AA228 Project 2 - Reinforcement Learning

Shane Dirks

November 9, 2018

Abstract

For this project I utilized a straightforward SARSA-λ algorithm to calculate the Q values for each state-action pair in the training data. Once the Q values were calculated I created a policy by selecting the optimal action where data was available. If a given state was not present in the training data a random action was selected.

1 Overview

The calculation of the Q values using SARSA-λ was relatively straight-forward. I utilized a static α value of 0.1 and a static γ of 0.5, both of which were chosen relatively arbitrarily after a few tests. The λ values were kindly given to us along with the training data. Once the Q values were calculated, selection of actions for states visited in the training data was straight-forward. For the medium and large data sets, the policies had to specify actions for states that never appeared in the training data. I implemented a simple function that would randomly choose an action for these states, but a number of interpolation techniques could be implemented with relative ease given a bit more time.

2 Performance

Computation time was not nearly as big of a concern for this project compared to project 1. The calculation of Q values using SARSA-λ finished in under a minute for all data sets using the first implementation I wrote, no optimization passes needed. For the small data set the creation of the policy took only fractions of a second. However, when it came to the medium and large data sets, the creation of the policies took a significantly longer time than the calculation of the Q values. (See the table 1 for all calculation times.)

Data Set	**Q Value Time**	**Total Time**
Small	8.451	8.482
Medium	34.200	1278.837
Large	44.093	749.288

Table 1: Times to calculate the Q values and total execution time for each data set in seconds.

The policies themselves performed surprisingly well, especially considering the fact that the large and medium policies contained a not-insignificant number of random actions. The scores can be seen in the table 2.

Data Set	**Score**
Small	13.6127
Medium	87.0031
Large	1197.8635

Table 2: Scores for each policy submitted (taken from the leaderboard).