Berthouze, L., Prince, C. G., Littman, M., Kozima, H., and Balkenius, C. (2007).
Proceedings of the Seventh International Conference on Epigenetic Robotics: Modeling
Cognitive Development in Robotic Systems. *Lund University Cognitive Studies, 135*.

When Less is More: Sensor resolution and learning

Josh de Leeuw and Ken Livingston
Cognitive Science Program, Vassar College
Poughkeepsie, NY 12604
jodeleeuw@vassar.edu livingst@vassar.edu

Abstract

Simulated mobile robots running a standard Q-learning algorithm learned basic navigational behavior in two different environments varying in complexity. Sensor resolution was manipulated to examine the effects of different levels of information on the learning task. In the simpler of the two environments the lowest sensor resolution resulted in the most efficient learning. In the more complex environment higher resolution sensors resulted in better learning. This is the first of a series of studies designed to explore the joint effects of environmental complexity, task complexity, and sensory resolution on autonomous learning.

1. Introduction

The *Markov Decision Process* approach to autonomous learning works well when the environment can be fully characterized, but for mobile robots operating in natural environments, even highly simplified natural environments, it can be difficult to construct such a characterization. Standard Q-learning techniques (Watkins, 1989) require that one describe a space of possible state-action pairs, but if states are given by outputs from continuously varying sensors (e.g., sonar, infra-red emitter detectors, laser range-finders), the number of possible states quickly becomes uncountably large. There are a number of well-documented solutions to this problem, including the use of various approximation functions. For the most part researchers appear to focus on methods for approximation, including using nerual network models (Lin, 1992), and different types of sparse-coarse-coded function approximators (Santamaria et al., 1998), that can generalize across the whole state space. But what degree of approximation will produce the most efficient learning? Our goal was to begin an exploration of the relationship between the complexity of the environment to be mastered and the number of sensor sub-ranges or bins during learning of a simple task. Creating artificial sensor categories is a basic strategy for function approximation (McCallum, 1995) that allows for direct control over how much detail is captured by the robot. Our hypothesis was that as environmental complexity increased so would the optimal number of bins.

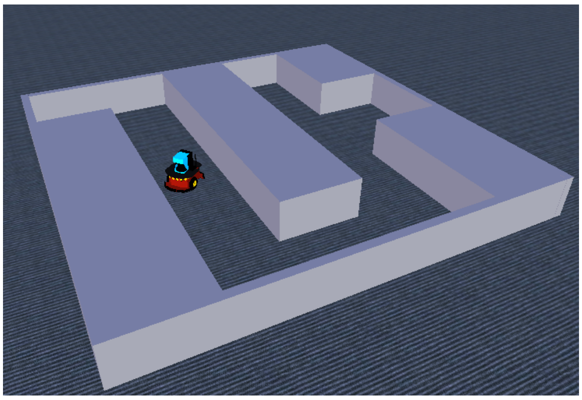

Figure 1: The simulated robot in the maze environment

2. Experiment

Microsoft Robotics Studio[1] was used to create a simple walled arena in which a model of a Pioneer-3DX robot was placed. Three range-finding laser sensors were simulated, one looking straight ahead at 0 degrees, and two others at plus and minus 45 degrees. Continuous ranging was sliced into 2, 4, 6, 8, and 10 bins. Bump sensors were also simulated on both the front and rear of the robot. Two versions of the environment were created, one empty of objects resulting in a wide open space surrounded by a barrier, and a second that resembled a maze with narrow corridors (see Figure 1). The action space consisted only of forward, backward, left turn, right turn, and stop. The Q-learning update rule used was

$$Q\left(s_t, a_t\right) + \alpha\left[r_t + \gamma max_a Q\left(s_{t+1}, a\right) - Q\left(s_t, a_t\right)\right]$$

with α (learning rate) = .30 and γ (discount rate) = .50. Positive reward values were assigned when the robot moved forward or backward. Turns in place were rewarded initially, though the reward for each successive turn was decreased and could become neg-

