In Twentieth Edition of the AAAI Robotics Workshop and Exhibition, San Francisco, CA, USA, August 2011.

Learning from Demonstration in Spatial Exploration

J. Pablo Munoz[1], **Arif T. Ozgelen**[2] and **Elizabeth I. Sklar**[1,2]
[1]Brooklyn College
[2]The Graduate Center
The City University of New York, New York, USA
jpablomch@gmail.com,tunaozgelen@gmail.com,sklar@sci.brooklyn.cuny.edu

Abstract

We present the initial stage of our research on *Learning from Demonstration* algorithms. We have implemented an algorithm based on Confident Execution, one of the components of the Confidence-Based Autonomy algorithm developed by Chernova and Veloso. Our preliminary experiments were conducted first in simulation and then using a Sony AIBO ERS-7 robot. So far, our robot has been able to learn crude navigation strategies, despite limited trials. We are currently working on improving our implementation by including additional features that describe more broadly the state of the agent. Our long term goal is to incorporate Learning from Demonstration techniques in our *HRTeam* (human/multi-robot) framework.

Introduction

Learning from Demonstration techniques are used in robotics to teach an embodied agent (i.e., a robot) which *action* is associated with a particular *state*, resulting in the derivation of a *policy*. In some implementations, the agent generalizes from knowledge it has gained previously, ultimately performing actions when encountering states that are similar to one of the states in its policy.

We have chosen to explore an algorithm based on the *Confidence-Based Autonomy (CBA)* technique designed by Sonia Chernova and Manuela Veloso (2007; 2009). Their method comprises two components: *Confident Execution* and *Corrective Demonstration*. Our implementation has been mostly focused on the first component. Confident Execution compares the current world state of the agent with the set of known states encoded in the available policy. From this comparison, a confidence value is obtained. If the confidence value is below the confidence threshold previously defined, a demonstration is requested in which a human tutor effects the correct action and the robot learner observes. Otherwise, the action associated with the most similar state in the policy is executed. By requesting a demonstration, the agent obtains the action that is associated with the current ambiguous or unknown state, and this new state-action pair is then added to the policy.

Confident Execution implementation

We implemented our version of Learning from Demonstration using the Sony AIBO ERS-7 robot and the Tekkotsu framework (Touretzky and Tira-Thompson 2005). Chernova and Veloso also used the AIBO robot to test their Learning from Demonstration algorithms (Chernova and Veloso 2007). The main differences between the implementations are as follows. First, the Chernova and Veloso implementation learns a *decision boundary* over time. This value is the threshold for the confidence used to determine if the robot should act or if it should request a demonstration. Our implementation used a constant value[1]. Second, the Chernova and Veloso implementation triggers the confidence-based decision making by selecting the nearest neighbor in state space. In contrast, our implementation selects the nearest obstacle. These implementation choices were made to help us develop our initial solution within a simplified, fixed environment. Future work will compare the two methods.

Our implementation of Confidence-Based Autonomy makes decisions using data collected from the AIBO's infrared (IR) sensors. The AIBO has 2 IR sensors located in its head. These point forward with ranges from 50 mm to 500 mm and from 200 mm to 1500 mm, respectively. Using the readings from these two sensors, along with information regarding the position of the AIBO's head, our controller creates a world state for the AIBO which models the distance to the nearest objects around the robot.

We set the distance threshold to 1500 mm, and only for the front sensor. This means that if the front sensor returns a reading within the threshold, the robot checks its policy. Otherwise, the AIBO keeps moving forward. The head of the AIBO moves continuously from one side to the other, stopping for 1 second in each direction ($-90°$, $0°$, and $90°$). If the distance returned by the front sensor is less than the distance threshold, then the robot checks if there is already a state in its policy that satisfies the conditions, and returns its associated action, i.e., a state in the policy corresponds to a confidence higher than the confidence threshold set. Otherwise, the AIBO comes to a complete stop, and a more thorough reading of the sensors is performed, checking that the reading effectively corresponds to the side that the robot

[1]A value of 0.93 was used for the preliminary experiments discussed here.

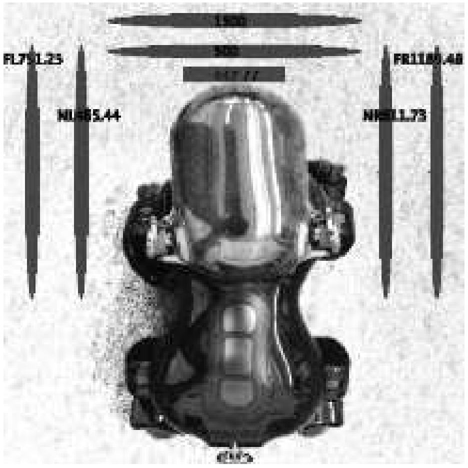

Figure 1: Graphical User Interface used to show the current world state of the AIBO

needs to check. This is done by confirming that the joints in the head are in the correct position for gathering the necessary data. After receiving a demonstration from the user, the new state-action pair is added to the robot's policy.

