International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:3, No:12, 2009

Avoiding Catastrophic Forgetting by a Dual-Network Memory Model Using a Chaotic Neural Network

Motonobu Hattori

Abstract—In neural networks, when new patterns are learned by a network, the new information radically interferes with previously stored patterns. This drawback is called catastrophic forgetting or catastrophic interference. In this paper, we propose a biologically inspired neural network model which overcomes this problem. The proposed model consists of two distinct networks: one is a Hopfield type of chaotic associative memory and the other is a multilayer neural network. We consider that these networks correspond to the hippocampus and the neocortex of the brain, respectively. Information given is firstly stored in the hippocampal network with fast learning algorithm. Then the stored information is recalled by chaotic behavior of each neuron in the hippocampal network. Finally, it is consolidated in the neocortical network by using pseudopatterns. Computer simulation results show that the proposed model has much better ability to avoid catastrophic forgetting in comparison with conventional models.

Keywords—catastrophic forgetting, chaotic neural network, complementary learning systems, dual-network.

I. Introduction

It is well known that when a neural network is trained on one set of patterns and then attempts to add new patterns to its repertoire, catastrophic interference, or the complete loss of all of its previously learned information may result [1]-[6]. This type of radical forgetting is unacceptable both for a model of human memory and for practical engineering applications. In order to avoid this implausible failure, numerous researchers have studied on this problem (see [4] for a review). Among them, French [3] and Ans and Rousset [6] independently developed dual-network architectures which are composed of two multilayer neural networks learned by the backpropagation algorithm. Their models are based on the principle of two separate pattern processing areas: one for early-processing and the other for long-term storage. In general, once training patterns have been learned by a network, it is natural to assume that the original patterns are no longer available. So, information is transfered back and forth between two networks by means of pseudopatterns in their dual-network models. They have shown that their models exhibit gradual forgetting by using pseudopatterns.

From a neuropsychological point of view, McClelland, McNaughton and O'Reilly [7] suggested that to alleviate catastrophic forgetting in the human brain two separate areas were evolved: the hippocampus and the neocortex. They suggested that the neocortex may be optimized for the gradual discovery of the shared structure of events and experience, and that the hippocampus provides a mechanism for rapid acquisition of new information and serves as a teacher to the neocortex after the initial acquisition. Since the hippocampus slowly trains the neocortex, new patterns do not interfere with previously stored patterns and are interleaved with them. They called this hippocampal-neocortical system as complementary learning systems.

In this paper, we proposed a novel dual-network memory model inspired by the complementary learning systems theory. Our dual-network memory model is composed of a Hopfield network and a multilayer neural network. Since the Hopfield network use Hebbian learning to store training patterns, it acquires new information quite rapidly. Moreover, information transfer from the Hopfield network to the multilayer neural network is carried out by chaotic recall of the Hopfield network. Owing to this, the original patterns learned by the network may be available for learning of the multilayer neural network. A number of computer simulation results show effectiveness of the proposed dual-network memory model.

II. Conventional dual-network memory model

French [3] and Ans and Rousset [6] independently proposed dual-network memory models. Even though their models differ in a certain number of respects, the essence of them is largely the same. Both models consist of two coupled multilayer neural networks: a hippocampal network and a neocortical network. The hippocampal network is for early-processing and the neocortical network is for long-term storage. That is, a new input is given to the hippocampal network and is stored there with previously learned information at first, then information stored by the hippocampal network is transfered to the neocortical network (memory consolidation). Information is transferred from one network to the other by pseudopatterns (see Fig.1).

In the conventional dual-network models, when a new pattern to be learned is given to the hippocampal network, it is learned with a set of neocortical pseudopatterns. A set of neocortical pseudopatterns is created by a random input and the output of the neocortical network after that input has been sent through it. Since this set of pseudopatterns reflects the previously learned patterns, a new pattern is interleaved with previously learned information. Hence, this technique reduces catastrophic forgetting. In Ans and Rousset's model,

M. Hattori is with Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, JAPAN.
E-mail: *m-hattori@yamanashi.ac.jp,* phone & fax: +81-55-220-8766.

pseudopatterns are also used to transfer the newly learned information from the hippocampal network to the neocortical network. To do this, hippocampal pseudopatterns are created by the hippocampal network and are then learned by the neocortical network.

