

Memory-based Simultaneous Learning of Motor, Perceptual, and Navigation Skills

Paul Robertson (paulr@bbn.com)

BBN Technologies
10 Moulton street
Cambridge, MA 02138, USA
email: paulr@bbn.com

Abstract. A learning approach and associated experiments are described that allow robotic devices to learn the relationship between control actions and sensory feedback, to acquire skills, and to adapt to a changing environment. The approach frames the motor skill learning problem as one involving the co-evolution of perceptual and actuation structures. A memory-based learning approach is taken and a Cached Event String Automata (CESA) is introduced. The approach supports the learning of complex sensory-motor skills from a very small number of trials with incremental skill refinement. A key aspect of the learning approach advocated in this paper is the use of self-motivated exploration to refine the learning of sensory-motor skills.

Keywords: Artificial Intelligence, Robotics, Adaptive Behavior, Learned Models, Memory-Based learning.

1 Introduction

Herb Simon [10] introduced us to the idea that an ant walking through a complex environment may exhibit complex behavior but that the complex behavior is largely a result of the environment and that the ant may implement rather simple control laws. Rod Brooks [4] described a hierarchical architecture that would allow the implementation of simple control laws that interact with the environment to produce complex and behavior robust. Brooks argued against models by claiming that the best model of the world was the world itself. Richard Feynman [5] gave us an anecdote about an early, informal, experiment that he performed with ants in order to understand how they learned efficient routes between food and colony even though the process of discovering the food resulted from a random exploration process. The key involved the use of pheromones that would be left in the environment that the ants could follow to and from the food. Errors made by ants in following the correct path would sometimes lead to an ant taking a shortcut and accidentally getting back on the path. Such shorter paths would be reinforced by pheromone trails and so the path would gradually be optimized over time and sooner or later would look fairly optimal. Feynman's ants used the world as their model in the Brooks sense and learned paths by annotating the world using chemicals.

By remembering the perceived world as sequences of perceptions we can simulate the process of learning described by Feynman and by doing so learn to perform skills. By remembering everything, we can find what we need when we need it. While this kind of learning may seem inefficient it has many advantages. David Marr [6] noted:

“... the problem is not totally intractable despite the huge sizes of all the relevant event spaces. The reason is that only a very small proportion of the possible events can ever actually occur, simply because of the length of time for which a brain lives. This means, first, that the memory can be quite coarse; and secondly, that if anything much happens twice, it is almost certain to be significant.”

In this paper we describe experiments that allow a robot to learn (1) the relationship between motor actuation and perception; (2) skills for navigating in the robot’s world; and (3) learning cognitive maps of the robots world that facilitate path planning.

Generalization allows learned skills to be applied in new circumstances.

Our approach uses a memory based learning approach [7] in which memories of sensory inputs and actions are recorded for later use (in a cache). These memories are hierarchically organized and generalized. Skills are applied to new situations by finding relevant memories and replaying them. The cache is populated both from action/event sequences resulting from goal directed behavior and self-motivation [1, 8, 9] driven by a meta-strategy of exploration in poorly represented parts of the event space as a low priority process.

In the following sections we introduce ‘Cached Event String Automata’ (CESA), describe algorithms that permit CESA to learn behavioral skills in real world environments, demonstrate the approach on a simple robotic platform, and discuss performance implications of the approach.

2 Cached Event String Automata

2.1 Overview of the Approach

Sequences of sensor inputs such as vision, leg position feedback and haptics; and actions such as actuation commands are recorded. A goal is specified by presenting a solution in perceptual terms with relevant details of the goal state present. We refer to these sequences of perceptual and actuation events as event strings – or strings for short. Our approach stores event strings in a manner that permits rapid retrieval and new complex activities are formed by stitching together fragments of generalizations of prior strings.

String stitching looks up previously learned sequences that when joined together get from the current state to the final state. Any action taken by the robot represents a choice on the part of the robot to perform that action. Strings are therefore hinged at action points. We consider sequences to be Markovian in that it doesn’t matter how a robot arrived at a given state.

2.2 Surprise-Based Exploration

Our learning approach depends upon reusing fragments of old skills in order to build new ones. The approach therefore depends upon there being pre-existing skills. An important question therefore is ‘where do the initial skills come from?’. There are a number of potential solutions to the problem of prior skills:

1. The initial sequences could be hand generated;
2. The robot could be tele-operated through a certain skill a number of times – each time recording the sequences of observations and actions; and
3. The robot could explore its world by itself.

(1) is probably not feasible, (2) is potentially very useful and is similar to the way we teach humans. (3) provides a way for the robot to automatically adapt when the environment changes. Our experiments to date have focused mostly on (3). Current experiments with sensory-motor learning with a humanoid robot are exploring (2).

Surprise-based learning is implemented as a low-priority process that is used to soak up free time resources such as when the robot is already in a goal state and is thus free to play.

When in a goal state (ie: nothing to do) the robot randomly (weighted by interest) choose to:

1. Do nothing.
2. Pick one of the available actions.

Existing strings predict an outcome (next element on the string). If the string base accurately predicts the outcome the interest level in the action is decreased. Otherwise the interest level (Surprise) is increased. When the robot encounters the opportunity to make an action for which it has a high interest it may decide to experiment with the action with increased probability proportional to the level of interest that has accumulated from prior encounters. As the robot experiments with actions of interest the cache fills up with strings that predict the outcomes of the actions and as they become more accurate the level of interest is decreased and the system gradually settles down to no exploration because of lack of surprise.

Whenever an action yields an unpredicted response the surprise is noted and later when the context is appropriate exploration will take place.

