A Neurobiologically Motivated Model for Self-organized Learning

Frank Emmert-Streib*

Institut für Theoretische Physik, Universität Bremen,
Otto-Hahn-Allee, 28334 Bremen, Germany
fes@stowers-institute.org

Abstract. We present a neurobiologically motivated model for an agent which generates a representation of its spacial environment by an active exploration. Our main objectives is the introduction of an action-selection mechanism based on the principle of self-reinforcement learning. We introduce the action-selection mechanism under the constraint that the agent receives only information an animal could receive too. Hence, we have to avoid all supervised learning methods which require a teacher. To solve this problem, we define a self-reinforcement signal as qualitative comparison between predicted an perceived stimulus of the agent. The self-reinforcement signal is used to construct internally a self-punishment function and the agent chooses its actions to minimize this function during learning. As a result it turns out that an active action-selection mechanism can improve the performance significantly if the problem to be learned becomes more difficult.

1 Introduction

The understanding of the action-perception cycle (APC) of an animal is an outstanding problem. If known, it would describe in a principle way the dynamical interplay between sensory processing, memory organization, prediction and action selection mechanism of an animal [10]. Due to the intertwined interactions of the whole system, the action-perception cycle is an systems theoretical approach to understand the neuronal organization of animals rather then a reductionistic one. Although known for a long time, we are still in the beginning to decipher its functional principle. We think, that the key to the understanding of the APC is the investigation of the action-selection mechanism, because it is by far underrepresented in the literature compared to all other parts contributing to the APC. In this paper we focus on one variant of the APC dealing with the spacial exploration of an animal in its environment.

We study the formation of an internal representation of an environment under the influence of different action-selection mechanisms. The major objective of this paper is the introduction of an action-selection mechanism which is based on

* Present address: Stowers Institute for Medical Research,1000 E. 50th Street, Kansas City, MO 64110, USA.

A. Gelbukh, A. de Albornoz, and H. Terashima (Eds.): MICAI 2005, LNAI 3789, pp. 415–424, 2005.

a principle we call *self-reinforcement learning* enabling the agent to self-organized learning. In contrast to reinforcement learning [8] which is also called minimal supervised learning because a teacher is necessary to feed a qualitative signal back to the animal evaluating the last action, our self-reinforcement learning paradigm allows the animal to generate a reinforcement signal itself by comparison between predicted and perceived stimulus. For this reason, we call this signal self-reinforcement signal to emphasize that learning occurs completely unsupervised. Hence, our approach is also different to, e.g., Active Learning which was so far only used in a supervised or at least semi-supervised context [11]. We use for the remaining parts of the action-perception cycle a similar model independently invented by Herrmann et al. [4] and Oore et al. [5] introduced as hippocampus model. This paper is organized as follows: In the next section 2 we define our model and emphasize especially the part of decision making. In section 3 we present results and in 4 we conclude with a discussion.

2 The Model

2.1 Environment

The environment of the agent is an one-dimensional lattice of size N with reflecting boundary conditions. Each site of the lattice $\hat{x} \in \{1, \ldots, E\}$ is assigned an observable symbol $\hat{s} \in \{0, 1\}$, e.g. the color of the site. The position $\hat{x}$ of the agent on the grid is not directly available to the agent but only the observable symbols representing stimuli. However, there is an observation error p_s^{oe} of a symbol $\hat{s}$ on a grid position $\hat{x}$ due to an external influences, e.g. lighting conditions. We model this observation error by a Markov process given by

$$P(s|\hat{s}) = \begin{pmatrix} 1 - p_{s=1}^{\mathrm{oe}} & p_{s=2}^{\mathrm{oe}} \\ p_{s=1}^{\mathrm{oe}} & 1 - p_{s=2}^{\mathrm{oe}} \end{pmatrix} \tag{1}$$

For reasons of simplicity we assume that Eq. 1 is independent of the spatial position. The observed symbol s by the agent is obtained from

$$P(s) = \sum_{\hat{s}} P(s|\hat{s}) P(\hat{s}) \tag{2}$$

Here $P(\hat{s}) = \delta_{\hat{s}, \hat{s}=f(\hat{x})}$, with a mapping f from the position of the agent $\hat{x}$ to the observable symbol $\hat{s}$ on that site. In the following we consider only the case of a symmetric observation error $p_{s=1}^{\mathrm{oe}} = p_{s=2}^{\mathrm{oe}}$.

Because the environment of the agent is one-dimensional we allow the agent only to select between two different actions, $a_t \in \{-1, 1\}$. In analogy to the observation error we introduce a position error p^{pe} occurring if the agent tries to change its grid position due to an external influences, e.g. friction. Here, we assume a homogenes environment to keep things simple and additionally, we assume the transition probability $P(\hat{x}_{t+1}|\hat{x_t}', a_{t+1})$ to be stationary. The new grid position $\hat{x}_{t+1}$ is obtained by

$$P(\hat{x}_{t+1}) = \sum_{\hat{x_t}'} P(\hat{x}_{t+1}|\hat{x_t}', a_{t+1}) P(\hat{x_t}') \tag{3}$$