Lecture 10 - Planning under Uncertainty (I)

Jesse Hoey
School of Computer Science
University of Waterloo

July 3, 2022

Readings: Poole & Mackworth (2nd ed.)Chapter 9.1-9.3

Recall from lecture 8a

expected value of a function on X, $V(X)$:

$\mathbb{E}(V) = \sum_{x \in Dom(X)} P(x)V(x)$

where $P(x)$ is the probability that $X = x$.

This is useful in decision making, where $V(X)$ is the utility of situation X.

Bayesian Decision Making

Bayesian decision making is then

$\mathbb{E}(V(\texttt{decision})) = \sum_{outcome} P(outcome|decision)V(outcome)$

Can also add context so $V(\texttt{decision}, context)$ is the value of `decision` in situation `context`

$\mathbb{E}(V(\texttt{decision}, \texttt{context}) = \sum_{outcome} P(outcome|\texttt{decision}, \texttt{context})V(outcome)$

In this lecture, we will explore V, and then $\mathbb{E}(V)$

Preferences

- Actions result in outcomes
- Agents have preferences over outcomes
- A (decision-theoretic) rational agent will do the action that has the best outcome for them
- Sometimes agents don't know the outcomes of the actions, but they still need to compare actions
- Agents have to act (doing nothing is often a meaningful action).

Preferences Over Outcomes

If o_1 and o_2 are outcomes

- $o_1 \succeq o_2$ means o_1 is at least as desirable as o_2 (weak preference)
- $o_1 \sim o_2$ means $o_1 \succeq o_2$ and $o_2 \succeq o_1$. indifference
- $o_1 \succ o_2$ means $o_1 \succeq o_2$ and $o_2 \not\succeq o_1$ strong preference

Lotteries

- An agent may not know the outcomes of their actions, but only have a probability distribution of the outcomes.
- A lottery is a probability distribution over outcomes. It is written

  $$[p_1 : o_1, p_2 : o_2, \dots, p_k : o_k]$$

  where the o_i are outcomes and $p_i > 0$ such that

  $$\sum_i p_i = 1$$

  The lottery specifies: outcome o_i occurs with probability p_i.
- When we talk about outcomes, we will include lotteries.

Properties of Preferences

- Completeness: Agents have to act, so they must have preferences:

  $$\forall o_1 \forall o_2 \; o_1 \succeq o_2 \text{ or } o_2 \succeq o_1$$

- Transitivity: Preferences must be transitive:

  $$\text{if } o_1 \succeq o_2 \text{ and } o_2 \succeq o_3 \text{ then } o_1 \succeq o_3$$

- Monotonicity: An agent prefers a larger chance of getting a better outcome than a smaller chance:
  - If $o_1 \succ o_2$ and $p > q$ then

    $$[p : o_1, 1-p : o_2] \succ [q : o_1, 1-q : o_2]$$

Properties of Preferences (cont.)

Continuity: Suppose $o_1 \succ o_2$ and $o_2 \succ o_3$, then there exists a $p \in [0, 1]$ such that

$$o_2 \sim [p : o_1, 1-p : o_3]$$

See `worked example 1 video lecture10a-wx1`

Decomposability: (no fun in gambling). An agent is indifferent between lotteries that have same probabilities and outcomes.
Substitutability: if $o_1 \sim o_2$ then the agent is indifferent between lotteries that only differ by o_1 and o_2.

