CS 188 Introduction to Artificial Intelligence
Spring 2024 Note 7

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

Credit (Machine Learning and Logic notes): Some sections adapted from the textbook *Artificial Intelligence: A Modern Approach*.

Last updated: August 26, 2023

A Knowledge Based Agent

Imagine a dangerous world filled with lava, the only respite a far away oasis. We would like our agent to be able to safely navigate from its current position to the oasis.

In reinforcement learning, we assume that the only guidance we can give is a reward function which will try to nudge the agent in the right direction, like a game of 'hot or cold'. As the agent explores and collects more observations about the world, it gradually learns to associate some actions with positive future reward and others with undesirable, scalding death. This way, it might learn to recognize certain cues from the world and act accordingly. For example, if it feels the air getting hot it should turn the other way.

However, we might consider an alternative strategy. Instead let's tell the agent some facts about the world and allow it to reason about what to do based on the information at hand. If we told the agent that air gets hot and hazy around pits of lava, or crisp and clean around bodies of water, then it could reasonably infer what areas of the landscape are dangerous or safe based on its readings of the atmosphere. This alternative type of agent is known as a **knowledge based agent**. Such an agent maintains a **knowledge base**, which is a collection of logical **sentences** that encodes what we have told the agent and what it has observed. The agent is also able to perform **logical inference** to draw new conclusions.

The Language of Logic

Just as with any other language, logic sentences are written in a special **syntax**. Every logical sentence is code for a **proposition** about a world that may or may not be true. For example the sentence "the floor is lava" may be true in our agent's world, but probably not true in ours. We can construct complex sentences by stringing together simpler ones with **logical connectives** to create sentences like "you can see all of campus from the Big C *and* hiking is a healthy break from studying". There are five logical connectives in the language:

- $\neg$, **not**: $\neg P$ is true *if and only if (iff)* P is false. The atomic sentences P and $\neg P$ are referred to as **literals**.
- $\wedge$, **and**: $A \wedge B$ is true *iff* both A is true and B is true. An 'and' sentence is known as a **conjunction** and its component propositions the **conjuncts**.

- $\lor$, **or**: $A \lor B$ is true *iff* either A is true or B is true. An ’or’ sentence is known as a **disjunction** and its component propositions the **disjuncts**.

- $\Rightarrow$, **implication**: $A \Rightarrow B$ is true unless A is true and B is false.

- $\Leftrightarrow$, **biconditional**: $A \Leftrightarrow B$ is true *iff* either both A and B are true or both are false.

P	Q	$\neg P$	$P \land Q$	$P \lor Q$	$P \Rightarrow Q$	$P \Leftrightarrow Q$
false	*false*	*true*	*false*	*false*	*true*	*true*
false	*true*	*true*	*false*	*true*	*true*	*false*
true	*false*	*false*	*false*	*true*	*false*	*false*
true	*true*	*false*	*true*	*true*	*true*	*true*

Figure 7.8 Truth tables for the five logical connectives. To use the table to compute, for example, the value of $P \lor Q$ when P is true and Q is false, first look on the left for the row where P is *true* and Q is *false* (the third row). Then look in that row under the $P \lor Q$ column to see the result: *true*.