MACHINE DATA AND LEARNING

KUSHAGRA AGARWAL 2018113012
ASSIGNMENT 5 PART 2 REPORT

POMDP FORMULATION

Entities involved are **Agent, and Target**. The coordinates of both the agent and the target along with the call being on/off decide the state for the given POMDP.

The states therefore are represented by a tuple:

(Agent Coordinates, Target Coordinates, Call Value)

Both the Agent and the Target are confined to move inside a 3*3 grid, hence their coordinates can be in the following range:

0<= Ax, Tx <=2; 0<= Ay, Ty <=2

Here **Ax** is the x coordinate of the agent, **Ay** is the y coordinate for the agent.

Also **Tx** is the x coordinate of the target and **Ty** is the y coordinate of the target.

Also the call can be either 0 or 1 (with 1 denoting on). Therefore:

S = ({Ax,Ay}, {Tx,Ty}, call)

The grid with the corresponding values of x and y shown:

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

The total number of states possible are (3*3) * (3*3) * (2) = **162**

The actions allowed are : ***{Stay, Up, Left, RIght, Down}***

The value of x is 1- ((2018113012 % 40) + 1) / 100 = 1- 13 /100 = **0.87**

The Observations allowed are: ***{ o1, o2, o3, o4, o5, o6}***

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.

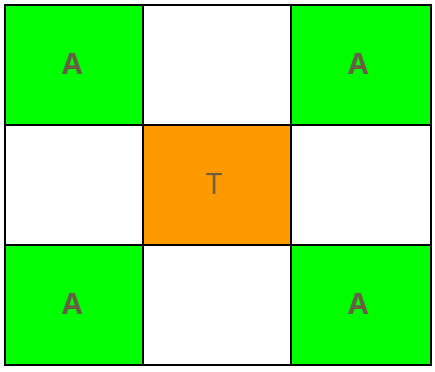
o6 is observed when the target is not in the 1 cell neighbourhood of the agent.

The Rewards are as following:

a) The step cost is -1
b) The reward on reaching the same cell as the target while the call is on is (2018113012%100)+10 = **22**

Q1) If you know the target is in (1,1) cell and your observation is o6 , what will be the initial belief state?

Sol) The target is in (1,1) and for o6 to occur, the agent and the target must not be in the 1 cell neighbourhood of each other. Therefore the possible positions for the agent are:



Now these correspond to the coordinates: (0,0), (2,0), (0,2), (2,2)

But the call values can be 0 or 1. Hence the possible states which satisfy this condition are:

1) ({0,0}, {1,1}, 0)
2) ({0,0}, {1,1}, 1)
3) ({0,2}, {1,1}, 0)
4) ({0,2}, {1,1}, 1)
5) ({2,0}, {1,1}, 0)
6) ({2,0}, {1,1}, 1)
7) ({2,2}, {1,1}, 0)
8) ({2,2}, {1,1}, 1)

All of these states are equally probable, hence we assign them the same probabilities= 1/ (N) where N defines the total number of such equally probable states. Here **N=8**

Hence the start state or the initial belief state contains of the 8 states previously listed each with a probability of ⅛= 0.125

Hence the initial belief vector is as follows:

b[s] = { 0.125 for the 8 states

0 for the rest }

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?

Sol) The agent is in (0,1) and the target is in 1 neighbourhood and the call is off: therefore the possible states for the target are: (0,2), (1,1), (0,0) and (0,1) itself.

