Markov Decision Process Problem Statement

Zhi29

May 2019

1 Setup

Consider a simple robot in a 2D grid world of length L and width W. That is, the robot can be located at any lattice point (x,y): $0 \leq x < L, 0 \leq y < W; x, y \in N$. At each point, the robot can face any of the 12 headings identified by the hours on a clock $h \in \{0, 1, ..., 11\}$, where $h = 0$ represent 12 o'clock is pointing up(i.e. positive y) and $h = 3$is pointing to the right (positive x)

In each time step, the robot can choose from several actions. Each singular action will consist of a movement followed by a rotation.

1. The robot can choose to take no motion at all, staying still and neither moving nor rotating.

2. Otherwise the robot can choose to either move "forwards" or "backwards". a. This may cause a pre-rotation error, see below. b. This will cause the robot move one unit in the direction it is facing, rounding to the nearest cardinal direction.

That is if the robot is pointing towards either 2, 3, or 4 and opts to move "forwards", the robot will move one unit in the +x direction. SImilarly, if the robot is pointing towards wither 11, 0, or 1 and opts to move "backwards", it will move one unit in the -y direction.

3. After the movement, the robot can choose to turn left, not turn, or turn right. A left(counter clockwise) turn will decrease the heading by 1(mod 12); right(clockwise) will increase the heading by 1 (mod 12). The robot can also keep the heading constant.

4. Attempting to move off of the grid will result in no linear movement, but the rotation portion of the action will happen.

Note that aside from the at edges of the grids, the robot can only rotate if it also moves forwards or backwards; it can move without rotating though.

The robot has an error probability p_e: if the robot chooses to move, it will first rotate by +1 or -1(mod 12) with probability p_e each, before it moves. It will not pre-rotate with probability $1 - 2p_e$. Choosing to stay still (take no motion) will not incur an error rotation.

The environment is:

The rewards for each state are independent of heading angle (or action taken). The border states $\{x = 0, x = L, y = 0, y = W\}$(red marked X) have

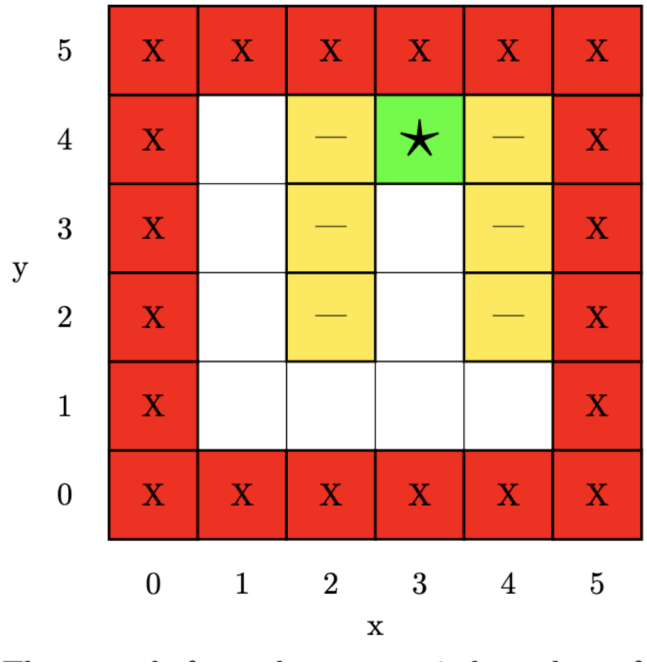

Figure 1: Environment

reward -100. The lane markers (yellow, marked -) have reward -10. The goal square (green, marked *) has reward +1. Every other state has reward 0.