Lab 8: Markov Decision Process

AI701: Artificial Intelligence

October 2021

1 Introduction

In this lab, you will implement Markov decision process by algorithms: Policy evaluation and Value iteration. These algorithms will be built to solve practical decision making.

Suppose we want to implement the task of a robot walking to the destination over obstacles. As shown in the following Figure 1, the robot is located in a 4×3 grid, the black area represents the obstacle, and the destination is the grid (3, 2) in the Upper right corner, we use +1 to represent it, and to avoid walking on the grid (3, 1), we use -1 to represent it. Now we consider this problem from the perspective of the MDP:

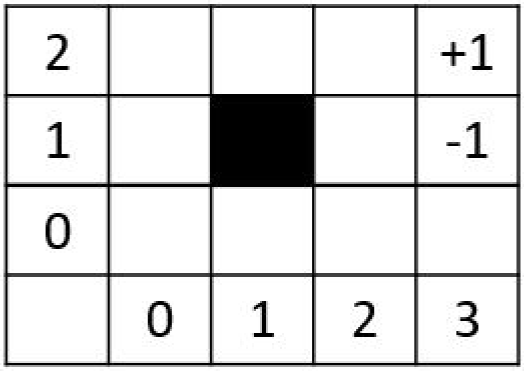

Figure 1: Example of Markov Decision Processes

1. **States:** The robot can exist in any of the 11 grids, so there are a total of 11 states, and the states set S represents the position that it can reach.
2. **Actions:** The actions the robot can make are up, down, left and right, actions set $A = \{U, D, L, R\}$.
3. **State transition distribution (P_{sa}):** Assuming that the core design of the robot's behavior is not so accurate, the robot may go off-track or

travel less accurately after receiving the relevant instructions. In order to simplify the analysis, establish the robot's dynamic model as follows Figure 2.

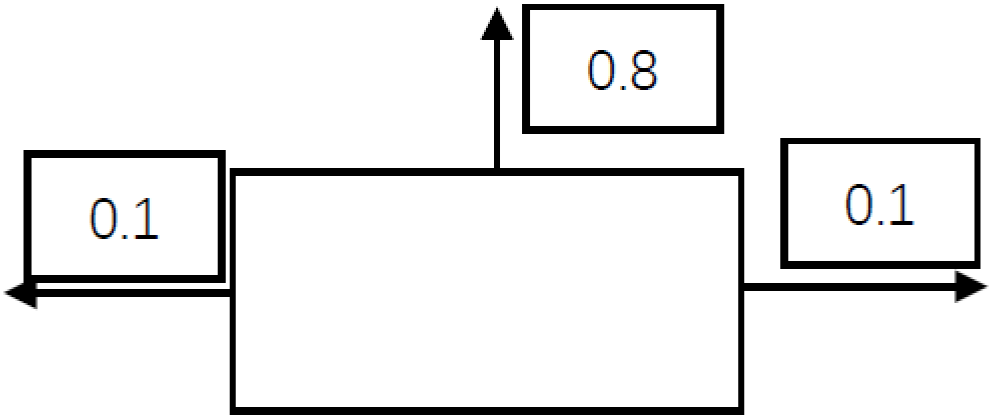

Figure 2: Robot activity model

When the robot is commanded to walk upwards, it has a probability of 0.1 to move to the left or right, and a probability of 0.8 to move to the specified direction. When the robot walks to a wall or to a non-adjacent grid, its probability is 0 .

4. **Reward:**The reward function can be set as follow.

   $R((3,2)) = +1.$

   $R((3,1)) = -1.$

   $R(s) = -0.04$ for other states s.

5. **Discount factor:** $\gamma \in [0,1]$.

2 Implement Markov Decision Processes

The dynamic process of MDP is as follows: the initial state of an agent is s_0, and then select an action a_0 from A to execute it. After execution, the agent randomly transfers to the next state s_1 according to the probability of P_{sa}, $s_1 \in P_{s_0 a_0}$. Then perform an action a_1 then move to state s_2, and then perform a_2 ...

Your task is to implement the state representation, transition model and optimal policy needed based on the case above. You also need to implement policy evaluation and value iteration algorithms to solve Markov decision processes.