Robot Learning in a Simple Environment

Nicholas Westbury and Joseph Modayil

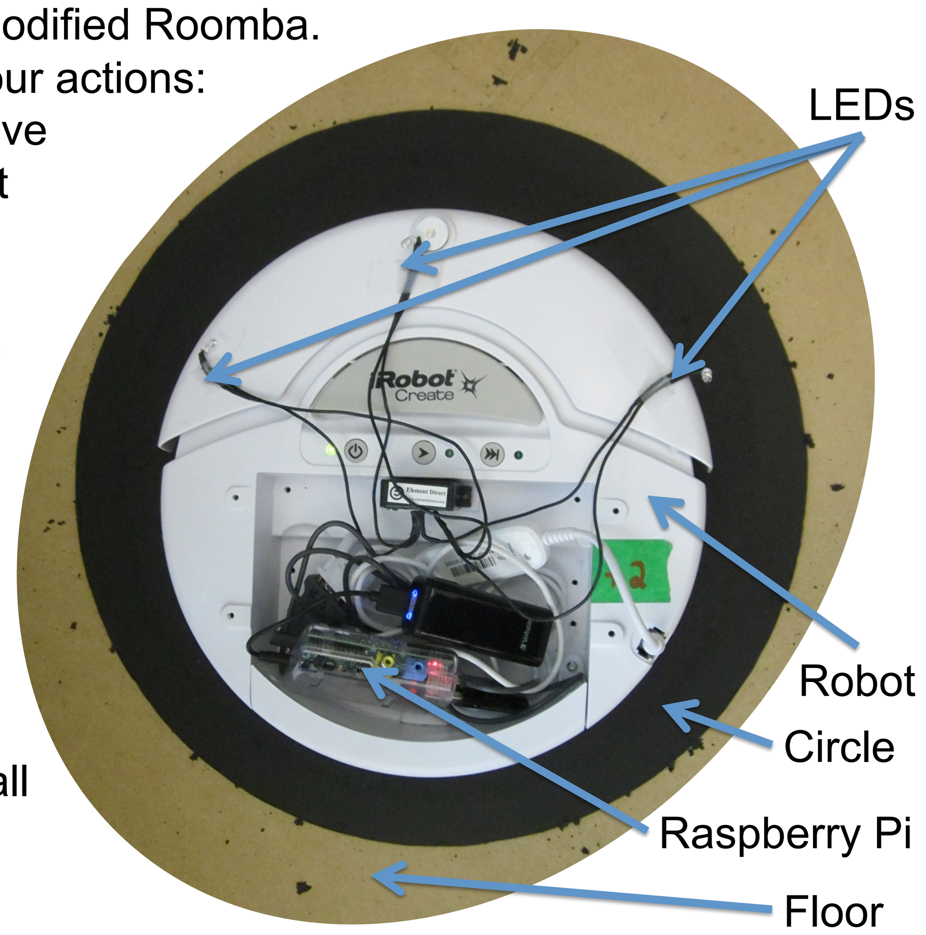
University of Alberta RLAI Lab

Introduction:

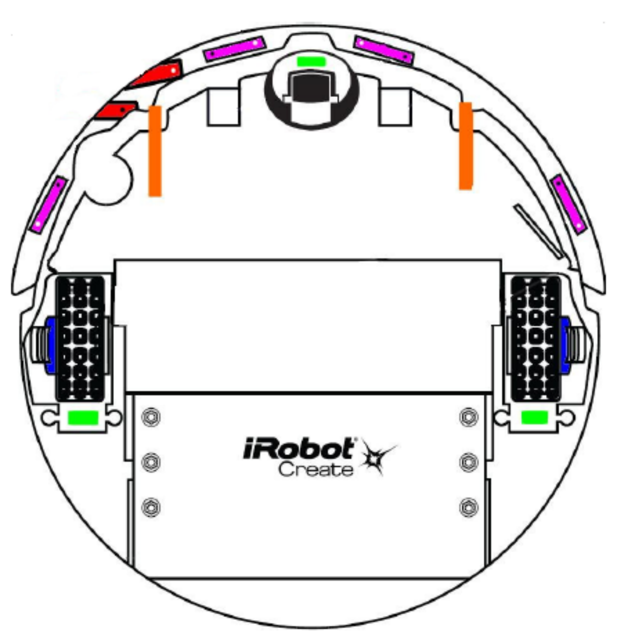
The goal of this project was to study how a robot can learn in a simple environment using two reinforcement learning algorithms. First, the robot learned to predict how far it could travel forward until it detected the edge of its environment. Second, the robot learned how to enter a particular sensory configuration from which it could travel forward the furthest. The robot learns incrementally without knowledge of its shape or a model of its environment.

