CAP 5636 - Advanced Artificial Intelligence

Introduction

This slide-deck is adapted and extended from the one used by Chelsea Finn at CS221 at Stanford.

Motivating artificial intelligence

- It is generally not hard to motivate AI these days. There have been some substantial success stories. A lot of the triumphs have been in games, such as Jeopardy! (IBM Watson, 2011), Go (DeepMind's AlphaGo, 2016), Dota 2 (OpenAI, 2019), Poker (CMU and Facebook, 2019), ChatGPT (2023)
- On non-game tasks, we also have systems that achieve strong performance on reading comprehension, speech recognition, face recognition, and medical imaging benchmarks.
- Unlike games, however, where the game is the full problem, good performance on a benchmark does not necessarily translate to good performance on the actual task in the wild. Just because you ace an exam doesn't necessarily mean you have perfect understanding or know how to apply that knowledge to real problems.
- So, while promising, not all of these results translate to real-world applications.

Dangers of AI

- From the non-scientific community, we also see speculation about the future: that it will bring about sweeping societal change due to automation, resulting in massive job loss, not unlike the industrial revolution, or that AI could even surpass human-level intelligence and seek to take control.
    - Prominent people warning about AI:
        - industry: Elon Musk, Bill Joy
        - science: Stephen Hawking, Max Tegmark
        - philosophers: Nick Bostrom
- While these are extreme views, there is no doubt that AI is and will continue to be transformational. We still don’t know exactly what that transformation will look like.

History of AI

- How did we get here? The name artifical intelligence goes back to a summer in 1956. John McCarthy, who was then at MIT but later founded the Stanford AI lab, organized a **workshop at Dartmouth College** with the leading thinkers of the time, and set out a very bold proposal...to build a system that could do it all
  - Attendees: John McCarthy, Marvin Minsky, Claude Shannon, etc.
  - Aim for general principles: Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it.

Early era (1950-1960s)

- **Checkers** (1952): Samuel’s program learned weights and played at strong amateur level
- **Problem solving** (1955): Newell & Simon’s Logic Theorist: prove theorems in Principia Mathematica using search + heuristics; later, General Problem Solver (GPS)

Perspective on the early successes

- While they did not solve it all, there were a lot of **interesting programs** that were created: programs that could play checkers at a strong amateur level, programs that could prove theorems.
- For one theorem Newell and Simon's Logical Theorist actually found a proof that was more elegant than what a human came up with. They actually tried to publish a paper on it but it got rejected because it was not a new theorem; perhaps they failed to realize that the third author was a computer program.
- From the beginning, people like John McCarthy sought **generality**, thinking of how commonsense reasoning could be encoded in logic. Newell and Simon's General Problem Solver promised to solve any problem (which could be suitably encoded in logic).

Overwhelming optimism...

It was a time of high optimism, with all the leaders of the field, all impressive thinkers, predicting that AI would be "solved" in a matter of years.

- Machines will be capable, within twenty years, of doing any work a man can do. (Herbert Simon)
- Within 10 years the problems of artificial intelligence will be substantially solved. (Marvin Minsky)
- I visualize a time when we will be to robots what dogs are to humans, and I’m rooting for the machines. (Claude Shannon)

...underwhelming results: AI Winter #1

- Despite some successes, certain tasks such as machine translation were complete failures. Example:
  - English: The spirit is willing but the flesh is weak
  - Russian: The vodka is good but the meat is rotten
- 1966: ALPAC report cut off government funding for machine translation, first AI winter
  - Won't be the last.

What went wrong?

- It turns out that the real world is very complex and most AI problems require a lot of compute and data.
- The hardware at the time was simply too limited both compared to the human brain.
- Also, casting problems as general logical reasoning meant that the approaches fell prey to the exponential search space, which no possible amount of compute could really fix.
    - Even if you had infinite compute, AI would not be solved. There are simply too many words, objects, and concepts in the world, and this information has to be somehow encoded in the AI system.

Contributions of the early era

- Though AI was not solved, a few generally useful technologies came out of the effort, such as Lisp
    - some claim it as still the world's most advanced programming language in a sense
- One particularly powerful paradigm is the separation between what you want to compute (**modeling**) and how to compute it (**inference**)

Knowledge-based systems (70-80s)

In the seventies and eighties, AI researchers looked to knowledge as a way to combat both the limited computation and information problems. If we could only figure out a way to encode prior knowledge in these systems, then they would have the necessary information and also have to do less compute.

Expert systems

Instead of the solve-it-all optimism from the 1950s, researchers focused on building narrow practical systems in targeted domains. These became known as expert systems. The idea was to elicit specific domain knowledge from experts in form of rules:

```
    if [premises] then [conclusion]
```

- DENDRAL: infer molecular structure from mass spectrometry
- MYCIN: diagnose blood infections, recommend antibiotics
- XCON: convert customer orders into parts specification; saved DEC $40 million a year by 1986

AI Winter #3

Contributions:

- First real application that impacted industry
- Knowledge helped curb the exponential growth of choices

Problems:

- Knowledge is not deterministic rules, need to model uncertainty
- Requires considerable manual effort to create rules, hard to maintain (**knowledge engineering**)

The technology ran into limitations and failed to scale up to more complex problems. Due to plenty of overpromising and underdelivering, the field collapsed again.