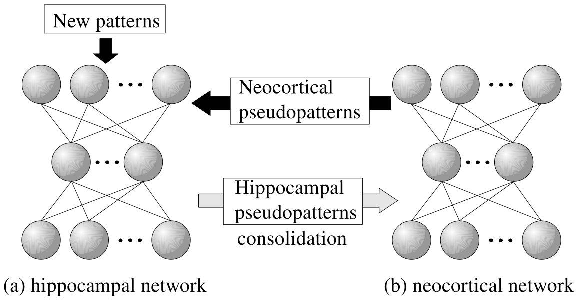

Fig. 1. Structure of the conventional dual-network memory model.

III. Dual-network memory model using chaotic neural network

The fundamental difference between the conventional dual-network memory models and our proposal is that the hippocampal network is now implemented by a chaotic neural network [8]. It is known that chaotic neural networks can dynamically retrieve stored patterns from a random input. Although previously learned original patterns are not available in the conventional models, they may be extracted from chaotic recall in the chaotic neural network. Therefore, we can significantly reduce catastrophic forgetting of originally learned information in the proposed dual-network. Figure 2 shows the structure of the proposed dual-network memory model.

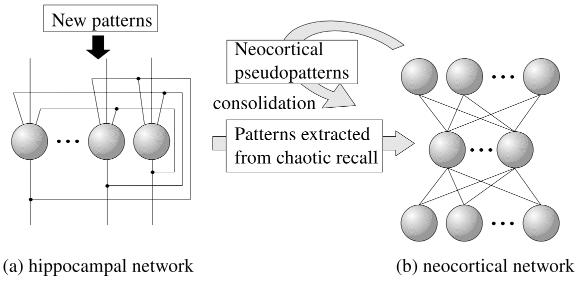

Fig. 2. Structure of the proposed dual-network memory model.

Let $\boldsymbol{X}^{(k)}$ be the kth new pattern to be stored, where $\boldsymbol{X}^{(k)} \in \{-1,1\}^N$ and $\boldsymbol{X}^{(k)} = (X_1^{(k)}, \cdots, X_N^{(k)})^T$. In the proposed model, it is leaned by the hippocampal network using the following Hebbian learning with forgetting:

$$w_{ij}(t+1) = \gamma \cdot w_{ij}(t) + X_i^{(k)} X_j^{(k)} \tag{1}$$

where w_{ij} denotes the connection weights from the jth neuron to the ith neuron and holds $w_{ij} = w_{ji}$ and $w_{ii} = 0$. Since Eq.(1) is a sort of Hebbian learning, the proposed hippocampal network acquires new patterns much more rapidly in comparison with the conventional ones learned by the backpropagation algorithm. The forgetting factor γ in Eq.(1) is a constant between 0 and 1. Owing to the use of γ, only patterns recently given remain in the hippocampal network.

Since the hippocampal network is composed of chaotic neurons, the dynamics of the ith neuron in the hippocampal network is represented by the following equations [8], [9]:

$$x_i(t+1) = f\{\eta_i(t+1) + \zeta_i(t+1)\} \tag{2}$$

$$\eta_i(t+1) = k_m \eta_i(t) + \sum_{j=1}^{N} w_{ij} x_j(t) \tag{3}$$

$$\zeta_i(t+1) = k_r \zeta_i(t) - \alpha x_i(t) + a_i \tag{4}$$

where $x_i(t+1)$ shows the output of the ith neuron at $t+1$, k_m and k_r are damping factors of refractoriness, α is a scaling factor of the refractoriness, a_i is an external input parameter, and $f(\cdot)$ show the following output function:

$$f(u) = \frac{1}{1+\exp(-u/\epsilon)} \tag{5}$$

where ϵ is the steepness parameter.

