COMPSCI 270: Artificial Intelligence

http://www2.cs.duke.edu/courses/spring19/compsci270/

Introduction

Instructor: Vincent Conitzer
Basic information about course

• TuTh 10:05-11:20am, French Science 2231
• Text: Artificial Intelligence: A Modern Approach
• Our amazing TA team:

   Alper  Michael  Sahiti  Alex  Carolyn
   Colter  Faris  Gideon  Peter  Will
Prerequisites

• Comfortable *programming* in general-purpose programming language – we’ll do assignments in Python and have a bootcamp

• Some knowledge of *algorithmic concepts* such as running times of algorithms

• Ideally, some familiarity with *probability* (we will go over this from the beginning but we will cover the basics only briefly)

• Not scared of *mathematics*; ideally, some background in discrete mathematics, able to do simple mathematical proofs
Grading

• Assignments: 30%
  – May discuss with another person; writeup and code must be your own
• Midterm exams: 40%
• Final exam: 30%
Some highly visible recent AI successes in games

Watson defeats Jeopardy champions (2011)


AlphaGo defeats Go champion (2016)

CMU’s Libratus defeats top human poker players (2017)
Typical picture in news articles

BusinessInsider reporting on the poker match…
Worries about AI - superintelligence

Nick Bostrom writes

Elon Musk influences
donates to

Yuval Noah Harari (Oct 2018): “for every dollar and every minute we invest in improving AI, we would be wise to invest a dollar and a minute in exploring and developing human consciousness.”

Max Tegmark writes

is co-founded by

future of life institute
Hard problems of consciousness – one perspective

1: world with creatures simulated on a computer

2: displayed perspective of one of the creatures

- To get from 1 to 2, need additional code to:
  - A. determine in which real-world colors to display perception
  - B. which agent’s perspective to display

- Is 2 more like our own experience than 1? If so, are there further facts about consciousness, perhaps beyond physics as we currently understand it?
Worries about AI - near term

- technological unemployment

- Intermediate stage en route to driverless cars.

- Autonomous vehicles – legal and other issues

- Autonomous weapon systems...

---

OH MY GOD, WHY DO YOU EVEN HAVE ALL THESE?
WHAT'S WRONG WITH YOU?
WE'RE LAUNCHING THEM INTO THE SUN.

THE MOMENT THE COMPUTERS CONTROLLING OUR NUCLEAR ARSENALS BECAME SENTIENT...
Artificial intelligence: where's the philosophical scrutiny?

AI research raises profound questions—but answers are lacking

by Vincent Conitzer / May 4, 2016 / Leave a comment

Natural Intelligence Still Has Its Advantages

AI is disruptive, but it hasn’t rendered humanity obsolete.

By Vincent Conitzer
Aug 28, 2018 6:27 p.m. ET

The late Stephen Hawking warned that artificial intelligence could “spell the end of the human race.” Already, AI systems are starting to outperform people in domains ranging from board games to speech recognition. Is humanity on the way out?

For those not working in AI, it can be difficult to interpret achievements in the field. Take, for example, Watson’s 2011 victory over human “Jeopardy!” champions Brad Rutter and Ken Jennings. This was a stunning achievement because while it should surprise nobody that Watson had access...
What is artificial intelligence?

• Popular conception driven by science fiction
  – Robots good at everything except emotions, empathy, appreciation of art, culture, …
    • … until later in the movie.

• Current AI is also bad at lots of simpler stuff!

• There is a lot of AI work on thinking about what other agents are thinking
Real AI

• A serious science.

