MDL Assignment 3, Part 1

Rutvij Menavlikar (2019111032)
Tejas Chaudhari (2019111013)

April 16, 2021

Partially Observable Markov Decision Process (POMDP)

A Partially Observable Markov Decision Process (**POMDP**) is a generalization of a Markov Decision Process (**MDP**). MDPs assume that the environment is fully observable. This means that the agent always knows which state it is in. But, when the environment is only partially observable, the agent does not necessarily know which state it is in. In cases like these, the agent uses POMDP to find the optimal policy.

A POMDP has the same elements as an MDP, the transition model $P(s'|s,a)$, actions $A(s)$, and reward function $R(s)$ but, it also has a sensor model $P(e|s')$, where e is the perceived evidence.
Here, the utility of a state s and the optimal action in s depend not just on s, but also on how much the POMDP agent knows when it is in s.
For the agent to determine the certainty of it being in a state s, a **belief state** b is stored. Technically, this belief state b is a probability distribution over all possible states the agent can be in, where, for a state s, $b(s)$ is the probability of the agent being in state s.

Values used for solving the problem

Roll number used: 2019111032
x = $1 - \frac{(1032 \bmod 30)+1}{100} = 0.87$
y = $(32 \bmod 4)+1 = 1$

Probability of actions

	Success	Failure
LEFT	0.87	0.13
RIGHT	0.87	0.13

Probability of observed states for given states

Columns → Actual state
Rows → Observed states
Table → P(Observed state | Actual state)

	Green	Red
Green	0.8	0.05
Red	0.2	0.95

Formula for calculating next Belief state

$b(s)$ → Previous belief state
$b'(s')$ → New belief state a → Action by agent
α → Normalizing constant
e → Perceived evidence

$$b'(s') = \alpha P(e|s') \sum_s P(s'|s,a) b(s)$$

The actions

Initial Beliefs

S1	S2	S3	S4	S5	S6
0.3333	0	0.3333	0	0	0.3333

Action 1 Right and observed Green

$b'(S1) = 0.0500 \times (0.13 \times 0.3333 + 0.13 \times 0 + 0 \times 0.3333 + 0 \times 0 + 0 \times 0 + 0 \times 0.3333) = 0.0021$
$b'(S2) = 0.8 \times (0.87 \times 0.3333 + 0 \times 0 + 0.13 \times 0.3333 + 0 \times 0 + 0 \times 0 + 0 \times 0.3333) = 0.2666$
$b'(S3) = 0.0500 \times (0 \times 0.3333 + 0.87 \times 0 + 0.3333 \times 0 + 0.13 \times 0 + 0 \times 0 + 0 \times 0.3333) = 0$
$b'(S4) = 0.8 \times (0 \times 0.3333 + 0 \times 0 + 0.87 \times 0.3333 + 0 \times 0 + 0.13 \times 0 + 0 \times 0.3333) = 0.2319$
$b'(S5) = 0.8 \times (0 \times 0.3333 + 0 \times 0 + 0 \times 0.3333 + 0.87 \times 0 + 0 \times 0 + 0.13 \times 0.3333) = 0.0433$
$b'(S6) = 0.0500 \times (0 \times 0.3333 + 0 \times 0 + 0 \times 0.3333 + 0 \times 0 + 0.87 \times 0 + 0.87 \times 0.3333) = 0.0145$

Normalization factor $(\alpha) = 1.8181$
After normalizing, new beliefs:

S1	S2	S3	S4	S5	S6
0.0039	0.4848	0	0.4218	0.0630	0.0263

Action 2 Left and observed Red

$b'(S1) = 0.95 \times (0.87 \times 0.0039 + 0.87 \times 0.4848 + 0 \times 0 + 0 \times 0.4218 + 0 \times 0.0630 + 0 \times 0.0263) = 0.4039$
$b'(S2) = 0.1999 \times (0.13 \times 0.0039 + 0 \times 0.4848 + 0.87 \times 0 + 0 \times 0.4218 + 0 \times 0.0630 + 0 \times 0.0263) = 0.0001$
$b'(S3) = 0.95 \times (0 \times 0.0039 + 0.13 \times 0.4848 + 0 \times 0 + 0.87 \times 0.4218 + 0 \times 0.0630 + 0 \times 0.0263) = 0.4085$
$b'(S4) = 0.1999 \times (0 \times 0.0039 + 0 \times 0.4848 + 0.13 \times 0 + 0 \times 0.4218 + 0.87 \times 0.0630 + 0 \times 0.0263) = 0.0109$
$b'(S5) = 0.1999 \times (0 \times 0.0039 + 0 \times 0.4848 + 0 \times 0 + 0.13 \times 0.4218 + 0 \times 0.0630 + 0.87 \times 0.0263) = 0.0155$
$b'(S6) = 0.95 \times (0 \times 0.0039 + 0 \times 0.4848 + 0 \times 0 + 0 \times 0.4218 + 0.13 \times 0.0630 + 0.13 \times 0.0263) = 0.0110$

Normalization factor $(\alpha) = 1.1762$
After normalizing, new beliefs:

S1	S2	S3	S4	S5	S6
0.4751	0.0001	0.4805	0.0129	0.0182	0.0129

Action 3 Left and observed Green

$b'(S1) = 0.0500 \times (0.87 \times 0.4751 + 0.87 \times 0.0001 + 0 \times 0.4805 + 0 \times 0.0129 + 0 \times 0.0182 + 0 \times 0.0129) = 0.0206$
$b'(S2) = 0.8 \times (0.13 \times 0.4751 + 0 \times 0.0001 + 0.87 \times 0.4805 + 0 \times 0.0129 + 0 \times 0.0182 + 0 \times 0.0129) = 0.3838$
$b'(S3) = 0.0500 \times (0 \times 0.4751 + 0.13 \times 0.0001 + 0 \times 0.4805 + 0.87 \times 0.0129 + 0 \times 0.0182 + 0 \times 0.0129) = 0.0005$
$b'(S4) = 0.8 \times (0 \times 0.4751 + 0 \times 0.0001 + 0.13 \times 0.4805 + 0 \times 0.0129 + 0.87 \times 0.0182 + 0 \times 0.0129) = 0.0627$
$b'(S5) = 0.8 \times (0 \times 0.4751 + 0 \times 0.0001 + 0 \times 0.4805 + 0.13 \times 0.0129 + 0 \times 0.0182 + 0.87 \times 0.0129) = 0.0103$
$b'(S6) = 0.0500 \times (0 \times 0.4751 + 0 \times 0.0001 + 0 \times 0.4805 + 0 \times 0.0129 + 0.13 \times 0.0182 + 0.13 \times 0.0129) = 0.0002$

Normalization factor $(\alpha) = 2.0903$
After normalizing, new beliefs:

S1	S2	S3	S4	S5	S6
0.0432	0.8024	0.0011	0.1310	0.0216	0.0004

Belief states obtained

	S1	S2	S3	S4	S5	S6
Initial	0.3333	0	0.3333	0	0	0.3333
After Action 1	0.0039	0.4848	0	0.4218	0.0630	0.0263
After Action 2	0.4751	0.0001	0.4805	0.0129	0.0182	0.0129
After Action 3	0.0432	0.8024	0.0011	0.1310	0.0216	0.0004