In the chaotic neural network, states of the network tend to remain in trained patterns for a relatively long period during chaotic recall. Therefore, we can extract stored patterns by a random input, observing chaotic recall and choosing states recalled for a long period.

Then extracted patterns are learned by the neocortical network with neocortical pseudopatterns. Here we propose to use two kinds of neocortical pseudopatterns: neocortical pseudopatterns I and II. One set of pseudopatterns, neocortical pseudopatterns I is created by the conventional manner: a random input and the output of the neocortical network. In contrast, neocortical pseudopatterns II is created as follows:

1) Reverse each element of the pattern extracted from the hippocampal network with a certain probability, P.
2) Give the pattern 1) to the neocortical network and obtain the output.
3) Repeat 1) and 2) until the predefined number of pairs of input and output is obtained. Make the obtained set neocortical pseudopatterns II.

Learning neocortical pseudopatterns II together may preserve information especially interfered by extracted patterns from the hippocampal network.

IV. Computer simulation results

In computer simulation, we used the following parameters: $\gamma = 0.7$ for Eq.(1), and $k_m = 0.10$, $k_r = 0.95$, $\alpha = 2.7$, $a_i = 0.8$ and $\epsilon = 0.1$ for Eqs.(2)-(5). The learning rate and the coefficient for the momentum term of the backpropagation algorithm was set to 0.01 and 0.9, respectively. Training patterns used are shown in Fig.3.

As shown in the figure, we used the same pattern as the input and the output. That is, dual-network memory models learned the identity map of a set of training patterns. Every time, two patterns are given to the hippocampal network and learned. In total, four pairs of patterns (*i.e.* eight patterns) were sequentially learned by the dual-network models.

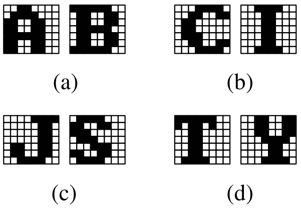

Fig. 3. Training patterns. {A,B}, {C,I}, {J,S} and {T,Y} are sequentially given to the hippocampal networks in order.

In oder to evaluate the network's ability to correctly reproduce the appropriate outputs for a given set of inputs, we define a performance measure, *goodness*. Let s_i be the bipolarized value of the ith output neuron of the neocortical network: we regard output activities less than and greater than 0 as bipolar values -1 and 1, respectively, and let t_i be the corresponding component of the desired pattern. Then the goodness g is defined as:

$$g = \frac{1}{N}\sum_{i=1}^{N} g_i \tag{6}$$

$$g_i = \begin{cases} 1 & \text{if } s_i = t_i \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where N is the number of output neurons. A goodness value of 1 indicates a perfect match between the calculated output and the desired one.

A. Performance of the conventional dual-network memory model

Figure 4 shows the mean goodness based on 20 trails when we varied the number of hippocampal pseudopatterns and that of the neocortical ones from 10 to 40.

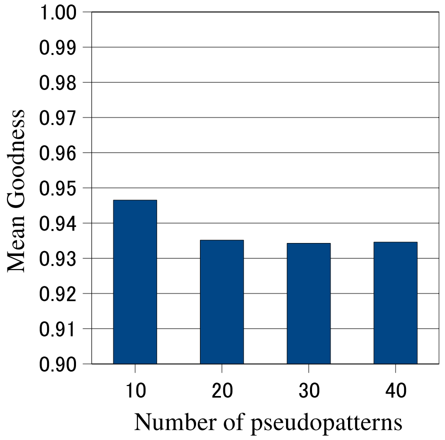

Fig. 4. Mean goodness of the conventional dual-network memory model.

As shown in Fig.4, the conventional dual-network memory model can reduce catastrophic forgetting to a certain extent. However, contrary to the results in [3], the performance was not improved even though we increased the number of pseudopatterns.

