Neurocomputing 134 (2014) 262–268

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A biologically inspired dual-network memory model for reduction of catastrophic forgetting

Motonobu Hattori

4-3-11 Takeda, Kofu, Yamanashi 400-8511, Japan

ARTICLE INFO

Article history:
Received 3 April 2013
Received in revised form
27 August 2013
Accepted 30 August 2013
Available online 14 January 2014

Keywords:
Catastrophic forgetting
Chaotic neural network
Complementary learning systems
Dual-network
Hippocampus
Neuronal turnover

ABSTRACT

Neural networks encounter serious catastrophic forgetting when information is learned sequentially, which is unacceptable for both a model of human memory and practical engineering applications. In this study, we propose a novel biologically inspired dual-network memory model that can significantly reduce catastrophic forgetting. The proposed model consists of two distinct neural networks: hippocampal and neocortical networks. Information is first stored in the hippocampal network, and thereafter, it is transferred to the neocortical network. In the hippocampal network, chaotic behavior of neurons in the CA3 region of the hippocampus and neuronal turnover in the dentate gyrus region are introduced. Chaotic recall by CA3 enables retrieval of stored information in the hippocampal network. Thereafter, information retrieved from the hippocampal network is interleaved with previously stored information and consolidated by using pseudopatterns in the neocortical network. The computer simulation results show the effectiveness of the proposed dual-network memory model.

1. Introduction

When a neural network is trained on one set of patterns and then when it later attempts to add new patterns to its repertoire, catastrophic interference or complete loss of all of its previously learned information may result. This type of radical forgetting is unacceptable for both a model of human memory and practical engineering applications.

McClelland et al. suggested that nature's way of avoiding catastrophic forgetting was the evolution of two separate areas of the brain: the hippocampus and neocortex [1]. They suggested that the neocortex may be optimized for gradual discovery of the shared structure of events and experiences, and that the hippocampal system is there to provide a mechanism for rapid acquisition of new information without interference with previously discovered regularities. In addition, they suggested that the hippocampal system serves as a teacher to the neocortex. This framework is called the complementary learning systems (CLS). Although McClelland et al. did not show a specific model for reducing catastrophic forgetting, French [2] and Ans and Rousset [3] independently developed dual-network architectures in the light of the CLS framework and *pseudorehearsal* proposed by Robins [4]. Their models are composed of two multilayer neural networks and implement the pseudorehearsal mechanism with an entire neural network system; information is transferred back and forth between two networks by means of pseudopatterns in their dual-network models. Although they showed that catastrophic forgetting can be suppressed, their models do not resemble the biological dual-network system, i.e., (1) their models are composed of two identical networks, and (2) because both networks learn with the backpropagation algorithm, their models are not suited for rapid acquisition of new information and both require a mechanism to avoid catastrophic forgetting.

In addition, we have previously proposed a dual-network memory model [5] inspired by the CLS theory [1]. This model is composed of two distinct networks: a hippocampal network for early processing and a neocortical network for long-term storage. This model employs a chaotic neural network [6] as the hippocampal network, and the information stored by the hippocampal network is transferred to the neocortical network by chaotic recall of the hippocampal network. Because previously learned original patterns can be extracted with chaotic recall, we have shown that our dual-network model can significantly reduce catastrophic forgetting. However, in this hippocampal network, we have to add some extra elements that take a "-1" value to each training pattern in order to avoid inverted versions of training patterns being recalled (discussed further in Section 2). This restriction is biologically implausible and we must use trial and error to determine how many elements must be added. Furthermore, the hippocampal network in our work is a Hopfield network (with chaotic neurons), and its storage capacity is extremely low. Therefore, recently, we proposed a much more biologically plausible hippocampal network and demonstrated its superior performance

E-mail addresses: m-hattori@yamanashi.ac.jp, moto-h@nifty.com

0925-2312/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.08.044

for extracting stored patterns and large storage capacity [7]. However, we did not investigate its characteristics as a dual-network memory. Hence, in this paper, we propose a novel dual-network memory model by replacing the original simple hippocampal network with a much more biologically plausible one [7], and compare its performance with that of the conventional model.

