Computer Science Department - Rutgers University **Spring 2020**

CS 440: Lecture Notes on Decisions and Utility Theory

16:198:440

Instructor: Wes Cowan

At this point in the class, we have extensively discussed the question of how to represent knowledge (certain or uncertain) and how to work with or manipulate it to produce new knowledge. The question we're shifting to now is this: how do we make decisions based on that data? Given what we know, how do we decide to, for instance, skip class or attend class in the face of a global pandemic? And, as is always the case in this course, how can we formulate this in a way that allows computers to make decisions as well?

1 Rational Preference

We model the idea of decision making through *preference*: an agent may prefer option A to option B (or B to A) or be indifferent between them. We can denote this idea as $A \succ B$, $B \succ A$, or $A \sim B$ respectively. Note, A and B can be anything - we are simply denoting the idea that the agent in question prefers one to the other.

People are of course welcome to their own preferences (and each person's preferences may differ from everyone else's). But we can say something about what *rational* preference ought to look like. That is, if an agent is rational in their preferences, their preferences should obey some reasonable conditions, the *Axioms of Rational Preference*. I can never remember all of these off the top of my head, but they capture key ideas about what it should mean to be rational. Three to drive the point home are these:

- **Completeness:** For any A and B, either $A \succ B$, $B \succ A$, or $A \sim B$. That is, any two objects are comparable under a system of rational preference. If this weren't the case, there would be things and situations that the agent would be incapable of deciding between and ultimately paralyzed by inaction.
- **Transitivity:** If $A \succ B$ and $B \succ C$, then under a system of rational preference, $A \succ C$. That is, if A is preferred to B, and B is preferred to C, then under a system of rational preference, A should be preferred to C. If this isn't the case, then the agent can be exploited in the following way: suppose, having C, someone offers the agent B. If the agent prefers B to C, then there ought to be a small price they are willing to pay in order to trade. Having B, the agent is then offered A - again, if they agent prefers A to B, then there ought to be a small price they are willing to pay in order to trade. Having A, the agent is then offered C once more by the seller. If C is preferred to A, then for some small price the agent will be willing to exchange A for C, thus returning them to their original state less however much money they paid out. This cycle could repeat endlessly, extracting endless wealth from the agent. The only way to break out of this money pump is if the agent's preferences satisfy transitivity.
- **Continuity or Monotonicity:** Suppose that $A \succ B \succ C$. Consider then, the following lottery or game G_p: with probability p, the agent will receive A, but with probability $1 - p$ the agent will receive C. When $p = 1$, $G_1 \succ B$ (as G_1 is just A). However, when $p = 0$, it is clear that $B \succ G_0$ (as G_0 is just C). Under a system of rational preference, there is some threshold p^* such that if $p > p^*$, then $G_p \succ B$, and if $p < p^*$, then $B \succ G_p$.

While there are others, in general the axioms of rational preference are necessary and capture the idea that if a system of preference *does not* satisfy them, then schemes such as the money pump above are available to extract infinite resources from the subject.

2 Utility Functions

Given a system of rational preference, how can we model it, computationally, and equip a computer or AI with one? The main result that will allow us to move forward here is the existence of *utility functions*. We have the following:

If an agent has a rational system of preference among elements of a set S, then there exists a real valued **utility function** $U : S \mapsto \mathbb{R}$ such that

$$A \succ B \leftrightarrow U(A) > U(B), \tag{1}$$

and $A \sim B$ if and only if $U(A) = U(B)$.

In this way, the utility function allows us to numerically encode the system of preference (assigning to every option a utility), and in doing so translates preference comparisons to numerical comparisons.

Additionally, we have the following representational theorem: given a game or lottery G with the following outcomes, that A_i occurs with probability p_i, the utility of the game is given by the expected utility of its outcomes. That is,

$$U(G) = \sum_i p_i U(A_i). \tag{2}$$

In this way, complex or uncertain decisions can be reduced to looking at the expected utility of their outcomes. Our general approach to equipping AIs with decision making will therefore be to give it a utility functions, and then let it make decisions via maximizing the utility of its outcome or expected outcome.

