A model of memory formation in the pair-association task

Atsuo SUEMITSU, Masahiko MORITA***

*Doctoral Program in Engineering, University of Tsukuba
1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8573, Japan
**Institute of Engineering Mechanics and Systems, University of Tsukuba
1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8573, Japan

Email:sue@bcl.esys.tsukuba.ac.jp

Abstract

Neurons related to pair-association memory have been found in the inferotemporal cortex of monkeys, but their activities do not accord with existing neural network models. We have previously shown that a neural network consisting of excitatory-inhibitory cell pairs is able to explain these neuronal activities; however in order to form the memory, it required an external learning signal. In the present paper, we supplement this model with another network to generate the learning signal. By simply inputting paired patterns in order, this model forms pair-association memories using the interaction between the two networks.

1. Introduction

In the inferotemporal cortex of monkeys, interesting neuronal activities related to pair-association memory have been reported [1]. This finding is very important when considering how long-term memories are structured and retrieved in the brain, but such activities are difficult to explain by conventional neural network models.

We previously reported [2] that a network consisting of pairs of excitatory and inhibitory cells explains the above empirical data well provided that an appropriate learning signal is given externally to the model. This learning signal is a spatiotemporal pattern that is necessary to form the memory. However, since no external learning signal is given to the monkey in the actual task, it should be generated from stimuli in the brain.

To solve this problem, we modify our model by adding a supplementary network, in which a pair of input patterns is transformed into a spatiotemporal pattern. We also show by computer simulation that this model has the ability to perform the pair-association task.

2. Background

Sakai and Miyashita [1] studied the inferotemporal neurons of monkeys by conducting the following experiment.

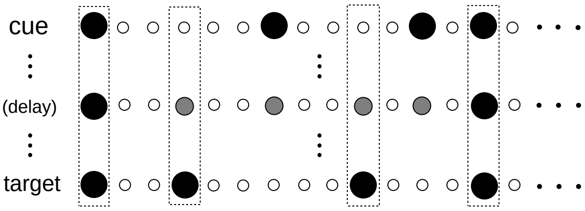


Fig. 1: Illustration of transition in the firing pattern during recall.

They first generated many pairs of figures, and trained monkeys to associate paired figures with each other. Then they measured neuronal activities in a delayed matching task, where one of the paired figures (cue figure) is presented for a short time and the monkey judges whether the figure presented after a delay period is the match (target figure) or not.

The results were basically the same as those from aprevious study by Miyashita and Chang [3], where the monkey memorized figures separately and sparse coding was found to be used for representing learned figures. However, two kinds of neurons exhibiting distinctive activities, called “pair-coding” and “pair-recall” neurons, were newly observed.

Pair-coding neurons exhibit a selective response to both figures of a pair, and exhibit sustained activity during the delay period. Pair-recall neurons did not exhibit a response to the cue figure, but gradually increase their activity during the delay period and exhibit the maximum activity when the target figure is presented.

The above result can be interpreted as follows (see Fig. 1). Each figure is represented by a sparse firing pattern of a neuron group. This pattern does not depend on pictorial features of the figure, rather, paired cue and target figures are encoded into mutually similar patterns and the overlap corresponds to pair-coding neurons. During the delay period, the firing pattern changes gradually from

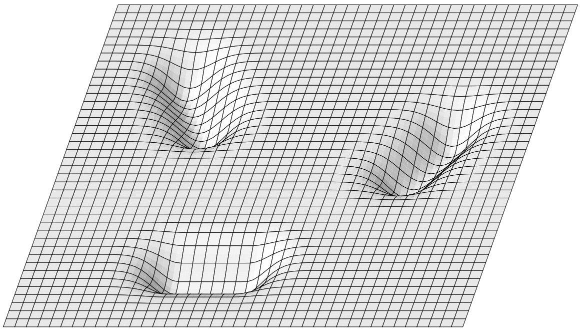


Fig. 2: Schematic energy landscape of a network storing paired associates.

the cue-coding pattern to the target-coding pattern; in this process, some neurons act as pair-recall neurons.

In order to realize such a gradual shift in firing pattern stably, it is thought that not only the cue-coding and target-coding states of the network, but also the entire path connecting them should be smooth and attractive, or at the bottom of an energy gutter, as schematically depicted in Fig. 2. In this figure, the x-y surface represents the state space of the network, and the z-axis represents the energy; three gutters corresponding to three pairs are drawn.

