RESEARCH MASTER INTERNSHIP

Department of Complex Systems Engineering

Supervisor : Dennis WILSON (ISAE), Emmanuel RACHELSON (ISAE), Timothée MASQUELIER (CerCo)

Location : ISAE-Supaero, Toulouse

Tel. : +33 5 61 33 81 27

E-mail. : dennis.wilson@isae.fr

INTERNSHIP DESCRIPTION

Domain : Artificial Intelligence, Computational Neuroscience

Title : **REWARD-MODULATED STDP FOR REINFORCEMENT LEARNING**

Spike-timing-dependent plasticity (STDP) is a well-studied biological learning rule for the change in synaptic efficacy in the brain based on the difference in spike time between the pre-synaptic and post-synaptic neurons. Reward-modulated STDP (RM-STDP) modifies the synaptic weight change determined by STDP based on a reward signal which is propagated throughout the network. RM-STDP has been successfully used as the learning rule in artificial spiking neural networks (SNN) to achieve competitive classification on the MNIST image classification task [1].

The temporal nature of SNNs and the semi-supervised nature of RM-STDP make this learning method a good match for the reinforcement learning (RL) domain, where many problems are temporal in nature and learning is semi-supervised. Despite this, there is little work in the application of RM-STDP to RL tasks such as control or locomotion [2]. One of the challenges of applying RM-STDP to RL is credit assignment, especially when the reception of a reward signal is temporally distant from the causal action. While there have been proposed solutions to the distal reward problem, they have only been demonstrated on simple benchmark tasks at the level of neural activity [3].

The goal of this project is to apply RM-STDP to classic RL tasks such as cart-pole balancing. This first application will open the door for more complex applications of RM-STDP and allow for high-level RL on the multitude of hardware based on SNNs. The ideal candidate for this project will have experience in RL and neural networks, and either an understanding of or a desire to learn about neuroscience.

References :

[1] Mozafari, Milad, et al. "Bio-inspired digit recognition using reward-modulated spike-timing-dependent plasticity in deep convolutional networks." *Pattern Recognition* 94 (2019): 87-95.

[2] Wilson, Dennis G., et al. "Learning aquatic locomotion with animats." Artificial Life Conference Proceedings 14. MIT Press, 2017.

[3] Izhikevich, Eugene M. "Solving the distal reward problem through linkage of STDP and dopamine signaling." Cerebral cortex 17.10 (2007): 2443-2452.

Methods: Spike-timing-dependent plasticity, reinforcement learning

50 % Theoretical Research	0 % Applied Research	50 % Experimental Research

Possibility to go on a Ph.D.: o Yes o No

APPLICANT PROFILE

Desired knowledge: Reinforcement learning, artificial neural networks, spiking neural networks

Languages/Systems : Julia, Python, Unix

Applications should be sent by e-mail to the supervisor.