Biologically Inspired Cognitive Architectures (2016) xxx, xxx–xxx

Available at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/bica

RESEARCH ARTICLE

Reinforcement learning in a bio-connectionist model based in the thalamo-cortical neural circuit

Mario Andrés Chalita *, Diego Lis, Agustín Caverzasi

National University of Córdoba, Argentina

Received 7 November 2015; received in revised form 17 February 2016; accepted 29 March 2016

KEYWORDS
Reinforcement learning;
Sparse coding;
Thalamo-cortical;
Cognition;
Connectionism;
Fuzzy neural network

Abstract
In a previous study, we presented a program to simulate a particular dynamic of the thalamo-cortical biological system. The method used was called bio-connectionism which linked the thalamo-cortical mechanism reproduced with animal perception. In this presentation, a reinforcement learning program is supported by this mechanism. In a game world designed to test the model developed, the agent is assigned to a character that must learn by trial and error from its own experience upon recognition of aversive and appetitive patterns. The results confirm, support and extend the notion of configuration, a term familiar with sparse coding principles. If, as it is documented, this mechanism observed in sensory areas can be thought as condition of perception, the brain areas taken together – each in its interaction with a respective sub-thalamic nucleus – are suspected to be considered as condition of cognition. We introduce some philosophical questions derived from the experimental results in the discussion section.

Introduction

Reinforcement Learning without Bio-connectionism is empty; Bio-connectionism without Reinforcement Learning is blind. We adapted the famous phrase from Immanuel Kant of his Critique of Pure Reason (see Kant & Guyer, 1998) to present a reinforcement learning design based on a model developed in a previous work. Back then, we presented a program to simulate a specific dynamic of the thalamo-cortical biological system which we named "configuration". The method used was baptized bio-connectionism, and the configuration of the thalamo-cortical system was linked with animal perception (Chalita & Lis, 2015). There were some outstanding issues, such as the question regarding

* Corresponding author at: Av. Colón 1880, segundo piso departamento 14, Córdoba, Cba. C.P. 5000, Argentina.
E-mail address: 9marioandres6@gmail.com (M.A. Chalita).

http://dx.doi.org/10.1016/j.bica.2016.03.001
2212-683X/

the possibility of extending the simulation of the whole brain, not just the primary sensory area. In this opportunity we present an extended system able to learn by trial and error from its own experience in an unpredictable environment, so that we suit our explanation in the area of the reinforcement learning (RL), a branch of machine learning that refers to the so called "hedonic" learning system (Mnih et al., 2015; Sutton & Barto, 1998). The application of RL in a bio-connectionist model has its antecedents in many previous studies that used RL to support hypotheses on the functioning of biological structures (see for example Chase, Kumar, Eickhoff, & Dombrovski, 2015; Clarke et al., 2014; Collins, Brown, Gold, Waltz, & Frank, 2014; Gläscher, Daw, Dayan & O'Doherty, 2010; Glimcher, 2011; Holroyd & Coles, 2002; Jitsev, Morrison, & Tittgemeyer, 2014; Lucantonio, Caprioli, & Schoenbaum, 2014; McDannald, Lucantonio, Burke, Niv, & Schoenbaum, 2011; Schultz, Dayan, & Montague, 1997; Senn & Pfister, 2014). In this opportunity we show a program based on the representation of not only the primary sensory area but also the main neocortical ones, and we also develop a virtual brain called the agent.

In a common RL program, there is a mechanism of reward and punishment without the need to specify how the task is to be achieved (Kaelbling, Littman, & Moore, 1996). Even in some cases there is not a task to be achieved, but just effects that the agent tends to avoid and effects such agent tends to maintain; in those cases we say the agent is just trying to survive, because we interpret the former are for some reason bad for it and the latter are good for it. The goodness of an action, in this sense, does not depend on a value function which estimates the future accumulated reward by taking such action (contrast with Rao, Bu, Xu, Wang, & Yin, 2009). This is in fact the only operation rule of the animal brain we want to reflect herein.

