A property of associative memory models with replacing units

Akira Date
Faculty of Engineering
University of Miyazaki
Miyazaki 889-2192
date@cs.miyazaki-u.ac.jp

Koji Kurata
Faculty of Engineering
University of the Ryukyus
Okinawa 903-0213
kurata@mibai.tec.u-ryukyu.ac.jp

Abstract

This paper describes a property of associative memory networks in which a number of units are replaced when the networks learn. In our network, every time the network learns a new item or pattern, a number of neurons die and the same number of neurons are born. It is shown that the memory capacity of the network depends on the number of replaced units, and that the memory capacity is maximized when the number of replaced units is optimal. The optimal number of replaced units is small and seems to be independent of the network size. Although our model was not motivated by higher nervous function, the results suggest that small number of newly born neurons might be optimal in some sense for the distributed memory system.
Keywords: associative memory, collective memories, catastrophic interference, new neuron

1 Introduction

The purpose of this paper is to describe a property of associative memory networks in which a number of units are replaced when the networks learn a new item. The neural network models of associative memory have been studied extensively since 1960's [1, 2, 3, 4, 5]. Most familiar learning for association among memories is the Hebb rule or a *correlation*-based learning in which each connection weight between two neurons are modified according to the correlation between the activities of the neurons. The learning is *local*, and easy to implement in the circuits. There exists a memory capacity for the network. The addition of new memories beyond the capacity overloads the system and makes all memory states irretrievable (catastrophic forgetting) unless there is a provision for forgetting old memories.

If the dynamics of weight connections of the network have decay [3, 4] or saturation[1], the catastrophic forgetting does not occur and the network can keep recent memories. Here we describes the alternative network which share the same properties; avoidance of catastrophic forgetting, keeping recent memories. In our network, every time the network learns, neurons die and are born. The number of replaced units was varied in the experiment to maximize the memory capacity. We found that the optimal number of replaced units was small and independent of the network size. Although our model was not motivated by higher nervous function, the results might suggest that small number of newly born neurons is enough and optimal for a temporal memory system such as the hippocampus.

2 Associative Memory

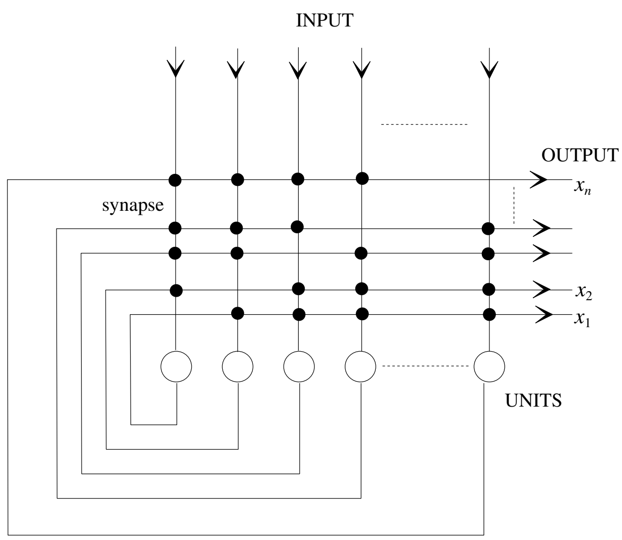

Figure 1: A network of associative memory

2.1 Network dynamics

In our network of associative memory (Fig.1, [1]), each unit, j, has two states, and is described by a variable

