PhD proposal, Computer Science

Motivated Multi scale Self-Organization for the emergence of coordinated sensorimotor behaviors

1 Contacts and location

The PhD thesis will be supervised by :

- Alain Dutech (alain.dutech@inria.f), HdR in computer science
- Jeremy Fix (jeremy.fix@centralesupelec.fr), PhD in computer science

The PhD thesis will be carried out, administratively, within the Biscuit Team (LORIA) and physically at the LORIA (Vandoeuvre-les-Nancy) laboratory and within the Metz campus of CentraleSupélec.

2 Context

A long standing question in cognitive science is how an agent can build up from its experience to develop new sensorimotor skills in an open-ended exploration and learning from the interactions with its environment. Some authors suggest that an agent has some innate sensorimotor primitives (e.g. central pattern generators) (Mussa-Ivaldi and Solla, 2004; Ijspeert, 2008). These primitives, or reflexes, allow the agent to initiate some basic interactions with the environment. One challenging question is to know how the agent can build on these interactions, their regularities but also their unexpected outcomes, to develop new skills and behaviors, in order to increase its "mastery" of the environment. While artificial intelligence initially focused on the abstract ability of reasoning, this challenge is at the core of the more recent trend of embodied cognition (Brooks, 1991; Nöe, 2004; Warren, 2006) that stresses the importance strong coupling between the brain, the body and the environment in which the agent evolves. The environment exposes to the agent some opportunities of action but the possibility of the agent to benefit from these opportunities strongly depend on the motor capabilities of the agent. There is therefore an interdependence between the environment and the brain/body of the agent and some mechanisms should allow the agent to capture and build up on this coupling.

There is a mathematically grounded theory, known as reinforcement learning, which addresses the question of how an agent can learn to behave in an unknown (model free) environment (Sutton and Barto, 2016). This theory formalizes the notions of reward, policy, value functions and provides algorithms allowing an agent to learn an optimal behavior with respect to the expected reward it can receive. The models within the mathematical theory of Reinforcement Learning are originally formulated with a granularity in time at the level of the interaction. There have been recently some extension to deal with continuous time (Doya, 2000) but there remains some difficulties regarding the biological plausibility of these models. Even if the biological plausibility is not necessarily a goal of the PhD thesis, it remains that the fact that the current models in reinforcement learning are not biologically plausible suggests that there might be alternative models to account for learning of sensorimotor behaviors.

In the Neuroscience litterature, there is a wealth of evidence that several deep structures in the primate brains, known as the basal ganglia, are involved in reward guided sensorimotor learning. Various factors contribute to this learning of behavior. It is well known since the pioneering work of (Schultz et al., 1997) that unexpected rewarding events trigger responses of the dopaminergic neurons in the *ventral tegmental area*, a structure known to be in close relationship with structures

of the basal ganglia. Several work stress the link between these structures and the theory of reinforcement learning (Niv, 2009). There are other neurotransmitters that might guide learning (Doya, 2002). Some authors also suggest that there is an interest in guiding learning by some form of curiosity which shapes how an agent confronts itself with sensorimotor abilities that are not yet mastered but still learnable (Kalplan and Oudeyer, 2007). In addition to be involved in reinforcement learning, works in Neuroscience indicate that the behaviors might categorized in two categories : habits and goal directed behaviors (Graybiel, 2005; Yin and Knowlton, 2006). Goal directed behaviors, which are sensitive to the desirability of the outcome and the contingency relationship between the executed action and an obtained outcome, seem to be transferable toward habits which are less flexible but more quickly executed. These two types of behavior and their neural substrate are indications on how an agent could increase its vocabulary of sensorimotor behaviors.

3 Subject

During this thesis, we would like to work with learning models that evolve continuously in time and that are compatible with the general framework of Reinforcement Learning and see how these models could allow an agent to build more abstract and complex behaviors.

For example, several works have shown that models specified within the Dynamic Systems Theory (DST) (Kelso, 1995) can deal with simple sensorimotor behaviors (Beer, 1995; Spencer et al., 2011). Despite these success, several questions remain opened and these are the kind of questions we would like to address in this thesis.

Thus, a non exhaustive list of challenging questions we would like to address could be:

- while we know that central pattern generators might be candidates for providing motor primitives, it is still unclear how a distributed model can build up on these motor primitives sensorimotor behaviors on a longer time scale;
- how these higher level, more abstract, sensorimotor behaviors could be integrated by the model to articulate its behavior ?
- how the system could learn even more complex behaviors, in some way, gradually bootstrapping its sensorimotor capabilities ?
- can we make operant the theories of (Warren, 2006; Keijzer, 2001) that advocate anticipative behaviors;
- is it possible to take inspirations from works on hierarchical reinforcement learning using options (Sutton et al., 1999; Dietterich, 2000) in a continuous framework ?

4 Expected skills

The PhD student is expected to be open to unconventional bio-inspired computing with some backgrounds in Machine Learning. As the work of the PhD has its root in cognitive science, we expect the candidate to be at least curious about philosophy, psychology and biology. This PhD thesis is not purely theoretical and we expect the candidate to experiment the developed models on a simulated or real robotic platform. Therefore, programming skills (`C++`, `Python`) and familiarity with a `Unix` operating system are of primary interest. Also a prior experience with the Robot Operating System (ROS) and simulated (e.g. V-REP) or real robotic platforms is a plus.

References

Beer, R. D. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial intelligence*, 72(1–2):173–215.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence Journal*, 47:139–159.

Dietterich, T. (2000). Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research (JAIR)*, 13:227–303.

Doya, K. (2000). Reinforcement learning in continuous time and space. *Neural Computation*, 12:219–245.

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, 15(4–6):495–506.

Graybiel, A. (2005). The basal ganglia: learning new tricks and loving it. *Current Opinion in Neurobiology*, (15):638—644.

Ijspeert, A. (2008). Central pattern generators for locomotion control in animals and robots: a review. *Neural Networks*, 21(4):642–653.

Kalplan, F. and Oudeyer, P. (2007). In the search of the neural circuits of intrinsic motivation. *Frontiers in Neuroscience*, 1(225–236).

Keijzer, F. (2001). *Representation and behavior*. MIT Press.

Kelso, J. S. (1995). *Dynamic Patterns*. MIT Press.

Mussa-Ivaldi, F. and Solla, S. (2004). Neural primitives for motion control. *IEEE Journal of Oceanic Engineering*, 29(3):640–650.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3):139–154.

Nöe, A. (2004). *Action in Perception*. MIT Press.

Schultz, W., Dayan, P., and Montague, P. (1997). A neural substrate of prediction and reward. *Science*, 275:1593—1599.

Spencer, J. P., Perone, S., and Buss, A. T. (2011). Twenty years and going strong: A dynamic systems revolution in motor and cognitive development. *Child Dev Perspect*, 5(4):260–266.

Sutton, R. and Barto, A. (2016). *Reinforcement learning: An Introduction*. 2nd edition, in progress.

Sutton, R., Precup, D., and Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, (112):118–211.

Warren, W. H. (2006). The dynamics of perception and action. *Psychological Review*, 113(2):358–389.

Yin, H. and Knowlton, B. (2006). The role of the basal ganglia in habit formation. *Nature Reviews, Neuroscience*, (7):464—476.