Guest Lecture: Partially Observable Markov Decision Processes (POMDPs)

Presenter: Ekhlas Sonu
For CSCI/ARTI 8920, Spring 2013.
Instructor: Prashant Doshi.
Markov Decision Processes (MDP)

\(<S, A, T, R>\)

- **S**: Set of states of the environment
- **A**: Set of agent's action
- **T**: \(S \times A \rightarrow \Delta(S)\), Transition function
- **R**: \(S \times A \rightarrow \mathbb{R}\), Reward Function

- At each time step, the agent can perfectly observe the state of the environment

![Diagram showing states and transitions](image)

All other states yield -0.04 at every step
Three horizon policy for the maze problem:

- **T=0**: With 3 steps to go, follow horizon 3 policy
- **T=1**: With 2 steps to go, switch to horizon 2 policy
- **T=2**: With 1 step to go, switch to horizon 1 policy

At each step the agent knows the state exactly
Partial Observability

• The agent is unable to observe the physical state exactly
  • At any given time, the agent is unable to recognize its current location exactly

• Imperfect sensory inputs:
  • Agent gets faulty observation due to defects in sensors, or different lighting conditions in the corridor could cause agent to get varied observations

• Indistinguishable states:
  • Different states may yield observation, e.g. same colored walls

• How do we account for the lack of knowledge of current state?
Partially Observable Markov Decision Processes (POMDPs)

\[<S, A, T, \Omega, O, R, \gamma, OC>\]

- **S**: Set of states of the environment
- **A**: Set of agent's action
- **T**: \(S \times A \rightarrow \Delta(S)\), Transition function
- **\(\Omega\)**: Set of observations received by the agent
- **O**: \(S \times A \rightarrow \Delta(\Omega)\), Observation function
- **R**: \(S \times A \rightarrow \mathbb{R}\), Reward Function
- **\(\gamma\)**: Discount factor
- **OC**: Optimality criterion: How many horizons to go?
POMDP Illustration: The Tiger Problem

S: \{TL, TR\}
A: \{L, OL, OR\}
Ω: \{GL, GR\}

Transition Function:

<table>
<thead>
<tr>
<th>T(L)</th>
<th>s'=TL</th>
<th>s'=TR</th>
</tr>
</thead>
<tbody>
<tr>
<td>s=TL</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>s=TR</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T(OL/OR)</th>
<th>s'=TL</th>
<th>s'=TR</th>
</tr>
</thead>
<tbody>
<tr>
<td>s=TL</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>s=TR</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Observation Function:

<table>
<thead>
<tr>
<th>O(L)</th>
<th>O=GL</th>
<th>O=GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>s=TL</td>
<td>0.85</td>
<td>0.15</td>
</tr>
<tr>
<td>s=TR</td>
<td>0.15</td>
<td>0.85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>O(OL/OR)</th>
<th>O=GL</th>
<th>O=GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>s=TL</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>s=TR</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Representing Uncertainty in Knowledge

• Since the state is not observable at any given time, how does the agent represent the information about the environment?
Representing Uncertainty in Knowledge

• Since the state is not observable at any given time, how does the agent represent the information about the environment?

• Belief
  
  • Probability distribution over the states: \( \Delta(S) \)
    
    – Tiger problem: \{P(TL), P(TR)\}
  
  • At each time step, the decision to make is given current belief and number of steps to go, what is the optimal action to take?
Representing Uncertainty in Knowledge

• Since the state is not observable at any given time, how does the agent represent the information about the environment?

• Belief
  • Probability distribution over the states: $\Delta(S)$
    – Tiger problem: \{P(TL), P(TR)\}
  • At each time step, the decision to make is given current belief and number of steps to go, what is the optimal action to take?

• A complete solution to POMDP maps the entire set of beliefs to a corresponding policy:
  • $B \rightarrow \Pi$

• How many beliefs are there?
Acting in a Partially Observable Environments

1. Agent computes the optimal action according to its belief
2. Agent performs the action that may alter the state of the environment
3. Agent receives observation from the environment
4. Agent updates its belief given action and observation
5. Goto to step 1.
Belief Update:

\[ B.U.(b^t, a, o) \]

\[ b^{t+1}(s^{t+1}) = Pr(s^{t+1}|b^t, a, o) \]
Belief Update: \[ \text{B.U.}(b^t, a, o) \]

\[ b^{t+1}(s^{t+1}) = Pr(s^{t+1}|b^t, a, o) \]

\[ = \frac{Pr(o|s^{t+1}, b^t, a) \cdot Pr(s^{t+1}|b^t, a)}{Pr(o|b^t, a)} \]
Belief Update

Belief Update: \( B.U.(b^t, a, o) \)

\[
B^{t+1}(s^{t+1}) = Pr(s^{t+1}|b^t, a, o)
\]

\[
= \frac{Pr(o|s^{t+1}, b^t, a) \cdot Pr(s^{t+1}|b^t, a)}{Pr(o|b^t, a)}
\]

