memory are reorganized. It may be done by the interactions between the hippocampal system and the neocortex [1, 12]. To know principles or roles of (at least) this two separate systems is another
key issue for understanding primate memory system. The error-correcting functioning was also not considered in this paper. If stable states selected sequentially in R2 are too close, a input with noise in R1 may not recall a correct states in R2. (suppose a input on item 7 with noise in Fig.3). Morita's network is very complicated, but he demonstrated the oscillatory mechanism of external signal to each neuron and the learning between input patterns and internally created attractors [15] (the attractors generated by his network seem completely independent of generated order).

We cannot reasonably expect to invent a solution to approach the general application of human thought. Therefore it would seem to be expedient to look to the neural science for clue, such as the evidence that the single neurons in the inferior temporal cortex tend to respond selectively the pictorial stimuli exposed sequentially [13]. It is in this spirit that the model here, with the random coding, has been developed. Models have been also formulated with the modified Hebbian learning rule [4,9] or trace mechanism [8]. In any cases, if two objects are extremely different in spatial patterns but presented sequentially, internal representations of them become to be similar. The point that I wish to make here is that spontaneous activities of neurons might play an important role not only during prenatal development [18], but also for primal information processing in memory. And these activities or initial internal representations would be effective to form a episodic memory which consists of a evidence including a concept of time, rather than to associate different 2-D views of a 3-D object. Latter may be encoded effectively by relationships between parts [5].

We have benefited from the advice and encouragement of Shun-ich Amari and Koji Kurata, although the present work could not reflect their influence.

References


