CSCI 446 Artificial Intelligence Project 4 Design Report

Roy Smart Nevin Leh Brian Marsh

December 1, 2016

1 Introduction

2 The Racetrack Problem

3 Reinforcement Learning Algorithms

3.1 Value Iteration

4 Q-Learning

4.1 Description

Q-learning is a model-free reinforcement learning method for determining optimal action-selection policies [Russell and Norvig, 2010]. An agent using this method leverages a quantity known as the Q-value to derive the optimal action a for each state s. The Q-value, $Q(s, a)$ describes the expected utility for every action in every state within the environment and is learned by the agent using temporal difference learning. An expression to calculate the Q-value is given by [Russell and Norvig, 2010] as

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[R(s) + \gamma \max_{a'} Q(s',a') - Q(s,a) \right] \tag{1}$$

where a is the action that was executed in state s that resulted in state s' and $R(s)$ is the reward function. The constants α and γ are known respectively as the learning rate and the discount factor. Equation 1 is used as an update rule to adjust the value of $Q(s, a)$ for each action-state pair in every time trial undertaken by the agent. Using this simple update rule and randomly initialized Q-values for every action-state, an agent can learn how to navigate an environment.

4.2 Design Implications

Since the racetracks are reasonably small, we can afford to sore the table of Q-values in memory. We will base our Q-learning agent off of the function Q-Learning-Agent provided by [Russell and Norvig, 2010]. To ease the training time, we will incrementally train the Q-learning agent by beginning training close to the finish line, and then increasing the distance as the agent learns each section.

5 Software Design

6 Experiment Design

For each agent we will generate learning curves that describe the number of steps required to reach the goal state vs. the number of training iterations. We will also track the number of times each agent runs into a wall and needs to be restarted. Each agent will be trained at least 10 times to determine reliable learning curves.

7 Summary

References

[Russell and Norvig, 2010] Russell, S. and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Pearson Education, Upper Saddle River, New Jersey 07458, 3rd edition.