CS 188: Artificial Intelligence

Rational Decisions

Instructor: Stuart Russell and Peyrin Kao

University of California, Berkeley
Maximum Expected Utility

- **Principle of maximum expected utility:**
  - A rational agent should choose the action that maximizes its expected utility, given its knowledge.

- **Questions:**
  - Where do utilities come from?
  - How do we know such utilities even exist?
  - How do we know that averaging even makes sense?
  - What if our behavior (preferences) can’t be described by utilities?
The need for numbers

- For worst-case minimax reasoning, terminal value scale doesn’t matter
  - We just want better states to have higher evaluations (get the ordering right)
  - The optimal decision is invariant under any *monotonic transformation*

- For average-case expectimax reasoning, we need *magnitudes* to be meaningful
Utilities

- Utilities are functions from outcomes (states of the world) to real numbers that describe an agent’s preferences.

- Where do utilities come from?
  - In a game, may be simple (+1/-1)
  - Utilities summarize the agent’s goals
  - Theorem: any “rational” preferences can be summarized as a utility function

- We hard-wire utilities and let behaviors emerge
  - Why don’t we let agents pick utilities?
  - Why don’t we prescribe behaviors?
Utilities: Uncertain Outcomes

Getting ice cream

Get Single

Get Double

Oops

Whew!
An agent must have preferences among:

- Prizes: $A$, $B$, etc.
- Lotteries: situations with uncertain prizes
  \[ L = [p, A; (1-p), B] \]

Notation:

- Preference: $A > B$
- Indifference: $A \sim B$
<table>
<thead>
<tr>
<th>PLAN A</th>
<th>PLAN B</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ø</td>
<td>Ø</td>
</tr>
<tr>
<td>Ø</td>
<td>Ø</td>
</tr>
</tbody>
</table>
We want some constraints on preferences before we call them rational, such as:

Axiom of Transitivity: \((A > B) \land (B > C) \implies (A > C)\)

For example: an agent with intransitive preferences can be induced to give away all of its money

- If \(B > C\), then an agent with \(C\) would pay (say) 1 cent to get \(B\)
- If \(A > B\), then an agent with \(B\) would pay (say) 1 cent to get \(A\)
- If \(C > A\), then an agent with \(A\) would pay (say) 1 cent to get \(C\)
Rational Preferences

The Axioms of Rationality

Orderability:
\[(A > B) \lor (B > A) \lor (A \sim B)\]

Transitivity:
\[(A > B) \land (B > C) \Rightarrow (A > C)\]

Continuity:
\[(A > B > C) \Rightarrow \exists p \ [p, A; 1-p, C] \sim B\]

Substitutability:
\[(A \sim B) \Rightarrow [p, A; 1-p, C] \sim [p, B; 1-p, C]\]

Monotonicity:
\[(A > B) \Rightarrow (p \geq q) \iff [p, A; 1-p, B] \geq [q, A; 1-q, B]\]

Theorem: Rational preferences imply behavior describable as maximization of expected utility
Theorem [Ramsey, 1931; von Neumann & Morgenstern, 1944]

Given any preferences satisfying these constraints, there exists a real-valued function $U$ such that:

$$U(A) \geq U(B) \iff A \geq B$$

$$U([p_1, S_1; \ldots; p_n, S_n]) = p_1 U(S_1) + \cdots + p_n U(S_n)$$

I.e. values assigned by $U$ preserve preferences of both prizes and lotteries!

Optimal policy invariant under positive affine transformation $U' = aU + b$, $a > 0$

Maximum expected utility (MEU) principle:

Choose the action that maximizes expected utility

Note: rationality does not require representing or manipulating utilities and probabilities

- E.g., a lookup table for perfect tic-tac-toe
Human Utilities

Spin the wheel or pay $ to pass.
Utilities map states to real numbers. Which numbers?

Standard approach to assessment (elicitation) of human utilities:
- Compare a prize $A$ to a standard lottery $L_p$ between
  - “best possible prize” $u_T$ with probability $p$
  - “worst possible catastrophe” $u_\perp$ with probability $1-p$
- Adjust lottery probability $p$ until indifference: $A \sim L_p$
- Resulting $p$ is a utility in $[0,1]$

**Human Utilities**

- Pay $50
- Instant death

$0.999999 \sim 0.000001$

No change

Pay $50
Money

- Money *does not* behave as a utility function, but we can talk about the utility of having money (or being in debt).
- Given a lottery $L = [p, X; (1-p), Y]$
  - The *expected monetary value* $\text{EMV}(L) = pX + (1-p)Y$
  - The utility is $U(L) = pU(X) + (1-p)U(Y)$
  - Typically, $U(L) < U(\text{EMV}(L))$
  - In this sense, people are *risk-averse*
- E.g., how much would you pay for a lottery ticket $L = [0.5, 10,000; 0.5, 0]$?
  - The *certainty equivalent* of a lottery $\text{CE}(L)$ is the cash amount such that $\text{CE}(L) \sim L$
  - The *insurance premium* is $\text{EMV}(L) - \text{CE}(L)$
  - If people were risk-neutral, this would be zero!
Post-decision Disappointment: the Optimizer’s Curse

- Usually we don’t have direct access to exact utilities, only estimates
  - E.g., you could make one of $k$ investments
  - An unbiased expert assesses their expected net profit $V_1,\ldots,V_k$
  - You choose the best one $V^*$
  - With high probability, its actual value is considerably less than $V^*$

- This is a serious problem in many areas:
  - Future performance of mutual funds
  - Efficacy of drugs measured by trials
  - Statistical significance in scientific papers
  - Winning an auction

Suppose true net profit is 0 and estimate $\sim N(0,1)$; Max of $k$ estimates:
Utilities of Sequences
Utilities of Sequences

- What preferences should an agent have over prize sequences?

