MDL Assign3 Part2 by Alapan Sau  Apr 26, 2021

MDL Assignment 3 | Part 2

In this assignment, we use `sarsop` to construct the policy file of the POMDP. The roll number used for the assignment calculation is `2019101033`.

Question 1

Given:

- The target is in $(1, 0)$
- The observation is $o6$

As the observation is $o6$, the agent can be placed in any position as long as it is not in the same position as the target or in the one neighbourhood of the target, that is, the agent can be any position excluding $(1, 0)$, $(0, 0)$, $(1, 1)$ and $(0, 2)$.

The initial belief state can be obtained by giving each state of the form:

$$S = ((x, y), (1, 0), 0 \ or \ 1) \ \ where \ \ (x, y) \notin \{(1, 0), \ (0, 0), \ (1, 1), \ (0, 2)\}$$

a uniform probability.

Question 2

Given :

- The agent is in $(1, 1)$.
- The target is the one neighbourhood of the agent
- The target is not making a call

The states satisying all the given conditions are as follows:

- $((1, 1), (1, 0), 0)$
- $((1, 1), (1, 2), 0)$
- $((1, 1), (0, 1), 0)$
- $((1, 1), (1, 1), 0)$

The intial belief state can be obtained by giving each of the above states a uniform

The initial belief state can be obtained by giving each of the above states a uniform, non zero probability and all other states a zero probability.
The initial belief state can be defined as follows:

$$b(s) = 1/4 \;\; if \;\; s \in \{((1,1),(1,0),0),\; ((1,1),(0,1),0),\; ((1,1),(1,2),0),\; ((1,1),(1,1),0)\}$$
$$b(s) = 0 \;\; otherwise$$

Question 3

Q1 simulation

```
(venv) root@deepnote:~/work/sarsop/src # ./pomdpsim --policy-file out.policy --simLen 100 --simNum 1000 ../../q1.pomdp

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            21.7474
 200            20.9105
 300            21.2366
 400            20.5585
 500            20.4292
 600            20.6085
 700            20.6063
 800            20.7086
 900            20.6488
 1000           20.6135
-----------------------------------

Finishing ...

------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
------------------------------------------------------------
 1000           20.6135           (19.6037, 21.6232)
------------------------------------------------------------
```

Q2 simulation

```
(venv) root@deepnote:~/work/sarsop/src # ./pomdpsim --policy-file out.policy --simLen 100 --simNum 1000 ../../q2.pomdp

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            25.4325
 200            24.5069
 300            24.2511
 400            24.2408
 500            23.8644
 600            24.1701
 700            24.1777
 800            24.294
 900            24.5462
 1000           24.4417
----------------------------------

Finishing ...

-----------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
-----------------------------------------------------------
 1000           24.4417            (23.4786, 25.4047)
-----------------------------------------------------------
```

The Expected Total Utilities of 1000 simulations are as follows :

- $Q1 : 21.3032$
- $Q2 : 24.1229$

In $Q1$, the agent was strictly outside the one neighborhood of the target whearas, in $Q2$ the target is strictly in the one neighborhood of the agent.
As a result, the agent is expected to take `more steps` to reach the target. Each step has a negative cost of -1, causing the Expected utility to be `higher` in the $Q2$.

Question 4

Given,

- $P(Agent\ Position = (0,0)) = 0.4$
- $P(Agent\ Position = (1,3)) = 0.6$
- $P(Target\ Position = (0,1)) = 0.25$
- $P(Target\ Position = (0,2)) = 0.25$
- $P(Target\ Position = (1,1)) = 0.25$
- $P(Target\ Position = (1,1)) = 0.25$

It does not matter whether the call is on or not, because no observation detects it, and we are not given any information regarding it. Therefore, the positions of the agent and target, observation and the initial belief state values are as follows:

Positions	Observation	Probability
((0,0),(0,1))	*o2*	*0.1*
((0,0),(0,2))	*o6*	*0.1*
((0,0),(1,1))	*o6*	*0.1*
((0,0),(1,2))	*o6*	*0.1*
((1,3),(0,1))	*o6*	*0.15*
((1,3),(0,2))	*o6*	*0.15*
((1,3),(1,1))	*o6*	*0.15*
((1,3),(1,2))	*o4*	*0.15*

Let O be the observation observed. Then, O can take one of the values $o2, o4, o6$

$\therefore P(O = o2) = 0.1$
$\therefore P(O = o4) = 0.15$
$\therefore P(O = o6) = 0.1 + 0.1 + 0.1 + 0.15 + 0.15 + 0.15 = 0.75$
Hence, $o6$ is clearly the most likely observation.

Question 5

The number of policy trees can be obtained by using the formula:

$$Number\ of\ Trees = |A|^n$$

$where$

$$|A| = Number\ of\ Actions$$

$$n = \sum_{r=0}^{T-1} |O|^r = \frac{|O|^T - 1}{|O| - 1}$$

Substituting the values,

$$N = 5^{\frac{6^T-1}{6-1}} = 5^{\frac{6^T-1}{5}}$$

```
-------------------------------------------------------------------------------
 Time   |#Trial |#Backup |LBound    |UBound    |Precision  |#Alphas |#Beliefs
-------------------------------------------------------------------------------
 0.15    44      441      24.7733    24.7743    0.000999018 163      99
-------------------------------------------------------------------------------
```

The Trials required are 44, so if we assume $T = 44$,

$$N = 5^{\frac{6^{44}-1}{5}}$$