Machine, Data, and Learning

Assignment - 3
POMDP

Roll Nos.: 2019101086 and 2019101105

Introduction

Cells and states in POMDP

(1, 0)	(1, 1)	(1, 2)	(1, 3)
(0, 0)	(0, 1)	(0, 2)	(0, 3)

Total number of states : 128 (Agent position (2*4) x Target position (2*4) x Call (True/ False))
Each state is mapped to a number:

$$
\begin{aligned}
& Call \\
+ \quad & Target_{position\_column} * 2 \\
+ \quad & Target_{position\_row} * 8 \\
+ \quad & Agent_{position\_column} * 16 \\
+ \quad & Agent_{position\_row} * 64
\end{aligned}
$$

Q1:

Given Target is in (1,0) and Observation is o6. The target is not in the one-cell neighborhood of the agent.
Possible positions for Agent are (0,1) ,(1,2) , (0,2) ,(1,3) ,(0,3)
Number of possible states = (Number of positions of target = 1) * (Number of positions of agent = 5) * (Number of values of call = 2 (T/ F)) = 10

All the ten states are equiprobable.
Initial Belief state:

States
((0,1) , (1,0) , 1),

((0,1) , (1,0) , 0),
((0,2) , (1,0) , 1),
((0,2) , (1,0) , 0),
((0,3) , (1,0) , 1),
((0,3) , (1,0) , 0),
((1,2) , (1,0) , 1),
((1,2) , (1,0) , 0),
((1,3) , (1,0) , 1),
((1,3) , (1,0) , 1)
have belief value 0.1 and the remaining states have belief value 0.

Q2:

Given Agent is in $(1, 1)$ and Target is in the one-cell neighborhood, i.e., four possible states for the neighborhood are $(1, 1), (1, 0), (0, 1), (1, 2)$, and given target in not making a call, so the number of possible states is 1*4*1 = 4.

The states

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

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

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

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

Have value 0.25. Other states will have a belief value as 0

Q3:

Expectations were calculated using **pomdsim** command with arguments **--simLen 100 and --simNum 1000** using the generated .pomdp and .policy files.
The expected value for Q1: **14.2337**
The expected value for Q2: **28.3098**

Q1:

```
~/I/S/M/A/part2 ❯❯❯ ../sarsop/src/pomdpsim --simLen 100 --simNum 1000 --policy-file out.policy 20191010186_2019101105.pomdp

Loading the model ...
  input file   : 20191010186_2019101105.pomdp

Loading the policy ...
  input file   : out.policy

Simulating ...
  action selection :  one-step look ahead

----------------------------------
 #Simulations  | Exp Total Reward
----------------------------------
 100            13.8989
 200            14.1858
 300            14.6717
 400            14.4178
 500            14.3882
 600            14.2027
 700            14.366
 800            14.3045
 900            14.3179
 1000           14.2337
----------------------------------

Finishing ...

------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
------------------------------------------------------------
 1000           14.2337           (13.5295, 14.9378)
------------------------------------------------------------
```

Q2:

```
~/I/S/M/A/part2 ❯❯❯ ../sarsop/src/pomdpsim --simLen 100 --simNum 1000 --policy-file out.policy q2.pomdp

Loading the model ...
  input file   : q2.pomdp

Loading the policy ...
  input file   : out.policy

Simulating ...
  action selection :  one-step look ahead

----------------------------------
 #Simulations  | Exp Total Reward
----------------------------------
 100            28.7763
 200            29.0511
 300            28.8957
 400            28.3778
 500            28.4331
 600            28.5954
 700            28.655
 800            28.42
 900            28.2621
 1000           28.3098
----------------------------------

Finishing ...

------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
------------------------------------------------------------
 1000           28.3098           (27.6177, 29.0018)
------------------------------------------------------------
```

Q4:

- The Agent is in (0,0) with a probability of 0.4
    - The Target is in (0,1) with a probability of 0.25
        - The observation is o2 with a probability of 1.0
        - This probability of o2, in this case, is 0.25 * 0.4 = 0.1
    - The Target is in (0,2) with a probability of 0.25
        - The observation is o6 with a probability of 1.0
        - This probability of o6, in this case, is 0.25 * 0.4 = 0.1
    - The Target is in (1,1) with a probability of 0.25
        - The observation is o6 with a probability of 1.0
        - This probability of o6, in this case, is 0.25 * 0.4 = 0.1
    - The Target is in (1,2) with a probability of 0.25
        - The observation is o6 with a probability of 1.0
        - This probability of o6, in this case, is 0.25 * 0.4 = 0.1

- The Agent is in (1,3) with a probability of 0.6
    - The Target is in (0,1) with a probability of 0.25
        - The observation is o6 with a probability of 1.0
        - This probability of o6, in this case, is 0.25 * 0.6 = 0.15
    - The Target is in (0,2) with a probability of 0.25
        - The observation is o6 with a probability of 1.0
        - This probability of o6, in this case, is 0.25 * 0.6 = 0.15
    - The Target is in (1,1) with a probability of 0.25
        - The observation is o6 with a probability of 1.0
        - This probability of o6, in this case, is 0.25 * 0.6 = 0.15
    - The Target is in (1,2) with a probability of 0.25
        - The observation is o4 with a probability of 1.0
        - This probability of o4, in this case, is 0.25 * 0.6 = 0.15

o6 : 0.1 * 3 + 0.15 * 3 = 0.75
o4: 0.15
o2: 0.1
Observation o6 is most likely to be observed as it has the highest probability.

Q5:

```
~/I/S/M/A/part2 ❯❯❯ ../sarsop/src/pomdpsol q5.pomdp

Loading the model ...
  input file   : q5.pomdp
  loading time : 0.20s

SARSOP initializing ...
  initialization time : 0.00s

--------------------------------------------------------------------------------
 Time   |#Trial |#Backup |LBound    |UBound    |Precision  |#Alphas |#Beliefs
--------------------------------------------------------------------------------
 0      0       0        11.13      24.06      12.93       5        1
 0.01   11      51       20.0097    20.1024    0.0927502   30       13
 0.01   18      103      20.0695    20.093     0.0234803   52       22
 0.02   23      151      20.0816    20.0912    0.00954044  65       31
 0.02   28      201      20.0855    20.0897    0.00420003  84       42
 0.03   32      250      20.0869    20.0892    0.0022028   93       57
 0.04   35      300      20.0874    20.089     0.00160091  109      69
 0.06   38      350      20.0878    20.0888    0.00104404  129      84
 0.06   40      375      20.0879    20.0888    0.000879535 148      91
--------------------------------------------------------------------------------

SARSOP finishing ...
  target precision reached
  target precision  : 0.001000
  precision reached : 0.000880

--------------------------------------------------------------------------------
 Time   |#Trial |#Backup |LBound    |UBound    |Precision  |#Alphas |#Beliefs
--------------------------------------------------------------------------------
 0.06   40      375      20.0879    20.0888    0.000879535 148      91
--------------------------------------------------------------------------------

Writing out policy ...
  output file : out.policy
```

P : Number of policy trees

$$P = |A|^N$$

N : Number of nodes in the tree

$|A|$: Number of Actions = 5

$|O|$: Number of observations = 6

T : Horizon of the POMDP = 40 from #Trial

$$N = \sum_{i=0}^{T-1} |O|^i \;=\; (|O|^T - 1)/\,(|O| - 1)$$

$$N \;=\; (6^{40} - 1)/\,(6 - 1) \;=\; 2.673498907768746813567769195315 \times 10^{30}$$

$$P \;= 5^{(6^{40}-1)/\,5}$$