Artificial Intelligence: Reinforcement Learning

How do we design machines that are able to act at the same capability of a human being. This report attempts to address this question. Knowledge of elementary probability theory will be assumed.

Jacob Denson

Chapter 1

What is AI?

Artificial Intelligence is the science of designing intelligent agents – actors on an environment. How do we define intelligence? In this class we define it to be the capability to achieve goals in the world. We do not spend time philosophically adressing this question. This definition thus becomes a tool rather than a philosophy – we can actually measure our intelligence.

By using this definition, it becomes simple to test whether a machine is intelligent. Given a task, we define a performance metric. An intelligent machine is one that scores relatively high on this metric. Rationality is choosing the optimal decisions given what it knows in order to maximize the metric. An important biproduct of the definition is that we do not need to understand how an agent works in order to understand why it is intelligent: Any strategy is fair game. If you would rather understand how the computation of a human brain works, artificial intelligence is not the field for you - take up computational neuroscience instead.

To be more specific, we define an agent as anything which can percieve an environment and act accordingly. An agent acts through its actuators, and percieves with its perceptions extracted from the environment, as shown in the diagram below. A percept sequence is the sequence of perceptions that an agent percieves. This should be (and has to be) all that an agent needs to make a rational decision. An agent function maps a percept sequence to an action – it corresponds to the agent making a decision. The way an agent behaves induces a sequence of actions which we can evaluate with a performance metric. The metric should be designed such that a rational agent, which selects the action expected to maximize its performance metric given the percept sequence and all built-in knowledge the agent has, should behave completely correctly in its environment. Note that a rational agent is not omniscient – it does not need to act in the best way given full knowledge of the environment, but the best expected knowledge given previous knowledge.
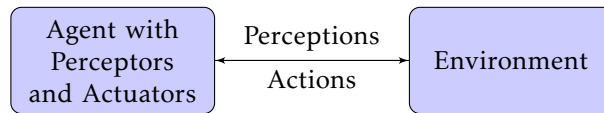
People may object to the definition established – a pocket calculator shouldn't be intelligent just because it responds intelligently to an input. We respond by stating that intelligence is relative, a pocket calculator is more intelligent that an abacus. Human intelligence is just the ability to be capable on the range of tasks a human is able to do well in.

One case study to aid in this understanding is Turing's definition of artificial intelligence via his 'Turing test'. A human interrogator attempts to determine whether the machine he is talking into has a human on the other side, or a robot impersonator. A machine is more intelligent than another if it tricks more people on average into thinking that it is a human. We do not care how the machine does it, as all aspects of the machine are hidden to the interrogator, but we do want the task to be performed as best as possible.

There are many aspects to building intellligence. We list the main topics here:

- Decision Making

- Learning

- Planning

- Representation of Knowledge

Early on, planning and representation were important, whereas recently decisions and learning have gained more importance. Though we will touch on the first two, we will mainly focus on decisions and learning, using the others to supplement the learning process.

1.1   Environments

Environments come in various shapes and sizes. We list various qualities here:

- A fully observable environment is an environment where all knowledge required and relevant to the agent is seen by its perceptors – that is, its whole state is known. A partially observable environment is where some relevant information is know, and a unobservable environment is where an agent has no sensors to detect the environment.

- A single agent environment contains one agent interacting with the environment. A multiagent environment contains numerous agents all interacting to achieve their goal. Agents should be considered actors that modify their actions to maximize a performance metric. Not all agents

may be following the same performance metric. In a multiagent environment, agents may be competitive, where an agents performance metric is maximized by minimizing all other agents metrics. Cooperative environments involve actions that maximize the performance metric of all other agents involve.

- In deterministic environments, an agents actions result in direct results in the environment. Stochastic or nondeterministic environments give uncertain results to an agents actions. This may be do to the random nature of the task or the unobtainable knowledge of the task.

- Episodic environments are environments where an agents percept sequence may be divided into atomic episodes where decisions in one episode do not impact the other. Continuous environments cannot be.

- An environment is static if it does not change when an agent is deciding actions to take. A dynamic environment can be seen as continually asking the agent to perform at all available times, and acting as if the agent chose to do nothing if it does not respond instantaneously.

- Discrete environments work in a finite or countable way. Continuous environments have the quality of a stream of states, action, or time associated with them.

The main focus in this class will be stochastic, single agent, discrete environments. Episodic and continuous tasks will both be considered.

Chapter 2

Utility?

Though the metric assigned to an agent's performance may not be given, the agent may still have a way to approximate this metric in order to perform better. A football team choose to play defensively when they know their metric is higher than the other teams. We call this a utility. Though we know the current metric in some problems in football match, this is not the fully calculate metric so is a utility. It would be almost unobtainable to get a full metric before decisions are made in all but the simplest environments.

One method an agent may attempt to obtain rationality is to rank various agent functions by a preference such that the top preference corresponds to an optimal score with the metric. Utility theory attempts to represent and reason with preferences. Decision theory takes uncertainty into account. The main problem is to choose the action that yields the hgihest expected utility, averaged over all outcomes of an outcome. Control theory takes a dynamic approach to maximizing the utility function over time. We will use many the ideas from these fields to organize the theory of reinforcement learning. Our first toy problem for introducing the concepts of reinforcement learning is the one-armed bandit problem.