[1] Available at http://www.microsoft.com/robotics

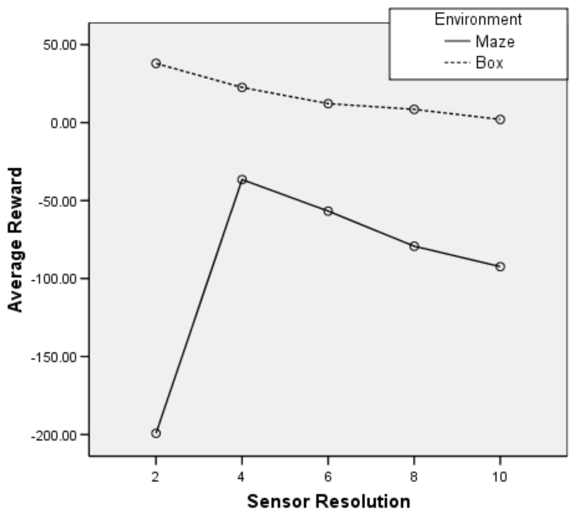

Figure 2: Agents were more competent in the open environment than the maze. Two bin sensor systems were most effecient in the open environment, but failed in the maze environment.

ative. Stops were negatively valued with the magnitude increasing the longer the robot stopped. Collisions were assigned the greatest negative reward. There were eight runs in each of 10 conditions (2 environments x 5 levels of resolution). A run consisted of 1000 iterations of the Q-learning algorithm, one every half-second for a total of just over eight minutes of learning. Each run is the functional equivalent of running a new robot with different start parameters through the simulation. This allows data from each robot to be treated as independent for purposes of subsequent data analysis.

3. Results and Discussion

Rewards earned over 1000 iterations were averaged into 10 trial blocks and subjected to a 2 (environmental complexity) by 5 (number of bins) repeated measures ANOVA. As Figure 2 shows, the open environment was significantly easier for robots to master than the maze ($F(1,70) = 90.838$, $MSE = 2396238.117$, $p < .001$). Post hoc t tests (alpha = .05 adjusted) showed that in the open environment the two bin sensor system performed significantly better than all others. However, in the maze environment the two bin system failed to learn at all, while the other systems all performed significantly better than the two bin system.

As predicted, the more complex environment demands greater sensory detail for efficient learning. In this particular experiment, the two bin sensor system was unable to learn in the maze environment. Yet the two bin system outperformed all others in the open environment, suggesting that an ideal learning agent would be able to tune its sensory system actively in response to environmental conditions. These findings have led us to begin a search for algorithms analogous to mechanisms of attention in biological agents. The fundamental biological limits of resolution for a sensor are not easily modified, but learning can be much more efficient if only such information as is necessary for a task is attended. Furthermore, as the agent acquires competent behavior in a given environment, it may be beneficial to increase sensory resolution as part of the development of the agent, assuming that the algorithms in place can generalize from old information to new information (Gomez et al., 2004). Viewed in this light, initial restrictions on sensor resolution that allow for more effecient learning may be an important part of ontogenetic development (Turkewitz and Kenny, 1982). In work now under way we seek to exploit this principle in robots by finding algorithms that tune sensor resolution to the demands of task and environment, based on prior learning as well as the rate of progress in current learning.

References

Gomez, G., Lungarella, M., Hotz, P., Matsushita, K., and Pfeifer, R. (2004). Simulating development in a real robot: On the concurrent increase of sensory, motor, and neural complexity. In *Proceedings of the Fourth International Workshop on Epigenetic Robotics*, pages 1–24.

Lin, L. (1992). Self-improving reactive agents based on reinforcement learning. *Machine Learning.*

McCallum, A. (1995). Instance-based utile distinctions for reinforcement learning with hidden state. In *Proceedings of the International Conference on Machine Learning*, pages 387–395.

Santamaria, J., Sutton, R., and Ram, A. (1998). Experiments with reinforcement learning in problems with continuous state and action spaces. *Adaptive Behavior.*

Turkewitz, G. and Kenny, P. (1982). Limitations on input as a basis for neural organization and perceptual development: a preliminary theoretical statement. *Developmental Psychobiology.*

Watkins, C. (1989). *Learning from delayed rewards.* University of Cambridge, England.