The name of reinforcement learning presents a dual problem: the agent must learn both how to recognize patterns and which of those patterns likes it and which ones do not like it. Pattern recognition, indeed, must be necessarily solved when taking over the last issue. This does not mean the reinforcement is delayed with respect to the recognition of the reinforced pattern; both are expected to be synchronous. The distinction between two aspects of learning occurs at the level of programming, of the algorithms necessary for the operation of the system, in short, it is about the conditions of learning and not about learning itself. We consider the problem of recognition of patterns already settled in our previous work (Chalita & Lis, 2015). In the program presented herein, we extend those results and explore the second issue: hedonism. Which is, briefly, the difference between recognition of patterns and what we call hedonism? In those terms, it is the same difference we detect between learning and the reinforcement of learning. So hedonism can be considered, according to the name of this subject, reinforcement in the recognition of patterns. We follow rigorously this definition of RL and show how in our design the reinforcement of learning brings about hedonic effects as a result.

The need to use a neural network framework to bring biology closer to RL is also manifested thoroughly by previous documentation (Fan, Tian, & Sengul, 2014; Faußer & Schwenker, 2015a, 2015b; Hill, Marquez, O'Connor, & Remus, 1994; Lin & Lee, 1991; Nakano, Otsuka, Yoshimoto, & Doya, 2015; Noel & Pandian, 2014; Senn & Pfister, 2014; Teng, Tan, Starzyk, Tan, & Teow, 2014; Zhou, Hu, Gao, & Wang, 2014 are some of them). The programmed brain is based on three paradigms synergistically combined: (1) artificial neural networks, (2) bio-connectionism and (3) reinforcement learning. The first one is the language with which the architecture of the agent is drawn; in this case, a recurrent neural network (Du & Swamy, 2014; Duell, Udluft, & Sterzing, 2012; Fausett, 1994). The second one supports the kernel conceptual frame from where it is possible to investigate the link between the agent and cognitive issues. The third one provides the opportunity to test the agent in its interaction with an environment and check their main expected effects according to the theory. The agent is kept as close as possible from biology through the three paradigms mentioned that are, in other words, (1) the biological shape that makes possible the other two ones: (2) the knowledge of the environment and (3) its relationship with it.

Bio-connectionist modeling

A multilayer feed-forward network architecture (Han, Wang, & Qiao, 2014; Hornik, Stinchcombe, & White, 1989), such that the neurons of each layer project to neurons of another layer or layers allows a direct translation to the biological neural system considering a layer as a neural nucleus and, in our case, as a cortical layer as well. In order to differentiate cortical layers from the layer in the bio-connectionist design, the latter will be called station from now on. Programmed neurons are grouped in stations connected lineally in sets, and forms a closed circuit since the neurons of the last station connect with the neurons of the first station of the set. Each station projects to the following station while it is stimulated by the previous station (see Fig. 1). The existence of external stimulation to one of these stations (called for that reason station zero, S0) is the second condition to get what we have called a configuration system (Chalita & Lis, 2015), i.e. a system able to form patterns or cell assemblies (Lansner, 2009).

As a result of our particular reading of the biological nervous system, we differentiate two directions of the neural projections: radial and tangential. Cortico-thalamic and thalamo-cortical neural projections are considered radial, whereas projections among different cortical areas are considered tangential. The closed circuit mentioned above is a case of a radial circuit, which simulates connections between a core nucleus of the thalamus (Jones, 2001) and a neocortical layer, and among the neocortical layers together in a same cortical area. In a tangential sense, this radial circuit can represent, for example, the primary sensory area, or any area of the cortex, since it is known that internal divisions of the thalamus have a correspondence with the map of the neocortex and practically all thalamic subnuclei have feedback projections with a specific neocortical area (Deschênes, Veinante, & Zhang, 1998), by following the so called principle of reciprocity (Diamond, Jones, & Powell, 1969). According to this, the whole tangential circuit can be understood as composed of a set of sensory radial circuits, a set of association radial

Please cite this article in press as: Chalita, MA et al., Reinforcement learning in a bio-connectionist model based in the thalamo-cortical neural circuit, *Biologically Inspired Cognitive Architectures* (2016), http://dx.doi.org/10.1016/j.bica.2016.03.001