Biologically Inspired Cognitive Architectures (2015) 11, 65–74

Available at www.sciencedirect.com

ScienceDirect

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

A neuroscience inspired gated learning action selection mechanism

Klaus Raizer *, Ricardo R. Gudwin

University of Campinas, School of Electrical and Computer Engineering, Campinas, SP, Brazil

Received 7 November 2014; accepted 7 November 2014

KEYWORDS
Action selection;
Neural networks;
Machine learning;
Reinforcement learning;
Genetic algorithm

Abstract
This paper presents an algorithm for action selection, in the context of intelligent agents, capable of learning from rewards which are sparse in time. Inspiration for the proposed algorithm was drawn from computational neuroscience models of how the human prefrontal cortex (PFC) works. We have observed that this abstraction provides some advantages, such as the representation of solutions as trees, making it human-readable, and turning the learning process into a combinatorial optimization problem. Results for it solving the 1-2-AX working memory task are presented and discussed. We also argue the pros and cons of the proposed algorithm and, finally, address potential future work.

Introduction

This paper presents an algorithm for action selection in the context of intelligent agents capable of learning from sparse in time rewards. Inspiration for the proposed algorithm was drawn from computational neuroscience models of how the human prefrontal cortex (PFC) works.

The mathematical and algorithmic study of how the human conscious mind solves the problem of selecting the next action to be taken has produced many interesting results (Baars & Franklin, 2009; Reggia, 2013). By controlling and managing other cognitive processes ''executive functions'' are those responsible for what is usually considered to be ''intelligent behavior''. They are a ''macroconstruct'' (Alvarez & Emory, 2006), in the sense that multiple sub-processes must work in conjunction to solve complex problems. The term ''executive function'' is therefore used as an umbrella for a wide range of cognitive processes and sub-processes (Chan, Shum, Toulopoulou, & Chen, 2008), with the most prominent being *action selection*, *planning*, *selective attention* and *learning* (Baars & Gage, 2010; Frank & Badre, 2012; Fuster, 2008).

In their nature, executive functions are mostly future directed and goal oriented, whilst exerting supervisory control over all voluntary activities. They deal with prospective actions and deliberate plans to achieve goals which can be

* Corresponding author.
E-mail addresses: klaus@dca.fee.unicamp.br (K. Raizer), gudwin@dca.fee.unicamp.br (R.R. Gudwin).

http://dx.doi.org/10.1016/j.bica.2014.11.016
2212-683X/

defined by the executive itself. In a sense, this is what the frontal lobe, in particular the prefrontal cortex, does for humans (Baars & Gage, 2010; Chersi, Ferrari, & Fogassi, 2011).

In this work we propose an action selection algorithm inspired by the computational neuroscience model described in the Leabra framework (Hazy, Frank, & O'Reilly, 2007; O'Reilly, Munakata, Frank, Hazy, & Contributors, 2012). With its PBWM (Prefrontal Cortex Basal Ganglia Working Memory) algorithm, Leabra models how the human PFC interacts with basal ganglia in order to learn from rewards separated in time and select the most appropriate action given a particular stimulus. In other words it performs, with the exception of *planning*, all major executive functions.

The PBWM mechanism strives to follow the biological cognitive process as closely as possible. The present work, however, focuses more on the development of a new algorithm, which is also biologically inspired but ultimately designed to be used in the development of intelligent agents. In order to do so, the inner workings of the PBWM mechanism (Hazy et al., 2007) were abstracted in the form of an action selection algorithm, whose behavior towards new stimuli is defined by an optimized tree structure. We have observed that this abstraction provided a number of advantages, such as:

- Representing solutions as trees, instead of neural networks, allows one to see the knowledge it encodes in a direct manner.
- This method turned the learning process into a combinatorial optimization problem. This potentially makes the use of different optimization techniques straightforward.

