Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 12 (2011) 144–153

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Learning to use episodic memory

Action editor: Andrew Howes

Nicholas A. Gorski *, John E. Laird

Computer Science & Engineering, University of Michigan, 2260 Hayward St., Ann Arbor, MI 48109-2121, USA

Received 22 December 2009; accepted 29 June 2010
Available online 8 August 2010

Abstract

This paper brings together work in modeling episodic memory and reinforcement learning (RL). We demonstrate that is possible to learn to use episodic memory retrievals while simultaneously learning to act in an external environment. In a series of three experiments, we investigate using RL to learn what to retrieve from episodic memory and when to retrieve it, how to use temporal episodic memory retrievals, and how to build cues that are the conjunctions of multiple features. In these experiments, our empirical results demonstrate that it is computationally feasible to learn to use episodic memory; furthermore, learning to use internal episodic memory accomplishes tasks that reinforcement learning alone cannot. These experiments also expose some important interactions that arise between reinforcement learning and episodic memory. In a fourth experiment, we demonstrate that an agent endowed with a simple bit memory cannot learn to use it effectively. This indicates that mechanistic characteristics of episodic memory may be essential to learning to use it, and that these characteristics are not shared by simpler memory mechanisms.

Keywords: Reinforcement learning; Episodic memory; Cognitive architecture; Artificial intelligence; Intelligent agents

1. Introduction

In this paper, we study a mechanism for learning to use the retrieval of knowledge from episodic memory. This unifies two important related areas of research in cognitive modeling. First, it extends prior work on the use of declarative memories in cognitive architecture where knowledge is accessed from declarative memories via deliberate and fixed cued retrievals (Anderson, 2007; Nuxoll & Laird, 2007; Wang & Laird, 2006) by exploring mechanisms for learning to use both simple and conjunctive cues. Second, it extends work on using reinforcement learning (RL) (Sutton & Barto, 1998) to learn not just control knowledge for external actions, but also to learn to control access to internal memories, expanding the range of behaviors that can learned by RL.

Earlier work has investigated increasing the space of problems applicable to RL algorithms by including internal memory mechanisms that can be deliberately controlled: Littman (1994) and Peshkin, Meulaeu, and Kaelbling (1999) developed RL agents that learned to toggle internal memory bits; Pearson, Gorski, Lewis, and Laird (2007) showed that an RL agent could learn to use a simple symbolic long-term memory; and Zilli and Hasselmo (2008) developed a system that learned to use both an internal short-term memory and an internal spatial episodic memory, which could store and retrieve symbols corresponding to locations in the environment. All four cases demonstrated a functional advantage from learning to use memory.

Our work significantly extends these previous studies in three ways: first, our episodic memory system automatically captures all aspects of experience; second, our system learns not only when to access episodic memory, but also learns to construct conjunctive cues and when to use them; and third, it takes advantage of the temporal structure of

* Corresponding author. Tel.: +1 734 763 0150; fax: +1 734 763 1260.
E-mail addresses: ngorski@umich.edu (N.A. Gorski), laird@umich.edu (J.E. Laird).

1389-0417/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2010.08.001

episodic memory by learning to advance through episodic memory when it is useful (this property is also shared by the Zilli & Hasselmo system, but for simpler task and episodic memory representations).

Our studies are pursued within a specific cognitive architecture, namely Soar (Laird, 2008), which incorporates all of the required components: perceptual and motor systems for interacting with external environments, an internal short-term memory, a long-term episodic memory, an RL mechanism, and a decision procedure that selects both internal and external actions. In comparison, ACT-R (Anderson, 2007) has many similar components but does not have an explicit episodic memory. Its long-term declarative memory stores only individual chunks, and it does not store episodes that include the complete current state of the system. To do so would require storing the contents of all ACT-R's buffers as a unitary structure, as well as the ability to retrieve and access them, without having the retrieved values being confused with the current values of those buffers. Moreover, ACT-R's declarative memory does not inherently encode the temporal structure of episodic memory, where temporally consecutive memories can be recalled (Tulving, 1983). While the work presented in this paper is specific to learning to use an episodic memory, similar work could be pursued in the context of ACT-R by learning to use its declarative memory mechanism. However, we are unaware of existing work in that area, and even if there were, it would fail to engage the same issues that arise with episodic memory.

2. Background

Soar includes an episodic memory that maintains a complete history of experience (Nuxoll & Laird, 2007), implemented so as to support efficient memory storage and retrieval (Derbinsky & Laird, 2009). Soar's working memory is a relational graph structure, consisting of nodes and links, similar to the structure of a semantic network. Complete "snapshots" of working memory are automatically stored in episodic memory following every processing cycle.

To retrieve an episode, a *cue* is created in working memory by the application of Soar's procedural knowledge, which is encoded as production rules (Laird, 2008). A cue is a partial specification of an episode, created in a special part of working memory. The episode that best matches the cue is retrieved to working memory. The degree of match is based on the number of elements in the cue found in an episode. If there are multiple episodes with the same degree of match, the most recent of those episodes is retrieved. Once an episode is retrieved to working memory, other knowledge (such as procedural knowledge) can access it. This style of cue-based retrieval process is similar to ACT-R's declarative memory retrieval process where procedural knowledge creates a cue in a retrieval buffer, and the declarative memory mechanism retrieves the appropriate chunk from the long-term store.

We refer to this type of cue-based retrieval as *deliberate*, to contrast it with *spontaneous* or automatic retrieval processes. A spontaneous retrieval process is automatic and depends on all the structures in working memory. Thus, an agent with spontaneous retrieval lacks control over when retrievals take place and what aspects of the situation are the basis for retrieval, whereas with deliberate control, the agent can control when episodic memory retrievals are initiated and what cues are the basis for retrieval.

After performing a cue-based retrieval, the agent can utilize the temporal structure of episodic memory and retrieve the next episode, providing a mechanism for the agent to move forward through its memories. This allows an agent to recall sequences of experiences.

Previously, Nuxoll (2007) created agents that used episodic memory to support a variety of capabilities. In that work, agents were given hard-coded procedural knowledge that specified when cues should be created for episodic memory, which structures should be used for cueing retrievals, and how to condition behavior based on the retrieved knowledge. The procedural knowledge was not tuned via learning (such as RL), so the agents *used* episodic memory, but did not *learn to use it*.

In this research, rather than endow agents with preexisting fixed control knowledge, we investigate: learning when to access episodic memory, learning what structures to use as cues, and learning how to condition behavior on the retrieved knowledge. All of these processes are using episodic memory, and this work then learns to use episodic memory in three different senses.

3. Well World

In order to explore how an agent might learn to use an internal episodic memory, we constructed several tasks within an artificial domain we call "Well World." The domain is simple enough to be tractable for an RL agent, but rich enough such that episodic memory can potentially improve performance.

The goal in Well World is to satisfy two internal drives: thirst and safety. Thirst is the agent's primary drive, and it seeks to satisfy that above safety. Thirst is satisfied by consuming water at a well that contains it, while safety is satisfied by consuming the safety resource at the shelter location.

Fig. 1 shows the base Well World environment. In the base configuration, there are three locations. At two of the locations, there are wells; at the third, shelter. In each location, the agent observes a set of attributes and values specific to that location, but does not perceive information

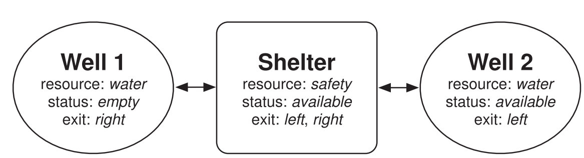

Fig. 1. Objects, resources, and adjacency in Well World.