Assignment5 - Part2
2018101075

Value of x = 1 - (075 % 40 + 1) / 100 = 0.64

Following grid is used for this assignment

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

Discount Factor = 0.5

Note: In the assignment, After reaching state (x, x, 1), I have considered ‘call’ off before taking the next step.

1. Each state of the POMDP is represented as a tuple (Agent Position, Target Position, Call). Here agent position and target position are tuples of position in the grid. Target is in cell (1, 1) and observation is o6 which means states can be

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

with equal probability.
Therefore, the probability of the above states is 1 / 8 = 0.125 and 0 for others.

Belief state = [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, 0.125, 0.125, 0, 0, 0, 0, 0, 0, 0, 0]

where state (i, j, k) is mapped to whole number by function
f((i, j, k)) = k + 2*j[1] + 2*3*j[0] + 2*3*3*i[1] + 2*3*3*3*i[0].

2. Agent is in (0, 1) and the target is in the neighborhood and is not making a call. Therefore, initial state can be

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

with equal probability.
Therefore, the probability of the above states is 1 / 4 = 0.25 and 0 for others.

Belief state = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.25,
0, 0.25, 0, 0.25, 0, 0, 0, 0.25, 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, 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]

where state (i, j, k) is mapped to whole number by function
f((i, j, k)) = k + 2*j[1] + 2*3*j[0] + 2*3*3*i[1] + 2*3*3*3*i[0].

3. For question1,

    Expected Reward / Expected utility = 11.8529

   For question2,

    Expected Reward / Expected utility = 25.8466

for simLen = 100 and simNum = 100

4. If agent is at (0, 1) with probability 0.6, then probabilities of observations are:

   1. o3 = 0.6 * 0.25 as target can be at (0, 0) with probability 0.25
   2. o5 = 0.6 * 0.25 as target can be at (0, 2) with probability 0.25
   3. o6 = 0.6 * 0.5 as target can be at (2, 0) or (2, 2) with probability 0.25 + 0.25 = 0.5
   4. 0 otherwise

If agent is at (2, 1) with probability 0.4, then probabilities of observation are:

   1. o3 = 0.4 * 0.25 as target can be at (2, 0) with probability 0.25
   2. o5 = 0.4 * 0.25 as target can be at (2, 2) with probability 0.25

3. o6 = 0.4 * 0.5 as target can be at (0, 2) or (0, 0) with probability 0.25 + 0.25 = 0.5.
4. 0 otherwise

Therefore final probability for each observation is:

1. o3 = 0.25
2. o5 = 0.25
3. o6 = 0.5
4. 0 otherwise

Therefore, observation 6 is most likely to be observed.

5. Number of Policy Tree is calculated using the following formula:

Total number of nodes in a policy tree is given by:

$$N = \sum_{\{i=0 \text{ to } T-1\}} |O|^i = (|O|^T - 1) \div (|O| - 1)$$

here,

N = number of nodes in tree
T = time horizon of PoMDP (height of tree)
$|O|$ = number of observations = 6
$|A|$ = number of actions = 5

So, number of policy trees possible would include every action possible in every node, i.e.
Number of trees = $|A|^N$

Number of trees is dependent on the horizon T. The value of P increases with increase in the value of T.