CS7.301 (Machine Data and Learning)

Vikrant Dewangan
Roll Number - 2018111024

30th April, 2020
Assignment - 5, part B

Parameters Involved

The following parameters were involved -

- x = 1 - (((2018111024) % 40 + 1) / 100)
  = **0.75**
  Thus the agents in moves in the direction of its action with probability **0.75** and opposite to the direction with probability **0.25** (or in the same cell if at border cells)
- Reward = 2018111024 % 100 + 10
  = **34**
  Thus the agent gets a reward of **+34** when it reaches the target before call is Off.
- If the agent transitions from any state to a terminal state, it immediately shuts the call. Now agent may or may not start a new call :
  if agent_state == target_state and call_state == 1:
      call_change_prob = [0.4, 0.6]
  else if call_state == 1:
      call_change_prob = [0.4, 0.6]
  else:
      call_change_prob = [0.2, 0.8]
  Note: call_change_prob[0] is probability of change initial state and call_change_prob[1] of no change.
- Numbering of agent's or target's states -

2	5	8

1	4	7
0	3	6

With each state being (agent state, target state, call state) tuple, we get a total 9 x 9 x 2 = 162 states.

1. Question 1

$A_?$	X	$A_?$
X	T	X
$A_?$	X	$A_?$

$A_?$: Agent uncertainty
X: Certainty that target or agent not present
T: Uncertainty of presence

Since upon observing o6, we are sure that the target is not present in cells (0,1), (1,0), (2,1) or (1,2). Thus they are certainly confident that the target must be in states (0,0), (2,0), (0,2) or (2,2). Thus initially, the target can be in either of the states with probability **0.25**. Thus we can compute the initial belief state as -

$$b(\{s_{agent}, s_{target}, call\}) = \begin{cases} 0.125 & \text{when } s_{agent} = (0,0) \text{ or } (0,2) \text{ or } (2,0) \text{ or } (2,2), \\ & \textbf{and } s_{target} = (1,1) \\ & \textbf{and } call = \text{On or Off} \\ 0 & \text{otherwise} \end{cases}$$

2. Question 2

?	X	X
A ?	?	X
?	X	X

A: Agent
X: Certainty that target not present
?: Uncertainty of presence

Since initially, it is known that the target is in one of the states (0,0), (1,1), (0,1) or (1,0) and that the call is **Off**, the initial belief state will look like -

$b(\{s_{agent}, s_{target}, call\}) = 0.25$ when s_{target} = (0,0) or (0,1) or (1,0) or (0,1)
and s_{agent} = (0,1)
and call = Off
$= 0$ otherwise

3. Question 3

3.1 Question 1

The policy file generated in Question 1 via sarsop is used to generate the simulation using simulation length as **100** and simulation number as **1000** -

#Simulations	Exp Total Reward
100	4.88268
200	5.00939
300	4.95585
400	4.88391
500	4.80111
600	4.71943
700	4.73074
800	4.81126
900	4.82644
1000	4.76012

#Simulations	Exp Total Reward	95% Confidence Interval
1000	4.76012	(4.58101, 4.93924)

The expected total reward is my expected utility, that is **4.76012**.

3.2 Question 2

The policy file generated in Question 1 via sarsop is used to generate the simulation using simulation length as **100** and simulation number as **1000** -

#Simulations	Exp Total Reward
100	10.4015
200	10.3102
300	10.3067
400	10.2584
500	10.2487
600	10.2509
700	10.2157
800	10.2241
900	10.2023
1000	10.1891

#Simulations	Exp Total Reward	95% Confidence Interval
1000	10.1891	(10.0563, 10.3219)

The expected total reward is my expected utility, that is **10.18912**

4. Question 4

$T_?$	X	$T_?$
$A_?$	X	$A_?$

$T_?$	X	$T_?$

$A_?$: Agent position uncertainty
X: Certainty that target or agent not present
$T_?$: Target position uncertainty
Thus we get our belief state as -

$$b(\{s_{agent}, s_{target}, call\}) = \begin{cases} 0.075 & \text{when } s_{target} = (0,0) \text{ or } (0,2) \text{ or } (2,0) \text{ or } (2,2), \\ & \textbf{and } s_{agent} = (0,1) \\ & \textbf{and } call = \text{On or Off} \\ 0.05 & \text{when } s_{target} = (0,0) \text{ or } (0,2) \text{ or } (2,0) \text{ or } (2,2), \\ & \textbf{and } s_{agent} = (2,1) \\ & \textbf{and } call = \text{On or Off} \end{cases}$$

Thus, we have total 8 different different scenarios -

Agent	Target	Probability	Observation
0,1	0,0	0.15	o3
0,1	0,2	0.15	o5
0,1	2,0	0.15	o6
0,1	2,2	0.15	o6
2,1	0,0	0.1	o6
2,1	0,2	0.1	o6
2,1	2,0	0.1	o3
2,1	2,2	0.1	o5

$O(o3) = 0.15 + 0.1$
$= 0.25$
$O(o5) = 0.15 + 0.1$
$= 0.25$
$O(o6) = 0.15 + 0.15 + 0.1 + 0.1$
$= 0.5$
Thus o6 is the most likely.

5. Question 5

We have the number of nodes possible =

$$N = \sum^{T-1} |O|^i$$
$$= (|O|^T - 1) / (|O| - 1)$$
$$= (6^T - 1) / (5)$$

For the number of policy tree, we get -

$$Trees = |A|^N$$
$$= 5^N$$

Depending upon the horizon(T) that we choose to stop the POMDP on, we can get a different number of trees. For example,

if T = 1,

Trees = 5

If T = 2,

Trees = 7

and so on