Automatic Development from Pixel-level Representation to Action-level Representation in Robot Navigation

Jefferson Provost
Center for the Neural Basis of Cognition and
Department of Neuroscience, University of Pittsburgh
115 Mellon Institute
Pittsburgh, PA 15213
jp@cnbc.cmu.edu

Many important real-world robotic tasks have *high diameter*, that is, their solution requires a large number of primitive actions by the robot. For example, they may require navigating to distant locations using primitive motor control commands. In addition, modern robots are endowed with rich, high-dimensional sensory systems, providing measurements of a continuous environment. Reinforcement learning (RL) has shown promise as a method for automatic learning of robot behavior, but current methods work best on low-diameter, low-dimensional tasks. Because of this problem, the success of RL on real-world tasks still depends on human analysis of the robot, environment, and task to provide a useful sensorimotor representation to the learning agent.

A new method, Self-Organizing Distinctive-state Abstraction (SODA) Provost, Kuipers, & Miikkulainen (2006); Provost (2007) solves this problem, by bootstrapping the robot's representation from the *pixel-level* of raw sensor input and motor control signals to a higher *action-level* consisting of distinctive states and extended actions that move the robot between these states. These new states and actions move the robot through its environment in large steps, allowing it to learn to navigate much more easily and quickly than it would using its primitive actions and sensations. SODA requires no hand-coded features or other prior knowledge of the robot's sensorimotor system or environment, and learns an abstraction that is suitable for supporting multiple tasks in an environment.

Given a robot with high-dimensional, continuous sensations, continuous actions, and one or more reinforcement signals for high-diameter tasks, the agent's learning process consists of the following steps.

1. **Define Primitive Actions.** The agent defines a set of discrete, short-range, local actions to act as the primitive motor operations. They are a fixed discretization of a learnable abstract motor interface consisting of a set of "principal motor components".

2. **Learn High-level Perceptual Features.** The agent explores using the primitive actions, and feeds the observed sensations to a growing-neural gas network Fritzke (1995) that organizes its units into a set of high-level perceptual features. These features are prototypical sensory impressions used both as a discrete state abstraction suitable for tabular reinforcement learning, and a set of continuous perceptual features for continuous control.

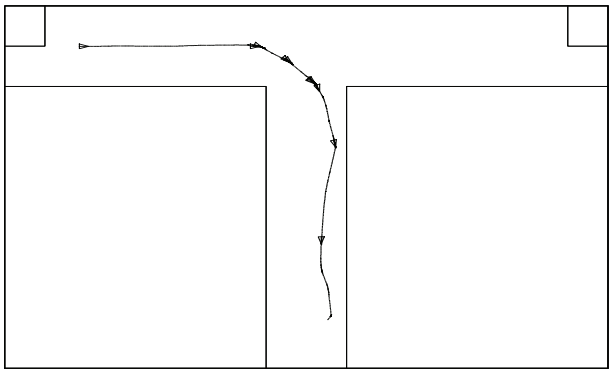

Figure 1: **Navigation using SODA**. This trace shows the path of a robot navigating in a T-shaped maze environment using SODA. The triangles show the starting points of the learned high-level actions. The narrow line shows the path of the robot using low-level motor commands. A path that required hundreds of actions in the lower, pixel-level representation requires only ten in the high-level action-based representation.

3. **Learn High-level Actions.** Using these new features, the agent defines perceptually distinctive states as points in the robot's state space that are the local best match for some perceptual feature, and creates actions that carry it from one distinctive state to another. The actions are drawn from the Spatial Semantic Hierarchy Kuipers (2000), a topological theory of spatial navigation. They are constructed as compositions of (1) trajectory-following control laws that carry the agent into the neighborhood of a new distinctive state, and (2) hill-climbing control laws that climb the gradient of a perceptual feature to a distinctive state. These actions are formulated as Options Sutton, Precup, & Singh (1999), whose control policies are learned through hierarchical reinforcement learning.

4. **Learn Tasks.** The agent attempts to learn its task using Sarsa, or another well-known and simple reinforcement learning method with a tabular state and action representa-

tion, using the perceptual categories in the self-organizing feature map as the state space, and the the new high-level actions as actions.

