REPORT

Assignment 5 Part 2

Shreeya Pahune

2018113011

POMDP

Each state of the POMDP is represented as a tuple (Agent Position, Target Position, Call) Where Agent Position is given by (x_1, y_1) where $0 \leq x_1 \leq 2$ and $0 \leq y_1 \leq 2$, Target Position is given by (x_2, y_2) where $0 \leq x_2 \leq 2$ and $0 \leq y_2 \leq 2$ and Call is denoted by a variable c which is 1 if the call is on else 0

We follow the following grid for finding coordinates of agent and target

0 , 0	0 , 1	0 , 2
1 , 0	1 , 1	1 , 2
2 , 0	2 , 1	2 , 2

Therefore there are a total of 162 states where each state is given as $\{ x_1, y_1, x_2, y_2, c\}$

Moreover there are 6 observations given:

o1 is observed when the target is in the same cell as the agent.

o2 is observed when the target is in the cell to the right of the agent's cell.

o3 is observed when the target is in the cell below agent's cell

o4 is observed when the target is in the cell to the left of agent's cell

o5 is observed when the target is in the cell above the agent's cell.

o6 is observed when the target is not in the 1 cell neighbourhood of the agent.

X = 1 - (((LastThreeDigitsOfRollNumber)%40 + 1) / 100) = 1 - (011%40 + 1)/100 = 0.88

Reward for reaching the target before call is turned off = (<RollNumber>%100 + 10) = 2018113011%100 + 10 = 21

Question 1

Given: Target position = (1 , 1) , observation = o6

Since the Agent observes o6 it means that the target is not present in the 1 cell neighbourhood, there are 4 possible position of the agent in the grid as shown below:

A		A
	T	
A		A

We don't know if the call is on or off. Hence there are a total of 4 * 2 = 8 out of 162 states where the agent is present initially. The agent has equal probability of being in all of these 8 states therefore the initial belief state is given by:

$$b[s] = \frac{1}{8} \text{ if } x_1 \in \{0, 2\} \text{ and } y_1 \in \{0, 2\} \text{ and } (x_2, y_2) = (1, 1) \text{ and } c \in \{0, 1\}$$
$$0 \text{ otherwise}$$

Question 2

Given: Agent Position = (0 , 1) , call = 0, observation != o6

We have a fixed location for the agent. The target has to be in the state such that the observation is not o6 so Target can be in one of the given four states

T	A(T)	T
	T	

The target can be present in the states (0, 0), (0 , 1), (0, 2) and (1, 1) with equal probability of 0.25 (¼)

Therefore the initial belief state will have 0.25 for the states {0,1,x_2, y_2, 0} where (x_2 , y_2) is one of { (0, 0), (0 , 1), (0, 2) , (1, 1) }

Question 3

The expected utility is calculated using:

$$r(b) = \sum_{s \in S} b(s) * R(s)$$

Question 1

Since the action and target are never in the same cell the reward is always -1

r(b) = 8 * (⅛) * (-1) = -1 (8 states will have their belief as ⅛ and the rest will be 0 as given in question 1)

Expected Utility = -1

Question 2

Since the call is turned off the reward will be always -1 again even though there is one state where agent and target are present in the same cell.

Therefore r(b) = 4 * (¼) * (-1) = -1 (4 states will have their belief as 1/4 and the rest will be 0 as given in question 2)

Expected Utility = -1

Question 4

T	A(0.6)	T
T	A(0.4)	T

Target is present in the for corners with equal probability which is 1/4

The agent is present in the cell (0,1) with 0.6 probability and (2,1) with 0.4 probability.

It will observe O2 when in (0,1) with ¼ probability and when in (2,1) with ¼ probability. Therefore O2 will be observed with a probability 0.6 * 0.25 + 0.4 * 0.25 = 0.25 times

Similarly It will observe O4 when in (0,1) with ¼ probability and when in (2,1) with ¼ probability. Therefore O4 will be observed with a probability 0.6 * 0.25 + 0.4 * 0.25 = 0.25 times

It will observe O6 when in (0,1) with a probability of ½ (when the T is in the two corners of x = 2 row) and when in (2,1) with a probability of ½ (when T is in the two corners of x = 0 row) So O6 will be observed with a probability 0.6*0.5+0.4*0.5 = 0.5

Therefore the most common observation is o6

Question 5

The number of trees is given by $|A|^N$

Where |A| = number of actions and N = $\sum_{i=0}^{T-1} \frac{|O|^i - 1}{|O| - 1}$ where T is the horizon, O is the number of observations.

As we can see the number of trees increases as the horizon increases and since the horizon is not a fixed value in SARSOP as it depends on the precision value obtained.Hence, the number of policy trees obtained is not a fixed value.