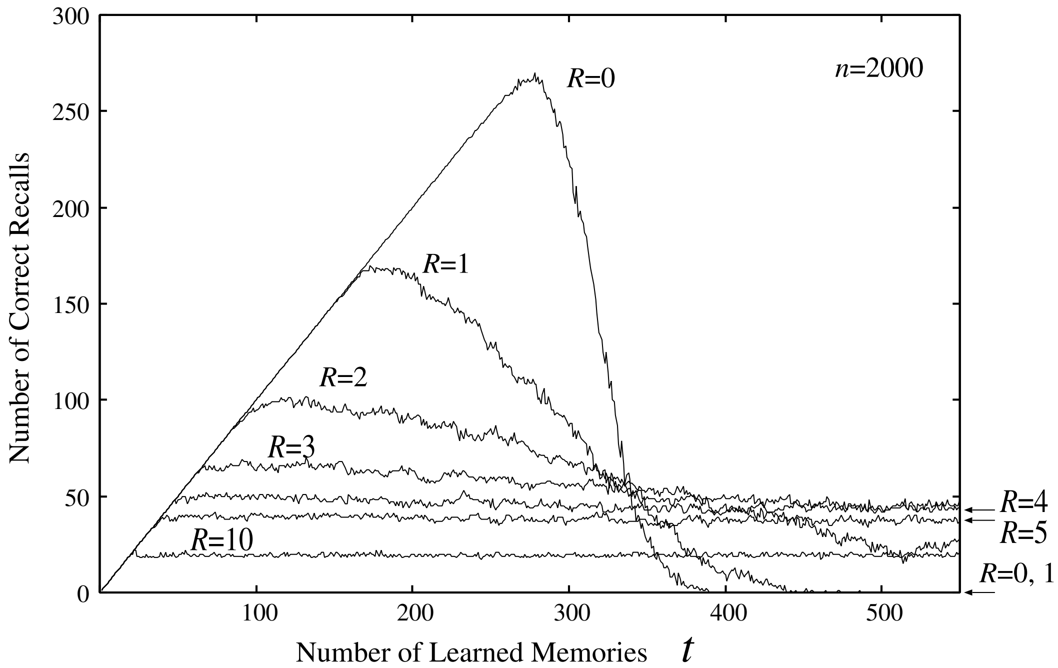

Figure 2: Number of memories correctly recalled as a function of number of learned ones.

$x_j = \pm 1$. The instantaneous state of the system of n units can be thought of as an n-dimensional vector having components x_i. The units are inter-connected by a network of synapses, with a synaptic strength w_{ij} from unit j to unit i. The instantaneous input to unit i is

$$u_i = \sum_{j=1}^{n} w_{ij} x_j \tag{1}$$

where x_j is the present state (± 1) of unit j. The state of the system changes in time; Each unit i readjust its state, setting $x_i = \pm 1$ according to whether u_i, the input to i at that moment is greater or less than zero. The units act asynchronously in a random order. This algorithm defines the time evolution of the state of the system. For any symmetric connection matrix $\{w_{ij}\}$; $w_{ij} = w_{ji}$, there are stable states of the network of units; Starting from any arbitrary initial state, the system reaches a stable state and ceases to evolve [1].

2.2 Content addressable memory

The Hebb rule has served as the starting point for the study of information storage in simplified models. Suppose we wish to store the set of states $\boldsymbol{x}^\mu, \mu = 1, 2, \cdots m$. To learn a new memory $\boldsymbol{x}^1$, increment w_{ij} by

$$\Delta w_{ij} = x_i^1 x_j^1. \tag{2}$$

This learning process is local; the increment for connection w_{ij} does not depend on the global structure of the state or past memories, but only on x_i^1 and x_j^1. It is fast and does not need to learn each memory repeatedly.

This network now functions as an associative memory. If started from an initial state which resembles somewhat state $\boldsymbol{x}^1$ and which resembles other $\boldsymbol{x}^\mu (\mu \neq 1)$ very little, the state will evolve to the state $\boldsymbol{x}^1$. The state $\boldsymbol{x}^1$ is evocable memory, and the system correctly reconstructs an entire memory from any initial partial information, as long as the partial information was sufficient to identify a single memory. Detailed properties of the collective operation of this network have been studied extensively [1, 2, 3, 4, 5].

2.3 Catastrophic forgetting or interference

Computer modeling of memory storage according to equation (2) was carried out for $n = 2000$. 550 random memory states $(\boldsymbol{x}^1, \cdots, \boldsymbol{x}^{550})$ were chosen in each of which half of units were in the active firing state on the average. The network learned one by one by updating w_{ij} incrementally, i.e., new memories were continually added to w_{ij}. There is a memory capacity for this network [1, 2, 5]; About $280 (= 0.14n)$ states can be simultaneously memorized without error in recall. The addition of new memories beyond the capacity overloads the system and makes all memory states irretrievable (catastrophic forgetting) unless

