A Practical Comparison of Three Robot Learning from Demonstration Algorithms

Russell Toris
Worcester Polytechnic Institute
Worcester, MA
USA
rctoris@wpi.edu

Halit Bener Suay
Worcester Polytechnic Institute
Worcester, MA
USA
benersuay@wpi.edu

Sonia Chernova
Worcester Polytechnic Institute
Worcester, MA
USA
soniac@wpi.edu

ABSTRACT

Research on robot learning from demonstration has seen significant growth in recent years, but existing evaluations have focused exclusively on algorithmic performance and not on usability factors, especially with respect to naïve users. Here we present findings from a comparative user study in which we asked non-experts to evaluate three distinctively different robot learning from demonstration algorithms – Behavior Networks, Interactive Reinforcement Learning, and Confidence Based Autonomy. Participants in the study showed a preference for interfaces where they controlled the robot directly (teleoperation and guidance) instead of providing retroactive feedback for past actions (reward and correction). Our results show that the best policy performance in most metrics was achieved using the Confidence Based Autonomy algorithm.

Categories and Subject Descriptors

I.2 [**Artificial Intelligence**]: Robotics

General Terms

Human Factors

Keywords

Learning from Demonstration, Human-Robot Interaction

1. INTRODUCTION

Algorithms for robot *Learning from Demonstration* (LfD) seek to enable human users to expand the capabilities of robots through interactive teaching instead of explicit programming. Specifically, LfD methods enable new robot behaviors to be learned based on demonstrations of the desired actions performed by the teacher. The broad research goal is to develop an *intuitive* method of programming that is accessible to untrained, naïve, users [1]. Research in this area lacks a comparative study, especially among higher-level, symbolic, learning algorithms that focus on learning behavioral or task representations. Performance evaluations are typically focused on learning time compared against non-interactive methods.

This lack of a comparison leaves many critical questions unanswered. Without a comparison of existing methods, future researchers have no guidance with respect to what learning algorithms and interaction techniques are most effective, and what underlying algorithmic assumptions accurately reflect the real-world use case. These problems are further heightened by the fact that many LfD techniques have only been evaluated through use by expert roboticists.

To further investigate these issues, we present a comparison study of three established learning from demonstration algorithms, each representing a distinctively different variant of policy learning and utilize different demonstration techniques:

- **Interactive Reinforcement Learning (Int-RL):** A system model policy learning technique in which the reward signal is provided by the human user at runtime. The teacher interacts with the robot by providing reward and guidance input through an on-screen interface. [4]
- **Behavior Networks (BNets):** A planning-based policy learning technique. The teacher interacts with the robot through kinesthetic teaching, a type of *experienced demonstration*, in which the user physically guides the robot through the task. [3]
- **Confidence Based Autonomy (CBA):** A mapping approximation technique based on a classification algorithm. The teacher interacts with the robot by selecting the next action or correcting a past action choice through an on-screen interface. [2]

To compare the three algorithms, 31 participants were recruited to use each of the three methods to teach a robotic agent to accomplish a simple task. To accomplish this, we designed the *DBug domain* for the Aldebaran Nao humanoid robot, which required the robot to catch and remove small moving HEXBUG toys from a table.

2. THE DBUG EXPERIMENT

Within the DBug domain, the robot was given 3 preprogrammed actions: picking up a bug at a center pickup location (marked on the table with an 'X'), sweeping a bug inwards towards the pickup location, and waiting for 1 second (i.e. doing nothing). The wait action was an important action to include as it was useful in teaching the robot when it was not appropriate to either sweep or pickup. Using images from Nao's on-board camera and image processing with the OpenCV library, the robot was able to continuously track the x and y position of all bugs in its field of view (a 320x240 pixel image). The state space was defined

HRI'12, March 5–8, 2012, Boston, Massachusetts, USA.
ACM 978-1-4503-1063-5/12/03.

Table 1: **Number of responses to "How well did the robot learn the task?"**

	Not at All	Not Well	Well	Very Well
Int-RL	4	9	17	1
BNets	4	17	9	1
CBA	1	4	17	9

Table 2: Number of responses to "How quickly did the robot learn the task?"