Although it is difficult to form such a landscape using conventional attractor neural networks, as they usually have a rippled energy landscape [4], we were able to realized it by introducing local inhibition cells [2], successfully reproducing similar activities to those of the above inferotemporal neurons. However, this model is insufficient because we used a man-made spatiotemporal pattern as a learning signal, and gave it externally to the network.

For a neural network to automatically generate an appropriate learning signal, it should change its output pattern successively yet slowly while preserving cue pattern information, which is difficult for any single isolated network. We found, however, that this problem can be solved by utilizing two interactive networks, as described in the following section.

3. Model

3.1 Structure

The model is composed of two interconnected networks, as shown in Fig. 3. Memories are formed in the storage network N_1; the network N_2 receives the input pattern $\boldsymbol{s}$ and the output pattern $\boldsymbol{x}$ of N_1, generates the learning signal $\boldsymbol{r}$, and sends it back to N_1.

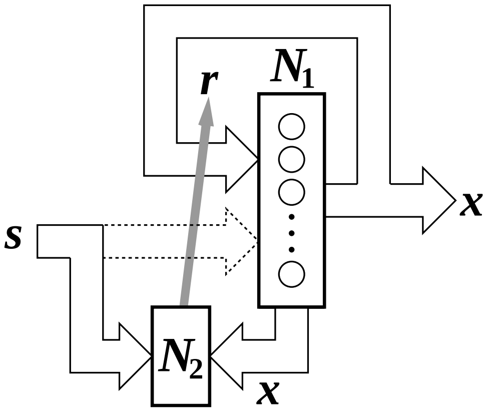


Fig. 3: Block diagram of the model.

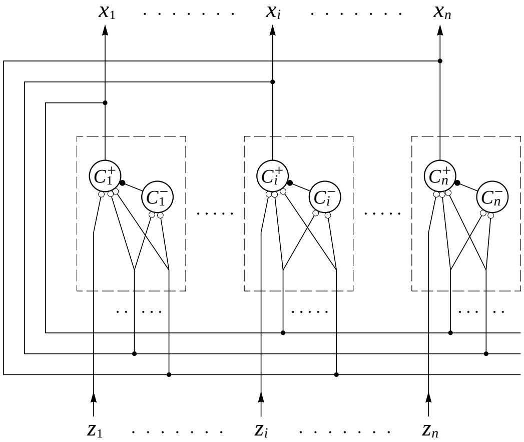


Fig. 4: Structure of the storage network N_1.

The storage network N_1 consists of pairs of excitatory and inhibitory cells, as shown in Fig. 4. A pair of cells surrounded by a broken line comprises a single unit, where the excitatory cell C_i^+ emits the output of the unit and the inhibitory cell C_i^- sends a strong inhibitory signal to C_i^+. Both cells receive recurrent inputs from other units. In mathematical terms,

$$
\begin{aligned}
y_i &= f\left(\sum_{j=1}^{n} w_{ij}^{-} x_j - \theta\right), && (1)\\
\tau \frac{du_i}{dt} &= -u_i + \sum_{j=1}^{n} w_{ij}^{+} x_j - w_i^{*} y_i + z_i, && (2)\\
x_i &= f\left(u_i\right), && (3)
\end{aligned}
$$

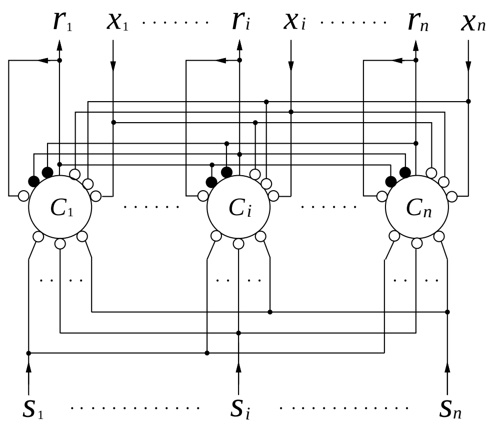


Fig. 5: Structure of learning signal generation network N_2.

where x_i and y_i are the outputs of C_i^+ and C_i^-, respectively, u_i is the potential, z_i is the external input, w_{ij}^+ and w_{ij}^- are synaptic weights from the j-th unit to C_i^+ and C_i^-, respectively, w_i^* represents the efficiency of the inhibitory synapse from C_i^- to C_i^+, and θ is a positive constant.

The activation function $f(u)$ of each cell is a monotonic sigmoid function that increases from 0 to 1. However, the input-output characteristics of the unit are nonmonotonic, that is, output x increases with the total input when input is small, but decreases when it becomes large and the inhibitory cell emits a large output.

