A neural model of frontostriatal interactions for behavioral planning and action chunking

Nicola De Pisapia, Nigel Goddard

Institute for Adaptive and Neural Computation, University of Edinburgh, 5 Forrest Hill, Edinburgh EH1 2QL, UK

Abstract

Neurobiology and neuropsychology sustain the view of the Prefrontal Cortex as a key region, in connection with others, for behavioral planning, i.e. for the generation and selection of goal oriented complex courses of actions. We investigate the role of the Prefrontal/Basal Ganglia system using a reward based computational model. In particular, we focus on how these two regions learn incrementally to chunk sequences of actions, thus allowing fast and hierarchically structured planning. We test a model of chunking in a motor and in a cognitive planning problem.

Keywords: Prefrontal cortex, Basal Ganglia, Dopamine, Planning, Chunking

Summary

The Prefrontal Cortex (PFC) is known to be involved in generating and selecting purposeful complex sequences of actions, i.e. in behavioral planning of how to reach goals in a changing environment [5]. Anatomically, the PFC receives bottom-up connections from other cortical and subcortical areas, so it can get information about the current state of the environment and bodily needs. The PFC also sends top-down connections to the same brain areas, thus checking and evaluating possible outcomes of courses of actions as selected from the available repertoire, updating internal models of the environment (i.e. how the environment changes according to which actions are chosen), and reorganizing these areas depending on previous rewards or punishments [4].

In this paper we focus on the PFC/Basal Ganglia (BG) system, and we propose a model of how it might implement a neural mechanism for action chunking in the service of planning. The BG, and in particular the Striatum, which constitutes the input nucleus receiving connections from the PFC, are known to be involved in the acquisition of sensory motor habits and behavioral repertoires [1]. Experimental data show an high degree of modifiability in the corticostriatal connections, and the key role of the dopaminergic signal as an error in the prediction of rewards. This has allowed other researchers to construct computational models of the BG using Reinforcement Learning (RL), and in particular actor/critic models and Temporal Difference (TD) learning [2]. TD models learn to associate states of the environment with a value function $V(t)$, where t is time, which predicts future rewards. The value function allows the system to select which actions lead to the maximal reward, using an error signal $\delta(t) = r(t) + \gamma V(t+1) - V(t)$ in the expectations of the rewards, where r is the reward, and $\gamma<1$ is a discount factor. This signal resembles actual dopaminergic activity in the BG, and it allows incremental learning about reward expectations. These previous models have investigated how the dopaminergic signal might construct in the BG sequences of goal oriented actions, i.e. motor habits. In the present work we investigate further, and make the hypothesis that a motor habit, once planned and several times successful in achieving its goal, is chunked into a higher order action. Entire hierarchies of action chunks are built in this way, where higher level chunks are sequences of lower level chunks. This hierarchical constructive development allows the agent fast and efficient planning, because when new and complex goals must be achieved, the information about how action chunks can be used to reach subgoals doesn't need to be reprocessed. To investigate such possibility, we use TD together with Temporal Abstraction (TA), a recent theoretical development in RL [3]. TA is a mathematical description of how to construct hierarchies of chunks. If an agent has to reach a goal state, it is allowed to choose among different "options", by predicting in internal models what will be the final state of the environment if it follows whole courses of actions, which can be temporally extended, stochastic, and contingent on previous events. An option is a closed-loop policy: once initiated, it decides which actions to take for some time, and then terminates. When an option terminates, the agent can choose to start others, until it reaches its goal state. In this hierarchical and temporally abstract decision making process, the agent achieves faster planning than if using just one level of time scales, simply because each option, when considered at its higher level, takes one time step, even if its actual execution might take several steps.

In the present work we connect neural responses to high level behavior in complex and changing environments by modeling modules that build and use action chunks of this kind. In particular, a module where internal models of the world are generated and maintained in short term memory (STM) is interpreted as the Dorsolateral PFC; a module where internal models are evaluated in respect to context and needs is interpreted as the Orbitofrontal and Ventromedial PFC regions; a module where the actions, under the directives of the two other modules, are incrementally chunked and organized following a hierarchical constructive development is interpreted as the Striatum.

A consequence of this model is that perceptual-motor and intellectual skills can be acquired and processed in essentially the same way. In fact, we tested our model both in a motor and in a cognitive domain. In both tests the agent has a limited computational capacity in STM, therefore it plans considering only a limited number of states of the environment per trial. In the first test, an agent is given several tasks consisting of reaching different goal locations in a grid world. In the second test, an agent plays the Tower of Hanoi, a game used in cognitive psychology to study planning capacities in humans (there are five disks of different size and three pegs. Any disk can be moved to any empty peg or onto any other disk of bigger size. Given any starting configuration, the task is to reach a goal configuration in the shortest number of moves). Initially our agents, in both tests, are capable only of basic actions, and therefore of achieving only simple goals, given the limited STM. But as learning takes place, and several rewards are accumulated, the agents become capable of chunking some useful sequences of actions, and therefore of planning how to achieve more and more complex goals.

References

[1] A. M. Graybiel, The basal ganglia and chunking of action repertoires, Neurobiology of learning and memory 70 (1998), 119-136.

[2] P. R. Montague, P. Dayan, T. J. Sejnowski, Framework for mesencephalic dopamine system based on predictive hebbian learning, Journal of Neuroscience 16 (1996), 1936-1947.

[3] D. Precup, Temporal abstraction in reinforcement learning, PhD thesis (2000), University of Massachusetts Amherst.

[4] W. Schultz, Multiple reward signals in the brain, Nature 1 (2000), 199-207.

[5] J. Tanji, L. Hoshi, Behavioral planning in the prefrontal cortex, Current Opinion in Neurobiology 11 (2001), 164-170.