We positioned the AIBO in a maze built in our lab, and the goal was to teach the AIBO how to explore all the parts of the maze without crashing into walls. For these experiments we created a Graphical User Interface (GUI) (Figure 1) to help the user understand the actual state of the AIBO and be able to provide a meaningful demonstration from which the robot could learn. The lines on the sides and front indicate the readings of the IR sensors, which turn red if the value is below the robot policy's distance threshold. When the AIBO encounters an ambiguous or unknown state, the GUI displays a dialog to ask the user for a demonstration. This dialog box includes the three possible actions that the agent can take: go forward, turn left, and turn right. The user—human tutor—demonstrates the action she wishes the robot to take. The AIBO observes the action, then it performs it, and then resumes gathering information to build its current world state, which again will be checked against its policy for subsequent actions.

Initial results and challenges

We ran two sets of preliminary experiments. During our initial experiments, we focused on teaching the AIBO to avoid objects closer than 400 mm. With the learned policy, our AIBO explored the maze in the lab successfully: it entered rooms, walked through the halls and never crashed into any walls. During the second set of experiments, we increased the distance to objects to avoid to values greater than 400 mm. Here, we encountered problems, especially due to the AIBO incorrectly determining the distance from its head to the floor and confusing the floor with the wall. This resulted in the robot taking the wrong action for particular situations and crashing into the maze wall. We tried to alleviate this problem by raising the robot's head, which worked for determining distances greater than 400 mm, but unfortunately reduced the robot's accuracy in reacting to objects closer than 400 mm.

We faced additional complications during the experiments. For instance, the AIBO's sensors had a few inaccurate readings, primarily due to the vibration of the head (where the sensors are located) when the robot is walking. Other causes of faulty sensor readings included low power level of the robot's batteries and communication errors between the robot and the GUI.

Next steps

Currently, we are focusing on improving the quality of the robot's world state, customized to its set of likely actions within the task environment. For example, we are decreasing the angle of rotation of the robot's head (to range from $45°$ to $-45°$) since the robot will not move in right angle turns. Also, we are experimenting with averaging over a sliding window of sensor readings instead of using a single reading, to mitigate communication problems.

Future work involves learning more complex behaviors and representing a more robust world state. Our long term goal is to incorporate Learning from Demonstration techniques in our *HRTeam* (human/multi-robot) framework (Sklar et al. 2010; 2011). In that work, we are applying Learning from Demonstration to robot localization by using learned confidence values as input to a particle filter through with the robot identifies its position in the maze.

Acknowlegements

This work was supported by NSF #CNS-0851901 and CUNY CIRG #1642.

References

Chernova, S., and Veloso, M. 2007. Confidence-based policy learning from demonstration using gaussian mixture models. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, AAMAS '07, 233:1–233:8. New York, NY, USA: ACM.

Chernova, S., and Veloso, M. 2009. Interactive policy learning through confidence-based autonomy. *Journal of Artificial Intelligence Research* 34(1):1–25.

Sklar, E.; Parsons, S.; Epstein, S.; Ozgelen, A. T.; Rabanca, G.; Anzaroot, S.; Gonzalez, J.; Lopez, J.; Lustig, M.; Ma, L.; Manashirov, M.; Munoz, J. P.; Salazar, S. B.; and Schwartz, M. 2010. Developing a Framework for Team-based Robotics Research. In *Nineteenth Edition of the AAAI Robotics Workshop and Exhibition*.

Sklar, E.; Ozgelen, A. T.; Munoz, J. P.; Gonzalez, J.; Manashirov, M.; Epstein, S. L.; and Parsons, S. 2011. Designing the HRTeam Framework: Lessons Learned from a Rough-and-Ready Human/Multi-Robot Team. In *Proceedings of the Workshop on Autonomous Robots and Multirobot Systems (ARMS) at AAMAS*.

Touretzky, D. S., and Tira-Thompson, E. J. 2005. Tekkotsu: a framework for aibo cognitive robotics. In *Proceedings of the 20th national conference on Artificial intelligence - Volume 4*, 1741–1742. AAAI Press.