A Biologically Inspired Working Memory Framework for Robots

Joshua L. Phillips and David C. Noelle

({JOSHUA.L.PHILLIPS,DAVID.NOELLE}@VANDERBILT.EDU)

Vanderbilt University

Nashville, TN USA

Abstract

This work focuses on a particular neurocomputational account of working memory function that has been used to explain a wide range of working memory phenomena in terms of interactions between the prefrontal cortex and the mesolimbic dopamine system. Using the mechanisms prescribed by this theory, we have constructed a software toolkit for creating working memory modules for use in robotic control systems. The challenges faced by embodied robots are similar to those experienced by humans in everyday living, making this domain useful for testing the utility and scalability of this computational theory of working memory. We report the results of a feasibility study, involving a robotic version of the delayed saccade task, and we discuss future plans to test our working memory model in the context of robot control and learning.

Motivation

•Research in the cognitive sciences has provided substantial insight into the nature of working memory and the biological mechanisms that produce it.

•In the work, we have focused on one particular computational account of working memory function.

•Taking an unorthodox approach, We have begun the process of embedding this working memory mechanism into robotic control systems, with the goal of using robotic platforms as challenging test beds for our computational theory of working memory function.

•The difficulties that arise when robots interact with the world are representative of the tasks that humans encounter daily.

Working Memory

•Working memory is a central component of almost all theories of human cognition.

•Many theories of working memory exist (Miyake and Shah, 1999).

•We focus on a particular neurocomputational account of working memory which is described as a system that stores a small number of "chunks" of information, protecting them from interference from other processing systems and positioning them so as to directly influence the generation of behavior (Goldman-Rakic, 1987).

Neurocomputational Basis of Working Memory

•Regions of prefrontal cortex (PFC) play an important role in working memory (Goldman-Rakic, 1987).

•Dense recurrent connections in PFC are thought to support active maintenance of high firing rates through mutual excitation (Camperi and Wang, 1998).

•Recurrent excitation is not a sufficiently flexible mechanism to account for the fluidity of working memory function

Intelligent Updating

•Rroper control of updating is learned from experience.

•Retention of a particular kind of informational chunk in a given situation that results in reward, makes the system more likely to retain similar chunks in similar situations, a form of reinforcement learning.

•One candidate for a neural substrate for reinforcement learning involves the mesolimbic dopamine system.

•Recordings of dopamine cell firing in awake behaving animals suggest that dopamine cells fire in response to changes in expected future reward (Shultz et al., 1997).

•Such a measure of change in expected future reward is a key component of a machine learning algorithm known as temporal difference (TD) learning (Sutton, 1988).

•Thus, the midbrain dopamine system is well positioned to assist in the learning of the appropriate timing for covert actions, such as working memory updating (Braver and Cohen, 2000).

•This model has been successfully used to account for a variety of working memory phenomena, including deficits seen in schizophrenia and under focal frontal lesions (Braver and Cohen, 2000; O'Reilly et al., 2002).

Temporal Difference Learning (Sutton, 1988)

•The temporal difference (TD) learning algorithm (Sutton, 1988) is a powerful method for learning to select actions based on reinforcement signals: sporadic, scalar measures of how "good" or "bad" the current situation is.

•The goal is not to estimate this value, however, but to estimate the value function, V(s), of the situation in terms of the likely reward to be received in the future.

•For example, when playing chess it is sometimes desirable to sacrifice one of your pieces (a pawn) in order to win the game.

State-Value Estimation:

$$V(s)=\gamma^0 r(s)+\gamma^1 r(s+1)+\ldots+\gamma^n r(s+n)$$

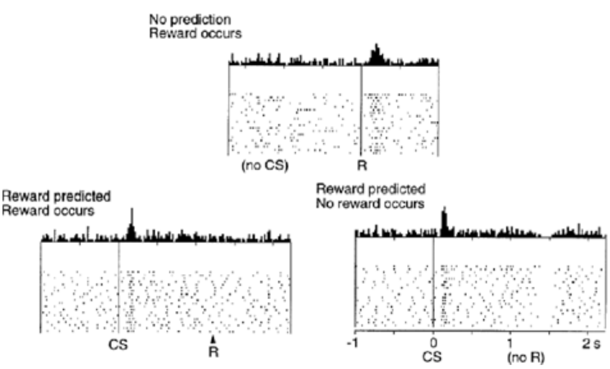
$$V(s)=\gamma^0 r(s)+\gamma^1 V(s+1)$$

$$V(s)=r(s)+\gamma V(s+1)$$

TD Error:

$$\delta(s)=\left[r(s)+\gamma V(s+1)\right]-V(s)$$

Dopamine Response to Conditioned Stimulus (CS) and Reward (R) (Shultz et al., 1997)



Working Memory in Robots

•We have generated an abstraction of our working memory model in the form of an open source software library that may be embedded in robot control software.

•As an initial demonstration of the functionality of this library, we have simulated a robotic version of a common neuroscientific test of spatial working memory: the *delayed saccade task*.

The Working Memory Toolkit

The Working Memory Toolkit (WMtk), is a software library which is general and flexible enough to be used on a variety of robotic platforms. The toolkit is written in ANSI C++ and consists of a set of classes and methods for constructing a working memory system that uses TD learning to select WM contents.

