Intelligent Robotic Systems - Assignment 3

Ola Ringdahl and Polina Kurtser

November 22, 2021

Question 1 [15 pts]

Define and explain the following terms [3pts/term]:

- Markovian property of a stochastic process.
- Markov chain
- Markov decision process (MDP).
- Value iteration
- Policy and Policy iteration.

Question 2 [45 pts]

Figure 1: $3X4$ planar gridworld with one obstacle and two absorbing (terminal) states

Given is a robot in a start state of a $3X4$ planar gridworld (Figure 1) with one obstacle and two absorbing (terminal) states. In the terminal states the robot receives rewards of $+1$ (goal state) and -1, respectively, whereas for all other states a reward of r is collected. In this gridworld the robot can perform four actions $f = \{N; E; S; W\}$. The probability for the commanded direction to occur is p_{act}, but with probability $1 - p_{act}$ the robot moves at right angles to the intended direction(e.g. if a command "N" was given from the starting point there is a p_{act} chance to end up at the cell above the start and a $1 - p_{act}$ chance to end up to the right. It is assumed that the boundaries are reflective, i.e., if the robot hits a wall or obstacle it remains in the same grid cell.

- Draw a graphical model of the MDP for $p_{act} = 1$, and $r = -0.04$. Provide an intuitive solution for the MDP and explain it. [5 pt]
- Draw a graphical model of the MDP for $p_{act} = 0.8, r = -0.04$ and $a = "N"$. Showing all states, transition probabilities and rewards. [10 pt]
- Evaluate the policy in Figure 2 for $p_{act} = 0.8, r = -0.04$. [15 pt]
- Solve the MDP using value iteration with $\gamma = 1, p_{act} = 1, r = -0.04$. Provide optimal policy and the value of that policy for each state. Choose a termination threshold you find reasonable and explain your choice. [15pt]

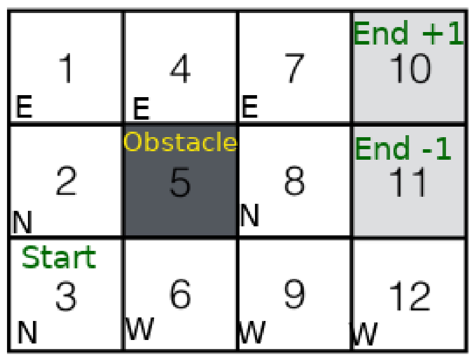

Figure 2: $3X4$ planar gridworld with one obstacle and two absorbing (terminal) states and a policy

Question 3 [40 pt]

Implement a Monte-Carlo GLIE, first-visit, incremental control algorithm on the girdworld in Figure 1 with discount factor $\gamma = 0.9$. For that you will need to:

- Implement an episode generator function that takes as input values of r and p_act and generates a random episode path. Attach a snapshot of the code with basic documentation [10 pt]
- Implement a gain G_t calculation function that takes in an episode generated above and discount factor γ and returns the episode gain. Attach a snapshot of the code with basic documentation [5 pts]

- Use the functions above to implement a Monte-Carlo GLIE, first-visit, incremental control algorithm with discount factor $\gamma = 0.9$, step reward $r = -0.04$, and $p_act = 0.8$. Find a suitable step-size parameter α and ϵ decay-rate for GLIE. Run the implementation 3 times and present the solutions acquired each time. Explain the result. [25 pt]

Submission

A report will be submitted through moodle no later then 05/12/2021.