Learning to Use Memory

Nicholas A. Gorski (ngorski@umich.edu)
Computer Science & Engineering, University of Michigan
2260 Hayward St., Ann Arbor MI 48109-2121 USA

Keywords: Reinforcement learning; memory; working memory; declarative memory.

Introduction

Reinforcement learning (RL) provides a general approach to support intelligent agents that learn to act in their environments (Sutton & Barto, 1998). The foundational reinforcement learning algorithms of Q-Learning and SARSA, however, are purely reactive and thus not generally applicable to problems in which knowledge must be maintained in memory.

My research focuses on investigating how memory can extend the range of possible behaviors that RL can achieve, and in particular how RL agents can learn to use biologically-inspired memory models. In this context, *using memory* has two senses: first, making use of the knowledge that is retrieved from memory in order to better perform the task at hand, thus making use of the declarative knowledge from memory; second, selecting actions (such as encoding, storage and retrieval) over memory as appropriate for the task, thus using memory through procedural knowledge. One view of this research is that it is an attempt to discern which procedural knowledge over memory *must* be architectural and which *may* be adaptive.

Some prior work has begun to investigate this direction. We demonstrated that it is possible to learn to use a human-inspired episodic memory model in certain specific cases, but that in others an agent cannot learn the optimal control strategy (Gorski & Laird, 2009). Other researchers have also found that RL agents endowed with episodic and working memory models can learn to achieve some tasks, but not others (e.g. Zilli & Hasselmo, 2007).

My primary research question is: how and when can RL be used to learn to use memory? To address this in my thesis, I will perform a comprehensive empirical exploration of learning to use memory in order to better understand the dynamics that arise when an RL agent is endowed with an internal memory model. I will identify characteristics of tasks that can be explored independently across sets of parameterized problems. My initial exploration will begin with three memory models: a simple bit memory model, a gated working memory model (inspired by human working memory), and an associative memory model (inspired by human episodic memory). I precede a more detailed discussion of my research plans with an overview of my research to date.

Progress to Date

My research initially focused on learning to use Soar's episodic memory model (Derbinsky & Laird, 2009; Laird, 2008). Nuxoll (2007) had previously identified a set of cognitive capabilities that could be supported by episodic memory, and demonstrated agents that performed a subset of these capabilities. However, these agents required significant background knowledge and performed no learning. We studied whether it was necessary to provide the knowledge to utilize these cognitive capabilities, or whether RL could learn to use episodic memory in specific ways, and eventually performed specific cognitive capabilities solely as an emergent response to environmental and architectural constraints and pressures.

We succeeded in demonstrating agents that learned to perform two specific cognitive capabilities: virtual sensing, in which an agent uses episodic memory to recall a portion of the environment state that it cannot directly perceive; and remembering past actions, in which an agent uses knowledge of past actions to guide current behavior (Gorski & Laird, 2009).

In the course of this work, we found three interesting results. First, trivial-seeming changes to the environment had dramatic effects on how well agents were able to learn to use memory. Similarly, it can be very difficult to construct a task that is "just right" such that it elicits the desired cognitive capability and in which an agent uses memory in desired way.

Second, it is significantly easier to learn to perform virtual sensing than to use the knowledge that results from remembering past actions. When learning to perform virtual sensing, the agent was retrieving knowledge from memory that was a reliable indicator of the state of the environment, regardless of the duration of the agent's existence. However, knowledge of past actions was useful only after the agent had converged to a relatively stable behavior in the environment, as the knowledge that was retrieved was more sensitive to interference effects of taking a related action at an inopportune time.

Third, in certain settings agents converged to nearly optimal behaviors, but used episodic memory essentially as a single bit of memory (similar to the bit memory of Littman, 1994). Even though the learned behavior was suboptimal, it was a sufficiently stable equilibrium such that the agent was not able to find the globally optimal behavior through additional exploration.

The third result motivated us to explore using a bit memory model in the same domain (Gorski & Laird, forthcoming). In this work, we determined that while bit memory was sufficiently capable of being used to represent the optimal policy when the agent was provided with some initial background knowledge, the agent could not learn to use bit memory effectively. We additionally identified important ways in which bit memory differed from the episodic memory model.

Agents learning to use memory were sensitive to small changes in the task specification; furthermore, the behaviors of agents using different memory models were very different in the same domain. These results motivated a more comprehensive exploration of the space of tasks and memory models.

