Neural Networks 22 (2009) 738–747

Contents lists available at ScienceDirect

Neural Networks

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

2009 Special Issue

Here and now: How time segments may become events in the hippocampus

András Lőrincz*, Gábor Szirtes

Department of Software Technology and Methodology, Eötvös Loránd University, Pázmány sétány 1/C, Budapest, H-1117, Hungary

ARTICLE INFO

Article history:
Received 6 May 2009
Received in revised form 25 May 2009
Accepted 25 June 2009

Keywords:
Place cells
Grid cells
Hippocampus
Independent Process Analysis
Time series model

ABSTRACT

The hippocampal formation is believed to play a central role in memory functions related to the representation of events. Events are usually considered as temporally bounded processes, in contrast to the continuous nature of sensory signal flow they originate from. Events are then organized and stored according to behavioral relevance and are used to facilitate prediction of similar events. In this paper we are interested in the kind of representation of sensory signals that allows for detecting and/or predicting events. Based on new results on the identification problem of linear hidden processes, we propose a connectionist network with biologically sound parameter tuning that can represent causal relationships and define events. Interestingly, the wiring diagram of our architecture not only resembles the gross anatomy of the hippocampal formation (including the entorhinal cortex), but it also features similar spatial distribution functions of activity (localized and periodic, 'grid-like' patterns) as found in the different parts of the hippocampal formation. We shortly discuss how our model corresponds to different theories on the role of the hippocampal formation in forming episodic memories or supporting spatial navigation. We speculate that our approach may constitute a step toward a unified theory about the functional role of the hippocampus and the structure of memory representations.

1. Introduction

Although our senses receive an enormous amount of information at every time instant, we have the remarkable ability to filter out, organize and store only those pieces of information that might be relevant from behavioral, physical or cognitive aspects. In addition, sensory information processing is believed to facilitate *prediction* (Bialek, Nemenman, & Tishby, 2001) of behaviorally relevant changes of observations including both internal and external variables. How this prediction actually works is an open question, but it is generally assumed that it is based on the creation of internal representations of lower complexity. In the temporal domain such condensed representation may lead to the notion of events. An *event* may intuitively be defined as a primary cause that results in temporally bounded change of a given state or condition behind the observations. For example, an animal may be still or moving fast, and anything (detecting a predator or a potential mating partner) that can trigger a switch between these states could be seen as an event. By learning causal relationships between events and the resulting changes, it becomes possible to predict succeeding states by detecting a particular event. However, in contrast to the widely used coding mechanisms where codewords can easily be distinguished, we receive a continuous flow of sensory signals. Due to the limited memory capacity, an efficient segmentation mechanism is required to help encode the incoming signals. In Section 2 we formalize our assumptions on the sensory signals and – based on the notion of *statistical independence* – we show how the resulting model can be used for segmentation. Statistical independence is often coupled with sparse coding (Földiák, 2002; Olshausen & Field, 1997; Seeger, 2008), which has been suggested as the underlying neural mechanism for optimal reconstruction regarding redundancy reduction and stimulus reconstruction. What it means is that most (natural) stimuli can be decomposed into a finite set of features which are sparse (that is they are 'silent' in most of the time), but when they can be detected, their contribution to the overall signal is quite important. This sparsity can then be used as a time stamp that marks the start and end points of state transition processes. In Section 3 we translate the algorithms into a connectionist network in which parameter tuning can only be realized via biologically plausible local interactions. In Section 4 we briefly sketch the functional parallels between our computational model and the hippocampal region (HR). For decades, hippocampus research has been dedicated to study either its role in episodic memory or its role in spatial navigation (Eichenbaum, 2000; O'Keefe & Nadel, 1978; Vargha-Khadem et al., 1997). Recently, a new line of research has emerged that attempts to unite the seemingly orthogonal theories developed in these two parallel tracks (e.g. Mizumori (2006)). Our model is essentially a signal encoding system dealing with time series and representations; thus it seems to belong to the first venue. To see how it behaves when there is space related

* Corresponding author. Tel.: +36 20 324 2453; fax: +36 1 381 2140.
E-mail addresses: andras.lorincz@elte.hu (A. Lőrincz), szirtes@inf.elte.hu (G. Szirtes).

0893-6080/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2009.06.020

information in the sensory signals (which is of fundamental importance in spatial navigation problems), we present some simulation results in Section 5 about the dynamics of the model when applied on inputs with explicit spatial dependence. Finally, in Section 6 we shortly discuss the results as well as the limitations of the current model and propose a mechanism that may support and extend our model even if such explicit spatial dependence of the inputs cannot be assumed. One of the most interesting issues is the relation between our information theoretically motivated model and theories on the sensory-motor integration process proposed to form predictive internal models of the world. We also formalize some predictions on the functioning and dynamics of the HR. A short version of this paper with a different emphasis has been accepted at IJCNN'09 (Lőrincz & Szirtes, 2009).

