Distributed simulation of a spiking temporal-difference learning model based on dopamine-modulated plasticity

Wiekbe Potjans[1,2], Abigail Morrison[1], Markus Diesmann[1,3]

1 RIKEN Brain Science Institute, Wako City, Saitama, Japan

2 Institute of Neurosciences and Medicine, Research Center Juelich, Juelich, Germany

3 Brain and Neural Systems Team, RIKEN Computational Science Research Program, Wako City, Saitama, Japan

E-Mail: wiebke_potjans@brain.riken.jp

An open question in the field of computational neuroscience is how higher organisms learn in environments with sparse rewards. We propose a spiking neural network model implementing temporal-difference (TD) learning based on dopamine modulated plasticity. For this purpose, we developed a general framework that enables distributed simulations of spiking neural networks with neuromodulated plasticity. Our network is able to learn a grid-world task with learning speed and equilibrium performance similar to a discrete-time TD algorithm.

1 Introduction

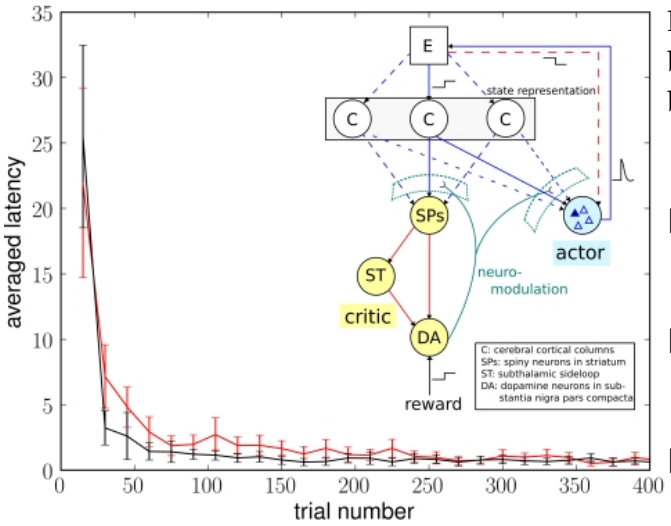
Making predictions about future rewards and adapting the behavior accordingly is crucial for any higher organism. One theory specialized for prediction problems is temporal-difference learning. Experimental findings suggest that TD learning is implemented by the mammalian brain. In particular, the resemblance of dopaminergic activity to the TD error signal [1] and the modulation of corticostriatal plasticity by dopamine [2] lend support to this hypothesis. We recently proposed the first spiking neural network model to implement actor-critic TD learning [3], enabling it to solve a complex task with sparse rewards. However, this model calculates an approximation of the TD error signal in each synapse, rather than utilizing a neuromodulatory system.

2 Model and simulation technology

Here, we propose a spiking neural network model (see figure inset) which dynamically generates a dopamine signal based on the actor-critic architecture proposed by Houk [4]. This signal modulates as a third factor the plasticity of the synapses encoding value function and policy. The implementation of neuromodulated plasticity in large-scale network simulations is challenging because the dynamics of these networks is commonly defined on the connectivity graph without explicit reference to the embedding of the nodes in physical space (e.g. NEST [5]). Therefore, we developed a general framework to simulate models with closed functional circuits that simultaneously represent synapses with neuro-modulated plasticity and neurons that release a neuromodulator in particular target regions.

3 Results

In our model learning is implemented on the level of synaptic plasticity. In this study, we would like to investigate the question of how the microscopic level of learning can be related to system-level learning. Therefore we test the learning behavior of our model in simulations performed in NEST [5], where an agent has to find a single rewarded state from a random position in a two dimensional grid-world task. During the learning process the synaptic weights implementing the value function and the policy develop such that they



reflect the proximity to the reward and an optimal policy towards the reward, respectively. Learning can also be seen on the macroscopic level. Learning is achieved in less than 100 trials and stays stable for at least 400 trials. The learning speed and equilibrium performance are comparable to those of a discrete time algorithmic TD learning implementation (see figure; red: spiking network model, black: discrete time algorithm).

5 Outlook

Our model investigates the concept of temporal-difference learning in the basal ganglia, a small subpart of the brain important for reward learning. However, the basal ganglia are functionally embedded in a closed loop in the brain with a major input coming from the cerebral cortex. A network modeling such a closed loop requires high-end supercomputers as the number of neurons, synapses and computational load increases by at least one order of magnitude. For a cortex model of one cubic millimeter investigated intensively in our group as well as for networks of the same size implementing dopamine-modulated plasticity linear scaling has been demonstrated up to at least 1000 processors showing that they are likely to be suitable applications for high-end supercomputers such as Japan's Next-Generation Supercomputer. Such closed-loop models can be used to perform lesion studies and to investigate the relation between electrophysiological studies and behavior.

References

[1] Schultz W, Dayan P and Montague PR, A neural substrate of prediction and reward, Science 275: 1593-1599, 1997

[2] Reynolds JN, Hyland BI and Wickens JR, A cellular mechanism of reward-related learning, Nature 413: 67-70, 2001

[3] Potjans W, Morrison A, Diesmann M, A spiking neural network model of an actor-critic learning agent, Neural Computation 21: 301-339, 2009

[4] Houk JC, Adams JL and Barto AG, A model of how the basal ganglia generate and use neural signals that predict reinforcement. Cambridge, MA:MIT Press; 1995

[5] Gewaltig M-O and Diesmann M: NEST (neural simulation tool). Scholarpedia 2(4), 1430, 2007

Acknowledgements

Partially funded by Next-Generation Supercomputer Project of MEXT, Japan, EU Grant 15879 (FACETS), BMBF Grant 01GQ0420 to BCCN Freiburg, and the Helmholtz Alliance on Systems Biology.