2.1 What is Utility? Where does Utility come from?

A reasonable question at this point is this - what is the utility of an outcome? What does it represent? In one sense, it represents the value that we are willing to assign to a given thing. Frequently, we measure this in terms of money: if I am willing to pay \$20 for A, but only \$5 for B, then I prefer A to B because its value/utility to me is higher. Notice additionally that this suggest that utility can be *subjective* - the utility I assess may be different from the utility that you assess. The connection to expected utility then ties back to the subjectivity of probability as well, from a Bayesian perspective - if my assessment of belief/probabilities differs from yours, we may have different assessments of the utility of certain outcomes.

Consider the following game: I will flip a fair coin, and if it comes up heads, you win \$1, but if it comes up tails, you lose \$1. According to the representation theorem, we have that

$$U(G) = \frac{1}{2}U(\$1) + \frac{1}{2}U(-\$1). \tag{3}$$

In this case, if your utility of having a dollar is greater than your utility of losing a dollar, then the utility of the game is positive and it is probably worth playing. For most people, the prospect of winning and losing a dollar are about the same in magnitude and so the expected utility of this game should be around 0 - not a lot differentiates playing this game from choosing not to play the game.

However, consider modifying the game in the following way: I will flip a fair coin, and if it comes up heads, you win $\$10,000$, and if it comes up tails, you lose $\$10,000$. Most students I've posed this game to have declined to play - in essence, losing $\$10,000$ hurts more than winning $\$10,000$ would feel good. From the representation theorem, we see that

$$U(G) = \frac{1}{2}U(\$10,000) + \frac{1}{2}U(-\$10,000), \tag{4}$$

hence, if the overall utility is negative (the game is not preferred to not playing), then we see that $U(\$10,000)$ is actually smaller than $U(-\$10,000)$ is negative, giving an overall negative value. This numerically captures the idea of losing that much money hurting more. But if the agent is such that losing that much money isn't a serious issue (for instance if they have money to spare), the overall utility of the game may still be 0 or even positive.

Where do utility functions come from? How are they defined? In general, if x is an amount of money, it may be that $U(x) = x$, but as the above shows this may not necessarily be the case. As utility functions need to capture all the intricacies of a person's system of preference, they can be quite complicated. This is an active area of economic and psychological study (as is non-rational preference) but a frequent approach is to present a person with a system of comparisons (in much the spirit as the previous examples above) and ask which they prefer. The numeric utility function can then be approximated from the survey results.

However, we will generally take a much simpler approach to defining utility functions for AIs. Typically utility will be defined in terms of a simple function of expected rewards or returns (in the manner of $U(x) = x$ for x a monetary value). Some complications will arise in the treatment of multi-stage or sequential decisions (where the result of one decision may affect the outcomes of future decisions), but we will deal with this as they arise. In general, there are two ways of thinking about the construction of utility functions.

- **Descriptive:** In general, descriptive utility functions try to capture some objective assessment of value, such as assigning monetary values to outcomes, or points (as in games), etc. In these cases, some decision-environment is well-understood (cost/rewards are known, etc), and we are trying to figure out how best to make decisions in this environment.
- **Prescriptive:** In general, prescriptive utility functions occur when we know what kind of behavior we want (such as efficiently removing children from burning buildings, etc), and we are trying to build a computational framework to give rise to that behavior (such as penalizing actions that resulted in loss of property or life, and rewarding actions that result in saving lives). In these cases, we are essentially trying to formalize intuitively understood environments and behaviors so that AIs can process them.

Either way, we can introduce the general decision making principle of acting to achieve the *maximum expected utility (MEU)*. That is, at every step, take the action that produces the maximum value of $U(\text{action})$. We will then look at how this utility may be calculated in terms of what is known about the environment.

3 An Example and the Value of Information

Suppose that you are a professor of modern history at the University of Washington in Seattle, and that your area of research is on unsolved mysteries. You know that in 1971, DB Cooper hijacked a plane, extorted $\$200,000$ from authorities, and then jumped from the plane between Portland and Seattle, to never be seen again. Through extensive research, you've determined with certainty that DB Cooper buried his stash in one of 50 plots of land, where it has remained to this day. Fortunately for you, all 50 plots are up for sale.

