Assignment 3 - Part 2

Aakash Aanegola - 2019101009

Aakash Jain - 2019101028

For the purpose of this assignment, the following notation was used

$Sr_ac_ar_tc_tx$ where

- r_a refers to the row that the agent is in
- c_a refers to the column that the agent is in
- r_t refers to the row that the target is in
- c_t refers to the column that the target is in
- x refers to the call status (0 indicates off and 1 indicates on)

The indexing starts from the top left corner, and moving down increases the row count by 1 and moving right increases the column count by 1.

Q1. The initial belief state is composed of 128 entries as that is the number of states, and hence we only indicated the probabilities of the states that are non-zero.

$S01100, S02100, S03100, S12100, S13100$
$S01101, S02101, S03101, S12101, S13101$

The above listed states are the possible states and each of them possess a probability of 0.1 in the belief state (assuming an equi-probable distribution). The optimal policy file has been attached.

Q2. In a similar fashion to the above belief state,

$S11110, S11100, S11010, S11120$ are the possible states, and assuming that they are equi-probable each of them will have a probability of 0.25.

Q3. To obtain the expected utility we used the `pomdpeval` tool which runs a simulation to obtain the results. The `pomdpeval` tool returns the expected total reward, which is equivalent to the expected utility.

```
-------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
-------------------------------------------------------------
 100           13.4664            (11.7075, 15.2253)
-------------------------------------------------------------
```

From the above image obtained after running `pomdpeval` we can see that the expected total reward is 13.47 and hence the expected utility is the same. However they have also given us a 95% confidence interval which indicates that this tool has some error in calculation, and the expected utility will be within the given interval 95% of the time.

```
-------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
-------------------------------------------------------------
 100           27.7944            (26.755, 28.8339)
-------------------------------------------------------------
```

Using the same method above we can see both the expected total reward (expected utility) and the confidence interval for the same (expected reward = 27.79).

Q4. The probability distribution is as follows:

- $O1$: 0.0
- $O2$: 0.1
- $O3$: 0.0
- $O4$: 0.15
- $O5$: 0.0
- $O6$: 0.75

This is because if the agent is in the position $(0, 0)$ then it will observe $O2$ only if the target is in $(0, 1)$ and will observe $O6$ otherwise. If the agent is in the position $(1, 3)$ then it will observe $O4$ only if the target is in the position $(1, 2)$ and will observer $O6$ otherwise.

Q5. On creating the policy for the given belief state, the following was observed

```
-------------------------------------------------------------------------------
 Time   |#Trial |#Backup |LBound    |UBound    |Precision  |#Alphas |#Beliefs
-------------------------------------------------------------------------------
 0.02   15      81       5.26253    5.26351    0.000973218 40       18
-------------------------------------------------------------------------------
```

The number of trials indicates the depth of the tree and can be used as a replacement for the time horizon, and if we use the formula to calculate the number of trees,

$|A| = 5$

$|O| = 6$

$|T| = 15$

And hence the number of policy trees is A^N where $N = \frac{|O|^T - 1}{|O| - 1}$.

Calculating this number, we get $5^{94036996915}$ which is a very large number.