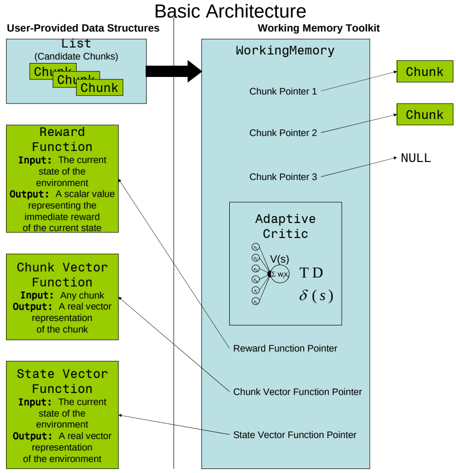
•On each time step of the task, a new list of candidate chunks is given to the `WorkingMemory` object by the robot control system.

•The system examines every possible subset of the collection of chunks that can fit within the limited capacity of the working memory.

•The combination of chunks that yields the highest estimate of future reward is the one that is selected, and all of the chunks in that subset are retained.

•The temporal difference error is then calculated to drive learning, using the reward function which determines the inherent "goodness" of the current state.

Basic Architecture

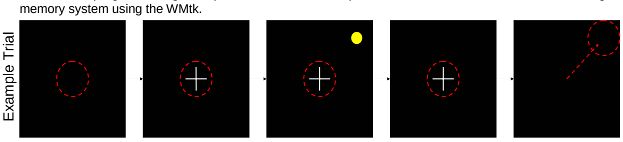


The Delayed Saccade Task

•In this task, the agent is expected to fix its gaze on an object in the center of the screen (a crosshair). Then another object (a brightly colored dot) is presented in the periphery for a brief period of time. Finally, once a "go" signal is provided (the crosshair vanishing), the agent is expected to shift its gaze to where the peripheral object had appeared earlier.

•Rather than program the agent to perform this task, we required it to learn correct behavior via a working memory system using the WMtk.

Example Trial



Simulation

•The capacity was set to three chunks, which was more than what was needed for this task.

•Three different kinds of chunks were considered for retention. These chunk types were:
- •"remember the location of the crosshair" (cross chunk)
- •"remember the location of the dot" (target chunk)
- •"remain fixated on whatever you're looking at" (fixation chunk)

•These chunks were generated by the agent's control system based on the current state of the environment.

•With regard to the reward function, a scheme was used that both matched standard practice in the primate laboratory and matched reward functions found in the reinforcement learning literature.

Results

•Average number of trials taken to reach criterion (20 correct trials in a row): 459.3 ± 25.4, n=1000.

•This number of trials is not unreasonable. It is much less than the number of trials typically needed to train monkeys on this task.

Automatic Processes and Controlled Processes

•The control system contained two parts:
- • Automatic Processes – behaviours automatically engaged unless actively blocked by WM contents.
- •Controlled Processes – behaviors driven by the presence of working memory chunks.

•The automatic behaviors of the agent were very basic, and they would not be able to reliably perform the delayed saccade task on their own.

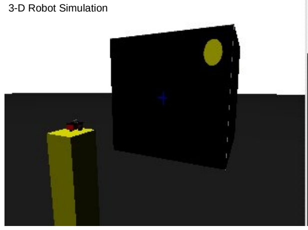
•Correct behavior would only be consistently produced if the working memory system learned, from experience, to retain appropriate chunks.

Future Work

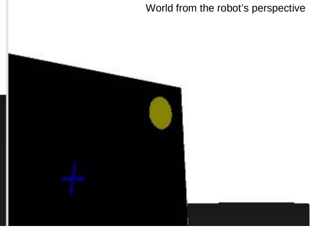
•Since the submission of this paper, we have moved this work into a full 3-D robot simulation program with very similar results as above. We hope to move this to actual hardware soon.

•We are also currently integrating the WMtk with an object recognition system embedded in a mobile robot for learning what environmental objects make for good landmarks (objects that are easily reacquired and are useful for localization).

3-D Robot Simulation



World from the robot's perspective



Guinness – Landmark Bot

References

Braver, T. S. and Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In Monsell, S. and Driver, J., editors, *Control of Cognitive Processes,* volume 18 of *Attention and Performance*, chapter 31, pages 713-737. MIT Press.

Camperi, M. and Wang, X.-J. (1998). A model of visuospatial working memory in prefrontal cortex: Recurrent network and cellular bistability. *Journal of Computational Neuroscience*, 5:383-405.

Goldman-Rakic, P. S. (1987). Circuitry of the prefrontal cortex and the regulation of behavior by representational knowledge. In Plum, F. and Mountcastle, V., editors, *Handbook of Physiology*, pages 373-417. American Physiological Society, Bethesda, MD.

Miyake, A. and Shah, P., editors (1999). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge University Press, Cambridge.

O'Reilly, R. C., Noelle, D. C., Braver, T. S., and Cohen, J. D. (2002). Prefrontal cortex and dynamic categorization tasks: Representational organization and neuromodulatory control. *Cerebral Cortex*, 12:246-257.

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

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, 3:9-44.