AAAI Technical Report FS-12-07
Robots Learning Interactively from Human Teachers

Training Wheels for the Robot: Learning from Demonstration Using Simulation

Nathan Koenig
Open Source Robotics Foundation
Mountain View, CA

Maja Matarić
University of Southern California
Los Angeles, CA

Introduction

Learning from demonstration (LfD) is a promising technique for instructing/teaching autonomous systems based on demonstrations from people who may have little to no experience with robots (Billard et al. 2008). An important aspect to LfD is the communication method used to transfer knowledge from an instructor to a robot. The communication method affects the complexity of the demonstration process for instructors, the range of tasks a robot can learn, and the learning algorithm itself.

We have designed a graphical interface and an instructional language to provide an intuitive teaching system. The drawback to simplifying the teaching interface is that the resulting demonstration data are less structured, adding complexity to the learning process. This additional complexity is handled through the combination of a minimal set of predefined behaviors and a task representation capable of learning probabilistic policies over a set of behaviors. The predefined behaviors consist of finite actions a robot can perform, which act as building blocks for more complex tasks. Example behaviors include *move to*, *pick up*, and *put down*. Behaviors operate in conjunction with a feature, which is an object the robot can observe and manipulate.

A series of behaviors and features from a group of instructors is used to generate a generalized policy for a task. The policy is represented by a set of decision networks, an extension of Bayesian networks, that incorporate the ability to probabilistically choose actions based on state information. We allow for error in the teaching and learning process by providing a mechanism to refine decision networks during autonomous operation. This technique effectively reduces the number of complete demonstrations required to accurately learn a task.

Teaching Interface and Task Learning

Teaching interfaces use a variety of forms, including manual manipulation (Hersch et al. 2008), teleoperation via joysticks (Grollman and Jenkins 2007; Chernova and Veloso 2007), graphical interfaces (Chernova and Veloso 2008), external observations (Schaal, Ijspeert, and Billard 2004; Pollard and Hodgins 2002), and sensors placed on the instructor (Ijspeert, Nakanishi, and Schaal 2002). A particular method may be appropriate for a type of robot or learning task. These methods also trade off teaching complexity and complexity of the learning algorithm.

This work emphasizes ease of use for instructors and generality of use with multiple robot platforms. The graphical interface meets these needs by providing a configurable interface that is familiar to people who have some computer experience. The instructor uses the graphical interface to observe the state of the world and send instructions to the robot, as shown in Figure1.

The instructor commands the robot by building sentences using an instructional language that is composed of a behavior, a feature, and an optional modifier-feature combination that specifies how the behavior is performed relative to the feature. An example instruction is *put down the salt on the table*, where *put down* is the behavior, *salt* is the feature, and *on the table* is the optional modifier-feature combination.

The graphical interface guides the instructor through the process of constructing a sentence by sequentially asking the instructor for the next part of the instruction. This process makes instruction generation clear to the user, and an undo option allows the user to easily fix mistakes. Once the instruction is complete, it is sent to the robot for execution. During execution, the user observes the robot's progress and any textual feedback. The process of building instructions continues until the instructor decides the demonstration is complete.

During a demonstration, the robot records the instructions and state information. After one or more instructors have provided demonstrations, the recorded information is used to learn a task policy represented by a set of decision networks. The decision network learning is detailed in (Koenig, Takayama, and Matarić 2010).

Experimental Setup

The graphical interface and learning system was used to teach a robot the process of cooking mushroom risotto. The demonstration environment consisted of a simulated kitchen, PR2 robot, ingredients, and utensils. Simulation was used to reduce teaching time, provide a stable environment, and facilitate error correction through an *undo* feature that reverses instructions.

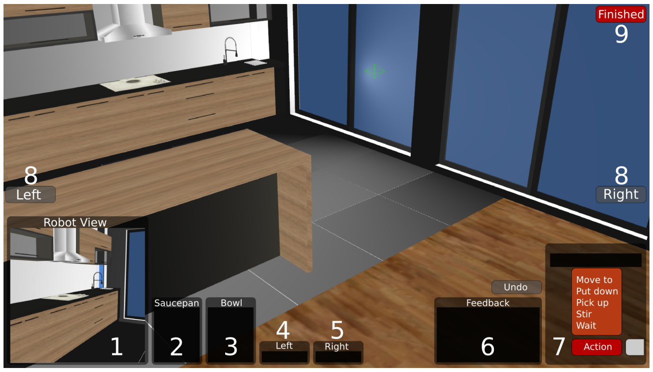

Figure 1: Simulated kitchen environment with teaching interface. 1. Robot camera view, 2. Contents of saucepan, 3. Contents of bowl, 4. Left gripper, 5. Right gripper, 6. Robot feedback, 7. Instruction interface, 8. Change view left/right, 9. Demonstration complete.