The rest of this paper is organized as follows. In Section 2, we briefly review the conventional dual-network memory model [5]. Section 3 explains the proposed dual-network memory model. In Section 4, we show the experimental results. Finally, conclusions are given in Section 5.

2. Conventional dual-network memory model using a chaotic neural network

2.1. Outline

Fig. 1 shows the structure of our conventional dual-network memory model [5]. It consists of two coupled neural networks: a hippocampal network, composed of chaotic neurons [6], and a neocortical network. Introducing chaotic neurons to the hippocampal network is not biologically implausible because the chaotic activities of the CA3 region have been observed [8]. In our conventional dual-network model, a new input is provided to the hippocampal network and is stored there initially, and then, this information stored by the hippocampal network is transferred to the neocortical network (memory consolidation).

Chaotic neural networks can dynamically retrieve stored patterns from a random input. Hence, previously learned original patterns can be extracted by chaotic recall in the hippocampal network. To transfer information from the hippocampal network to the neocortical one without severe catastrophic forgetting, extracted patterns are learned by the neocortical network with 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 processed. Because this set of pseudopatterns reflects the previously learned patterns, new patterns extracted from the hippocampal network are interleaved with previously learned information. Thus, catastrophic forgetting can be reduced.

2.2. Learning of hippocampal network

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

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

where w_{ij} denotes the connection weights from the jth neuron to ith neuron and holds $w_{ij} = w_{ji}$ and $w_{ii} = 0$. Because Eq. (1) is a form of Hebbian learning, the proposed hippocampal network acquires new patterns much more rapidly in comparison with the conventional ones learned by the backpropagation algorithm [2,3]. The forgetting factor γ in Eq. (1) is a constant between 0 and 1. Because of the use of γ, only recent patterns remain in the hippocampal network.

Because the hippocampal network is composed of chaotic neurons, the dynamics of the ith neuron in the hippocampal network is represented by the following equations [6,9]:

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

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

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

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

$$f(u) = \tanh\left(\frac{u}{\varepsilon}\right), \tag{5}$$

where ε is the steepness parameter.

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

2.3. Learning of neocortical network

The extracted patterns from the hippocampal network are learned by the neocortical network with neocortical pseudopatterns. In the conventional model, two types of neocortical pseudopatterns are used: neocortical pseudopatterns I and II [5]. Neocortical pseudopattern I is created in the conventional manner [2,3], with a random input and the corresponding output of the neocortical network. In contrast, neocortical pseudopattern 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 from (1) to the neocortical network and obtain the output.
3. Repeat (1) and (2) until a predefined number of pairs of inputs and outputs are obtained. Define the obtained set as neocortical pseudopattern II.

Learning neocortical pseudopattern II together may preserve information that would otherwise be interfered by the extracted patterns from the hippocampal network. In Ref. [5], we have shown that using both pseudopatterns I and II is more effective than using only one or the other.

2.4. Biologically implausible restriction

In the conventional dual-network memory model, it was necessary to add "−1"(white) elements to each training pattern [5]. Fig. 2 shows an example of training patterns used for the conventional hippocampal network. To avoid inverted versions of training patterns being recalled, these extra "−1" elements have to be cramped during recall. This restriction is not only biologically implausible but also undesirable because we must use trial and error to determine the number of extra elements. Moreover, because the conventional hippocampal network is a Hopfield

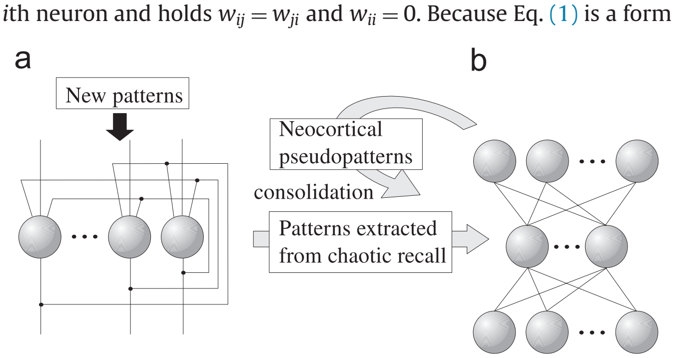

Fig. 1. Structure of the conventional dual-network memory model [5]: (a) hippocampal network and (b) neocortical network.