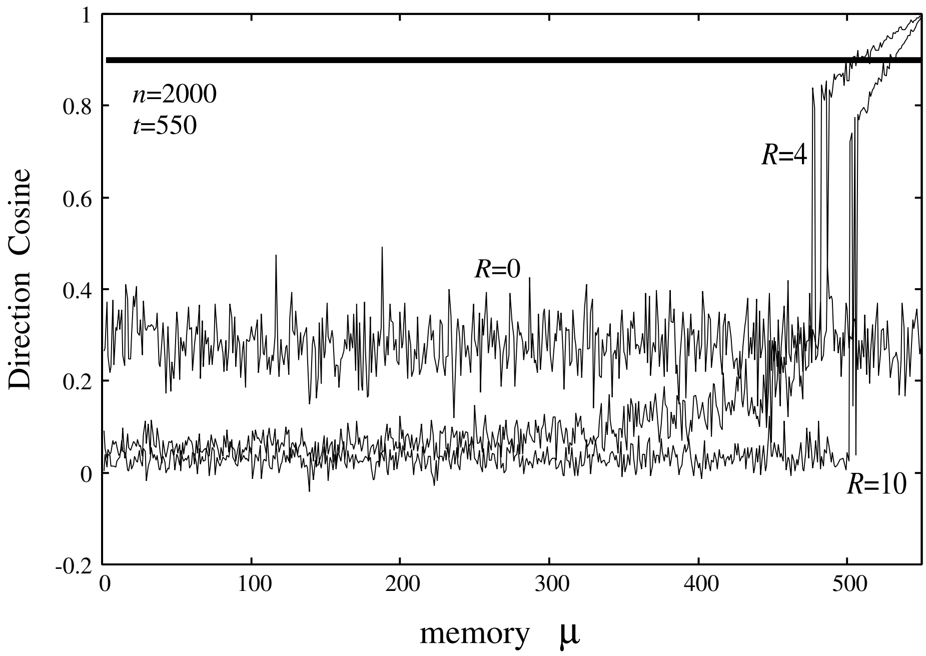

Figure 3: Closeness between each learned state $\boldsymbol{x}^\mu$ and the recalled state for each memory, $\boldsymbol{x}^\mu, \mu = 1, \cdots, 550$.

there is a provision for forgetting old memories.

The network learned 550 memories one by one in which a memory $\boldsymbol{x}^t$ was learned at time t. We tested at each time t, whether network can recall each learned memory without error. Each learned state $\boldsymbol{x}^\mu$, $\mu = 1, \cdots, t$ was given as a starting state to the network. The network quickly reached a stable state $\boldsymbol{x}$. We compared $\boldsymbol{x}^\mu$ to $\boldsymbol{x}$ thus obtained. The proximity of the two states was measured by retrieval overlap

$$a = \frac{1}{n} \sum_{i=1}^{n} x_i^\mu x_i.$$

We counted the number of successfully recalled memories in which the proximity was larger than 0.9. A curve of $R = 0$ in Fig. 1 illustrates how many states there were remembered without error in recall at time t; The network was able to remember all of memorized patterns for $t < 280$, where it was able to learn, But at $t = 280$, forgetting started, and any memories could not be remembered correctly after $t = 380$., i.e., the network could not recall even the most recent memory.

3 Results

For simplicity, we assume the following throughout the present experiment: Every time the network learns a new memory, a number of neurons, say R neurons, die and the same number of neurons are born; we assumed that the total number of neurons does not change in time. It corresponds to reset connection weights $w_{ij} = w_{ji} = 0, j = 1, \cdots, n$ for replaced neurons i. Units were replaced from the oldest one, i.e., they were replaced always in the same order. Simulation were carried out on a computer for $n = 2000$ with varying R, number of replacing units. Results are shown in Fig.1. Larger was the number of replacing units R, earlier the start of forgetting. For $R = 1$, forgetting started at around $t = 180$ and for $R = 3$ it started at around $t = 60$. Our interest was the properties of the network whose structure had already been stable ($t > 500$). The network for $R = 0, 1$ could not recall any memory at $t > 420$. For large $R \geq 2$, the network could recall a number of memories; 44 memories were recalled successfully for $R = 4$.