	Very Slowly	Slowly	Quickly
Int-RL	11	16	4
BNets	8	11	12
CBA	3	9	19

to be the x and y position of the bug with the shortest Euclidean distance to the pickup zone. This state space was considered to be large enough such that the solution was not trivial, yet small enough that a user could make reasonable progress teaching the agent within a short period of time.

Upon arrival, the participant was given a broad overview of the task that was being trained as well as the capabilities of the Nao. Following the introduction, the participant used each algorithm for 10 minute intervals in a randomized order to train the robot. After each training session the participant was asked to fill out a brief questionnaire relating to the method that they had just used. At the conclusion of their session, a final general questionnaire was given which asked comparative questions about the three methods.

3. SUMMARIZED RESULTS

To evaluate the final learned policies for the three algorithms for each user, we reconstruct the final policy of each algorithm at the end of each training session using data collected during the experiments. These models were then queried for an action for every possible state. On average, our subjects taught the pick-up action and the sweep action using CBA with a higher accuracy as compared to the other methods followed by Int-RL. Subjects taught the wait action with higher accuracy using BNets. It is worth noting that the wait action was implicitly taught in BNets, that is, if the teacher doesn't demonstrate any action by moving the arms of the robot, then the agent learned to wait, for instance when there is no bug to pick up in sight. Based on the lower accuracy rate of the other two actions, we think that this implicitness is the reason why people had more success with BNets for this specific action.

In addition to assessing the user's performance with each algorithm, we are interested in looking at trends in the survey data. After each training session, the user filled out a algorithm-specific survey. In particular, two general questions appeared on each of the surveys: *How well did the robot learn the task?* and *How quickly did the robot learn the task?*. The total number of responses for each possible choice in each question are listed in Table 1 and Table 2.

When asked *Which method did you find easiest to learn?* 24 people answered CBA, 7 people answered BNets, and nobody said Int-RL. We understand from this that most people actually wanted to tell the robot what to do at each moment using a graphical interface and a mouse. When asked *If you were to teach a new task to a robot, which method would you prefer to use?*, 25 people answered CBA, 5 people answered BNets, and 1 person chose Int-RL. When we asked for the order of general preference of the methods, CBA was ranked first with 27 votes, Int-RL became second with 18 votes, and BNets ended-up being the last with 16 votes.

The general survey included space where participants were able to write free-form feedback about the algorithms in the study. Three common responses from the surveys include the need for additional feedback from the robot regarding what it was sensing, the need for better understanding of what the robot is "thinking", and the need of communicating with the robot more naturally (e.g. through speech). These three points, especially the importance of feedback, are surely not unprecedented findings; however, their importance is often underestimated in existing LfD methods. Our aim is to draw attention to these factors for future research.

4. CONCLUSION

Our results show that users were most proficient in and satisfied with teaching using the Confidence-Based Autonomy algorithm. Based on our participants' comments, this is mainly because naïve teachers like to see which actions are available to them and give commands directly to the robot. Although a lot of the participants liked the way they interacted with the robot for BNets (kinesthetic teaching), they were not happy with the performance of the result. In case of Int-RL, participants indicated that they could not see a satisfactory improvement in the action decision of the agent. This was mostly caused by the inherently slow learning rate of the algorithm; however, to the participants, this factor was not obvious. Our results tell us that it is important to keep the transparency of the robot's knowledge to its teacher during the whole process. Techniques for communicating the knowledge state of the robot in a effective way remains an important area for future work.

5. REFERENCES

[1] B. D. Argall, S. Chernova, M. Veloso, and B. Browning. A survey of robot learning from demonstration. *Robot. Auton. Syst.*, 57:469–483, May 2009.

[2] S. Chernova and M. Veloso. Interactive policy learning through confidence-based autonomy. *J. Artificial Intelligence Research*, pages 1–25, 2009.

[3] M. N. Nicolescu and M. J. Mataric. Natural methods for robot task learning: instructive demonstrations, generalization and practice. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, AAMAS '03, pages 241–248, New York, NY, USA, 2003. ACM.

[4] A. L. Thomaz and C. Breazeal. Adding guidance to interactive reinforcement learning. In *Proceedings of the Twentieth Conference on Artificial Intelligence (AAAI)*, 2006.