Figure 5 shows the best recall result of the conventional model in the experiment: goodness was 0.973. The ability to avoid forgetting was not particularly well in the early items such as A, B and C.

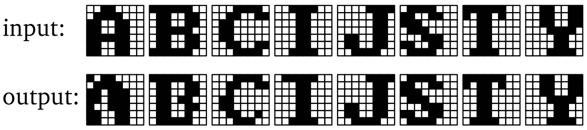

Fig. 5. The best recall result of the conventional dual-network memory model: goodness= 0.973. 10 pseudopatterns were used for both the hippocampal and the neocortical network.

B. Performance of the proposed dual-network memory model

In the proposed hippocampal network, in oder to avoid inverted version of training patterns being recalled, we added 24 elements that took -1 to each training pattern when learning. Then, states of the corresponding 24 neurons in the hippocampal network were cramped at -1 during chaotic recall (see Fig.6). These additional elements were removed when extracted patterns were sent to the neocortical network.

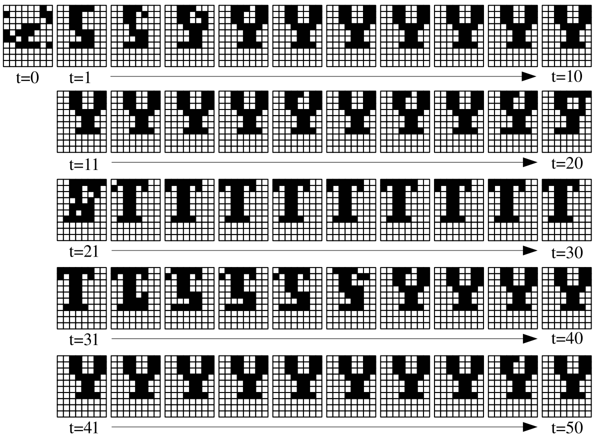

Fig. 6. An example of chaotic recall of the hippocampal network in the proposed dual-network memory model.

In chaotic recall of the proposed model, we incremented the time t after all neurons updated theirs states asynchronously. To extract patterns from chaotic recall, a random input is given to the hippocampal network, and then we examined each output until $t = 50$ after bipolarizing it. We extracted patterns when the bipolarized outputs were unchanged more than 5 times. Figure 6 shows an example of chaotic recall of the proposed hippocampal network after the forth pair of training patterns {Y,T} was learned. As seen in the figure, states of the network remain in the trained patterns for a long time. In addition, owing to the use of the forgetting factor γ, only patterns recently learned were recalled.

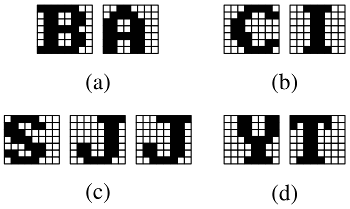

Fig. 7. An example of patterns extracted from chaotic recall in the proposed dual-network memory model.

Figure 7 shows an example of extracted patterns from chaotic recall. Although a pattern which was slightly different from the corresponding training pattern J was additionally extracted in Fig.7(c), all training patterns were extracted successfully.

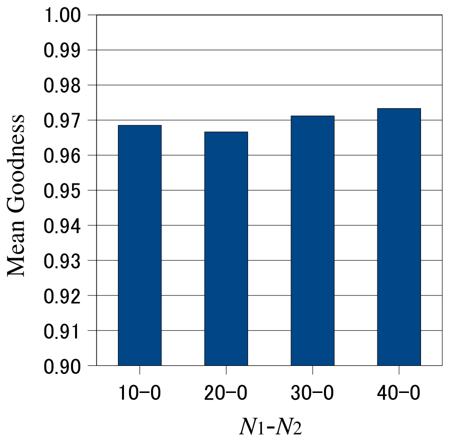

Fig. 8. Mean goodness of the proposed dual-network memory model when only pseudopatterns I was used.