A fifteen-step tutorial guided participants through each component of the graphical interface. Upon completion of the tutorial, participants received a printed mushroom risotto recipe. At this point, each the participant was free to instruct the robot. Following the demonstration, each participant completed a survey that collected demographic information.

Results

Thirty three participants completed the demonstration, 26 male and 7 female, with an age range of 23 to 67. The average time to complete the demonstration was 27 minutes, with a standard deviation of 10 minutes. In contrast, a PR2 robot took roughly three to four times as long to complete the task due to the pace of perception and slowness of the movements for safety.

As expected, the instructions sent to the robot were similar across participants in the beginning of the demonstration. The instructions diverged over time as participants chose to complete the recipe using a different order of instructions. This divergence made it more difficult to learn correct decision networks. As a result, a few networks produced incorrect behavior when used on an autonomous robot.

The incorrect networks can be discovered by the system itself or by a human observer. An incorrect decision network can be self-discovered when multiple behaviors have the same probability. In these cases, the system is incapable of choosing the best behavior and must ask for help. In all other instances, the system chooses what it believes to be the best behavior, which may not in fact be what a human observer would select.

In both cases, a human may intervene and provide the system with the correct behavior. This new information is incorporated directly into the decision network through an update process.

Discussion and Future Work

We have developed a system that provides an intuitive interface for instructing a robot in time-extended tasks. The interface requires no prior knowledge of robotics. Teaching in simulation provides a less time consuming teaching experience with the ability to gracefully fix errors. The demonstration data are used to generate decision networks that represent the task. These networks can be refined when the system operates autonomously.

Due to the decision network representation, we are constrained to high-level tasks that are categorized as a time-extended series of behaviors. The process of solving decision networks does not operate on the time scale necessary for joint control. We are also limited by the range of features and behaviors available to the robot. The diversity of features is ever increasing, but is still limited by current work in object perception and recognition.

Our future work will incorporate knowledge transfer across tasks to allow a robot to utilize decision networks from one task to reduce teaching time for a new task. We will also study the use of graphical interfaces as a teaching tool, and determine how best to provide two-way communication between an instructor and a robot.

References

Billard, A.; Calinon, S.; Dillmann, R.; and Schaal, S. 2008. *Robot Programming by Demonstration*. MIT Press. chapter 59.

Chernova, S., and Veloso, M. 2007. Confidence-based policy learning from demonstration using gaussian mixture models. In *International Conference on Autonomous Agents and Multiagent Systems*.

Chernova, S., and Veloso, M. 2008. Teaching collaborative multi-robot tasks through demonstration. In *IEEE-RAS International Conference on Humanoid Robots*.

Grollman, D. H., and Jenkins, O. C. 2007. Learning robot soccer skills from demonstration. In *International Conference on Development and Learning*, 276–281.

Hersch, M.; Guenter, F.; Calinon, S.; and Billard, A. 2008. Dynamical system modulation for robot learning via kinesthetic demonstrations. *IEEE Transactions on Robotics* 24(6):1463–1467.

Ijspeert, J. A.; Nakanishi, J.; and Schaal, S. 2002. movement imitation with nonlinear dynamical systems in humanoid robots. In *international conference on robotics and automation (icra2002)*.

Koenig, N.; Takayama, L.; and Matarić, M. 2010. Learning from demonstration: A study of visual and audiory communication and influence diagrams. In *International Symposium on Experimental Robotics (ISER'10)*.

Pollard, N., and Hodgins, J. K. 2002. Generalizing demonstrated manipulation tasks. In *Workshop on the Algorithmic Foundations of Robotics*.

Schaal, S.; Ijspeert, A.; and Billard, A. 2004. *Computational approaches to motor learning by imitation*. Number 1431. Oxford University Press. 199–218.