Exercise 0: An Invitation to Reinforcement Learning

Four Rooms Problem

A picture containing text, shoji, crossword puzzle

Description automatically generated

Prof: Robert Platt

Date: September 17th, 2021

Name: Guanang Su

Part 2. Manual Policy

This is some procedures running in the manual policy with the user interface. The user with go from (0, 0) and try to reach the final state (10, 10). By using l(left), r(right), u(up), d(down) to control the agent to move with keyboard user input. It will show you the location after each move and the action you take after noise with a reward. After the user arrive at (10, 10), the agent will go back to the starting state and run again until you reach the maximum trails number or killed the process with “crtl+C”.

Text

Description automatically generated

Text

Description automatically generated

Part 3. Random Policy

This is a figure of the random policy with running 10 trails and 10,000 steps on each trail. The trail is in dotted colored lines and perform as a black solid line with the mean of each trail on each step. The mean result of the policy achieves a cumulative reward slightly over 8, which means that on 10000 steps, it reaches the goal approximately 8 times. It takes about 125 steps to arrive the final state.

Chart

Description automatically generated

I would say that random policy is much worse than the manual policy, usually, I took about 25-30 runs to arrive at the star position, (10, 10). The reason that come up with the result in my random policy, the system picks a random action from the four direction each time from the current state, so the step is not controllable since it is random generated, and the right position is up and right on the corner. Compared to the random one, the manual policy is on the user input. User could decide their purposed action by watching the grid and wall to control the agent to move closer to the final position with a shorter path. So the manual policy would perform better than the random one.

Part 4. Two more policies and comparison between all policies

Chart, scatter chart

Description automatically generated

Chart

Description automatically generated

Graphical user interface, chart

Description automatically generated