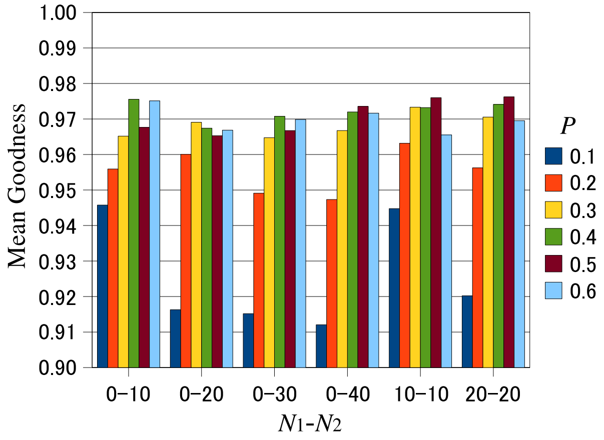

Fig. 9. Mean goodness of the proposed dual-network memory model when pseudopatterns II was used.

Figure 8 shows the mean goodness based on 20 trails when only pseudopatterns I was used. In Fig.8, N_1 and N_2 show the number of pseudopatterns I and that of pseudopatterns II, respectively. In contrast to the results of the conventional model shown in Fig.4, the mean goodness was much improved by using the chaotic neural network as the hippocampal network.

Figure 9 shows the mean goodness of the proposed model based on 20 trials when pseudopatterns II was used. In this figure, P denotes the probability of reversing each element of the pattern extracted from the hippocampal network. In this experiment, the highest mean goodness value was 0.976 when $(N_1, N_2) = (20, 20)$ and $P = 0.5$. We can see that using pseudopatterns II in the proposed model even improves the mean goodness in comparison with the results in Fig.8. Moreover, as shown in Fig.9, using both pseudopatterns I and II may be more effective and a desirable value of P may be between 0.4 and 0.5.

Figure 10 shows the best recall result of the proposed model in the experiment: goodness was 0.998. In contrast to the result of the conventional model shown in Fig.5, the proposed model avoid catastrophic forgetting very well especially for the patterns trained early.

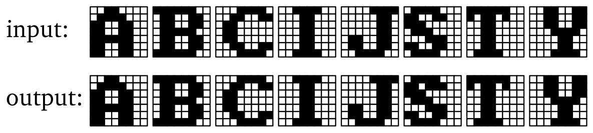

Fig. 10. The best recall result of the proposed dual-network memory model: goodness= 0.998. N_1–N_2 = 20–20 and $P = 0.5$.

V. Conclusions

In this paper, we have proposed a novel dual-network memory model inspired by the complementary learning systems theory [7]. The proposed model consists of two distinct networks: one is a Hopfield type of chaotic associative memory and the other is a multilayer neural network. These networks corresponds to the hippocampus and the neocortex of the brain, respectively. The proposed model have the following features:

1) Since the hippocampal network is learned by Hebbian learning, new information is stored quite rapidly.
2) Using the forgetting factor γ, the hippocampal network tends to store only recent patterns. That is, the hippocampal network works as a short term memory, while the neocortical network learned by the backpropagation algorithm works as a long term memory.
3) Since information transfer from the hippocampal network to the neocortical network is carried out by using chaotic recall of the hippocampal network, original patterns learned by the hippocampal network may be available for learning of the neocortical network. This can much contribute to reduce catastrophic forgetting.
4) Using both pseudopatterns I and II even reduce catastrophic forgetting.

As is well known, Hopfield networks learned by Hebbian learning have only a small storage capacity. In the proposed model, however, since the hippocampal network works as a short term memory and patterns extracted are sent to the neocortical network in a certain short period, a small storage capacity of the hippocampal network doesn't become a problem in the proposed model. Although the hippocampus and the neocortex is not strictly modeled in the proposed architecture, using the recurrent structure in the hippocampal network seems to be more plausible than the conventional network because the hippocampal CA3 has recurrent connections. In addition, information transfer from the hippocampal network to the neocortical network by chaotic recall might relate to the memory consolidation during sleep [10]. We have already investigated effects of the hippocampal neurogenesis in a hippocampal model [11]. In the future research, we will introduce the neuronal turnover into our dual-network memory model.