Research Plan

In order to understand the dynamics of learning to use memory, I propose a methodical and comprehensive empirical exploration of the space of possible tasks and memory models. As the space of possible tasks and memory models is infinite, it will be necessary to focus my empirical study on a particular set of tasks and memory models, which will be used to draw conclusions that can apply to tasks and memory more generally.

The tasks that I will explore have been selected on the basis of understanding how varying specific aspects (or characteristics) of a task affect the ability to learn to use memory in it. We have identified a very simple task, inspired by T-Maze tasks from the experimental psychology literature, that can be parameterized across independent dimensions. When these dimensions correspond to characteristics that are relevant to how memory must be used in a task, then observing the behavior that emerges in those tasks will inform how learning to use memory scales and what patterns of behavior take place in the course of the learning process.

The task characteristics that we are primarily interested in are those that directly relate to how knowledge must be retained while performing a task (we refer to this knowledge that must be maintained over time as *salient knowledge*). These characteristics include:

- The temporal delay between when salient knowledge is acquired and a task action that depends on it
- The quantity of salient knowledge that must be maintained simultaneously in a task
- The number of actions in a task that depend on salient knowledge.

I have identified a preliminary set of tasks that are parameterized along these relevant characteristics.

Exploring the space of memory models will require a different approach. While it is possible to design tasks that isolate individual characteristics and explore them over a parameterized task set, a given memory model cannot exist without architecturally committing to a number of simultaneous points in the various dimensions that define a memory model. Therefore, we will explore the space of memory models using a top-down approach.

We will explore bit memory, gated working memory, and an associative long-term memory in the context of the set of tasks discussed above. In a first pass, we will perform a comprehensive sweep exploring artificial agents that learn to use each memory model across all tasks (the cross product of memory models and tasks). After analyzing the results of this study, we will then modify the three memory models in an attempt to explore functional differences that they exhibit when an agent learns to use them, so as to be able to determine which characteristics of memory are directly responsible for supporting the necessary learning behavior, or not supporting it.

Throughout my investigation, my focus will be on the dynamics that arise between memory and task. I intend to be agnostic regarding specific RL algorithms as much as possible, and consistently apply the same algorithm (e.g. SARSA, Sutton & Barto, 1998) in all of my experiments.

My evaluation will focus on two issues: how agent performance scales with characteristics of task, and which characteristics of memory are most directly tied to which task characteristics.

Although my research is grounded in the field of artificial intelligence, I aim to draw conclusions from my work that inform cognitive scientists as to the nature of how procedural knowledge that uses memory (both controls it and makes use of the knowledge from it) can be learned. While most memory models assume some architectural basis for certain internal actions over memory, such as encoding and storage to long-term declarative memory, the procedural knowledge that governs memory retrievals and how that retrieved memory impacts task performance is adaptive. By better understanding in which tasks it is computationally feasible to learn to use specific memory models, we might better understand the constraints on human memory (and learning). In the field of artificial intelligence, learning to use memory is one approach to answering challenging problems of overcoming tasks with incomplete information while maintaining responsive learning and decision making.

References

Derbinsky, N. & Laird, J.E. (2009). Efficiently Implementing Episodic Memory. *Proceedings of the 8th Intl. Conf. on Case-Based Reasoning*. Seattle, WA.

Gorski, N.A. & Laird, J.E. (2009). Learning to Use Episodic Memory. *Proceedings of the 9th Intl. Conf. on Cognitive Modeling*. Manchester, UK.

Gorski, N.A. & Laird, J.E. (forthcoming). Learning to Use Episodic Memory. *Under Review.*

Laird, J.E. (2008). Extending the Soar Cognitive Architecture. *Proceedings of the 1st Conf. on Artificial General Intelligence*.

Littman, M.L. (1994). Memoryless Policies: Theoretical Limitations and Practical Results. *Proceedings of the 3rd Intl. Conf. on Simulation and Adaptive Behavior.*

Nuxoll, A. (2007). Enhancing Intelligent Agents with Episodic Memory. Ph.D. diss., Computer Science & Engineering, U. of Michigan, Ann Arbor.

Sutton, R.S. & Barto, A.G. (1998). Reinforcement Learning: An Introduction. Cambridge: MIT Press.

Zilli, E.A. & Hasselmo, M.E. (2007). Modeling the Role of Working Memory and Episodic Memory in Behavioral Tasks. *Hippocampus, 18*, 193-209.