Bayes' Rule
Belief Update

Belief Update: \( B.U.(b^t, a, o) \)

\[
b^{t+1}(s^{t+1}) = Pr(s^{t+1}|b^t, a, o)
\]

\[
= \frac{Pr(o|s^{t+1}, b^t, a) \cdot Pr(s^{t+1}|b^t, a)}{Pr(o|b^t, a)}
\]

\[
= \eta \cdot Pr(o|s^{t+1}, a) \cdot Pr(s^{t+1}|b^t, a)
\]
Belief Update:

\[ B.U.(b^t, a, o) \]

\[ b^{t+1}(s^{t+1}) = Pr(s^{t+1}|b^t, a, o) \]

\[ = \frac{Pr(o|s^{t+1}, b^t, a) \cdot Pr(s^{t+1}|b^t, a)}{Pr(o|b^t, a)} \]

\[ = \eta \ Pr(o|s^{t+1}, a) \cdot Pr(s^{t+1}|b^t, a) \]

\[ = \eta \ Pr(o|s^{t+1}, a) \sum_{s^t} b(s^t) Pr(s^{t+1}|s^t, a) \]
Belief Update:

\[ B.U.(b^t, a, o) \]

\[ b^{t+1}(s^{t+1}) = Pr(s^{t+1}|b^t, a, o) \]

\[ = \frac{Pr(o|s^{t+1}, b^t, a) \cdot Pr(s^{t+1}|b^t, a)}{Pr(o|b^t, a)} \]

\[ = \eta \cdot Pr(o|s^{t+1}, a) \cdot Pr(s^{t+1}|b^t, a) \]

\[ = \eta \cdot Pr(o|s^{t+1}, a) \sum_{s^t} b(s^t) Pr(s^{t+1}|s^t, a) \]

\[ \therefore b^{t+1}(s^{t+1}) = \eta \cdot O(s^{t+1}, a, o) \sum_{s^t} b(s^t) T(s^t, a, s^{t+1}) \]
Solving POMDP

Bellman Equation for MDP

\[ V_h(s^t) = \max_a \left\{ R(s, a) + \gamma \sum_{s^{t+1}} P(r(s^{t+1} | s^t, a)) V_{h-1}(s^{t+1}) \right\} \]

Bellman Equation for POMDP:

Expected Reward of an optimal horizon \( h \) policy at time step \( t \):

\[ V^h(b^t) = \max_a \left\{ \sum_s b^t(s) R(s, a) \right. \]
\[ \left. + \gamma \sum_o Pr(o | b^t, a) V^{h-1}(B.U(b^t, a, o)) \right\} \]
### Illustration: The Tiger Problem Perspective

- **S:** \{TL, TR\}
- **A:** \{L, OL, OR\}
- **Ω:** \{GL, GR\}

#### Transition Function:

<table>
<thead>
<tr>
<th>(T(L))</th>
<th>(s' = TL)</th>
<th>(s' = TR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(s = TL)</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(s = TR)</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(T(OL/OR))</th>
<th>(s' = TL)</th>
<th>(s' = TR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(s = TL)</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(s = TR)</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

#### Observation Function:

<table>
<thead>
<tr>
<th>(O(L))</th>
<th>(O = GL)</th>
<th>(O = GR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(s = TL)</td>
<td>0.85</td>
<td>0.15</td>
</tr>
<tr>
<td>(s = TR)</td>
<td>0.15</td>
<td>0.85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(O(OL/OR))</th>
<th>(O = GL)</th>
<th>(O = GR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(s = TL)</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(s = TR)</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

#### Reward Function:

<table>
<thead>
<tr>
<th>Act/St</th>
<th>TL</th>
<th>TR</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>OL</td>
<td>-100</td>
<td>10</td>
</tr>
<tr>
<td>OR</td>
<td>10</td>
<td>-100</td>
</tr>
</tbody>
</table>

Given at \(t=0\), \(b = <0.5, 0.5>\), what is the optimal horizon \(H\) policy?

**H=1:**

\[
V^1_{OL}(b) = 0.5 \times -100 + 0.5 \times 10 = -45
\]

\[
V^1_{OR}(b) = 0.5 \times 10 + 0.5 \times -100 = -45
\]

\[
V^1_L(b) = 0.5 \times -1 + 0.5 \times -1 = -1
\]

\[
V^1_{L}(b) = 0.5 \times 0.85 + 0.5 \times 0.15 = 0.5 \times -1 + 0.5 \times -1 = -1
\]

\[
V^1_{OL}(b) = 0.5 \times -100 + 0.5 \times 10 = -45
\]

\[
V^1_{OR}(b) = 0.5 \times 10 + 0.5 \times -100 = -45
\]
Illustration: The Tiger Problem Perspective

S: {TL, TR}       A: {L, OL, OR}       Ω: {GL, GR}

Transition Function:

<table>
<thead>
<tr>
<th>T(L)</th>
<th>s'=TL</th>
<th>s'=TR</th>
</tr>
</thead>
<tbody>
<tr>
<td>s=TL</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>s=TR</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T(OL/OR)</th>
<th>s'=TL</th>
<th>s'=TR</th>
</tr>
</thead>
<tbody>
<tr>
<td>s=TL</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>s=TR</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Observation Function:

