BMC Neuroscience

Poster presentation

Open Access

Direct reinforcement learning, spike time dependent plasticity and the BCM rule

Dorit Barash[1] and Ron Meir*[2]

Address: [1]IBM Haifa Research Lab, Mount Carmel, Haifa 31905, Israel and [2]Department of Electrical Engineering, Technion, Haifa 32000, Israel

Email: Ron Meir* - rmeir@ee.technion.ac.il

* Corresponding author

from Sixteenth Annual Computational Neuroscience Meeting: CNS*2007
Toronto, Canada. 7–12 July 2007

Published: 6 July 2007

BMC Neuroscience 2007, **8**(Suppl 2):P197 doi:10.1186/1471-2202-8-S2-P197

Learning agents, whether natural or artificial, must update their internal parameters in order to improve their behavior over time. In reinforcement learning, this plasticity is influenced by an environmental signal, termed a reward, which directs the changes in appropriate directions. We model a network of spiking neurons as a Partially Observed Markov Decision Process (POMDP) and apply a recently introduced policy learning algorithm from Machine Learning to the network [1]. Based on computing a stochastic gradient approximation of the average reward, we derive a plasticity rule falling in the class of Spike Time Dependent Plasticity (STDP) rules, which ensures convergence to a local maximum of the average reward. The approach is applicable to a broad class of neuronal models, including the Hodgkin-Huxley model. The obtained update rule is based on the correlation between the reward signal and local data available at the synaptic site. This data depends on local activity (e.g., pre and post synaptic spikes) and requires mechanisms that are available at the cellular level. Simulations on several toy problems demonstrate the utility of the approach. Like most stochastic gradient based methods, the convergence rate is slow, even though the percentage of convergence to global maxima is high. Additionally, through statistical analysis we show that the synaptic plasticity rule established is closely related to the widely used BCM rule [2], for which good biological evidence exists. The relation to the BCM rule captures the nature of the relation between pre and post synaptic spiking rates, and in particular the self-regularizing nature of the BCM rule. Compared to previous work in this field, our model is more realistic than the one used in [3], and the derivation of the update rule applies to a broad class of voltage based neuronal models, eliminating some of the additional statistical assumptions required in [4]. Finally, the connection between Reinforcement Learning and the BCM rule is, to the best of our knowledge, new.

References

1. Baxter J, Bartlett PL: **Infinite-horizon policy-gradient estimation.** *Journal of Artificial Intelligence Research* 2001, **15:**319-350.
2. Bienenstock EL, Cooper LN, Munro PW: **Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex.** *J Neurosci* 1982, **2(1):**32-48.
3. Bartlett PL, Baxter J: **Hebbian synaptic modifications in spiking neurons that learn.** In *Technical report, Reasearch School of Information Sciences and Engineering Australian National University*; 1999.
4. Xie X, Seung HS: **Learning in neural networks by reinforcement of irregular spiking.** *Physical Review E* 2004, **69:**041909.