• **General-purpose AI** like the robots of science fiction is incredibly hard
  – Human brain appears to have lots of special and general functions, integrated in some amazing way that we really do not understand (yet)

• **Special-purpose AI** is more doable (nontrivial)
  – E.g., chess/poker/Go playing programs, logistics planning, automated translation, speech and image recognition, web search, data mining, medical diagnosis, keeping a car on the road, ... ... ... ... ...
Definitions of AI

<table>
<thead>
<tr>
<th>Systems that think like humans</th>
<th>Systems that think rationally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systems that act like humans</td>
<td>Systems that act rationally</td>
</tr>
</tbody>
</table>

- We will mostly follow “act rationally” approach
  - Distinction may not be that important
    - acting rationally/like a human presumably requires (some sort of) thinking rationally/like a human,
    - humans much more rational anyway in complex domains

focus on action
sidesteps philosophical
issues such as “is the
system conscious” etc.

if our system can be more rational than humans in some cases, why not?
“Chinese room” argument [Searle 1980]

- Person who knows English but not Chinese sits in room
- Receives notes in Chinese
- Has systematic English rule book for how to write new Chinese characters based on input Chinese characters, returns his notes
  - Person=CPU, rule book=AI program, really also need lots of paper (storage)
  - Has no understanding of what they mean
  - But from the outside, the room gives perfectly reasonable answers in Chinese!
- Searle’s argument: the room has no intelligence in it!
Turing Test

- (Human) judge communicates with a human and a machine over text-only channel,
- Both human and machine try to act like a human,
- Judge tries to tell which is which.
- Numerous variants
- Loebner prize
  - 2016-2018 winner: [http://www.square-bear.co.uk/mitsuku/nfchat.htm](http://www.square-bear.co.uk/mitsuku/nfchat.htm)

Turing Test on unsuspecting judges

• It is possible to (temporarily) fool humans who do not realize they may be talking to a bot

• ELIZA program [Weizenbaum 66] rephrases partner’s statements and questions (~psychotherapist)
five: just asking if richard can tell me about some good lobster diving spots that's it if this is a problem then forget iot i dont know what's going on here

AOLiza: that is interesting. please continue.

five: continue what just ask richard when he gets home about some lobster diving up hat way that's all if you got a problem with it then forget it, this isn't worth it i dont know what's been going on or what bev has told you but i havent done anything to anybody iam the one that got used and dumped

AOLiza: tell me more about that.

five: about what

AOLiza: what do you think?

five: there's nothing to tell she used me then got tired of me said i wasn't her type anymore i was possessive and controlling

AOLiza: perhaps i already know you were possessive and controlling.

five: i was huh you nuts i gave her all the room she wanted never told nhexerre what to do or where to go
Is Turing Test the right goal?

• “Aeronautical engineering texts do not define the goal of their field as making ‘machines that fly so exactly like pigeons that they can fool even other pigeons.’” [Russell and Norvig]
Lessons from AI research

• Clearly-defined tasks that we think require intelligence and education from humans tend to be doable for AI techniques
  – Playing chess, drawing logical inferences from clearly-stated facts, performing probability calculations in well-defined environments, …
  – Although, scalability can be a significant issue

• Complex, messy, ambiguous tasks that come naturally to humans (in some cases other animals) are much harder…

• … though recent years have seen remarkable progress, especially in machine learning for narrow domains
  – Image recognition, speech recognition, reinforcement learning in computer games, self-driving cars

• AI systems still lack: broad understanding of the world, common sense, ability to learn from very few examples, truly out-of-the-box creativity…

• We don’t understand consciousness. (Does it matter for AI?)
Some areas where humans shine

• Coming up with reasonably good solutions in complex messy environments

• Adapting/self-evaluation/creativity (“My usual approach to chess is getting me into trouble against this person… Why? Is there something entirely different I can do?”)

• Analogical reasoning, transfer learning (applying insights from one domain to another)

• Explaining our reasoning

• Tasks that require a broad understanding of the (human) world

• Knowing what it’s like to be human

• Humor

• …
Early history of AI

• 50s/60s: Early successes! AI can draw logical conclusions, prove some theorems, create simple plans… Some initial work on neural networks…

• Led to overhyping: researchers promised funding agencies spectacular progress, but started running into difficulties:
  – **Ambiguity**: highly funded translation programs (Russian to English) were good at syntactic manipulation but bad at disambiguation
    • “The spirit is willing but the flesh is weak” becomes “The vodka is good but the meat is rotten”
  – **Scalability/complexity**: early examples were very small, programs could not scale to bigger instances
  – Limitations of **representations** used
History of AI...

