CS 188   Introduction to Artificial Intelligence
Spring 2024   Note 1

Author (all other notes): Nikhil Sharma

Author (Bayes' Nets notes): Josh Hug and Jacky Liang, edited by Regina Wang

Author (Logic notes): Henry Zhu, edited by Peyrin Kao

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 interact with the environment and act on it using actuators. As a very simple example, the environment for a checkers agent is the virtual checkers board on which it plays against opponents, where piece moves are 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 - 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 has 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 different 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 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.