SODA has been tested in a simulated robot with a 180-element laser rangefinder, an eight-point compass, and a drive-and-turn motor system with realistic motor noise. The agent learned in two environments, the 10 m $\times$ 6 m T-maze in Figure 1, and a realistic, 40 m $\times$ 40 m simulation of a floor of the ACES building at the University of Texas at Austin. In each environment, the agents using SODA were able to learn to navigate dramatically faster than agents using primitive actions. Moreover, SODA gave an even greater advantage in the large environment than in the small one. In addition, the learned representation contained many meaningful large-scale actions, such as trajectory-following down a corridor to an intersection. These actions greatly reduce the diameter of navigation tasks. An example navigation trace showing the starting points of the large scale actions is shown in Figure 1. In this example, a navigation task requiring more than 300 primitive actions is accomplished with ten abstract actions.

Beyond the improvements in task diameter, a large part of the power of SODA comes from the fact that it converts a high-dimensional, continuous, fine-grained sensorimotor representation into an atomic, discrete, coarse grained symbolic representation suitable as input to many simple and well-studied learning algorithms. This allows the agent to learn to navigate in a continuous environment using ordinary tabular reinforcement learning instead of more complex function approximation methods that must be tailored for individual problems. This representation change is accomplished by delegating the details of low-level control to the well-defined and circumscribed trajectory-following and hill-climbing options. These options use function approximation for policy learning, but because each option encapsulates a simple, spatially constrained subtask, they are easy to specify and easy to learn.

Furthermore, SODA's continuous-to-discrete abstraction facilitates its use as a building-block in a larger *Bootstrap Learning* process Kuipers *et al.* (2006) for robot learning, in which the representations learned at lower levels are used as input to higher level learning processes. For example, in partially observable environments, SODA's representation is readily processed by existing methods such as POMDPs or Predictive State Representations Littman, Sutton, & Singh (2002), without the need to adapt these methods to continuous spaces.

In addition to investigating and evaluating SODA as a substrate for higher-level bootstrap learning, planned future work on SODA includes the formulation of a "factored SODA" for use in robots with very large and complex sensorimotor systems for which a single set of perceptual categories is insufficient for supporting intelligent behavior. Such a system would have separate sets of perceptual features for different groups of closely-related sensors, each set of perceptual features would have its corresponding set of actions, that move the robot with respect to those features. Such a system would have more available actions to the agent, and would abstract its input into a set of tuples, rather than a single perceptual category. Nevertheless, factored SODA should still be amenable to bootstrapping to many higher level methods.

To summarize, SODA automatically changes a robots representation of its world from a continuous, high-dimensional representation of raw sensorimotor experience into one of symbolic percepts and large-scale, abstract actions. This representation change allows robotic agents to learn to navigate quickly using off-the-shelf reinforcement learning algorithms, and is amenable for use in bootstrapping to still higher representation levels.

Acknowledgements

This work was supported by the generous support of the National Science Foundation (NSF grants IIS-0413257 and EIA-0303609) and National Institute of Mental Health (NIMH grant R01-MH66991).

References

Fritzke, B. 1995. A growing neural gas network learns topologies. In Tesauro, G.; Touretzky, D. S.; and Leen, T. K., eds., *Advances in Neural Information Processing Systems 7*. Cambridge, MA: MIT Press. 625–632.

Kuipers, B.; Beeson, P.; Modayil, J.; and Provost, J. 2006. Bootstrap learning of foundational representations. *Connection Science* 18(2).

Kuipers, B. J. 2000. The spatial semantic hierarchy. *Artificial Intelligence* 119:191–233.

Littman, M.; Sutton, R. S.; and Singh, S. 2002. Predictive representations of state. In *Advances in Neural Information Processing Systems*, volume 14, 1555–1561. MIT Press.

Provost, J.; Kuipers, B. J.; and Miikkulainen, R. 2006. Developing navigation behavior through self-organizing distinctive-state abstraction. *Connection Science* 18.2.

Provost, J. 2007. *Reinforcement Learning in High-Diameter, Continuous Environments*. Ph.D. dissertation, The University of Texas at Austin, Austin, Texas.

Sutton, R. S.; Precup, D.; and Singh, S. 1999. Between MDPs and SMDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence* 112:181–211.