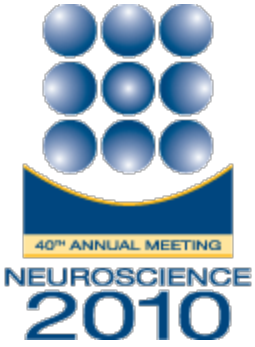
Print this Page



Presentation Abstract

Program#/Poster#:	711.22/NNN45
Title:	Reinforcement learning with a value system in an attractor neural network
Location:	Halls B-H
Presentation Time:	Tuesday, Nov 16, 2010, 2:00 PM - 3:00 PM
Authors:	***M. RIGOTTI**, S. FUSI; Dept. of Neurosci., Columbia Univ., NEW YORK, NY
Abstract:	The prediction of future events is key for the survival of organisms in a dynamic environment. It is essential for deciding between alternative courses of action, some of which may lead to food or other rewarding outcomes, others which may result in danger or loss of resources. Reinforcement Learning (RL) offers a theoretical framework for formalizing the problem of predicting the result of interactions with the environment and adapting behavior so as to choose an optimal course of action. Apart from its formal appeal, RL's popularity in the neuroscience community is at least partly due to its success in modeling neural data. Temporal-difference (TD) learning algorithms in particular attribute a well-defined function to dopamine neurons by interpreting their activity as encoding a prediction error signal quantifying the mismatch between the predicted and the observed reward due

to a perceived state of the environment (the so-called reward-prediction error theory of dopamine). Considerable effort has been recently devoted to the extension of the RL formalism to the case in which the environment is only partially observable, so that its state is not uniquely determined by the interaction with the agent.
As an alternative to the Partially Observable Markov Decision Process (POMDP) perspective, we studied a neural network model representing internal states (as opposed to environment states) and their values as attractors of the neural dynamics. We implemented an extension of an actor/critic architecture where the short-lived phasic reward-prediction error signal is utilized (by the critic) to update the time-extended value of internal states and (by the actor) to modify a probabilistic policy at every state. In our model the probabilistic policy is defined by the transition probabilities between internal states conditioned on an observed external stimulus from the environment. Although in general with an approximate suboptimal strategy, this gives a natural way to reactively cope with a partially observable environment and actuate an initial exploration phase.
Motivated by previous work [Rigotti et al., 2010] we assume that this exploratory phase mediates the creation of new attractors representing novel internal states which encode the observed stimulus statistics.
We analyse the interaction and the convergence of RL with this unsupervised mechanism for the creation of new states in some simple classical conditioning and context-dependent spatial navigation tasks, and argue that it could be an effective way to extend standard RL to navigate realistic dynamic environments requiring context-dependent courses of action.

Disclosures: **M. Rigotti**, None; **S. Fusi**, None.

Keyword(s): DECISION MAKING

REINFORCEMENT LEARNING

Support: DARPA SyNAPSE HR0011-09-C-0002

NIMH grant 2RO1MH58754

[Authors]. [Abstract Title]. Program No. XXX.XX. 2010 Neuroscience Meeting Planner. San Diego, CA: Society for Neuroscience, 2010. Online.