Probabilistic planning: Quality criteria for plans

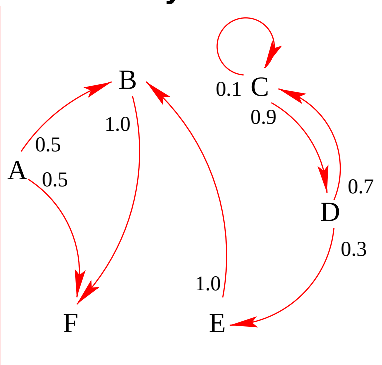
1. Plan reaches goals with probability 1.
2. Plan reaches goals with minimal expected cost.
3. Plan reaches goals with maximal probability.
4. Plan produces maximal expected rewards.

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



	A	B	C	D	E	F
A	0	0.5	0	0	0	0.5
B	0	0	0	0	0	1.0
C	0	0	0.1	0.9	0	0
D	0	0	0.7	0	0.3	0
E	0	1.0	0	0	0	0
F	0	0	0	0	0	0

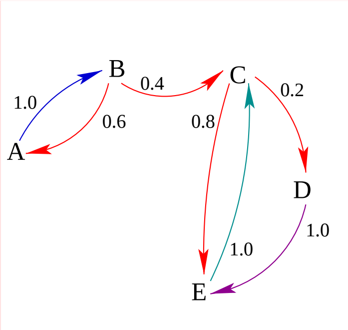
Probability distributions: example

Successors of state s (with $s \models a \wedge b \wedge c$) w.r.t. operator $\langle a, (0.1\neg a|0.9\neg b) \wedge (0.8\neg c|0.2c)\rangle$:

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

one successor state satisfies $\neg a \wedge b \wedge \neg c$, the second $a \wedge \neg b \wedge \neg c$, the third $\neg a \wedge b \wedge c$, the fourth $a \wedge \neg b \wedge c$.

Transition probabilities under a plan



$J = (0.9, 0.1, 0, 0, 0)^T$

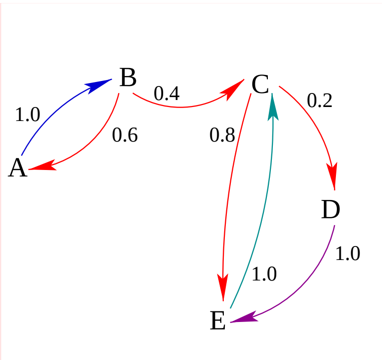
	A	B	C	D	E
A	0	1.0	0	0	0
B	0.6	0	0.4	0	0
C	0	0	0	0.2	0.8
D	0	0	0	0	1.0
E	0	0	1.0	0	0

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.

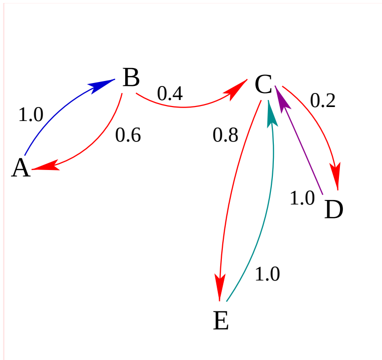
J	probability distribution initially
JM	after 1 action
JMM	after 2 actions
$JMMM$	after 3 actions
$\vdots$	
JM^i	after i actions

Probabilities of states under a plan



t	A	B	C	D	E
0	0.900	0.100	0.000	0.000	0.000
1	0.060	0.900	0.040	0.000	0.000
2	0.540	0.060	0.360	0.008	0.032
3	0.036	0.540	0.056	0.072	0.296
4	0.324	0.036	0.512	0.011	0.117
5	0.022	0.324	0.131	0.102	0.421
6	0.194	0.022	0.550	0.026	0.207
7	0.013	0.194	0.216	0.110	0.467
8	0.117	0.013	0.544	0.043	0.283
9	0.008	0.117	0.288	0.109	0.479
10	0.070	0.008	0.525	0.058	0.339
$\vdots$					
	0.000	0.000	0.455	0.091	0.455
	0.000	0.000	0.455	0.091	0.455

Probabilities of states under a plan (periodic)



t	A	B	C	D	E
0	0.900	0.100	0.000	0.000	0.000
1	0.060	0.900	0.040	0.000	0.000
2	0.540	0.060	0.360	0.008	0.032
3	0.036	0.540	0.064	0.072	0.288
4	0.324	0.036	0.576	0.013	0.051
5	0.022	0.324	0.078	0.115	0.461
6	0.194	0.022	0.706	0.016	0.063
7	0.013	0.194	0.087	0.141	0.564
$\vdots$					
	0.000	0.000	0.900	0.020	0.080
	0.000	0.000	0.100	0.180	0.720
	0.000	0.000	0.900	0.020	0.080
	0.000	0.000	0.100	0.180	0.720

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 $r_1 + cr_2 + c^2 r_3 + c^3 r_4 + \ldots + c^{k-1} r_k + \cdots$

Probabilistic planning: definition

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

- P is a finite set of state variables,
- I represents a probab. distribution on states (valuations of P),
- O is a set of operators on P, and
- R assigns every operator and valuation a reward (a real number).

Representing probab. distributions and rewards

I is a set $\{\langle \phi_1, p_1 \rangle, \langle \phi_2, p_2 \rangle, \ldots, \langle \phi_n, p_n \rangle\}$ that expresses a probability distribution over valuations of P. (Default is 0.0.)

We require that $\phi_i \models \neg \phi_j$ for every $\{i, j\} \subseteq \{1, \ldots, n\}$.

$R(o)$ for every $o \in O$ is a set $\{\langle \phi_1, r_1 \rangle, \langle \phi_2, r_2 \rangle, \ldots, \langle \phi_m, r_m \rangle\}$ that expresses the rewards obtained when o is applied: if o is applied in s and $s \models \phi_k$, then reward is r_k. (Default is 0.0.)

We require that $\phi_i \models \neg \phi_j$ for every $\{i, j\} \subseteq \{1, \ldots, m\}$.

Expressing goals in terms of rewards

Let G be a set of goal states. Modify the operators and the reward function as follows.

- Replace $\langle c, e \rangle$ by $\langle c \wedge \neg G, e \rangle$ and add operator $o_g = \langle G, \top \rangle$.
- Rewards are $R(o) = \{\langle G, 1.0 \rangle\}$ for all operators.
- Expected average reward is 1.0 iff goals always reached.
- For discounted rewards no exact correspondence.

Definition of stochastic transition systems

DEFINITION A *stochastic transition system with rewards* is a 4-tuple $\langle S, A, p, R \rangle$ where

- S is a finite set of states,
- A is a finite set of actions,
- p is a partial function that maps each state $s \in S$ and action $a \in A$ to a probability distribution on S (notation $p(s'|s, a)$),
- $R : S \times A \to \mathcal{R}$ is a reward function which maps each state $s \in S$ and action $a \in A$ to real number.

Stochastic transition systems are described by $\langle P, I, O, R \rangle$.