Proceedings of the 2007 IEEE Symposium on
Foundations of Computational Intelligence (FOCI 2007)

Designing the Reward System: Computational and Biological Principles

Kenji Doya
Okinawa Institute of Science and Technology

Abstract— **In the standard framework of optimal control and reinforcement learning, a "cost" or "reward" function is given a priori, and an intelligent agent is supposed to optimize it. In real life, control engineers and machine learning researchers are often faced with the problem of how to design a good objective function to let the agent attain an intended goal efficiently and reliably. Such meta-level tuning of adaptive processes is the major challenge in bringing intelligent agents to real-world applications.**

While development of theoretical frameworks for 'meta-learning' of adaptive agents is an urgent engineering problem, it is an important biological question to ask what kind of meta-learning mechanisms our brain implements to enable robust and flexible control and learning. We are putting together computational, neurobiological, and robotic approaches to attack these dual problems of computational theory and biological implementation of meta-learning.

In the computational front, our major strategies have been parallelism and evolution. We derived a parallel learning framework, CLIS (concurrent learning by importance sampling), in which multiple adaptive agents sharing the same behavioral experience (i.e., many brains sharing one body) complete to take in charge of actions, but learn concurrently from fellow agents' successes and failures by appropriate weighting. We also developed an embodied evolution scheme in which reward functions for multiple objectives are optimized to match environmental constraints. We are verifying these computational frameworks using a robotic platform called "Cyber Rodents", which realize self-preservation by catching battery packs and self- reproduction in software by IR transmission of their genetic codes.

On the biological side, our major focus is on the selection of appropriate temporal horizon for prediction and learning. In our life, whether to enjoy immediate reward or to look ahead for long- term profit is often a difficult decision (e.g., relax on the beach or work on the lecture). In reinforcement learning theory, this trade- off is dealt with by the setting of the temporal discounting parameter. After reviewing a wealth of neurobiological studies, we conjectured that the serotonin, one of the major neuromodulators, regulates the temporal discounting parameter of the brain. In our functional brain imaging studies, we found that our brain has parallel pathways for predicting immediate and future rewards and that they are differentially modulated by the activity of serotonin. Our recent neural recording studies in rats also showed that the serotonin neurons increase firing during delay periods when rats wait for expected future rewards. These are consistent with our conjecture.

To conclude, theoretical development for computational intelligence is critical for understanding the highly adaptive mechanisms of the brain, and more knowledge of the brain as a real instantiation of fluid intelligence sets higher goals for theoretical development.

REFERENCES

[1] Doya K. (2002). Metalearning and neuromodulation. Neural Networks, 15, 495-506.
[2] Doya K., Uchibe E. (2005). The Cyber Rodent project: Exploration of adaptive mechanisms for self-preservation and self-reproduction. Adaptive Behavior, 13, 149-160.
[3] Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., Yamawaki, S. (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. Nature Neuroscience, 7, 887-893.

1-4244-0703-6/07/$20.00 ©2007 IEEE