MDL Assignment 5-part 2

Tanvi Karandikar
2018101059

Mapping the grid to states 0 to 8

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

Mapping the actions:
0 ->stay
1-> up
2-> down
3->left
4->right

Mapping the observations: 0 to 5
i->o[i+1]

Mapping the states: 9*9*2=162 possible states
(cell of agent, cell of target, cell state (0 if off, 1 if on))
(a,t,c)-> a*18 + t*2 + c

NOTE: rule followed is, if agent,target in same state with call on, then it is instantaneously changed to off before taking next action. ie in this case new call probabilites are 0.6 to turn off, 0.4 to remain on

Question 1

Target is in (1,1) ie 4.
o6 means agent is in 2,8,0,6.
also call can be 1 or 0.

Hence the possible starting states will be
(0,4,0); (0,4,1); (2,4,0); (2,4,1); (6,4,0); (6,4,1); (8,4,0); (8,4,1)

Initial belief state will have all of these with same probability ie ⅛.
Rest all states will have inital belief state 0.

Policy file is attached. Inital beliefs have been taken into account by mapping above states to single integer representation.
They are specified by including the line:
start include: 8 9 44 45 116 117 152 153

Question 2

Agent is in (0,1) ie 1
One neighbourhood means within distance 1.
So, target is at cells 0,1,3,4. [NOTE: here we include 4 in the 1-neighbourhood]
Given call=0

Initial belief state will have all of these with same probability ie ¼.
Rest all states will have inital belief state 0.

Inital beliefs have been taken into account by mapping above states to single integer representation.
They are specified by including the line
So, possible states are (1,2,0), (1,4,0), (1,0,0), (1,1,0)
start include: 22 26 18 20

Question 3

Expectations were calculated by using the **--simLen 100 --simNum 1000 --policy-file** flag with **pomdpsim** program, and output file from **pomdpsol.**
Expected value for q1: 11.9014
Expected value for q2: 22.0953
Image of each output is below.

q1 pomdpsim output:

```
Loading the model ...
  input file   : ./q1.pomdp

Loading the policy ...
  input file   : ./out.policy

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

-----------------------------------
 #Simulations  | Exp Total Reward
-----------------------------------
 100           12.3406
 200           11.8009
 300           12.274
 400           12.1521
 500           12.0815
 600           12.0647
 700           11.9341
 800           11.982
 900           11.9139
 1000          11.9014
-----------------------------------

Finishing ...

-------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
-------------------------------------------------------------
 1000          11.9014            (11.4581, 12.3447)
-------------------------------------------------------------
```

q2 pomdpsim output:

```
Loading the model ...
  input file   : ./q2.pomdp

Loading the policy ...
  input file   : ./out.policy

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

----------------------------------
 #Simulations  | Exp Total Reward
----------------------------------
 100           21.2805
 200           22.2951
 300           22.3548
 400           22.3961
 500           22.2384
 600           22.1563
 700           22.0848
 800           22.1625
 900           22.0638
 1000          22.0953
----------------------------------

Finishing ...

-------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
-------------------------------------------------------------
 1000          22.0953           (21.6262, 22.5644)
-------------------------------------------------------------
```

Question 4

The agent has two possibilites: 1 and 7
Target has 4 possibilities: 0,2,6,8
Call has two: 0,1

Total possible states=2*4*2=16
We know what each observation means, so we can check for each of the possibilites

state,probability,observation
(1 0 0) 0.075 o3
(1 0 1) 0.075 o3
(1 2 0) 0.075 o5
(1 2 1) 0.075 o5
(1 6 0) 0.075 o6
(1 6 1) 0.075 o6
(1 8 0) 0.075 o6
(1 8 1) 0.075 o6
(7 0 0) 0.05 o6
(7 0 1) 0.05 o6
(7 2 0) 0.05 o6
(7 2 1) 0.05 o6
(7 6 0) 0.05 o3
(7 6 1) 0.05 o3
(7 8 0) 0.05 o5
(7 8 1) 0.05 o5

o3 -> 4 times 0.075*2 + 0.05*2=0.25 probability
o5 -> 4 times 0.075*2 + 0.05*2=0.25 probability
o6 -> 8 times 0.075*4 + 0.05*4=0.5 probablility

Clearly, we are most likely to observe observation **o6** since it has highest probability.

Question 5

On running **pomdpsol**

Time	#Trial	#Backup	LBound	UBound	Precision	#Alphas	#Beliefs
3.04	165	2163	14.7999	14.8009	0.000990751	1201	459

We will use the #Trial as T value for calculation.

The formula used in calculation is

How many policy trees, if |A| actions, |O| observations, T horizon:

- How many nodes in a tree:

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

How many trees:

$$|A|^N$$

Here |A|=5, |O|=6, T=165.

Thus, N=(6^165 -1)/(6-1)= approx 10^128

Now, |A|^N=5^(10^128) is the approximate number of policy trees obtained, which is a very large amount