- More or less? [1, 2, 2] or [2, 3, 4]

- Now or later? [0, 0, 1] or [1, 0, 0]
Theorem: if we assume stationary preferences:
\[ [a_1, a_2, ...] > [b_1, b_2, ...] \iff [c, a_1, a_2, ...] > [c, b_1, b_2, ...] \]
then there is only one way to define utilities:

- **Additive discounted utility:**
  \[
  U([r_0, r_1, r_2, ...]) = r_0 + \gamma r_1 + \gamma^2 r_2 + ... \\
  \text{where } \gamma \in [0,1] \text{ is the discount factor}
  \]
Invariance for sequences

- Invariance for utilities (reminder):
  Optimal policy invariant under positive affine transformation \( U' = aU + b, \ a > 0 \)

- Invariance for rewards:
  Optimal policy is also invariant under potential transformation:
  \[ R'(s,a,s') = R(s,a,s') + \gamma \Phi(s') - \Phi(s) \]
  where \( \Phi \) is *any function of state*

- These *shaping rewards* can massively speed up RL

- Soccer example: \( R(s,a,s') = +3 \) for a win, \(+1\) for a draw, \(0\) for a loss
  \[ \Phi(s) = (100 \times \text{goal difference}) + (\text{distance to goal} / D) + 0.1(\text{possession}) \]
Decision Networks
Decision Networks

Umbrella

Weather

Forecast
Decision Networks

- Decision network = Bayes net + Actions + Utilities

  - Action nodes (rectangles, cannot have parents, will have value fixed by algorithm)
  - Utility nodes (diamond, depends on action and chance nodes)

- Decision algorithm:
  - Fix evidence $e$
  - For each possible action $a$
    - Fix action node to $a$
    - Compute posterior $P(W|e,a)$ for parents $W$ of $U$
    - Compute expected utility $\sum_w P(w|e,a) U(a,w)$
  - Return action with highest expected utility

Bayes net inference!
Example: Take an umbrella?

- Decision algorithm:
  - Fix evidence $e$
  - For each possible action $a$
    - Fix action node to $a$
    - Compute posterior $P(W|e,a)$ for parents $W$ of $U$
    - Compute expected utility of action $a$: $\sum_w P(w|e,a) U(a,w)$
  - Return action with highest expected utility

Umbrella = leave

$$EU(\text{leave}|F=\text{bad}) = \sum_w P(w|F=\text{bad}) U(\text{leave},w)$$

$$= 0.34 \times 100 + 0.66 \times 0 = 34$$

Umbrella = take

$$EU(\text{take}|F=\text{bad}) = \sum_w P(w|F=\text{bad}) U(\text{take},w)$$

$$= 0.34 \times 20 + 0.66 \times 70 = 53$$

Optimal decision = take!
Decision network with utilities on outcome states

| W         | P(W) | P(F=bad | W) |
|-----------|------|--------|
| sun       | 0.7  | 0.17   |
| rain      |      | 0.77   |

Here, U is a true utility.

With an action node as parent, it is sometimes called a **Q-value**
Value of information

- Suppose you haven’t yet seen the forecast
  - EU(leave | ) = 0.7x100 + 0.3x0 = 70
  - EU(take | ) = 0.7x20 + 0.3x70 = 35

- What if you look at the forecast?
  - If Forecast=good
    - EU(leave | F=good) = 0.89x100 + 0.11x0 = 89
    - EU(take | F=good) = 0.89x20 + 0.11x70 = 25
  - If Forecast=bad
    - EU(leave | F=bad) = 0.34x100 + 0.66x0 = 34
    - EU(take | F=bad) = 0.34x20 + 0.66x70 = 53
  - P(Forecast) = <0.65,0.35>
  - Expected utility given forecast
    - = 0.65x89 + 0.35x53 = 76.4
- Value of information = 76.4-70 = 6.4
Video of Demo Ghostbusters with VPI
Value of information contd.

- General idea: value of information = \textit{expected improvement in decision quality} from observing value of a variable
  - E.g., oil company deciding on seismic exploration and test drilling
  - E.g., doctor deciding whether to order a blood test
  - E.g., person deciding on whether to look before crossing the road

- Key point: decision network contains everything needed to compute it!

- \( VPI(E_i \mid e) = \left[ \sum_{e_i} P(e_i \mid e) \max_a \text{EU}(a \mid e_i, e) \right] - \max_a \text{EU}(a \mid e) \)
VPI Properties

VPI is non-negative! $\text{VPI}(E_i | e) \geq 0$

VPI is not (usually) additive: $\text{VPI}(E_i, E_j | e) \neq \text{VPI}(E_i | e) + \text{VPI}(E_j | e)$

VPI is order-independent: $\text{VPI}(E_i, E_j | e) = \text{VPI}(E_j, E_i | e)$