String Hinge Spreading Activation A string is an instantiation of a path through a MDP. Actions represent decisions to follow a path through an MDP. A string can therefore be cut at any action (hinge) along the string. Take a string, light up the hinges and find follow on strings from the hinges to the goal. Finds solutions in order of least number of splices. Collect solutions until an adequate solution is found then invoke the best found up to that point.

A CESA is a 5-tuple $\langle \omega_1, \omega_g, A, \Omega, S_{A\omega} \rangle$

where: $\omega_1 \in W$ is the initial observation $\omega_g \in W$ is the goal observation A is the set of actions, Ω is the set of observations, $S_{A\omega}$ is a set of cached strings whose elements $e_i \in A \cup \Omega$, Learning takes a CESA and produces a new CESA' that changes the $S_{A\omega}$ by adding, removing and or otherwise modifying the entries.

3 Relationship of CESA to HMM's

We have already mentioned that the strings stored in the CESA memory are Markovian in nature and that actions in a string represent decision points. Each recorded string represents a path taken through an HMM. If there were a sufficient number of them the probabilities of transitions at the decision points could be calculated. A CESA then evolves into a non-compact representation of an HMM. If the environment changes the structure of the HMM can change as new sequences are experienced. By not compactly representing the HMM we are able to adapt to a changing environment without additional mechanism.

A single string allows the activity represented by that string to be learned. Generalization is performed by performing nearest neighbor matching. When there is only one string the nearest neighbor will be the that string but as more strings are cached the nearest neighbor match becomes more accurate and string splicing allows for learned fragments to be assembled into complex activities. As more strings are learned the probabilities at decision points are gradually estimated.

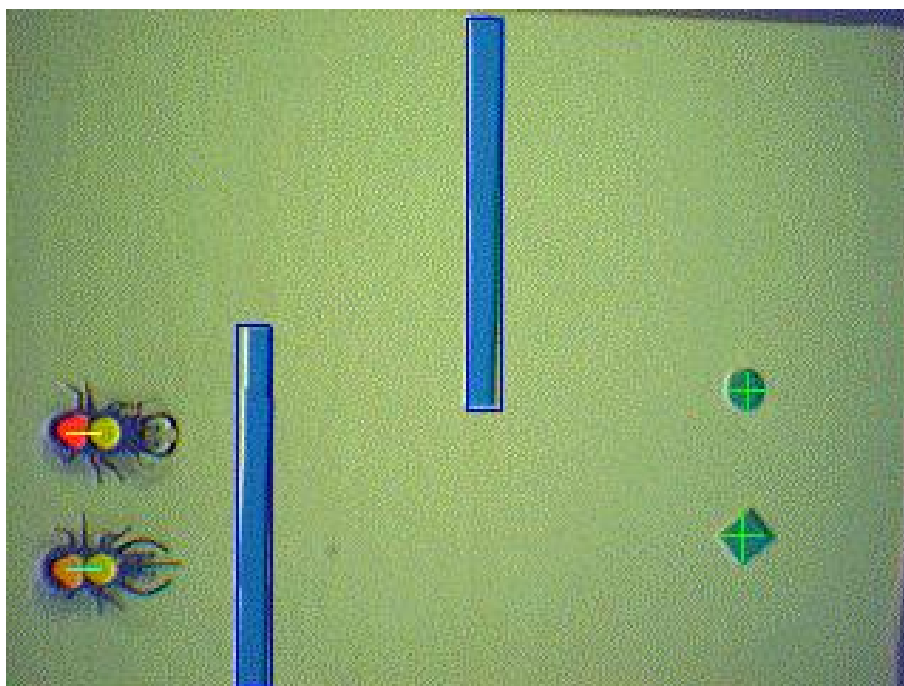


Fig. 1. Experimental Robot Learning Platform

The approach described above resembles spreading activation [2] and motor procedures encoded in human Basal Ganglia [3].

Experiments to date have successfully demonstrated the ability for a simple insect-like robot to simultaneously learn motor procedures, image interpretation procedures, and a cognitive map so that the robots can avoid obstacles as seen through their vision inputs and get from a starting position to a target position through a maze after sufficient trials in the maze.

4 Conclusions

In the workshop we will present experimental results of CESA learning on two robotic platforms, the insect-like robots shown in Figure 1, and a humanoid robot with 20 degrees of freedom. We will discuss techniques developed to achieve acceptable performance of CESA learning.

References

1. S. Singh, A. G. Barto and N. Chentanez. Intrinsically motivated learning of hierarchical collections of skills. In *Proceedings of the 3rd International Conference on Developmental Learning (ICDL 04)*, 2004.
2. J. R. Anderson. A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22:261–295, 1983.

3. G. S. Berns and T. J. Sejnowski. A computational model of how the basal ganglia produce sequences. *Journal of Cognitive Neuroscience*, 10.
4. R.A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2:14–23, 1986.
5. R.P. Feynman. *'Surely You're Joking, Mr. Feynman!'* (*Adventures of a Curious Character*). W. W. Norton and Company, 1985.
6. D. Marr. A theory for cerebral neocortex. *Proceedings of the Royal Society of London B*, 176:161–234, 1970.
7. Andrew Moore. A tutorial on kd-trees. Extract from PhD Thesis, 1991. Available from <http://www.cs.cmu.edu/simawm/papers.html>.
8. F. Kaplan P. Oudeyer and V.V. Hafner. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11, April 2007.
9. D. Precup R. S. Sutton and S. Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112:181–211, 1999.
10. H.A. Simon. *Sciences of the Artificial*. MIT Press, 1996.