Intelligent Systems (AI-2)

Computer Science cpsc422, Lecture 2

Jan, 7, 2015
Lecture Overview

Value of Information and Value of Control

Markov Decision Processes (MDPs)
- Formal Specification and example
- Define a policy for an MDP
<table>
<thead>
<tr>
<th>Problem</th>
<th>Static</th>
<th>Sequential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constraint Satisfaction</td>
<td>Query</td>
<td>Planning</td>
</tr>
</tbody>
</table>

**Environment**

<table>
<thead>
<tr>
<th>Deterministic</th>
<th>Stochastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arc Consistency</td>
<td>SLS</td>
</tr>
<tr>
<td>Search</td>
<td>Var. Elimination</td>
</tr>
<tr>
<td>Vars + Constraints</td>
<td></td>
</tr>
<tr>
<td>Search</td>
<td>Decision Nets</td>
</tr>
<tr>
<td>Logics</td>
<td>STRIPS</td>
</tr>
</tbody>
</table>

**Representation**

Reasoning Technique
Simple Decision Net

- Early in the morning. Shall I take my **umbrella** today? (I’ll have to go for a long walk at noon)
- Relevant Random Variables?
Polices for Umbrella Problem

- A **policy** specifies what an agent should do under each circumstance (for each decision, consider the parents of the decision node).

In the *Umbrella* case:

\[ D_1 \ ? \ T \ F \]

\[ pD_1 \]

- Rainy
- Cloudy
- Sunny

**How many policies?**

\[ \begin{array}{c}
\text{1pD}_1 \\
\text{1D}_1 \\
\end{array} \]

One possible Policy:

\[ \rightarrow R \ T \ F \ T \ldots \]

\[ \rightarrow C \ T \ F \ T \ldots \]

\[ \rightarrow S \ F \ F \ F \ T \ldots \]

3 policies
Value of Information

- Early in the morning. I listen to the weather forecast, shall I take my umbrella today? (I’ll have to go for a long walk at noon)
- What would help the agent make a better Umbrella decision?
Value of Information

- The value of information of a random variable $X$ for decision $D$ is: $EU(\text{knowing } X) - EU(\text{not knowing } X)$ the utility of the network with an arc from $X$ to $D$ minus the utility of the network without the arc.

- Intuitively:
  - The value of information is always $\geq 0$
  - It is positive only if the agent changes its policy
Value of Information (cont.)

- The value of information provides a bound on how much you should be prepared to pay for a sensor. How much is a **perfect** weather forecast worth?

- Original maximum expected utility: 77
- Maximum expected utility when we know Weather: 91
- Better forecast is worth at most: 74
Value of Information

- The value of information provides a bound on how much you should be prepared to pay for a sensor. How much is a perfect fire sensor worth?

Original maximum expected utility: \(-22.6\)
- Maximum expected utility when we know Fire: \(-2\)
- Perfect fire sensor is worth: \(20.6\)
Value of Control

• What would help the agent to make an even better *Umbrella* decision? To maximize its utility.

<table>
<thead>
<tr>
<th>Weather</th>
<th>Umbrella</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rain</td>
<td>true</td>
<td>70</td>
</tr>
<tr>
<td>Rain</td>
<td>false</td>
<td>0</td>
</tr>
<tr>
<td>noRain</td>
<td>true</td>
<td>20</td>
</tr>
<tr>
<td>noRain</td>
<td>false</td>
<td>100</td>
</tr>
</tbody>
</table>

• The *value of control* of a variable $X$ is: the utility of the network when you make $X$ a decision variable *minus* the utility of the network when $X$ is a random variable.
Value of Control

• What if we could control the weather?

  - Original maximum expected utility: 77
  - Maximum expected utility when we control the weather: 100
  - Value of control of the weather: 23
Value of Control

• What if we control Tampering?

• Original maximum expected utility: \(-22.6\)
• Maximum expected utility when we control the Tampering: \(-20.7\)
• Value of control of Tampering: \(1.9\)
• Let’s take a look at the policy
• Conclusion: do not tamper with fire alarms!

CPSC 422, Lecture 2
Lecture Overview

Value of Information and Value of Control

Markov Decision Processes (MDPs)
• Formal Specification and example
• Define a policy for an MDP
Combining ideas for Stochastic planning

• What is a key limitation of decision networks?

Represent (and optimize) only a fixed number of decisions

• What is an advantage of Markov models?

The network can extend indefinitely

Goal: represent (and optimize) an indefinite sequence of decisions
Decision Processes

Often an agent needs to go beyond a fixed set of decisions – Examples?

• Would like to have an ongoing decision process

**Infinite horizon problems**: process does not stop

Robot surviving on planet, Monitoring Nuc. Plant, . . . .

**Indefinite horizon problem**: the agent does not know when the process may stop

**Finite horizon**: the process must end at a give time $N$
Recap: Markov Models

- $|\text{dom}(S)| = n$
- $|\text{dom}(O)| = K$

HMM

$P(S_0) \sim \text{i.i.d.}$

$P(S_{t+1} | S_t) \sim \text{i.i.d.}$

$P(O_t | S_t) \sim \text{i.i.d.}$
Markov Models

- Markov Chains
- Hidden Markov Model
- Partially Observable Markov Decision Processes (POMDPs)
- Markov Decision Processes (MDPs)

Noisy Observations

Noisy Actions

Noisy Outcomes

Rewards
How can we deal with indefinite/infinite Decision processes?

We make the same two assumptions we made for….

The action outcome depends only on the current state \( S_t \) be the state at time \( t \)...

The process is stationary…

We also need a more flexible specification for the utility. How?

- Defined based on a reward/punishment \( R(s) \) that the agent receives in each state \( s \)

\[
\begin{align*}
S_0 & \rightarrow S_1 \rightarrow \ldots \rightarrow S_n \\
e \in & \mathbb{R} R_0 \rightarrow R_2 \rightarrow \ldots \rightarrow R_n
\end{align*}
\]
MDP: formal specification

For an MDP you specify:

- set $S$ of states and set $A$ of actions
- the process’ dynamics (or *transition model*)
  \[ P(S_{t+1} | S_t, A_t) \]
- The reward function
  \[ R(s, a, s') \]
  describing the reward that the agent receives when it performs action $a$ in state $s$ and ends up in state $s'$
- $R(s)$ is used when the reward depends only on the state $s$ and not on how the agent got there
- Absorbing/stopping/terminal state
  formal action $P(S_{ab} | a, S_{ab}) = 1$ \( R(S_{ab}, \emptyset, S_{ab}) = 0 \)
MDP graphical specification

Basically a MDP is a Markov Chain augmented with actions and rewards/values

\[ P(S_{t+1} | S_t, A_t) \]

\[ R(S_{t+1}, A_t, S_t) \]
When Rewards only depend on the state
Learning Goals for today’s class

You can:

- Define and compute Value of Information and Value of Control in a decision network
- Effectively represent indefinite/infinite decision processes with a Markov Decision Process (MDP)
Todo for this Fri

- assignment0 – Google Form
- Read textbook 9.5
  - 9.5.1 Value of a Policy
  - 9.5.2 Value of an Optimal Policy
  - 9.5.3 Value Iteration