Complex Task Learning from Unstructured Demonstrations

Scott Niekum
Department of Computer Science
University of Massachusetts Amherst
140 Governors Drive
Amherst, MA 01003

A simple system that allows end-users to intuitively program robots is a key step in getting robots out of the laboratory and into the real world. Although in many cases it is possible for an expert to successfully program a robot to perform complex tasks, such programming requires a great deal of knowledge, is time-consuming, and is often task-specific. In response to this, much recent work has focused on robot learning from demonstration (LfD) (Argall et al. 2009; Billard et al. 2008), where non-expert users can teach a robot how to perform a task by example. Such demonstrations eliminate the need for knowledge of the robotic system, and in many cases, require only a fraction of the time that it would take an expert to design a controller by hand.

Ideally, an LfD system can learn to perform and generalize complex tasks given a minimal number of demonstrations without requiring knowledge about the robot. Much LfD research has focused on the case in which the robot learns a monolithic policy from a demonstration of a simple task with a well-defined beginning and end. This approach often fails for complex tasks that are difficult to model with a single policy. Thus, structured demonstrations are often provided for a sequence of subtasks, or *skills*, that are easier to learn and generalize than the task as a whole, and which may be widely reusable across many tasks.

However, a number of problems are associated with segmenting tasks by hand and providing individual skill demonstrations. The most obvious of these is simply an issue of convenience, since the most natural way to demonstrate a task is by performing it continuously from start to finish. Dividing a task into component skills is not only time-consuming, but often difficult—an effective segmentation can require knowledge of the robot's kinematic properties, internal representations, and existing skill competencies. Since skills may be repeated within and across tasks, defining skills also requires qualitative judgements to be made about when two segments are similar enough to be considered the "same", or in deciding the appropriate level of granularity at which to perform segmentation. Clearly, users cannot be expected to manually manage this collection of skills as it grows over time.

For this reason, recent work has aimed at automating the segmentation process (Jenkins and Matarić 2004; Grollman and Jenkins 2010; Butterfield et al. 2010; Konidaris et al. 2010). Collectively, this body of work has addressed several key issues that are critical to any system that aims to learn increasingly complex tasks from unstructured demonstrations. In this document, we use *unstructured* to refer to demonstrations that are unsegmented, possibly incomplete, and may come from multiple tasks or skills. First, the robot must be able to recognize repeated instances of skills and generalize them to new settings. Segmentation should also be able to be performed without the need for *a priori* knowledge about the number or structure of skills involved in a task. Additionally, the robot should be able to identify a broad, general class of skills, including object manipulation skills, gestures, and goal-based actions. Finally, the representation of skill policies should be such that they can be corrected by the user and improved through practice.

Although many of these issues have been addressed individually in these previous research efforts, no system that we are aware of has jointly addressed them all in a principled manner in an unstructured setting. Our contribution is a framework that addresses all of these issues by integrating a principled Bayesian nonparametric approach to segmentation with state-of-the-art LfD and RL techniques as a first step towards a natural, scalable system that will be able to learn tasks of increasing complexity. Specifically, we propose to design a system that:

1. segments unstructured demonstrations into appropriate numbers of component skills, recognizes repeated skills across demonstrations and tasks, and generalizes these skills to new situations.
2. allows the user to provide unstructured, interactive corrections and feedback to the robot, without requiring any knowledge of the robot's underlying representation of the task or its component skills.
3. infers the user's intentions for each segmented skill and autonomously improves these skills using reinforcement learning.

To address trajectory segmentation and skill reuse, we propose the use of recent developments in Bayesian nonparametrics. Hidden Markov Models (HMMs) have a long history of use for interpreting time series data, but have been limited by several constraints, including the need to specify the number of modes *a priori*. We examine the Beta-Process

Autoregressive HMM (BP-AR-HMM) (Fox et al. 2009), a nonparametric extension of the HMM that can automatically infer an appropriate number of hidden states from data, while also allowing the sharing of modes across trajectories and the representation of rich time-series dependencies between observations. The BP-AR-HMM has been shown to successfully segment human motion capture data into activities which can be shared across trajectories, a process similar to our notion of skill parsing in unstructured demonstrations.

