CS7.301 Machine, Data, and Learning

Assignment 05B: Successive Approximations of the Reachable Space under Optimal Policies

Naren Akash R J

POMDP Environment Representation

(0,2)	(1,2)	(2,2)
(0,1)	(1,1)	(2,1)
(0,0)	(1,0)	(2,0)

Legend

	Agent
	Target
	Both

Question 01

The following information is given in the question.

- Target is in (1,1) cell of the grid.
- Agent is not in the one-cell neighborhood of the target.

From the above information, the agent can be in states (0,0), (0,2), (2,2) and (2,0). For each of the above states of the agent, the call status can be either on or off. Therefore, we get a total of 4 * 2 = 8 states.

(0,2)	(1,2)	(2,2)
(0,1)	(1,1)	(2,1)
(0,0)	(1,0)	(2,0)

The probability for each of the eight states is uniformly land is equal to 0.125. For all other states, we assign zero probability as their initial belief state.

States ((0,0), (1,1), 0), ((0,0), (1,1), 1), ((0,2), (1,1), 0), ((0,2), (1,1), 1), ((2,2), (1,1), 0), ((2,2), (1,1), 1), ((2,0), (1,1), 0), and ((2,0), (1,1), 1) each get the value 0.125 as their initial belief state.

Question 02

The following information is given in the question.

- Agent is in (0,1) cell of the grid.
- Target is in the one-cell neighborhood of the agent.
- Target is not making a call.

(0,2)	(1,2)	(2,2)
(0,1)	(1,1)	(2,1)
(0,0)	(1,0)	(2,0)

From the above information, we can say that the target is in one of the following cells: (0,0), (0,1), (0,2), and (1,1). The probability of the target to be in one of these states is uniform and is equal to 0.25.

States ((0,1), (0,0), 0), ((0,1), (0,1), 0), ((0,1), (0,2), 0) and ((0,1), (1,1), 0) have the value of 0.25 in as their initial belief state. The rest of the states will have value equal to 0.

Question 03

For Question 01, we obtain the expected total reward to be approximately equal to 5.542 for hundred simulations. In case of Question 02, we obtain the same as 10.585.

Question 04

The following information is given in the question.

- With probability 0.6, the agent is in (0,1).
- With probability 0.4, the agent is in (2,1).
- Target is in one of the following cells: (0,0), (0,2), (2,2) and (2,0).

The probability of the target to be in one of the above-mentioned states is uniform and is equal to 0.25.

(0,2)	(1,2)	(2,2)
(0,1)	(1,1)	(2,1)
(0,0)	(1,0)	(2,0)

- When the target is in (0,0),
  - If the agent is in (0,1), the agent is above the target. [o3]

    *Probability = 0.6 * 0.25 = 0.15*

  - If the agent is in (2,1), the agent is not in the target's neighborhood. [o6]

    *Probability = 0.4 * 0.25 = 0.10*

- When the target is in (0,2),
  - If the agent is in (0,1), the agent is below the target. [o5]

    *Probability = 0.6 * 0.25 = 0.15*

  - If the agent is in (2,1), the agent is not in the target's neighborhood. [o6]

    *Probability = 0.4 * 0.25 = 0.10*

- When the target is in (2,2),
  - If the agent is in (0,1), the agent is not in the target's neighborhood. [o6]

    *Probability = 0.6 * 0.25 = 0.15*

  - If the agent is in (2,1), the agent is below the target. [o5]

    *Probability = 0.4 * 0.25 = 0.10*

- When the target is in (2,0),
  - If the agent is in (0,1), the agent is not in the target's neighborhood. [o6]

    *Probability = 0.6 * 0.25 = 0.15*

  - If the agent is in (2,1), the agent is above the target. [o3]

    *Probability = 0.4 * 0.25 = 0.10*

Now, we sum up the probabilities for each of the six observations.

- Probability(o6) = 0.50
- Probability(o5) = 0.25
- Probability(o3) = 0.25
- For the rest, the probability is zero.

Question 05

Given,

- the number of nodes in the tree, N
- the height of the tree (i.e., horizon of the POMDP), T
- the number of observation (here, equal to 6), |O|
- the number of actions (here, equal to 5), |A|

The number of policy trees, $P = |A|^N = 5^N$

where, $N = \Sigma_i (|O|^T - 1) / (|O| - 1)$

$= \Sigma_i (6^T - 1) / 5$

A good value of T turns out to be equal to 262 which we obtained from *#Trail* after running the *pomdpsolve* script for the [dot]policy file created.

$N = \Sigma_{i = 0 \text{ to } 261}(6^{262} - 1) / 5$

$A = 5^N = 5^{\Sigma i = 0 \text{ to } 261(6\text{^}262 - 1) / 5}$

Assumptions

- The step cost for 'stay' action is assumed to be equal to zero.
- The probability of making a move in the agent's intended direction is 0.79.
- The discount factor is taken to be equal to 0.5.

Reference(s)

1. David Hsu, Wee Sun Lee, Nan Rong: A Point-Based POMDP Planner for Target Tracking *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2008
2. POMDP File Format: How to encode a POMDP problem for pomdp-solve? *Retrieved from* http://pomdp.org/code/pomdp-file-spec.html
3. SARSOP: Approximate POMDP Planning (APPL) Toolkit Available on https://github.com/AdaCompNUS/sarsop