SARSOP.md

SARSOP

> ### Team
>
> - Archit Jain (2019101053)
> - Pulkit Gupta (2019101078)

Description

Roll number	2019101078
Prob_to_move_up	0.1
Prob_to_move_down	0.1
Prob_to_move_right	0.1
Prob_to_move_left	0.1
Prob_to_move_stay	0.6
Prob_to_make_call	0.5
Prob_to_call_off	0.1
x	1 - (((1078)%30 + 1)/100) = 0.71
Prob_to_move_desired	0.71
Prob_to_move_opposite	0.29
Reward_each_step	-1 (not for action stay)
Reward_reach_target_call_on	48
discout_factor	0.5

- Grid

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

Mapping the grid from (x,y) to 4*x + y			

- Each POMDP state is represented as tuple `(Agent Position, Target Position, Call)`
- Total Number of states: 8*(8*2) = 128
- These states can be mapped as (A, T, C) = 16*A* + *2*T + C

SOLUTIONS

Solution 1

Initial Condition

- Target is at (1,0) = 4
- Observation is o6 (when the target is not in the 1 cell neighbourhood of the agent.)

Deduced

- Agent can be at -> (0,1) = 1; (1,2) = 6; (0,2) = 2; (1,3) = 7; (0,3) = 3
- Call -> 0 (off); 1 (on)
- POMDP states possible ->

1.	((0,1), (1,0), 0)	(1, 4, 0)	24
2.	((0,1), (1,0), 1)	(1, 4, 1)	25
3.	((1,2), (1,0), 0)	(6, 4, 0)	104
4.	((1,2), (1,0), 1)	(6, 4, 1)	105
5.	((0,2), (1,0), 0)	(2, 4, 0)	40
6.	((0,2), (1,0), 1)	(2, 4, 1)	41
7.	((1,3), (1,0), 0)	(7, 4, 0)	120
8.	((1,3), (1,0), 1)	(7, 4, 1)	121
9.	((0,3), (1,0), 0)	(3, 4, 0)	56
10.	((0,3), (1,0), 1)	(3, 4, 1)	57

These states are possible with equal probability 1/10. All other states have initial belief state = 0

Solution 2

Initial Condition

- Agent is at (1,1) = 5
- Call -> 0

Deduced

- Target can be at -> (1,0) = 4; (1,1) = 5; (0,1) = 1; (1,2) = 6
- POMDP states possible ->

1.	((1,1), (1,0), 0)	(5, 4, 0)	88
2.	((1,1), (1,1), 0)	(5, 5, 0)	90
3.	((1,1), (0,1), 0)	(5, 1, 0)	82
4.	((1,1), (1,2), 0)	(5, 6, 0)	92

These states are possible with equal probability 1/4. All other states have initial belief state = 0

Solution 3

Expectations were calculated by using the `--simLen 50 --simNum 500 --policy-file` flag with pomdpsi pomdpsol.

- For q1, Expected Utility = 7.02887

```
Loading the model ...
  input file   : ../../pulkit/2019101078.pomdp

Loading the policy ...
  input file   : ../../pulkit/2019101078.policy

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

-----------------------------------
 #Simulations  | Exp Total Reward
-----------------------------------
 50              6.5194
 100             6.3065
 150             6.52897
 200             6.98051
 250             7.10652
 300             7.11886
 350             7.20643
 400             7.01956
 450             7.16071
 500             7.02887
-----------------------------------

Finishing ...
```

```
------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
------------------------------------------------------------
 500             7.02887            (6.52925, 7.5285)
------------------------------------------------------------
```

- For q2, Expected Utility = 17.6657

```
Loading the model ...
  input file   : ../../pulkit/2019101078_b.pomdp

Loading the policy ...
  input file   : ../../pulkit/2019101078_b.policy

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

-----------------------------------
 #Simulations  | Exp Total Reward
-----------------------------------
 50              17.7864
 100             17.4281
 150             17.5706
 200             17.4942
 250             17.4508
 300             17.4938
 350             17.5747
 400             17.5289
 450             17.62
 500             17.6657
-----------------------------------

Finishing ...

------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
------------------------------------------------------------
 500             17.6657            (17.3531, 17.9782)
------------------------------------------------------------
```

Solution 4

Inital Condition

- Agent is at ->

	(x,y)	Mapped state	Probablity
1.	(0,0)	0	0.4

	(x,y)	Mapped state	Probablity
2.	(1,3)	7	0.6

- Target is at ->

	(x,y)	Mapped state	Probablity
1.	(0,1)	1	0.25
2.	(0,2)	2	0.25
3.	(1,1)	5	0.25
4.	(1,2)	6	0.25

Call can be ->

	Call value	Probability
1.	0	0.5
2.	1	0.5

Observations

	(A,T,C)	Probability	Observation
1.	(0, 1, 0)	0.4*(0.25*0.5) = 0.05	o2
2.	(0, 2, 0)	0.4*(0.25*0.5) = 0.05	o6
3.	(0, 5, 0)	0.4*(0.25*0.5) = 0.05	o6
4.	(0, 6, 0)	0.4*(0.25*0.5) = 0.05	o6
5.	(7, 1, 0)	0.6*(0.25*0.5) = 0.75	o6
6.	(7, 2, 0)	0.6*(0.25*0.5) = 0.75	o6
7.	(7, 5, 0)	0.6*(0.25*0.5) = 0.75	o6
8.	(7, 6, 0)	0.6*(0.25*0.5) = 0.75	o4
9.	(0, 1, 1)	0.4*(0.25*0.5) = 0.05	o2
10.	(0, 2, 1)	0.4*(0.25*0.5) = 0.05	o6
11.	(0, 5, 1)	0.4*(0.25*0.5) = 0.05	o6
12.	(0, 6, 1)	0.4*(0.25*0.5) = 0.05	o6
13.	(7, 1, 1)	0.6*(0.25*0.5) = 0.75	o6

	(A,T,C)	Probability	Observation
14.	(7, 2, 1)	0.6*(0.25*0.5) = 0.75	o6
15.	(7, 5, 1)	0.6*(0.25*0.5) = 0.75	o6
16.	(7, 6, 1)	0.6*(0.25*0.5) = 0.75	o4

- Observation o2 = 0.05*2 = 0.1
- Observation o4 = 0.75*2 = 1.5
- Observation o6 = 0.05*6 + 0.75*6 = 0.3 + 4.5 = 4.8
- Observation o6 is most likely to be observed as it has the maximum probability.

Solution 5

On running pompdsol for Q4 we get,

Time	#Trial	#Backup	LBound	UBound	Precision	#Alphas	#Beliefs
0.08	35	299	11.8972	11.8982	0.0009728	96	60

Number of Policy trees = $|A|^N$ where |A| = actions, N = Number of nodes in a tree. N = $(|O|^T - 1)/(|O| - 1$
Horizon In our case:

- |A| = 5
- |O| = 6
- T = 35 (We use #trial as T value)
- N = (6^35 - 1)/(6-1)
- $|A|^N = 5^{(6^{35} - 1)/(6-1)}$ is approx number of policy trees obtained.

Number of trees is dependent on the horizon T. The value of Policy trees increases with increase in the