CSCD84: Artificial Intelligence

Problem Set 3: MDP

Solution by: Ali Parchekani

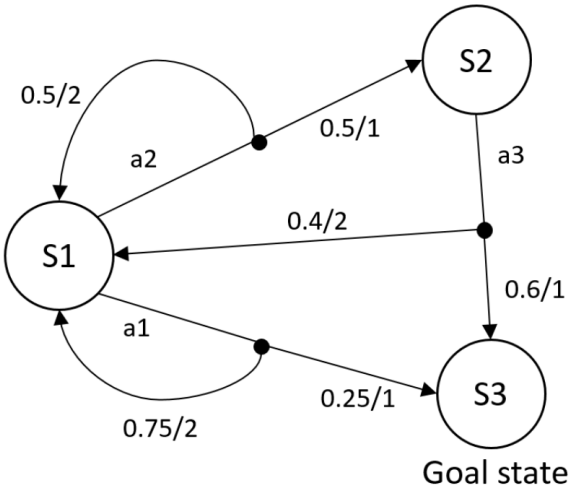
UTORid:

First and Last Name:

By turning in this assignment, I agree by the University of Toronto honor code and declare that all of this is my own work.

An MDP with a single goal state ($S3$) is given below.

1. Given the expected goal distances $c(S1) = 7, C(S2) = 4.2$, and $C(S3) = 0$, calculate the optimal policy for state S1.
2. Suppose that we want to follow a policy where we pick action a2 in state S1 and action a3 in state S2. Calculate the expected cost of S1 and S2 for this policy.



Solution:

1. We use $c(s, a)$ to denote the expected cost of reaching a goal state if one starts in state s, executes action a and then acts according to the policy. Since S3 is a goal state and S2 has only one available action, we only need to calculate $c(S1, a1)$ and $c(S1, a2)$ in order to decide whether to execute $a1$ or $a2$ at S1.

$$c(S1, a1) = 0.25(1 + c(S3)) + 0.75(2 + c(S1)) = 0.25(1 + 0) + 0.75(2 + 7) = 7 \quad (1)$$
$$c(S1, a2) = 0.5(1 + c(S2)) + 0.5(2 + c(S1)) = 0.5(2 + c(S1)) = 0.5(1 + 4.2) + 0.5(2 + 7) = 7.1 \quad (2)$$

Since $c(S1, a1) < c(S1, a2)$, in the optimal policy, we execute $a1$ at S1.

2. Since the given policy executes $a2$ at S1, we simply ignore $a1$ during our computation. We first generate the following set of equations.

$$c(S1) = c(S1, a2) = 0.5(1 + c(S2)) + 0.5(2 + c(S1)), \quad (3)$$
$$c(S2) = c(S2, a3) = 0.6(1 + c(S3)) + 0.4(2 + c(S1)), \quad (4)$$
$$c(S3) = 0.$$

Solving this system of equations would result in

$$c(S1) = 2.2/0.3 = 7.33,$$
$$c(S2) = 1.4 + 0.4c(S1) = 1.4 + 0.4(7.33) = 4.333.$$