To address the problems of LfD and policy improvement, we examine recent work at the intersection of control theory and reinforcement learning (RL) (Sutton and Barto 1998). LfD algorithms often use demonstration trajectories to construct fixed control policies that offer no mechanism for improvement through practice. By contrast, reinforcement learning has had much success learning and improving control policies through interaction with the environment. However, RL has only had limited success in general robotics applications, largely due to the costs of gathering data on physical robots and the difficulty of exploration in high-dimensional spaces. Dynamic Movement Primitives (DMPs) (Ijspeert, Nakanishi, and Schaal 2003) address these issues by providing a unified framework in which stable dynamic controllers can be created via LfD and improved through RL. DMPs have successfully been used to learn complex structured control tasks such as humanoid walking, but can also be used as building blocks for representing larger, multi-step tasks.

The combination and extension of these techniques will allow a robot to segment and identify repeated skills in human demonstrations, create baseline skill policies from demonstration segments, receive feedback from the user, improve skills through practice, and expand the skill library as needed. Together, these capabilities will be a major step toward open-ended robotic task demonstration by end-users.

Part of this work will be integrative in nature, but there are several major technical questions that must also be addressed, primarily stemming from the nature of working with unstructured demonstrations. First, how can we robustly infer preconditions, postconditions, and reward functions for skills when the underlying segmentations may be unreliable? Second, how might human feedback and corrections be able to help mitigate this difficulty? Third, how can tasks and skills be improved through practice when the preconditions, postconditions, and reward functions may also be unreliable?

Finally, we propose to validate our approach through experiments on the PR2 robot, a mobile manipulator developed by Willow Garage. The PR2 is built on the widely used open source Robot Operating System (ROS); all code developed in this project will be open source and made available to the growing community of users. In addition to working with the freely available PR2 simulator, we will have access to a physical PR2 robot through a partnership with the Robert Bosch Research and Technology Center, allowing us to test our approach in a variety of real world scenarios.

As of now, we have developed and tested the software infrastructure for demonstration collection on the PR2, trajectory parsing using BR-AR-HMMs, and DMP training, execution, and generalization. We have been able to leverage code from the ROS community, including code for tabletop detection, object recognition, robust grasp planning, and cartesian arm control, as well as a BP-AR-HMM implementation made available by Emily Fox[1].

Using this software base, and extending it considerably, we have preliminary results (currently in submission) that display an intelligent parsing and replay of demonstrations from a block stacking task performed in the simulator. By recognizing repeated skills within and across demonstrations, we are able to automatically detect a relevant coordinate frame for each skill, facilitating generalization to new situations. We show that the robot is able to successfully complete the task in novel configurations, despite the fact that the task requires multiple steps in various coordinate frames and that only unstructured full-task demonstrations were provided. We also demonstrate the ability of the robot to recognize previously learned skills, allowing it to draw on a library of skills to speed up learning. We are currently moving this experimental framework over to the physical PR2 at Bosch for a complex dishwasher loading task.

References

Argall, B.; Chernova, S.; Veloso, M.; and Browning, B. 2009. A survey of robot learning from demonstration. *Robotics and Autonomous Systems* 57(5):469–483.

Billard, A.; Calinon, S.; Dillmann, R.; and Schaal, S. 2008. *Handbook of Robotics*. Secaucus, NJ, USA: Springer. chapter Robot programming by demonstration.

Butterfield, J.; Osentoski, S.; Jay, G.; and Jenkins, O. 2010. Learning from demonstration using a multi-valued function regressor for time-series data. In *Proceedings of the Tenth IEEE-RAS International Conference on Humanoid Robots*.

Fox, E.; Sudderth, E.; Jordan, M.; and Willsky, A. 2009. Sharing features among dynamical systems with beta processes. *Advances in Neural Information Processing Systems 22* 549–557.

Grollman, D., and Jenkins, O. 2010. Incremental learning of subtasks from unsegmented demonstration. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 261–266. IEEE.

Ijspeert, A.; Nakanishi, J.; and Schaal, S. 2003. Learning attractor landscapes for learning motor primitives. *Advances in Neural Information Processing Systems 16* 1547–1554.

Jenkins, O. C., and Matarić, M. J. 2004. A spatio-temporal extension to Isomap nonlinear dimension reduction. In *Proceedings of the Twenty-First International Conference on Machine Learning*, 441–448.

Konidaris, G. D.; Kuindersma, S. R.; Barto, A. G.; and Grupen, R. A. 2010. Constructing skill trees for reinforcement learning agents from demonstration trajectories. In *Advances in Neural Information Processing Systems 23*.

Sutton, R. S., and Barto, A. G. 1998. *Reinforcement Learning: An Introduction*. MIT Press.

[1] http://stat.wharton.upenn.edu/~ebfox/software.html