<table>
<thead>
<tr>
<th>O(L)</th>
<th>O=GL</th>
<th>O=GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>s=TL</td>
<td>0.85</td>
<td>0.15</td>
</tr>
<tr>
<td>s=TR</td>
<td>0.15</td>
<td>0.85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>O(OL/OR)</th>
<th>O=GL</th>
<th>O=GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>s=TL</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>s=TR</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Reward Function:

<table>
<thead>
<tr>
<th>Act/St</th>
<th>TL</th>
<th>TR</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>OL</td>
<td>-100</td>
<td>10</td>
</tr>
<tr>
<td>OR</td>
<td>10</td>
<td>-100</td>
</tr>
</tbody>
</table>

Given at t=0, b = <0.5, 0.5>, what is the optimal horizon H policy?

H=1:

\[
V^1_{OL}(b) = 0.5 \times -100 + 0.5 \times 10 = -45
\]

\[
V^1_{OR}(b) = 0.5 \times 10 + 0.5 \times -100 = -45
\]

\[
V^1_{L}(b) = 0.5 \times R(TL, L) + 0.5 \times R(TR, L)
\]

\[
= 0.5 \times -1 + 0.5 \times -1 = -1
\]

Optimal
Illustration: The Tiger Problem Perspective

$H=2$: \[ V^2_a(b) = 0.5 \times R(TL, a) + 0.5 \times R(TR, a) + \gamma \sum_o Pr(o|b, a)V^1(B.U(b, a, o)); \forall a \]

\[ V^2(b) = \max_a V^2_a(b) \]

$H=3$: \[ V^3_a(b) = 0.5 \times R(TL, a) + 0.5 \times R(TR, a) + \gamma \sum_o Pr(o|b, a)V^2(B.U(b, a, o)); \forall a \]

\[ V^3(b) = \max_a V^3_a(b) \]

And so on...
Illustration: The Tiger Problem Perspective

T=0, H=3

\[ <0.5,0.5> \]

L

\[ \text{GL} \quad \text{GR} \]

? ?

T=1, H=2

\[ <0.5,0.5> \]

OL

\[ \text{GL} \quad \text{GR} \]

? ?

\[ <0.5,0.5> \]

OR

\[ \text{GL} \quad \text{GR} \]

? ?
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2
Illustration: The Tiger Problem Perspective

T=0, H=3

L

<0.5,0.5>

GL

<0.85,0.15>

GR

<0.15,0.85>

T=1, H=2

OL

<0.5,0.5>

GL

?

GR

?

OR

<0.5,0.5>

GL

?

GR

?
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

GL

GR

<0.5,0.5>

<0.5,0.5>

<0.5,0.5>

<0.85,0.15>

<0.15,0.85>

<0.5,0.5>

?}

?}

?}
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1

\[ V_L = ? \]
\[ V_{OL} = ? \]
\[ V_{OR} = ? \]
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1

\[ V_L = -1 \quad V_{OL} = -96.7 \quad V_{OR} = 6.7 \]
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1

\[ V_{OR} = 6.7 \]
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1

V_{OR} = 6.7  
V_{L} = -1  
V_{OL} = 6.7  
V_{L} = -1  
V_{OL} = -84.85  
V_{OR} = -6.5
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1

\[ V_{OR} = 6.7 \]

\[ V_L = -1 \]

\[ V_{OL} = 6.7 \]

\[ V_L = -1 \]

\[ V_{OL} = 6.7 \]

\[ V_L = -1 \]

\[ V_{OL} = -6.5 \]

\[ V_{OR} = -84.85 \]
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1

\( V_{OR} = 6.7 \)

\( V_L = -1 \)

\( V_{OL} = 6.7 \)

\( V_L = -1 \)

\( V_L = -1 \)
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1

V_{OR} = 6.7
V_{L} = -1

V_{OL} = 6.7
V_{L} = -1

V_{L} = -1
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1

V_{OR} = 6.7

V_{L} = 4.74

V_{OL} = 6.7

V_{L} = -1

V_{L} = -1

V_{L} = -1
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1

\[ V_{OR} = 6.7 \]

\[ V_L = -1 \]

\[ V_{OL} = 6.7 \]

\[ V_L = -1 \]

\[ V_L = -1 \]

\[ V_L = -2 \]
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1
Illustration: The Tiger Problem Perspective

T=0, H=3

T=1, H=2

T=2, H=1

$V_L = 3.74$

$V_L = 4.74$

$V_L = -1$

$V_L = -2$

$V_{OL} = -47$

$V_{OL} = 6.7$

$V_{OR} = -47$

$V_{OR} = 6.7$

$V_{OL} = 6.7$

$V_{OL} = 6.7$
Illustration: The Tiger Problem Perspective

T=0, H=3

V_L = 3.74

T=1, H=2

<0.85,0.15>

VL = 4.74

<0.5,0.5>

T=2, H=1

<0.97,0.03>

V_OR = 6.7

<0.5,0.5>

VL = -1

<0.03,0.97>

V_Ol = 6.7