Learning from a Single Demonstration: Motion Planning with Skill Segmentation

Scott Kuindersma[†‡] **George Konidaris**[‡] **Roderic Grupen**[†] **Andrew Barto**[‡]
Laboratory for Perceptual Robotics[†] Autonomous Learning Laboratory[‡]
Computer Science Department, University of Massachusetts Amherst
{scottk, gdk, grupen, barto}@cs.umass.edu

Abstract

We propose an approach to control learning from demonstration that first segments demonstration trajectories to identify subgoals, then uses model-based control methods to sequentially reach these subgoals to solve the overall task. Using this approach, we show that a mobile robot is able to solve a combined navigation and manipulation task robustly after observing only a single successful trajectory.

1 Introduction

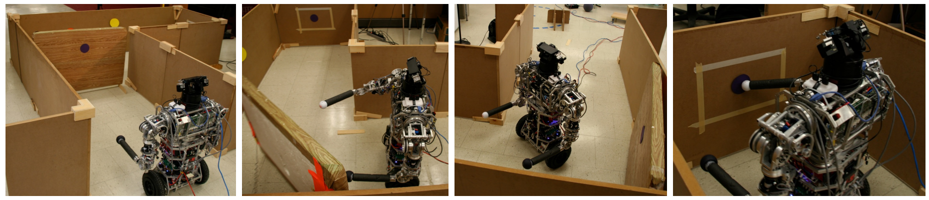
Learning from demonstration (LfD) is an approach to robot control programming that relies on example executions of a task, typically provided by a human teacher [1]. The demonstration data is typically available in the form of *trajectories*, which in our case consists of sequences of states, actions and rewards. In recent work, Konidaris et al. [3] introduced CST, a method for *constructing skill trees* from demonstration. The algorithm addresses the problem of having to represent complex behaviors monolithically by automatically segmenting trajectories into component skills with respective sensorimotor abstractions, and then merging these skills across trajectories to arrive at a tree structure rooted at the goal. Once the tree is constructed, the individual skill policies are fit using linear regression and are incrementally improved using reinforcement learning.

Our approach is to use CST to identify subgoals for *motion planning*, rather than learning policies directly. Motion planning is a broad term that encompasses a variety of methods for model-based robot control (e.g. [2, 4, 5]). We employ a simple closed-loop potential-based method to derive controllers using sensorimotor abstractions and subgoals identified by CST. Although this approach requires a reasonable model, where previous work with CST does not, in many instances we expect the resulting policies to achieve a higher level of robustness with fewer training examples.

2 Skill Acquisition from a Single Demonstration with the uBot

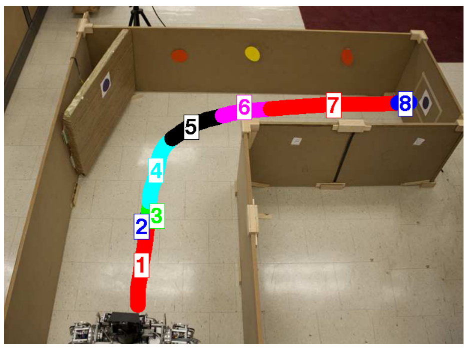
As in Konidaris et al. [3], we tested this approach using a sequential navigation and manipulation task using the uBot-5, a 13-DOF dynamically balancing mobile manipulator (see Figure 1). To simplify perception, the uBot used colored purple, orange and yellow circles placed throughout its environment as perceptually salient objects. The distances (obtained using onboard stereo vision) between the uBot to each landmark are computed at 8Hz and filtered. We assume a reward of -1 at each time step.

A single demonstration from an expert teleoperator was collected and used in the segmentation. The uBot is assumed to engage one of two motor abstractions at a time: either performing endpoint position control, or controlling the speed and angle of its forward motion using differential drive. Thus, six sensorimotor abstractions were made available during the segmentation, each containing either the differences between the right endpoint and object positions, or the distance to and angle between the robot's body and the object.



(a) (b) (c) (d)

Figure 1: Starting at the beginning of a corridor (a), the uBot must approach and push open a door (b), turn through the doorway (c), then approach and push a panel (d).



#	Description	Abstraction
1	Approach door	{purple, body}
2	Open door	{purple, endpoint}
3	Wait for door	{purple, body}
4	Drive past door	{orange, body}
5	Turn corner	{yellow, body}
6	Align with back panel	{orange, body}
7	Approach back panel	{purple, body}
8	Touch back panel	{purple, endpoint}

Figure 2: Identified skill segments and abstractions.

CST identified eight skill segments with appropriate abstractions (Figure 2). The skill policies are derived using the termination configuration of each segment to define a subgoal, $\boldsymbol{x}_{ref}$, to parameterize a quadratic potential function $\phi(\boldsymbol{x}) = (\boldsymbol{x}_{ref} - \boldsymbol{x})^2$. We assume that a model of the abstract sensorimotor dynamics is known, so we can derive the Jacobian vector $J_t = \frac{\partial \phi(\boldsymbol{x}_t)}{\partial \boldsymbol{u}} = \frac{\partial \phi(\boldsymbol{x}_t)}{\partial \boldsymbol{x}} \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{u}}$.[1] To calculate the desired control inputs, $\Delta \boldsymbol{u}_t$, we use the pseudoinverse of J_t to map displacements along the gradient of ϕ into action space, $\Delta \boldsymbol{u}_t = -K J_t^{\#} \phi(\boldsymbol{x}_t)$, where K is a diagonal gain matrix that can be fit using the demonstration data.

The resulting skills were evaluated by executing them in sequence starting at the beginning of the maze. The robot succeeded in solving the task in 10 out of 10 tries and exhibited greater invariance to initial conditions while requiring less data compared to the regression-based method [3]. Finally, we note that by segmenting into subtasks, we were able to apply a simple control algorithm that would be difficult or impossible to apply monolithically.

Acknowledgments

Scott Kuindersma is supported by a NASA GSRP fellowship. Andrew Barto and George Konidaris were supported by the Air Force Office of Scientific Research under grant FA9550-08-1-0418.

References

[1] Brenna D. Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469–483, May 2009.

[2] Oussama Khatib. Real-time obstacle avoidance for manipulators and mobile robots. *International Journal of Robotics Research*, 5(1):90–98, 1986.

[3] George Konidaris, Scott Kuindersma, Andrew Barto, and Roderic Grupen. Constructing skill trees for reinforcement learning agents from demonstration trajectories. In *Advances in Neural Information Processing Systems (NIPS)*, 2010.

[4] Steven M. LaValle and James J. Kuffner Jr. Rapidly-exploring random trees: Progress and prospects. In *Algorithmic and Computational Robotics: New Directions*, pages 293–308, 2000.

[5] Russ Tedrake, Ian R. Manchester, Mark M. Tobenkin, and John W. Roberts. LQR-Trees: Feedback motion planning via sums of squares verification. *International Journal of Robotics Research*, 29:1038–1052, July 2010.

[1] Note that the state and action spaces for each skill are determined by the segmentation.