Experimental Settings

- The environment is a dark circle placed on a light-colored floor.
- The robot is a modified Roomba.
- The robot has four actions: move forwards, move backwards, turn left and turn right.
- Every 1/10 of a second, it selects an action.



To right: The robot used to conduct the experiments equipped with LEDs and a small computer, a “Raspberry Pi”.



The robot’s sensors

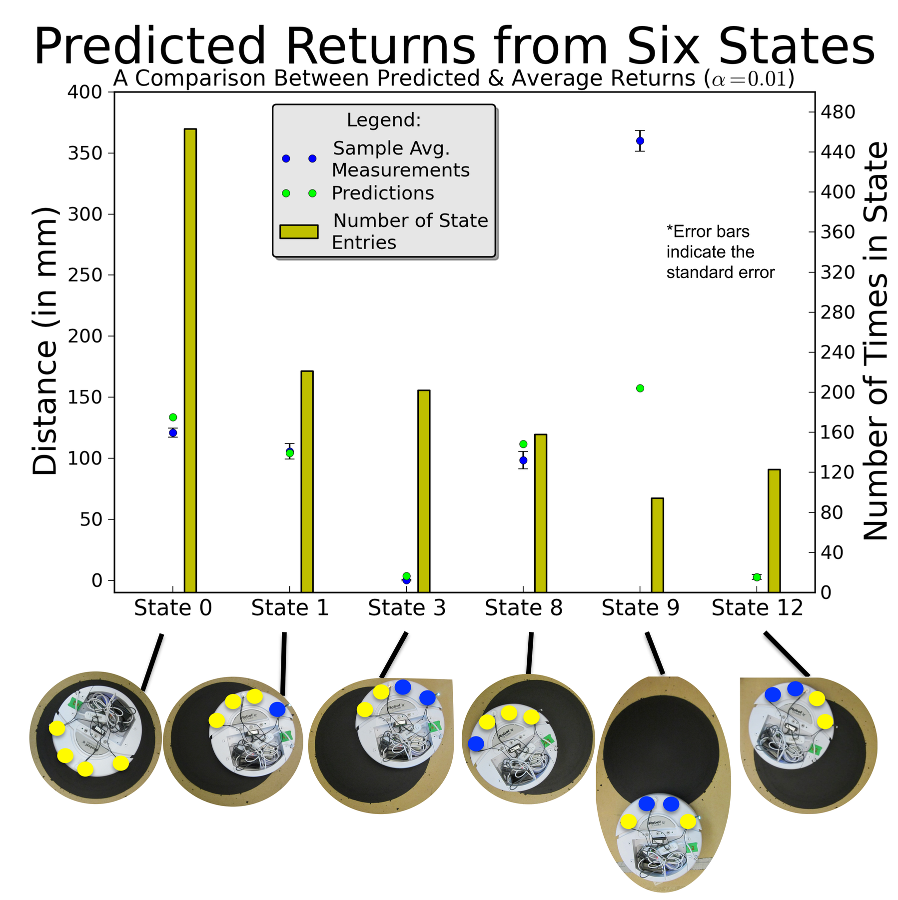
- The cliff sensors detect whether the sensor is above the circle (0) or the floor (1).
- The motor encoders measure the distance the robot has traveled in millimeters.
- The robot is forced to stop if either front cliff sensor detects the floor.