2. Identification of the sensory input

It is natural to interpret the observations ($x(t) \in \mathbf{R}^d$) as mixed signals emitted by a hidden (not directly observable) state variable ($s(t) \in \mathbf{R}^d$) that evolves in time thus forming a process. The simplest case is if linearity is assumed both for the mixing and the dynamics of the process. Accordingly, the observations and the hidden process may be written as:

$$x(t) = As(t)$$

$$s(t+1) = \sum_{i=0}^{I-1} F_i s(t-i) + \sum_{j=0}^{J-1} H_j e(t+1-j)$$

that is the observations are instantaneous mixtures of the state components whose evolution follows an autoregressive moving average process (ARMA) of order (I, J) with driving noise (or innovation process) $e(t) \in \mathbf{R}^d$. In general the driving noise components are assumed to be *temporally* independent and identically distributed (i.i.d.) stochastic variables. However, in accord with our causal definition of events we also assume that noise components (or at least their subgroups) are *spatially* (i.e., index-wise) independent. We assume that matrix $H_0 \in \mathbf{R}^{d\times d}$ is the identity matrix $I_d \in \mathbf{R}^{d\times d}$. The ARMA$(I, J)$ model comprises the contributions of previous states transferred by the predictive matrices F_i ($i = 0, \ldots, I-1$) and the different echoes of the driving noise transferred by matrices H_j ($j = 0, \ldots, J-1$). The goal is to find the $\hat{F}_i \in \mathbf{R}^{d\times d}$ ($i = 0, \ldots, I-1$) and $\hat{H}_j \in \mathbf{R}^{d\times d}$ ($j = 0, \ldots, J-1$) estimations of the hidden dynamics and the echo structure, respectively, and to learn to separately represent the estimated hidden state, $\hat{s}(t)$, and the independent subgroups of the estimated driving noise $\hat{e}(t)$.

Regarding the structure of F_i, $i = (0, \ldots, I-1)$, a special case seems relevant: *joint block-diagonal structure* implies dynamical sub-processes that do not *mix*. These hidden processes are independent in a dynamical sense so their identification could reduce the representational complexity.

For simplicity, we assume a hidden ARMA$(1, 0)$ = AR(1) process and the issue of delays will be discussed later. (On higher order hidden ARMA processes and post-nonlinear extensions see Lőrincz and Szabó (2007), Szabó, Póczos, and Lőrincz (2007a, 2007b) and Szabó, Póczos, Szirtes, and Lőrincz (2007), respectively.)

The key step to solve the identification problem is to recognize that the observation process is also an AR(1) process, if matrix A can be inverted:

$$x(t+1) = Mx(t) + n(t+1), \tag{1}$$

where the observation noise is $n(t+1) = Ae(t+1)$. According to the d-central limit theorem (Petrov, 1958) $n(t+1)$ is approximately Gaussian so the predictive matrix ($M = AFA^{-1}$) may be estimated by least-mean square approximations and then the wanted independent driving noise components can be extracted by applying Independent Component/Subspace Analysis (ICA/ISA) (Cardoso, 1998; Comon, 1994; Jutten & Herault, 1991) on the observation noise. An important result (Póczos, Szabó, Kiszlinger, & Lőrincz, 2007; Póczos & Lőrincz, 2006) is that – for a large class of source distributions – separation of independent subspaces (i.e., the ISA problem) can be solved in two steps. First, traditional ICA methods yield one dimensional components and second, the resulting components should be grouped to form independent subspaces. In turn, ISA of the estimated noise can simultaneously recover the estimated mixing matrix $\hat{A}$, the hidden state $\hat{s}(t) = \hat{A}^{-1}x(t)$ as well as the assumed independent subspaces of the multidimensional driving source components $\hat{e}(t) = \hat{A}^{-1}\hat{n}(t)$. If $\hat{A}$ is recovered, then the hidden predictive matrix can also be approximated ($\hat{F} = \hat{A}^{-1}\hat{M}\hat{A}$).

3. A connectionist network implementation

In this section we provide local learning rules for parameter estimation of the identification task described above. The resulting algorithms can be translated into a neural network in which 'activity' of a 'neuronal layer' is represented by a vector, connection weights between layers or within layer components (i.e., recurrent connections) are represented by matrices and neurons may realize nonlinear transformations of their inputs.

3.1. Assumptions and simplifications

For simplicity, rate coding (manifesting analogue values) and mixed, i.e., positive and negative weights (thus contradicting with Dale's Principle) are assumed throughout the derivations, but the proposed functioning can in principle be also realized by using either positive coding (Plumbley, 2002) or homogeneous connection systems (Parisien, Anderson, & Eliasmith, 2008). Inhibitions (or subtractions) are manifested by separate inhibitory populations within a layer using feedback or feed-forward inhibition.

We also assume that after each transformation the resulting entities (represented by a given layer) get decorrelated and normalized (i.e. they are subject to *whitening*). Whitening helps compressed encoding (Brand, 2006) and it speeds up ICA (Amari, Cichocki, & Yang, 1996; Cardoso & Laheld, 1996) if applied during preprocessing. Whitening may take place within a layer with the help of inhibitory recurrent connections for which biologically feasible learning rules can also be given (Földiák, 1989). Interestingly, whitening can also be realized by utilizing inter-layer feedback connections. The so called forward-inverse model (Kawato, Hayakawa, & Inui, 1993) or reconstruction network (Lőrincz & Buzsáki, 2000) assumes a *loopy* structure with a *bottom-up* (BU) and a *top-down* (TD) transformations in which representation of the input is used to regenerate an estimate of the input and different constraints can be put on the transformations. It can be shown that in a reconstruction network the two branches can learn to invert each other – at least in the pseudo-inverse sense – applying simple Hebbian rules that we introduce later. Both whitening and ICA can be realized in a reconstruction network for which the mathematical details and a possible implementation scheme will soon be provided.

The different algorithmic tasks and the related learning rules will be described in separate subsections below.

3.2. Innovation

The first algorithmic step is the estimation of the innovation: $\hat{n}(t+1) = x(t+1) - \hat{M}x(t)$. The corresponding cost function $J(\hat{M}) = \frac{1}{2}\sum_t |x(t+1) - \hat{M}\,x(t)|^2$ leads to the following Hebbian