Planning With Uncertainty Decision Networks

Arnav Gupta

November 27, 2024

Contents

1 Preferences **1**
1.1 Properties of Preferences 2
2 Rationality and Irrationality **3**
3 Decisions **3**
4 Decision Networks **4**
5 Sequential Decisions **4**
6 Policies **5**

1 Preferences

Let $V(x)$ be the **utility** of the situation X. **Bayesian decision making** is

$$\mathbb{E}(V(\text{decision})) = \sum_{\text{outcome}} P(\text{outcome} \mid \text{decision})V(\text{outcome})$$

Context can also be added to V so the utility is the value of the decision in the situation context.

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

Actions result in outcomes and agents have **preferences** over outcomes. A decision-theoretic **rational** agent will do the action that has the best outcome for them. Agents have to act (even if doing nothing) and compare actions without knowing the outcomes of actions.

If o_1 and o_2 are outcomes

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

Lottery: probability distribution over outcomes, denoted

$$[p_1 : o_1, \ldots, p_k : o_k]$$

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

$$\sum_i p_i = 1$$

1.1 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: if $o_1 \succeq o_2$ and $o_2 \succeq o_3$, then $o_1 \succeq o_3$

Monotonicity: an agent prefers a larger chance of getting a better outcome than a smaller chance, so if $o_1 \succ o_2$ and $p > q$ then

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

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

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

Decomposibility: 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

If preferences follow these properties, they can be measured by the function

$$\text{utility} : \text{outcome} \to [0, 1]$$

such that

- $o_1 \succeq o_2$ if and only if $\text{utility}(o_1) \geq \text{utility}(o_2)$
- utilities are linear with probabilities

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

2 Rationality and Irrationality

Rational agents act to maximize expected utility.

Humans are not rational and weight value differently for gains and losses (prospect theory).

3 Decisions

Making a decision on what an agent should do depends on

- agent's **ability**: options available to it
- agent's **beliefs**: how the world could be given the agent's knowledge, updating by sensing world
- agent's **preferences**: what the agent actually wants and tradeoffs present when there are risks

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

Decision variables: random variables that an agent gets to choose the values of, specifically it can choose any value in the domain

Expected utility of a decision leads to outcomes ω for utility function u is

$$\mathcal{E}(u \mid D = d_i) = \sum P(\omega \mid D = d_i) u(\omega)$$

Optimal single decision: decision $D = d_{\max}$ whose expected utility is maximal:

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

4 Decision Networks

Graphical representation of a finite sequential decision problem. Extension of belief networks to include decision variables and utility.

A decision network specifies

- what info is available when the agent has to act
- which variables the utility depends on

A random variable is drawn as an ellipse, with arcs into the node representing probabilistic dependence. A decision variable is drawn as a rectangle, with arcs into the node representing info available when the decision is made. A utility node is drawn as a diamond, with arcs into the node representing variables that the utility depends on.

Assume random variables $X_1, \ldots, X_n$, decision variables D, and utility dependent on $X_{i_1}, \ldots, X_{i_k}$ and D:

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

where parents(X_j) may include D. To find the optimal decision:

- create a factor for each conditional probability and for the utility
- multiply these together and sum out all random variables, creating a factor on D that gives the expected utility for each D
- choose the D with the maximum value in the factor

5 Sequential Decisions

An intelligent agent observes and acts iteratively, where subsequent actions can depend on what is observed, which depends on previous actions.

Often an action is carried out to provide info for future actions.

Sequential decision problem: consists of a sequence of decision variables $D_1, \ldots, D_n$ where each D_i has an info set of variables parents(D_i) whose value will be known when decision D_i is made

6 Policies

Specifies what an agent should do under each circumstance.

A **policy** is a sequence $\delta_1, \ldots, \delta_n$ of decision functions

$$\delta_i : \text{dom}(\text{parents}(D_i)) \to \text{dom}(D_i)$$

which means that when the agent has observed $O \in \text{dom}(\text{parents}(D_i))$ it will do $\delta_i(O)$.

A world ω satisfies a policy δ, denoted $\omega \models \delta$ if the decisions of the policy are those the world assigns to the decision variables.

The **expected utility** of a policy δ is

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

Optimal policy comes from maximizing expected utility.

To find an optimal policy:

1. create a factor for each conditional probability table and a factor for the utility
2. set remaining decision nodes to all decision nodes
3. multiply factors and sum out variables that are not parents of a remaining decision node
4. select and remove a decision variable D from the list of remaining decision nodes, picking one that is in a factor with only itself and some of its parents (no children)
5. eliminate D by maximizing, which returns the optimal decision function for D, $\arg\max_D f$, and a new factor to use $\max_D f$
6. repeat steps 3 to 5 until there are no more remaining decision nodes
7. eliminate the remaining random variables by multiplying th factors, which is the expected utility of the optimal policy
8. if any nodes were in evidence, divide by the $P(\text{evidence})$