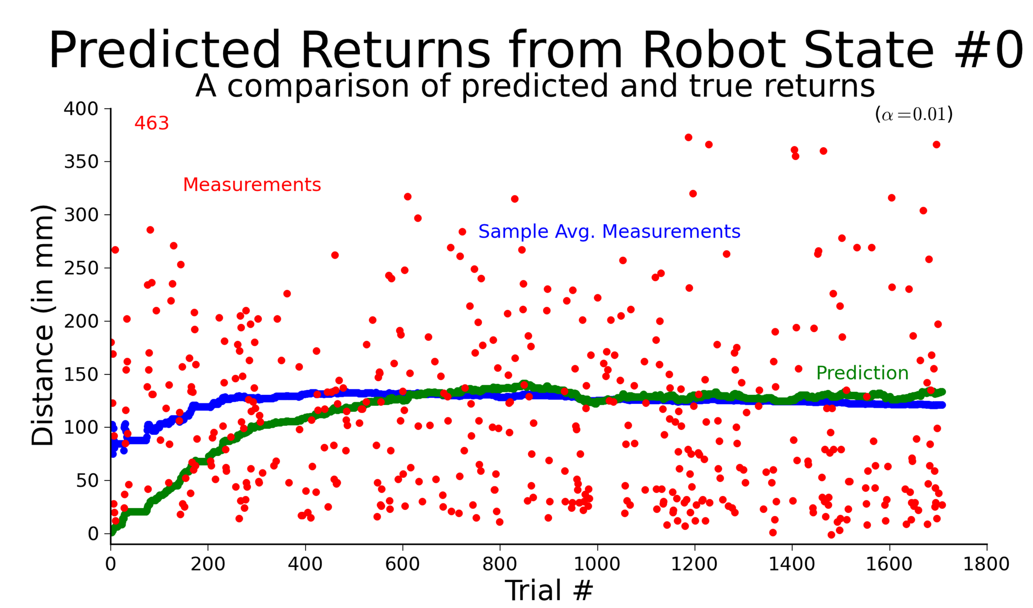
Prediction Experiment:

Objective: make the robot learn to predict how far it expects to move forward starting from a particular sensory configuration until one of its front sensors detects the floor.

Method:

- Each of the sixteen sensory configurations corresponds to a different state.
- The robot’s behavior alternates between wandering and advancing modes.
  - Advancing: moving forward until it reaches the edge and measures the distance
  - Wandering: takes random actions to get into different states
- The robot learns and measures while it is in advancing mode. The robot learns to predict with the TD(lambda) algorithm.





Results:

- Predictions from frequently visited states approach the average measurements for a run.
- State 9 had the highest sample average measurement but a lower-than-expected prediction due to lack of visits (see graph to right).

Control Experiment:

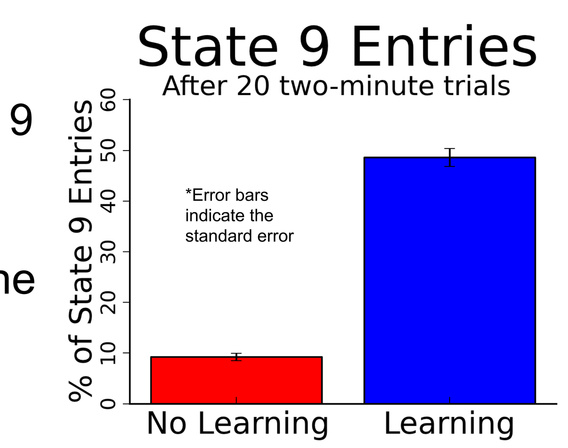
Objective: Make the robot get into the state 9 frequently

Method:

- The robot learns with the SARSA(lambda) algorithm
- Positive reward given only when the robot is in state 9

Results:

- The robot succeeds in getting into state 9 more than it does with no learning,
- The algorithm found that the moving backwards action maximized its reward



Future Direction:

The next step of the project is to combine the separate methods of locating the best state for the farthest advance and learning the way to enter the state. This should get more accurate predictions from infrequently visited states, such as state 9.

Acknowledgements

I would like to thank the dedication and help of my supervisor, Joseph Modayil, as well as the rest of the RLAI lab for their help and insight. I also wish to thank the HIP program for this opportunity.

References:

(1) Sutton, Richard S.; Barto, Andrew G., Reinforcement Learning: An Introduction, 1998, 342p.

(2) “iRobot Sensors”. Photograph. *Labview Hacker*. Aug. 6 201