Reinforcement Learning on a *TJ Pro™*

Motivation

The following is a modified excerpt from "Comparison of Reinforcement Learning Techniques for Automatic Behavior Programming" by J. Andrew Bagnell, Keith L. Doty, and A. Antonio Arroyo.

The reactionist's paradigm of machine intelligence emphasizes a direct coupling of behaviors to a robot's sensor and actuators. This approach and the architectures inspired by it, has lead to the development of sophisticated and robust agents, and moreover, has made that development an incremental and extensible process. [Brooks, 1986]

The implementation of the individual controllers required by a behavior-based robot remains a painstaking process of hand coding. Building our program based on the efforts of [Mahadevan and Connell, 1992] in automatic behavior programming, we present to you a reactive obstacle-avoidance controller for a TJ Pro™ robot using the techniques of *reinforcement learning* (RL).

[Kaebling, 1996] defines RL in the following way:
"Reinforcement learning is the problem faced by an agent that must learn behavior through trial-and-error interactions with a dynamic environment… It is appropriately thought of as a class of problems, rather than a set of techniques."

Use

The algorithm implement here is called Q-learning, invented by Christopher Watkins to solve RL problems. [Watkins and Dayan, 1989]

The controller develop for the TJ Pro™ was one that would enable it to explore it environment, but avoid any obstacles in its path. We implement this by rewarding the robot for traveling forward and punishing it for hitting obstacles with its bumper.

This, of course, means it is critical that your TJ Pro™ know that it is hitting an obstacle. If the bumper is not responding to the obstacles in its path, the robot can never hope to learn. If this happens in your experiments with TJ Pro™, run it on a higher friction surface, or with harder objects, or with weight on its back to ensure the bumper responds.

As long as the bumper works, you should be able to drop your TJ Pro™ in a new environment, and it should learn to maneuver around skillfully—give it time though! The more cluttered the environment with boxes and stuff, the more interesting! At first it will wander around maniacally, slamming into walls and boxes. Slowly, the robot will learn to turn away form obstacles; at first only after it hits them and then from further away. It will occasionally hit things, even after learning, since it executes random regularly to ensure it can try all actions in all possible conditions it might encounter. After all, if it doesn't explore all his options, it can't become an expert!

The rear IR emitter of the robot can be replaced by a visible LED so you can tell when the robot is making a random maneuver as opposed to a learned maneuver. Early in the learning process the robot makes a number of random manuevers that appear to be doing collision avoidance. Don't be fooled! After awhile the robot begins to seriously bang into objects and then, after a few minutes clearly avoids most obstacles.

At the robot has learned for about 5-10 minutes, change the environment radically and notice what happens! The robot bumping frequency may increase temporarily, but then should subside after a few minutes. The robot is learning the new environment!

So give it a try, and *ENJOY*!

For more information on Reinforcement learning and mobile robotics in general consult the following references:

[Mahadevan and Connell, 1992] Sridhar Mahadevan and Jonathan Connell, "Automatic Programming of Behavior-based Robots using Reinforcement Learning*", Artificial Intelligence , vol. 55, Nos. 2-3, pp. 311-365,* June, 1992 .

[Watkins and Dayan, 1992] Christopher J. C. H. Watkins and Peter Dayan. Q-Learning. Machine Learning, 8, 279-292, 1992.

[Brooks, 1986] Rodney Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1), 1986.

[Kaebling and Littman, 1996] Leslie P. Kaebling and Michael L. Littman, Reinforcement Learning: A Survery. Journal of Artificial Intelligence Research 4, Morgan Kauffman, 1996. http://www.cs.brown.edu/publications/techreports/reports/CS-94-39.html