3.2 Learning signal generation network

The structure of network N_2 is shown in Fig. 5. This network consists of n cells, corresponding one-to-one with the units of N_1. The i-th cell C_i receives the input pattern $\boldsymbol{s} = (s_1, \ldots, s_n)$ through a synaptic weight p_{ij} and outputs the learning signal r_i to the corresponding unit of N_1. The synaptic weight p_{ij} takes a random value such that N_2 functions as a random transformation network. Cell C_i also receives feedback signal x_j from every unit of N_1 through a random synaptic weight q_{ij}.

This network is a kind of competitive network, as C_i has a self-excitatory connection and lateral inhibitory connections. This permits only a few cells to emit large outputs, with the outputs of the other cells being almost zero. This indicates that output pattern $\boldsymbol{r}$ is a sparse vector.

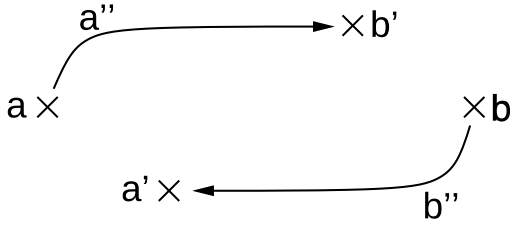


Fig. 6: Paths of the learning signal.

In mathematical terms,

$$\begin{aligned} \tau \frac{dv_i}{dt} &= -v_i + \sum_{j=1}^{n} p_{ij} s_j + \sum_{j=1}^{n} q_{ij} x_j \\ &\quad - \rho \sum_{j \neq i} r_j + \sigma r_i + h, \qquad (4) \\ r_i &= f(v_i), \qquad (5) \end{aligned}$$

where v_i denotes the potential of C_i, ρ and σ are positive constants representing the efficiency of lateral inhibition and self-excitation, respectively, and h is an offset.

3.3 Behavior of model

When we consider the behavior of the model in learning, the interaction between networks N_1 and N_2 is important. Since the output $\boldsymbol{r}$ of N_2 is the learning signal for N_1, generally $\boldsymbol{x}$ is similar to $\boldsymbol{r}$. When $\boldsymbol{r}$ moves, however, $\boldsymbol{x}$ follows slightly behind, moving continuously at a limited rate even if $\boldsymbol{r}$ jumps. Also, N_2 tends to preserve its current output because of the competitive property of cells, whereas the input from N_1 through random connections has an effect of driving the state of N_2 toward a certain direction, dependent on $\boldsymbol{x}$. Taking this interaction into account, let us consider the case for which we input cue pattern A into the model in the rest state (where the outputs of all cells are almost zero) and input target pattern B after a delay.

First, when input $\boldsymbol{s} = \mathsf{A}$ is maintained for some time, network N_2 outputs a, which is a randomly transformed pattern of A, and after a short delay, output $\boldsymbol{x}$ of N_1 becomes similar to a. This output is fed back into N_2, but it does not change $\boldsymbol{r}$ significantly while $\boldsymbol{s} = \mathsf{A}$.

When the external input is paused and $\boldsymbol{s} = 0$, however, the effect of feedback becomes dominant and $\boldsymbol{r}$ begins to move. This continues during the delay period, with $\boldsymbol{r}$ gradually decreasing its moving rate to reach pattern a", which is moderately similar to a. Then, by inputting B, $\boldsymbol{r}$ moves slowly

