Extra Credit Project 6: Reinforcement Learning

Due Dec. 11, 2017

Overview

In this project you will apply reinforcement learning to a simulated robot navigation problem, for which you will use eligibility traces and TD learning.

Your robot will move in a 20×20 gridworld, in which the squares are either unoccupied or contain an obstacle. Your robot can move to an adjacent unoccupied square in four directions (up, down, right, left). In any given gridworld, one unoccupied squared is designated as the *goal*. An episode begins by placing the robot on an unoccupied square sufficiently far from the goal, and then allowing it to act until it either reaches the goal or until a specified maximum number of actions are taken. If it reached the goal, it receives a reward of 100, otherwise it receives no reward.

Methods

(1) Implement functions to create gridworlds. You should be able to create a specific gridworld (e.g., either read from a file or specified as a data structure) or to create a random gridworld with a certain density of obstacles. Your functions should also allow either the designation of a square as the goal or random selection of an unoccupied square as the goal.

(2) Develop a function to estimate the $V(x,y)$ matrix by running multiple episodes from random starting states (on unoccupied squares), running until the goal is reached or a specified maximum number of steps is reached. Use the V matrix and an ε-greedy policy to choose actions.

(3) Plot a running average of the reward per episode so you can monitor learning. Also plot a running average the number of steps until a goal is reached (i.e., the lengths of the episodes).

(4) Define a function to plot a learned V matrix as, for example, a heat map.

(5) When your functions are implemented, demonstrate their operation on several hand-designed simple and complex gridworlds.

(6) Next demonstrate their operation on several random gridworlds of various densities.

(7) For both (5) and (6) experiment with various values of the hyperparameters γ, ε, η, and λ. Report your results.