Artificial neural networks

- Much of AI's history was dominated by the **logical tradition**, but there was another smaller camp, grounded in **neural networks** inspired by the brain.
- (Artificial) neural networks were introduced by a famous paper by McCulloch and Pitts, who devised a simple mathematical model and showed how it could be be used to compute arbitrary logical functions.
- Much of the early work was on understanding the mathematical properties of these networks, since computers were too weak to do anything interesting.

AI winter #2

- 1969: Perceptrons: A book by Marvin Minsky and Seymour Papert
- Explored many mathematical properties of Perceptrons (linear models) and showed that they could not solve some simple problems such as XOR.
- Even though this result says nothing about the capabilities of deeper networks, the book is largely credited with the demise of neural networks research, and the continued rise of logical AI.

First rebirth of neural networks

- In the 1980s, there was a renewed interest in neural networks.
- 1986: Rumelhardt, Hinton, Williams popularized **backpropagation**, a method for training multi-layer networks
- 1989 Yann LeCun built a system based on convolutional neural networks to recognize handwritten digits.
    - This was one of the first successful uses of neural networks, which was then deployed by the USPS to recognize zip codes

AI winter #4

- After the early successes of neural networks in mid-1980s, progress stalled
- A theoretical result showing that a neural network with a single hidden layer has the same expressive power as a network with arbitrary number of layers discouraged work in deep networks
  - Not the first (or last) case that misinterpreted theoretical results are an obstacle to progress
- It turned out that you can have systems that are much simpler to train, but they have better classifier performance (eg. support vector machines, inspired by statistical theory)

Second rebirth of neural networks (2012-current)

- Hindsight, what turned out that the limiting factor was is
  - Too little data
  - Too little computational power.
- The real break for neural networks came in the 2010s. With the rise of compute (notably GPUs) and large datasets such as ImageNet (2009), the time was ripe for the world to take note of neural networks.

The deep learning revolution

- AlexNet (2012) was a pivotal system that showed the promise of deep convolutional networks on ImageNet, the benchmark created by the computer vision community who was at the time still skeptical of deep learning.
- Many other success stories in speech recognition and machine translation followed
- AlphaGo (2016): deep reinforcement learning, defeat world champion Lee Sedol

Transformers and large language models

- A technical advance (Vaswani et al. Attention is all you need 2017) proposed **transformers** a combination of attention models and query architecture that turned out to learn better than previous models.
- Allowed the creation of a set of **large language models**: neural networks that are trained on extensive amount of language data, normally scraped from the internet.
- Very high performance on natural language tasks. Very good at combining and adapting learned knowledge.

The next AI winter: is it possible?

- This time we have real money making applications.
- Significant commitments of money, human effort and mindshare from:
    - Universities / Students
    - All the largest technology companies
    - Significant investment from venture capital and other investment funds
    - Nation states and their military and intelligence agencies
    - The open source community

The next AI winter: possible causes for it

- Societal opposition to AI
  - Almost unanimously negative treatment of AI in science fiction, for example.
  - No clear narrative of how we will use advanced AI for good
- Plateauing of progress
- Lack of progress as scaling LLMs sucks up all the creative energy
- Further progress becomes to expensive in terms of compute power
- Investment collapse as many investments do not generate expected return because:
  - Performance targets are not reached
  - Expenses higher then expected
  - Cheaper (or even free) competition

Two intellectual traditions

- Reflecting back on the past of AI, there have been two intellectual traditions that have dominated the scene: **logic** and **neuroscience** (at least initially).
- This debate is paralleled in cognitive science with connectionism and computationalism.
- While there are deep philosophical differences, perhaps there are deeper connections.
  - For example, McCulloch and Pitts' work from 1943 can be viewed as the root of deep learning, but that paper is mostly about how to implement logical operations.
  - The game of Go (and indeed, many games) can be perfectly characterized by a set of simple logic rules. At the same time, the most successful systems (AlphaGo) do not tackle the problem directly using logic, but appeal to the fuzzier world of artificial neural networks

AI is a melting pot of ideas

- Bayes rule (Bayes, 1763) from **probability**
- Least squares regression (Gauss, 1795) from **astronomy**
- First-order logic (Frege, 1893) from **logic**
- Maximum likelihood (Fisher, 1922) from **statistics**
- Artificial neural networks (McCulloch/Pitts, 1943) from **neuroscience**
- Minimax games (von Neumann, 1944) from **economics**
- Stochastic gradient descent (Robbins/Monro, 1951) from **optimization**
- Uniform cost search (Dijkstra, 1956) from **algorithms**
- Value iteration (Bellman, 1957) from **control theory**

Often, it is the new connections between these fields that are made and their application to important real-world problems that makes working on AI so rewarding.