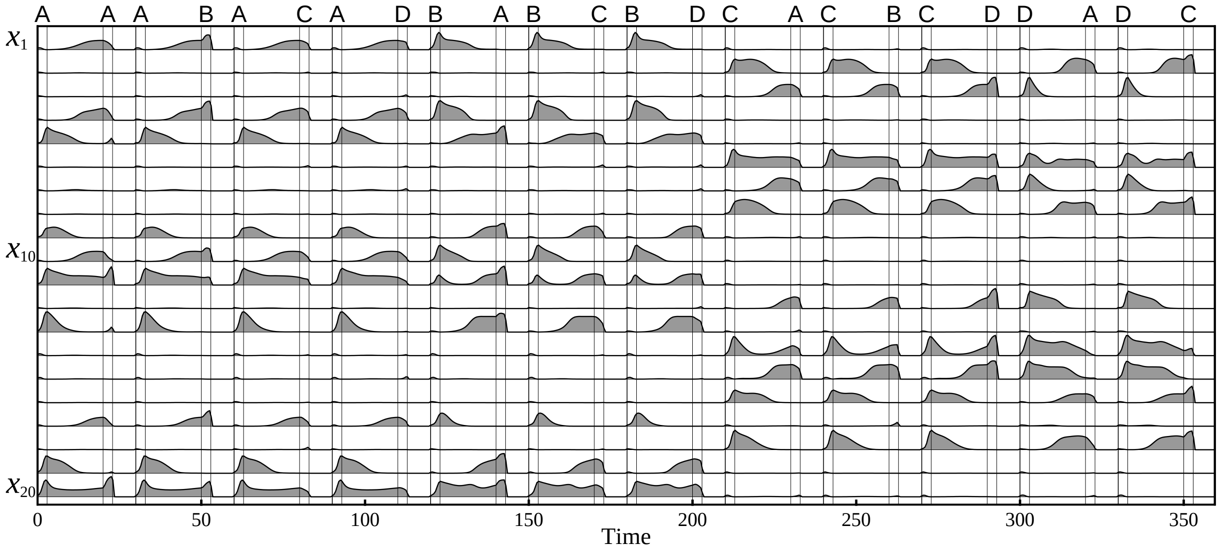


Fig. 7: Response of each unit to various input patterns.

toward b, which is the output pattern of N_2 to input B in the rest state, but does not reach b, stopping at b' (see Fig. 6).

In the same way, when we input B and A, in that order, a learning signal starting from b via b'' to a' is generated. It should be noted that the feedback from N_1 to N_2 is important in separating the two paths a $\to$ b$'$ and b $\to$ a$'$ as well as in regulating the moving rate of $\boldsymbol{r}$.

In parallel to the above process, N_1 is trained using $\boldsymbol{r}$. Specifically, N_1 receives $\boldsymbol{r}$ in the form $z_i = \lambda r_i$, where λ denotes the input intensity, and synaptic weights are modified according to

$$\tau' \frac{dw_{ij}^{+}}{dt} = -w_{ij}^{+} + \alpha r_i x_j, \tag{6}$$

$$\tau' \frac{dw_{ij}^{-}}{dt} = -w_{ij}^{-} - \beta_1 r_i x_j + \beta_2 x_i x_j + \gamma. \tag{7}$$

Here α, β_1, and β_2 are learning coefficients, γ is a positive constant representing the lateral inhibition between units, and τ' is a time constant of learning ($\tau' \gg \tau$). Coefficient α may be a positive constant, but the learning performance is better when α is a decreasing function of x_i; β_1 and β_2 are constants that satisfy $0 < \beta_1 < \beta_2$.

Through this learning, the network energy is lowered around the state specified by $\boldsymbol{r}$, and as a result of $\boldsymbol{r}$ moving successively at a slow pace, a gutter is engraved in the energy landscape along its track. In addition, a gentle flow in the same direction as the movement of $\boldsymbol{r}$ is generated at the bottom of the gutter.

Consequently after learning, the state of N_1 moves along the trajectory of $\boldsymbol{r}$ simply by just giving the initial state. That is, if we input A as a cue and a is sent to N_1 through N_2, N_1 shifts its state to b' during the delay period. If we input B then, the state of N_1 quickly changes to b, and thus N_1 responds more strongly to B than to any other pattern (see Fig. 8).