Acknowledgment

This work was supported by Grant-in-Aid for Scientific Research (19700217) to Dr. Hattori for the Ministry of Education, Culture, Sports, Science and Technology in Japan. The author would like to thank Mr. Kodai Saito for his experiments on the prototype of the proposed dual-network memory model.

References

[1] R. M. French, "Using semi-distributed representation to overcome catastrophic forgetting in connectionist network," *Proceedings of the 13th Annual Cognitive Science Society Conference*, pp.173-178, 1991.

[2] R. M. French, "Dynamically constraining connectionist networks to produce distributed, orthogonal representations to reduce catastrophic interference," Proceedings of the 16th Annual Cognitive Science Society Conference, pp.335-340, 1994.

[3] R. M. French, "Pseudo-recurrent connectionist networks: An approach to the "sensitivity-stability" dilemma," *Connection Science*, vol.9, no.4, pp.353-379, 1997.

[4] R. M. French, "Catastrophic forgetting in connectionist networks," *Trends in Cognitive Sciences*, vol.3, no.4, pp.128-135, 1997.

[5] R. M. French, B. Ans and S. Rousset, "Pseudopatterns and dual-network memory models: Advantages and shortcomings," In *Connectionist Models of Learning, Development and Evolution* (R. French and J. Sougné eds.), London, Springer, pp.13-22, 2001.

[6] B. Ans and S. Rousset, "Avoiding catastrophic forgetting by coupling two reverberating neural networks", *Academie des Sciences, Sciences de la vie*, vol.320, pp.989-997, 1997.

[7] J. McClelland, B. McNaughton and R. O'Reilly, "Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory", *Psychological Review*, vol.102, no.3, pp.419-457, 1995.

[8] K. Aihara, T. Takabe and M. Toyoda, "Chaotic Neural Networks," *Physics Letters A*, vol.144, no.6-7, pp.333-340, 1990.

[9] Y. Osana, M. Hattori and M. Hagiwara, "Chaotic Bidirectional Associative Memory," *Proceeding of the International Joint Conference on Neural Networks*, Washington D.C., vol.2, pp.816-821, 1996.

[10] G. Buzsáki, "Memory consolidation during sleep: a neurophysiological perspective," *Journal of Sleep Research*, vol.7, issue S1, pp.17-23, 1998.

[11] Y. Wakagi and M. Hattori, "A Model of Hippocampal Learning with Neuronal Turnover in Dentate Gyrus," *International Journal of Mathematics and Computers in Simulation*, issue 2, vol.2, pp.215-222, 2008.

[12] R. C. O'Reilly and J. W. Ruby, "Computational Principles of Learning in the Neocortex and Hippocampus," *HIPPOCAMPUS*, vol.10, pp.389-397, 2000.

[13] K. A. Norman and R. C. O'Reilly, "Modeling Hippocampal and Neocortical Contribution to Recognition Memory: A Complementary-Learning-Systems Approach," *Psychological Review*, vol.110, no.4, pp.611-646, 2003.

Motonobu Hattori was born in Tokyo, Japan on February 11, 1970. He received B.E., M.E. and Ph.D degrees in Electrical Engineering from Keio University in 1992, 1994 and 1997, respectively.

In 1997, he became a research associate of University of Yamanashi, Kofu, Japan. Since 2000, he has been an Associate Professor. From 2003 to 2004, he was a visiting researcher at Center for the Neural Basis of Cognition (CNBC) in Carnegie Mellon University, USA. His current research interest is in neural networks, reinforcement learning and evolutionary computation.

Dr. Hattori is a member of the IEEE, IEICE and IEE. He received Niwa Memorial Award in 1996.