You're faced with the following question: should you buy one of these plots in the hopes of finding DB Cooper's ransom money, or should you recognize that finding it is unlikely, and just save your money?

Translating this into the language of utility - what would the utility of buying a plot be vs of not buying a plot? Assuming that $U(x) = x$ for monetary values, we have that

$$U(\text{don't buy}) = 0. \tag{5}$$

To generalize a bit, suppose that \$$C$ is hidden somewhere among N plots, and each plot is on sale at price \$$p$. Assuming that each plot is equally likely to contain the ransom, we have that

$$\begin{aligned} U(\text{buy}) &= \mathbb{P}(\text{bought correct plot})U(\text{correct plot}) + \mathbb{P}(\text{bought incorrect plot})U(\text{incorrect plot}) \\ &= (1/N)U(\text{correct plot}) + (1 - 1/N)U(\text{incorrect plot}) \\ &= (1/N)(C - p) + (1 - 1/N)(-p) \\ &= C/N - p. \end{aligned} \tag{6}$$

In the above, we are taking the utility of a plot to be the money it contains, less the cost of buying it. We see that not buying has a utility of 0, and buying has a utility of $C/N - p$. Hence, as long as $C/N - p > 0$, or $p < C/N = \$4,000$ (in 1971 dollars), then buying one of the plots is a rational choice.

Note: We can also turn this reasoning on its head. If you are the one who is selling, it is irrational to sell a plot of land for less than \$4,000.

If $p = C/N$, then monetarily you should be indifferent to buying vs not buying - both actions have a utility of 0. Making a decision between them then may come down to more external factors that are not captured in a pure monetary utility function (such as fame/notoriety for putting the only unsolved case of air piracy in commercial aviation to rest). But an alternative utility function will capture this more precisely.

However, you realize that you could also pay one of your grad students to scope out one of the plots of land before buying it. Obviously this is risky (trespassing is a crime, and you don't want to give away a potential find) so you decide to have your student scope out at most one of the plots of land. If your student finds the ransom money, you'll buy the plot, but if the student doesn't find the ransom money, you know not to buy that plot and instead buy one of the others. What would the utility of this be?

$$\begin{aligned} U(\text{buy after search}) &= \mathbb{P}(\text{search success})U(\text{buy with success}) + \mathbb{P}(\text{search fail})U(\text{buy without success}) \\ &= \frac{1}{N}U(\text{buy with success}) + \left(1 - \frac{1}{N}\right)U(\text{buy without success}). \end{aligned} \tag{7}$$

If the search is successful, you get a total value of $C - p$ as before, $U(\text{buy with success}) = C - p$. If the search is not successful, though, then you know to only buy from the remaining $N - 1$ plots. Extrapolating from the previous case, the utility of this is going to be $U(\text{buy without success}) = C/(N - 1) - p$. This gives a total utility of

$$\begin{aligned} U(\text{buy after search}) &= \frac{1}{N}(C - p) + \left(1 - \frac{1}{N}\right)\left(\frac{C}{N-1} - p\right) \\ &= 2C/N - p \end{aligned} \tag{8}$$

Note that $U(\text{buy after search}) > U(\text{buy without search})$, in all situations - if it is rational to buy, it's definitely rational to have a student search the plot first! And notice as well, that this assessment is independent of whether or not the student actually finds anything. The act of testing a plot first is valuable, and rational.

This analysis can also be extended to consider how much you may have to pay your grad student to perform the test. If the grad student is going to charge you g, you can show that the total utility of buying after searching will be

$$U(\text{buy after search}) = 2C/N - p - g, \tag{9}$$

which is greater than $U(\text{buy without search})$ whenever $C/N > g$. That is, as long as your grad student would charge less than your expected return from buying the plot of land, it's worthwhile to send them.

4 Sequential or Multi-Stage Decisions

Where we will move next in lecture is to consider sequential or multi-stage decisions, where for instance the decision you make at one step influences what decisions you have available in the future. Utility can be complex in these situations, as the ‘expected outcome’ may depend on future decisions beyond the current one. In general, we will follow the general analysis / setting of **Markov Decision Processes** and this will serve as the basis for reinforcement learning.