National University of Singapore
School of Computing

Semester 1, AY2021-22

CS4246/CS5446 AI Planning and Decision Making

Issued: 22 Oct, 2021

Tutorial Week 11: Partially Observable Markov Decision Process

Guidelines

You may discuss the content of the questions with your classmates. But everyone should work on and be ready to present ALL the solutions.

Problem 1: 4×3 Grid World

[Modified from RN 3e17.13] We can convert the 4×3 world of Figure 17.1 into a POMDP by adding a noisy sensor instead of assuming that the agent knows its location exactly. Such a sensor might measure the number of adjacent walls, which happens to be 2 in all the nonterminal squares except for those in the third column, where the value is 1; a noisy version might give the wrong value with probability 0.1.

Let the initial belief state be b_0 for the 4×3 POMDP be the uniform distribution over the non-terminal states, i.e.,

$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	0
$\frac{1}{9}$	$\times$	$\frac{1}{9}$	0
$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

Calculate the exact belief state b_1 (rounded off to 5 decimal places) after the agent moves *Left* and its sensor reports 1 adjacent wall.

Problem 2: Complexity

[Modified from RN 3e 17.14] What is the time complexity of d steps of POMDP value iteration for a sensorless environment? Give an upper bound on the number of α-vectors generated in the process.

Problem 3: Captain Jack's Adventure

Captain Jack would like to go to Treasure Island (Island) but does not know the way. He knows that the Island is on his left (state s_1) with probability p and the Island is on his right (state s_2) with probability $1-p$. If he goes in the wrong direction, he would end up in Pirates Den (Den), a place that he wants to avoid badly. Captain Jack has three possible actions. He can go left (action a_1), go right (action a_2), or ask the Lighthouse Keeper (Keeper) at his current docking harbor (action a_3) whether to go left or right. If he goes in the correct direction, he gets a reward of 100 (e.g. $R(s_1, a_1) = 100$) but if he goes in the wrong direction he gets a penalty of -100 (e.g. $R(s_1, a_2) = -100$). The Keeper never lies, providing the observations left for Island on the left, and right for Island on the right. But asking the Keeper will cost -10 (i.e. $R(s_1, a_3) = R(s_2, a_3) = -10$).

(a) The value of a one-step plan taken in state s is simply the reward of taking the action a in state s: $R(s, a)$. Going left or right are terminal actions while asking the Keeper is non-terminal. Hence, two-step conditional plans can only start with the non-terminal action of asking the Keeper (a_3) followed by an observation and ends with taking another action.

  (i) How many two-step conditional plans that starts with action a_3 are there?

  (ii) There is only one non-dominated two-step conditional plan: draw (or clearly describe) the non-dominated two step conditional plan.

(b) The one-step plan consisting of asking the Keeper cannot be optimal. Hence there can be at most two non-dominated one-step plans. From part (a) of this question, we know that there is only one non-dominated two-step conditional plan, giving a total of 3 non-dominated one and two step plans.

  (i) Give the three α-vectors corresponding to the three non-dominated plans. Assume that the discount factor is $\gamma = 1$ (not discounted).

  (ii) Partition the beliefs into regions where each plan is optimal. Describe the regions.