Intro to Artificial Intelligence

CSCI 2400 “Part 1”
Instructor: David Byrd

Disclaimer: Lecture notes can’t and won’t cover everything I say in class. You should attend class each day and use these for review or reinforcement.

1 AI Background

Artificial Intelligence (AI) is really hot. Computers are playing complex video games, there are at least some self-driving cars on public roads, and virtual personal assistants have finally become useful! Being in AI/ML now feels like being a web programmer in 1994 (start of the first dot-com boom). It’s a good place to be. But what *is* AI actually?

1.1 What is AI?

We can broadly think of AI as an entity that successfully performs behaviors that a person might reasonably believe require intelligence. This could mean reasoning about facts, making decisions without hard-coded logic, or finding patterns or “deeper meaning” in data. There is an issue with this answer: it assumes you already understand “intelligence”. This line of thinking is where Alan Turing’s “imitation game” (now called the Turing Test) originated. [Discussion: how is the usual presentation of the Turing Test wrong? Answer: It omits the presence of a human actor pretending to be a computer!]

1.2 Why do we need AI?

There is a *lot* of effort spent finding efficient ways to quickly solve interesting problems. In an Algorithms class, you might discuss Dynamic Programming and Memoization, learn to formally evaluate and improve the runtime of algorithms, and more. But there are also many useful problems for which there *is no known solution* in polynomial time (NP-Hard). For these problems, one *cannot* simply “compute” an answer beyond the size of a toy problem.

The basic goal of AI from a computational (not societal) perspective is then: to obtain reasonably accurate estimates to NP-Hard problems in a reasonable amount of time. For example, a self-driving car has potentially thousands of input variables that can together take on millions, billions, or even more unique combinations (sensors, fuel level, occupant preferences, location data, ...). It is unreasonable for the car to evaluate all possible combinations of input to precisely calculate “the correct answer” to the problem of what it should do, never mind to calculate such an answer many times per second! An algorithm buying and selling stocks faces the same problem: potentially almost any event in the world can affect stock prices, plus we are dealing with human emotions (greed and fear) and more. Even when the question is simple: "should I buy or sell SPY today?", the answer is incredibly unclear.

1.3 Super-brief History of AI

- Strong AI
  - First goal of artificial intelligence research community (c. 1955)
  - Emulate a human completely
    - General and broad, can do anything a human can do (even experience emotions like love)
    - Beat the Turing Test (and much more)
  - In the 1950s, many researchers assumed this problem would be solved by the 1970s
  - Not solved
    - Big promises and relatively small results led to the cutting of AI funding by DARPA
    - Start of First AI Winter (c. 1975)
- Expert Systems
  - Acknowledged the mistake of immediately trying to "build a human" from scratch
    - Even humans don't work that way. We learn slowly by experience and emulation.
  - Instead, just *model* and *emulate* human processes
  - Focus on rule-based systems designed by human experts
    - *e.g.* teach the computer everything you know about medical diagnosis
  - Brittle systems with little generalization (can't extrapolate to new situations)
    - Big promises and relatively small results led to reduced funding and interest
    - Start of Second AI Winter (c. 1987)
- Weak AI
  - Finally understand the difficulty of the problem
  - Focus on solving small, specific problems only
    - Narrow, tailored to one task
  - Automated
    - The real key! These are the first AI systems that did not have to be manually "taught" by humans every step of the way.

        * This is where most of the current industry work is (automating things formerly done by humans).

- Machine Learning
  - *Not the same thing* as Artificial Intelligence! A subset of AI.
  - AI that self-improves by finding patterns in data, and acting on that information.
  - "An algorithm that improves performance at a specific task by exposure to data."

1.4 Types of AI problems

We can break AI problems down along two axes: is the AI trying to think rationally, or think like a human; is it trying to act rationally, or act like a human? (Think about the humans you know and you may agree that "rationally" and "like a human" are very different, indeed!)

1.4.1 Think like a human

We could also call this Cognitive Science. Designing such an AI requires formulating a hypothesis of *how* the human mind works. Thinking "like a human" means mistakes can be acceptable (humans make them) and thoughts need not be rational (humans often are not: emotion, intuition).

1.4.2 Act like a human

Relaxing the prior constraint allows us to simply craft an AI program that *does* what a human *does*, without concern for the internal process. The mechanism of *doing* can also be different. If a human would put the orange ball in the basket, and the AI does that as well, then it has "acted like a human", even if the robot uses an electromagnet to move the ball instead of a "hand".

The Turing Test falls into this category as well. We are unconcerned with how or why the AI decides to respond (via text chat) in a particular way, only that the output is sufficiently "human" as to deceive a human judge. Computer games also typically fall into this category. We usually *do not want* a perfect game AI. It isn't fun! (Imagine an FPS bot that snipes you between the eyes 100% of the time at the microsecond you enter maximum weapon range.) We want an AI opponent that plays *like a human opponent* of some particular skill level.

1.4.3 Think rationally

We now drop the requirement of emulating humans entirely. Instead, we want to create an AI that performs logically with sound reasoning. Instead of Kirk, our AI should be Spock. We may want it to capture a formal representation of facts, then manipulate those facts in a valid way to prove theorems or deduce conclusions.

- A "think rationally" example:

  - Socrates is a man.
  - All men are mortal.
  - Mortal entities can be killed.
    * Could I kill Socrates? (Yes.)
    * Could I kill Plato? (No idea what Plato is, can't answer.)
    * Could I kill Joan of Arc? (No idea what Joan of Arc is. Even if you add fact "Joan of Arc is a woman", *still* no idea, because it does not know that men and women share important qualities like mortality.)

- Another:
  - All dogs have four legs.
  - Spot has three legs.
    * Is Spot a dog? (No.)

Actually, Spot is a dog. He just had an accident. How do we make the rational-thinking agent allow for this? "Some dogs have four legs." That's pretty much a useless fact. "All dogs have four legs at birth." That's still not necessarily true (birth defects exist). We'd end up writing some kind of bizarre legal contract of rules trying to define all the "edge cases", and this is just for one simple concept!

This type of AI is great for chess, checkers, or Go, but does not perform well in the messy real world due to all those exceptions.

1.4.4 Act rationally

So we drop any notion of caring how the AI thinks at all, and simply ask it to *do the right thing*. More formally, we expect that it will perform the optimal action under all situations. This requires some definition of "what is optimal?", meaning we now need a measurement of "goodness". If thinking rationally was *making correct inferences*, then acting rationally is *choosing the actions that lead to the best outcomes*. Acting rationally will often require memory and learning abilities. (There exist many deep neural networks that can reliably identify a dog when shown a picture of one, but none of them have *any idea* "what is a dog?" in any sense that we would understand. We've side-stepped the need to answer that question.)

Important: A rational-acting AI's actions need *not* be "correct" from an omniscient perspective. It is sufficient that they be optimal from its own limited perspective. That is, the AI must always make the best decision *based on the information it has*. (An AI-driven robot that drives off a cliff while trying to reach some goal *seems* irrational, but if the robot lacked any kind of cliff-sensor, then the AI's actions were still rational. It can't accommodate what it can't perceive.)

Later, in the Machine Learning portion of the course, we won't even ask that our agent behave *correctly*. It will be enough that it simply behave *reasonably*.

Next up: Agents and Environments.

Acknowledgements and thanks to Professors Mark Riedl and Jim Rehg of the Georgia Tech School of Interactive Computing.