To see which 44 memories out of 550 were kept successfully, we tested the network at $t = 550$; each learned memory $\boldsymbol{x}^\mu$ was given as a initial state, and the closeness of stable state of the network and memories was measured. Results are shown in Fig.2. As we expected, only the newest memories were recalled correctly. Results of the networks for $R = 4$ and $R = 10$ show that the number of memory successfully retained (a network capacity) was dependent on R, the number of replacing units. It turned out that there was a optimal replacing number near $R = 3$ for $n = 2000$. Computer simulation was carried out for $n = 5000$ with varying R, we found that the optimal replacing number was the same as before, near $R = 3$. The number of success recalls depends on R, and the network for $R \approx 3$ was the optimal to maximize the number, which seems to be independent to the network size.

4 Discussion

To our surprise, the optimal $R \approx 3$ seemed to be independent to the size of the network. The proof of this conjecture might be possible under the following assumption: The present network is equivalent to the network in which weight connections have decay [3, 4] or saturation[1]. The networks share the same properties: avoidance of catastrophic forgetting, keeping recent memories.

The CA3 area of hippocampus is involved in associative memory recall [6, 7], and often modeled by a fully connected network in which each memory is represented by the activity of distributed and sparsely coded pattern. The models proposed so far for explanations of neurogenesis in hippocampus were layered networks which was designed based on known anatomical knowledge [8, 9, 10, 11, 12] in which hypothetical functions were assigned such as making sparse representation, enlarging a dimension of input signal, assigning distinct codes to similar inputs. In our simple system, replacing only a small number of units prevent from the catastrophic forgetting, maximizing a memory capacity. Memories in the hippocampus seems to move into the cortex within 6 weeks. The role of the hippocampus memory systems seems to retain recent limited number of memories. Our example also might answer from a mathematical viewpoint the question of why most newly born neurons die before they mature, and substantial number of replaced neurons observed is so small. It seems that the properties revealed in the present paper does not change by a slight modification of the model. (e.g., sparsely coded memory, the order of turnover)

Acknowledgments

This works was partly supported by Grant-in-Aid (16700232) from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT).

References

[1] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Science USA*, vol. 79, pp. 2554–2558, 1982.

[2] D. J. Amit, H. Gutfreund, and H. Sompolinsky, "Storing infinite numbers of pattern in a spin-glass model of neural networks," *Physical Review Letters*, vol. 55, pp. 1530–1533, 1985.

[3] J. P. Nadal, G. Toulouse, J. P. Changeux, and S. Dehaene, "Network of formal neurons and memory palimpsests," *Europhysics Letters*, vol. 1, pp. 535–542, 1986.

[4] M. Mézard, J. P. Nadal, and G. Toulouse, "Solvable models of working memories," *Journal of Physique*, vol. 47, pp. 1457–1462, 1986.

[5] S. Amari and K. Maginu, "Statistical neurodynamics of associative memory," *Neural Networks*, vol. 1, pp. 63–73, 1988.

[6] B. L. McNaughton and R. G. M. Morris, "Hippocampal synaptic enhancement and information storage within a distributed system," *Trends in Neurosciences*, vol. 10, pp. 408–415, 1987.

[7] K. Nakazawa, M. C. Quirk, R. A. Chitwood, M. Watanabe, M. F. Yeckel, L. D. Sun, A. Kato, C. A. Carr, D. Johnston, M. A. Wilson, and S. Tonegawa, "Requirement for hippocampal CA3 NMDA receptors in associative memory recall," *Science*, vol. 297, pp. 211–218, 2002.

[8] D. Marr, "Simple memory: a theory for archicortex," *Philosophical transactions of the Royal Society of London B*, vol. 262, pp. 23–81, 1971.

[9] D. Willshaw, "An assessment of marr's theory of the hippocampus as a temporary memory store," *Philosophical transactions of the Royal Society of London B*, vol. 329, pp. 205–215, 1990.

[10] S. Becker, "A computational principle for hippocampal learning and neurogenesis," *Hippocampus*, vol. 15, pp. 722–738, 2005.

[11] M. O. M. Guillermo A. Cecchi, "Computational models of adult neurogenesis," *Physica A*, vol. 356, pp. 43–47, 2005.

[12] L. Wiskott, M. Rasch, and G. Kempermann, "A functional hypothesis for adult hippocampal neurogenesis: Avoidance of catastrophic interference in the dentate gyrus," *Hippocampus*, vol. 16, pp. 329–343, 2006.