Incremental Reinforcement Learning for Autonomous Agents

Kevin R. Dixon

Electrical and Computer Engineering
Carnegie Mellon University

Prepared for the Ph.D. Qualifying Examination

The problem of autonomous agents learning a complex task in an unknown environment is considered. The agents can both sense their environment through sensors and affect their environment through a set of actions. Each agent attempts to learn a function mapping its input state space to its set of actions while maximizing the expected sum of a discounted external reward signal that the agent receives at every discrete time step. This function is called a policy.

There are two broad classes of algorithms for solving problems of this type. The first is called Dynamic Programming (DP). The second is called Reinforcement Learning (RL). In general, DP finds the optimal policy but suffers from the "curse of dimensionality". Through trial-and-error, RL focuses on finding sub-optimal policies, thereby avoiding the computational explosion of DP. Therefore, RL is particularly useful when the agent's state space is so large that DP becomes impractical. However, in many realistic problems, the size of the state space requires that function approximators (e.g., artificial neural networks) be used with RL, so that the agent need not explore its environment in as much detail because of generalization in the function approximator. Even when function approximators are used, sufficiently exploring the environment may prove prohibitive to the agent because of the sheer enormity of the state space, the low probabilities of entering certain states, or the cost of generating the data. To construct a reasonable policy in such problems within a satisfactory amount of time, the agent may need to bias the exploration of its environment to investigate the states that are "important" more thoroughly.

In this talk, I will present a method for biasing the exploration of a RL agent by incorporating previously learned tasks. These previously learned tasks can be used to "guide" the agent through a more complex task. This method has the ability to speed up learning and, since it is inherently incremental by nature, can lead to more robust policies. I will also present results of this method based on a group of mobile robots learning to play a "game" in a simulated environment and compare these results to previous methods.