Neural Networks 22 (2009) 687–696

Contents lists available at ScienceDirect

Neural Networks

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

2009 Special Issue

Online unsupervised formation of cell assemblies for the encoding of multiple cognitive maps

Utku Salihoglu[a,*], Hugues Bersini[a], Yoko Yamaguchi[b], Colin Molter[b,*]

[a] *IRIDIA-CoDE, Universite Libre de Bruxelles, 50, Av. F. Roosevelt, CP* 194/6 *1050 Brussels, Belgium*

[b] *Laboratory for Dynamics of Emergent Intelligence, RIKEN Brain Science Institute, 2-1 Hirosawa Wako, Saitama 351-0198, Japan*

ARTICLE INFO

Article history:
Received 5 May 2009
Received in revised form 30 May 2009
Accepted 25 June 2009

Keywords:
Cell assemblies
Retroaxonal learning
Hebbian learning
Recurrent network

ABSTRACT

Since their introduction sixty years ago, cell assemblies have proved to be a powerful paradigm for brain information processing. After their introduction in artificial intelligence, cell assemblies became commonly used in computational neuroscience as a neural substrate for content addressable memories. However, the mechanisms underlying their formation are poorly understood and, so far, there is no biologically plausible algorithms which can explain how external stimuli can be online stored in cell assemblies.

We addressed this question in a previous paper [Salihoglu, U., Bersini, H., Yamaguchi, Y., Molter, C., (2009). A model for the cognitive map formation: Application of the retroaxonal theory. In *Proc. IEEE international joint conference on neural networks*], were, based on biologically plausible mechanisms, a novel unsupervised algorithm for online cell assemblies' creation was developed. The procedure involved simultaneously, a fast Hebbian/anti-Hebbian learning of the network's recurrent connections for the creation of new cell assemblies, and a slower feedback signal which stabilized the cell assemblies by learning the feedforward input connections.

Here, we first quantify the role played by the retroaxonal feedback mechanism. Then, we show how multiple cognitive maps, composed by a set of orthogonal input stimuli, can be encoded in the network. As a result, when facing a previously learned input, the system is able to retrieve the cognitive map it belongs to. As a consequence, ambiguous inputs which could belong to multiple cognitive maps can be disambiguated by the knowledge of the context, i.e. the cognitive map.

1. Introduction

During the past sixty years, despite seminal observations suggesting the existence and the importance of complex dynamics in the brain (Babloyantz & Destexhe, 1986; Skarda & Freeman, 1987), fixed point dynamics has been the predominant regime used to describe brain information processing (Grossberg, 1992, for a review). More recently, the increasing power of computers and the development of new statistical mathematics demonstrated less equivocally the necessity to rely on more complex dynamics (e.g. Kenet, Bibitchkov, Tsodyks, Grinvald & Arieli, 2003). In that view, by extending classical Hopfield networks to encode cyclic attractors, the authors demonstrated that cyclic and chaotic dynamics could encompass several limitations of fixed point dynamics (Molter, Salihoglu & Bersini, 2007).

Still, the nature of the information processed was not defined. More than fifty years ago, Hebb proposed the cell assembly theory of cortical associative memory (Hebb, 1949). In this theory, each memory is defined by a cell assembly, i.e. a set of cells having strong synaptic weights between each other due to the well-known Hebbian rule of synaptic plasticity. The functional principles underlying that theory of memory has been formalized mathematically as attractor neural networks and is still a working concept in the neuroscience community for the understanding of how the brain works. The first part of this paper tries to conciliate these two views and propose the encoding of information in predefined cell assemblies, here noted CA(s), characterized by complex dynamics by relying on a very simple rate firing model.

To validate our model, two features are tested: first, the ability to recover the full information from partial stimulation (content addressability); second, the ability to maintain a memory of the stimulus in the network's dynamics (working memory). The working memory appears as a fundamental component in the realization of higher cognitive functions, and defines the ability

---

* Corresponding author.
E-mail addresses: usalihog@iridia.ulb.ac.be (U. Salihoglu), cmolter@brain.riken.jp (C. Molter).

0893-6080/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2009.06.025

to hold and manipulate limited amounts of information during short time periods (Baddeley & Hitch, 1974). The neural basis of the working memory has been widely investigated in primates with single cell recordings (Fuster, 1973; Fuster & Alexander, 1971; Rainer, Asaad & Miller, 1998) and neuro-imaging tools (Cohen et al., 1997). It was demonstrated that some of the cells which were responsive to the stimulus maintained their activity during a short period after stimulus offset. In response, several computational models have already shown that cell assemblies could work as working memory by actively holding a limited amount of information for a short time (e.g. Compte, Brunel, Goldman-Rakic & Wang, 2000; Durstewitz, Seamans & Sejnowski, 2000; Molter, Colliaux & Yamaguchi, in press; Mongillo, Barak & Tsodyks, 2008). Similarly, in our model, after stimulus removal, the dynamics can remain in a specific attractor. The novelty in our model will be the presence of strange attractors reached for ambiguous stimuli and leading to itinerancy among several attractors.