Theorem

If preferences follow the preceding properties, then preferences can be measured by a function

$$utility : outcomes \rightarrow [0, 1]$$

such that

- $o_1 \succeq o_2$ if and only if $utility(o_1) \geq utility(o2)$.
- Utilities are linear with probabilities:

$$utility([p_1 : o_1, p_2 : o_2, \ldots, p_k : o_k]) = \sum_{i=1}^{k} p_i \times utility(o_i)$$

(see proof in Book - proposition 9.3)

Rationality and Irrationality

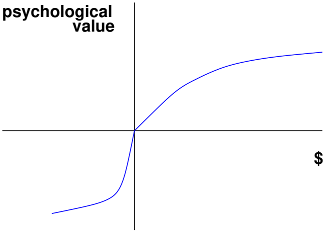
- Rational agents act so as to maximize expected utility:
  - Action a_1 leads to outcome $[o_1, \ldots, o_k]$ with probabilities $[p_1, p_2, \ldots, p_k]$
  - Action a_2 leads to outcome $[o_1, \ldots, o_k]$ with probabilities $[q_1, q_2, \ldots, q_k]$
  - if $\sum_{i=1}^{k} p_i \times utility(o_i) > \sum_{i=1}^{k} q_i \times utility(o_i)$ then action a_1 is the rational choice
- Humans are n not rational... What would you prefer

  $1,000,000 or [0.5 : $0, 0.5 : $2,000,000]?
- Would you prefer

  lose $100 or [0.5 : lose $0, 0.5 : lose $200]?

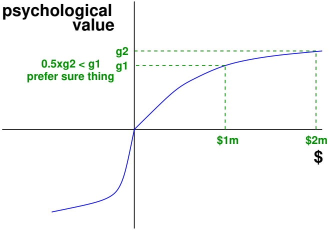
Prospect Theory - Tversky and Kahneman

Humans weight value differently for gains vs losses,



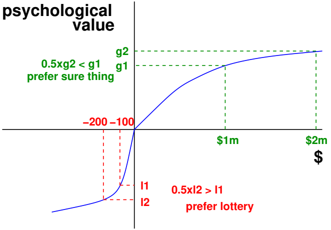
Prospect Theory - Tversky and Kahneman

Humans weight value differently for gains vs losses,
$1,000,000 or [0.5 : $0, 0.5 : $2,000,000]?
$g1$: psychological value of sure thing
$0.5 \times g2$: psychological value of lottery



Prospect Theory - Tversky and Kahneman

Humans weight value differently for gains vs losses,
lose $100 or [0.5 : lose $0, 0.5 : lose $200]
$l1$: psychological value of sure thing
$0.5 \times l2$: psychological value of lottery



Ultimatum Game

- Two-player game: agents A and B
- A gets $10
- A can offer B any amount $x = \$[0 - 10]$
- B can
  - accept: B gets x, A gets $10 - x$
  - reject: A and B both get 0
- rational choice: A offers B $\epsilon \to 0$, B accepts
- Humans: $x \approx \$4$

Making Decisions Under Uncertainty

What an agent should do depends on:

- The agent's ability — what options are available to it.
- The agent's beliefs — the ways the world could be, given the agent's knowledge. Sensing the world updates the agent's beliefs.
- The agent's preferences — what the agent actually wants and the tradeoffs when there are risks.

Decision theory specifies how to trade off the desirability and probabilities of the possible outcomes for competing actions.

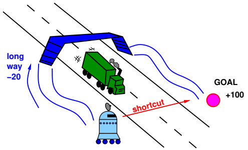
Single decisions

- Decision variables are like random variables that an agent gets to choose the value of.
- In a single decision variable, the agent can choose $D = d_i$ for any $d_i \in dom(D)$.
- Expected utility of decision $D = d_i$ leading to outcomes ω for utility function u
  $\mathcal{E}(u|D = d_i) = \sum P(\omega|D = d_i)u(\omega)$.
- An optimal single decision is the decision $D = d_{max}$ whose expected utility is maximal:

$$\mathcal{E}(u|D = d_{max}) = \max_{d_i \in dom(D)} \mathcal{E}(u|D = d_i).$$

Delivery Robot

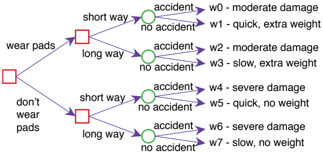
- To get to its goal, a robot can go one of two ways: a long, safe route and a shortcut.
- The robot can put on a set of pads before setting off.



