Abstract of the 13th Toyota Conference on Affective Minds (1999).

METALEARNING, NEUROMODULATION AND EMOTION

Kenji Doya
doya@ctr.atr.co.jp
ATR International; CREST, Japan Science and Technology Corp.

Recent advances in machine learning and artificial neural networks have enabled us to build robots and virtual agents that can learn a variety of behavioral tasks. However, their learning capabilities are strongly dependent on a number of *hyperparameters*, such as learning rates and model complexity. The permissible ranges of such hyperparameters are dependent on particular tasks and environments, making it necessary for a human expert to tune them, usually by trial and error. This is why most learning robots and agents to date can only work in the laboratories.

This is in a marked contrast with learning in even the most primitive animals, which can readily adjust themselves to unpredicted environments without any help by a supervisor. This commonsense observation suggests that the brain has a certain mechanism for *metalearning*, a capability of dynamically adjusting its own hyperparameters of learning. A candidate of such a regulatory mechanism in the brain is the diffuse neuromodulator systems that project from the midbrain and the brainstem toward the entire brain including the cerebral cortex and the cerebellum. Most notable of such neuromodulators are dopamine, serotonin, noradrenaline, and acetylcholine.

In order to understand the mechanism of metalearning in natural behaving systems, the theory of reinforcement learning (RL), which has been developed for artificial agents that learn to optimize their behaviors through interaction with the environment, could provide a comprehensive computational framework.

Central to the theory of reinforcement learning is the value function of a state:

$$V(s(t)) = \mathrm{E}[\, r(t) + \gamma\, r(t+1) + \gamma^2\, r(t+2) + \ldots]$$

where $r(t)$, $r(t+1)$, $r(t+2)$,… denote the reward acquired by following a certain action policy $s \rightarrow a$ starting from the initial state $s(t)$. The discount factor $0 \le \gamma \le 1$ specifies how far into the future rewards are taken into account. The optimal policy that maximizes the above expectation of cumulative reward is obtained by solving the Bellman equation:

$$V(s) = \mathrm{argmax}_a\, [\, r(s,a) + \gamma\, V(s'(s,a))]$$

where $s'(s,a)$ is the state reached by taking an action a at state s. What this equation says is that when taking an action a, both the

immediate reward $r(s,a)$ and the future cumulative reward $V(s'(s,a))$ should be taken into account.

The relative merit of taking an action a at state s

$$\delta(s,a) = r(s,a) + \gamma\, V(s'(s,a)) - V(s),$$

which is called the *temporal difference* (TD) signal, can be used both for action selection and value function learning. A common way of stochastic action selection to facilitate exploration is the Gibbs sampling method:

$$\mathrm{Prob}(a(t)=a_i) = \exp[\beta\, \delta(s(t),a_i)] / \Sigma_j \exp[\beta\, \delta(s(t),a_j)],$$

where β is a parameter that controls the randomness of action choice, called the inverse temperature.

The estimate of the value function is updated by

$$V(s(t)) := V(s(t)) + \alpha\, \delta(s(t),a(t))$$

where α is the learning rate.

Based on a large body of neurobiological data and computational modeling studies, I propose the following hypotheses:

1) The dopaminergic system encodes the relative merit δ.
2) The serotonergic system controls the time scale of evaluation γ.
3) The noradrenergic system controls the inverse temperature β.
4) The acetylcholinergic system controls the learning rate α.

The theory of reinforcement learning provides a clue as to how these hyperparameters should be adjusted in reference to the task and the environment. The above hypotheses lead to predictions about the effect of neuromodulators on learning behaviors, the environmental effects on the neuromodulatory systems, and the appropriate balance between the levels of neuromodulators. The comparison of such predictions with experimental data would help us better understand the metalearning mechanism of the brain.

Neurobiological studies of emotion have so far focused on the role of emotion as the 'emergency programs' of behavior, such as escaping and freezing. However, the role of emotion in modulating cognitive and behavioral learning systems is highly important; many of affective and mental disorders occur as a result of the 'runaway' of learning systems. Consideration of the emotion as the metalearning system enables a novel computational approach in which the studies of learning theory, autonomous agents, and the neuromodulatory systems can be bound together.