Neuromodulation and open-ended development

Frédéric Kaplan
Sony Computer Science Laboratory Paris
Developmental Robotics Group
6 rue Amyot 75005 Paris France
kaplan@csl.sony.fr

Pierre-Yves Oudeyer
Sony Computer Science Laboratory Paris
Developmental Robotics Group
6 rue Amyot 75005 Paris France
py@csl.sony.fr

Recent discoveries showing a convergence between patterns of the activity in the midbrain dopamine neurons and computational model of reinforcement learning have lead to an important amount of speculations about learning activities in the brain [5]. In particular actor-critic reinforcement learning architectures have been presented as relevant models to account for functional and anatomical subdivisions in the midbrain dopamine system. Central to some of these models is the idea that dopamine cells report the error in predicting expected reward delivery and that this information is used in two different ways. The value system learned by the critic is associated with projections from the ventral tegmental area to the amygdala and the orbitofrontal cortex. The action selection scheme of the actor is thought to be realized by dopamine pathways initiated in the substantia nigra pars compacta and projecting to the striatium, thus controlling the choice of actions during cortico-striato-thalamo-cortical loops. This popular model has raised some amount of controversies (e.g. [1]) but has definitively shown that artificial learning paradigms could lead to interesting new interpretations of neurophysiological data.

We want to emphasize the inspiring role that research in developmental robotics can play in this context. One of the important goal of this new research field is to understand which dynamics can lead to open-ended development *i.e.* how robots can be designed to continuously learn new skills of increasing complexity. Paradigms based on conditioning and external rewards have difficulties to account for the active nature of development and exploratory behaviors. Children in the first years of their life actively choose in which learning task they take part, avoiding situations that are too difficult for them or that have become too predictable. This suggests the existence of intrinsic motivations structuring learning activities. Proposing models for such motivations has become a major challenge for developmental robotics.

We argue that in order to realize autonomous mental open-ended development, reinforcement learning models could be interestingly associated with an internal reward system based on the maximization of learning progress. Several preliminary computational and robotic experiments show how intrinsic motivations enable the development of novel behaviors of increasing complexity (e.g. [3, 4]). These new models naturally lead to investigate how the basic actor-critic paradigm could be extended to account for an architecture capable of evaluating its own "learning progress". Studies suggesting that dopamine responses could be interpreted as reporting "prediction error" (and not only "reward prediction error") [2] may be taken into consideration for formulating new hypotheses about neural processes that could account for a system of intrinsic motivations.

References

[1] P. Dayan and W. Belleine. Reward, motivation and reinforcement learning. *Neuron*, 36:285–298, 2002.

[2] J-C. Horvitz. Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience*, 96(4):651–656, 2000.

[3] F. Kaplan and P-Y. Oudeyer. Maximizing learning progress: an internal reward system for development. In F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, editors, *Embodied Artificial Intelligence*, LNAI 3139, pages 259–270. Springer-Verlag, 2004.

[4] P-Y. Oudeyer and F. Kaplan. Intelligent adaptive curiosity: a source of self-development. In Luc Berthouze, Hideki Kozima, Christopher G. Prince, Giulio Sandini, Georgi Stojanov, G. Metta, and C. Balkenius, editors, *Proceedings of the 4th International Workshop on Epigenetic Robotics*, volume 117, pages 127–130. Lund University Cognitive Studies, 2004.

[5] W. Schultz, P. Dayan, and P.R. Montague. A neural substrate of prediction and reward. *Science*, 275:1593–1599, 1997.