- Goal is worth 100, taking the long route costs 20, and putting on pads costs 5.
- Accidents are costly, but less so if pads are worn.
- Accidents are more likely on the shortcut.

Decision Tree for Delivery Robot

The robot can choose to wear pads to protect itself or not.
The robot can choose to go the short way past the stairs or a long way that reduces the chance of an accident. Thus, the robot has two decision variables: Wear_Pads and Which_Way
There is one random variable of whether there is an accident.



Example Quantification

Which Way	Accident	Prob.
long	true	0.01
long	false	0.99
short	true	0.2
short	false	0.8

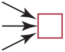
Which Way	Accident	Wear Pads	Value
long	true	true	30
long	false	true	75
long	true	false	0
long	false	false	80
short	true	true	35
short	false	true	95
short	true	false	3
short	false	false	100

Decision Networks

- A decision network is a graphical representation of a finite sequential decision problem.
- Decision networks extend belief networks to include decision variables and utility.
- A decision network specifies what information is available when the agent has to act.
- A decision network specifies which variables the utility depends on.

Decision Networks

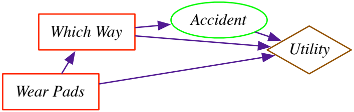
- A random variable is drawn as an ellipse. Arcs into the node represent probabilistic dependence.



- A decision variable is drawn as a rectangle. Arcs into the node represent information available when the decision is made.

- A utility node is drawn as a diamond. Arcs into the node represent variables that the utility depends on.

Single-stage decision networks: Robot Pads



$\mathcal{E}(which\_way, wear\_pads) =$
$\sum_{accident} P(accident|which\_way)U(which\_way, accident, wear\_pads)$

Finding the optimal decision

- Suppose the random variables are $X_1, \ldots, X_n$, decision variables are D, and utility depends on $X_{i_1}, \ldots, X_{i_k}$ and D:

$$\mathcal{E}(u|D) = \sum_{X_1,\ldots,X_n} P(X_1,\ldots,X_n|D) \times u(X_{i_1},\ldots,X_{i_k},D)$$
$$= \sum_{X_1,\ldots,X_n} \left[\prod_{j=1}^{n} P(X_j|parents(X_j))\right] \times u(X_{i_1},\ldots,X_{i_k},D)$$

  To find the optimal decision:
  - Create a factor for each conditional probability and for the utility
  - Multiply together and sum out all of the random variables
  - This creates a factor on D that gives the expected utility for each D
  - Choose the D with the maximum value in the factor.

Sequential Decisions

- An intelligent agent doesn't make a multi-step decision and carry it out without considering revising it based on future information.
- A more typical scenario is where the agent:
  observes, acts, observes, acts, . . .
- Subsequent actions can depend on what is observed.
  What is observed depends on previous actions.
- Often the sole reason for carrying out an action is to provide information for future actions.
  For example: diagnostic tests, spying.

Sequential decision problems

- A sequential decision problem consists of a sequence of decision variables $D_1, \ldots, D_n$.
- Each D_i has an information set of variables $parents(D_i)$, whose value will be known at the time decision D_i is made.

Policies

- A policy specifies what an agent should do under each circumstance.
- A policy is a sequence $\delta_1, \ldots, \delta_n$ of decision functions

$$\delta_i : dom(parents(D_i)) \rightarrow dom(D_i).$$

This policy means that when the agent has observed $O \in dom(parents(D_i))$, it will do $\delta_i(O)$.

