ASSIGNMENT 5(B)

MACHINE, DATA, AND LEARNING

MAY 2020

SUBMITTED BY

YASH BHANSALI
2018101068

The Grid is being represented in the following way.

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

Note: For the case where the agent reaches the target with the call ON, the call is turned OFF immediately, which is an intermediate step. For the next step, the call would be turned on with probability 0.4

1

Location of target : (1,1)
Observation: o6.

Possible location where the agent can be : [(0,0), (0,2), (2,0), (2,2)].

	A	
A	T	A
	A	

The possible states are :

[**((0,0), (1,1), 0),** **((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)**]

All these states are equally probable and the probability of all other states is 0.

Hence, the initial belief state can be given by the function:

b(s) ={⅛ if agent_position in [(0, 0), (0, 2), (2, 0), (2, 2)] and target_position == (1, 1)}

Otherwise b(s) = 0

Hence, for any state s where agent_position is one of [(0, 0), (0, 2), (2, 0), (2, 2)], and target_position is (1, 1) and call_active is either ON or OFF. This gives us 8 states, whose belief value is 1/8 and the belief value for the rest of the states is 0.

2

Possible cells for the target are (0,0); (0,1); (1,1) and (0,2). Each of the following states are equally likely:

[**((0,1), (0,0), 0),** **((0,1), (0,1), 0),** **((0,1), (1,1), 0),** **((0,1), (0,2), 0)**]

These states will have initial probability 1/4 and the rest of them have probability 0.

Hence, the initial belief state can be given by the function:

b(s) ={¼ if target_position in [(0, 0), (0, 1), (1, 1), (0, 2)] and agent_position == (0, 1)}

Otherwise b(s) = 0

Hence, for any state s where target_position is one of [(0, 0), (0, 1), (1, 1), (0, 2)], and agent_position is (0, 1) and call_active is OFF. This gives us 4 states, whose belief value is ¼ and the belief value for the rest of the states is 0.

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

3

Calculation for Q1

Simulations	Exp. Total Reward	95% Confidence Interval
100	**15.4316**	**(13.8573, 17.0059)**

Calculation for Q2

Simulations	Exp. Total Reward	95% Confidence Interval
100	**26.4793**	**(25.1033, 27.8553)**

4

- Agent is in (0,1) with probability 0.6 and in (2,1) with probability 0.4
- Target is in one of the 4 corner cells with equal probability i.e target is in one cell, out of [(0,0), (0,2), (2,0), (2,2)], with probability 0.25.

The list of all possible cases with the corresponding observation is as follows:

Target	**(0,1) [Pr=0.6] Agent**	**(2,1) [Pr=0.4] Agent**
(0,0) [P=0.25]	**o3**	**o6**
(0,2) [P=0.25]	**o5**	**o6**
(2,0) [P=0.25]	**o6**	**o3**
(2,2) [P=0.25]	**o6**	**o5**

The corresponding probabilities of each observation are as follows:

Target	**(0,1) [Pr=0.6]** **Agent**	**(2,1) [Pr=0.4]** **Agent**
(0,0) [Pr=0.25]	0.6*0.25=0.15	0.4*0.25=0.10
(0,2) [Pr=0.25]	0.6*0.25=0.15	0.4*0.25=0.10
(2,0) [Pr=0.25]	0.6*0.25=0.15	0.4*0.25=0.10
(2,2) [Pr=0.25]	0.6*0.25=0.15	0.4*0.25=0.10

The total probability of each state is as follows:

State	**Probability**
o1	**0**
o2	**0**
o3	0.15+0.10=**0.25**
o4	**0**
o5	0.15+0.10=**0.25**
o6	0.15+0.15+0.10+0.10=**0.5**

Hence, the most likely state is **o6.**

5

The number of policy trees P is given by:

$$P = |A|^N$$

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

If there are A actions, O observations and T horizon,

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 **536**.

Time	#Trial	#Backup	LBound	UBound	Precision	#Alphas	#Beliefs
57.02	536	7507	14.8299	14.8309	0.000999358	2876	1825

Number of trees = $5^{(6^{536}-1)/5}$