Incremental Life-long Learning by Demonstration

Nathan Koenig and Maja Matarić
Interaction Lab, Computer Science Dept.
Viterbi School of Engineering
University of Southern California
Los Angeles, CA 90089
nkoenig, mataric@usc.edu

Most robot systems are designed to operate within a single task domain. Each additional task or environment condition requires significant human effort. The goal of this work is to develop a robot control architecture that alleviates the human programming bottleneck by allowing the robot to learn, store, and reuse task policies. This architecture will allow a robot to continually adapt to new environments, users' desires, and its own changing physical capabilities.

Task level learning is the primary focus of this work, where *tasks* are defined as complex and long-term decision process. Demonstrations of tasks, performed by human teachers, boostrap the learning process. In [1], [5] a robotic arm learned the pendulum swing-up and juggling skills by seeding the learning process with human demonstrations. Although this learning is strongly applicable to joint trajectories, it does not extend well to task learning. Functional learning has also been applied at the task level, but is limited by task complexity [4]. To overcome these limitations, policies consisting of discrete actions are often learned [2], [3]. This approach is inherently limited to a small range of tasks due to the restricted action set.

The challenges described above necessitate the creation of a general purpose task learning architecture that can be applied to a range of tasks, robots, and environments. Tasks should not be limited by a predetermined set of actions, nor should they be constrained to a particular robot configuration. Furthermore, the system should be capable of continually learning new tasks, and intelligently applying prior knowledge toward these new tasks. Our work can be summarized as developing a system capable of *life-long learning of tasks* limited only by a robot's physical capabilities.

As a motivating example, consider a mobile manipulator that should ultimately accomplish a search and rescue task. Teaching or programming the robot to perform such a complex task is rather difficult due to the scope of the problem and dynamic environmental conditions. Instead, this task can be broken down into a few sub-tasks, such as search and human retrieval. Searching can be further decomposed into exploration and human identification. Finally, exploration can be defined as a combination of navigation tasks, such as hallway and doorway traversals, and maintenance of a map. As one descends the task hierarchy, the tasks become more general purpose. Navigation tasks can be applied to not only search and rescue, but also construction, tracking, and pursuit-evasion. Learning these sub-tasks first is beneficial, as they can be reused repeatedly. This incremental approach to learning is found throughout human societies, and should be equally applicable to the robot domain as well.

Teaching each of these tasks is performed by a human through a graphical interface. Commands are given in task space, and feedback is provided by visualization of sensor data. This training environment forces the human teacher to think in terms of what the robot can understand. During a demonstration, the robot student monitors and records information from its environment. The collected data are then used to create or modify a task.

Each task is represented by a decision network, a augementation of Bayesian networks. These networks have several appealing features including human readability, compact representation of conditional dependencies, and a simple refinement process. Complex tasks can be represented without an exponential growth in states, as seen in POMDPs and MDPs, or in branches, as in decision trees. Networks of this type can also be combined hierarchically, allowing for complex task models.

The teaching and learning processes result in a robot that is capable of performing one or more tasks based on the demonstrated generalized tasks and combination of those tasks. An operator can activate a task, at which time the robot chooses the most appropriate decision network from its repertoire. This choice is based upon the given task goal, and observable features. The chosen network is executed to completion, or error. During autonomous operation a human observer can offer feedback in the form of positive or negative rewards, indicating good and bad actions. These rewards help refine the decision network, by adjusting the utility functions.

References

[1] C. G. Atkeson and S. Schaal. Robot learning from demonstration. In *Machine Learning: Proceedings of the Fourteenth Int. Conf. (ICML '97)*, pages 12–20. Morgan Kaufmann, 1997.

[2] Darrin C. Bentivegna, Christopher G. Atkenson, and Gordon Cheng. A framework for learning from observation using primitives. *Robotics Society of Japan*, 22(2):28–33, 2004.

[3] Sonia Chernova and Manuela Veloso. Multi-thresholded approach to demonstration selection for interactive robot learning. In *3rd ACM/IEEE Inter. Conf. Human-Robot Interaction (HRI'08)*, Amsterdam, The Netherlands, March 2008.

[4] Daniel H Grollman and Odest Chadwicke Jenkins. Sparse incremental learning for interactive robot control policy estimation. In *IEEE Int. Conf. Robotics and Automation*, Pasadena, CA, USA, May 2008.

[5] S. Schaal and C. G. Atkeson. Robot juggling: An implementation of memory-based learning. *Control Systems Magazine*, 14(1):57–71, 1994.