Expected Utility of a Policy

- Possible world ω satisfies policy δ, written $\omega \models \delta$ if the decisions of the policy are those the world assigns to the decision variables. That is, each world assigns values to the decision nodes that are the same as in the policy.
- The expected utility of policy δ is

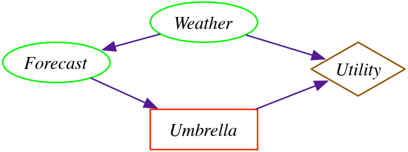
$$\mathcal{E}(u|\delta) = \sum_{\omega \models \delta} u(\omega) \times P(\omega),$$

- An optimal policy is one with the highest expected utility.

Finding the optimal policy

1. Create a factor for each conditional probability table and a factor for the utility.
2. Set remaining decision nodes $\leftarrow$ all decision nodes
3. Mutiply factors and sum out variables that are not parents of a remaining decision node.
4. Select and remove a decision variable D from list of remaining decision nodes:
   pick one that is in a factor with only itself
   and some of its parents (no children).
5. Eliminate D by maximizing. This returns:
   - the optimal decision function for D, $\arg\max_D f$
   - a new factor to use, $\max_D f$
6. Repeat 3-5 till there are no more remaining decision nodes.
7. Eliminate the remaining random variables. Multiply the factors: this is the expected utility of the optimal policy.
8. If any nodes were in evidence, divide by the P(evidence)

Umbrella Decision Network



You don't get to observe the weather when you have to decide whether to take your umbrella. You do get to observe the forecast.

Initial factors for the Umbrella Decision

Weather	Value
norain	0.7
rain	0.3

Weather	Fcast	Value
norain	sunny	0.7
norain	cloudy	0.2
norain	rainy	0.1
rain	sunny	0.15
rain	cloudy	0.25
rain	rainy	0.6

Weather	Umb	Value
norain	take	20
norain	leave	100
rain	take	70
rain	leave	0

Eliminating By Maximizing

f:

Fcast	Umb	Val
sunny	take	12.95
sunny	leave	49.0
cloudy	take	8.05
cloudy	leave	14.0
rainy	take	14.0
rainy	leave	7.0

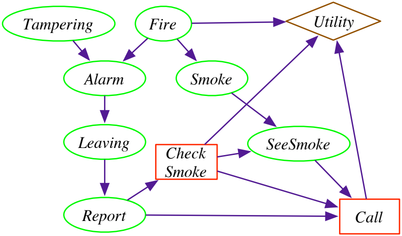
$\max_{Umb} f$:

Fcast	Val
sunny	49.0
cloudy	14.0
rainy	14.0

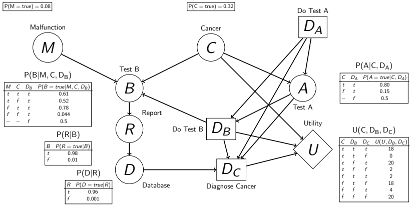
$\arg\max_{Umb} f$:

Fcast	Umb
sunny	leave
cloudy	leave
rainy	take

Decision Network for the Alarm Problem



Decision Network for the Cancer Problem



$P(B|M, C, D_B)$

M	C	D_B	$P(B = true \mid M, C, D_B)$
t	t	t	0.61
t	f	t	0.52
f	t	t	0.78
f	f	t	0.044
–	–	f	0.5

$P(R|B)$

B	$P(R = true \mid B)$
t	0.98
f	0.01

$P(D|R)$

R	$P(D = true \mid R)$
t	0.96
f	0.001

$P(A|C, D_A)$

C	D_A	$P(A = true \mid C, D_A)$
t	t	0.80
f	t	0.15
–	f	0.5

$U(C, D_B, D_C)$

C	D_B	D_C	$U(U, D_B, D_C)$
t	t	t	18
t	t	f	0
t	f	t	20
t	f	f	2
f	t	t	2
f	t	f	18
f	f	t	4
f	f	f	20

Next:

- Planning with uncertainty (Poole & Mackworth (2nd ed.)chapter 9.5)
- Reinforcement Learning (Poole & Mackworth (2nd ed.)chapter 12.1,12.3-12.9)