Assignment-5
Part-B Report

Jaidev Shriram (2018101012)

April 2020

1 Parameters

x = 0.87
reward = 22

2 Clarifications

Here, we have assumed that on reaching the final state, **the call is turned off immediately, and transition probabilities are calculated from this new state**. This is as per the clarification on Moodle.

3 Questions

3.1 If you know the target is in (1,1) cell and your observation is o6 , what will be the initial belief state?

The initial belief state will be:

*0 0 0 0 0 0 0 0 0.125 0.125 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0.125 0.125 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0*

0 0 0 0 0.125 0.125 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.125 0.125 0 0 0 0 0 0 0 0

where these 162 states are in increasing order of the form (AgentPos, TargetPos, isCallOn). Here, the probability of starting from a non neighbouring state of (1,1) is uniformly distributed. This would be the corner points, with two states per corner for the Call Status. Hence, there are a total of 8 states with a probability of $\frac{1}{8} = 0.125$ per such state.

The policy file for this has been attached

3.2 If you are in (0,1) and you know the target is in your one neighborhood and is not making a call what is your initial belief state?

We can consider one neighbourhood as the set of states that are one manhattan distance away from the agent position - $(0, 1)$. Hence, the possibilities for the target are:

1. $(0, 0)$
2. $(1, 1)$
3. $(0, 2)$
4. $(0, 1)$ - It can even be in the same location!

Hence, there are four possibilities for the target location. Since the call is known to be off, and agent position is known, there are only 4 states that are equally probably. Hence, the belief states are as follows:

1. b((0, 1), (0, 0), Off)) = $\frac{1}{4}$
2. b((0, 1), (1, 1), Off)) = $\frac{1}{4}$
3. b((0, 1), (0, 2), Off)) = $\frac{1}{4}$
4. b((0, 1), (0, 1), Off)) = $\frac{1}{4}$
5. belief of Remaining States = 0

3.3 What is the expected utility for initial belief states in questions 1 and 2?

We can find the expected utility for the initial belief states by running *pomdpeval* with the *.pomdp* file generated along with the policy file created from *pomdpsol*. Here, we use 100 simulations with 1000 steps in each simulation.

Question 1: Expected Utility = Expected Total Reward = 2.67458
Question 2: Expected Utility = Expected Total Reward = 6.60525

3.4 If your agent is in (0,1) with probability 0.6 and in (2,1) with probability 0.4 and the target is in the 4 corner cells with equal probability, which observation are you most likely to observe? Explain.

The states $(0, 1)$ and $(2, 1)$ are the two extreme positions in the central row of our grid. They are sandwiched between two corner positions.

If we know the probability of an observation given a state $P(o|s)$, then the probability of an observation is:

$$P(o) = \sum_{s} P(s) * P(o|s) \tag{1}$$

All the states of the system can be understoof as the cartesian product of possible agent positions, target positions. Since there is no observation specific to calls, we can ignore it altogether.

Let us consider each observation one by one. Note that the observations are 100% accurate, so $P(o|s)$ will be 0 or 1 only.

o1

o1 is observed when the target is in the same cell as the agent.

This is never the case as per the question. Hence, probability is 0.

o2

o2 is observed when the target is in the cell to the right of the agent's cell.

This is also never going to happen.

o3

o3 is observed when the target is in the cell below agent's cell

This is a likely scenario

$$P(o3) = 0.6 * 0.25 * 1 + 0.4 * 0.25 = 0.25 \tag{2}$$

o4

o4 is observed when the target is in the cell to the left of agent's cel

This is never going to happen

o5

o5 is observed when the target is in the cell above the agent's cell

This is a likely scenario:

$$P(o5) = 0.6 * 0.25 + 0.4 * 0.25 = 0.25 \tag{3}$$

o6

o6 is observed when the target is not in the 1 cell neighbourhood of the agent.

This is possible too.

Assume that the agent is at $(0, 1)$, then the target can be at $(2, 2)$ or at $(2, 0)$, both of which are not in the one neighbourhood of the agent. We can predict a similar situation for the other position of the agent.

Hence, for each chosen agent state, there are two target states with probability 0.25 each. Hence, total probability of the observation is:

$$P(o6) = 0.6 * (0.25 + 0.25) + 0.4 * (0.25 + 0.25) = 0.5 \tag{4}$$

Hence, it's clear that $o6$ is most likely to occur. Additionally, the observation is more likely to happen near $(0, 1)$ since that state has a higher probability.

3.5 How many policy trees are obtained in this case, explain?

Policy trees may be calculated as following:

$$N = \sum_{i=0}^{T-1} |O|^i = \frac{|O|^T - 1}{|O| - 1} \tag{5}$$

Number of trees = $\|\mathrm{A}\|^N$

Here A refers to the actions possible, O the observations possible, and T is the time horizon - the number of steps the agent takes.

In our situation, we have 5 actions, 6 observations, and the time horizon (trial count) is 156. This is because on running *pomdpsol*, it converged after 156 steps. The precision value at the end of these steps was 0.00099, below the target precision of 0.001.

In this case, there are infinite policy trees that are possible. This is because, as we increase the horizon, the number of nodes does not converge. Naturally, since our POMDP model doesn't have an absorbing state, this is possible. Hence, as we increase our horizon size, there'll be new policy trees, making an infinite total.