4. Computer Simulation

A computer simulation was carried out with parameters

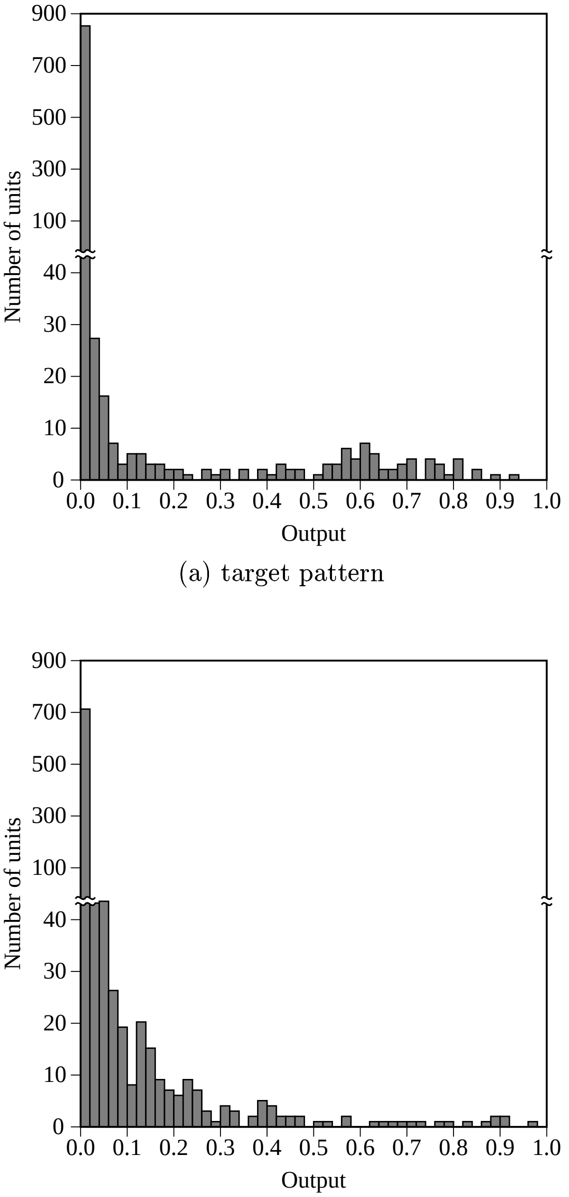
$$\tau = 10, \ \ \tau' = 50000\tau, \ \ \theta = 3, \ \ w_i^* = 10,$$
$$\lambda = 0.3, \ \ c = 10, \ \ \alpha' = 50, \ \ \beta_1 = 25,$$
$$\beta_2 = 50, \ \ \gamma = 0.05, \ \ \rho = 0.016, \ \ \sigma = 0.8.$$

First, we randomly generated 20 pairs of patterns, where each pattern is a 1000-dimensional sparse vector with 10% of elements being 1 and the rest 0. We then input these pairs of patterns in order and in reverse order, applying the above learning procedure. Training was repeated 20 times for each pair.

After learning, we tested the model by repeating a trial in which we gave a cue pattern to the model, input a test (target or non-target) pattern after a delay, and reset all cells. The response of the model is shown in Fig. 7, where the time course of the outputs of 20 units in N_1 is plotted. We can see that many units exhibit a large output when we input the correct target (in the second, fifth, tenth and twelfth trials), whereas the outputs are small for non-target input.

Histograms of the outputs of all units at the end of test input are shown in Fig. 8, where (a) and (b) are typical cases for target and non-target patterns, respectively. Although their averages are nearly equal, the two distributions are obviously different. In fact, the number of units in (a) with outputs greater than 0.5 is about 3 times larger than in (b). This indicates that the model is able to distinguish the target pattern.

We also see that the units exhibit similar activities to the inferotemporal neurons. For example, unit 20 responds to both A and B and sustains a large output during the delay period, and unit 1 exhibits no response to A but strong response to B with a gradually increasing output in the delay period. These activities correspond well to those of the above pair-coding and pair-recall neurons in the monkey brain.



(a) target pattern

(b) non-target pattern

Fig. 8: Distribution of x_i.

5. Concluding remarks

We have described a model that forms pair-association memory based on the interaction between two networks, and demonstrated that this model can not only perform a pair-association task but also explains the neuronal activities observed in the monkey inferotemporal cortex.

This model is biologically plausible, constructed on the basis of computational requirements, and to date, no other model can sufficiently explain the above empirical data. We therefore believe that the same principle underlies the neural mechanism of pair-association memory in the primate brain. We also believe that the learning signal generation network N_2 corresponds to the rhinal cortex based on lesion studies on this area [5,6], although further examinations are necessary, both theoretically and experimentally.

References

[1] Sakai, K. and Miyashita, Y.: Neural organization for the long-term memory or paired association, *Nature*, **354**, 152–155 (1991).

[2] Suemitsu, A. and Morita, M.: A neural network model of pair-association memory in the inferotemporal cortex, *The 6th International Conference on Neural Information Processing*, **2**, 790–794 (1999).

[3] Miyashita, Y. and Chang, H. S.: Neuronal correlate of pictorial short-term memory in the primate temporal cortex, *Nature*, **331**, 68–70 (1988).

[4] Morita, M.: Computational study on the neural mechanism of sequential pattern memory, *Cognitive Brain Research*, **5**, 137–146 (1996).

[5] Murray, E. A., Gaffan, D. and Mishkin, M.: Neural substrates of visual stimulus-stimulus association in rhesus monkeys, *J. Neuroscience*, **13**, 4549–4561 (1993).

[6] Higuchi, S. and Miyashita, Y.: Formation of mnemonic neural response to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions, *Proc. Natl. Acad. Sci. USA*, **93**, 739–743 (1996).