Reinforcement Learning and Spiking Neurones: from models to hardware

Team BISCUIT, Loria

Supervision : Alain DUTECH (HDR), Bernard GIRAU (HDR)
alain.dutech@loria.fr & bernard.girau@loria.fr

Last change: May 6, 2019

1 Context

The BISCUIT[1] team of the Loria[2] laboratory studies computational paradigms where calculations are adaptive, distributed and decentralized, carried out by populations of simple computing units that communicate mainly with their close neighbors. These properties are compatible with the implementation of unsupervised – but not un-guided – self-organization principles to tackle difficult problems such as situated cognitive computation, autonomous robotics, adaptive allocation of computation resources, etc.

These characteristics also make it possible to consider a better use of so-called "neuro-morphic" processors that are ermerging (IBM Truenorth, Intel Loihi, etc.). These processors are based on neuro-inspired principles that respect the constraints of the paradigms we are studying, and can benefit from self-organization mechanisms – not supervised but guided – that we are developing, both in terms of applications and neuromorphic resources management.

This is why the BISCUIT team is committed to designing unsupervised and guided learning architectures and algorithms for spatialized and decentralized computing populations while remaining as close as possible to the constraints and characteristics of the Hardware. The subject of the proposed doctoral thesis is a additonal step in that direction.

2 Goals

The main goal of this thesis is to explore the use of mechanisms at the crossroads of neuro-inspired calculation and reinforcement learning within the framework of so-called *neuromorphic* architectures.

The general framework of *reinforcement learning* (or RL) (Sutton and Barto, 1998) proposes a theoretical basis for decision making in uncertainty. Within the BISCUIT team, it is our main prospect in exploring how to guide the processes of self-organization. The classical algorithms are mainly dedicated to discrete and centralized approaches that are not very compatible with our computation paradigms. Similarly, the mechanisms currently in place in in deep reinforcement learning are based on gradient descents and cannot adapt to our constraints of decentralization and non-supervision.

This is the reason why we want to further explore one of the learning mechanisms coming from the connectionism world where decentralization and population coding are easier. This mechanism is called Spike-Timing Dependent Plasticity (or STDP) (Markram et al., 1997; Bi and Poo, 1998), an unsupervised learning rule that can be modulated, for example by taking into account a reinforcement signal, as in classical reinforcement learning.

Several studies have already been carried out in this direction (see, for example , (Florian, 2007; El-Laithy and Bogdan, 2011)), but these works are still few in number. Moreover, they were

[1] Bio-Inspired Situated Cellular and Unconventional Information Technology, http://biscuit.loria.fr/
[2] www.loria.fr

conducted independently of the current advances in neuromorphic processors, some of which are very recent (Intel chip "Loihi", (Davies et al., 2018)). However, a strong trend of recent neuromorphic processors is precisely the implementation on a chip of configurable STDP mechanisms, which allows the adaptability of the implanted models while being based on decentralized and local learning rules. The purpose of this thesis is therefore to study this family of algorithms by taking into account the computational paradigms studied by the BISCUIT team and in the perspective of the emergence of neuromorphic circuits. This will include :

- conduct a literature review on STDP and RL;
- explore the capacity of existing STDP models modulated by RL to allow the learning of our neural models (mainly self-organizing map and dynamical neural fields);
- propose adaptations of these algorithms compatible with the constraints imposed by neuromorphic processors;
- adapt these algorithms to the team's key issues, in particular the decentralized dynamic control of the allocation of computing resources on neuromorphic processors.

3 Working conditions and desired skills

The doctoral student will be welcomed at the Loria in Nancy, France. He or she will work under the supervision of Alain Dutech and Bernard Girau. Scientific collaboration with other team members is expected, as well as more general scientific discussions and collaborations with other members of the laboratory. The expected duration of the doctorate is three years.

In addition to advanced master's level computer skills, we expect solid foundations on the associated mathematical concepts (in particular probabilities and differential equations). The candidate should have some appetence for the design of digital circuits and artificial intelligence. Finally, the candidate, who holds a Master's degree in computer science or equivalent, must be creative, curious and autonomous. The team will provide a set of programming tools and all the support necessary for the technical aspects of the work, which will allow the doctoral student to focus on the scientific questions. Being comfortable with software design is also required, the code production will be done under Linux.

References

Bi, G.-q. and Poo, M.-m. (1998). Synaptic modifications in cultured hippocampal neurons: Depend ence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18(24):10464–10472.

Davies, M., Srinivasa, N., Lin, T., Chinya, G., Cao, Y., Choday, S. H., Dimou, G., Joshi, P., Imam, N., Jain, S., Liao, Y., Lin, C., Lines, A., Liu, R., Mathaikutty, D., McCoy, S., Paul, A., Tse, J., Venkataramanan, G., Weng, Y., Wild, A., Yang, Y., and Wang, H. (2018). Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 38(1):82–99.

El-Laithy, K. and Bogdan, M. (2011). A Reinforcement Learning Framework for Spiking Networks with Dynamic Synapses. *omputational Intelligence and Neuroscience*, 2011.

Florian, R. V. (2007). Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. *Neural Computation*, 19(6):1468–1502.

Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297):213–215.

Sutton, R. and Barto, A. (1998). *Reinforcement Learning*. Bradford Book, MIT Press, Cambridge, MA.