CS 188 Introduction to Artificial Intelligence
Summer 2023 Note 1

These lecture notes are based on notes originally written by Nikhil Sharma. Revised and edited by Ramanan Abeyakaran.

Agents

In artificial intelligence, the central problem at hand is that of the creation of a rational **agent**, an entity that has goals or preferences and tries to perform a series of **actions** that yield the best/optimal expected outcome given these goals. Rational agents exist in an **environment**, which is specific to the given instantiation of the agent. Agents use sensors to learn about the environment and act on it using actuators. Take a simple checker's agent for example: the environment for a checkers agent is the virtual checkers board on which it plays against opponents, and its piece moves are the actions. Together, an environment and the agents that reside within it create a **world**.

A **reflex agent** is one that doesn't think about the consequences of its actions, but rather selects an action based solely on the current state of the world. These agents are typically outperformed by **planning agents**, which maintain a model of the world and use this model to simulate performing various actions. Then, the agent can determine hypothesized consequences of the actions and can select the best one. This is simulated "intelligence" in the sense that it's exactly what humans do when trying to determine the best possible move in any situation - by thinking ahead.

To define the task environment we use the **PEAS** (**P**erformance Measure, **E**nvironment, **A**ctuators, **S**ensors) description. The performance measure describes what utility the agent tries to increase. The environment summarizes where the agent acts and what affects the agent. The actuators and the sensors are the methods with which the agent acts on the environment and receives information from it.

The **design** of an agent heavily depends on the type of environment the agents acts upon. We can characterize the types of environments in the following ways.

- In *partially observable* environments, the agent does not have full information about the state and thus the agent must have an internal estimate of the state of the world. This is in contrast to *fully observable* environments, where the agent maintains full information about their state.
- *Stochastic* environments have uncertainty in the transition model, i.e. taking an action in a specific state may have multiple possible outcomes with varying associated probabilities. This is in contrast to *deterministic* environments, where taking an action in a state has a single outcome that is guaranteed to happen.
- In *multi-agent* environments the agent acts in the environments along with other agents. For this reason the agent might need to randomize its actions in order to avoid being "predictable" by other agents.
- If the environment does not change as the agent acts on it, then this environment is called *static*. This is in contrast to *dynamic* environments that change as the agent interacts with it.

- If an environment has *known physics*, then the transition model (even if stochastic) is known to the agent and it can use that when planning a path. If the *physics are unknown* the agent will need to take actions deliberately to learn the unknown dynamics.

Utilities

Throughout our discussion of rational agents, the concept of utility came up repeatedly. In games, for example, Utility values are generally hard-wired into the game, and agents use these utility values to select an action. We'll now discuss what's necessary in order to generate a viable utility function.

Rational agents must follow the **principle of maximum utility** - they must always select the action that maximizes their expected utility. However, obeying this principle only benefits agents that have **rational preferences**. To construct an example of irrational preferences, say there exist 3 objects, A, B, and C, and our agent is currently in possession of A. Say our agent has the following set of irrational preferences:

- Our agent prefers B to A plus $1
- Our agent prefers C to B plus $1
- Our agent prefers A to C plus $1

A malicious agent in possession of B and C can trade our agent B for A plus a dollar, then C for B plus a dollar, then A again for C plus a dollar. Our agent has just lost $3 for nothing! In this way, our agent can be forced to give up all of its money in an endless and nightmarish cycle.

Let's now properly define the mathematical language of preferences:

- If an agent prefers receiving a prize A to receiving a prize B, this is written $A \succ B$
- If an agent is indifferent between receiving A or B, this is written as $A \sim B$
- A **lottery** is a situation with different prizes resulting with different probabilities. To denote lottery where A is received with probability p and B is received with probability $(1-p)$, we write

$$L = [p, A;\ (1-p), B]$$

In order for a set of preferences to be rational, they must follow the five **Axioms of Rationality**:

- *Orderability*: $(A \succ B) \vee (B \succ A) \vee (A \sim B)$
  A rational agent must either prefer one of A or B, or be indifferent between the two.
- *Transitivity*: $(A \succ B) \wedge (B \succ C) \Rightarrow (A \succ C)$
  If a rational agent prefers A to B and B to C, then it prefers A to C.
- *Continuity*: $A \succ B \succ C \Rightarrow \exists p\, [p, A;\ (1-p), C] \sim B$
  If a rational agent prefers A to B but B to C, then it's possible to construct a lottery L between A and C such that the agent is indifferent between L and B with appropriate selection of p.
- *Substitutability*: $A \sim B \Rightarrow [p, A;\ (1-p), C] \sim [p, B;\ (1-p), C]$
  A rational agent indifferent between two prizes A and B is also indifferent between any two lotteries which only differ in substitutions of A for B or B for A.

- *Monotonicity*: $A \succ B \Rightarrow (p \geq q \Leftrightarrow [p, A;\ (1-p), B] \succeq [q, A;\ (1-q), B]$
  If a rational agent prefers A over B, then given a choice between lotteries involving only A and B, the agent prefers the lottery assigning the highest probability to A.

If all five axioms are satisfied by an agent, then it's guaranteed that the agent's behavior is describable as a maximization of expected utility. More specifically, this implies that there exists a real-valued **utility function** U that when implemented will assign greater utilities to preferred prizes, and also that the utility of a lottery is the expected value of the utility of the prize resulting from the lottery. These two statements can be summarized in two concise mathematical equivalences:

$$U(A) \geq U(B) \quad \Leftrightarrow \quad A \succeq B \tag{1}$$

$$U([p_1, S_1;\ \ldots\ ; p_n, S_n]) \quad = \quad \sum_i p_i U(S_i) \tag{2}$$

If these constraints are met and an appropriate choice of algorithm is made, the agent implementing such a utility function is guaranteed to behave optimally. Let's discuss utility functions in greater detail with a concrete example. Consider the following lottery:

$$L = [0.5, \$0;\ 0.5, \$1000]$$

This represents a lottery where you receive \$1000 with probability 0.5 and \$0 with probability 0.5. Now consider three agents A_1, A_2, and A_3 which have utility functions $U_1(\$x) = x$, $U_2(\$x) = \sqrt{x}$, and $U_3(\$x) = x^2$ respectively. If each of the three agents were faced with a choice between participting in the lottery and receiving a flat payment of \$500, which would they choose? The respective utilities for each agent of participating in the lottery and accepting the flat payment are listed in the following table:

Agent	**Lottery**	**Flat Payment**
1	500	500
2	15.81	22.36
3	500000	250000

These utility values for the lotteries were calculated as follows, making use of equation (2) above:

$$\begin{aligned}
U_1(L) &= U_1([0.5, \$0;\ 0.5, \$1000]) = 0.5 \cdot U_1(\$1000) + 0.5 \cdot U_1(\$0) = 0.5 \cdot 1000 + 0.5 \cdot 0 = \boxed{500} \\
U_2(L) &= U_2([0.5, \$0;\ 0.5, \$1000]) = 0.5 \cdot U_2(\$1000) + 0.5 \cdot U_2(\$0) = 0.5 \cdot \sqrt{1000} + 0.5 \cdot \sqrt{0} = \boxed{15.81} \\
U_3(L) &= U_1([0.5, \$0;\ 0.5, \$1000]) = 0.5 \cdot U_3(\$1000) + 0.5 \cdot U_3(\$0) = 0.5 \cdot 1000^2 + 0.5 \cdot 0^2 = \boxed{500000}
\end{aligned}$$

With these results, we can see that agent A_1 is indifferent between participating in the lottery and receiving the flat payment (the utilities for both cases are identical). Such an agent is known as **risk-neutral**. Similarly, agent A_2 prefers the flat payment to the lottery and is known as **risk-averse** and agent A_3 prefers the lottery to the flat payment and is known as **risk-seeking**.

Summary

In this note, we discussed how an agent interacts with the environment through its sensors and its actuators. The agent function describes what the agent does in all circumstances. Rationality of the agent means that the agent seeks to maximize their expected utility. We discussed utility and lotteries, and how to reason about them. Finally, we defined our task environments using PEAS descriptions.