T	A, T	T
	T	

That gives us 4 possible locations for the target to be in, but this time we know that the call is off. Hence the number of states is also 4, namely:

1) ({0,1}, {0,0}, 0)
2) ({0,1}, {0,1}, 0)
3) ({0,1}, {1,1}, 0)
4) ({0,1}, {0,2}, 0)

All of these states are equally probable, hence we assign them the same probabilities= 1/ (N) where N defines the total number of such equally probable states. Here **N= 4**

Hence the start state or the initial belief state contains of the 8 states previously listed each with a probability of ¼ = 0.25

Hence the initial belief vector is as follows:

b[s] = { 0.25 for the 4 states

0 for the rest

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

Sol) 1) Expected Total Reward = 2.74268 after 1000 iterations

```
Loading the model ...
  input file   : ../../pomdps/kushagraQ1.pomdp

Loading the policy ...
  input file   : ../../policies/Kush1.policy

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

-----------------------------------
 #Simulations   | Exp Total Reward
-----------------------------------
 100            2.70288
 200            2.79868
 300            2.91059
 400            2.84032
 500            2.82453
 600            2.77595
 700            2.70162
 800            2.72497
 900            2.76642
 1000           2.74268
-----------------------------------

Finishing ...

-------------------------------------------------------------
 #Simulations   | Exp Total Reward | 95% Confidence Interval
-------------------------------------------------------------
 1000           2.74268            (2.58984, 2.89552)
-------------------------------------------------------------
```

2) Expected Total Reward = 6.98522 after 1000 iterations

```
Loading the model ...
  input file   : ../../pomdps/kushagraQ2.pomdp

Loading the policy ...
  input file   : ../../policies/Kush2.policy

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

----------------------------------
 #Simulations  | Exp Total Reward
----------------------------------
 100           7.69835
 200           7.47684
 300           7.21995
 400           7.11308
 500           7.10088
 600           7.10285
 700           7.06472
 800           7.03148
 900           6.99588
 1000          6.98522
----------------------------------

Finishing ...

------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
------------------------------------------------------------
 1000          6.98522            (6.82662, 7.14383)
------------------------------------------------------------
```

The expected utility is:

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

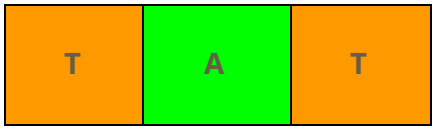
Expectations were calculated by using the following command:

```
./pomdpsim --simLen 100 --simNum 1000 --policy-file <policyfile>.policy <pomdpfile>.pomdp
```

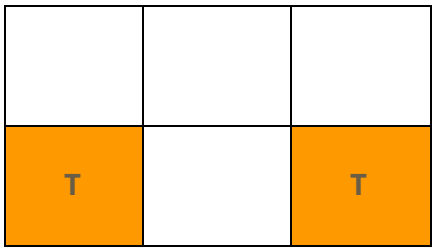
Q4) If your agent is in (0,1) with probability 0.6 and in (2,1) with probability 0.4 and the target is in the 4 corner cells with equal probability, which observation are you most likely to observe? Explain.

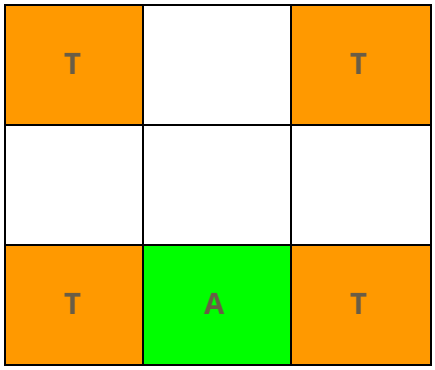
Sol) The agent is in (0,1) or (2,1) while the target is in (0,0) or (0,2), (2,0), (2,2) with equal probability. The two possibilities therefore are:

1) With probability = 0.6



2) With probability = 0.4





Calculating the probability of each observation state is as follows:

	X {Agent (0,1)}	Y {Agent(2,1)}	0.6*X + 0.4*Y
o1	0	0	0
o2	1/4	1/4	1/4
o3	0	0	0
o4	1/4	1/4	1/4
o5	0	0	0
o6	2 * ¼ = 1/2	2* ¼ = 1/2	1/2

o2 is observed when the agent is in (0,1) and the target is in (1,1) with prob =0.25

And when the agent is in (2,1) and target is in (2,2) with prob=0.25

Total prob= 0.6*0.25 + 0.4*0.25 = 0.25

Similar analysis holds for o4

o6 is observed when the agent is in (0,1) and the target is in (2,0) with prob 0.25 or (2,2) with prob 0.25. Therefore total = 0.5

Similarly total =0.5 when agent is in (2,1).

