MSc in Computer Science and Engineering

Planning, Learning and Decision Making

Homework 2. Markov decision problems

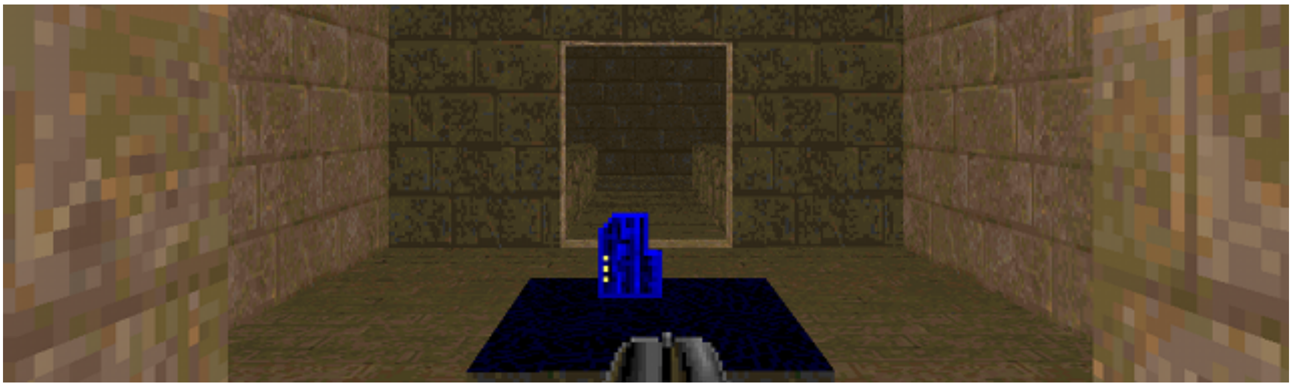

Figure 1: A "Doom" player is facing the blue key.

Consider the game "Doom",[1] where a player must survive a maze and reach the exit (see Fig. 1 for an example screenshot). In this homework, you will model an extremely simplified instance of a "Doom" game as a Markov decision problem.

◇

Suppose that the agent moves in the environment depicted in Fig. 2. At each step, the agent may move in any of the four directions—up, down, left, and right. Movement across a *grey* cell division succeeds with a 0.8 probability and fails with a 0.2 probability. Movements across colored cell divisions (blue or red) succeed with a 0.8 probability (and fail with a probability 0.2) *only if the agent has the corresponding colored key.* Otherwise, they fail with probability 1. When the movement fails, the agent remains in the same cell. Finally, assume that the action "left" always succeeds in taking the agent through the exit.

To get a colored key, the agent simply needs to stand in the corresponding cell. In other words, as soon as the agent stands on the cell of a colored key, you may consider that it holds that key thereafter.

[1] `https://en.wikipedia.org/wiki/Doom_II`

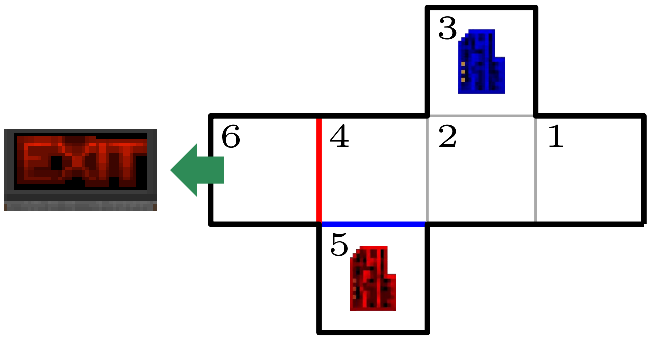

Figure 2: Environment where the agent must reach the exit (located in cell 6).

Exercise 1.

(a) Identify the state space, $\mathcal{X}$, and the action space, $\mathcal{A}$, for the MDP. You should treat "Exit" as an absorbing state. Moreover, you can assume that the agent never has the red key without the blue key and never reaches the goal without both keys.

(b) Write down the transition probability matrix for the action "right" and a (possible) cost function for the MDP. Make sure that the cost function is as simple as possible and verifies $c(x, a) \in [0, 1]$ for all states $x \in \mathcal{X}$ and actions $a \in \mathcal{A}$. Note, in particular, that the cost should depend only on *where the agent is.*

(c) Compute the cost-to-go function associated with the policy in which the agent always goes right, using a discount $\gamma = 0.9$. You can use any software of your liking for the harder computations, but should indicate all other computations.