Lecture 20: Conditional planning with probabilities

- Objectives of planning: success probabilities, costs/rewards
- Stochastic transition systems
- Probabilistic planning with full observability
- Valuing plans
Plan quality criteria

1. Plan reaches goals with probability 1.

2. Plan reaches goals with maximal probability.

3. Plan reaches goals with minimal expected cost.

4. Plan produces maximum expected rewards.
Probabilistic planning (MDPs)

• Objective is to gain highest possible rewards.

• No designated goal states. (Goals can be simulated with rewards!)

• Length of plan execution is infinite.

• Different criteria for defining what the desired plans are.
Probability distributions on successor states

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
<td>0</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Operators for probabilistic planning

DEFINITION A nondeterministic operator is \( \langle c, e \rangle \) where \( c \) is a formula and \( e \) is a nondeterministic effect defined as

1. \( p \) and \( \neg p \) for state variables \( p \in P \) are effects.

2. \( e_1 \land \cdots \land e_n \) is an effect if \( e_1, \ldots, e_n \) are effects.

3. \( c \triangleright e \) is an effect if \( c \) is a formula over \( P \) and \( e \) is an effect.

4. \( p_1e_1|\cdots|p_n e_n \) is an effect if \( e_1, \ldots, e_n \) for \( n \geq 2 \) are effects, \( p_i > 0 \) for all \( i \in \{1, \ldots, n\} \) and \( \sum_{i=1}^{n} p_i = 1 \).
Probability distribution expressed by an operator

$[e]_s$ is set of literals made true by effect $e$ in state $s$:

1. $[a]_s = \{\langle 1.0, \{a\} \rangle\}$ and $[-a]_s = \{\langle 1.0, \{-a\} \rangle\}$ for $a \in A$.

2. $[e_1 \land \cdots \land e_n]_s = \{\langle \prod_{i=1}^n p_i, \bigcup_{i=1}^n f_i \rangle | \langle p_1, f_1 \rangle \in [e_1]_s, \ldots, \langle p_n, f_n \rangle \in [e_n]_s \}$.

3. $[c' \triangleright e]_s = [e]_s$ if $s \models c'$ and $[c' \triangleright e]_s = \{\langle 1.0, \emptyset \rangle\}$ otherwise.

4. $[p_1 e_1 | \cdots | p_n e_n]_s = \{\langle p_1 \cdot p, e \rangle | \langle p, e \rangle \in [e_1]_s \} \cup \cdots \cup \{\langle p_n \cdot p, e \rangle | \langle p, e \rangle \in [e_n]_s \}$.
In (4) the union of sets is defined so that e.g. \( \{\langle 0.2, \{a\}\rangle\} \cup \{\langle 0.2, \{a\}\rangle\} = \{\langle 0.4, \{a\}\rangle\} \).

**EXAMPLE** successor of state \( s \) (satisfies \( a \land b \land c \)) w.r.t. \( \langle a, (0.1\neg a|0.9\neg b) \land (0.8\neg c|0.2c)\rangle \):

\[
\begin{align*}
[(0.1\neg a|0.9\neg b)]_s &= \{\langle 0.1, \{\neg a\}\rangle, \langle 0.9, \{\neg b\}\rangle\} \\
[(0.8\neg c|0.2c)]_s &= \{\langle 0.8, \{\neg c\}\rangle, \langle 0.2, \{c\}\rangle\} \\
[(0.1\neg a|0.9\neg b) \land (0.8\neg c|0.2c)]_s &= \{\langle 0.08, \{\neg a, \neg c\}\rangle, \langle 0.72, \{\neg b, \neg c\}\rangle, \\
&\quad \langle 0.02, \{\neg a, c\}\rangle, \langle 0.18, \{\neg b, c\}\rangle\}
\end{align*}
\]

one successor state satisfies \( \neg a \land b \land \neg c \), the second \( a \land \neg b \land \neg c \), the third \( \neg a \land b \land c \), the fourth \( a \land \neg b \land c \).
Probabilistic planning with full observability

- Problem is a generalization of the conditional planning with FO we have discussed earlier.

- Plans are mappings from states to actions.

- In some variants of the problem plan execution does not terminate.
Probabilistic planning with full observability

DEFINITION A *problem instance in probabilistic planning* is
\( \langle P, I, O, R \rangle \) where

- \( P \) is a finite set of state variables,
- \( I \) is a probability distribution on states (valuations of \( P \)),
- \( O \) is a set of operators on \( P \),
- \( R \) assigns every operator and valuation a reward.
Probabilistic planning with full observability: plans

A plan for \( \langle P, I, O, R \rangle \) assigns an operator \( o \in O \) to every state (valuation of \( P \)).

(Alternatively, we could use graph-like plans with branches etc.)

- Plan execution is exactly like in non-probabilistic conditional planning.
- Criteria that acceptable plans must fulfill are discussed later...
Transition probabilities under a plan

\[ J = (0.9, 0.1, 0, 0, 0)^T \]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0.6</td>
<td>0</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0.8</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Probabilities of states under a plan

Can be computed by matrix multiplication from the probability distribution for the initial states and the transition probabilities of the plan.

\[ J \] probability distribution initially
\[ JM \] after 1 action
\[ JMM \] after 2 actions
\[ JMMM \] after 3 actions
\[ \vdots \]
\[ JM^i \] after \( i \) actions
Probabilities of states under a plan

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.900</td>
<td>0.100</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>B</td>
<td>0.060</td>
<td>0.900</td>
<td>0.040</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>C</td>
<td>0.540</td>
<td>0.060</td>
<td>0.360</td>
<td>0.008</td>
<td>0.032</td>
</tr>
<tr>
<td>D</td>
<td>0.036</td>
<td>0.540</td>
<td>0.056</td>
<td>0.072</td>
<td>0.296</td>
</tr>
<tr>
<td>E</td>
<td>0.324</td>
<td>0.036</td>
<td>0.512</td>
<td>0.011</td>
<td>0.117</td>
</tr>
</tbody>
</table>

Jussi Rintanen
Probabilities of states under a plan (periodic)

A  B  C  D  E
0.900 0.100 0.000 0.000 0.000
0.060 0.900 0.040 0.000 0.000
0.540 0.060 0.360 0.008 0.032
0.036 0.540 0.064 0.072 0.288
0.324 0.036 0.576 0.013 0.051
0.022 0.324 0.078 0.115 0.461
0.194 0.022 0.706 0.016 0.063
0.013 0.194 0.087 0.141 0.564

...
Rewards/costs produced by a plan

A plan produces an infinite sequence of rewards/costs $r_1, r_2, r_3, \ldots$.

Alternative ways of valuing a plan:

1. sum of all rewards over a horizon of length $n$: $r_1 + r_2 + \cdots + r_n$

2. average rewards $\lim_{N \to \infty} \frac{\sum_{i=1}^{N} r_i}{N}$

3. discounted rewards $r_1 + cr_2 + c^2r_3 + c^3r_4 + \cdots + c^{k-1}r_k + \cdots$
Outline of next lectures

- Algorithms for solving probabilistic planning
- Extension of BDD-based techniques to probabilistic problems
- Partial observability