- 70s, 80s: Creation of expert systems (systems specialized for one particular task based on experts’ knowledge), wide industry adoption
- Again, overpromising...
- … led to AI winter(s)
  - Funding cutbacks, bad reputation
Modern AI

- More rigorous, scientific, formal/mathematical
- Fewer grandiose promises
- Divided into many subareas interested in particular aspects
- More directly connected to “neighboring” disciplines
  - Theoretical computer science, statistics, economics, operations research, biology, psychology/neuroscience, …
  - Often leads to question “Is this really AI”? 
- Some senior AI researchers are calling for re-integration of all these topics, return to more grandiose goals of AI
  - Somewhat risky proposition for graduate students and junior faculty…
Some AI videos

• Note: there is a lot of AI that is not quite this “sexy” but still very valuable!
  – E.g. logistics planning – DARPA claims that savings from a single AI planning application during 1991 Persian Gulf crisis more than paid back for all of DARPA’s investment in AI, ever. [Russell and Norvig]

• https://www.youtube.com/user/aaaivideocompetition
• https://www.youtube.com/watch?v=1JJsBFiXGl0
• https://www.youtube.com/watch?v=s6VIWDUHTa4
• https://www.youtube.com/watch?v=C5Xnxjq63Zg
• https://www.youtube.com/watch?v=ScXX2bndGJc
• https://www.youtube.com/watch?v=V1eYniJ0Rnk
This course

• Focus on general AI techniques that have been useful in many applications

• Will try to avoid application-specific techniques (still interesting and worthwhile!)

• Will try not to overlap with the machine learning course
Topics (and examples)

• Search
  – Solving a Rubik’s cube

• Constraint satisfaction/optimization problems
  – Scheduling a given set of meetings (optimally)

• Game playing
  – Playing chess or poker

• Logic, knowledge representation
  – Solving logic puzzles, proving theorems

• Planning
  – Finding a schedule that will allow you to graduate (reasoning backwards from the goal)

• Probability, decision theory, reasoning under uncertainty
  – Given some symptoms, what is the probability that a patient has a particular condition? How should we treat the patient?

note overlap among topics...
Lots of AI at Duke!

- **Vince Conitzer** (multiagent systems, game theory, economics)
- **Ron Parr** (reasoning under uncertainty, reinforcement learning, robotics)
- **Carlo Tomasi** (computer vision, medical imaging)
- **Cynthia Rudin** (machine learning (especially interpretable ML), data mining, knowledge discovery)
- **Alex Hartemink** (computational biology, machine learning, reasoning under uncertainty)
- **Bruce Donald** (computational biology & chemistry)
- **Kris Hauser** (robot motion planning and control, integrating perception and planning)
- **Missy Cummings** (humans+AI, unmanned vehicles)
- **Guillermo Sapiro** (computer vision, image analysis, machine learning)
- **Larry Carin** (machine learning, neural networks)
- **Katherine Heller** (statistics, machine learning)
- **Sayan Mukherjee** (statistics, machine learning)
- ...
Getting more involved

• Duke Robotics, Intelligence, and Vision (DRIV) seminar (=AI seminar)
  – Website: http://driv.cs.duke.edu/
  – Mailing list: https://lists.duke.edu/sympa/subscribe/drive

• Machine Learning at Duke
  – Website: http://machinelearning.duke.edu/
  – Mailing list: https://lists.duke.edu/sympa/subscribe/machine-learning

• “Moral AI”
  – http://moralai.cs.duke.edu/
  – https://lists.duke.edu/sympa/subscribe/moral_ai