The second part of the paper addresses an issue which was not raised in the previous working memory models: how to form these cell assemblies. To this end, we propose a novel unsupervised algorithm which creates cell assemblies based on the external stimulus. The procedure combines two biologically plausible mechanisms. First, the rapid Hebbian/anti-Hebbian learning of the network's recurrent connections to create the cell assemblies. Second, a slow feedback mechanism to organize the incoming connections for the stabilization (or destruction) of the cell assemblies. This retroaxonal feedback has been observed to occur on several levels in the brain (Buss, Sun & Oppenheim, 2006; Hamburger, 1992, 1993; Oppenheim, 1991) and has recently been suggested as a plausible mechanism for stabilizing neuronal activity (Harris, 2008). Results show that the obtained CAs exhibit similar behavior as the pre-encoded ones.

This algorithm is reminiscent of a long tradition of models promoting the unsupervised self organization of information in neural networks, such as the adaptive resonance theory (e.g. Carpenter & Grossberg, 1988; Grossberg, 1993) or the self organizing maps (e.g. Kohonen, 1982, 2001). However, our method differs radically regarding the nature of the dynamics expected. While in the former models the successful encoding/retrieval of information was characterized by simple dynamics (usually in the form of fixed point attractors), here, following the view that the presence of chaotic dynamics can boost the network's capacity (Molter & Bersini, 2003), complex dynamics was enforced as much as possible.

Finally, this model is proposed as a working paradigm for the formation of a cognitive map. In that view, a map results from the juxtaposition of the several cell assemblies associated with the environment's set of stimuli. To learn multiple maps, a context layer is added. The network can then recover the map to which a stimulus belongs. The context knowledge (from the previous stimuli or top-down control) can then help to identify noisy stimuli and more precisely can help to disambiguate external stimuli which could be associated with CAs from multiple maps. This model of cognitive map formation based on the creation of cell assemblies is an alternative to the view that a cognitive map is characterized by a continuous attractor (McNaughton, Battaglia, Jensen, Moser, E. & Moser, M. B, 2006; Samsonovich & McNaughton, 1997).

2. Network model architecture

2.1. Architecture and activation function

The basic structure of the network is constituted by two layers of neurons (Fig. 1). The first layer contains M units and represents the external stimulus. This layer feeds through the feedforward connections W_S the second layer of N neurons. That second layer, called the associative layer, contains recurrent connections W_R and accordingly, can be seen as a model of the dorsal CA3 network. As an extension of our previous model (Salihoglu, Bersini, Yamaguchi & Molter, 2009), an additional context layer of P units has been added. That layer receives connections from the associative layer (W_C) and feeds it in return (W_V).

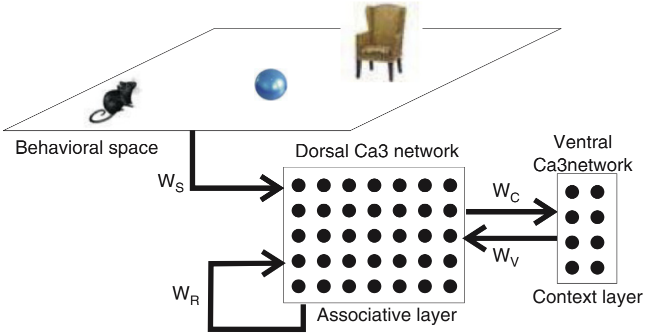

Fig. 1. Model architecture. The associative layer, modelling the dorsal CA3 network, receives feedforward connections from external stimuli. Additionally, the associative layer is linked to a context layer, which could be the ventral CA3 network.

Each cell of the system is represented by simple McCulloch and Pitts neurons evolving in discrete time step (McCulloch & Pitts, 1943). The network's update rule was chosen to be synchronous.[1] This enabled the use of matrix computation, and the state at time n of the cells in the associative layer ($\mathbf{X(n)}$) and in the context layer ($\mathbf{C(n)}$) is given by :

$$\begin{aligned} \mathbf{X}(n) &= F\left(W_R\mathbf{X}(n-1) + W_S\mathbf{I}(n-1) + W_V\mathbf{C}(n-1)\right) \\ \mathbf{C}(n) &= G\left(W_C\mathbf{X}(n-1)\right) \end{aligned} \tag{1}$$

where W_R, W_S, W_C and W_V are weight matrix, and $\mathbf{I}(n)$ is the input vector at time n.

Units in the associative layer used a sigmoid transfer function:

$$f(x_i) = \frac{\tanh(3x_i - 2) + 1}{2} \tag{2}$$

with the coefficients chosen such that the firing rate is bounded ($x \in [0, 1]$) and is nearly equal to zero in absence of any inputs ($f(0) \approx 0$; which means that no short term memory feature was implemented at the cell level). F is the vector formulation of the function f.

In the context layer, a multiple winner take all activation function was chosen. If we note K_i the total activity impinging a context cell c_i, the activity of that cell at time n is given by:

$$c_i(n) = \begin{cases} 1 & \text{if } \left(\dfrac{K_i(n)}{\sum_i^P K_i(n)}\right)^{10} > 0.1 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where P is the total number of cells in the context layer. This transfer function defines the function G in Eq. (1). The exponentiation enhances distance between good candidates and averaged ones.

---

[1] To enhance complex dynamics and to speed up the computations, (the convergence theorem to fixed point attractors in Hopfield networks required an asynchronous update rule (Hopfield, 1982)).