Reinforcement Learning with Spiking Neural Networks of Multi-Quadratic Integrate-and-Fire neurons

Antoine Debor, Pierre Sacré, and Guillaume Drion
Dept. of Electrical Engineering and Computer Science, University of Liège, Belgium
Email: {adebor, p.sacre, gdrion}@uliege.be

Artificial intelligence has progressed a lot these last decades, showing impressive and groundbreaking results. These advances are mainly due to the use of deep Artificial Neural Networks (ANNs) architectures dedicated to many cognitive tasks such as vision, language, or decision-making. However, these systems still struggle with adaptive behavior and continual learning, while being resource-hungry, not only in terms of computing resources, but also in terms of the amount of data and time that are needed to train them.

In contrast, such challenges are easily overcome by biological brains. Like ANNs, these are made of billions of interconnected neurons. However, big differences exist between the two, the brain being more complex than ANNs in many respects. Among these, one key difference lies in the way biological neurons communicate. Artificial neurons communicate through real values, while biological neurons communicate through spikes. These can be seen as rapid rises and falls of the membrane potential of the neurons, propagating the signal and the information through the brain. Neurons in the brain hence use a time-based way of communicating, contrary to ANNs. These spikes form trains which can exhibit various typical patterns, defining regimes for the spiking neuron. A neuron can see its regime change over time under the effect of neuromodulation, a key process in the brain involved in many cognitive aspects including learning and adaptive decision-making among others.

In order to narrow the gap with the brain, a new type of artificial neural networks, that are closer to biology, has emerged: the so-called Spiking Neural Networks (SNNs). Neurons in these networks are designed to reproduce the temporal dynamics of biological neurons, with varying levels of complexity, ranging from complex, quantitative models, to simple phenomenological models, *e.g.* the widely used Leaky Integrate-and-Fire (LIF) model. The LIF model is often preferred in machine learning contexts because of its ease of implementation and usage. Networks of LIF neurons have been proven successful in learning various Supervised Learning tasks, as well as in the Reinforcement Learning (RL) setting.

However, due to its lack of biological accuracy, the LIF neuron cannot account for the richness of temporal regimes a biological neuron shows. In contrast, the Multi-Quadratic Integrate-and-Fire (mQIF) model [1] has been developed as an extension of the LIF model that allows for bistability and for the bursting regime. It is thus an interesting alternative to the LIF that can be relatively simply implemented in SNNs using traditional frameworks, while offering sufficient physiological fidelity to express the different spiking regimes of a biological neuron, hence opening the door to neuromodulation.

Recently, the use of population coding in SNNs made of LIF neurons was explored [2], showing that it improved their performance in RL environments for continuous control, compared to ANNs with comparable architecture and to SNNs without population coding. This architecture, called PopSAN, made use of LIF neurons and it was shown that coding observations using as few as $T = 5$ neuronal time steps, *i.e.* the number of steps over which the neuronal model is simulated for one step in the environment, was sufficient for the network to succeed at the tasks. While being computationally appealing, this suggests that this architecture does not leverage the time-encoding capabilities SNNs have to offer.

In this work, we build on the PopSAN architecture by replacing LIF neurons with mQIF neurons, hence allowing for more richness in terms of spiking regimes. We evaluate this modified architecture on continuous control tasks with multibody agents in the MuJoCo simulator. We study how the training phase and performance of these modified agents are influenced by the modification of the neuronal model, as well as how the learned policies leverage the expressiveness of this model in terms of temporal encoding. In particular, we study how the performance of the policies is influenced by the number of steps T over which observations are encoded and actions are decoded.

This work hence paves the way for Spiking RL policies that leverage rich temporal dynamics and that can be neuromodulated to improve learning and adaptation.

References

[1] T. Van Pottelbergh et al., “Robust Modulation of Integrate-and-Fire Models“, Neural Computation, 2018.

[2] G. Tang et al., “Deep Reinforcement Learning with Population-Coded Spiking Neural Network for Continuous Control“, Proceedings of Machine Learning Research, 2021.

Acknowledgment This work was supported by the Belgian Government through the FPS Policy and Support (BOSA) grant NEMODEI.