Assignment 5 Part-2

Note: I have assumed that when the agent and target move to the same position, call is turned off before taking the next step. So (x,x,1) -> (x,x,0) has probability 0.6 instead of 1.

Grid Used:

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

x increases rightwards and y increases upwards.

Q1 .If you know the target is in (1,1) cell and your observation is o6 , what will be the initial belief state? Please submit the optimal policy file named <RollNumber>.policy for the POMDP taking into account the initial belief state you obtained.

For this question the belief state is:
Target = (1,1)
Agent = (2,0), (0,2), (0,0), (2,2)
Call = on or off
Therefore the probability is divided equally among 8 states(states are of form (agent, target,call)):
((0,0), (1,1), 0) # states are of form (agent, target, call)
((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 each of the above state getting initial value to be 0.125 and rest 0.

Q2. 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?

Here we have 4 equally likely states
Agent = (0,1)
Target = (1,1), (0,0), (0,1), (0,2)
Call = off
Therefore the probability is divided equally among 8 states(states are of form (agent, target,call)):
((0,1), (0,0), 0)
((0,1), (0,1), 0)

((0,1), (1,1), 0)
((0,1), (0,2), 0)
With each of the above state getting initial value to be 0.25 and rest 0.

Note: Assumed that (0,1) is considered to be part of one neighbourhood.

Q3. What is the expected utility for initial belief states in questions 1 and 2?

Using ./pomdpsim --simLen 100 --simNum 100 --policy-file out.policy 2018101015.pomdp to calculate expected utility

For Question 1

Simulations	Exp. Total Reward	95% Confidence Interval
100	3.5669	(3.01582, 4.11798)

For Question 2

Simulations	Exp. Total Reward	95% Confidence Interval
100	7.67101	(7.06869, 8.27334)

Q4. 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.

Let X : Agent is in (0,1) and Y: Agent is in (2,1).

P(X) = 0.6
P(o6 | X) = 0.5 # two corner cells on the far side
P(o3 | X) = 0.25 # below the cell
P(o5 | X) = 0.25 # above the cell

P(Y) = 0.4
P(o6 | Y) = 0.5 # two corner cells on the far side
P(o3 | Y) = 0.25 # below the cell
P(o5 | Y) = 0.25 # above the cell

P(o6) = P(X)P(o6|X) + P(Y)P(o6|Y) = 0.5
P(o3) = P(X)P(o3|X) + P(Y)P(o3|Y) = 0.25
P(o5) = P(X)P(o5|X) + P(Y)P(o5|Y) = 0.25

Probability to observe o6 is maximum.

Q5. How many policy trees are obtained in this case, explain?

Total number of nodes in a policy tree is given by

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

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

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

Using T as #Trial when running pomdpsol in Q1 which is 24.

Number of trees = $5^{(6^{24}-1)/5}$