Hence for o6 Total = 0.6*0.5 + 0.4*0.5 = 0.5

Therefore **the most probable observation is o6**

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

Sol) The number of trees = $|A|^N$

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

The number of trees increases as the horizon increases and since the horizon is not a fixed value in SARSOP (depends on the precision value) Therefore the number of policy trees is not fixed.

Now Assuming Trials to be the horizon we get an approximate number of policy trees for the Q1 policy

A = 5, O= 6, T= trials. The value of trials as obtained from the image below is 169.

Therefore:

N = (6 ^ 169 - 1) / (6-1) = 6^169/5 (approx) **=** $\mathbf{6*10^{130}}$

So, the **Number of trees** $\mathbf{= 5^{(6*10\text{^}130)}}$

Time	#Trial	#Backup	LBound	UBound	Precision	#Alphas	#Beliefs
0	0	0	0.805922	4.31555	3.50963	5	1
0.01	11	51	2.73823	2.8692	0.130973	47	19
0.01	19	100	2.76952	2.85988	0.0903669	65	31
0.02	25	150	2.78851	2.82067	0.0321612	86	40
0.03	30	200	2.79904	2.81941	0.020368	94	53
0.05	35	251	2.80393	2.81814	0.0142116	120	60
0.06	40	301	2.80464	2.81719	0.0125524	138	65
0.08	45	351	2.80467	2.81677	0.0121028	166	70
0.09	50	401	2.80468	2.8155	0.010817	200	75
0.12	54	450	2.80551	2.8147	0.00919123	240	86
0.15	58	500	2.80617	2.81419	0.0080158	265	98
0.19	63	553	2.8067	2.81401	0.00730642	289	111
0.21	68	601	2.80962	2.81367	0.00405063	261	115
0.25	72	651	2.80984	2.81345	0.00360502	281	124
0.28	76	701	2.80985	2.81322	0.00337497	280	128
0.31	80	750	2.80985	2.8132	0.00334492	295	132
0.35	84	805	2.8102	2.81318	0.00298073	314	139
0.38	88	855	2.8102	2.81316	0.00295986	333	143
0.42	92	901	2.81033	2.81311	0.00278002	324	150
0.44	96	950	2.81045	2.81306	0.00260841	334	154
0.48	100	1000	2.81045	2.81301	0.00255984	319	158
0.52	104	1050	2.81045	2.81297	0.00252136	333	163
0.56	109	1105	2.81045	2.81295	0.00249937	312	167
0.6	113	1153	2.81045	2.81272	0.00227312	326	172
0.67	116	1200	2.81054	2.81264	0.00210279	356	198
0.72	120	1251	2.81062	2.81243	0.00181466	384	207
0.77	123	1300	2.8107	2.81243	0.00173067	421	212
0.82	127	1351	2.81072	2.81239	0.00166871	456	216
0.88	131	1401	2.81089	2.81233	0.00144801	495	219
0.96	135	1450	2.81092	2.81232	0.00139482	522	237
1.05	138	1500	2.81102	2.81229	0.00126333	553	254
1.13	142	1555	2.81108	2.81225	0.00117681	571	264
1.18	145	1603	2.81109	2.81224	0.00115535	593	267
1.27	148	1650	2.81109	2.81223	0.00113804	640	276
1.34	152	1703	2.81109	2.8122	0.00110683	669	280
1.4	155	1750	2.81112	2.81219	0.00106865	696	285
1.5	159	1807	2.81112	2.81217	0.00105721	702	291
1.57	162	1851	2.81112	2.81217	0.00104955	746	295
1.67	165	1900	2.81113	2.81215	0.00102359	740	311
1.81	169	1953	2.81113	2.81213	0.000997796	777	327
1.81	169	1953	2.81113	2.81213	0.000997796	777	327