Details on how we achieved this can be seen in Section ''Methods''. The remainder of this paper is organized as follows. Section ''Motivation'' presents our initial motivations for developing this algorithm. Section ''The PBWM mechanism'' provides a brief description of how the original PBWM mechanism works, while describing the ''1-2-AX'' working memory[1] task used as a benchmark for validation. Section ''GLAS — A gated-learning action selection mechanism'' then describes our proposed gated-learning action selection (GLAS) mechanism. In Section ''Results'' we apply GLAS to learn the 1-2-AX working memory task and present training results given a particular sequence of events. The paper closes with Section ''Conclusions'', where we discuss obtained results. We also argue the pros and cons of the proposed algorithm and, finally, point to potential future work.

Motivation

The core motivation for developing this algorithm is to take advantage of what neuroscience, and more specifically computational neuroscience, has produced that could be useful for the development of artificial intelligent agents.

Specifically, adapting the PBWM mechanism to provide human-readable solutions was motivated by our previous work with behavior networks[2] and action selection (Raizer, Paraense, & Gudwin, 2012; Raizer, Rohmer, Paraense, & Gudwin, 2013).

As a matter of fact, not only should an agent be able to select the most relevant action at a given time, but it should do so while taking into consideration its future consequences. Traditional reinforcement learning mechanisms, such as variations of SARSA and Q-Learning (Russell & Norvig, 2003), have often been used to solve challenging problems in engineering and computer science (Bagnell & Schneider, 2001; Mahadevan & Connell, 1992; Riedmiller, Gabel, Hafner, & Lange, 2009; Stone, Sutton, & Kuhlmann, 2005; ZicoKolter & Ng, 2011). They lack, however, an ability the mammalian brain excels at: to bridge the gap between actions and late rewards (Bakker, Zhumatiy, Gruener, & Schmidhuber, 2003).

Let us take for instance the task of teaching a dog that taking a bath is a rewarding experience.

In Figs. 1 and 2, ''stimulus from senses'' represents the dog's perception of having a bath. Fig. 1(a) represents the use of a synchronous reward (reward is given while the dog is still perceiving the stimulus) and Fig. 1(b) represents the use of a late reward. We see therefore 4 bath episodes being represented here, and learning is represented by the dotted vertical line. If every time during bath its owner gives the dog a cookie (reward), a burst of dopamine neural firing happens in the dogs' brain. Since the stimulus of taking a bath is still active, the brain manages to correlate this reward with the beginning of the stimulus, linking the perception of taking a bath to being something good.

However, if the owner waits to give its dog a cookie long after each bath is finished, something like what is described in Fig. 1(b) could happen. In this case, there is nothing linking the moment of reward to the appearance of the stimulus.

Dogs, however, are mammals with highly developed prefrontal cortexes. The PFC works, among other things, as a temporary container for storing stimuli representations. Therefore, what really should happen in the previous case, is something like what we see in Fig. 2.

In this case, the PFC stimulus representation was held long enough for the brain to make the association. In other words, the represented stimulus, stored in PFC, acted as if it were the perceived stimulus coming from the dog's senses.

Traditional artificial neural networks, such as MLPs (multi layer perceptrons) and recurrent neural networks, are known to be bad at establishing a link between longer time lapses, since backpropagated errors tend to either explode or exponentially decay during training (Pérez-Ortiz, Gers, Eck, & Schmidhuber, 2003).

Computational models successful at solving this kind of problem are usually those based on gating mechanisms. A

[1] Working memory (WM) is a term, coined by behavioral neuroscience, which describes the cognitive capacity for storing and manipulating novel information for a short period of time (Baars & Gage, 2010). It was initially proposed by Allan Baddeley (Baddeley & Hitch, 1974), and it is believed that the PFC plays a critical role in active maintenance of WM information (Fuster, 2008).

[2] A behavior network is an action selection mechanism initially developed by Maes (1989), which is capable of selecting the most relevant action at the present time, while at the same time deliberating the consequences of those actions. Deliberation is made possible because each behavior has a list of preconditions, which hold propositions necessary for the behavior to become relevant, and lists of consequences. These lists contain human-readable propositions about the world state. For more details we refer to Maes' original paper.