4COSC008C
Trends in Computer Science

Week 8: Artificial Intelligence

Computer Science

- Computer Science is a rapidly changing field
- The way we produce and consume Data is changing rapidly.

- However, past challenges in Science and Technology remain
  - understanding the human brain
    - reasoning, cognition, creativity
  - Creating intelligent machines
    - Is this Possible? Feasible? Desirable?
    - What are the scientific, technical and ethical challenges?
  - arguably AI still poses the most interesting challenges and questions in Computer Science today, and relates more and more in the way we collect, manipulate and draw conclusions from Data

Today we will introduce:

- Intelligence and Artificial intelligence.
- Ethical concerns related to Artificial Intelligence

You will find useful to consider

Russell, S. J. and Norvig, P. (2016) *Artificial Intelligence: a modern approach* Boston: Pearson, 3rd edition, ch. 1, p.1-34 (accessible via library search).

Artificial Intelligence

- Intelligence:
- "The ability to acquire and apply knowledge and skills." (Oxford Dictionary)
  - in particular,
    - *the ability to solve novel problems*
    - *the ability to act rationally*
    - *the ability to act like humans*

- Artificial Intelligence
  - build and understand intelligent entities or agents
  - different approaches, including "engineering" versus "cognitive modeling" [Smyth, P. (2007), ICS271; still valid?]

Artificial Intelligence (John McCarthy, Stanford University)

- **Artificial Intelligence**

  It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.

- **Intelligence**

  Intelligence is, computationally, the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people, many animals and some machines.

- **Defining intelligence depends on its relationship to human intelligence**

  The problem is that we cannot yet specify the types of computational procedures we want to call intelligent. We understand some of the mechanisms of intelligence and not others.

- **More in:** http://www-formal.stanford.edu/jmc/whatisai/node1.html

Intelligence involves:

- The ability to interact with the real world
  - to perceive, understand (reason), and act
  - e.g. speech recognition and understanding; speech synthesis
  - image understanding
  - ability to take actions, have an effect

- The ability to Reason and Plan
  - modeling the external world, given input
  - solving new problems, planning, and making decisions
  - ability to deal with unexpected problems, uncertainties

Intelligence involves

- Learning and Adaptation
  - we are continuously learning and adapting
  - our internal models are always being “updated”
    - For example, a baby learning to classify and recognise animals

AI involves applications which are able to mimic or duplicate tasks which, if they were to be undertaken by a human, would require intelligence.

Academic Disciplines relevant to AI

• Philosophy	Logic, methods of reasoning, mind as physical system, foundations of learning, language, rationality.
• Mathematics	Formal representation and proof, algorithms, computation, (un)decidability, (in)tractability
• Probability/Statistics	modeling uncertainty, learning from data
• Economics	utility, decision theory, rational economic agents
• Neuroscience	neurons as information processing units.

Academic Disciplines relevant to AI

- Psychology/ Cognitive Science — how do people behave, perceive, process information, represent knowledge.
- Computer engineering — building fast computers/robots
- Control theory — design systems that maximize an objective function over time
- Linguistics — knowledge representation, grammars

What is involved in Intelligence?

- Perceiving, recognising, understanding the real world
- Reasoning and planning about the external world
- Learning and adaptation
- So what general principles should we use to achieve these goals?

Different types of AI

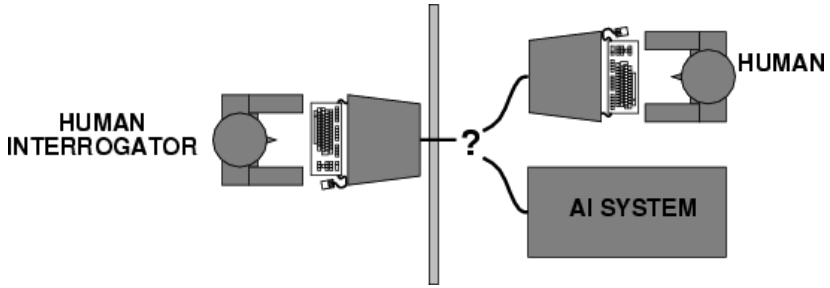
1. Modeling exactly how humans actually think
2. Modeling exactly how humans actually act
3. Modeling how ideal agents “should think”
4. Modeling how ideal agents “should act”

- Symbolic AI focuses on the last definition

Acting humanly- the Turing test

- Turing (1950) "Computing machinery and intelligence“
- "Can machines think?" → "Can machines behave intelligently?“
- Operational test for intelligent behavior: the Imitation Game

Acting humanly- the Turing test



- Suggests major components required for AI:
  - knowledge representation
  - reasoning,
  - language/image understanding,
  - learning

* Question: is it important that an intelligent system acts like a human?

Can computers speak like humans do?

- This is known as “speech synthesis”
  - translate text to phonetic form
    - e.g., “fictitious” -> fik-tish-es
  - use pronunciation rules to map phonemes to actual sound
    - e.g., “tish” -> sequence of basic audio sounds

- Difficulties
  - sounds made by this “lookup” approach sound unnatural
  - sounds are not independent
    - e.g. “act” and “action”
    - modern systems can handle this pretty well
  - a harder problem is emphasis or emotion
    - humans understand what they are saying
    - machines don’t: so they sound unnatural

Can computers speak like humans do?

- Conclusion:
  - NOT in the same way as humans
  - YES, for limited information gathering

  - What about applications like Siri or Cortana?

Can computers recognise speech?

- Speech Recognition:
  - mapping sounds from a microphone into a list of words
  - classic problem in AI, fairly difficult to duplicate or mimic human dialogue; a lot of progress and current interest

    - Recognising single words from a small vocabulary
    - systems can do this with high accuracy (order of 99%)
    - e.g. Bank details
      - limited vocabulary (personal name, account number)
      - computer tries to recognise you first, if unsuccessful hands you over to a human

Recognising human speech

- Recognising normal speech is much more difficult
  - speech is continuous: where are the boundaries between words?
    - e.g., “John’s car has a flat tire”
  - Different ways to combine words
    - can be many thousands of possible words/combinations (economy of language)
    - we can use **context** to help us understand what someone said
      - we can ‘correct’ input and replace ‘noise’
      - try telling a waiter in a restaurant:
        - “I would like some dream and sugar in my coffee”
  - background noise, other speakers, accents, colds, among others
  - on normal speech, modern systems are only about 60-70% accurate

Recognising human speech

- Conclusion:
  - NO, normal speech is too complex to fully accurately recognise
  - YES, for restricted problems (small vocabulary, single speaker)

- Understanding is different to recognition:
  - “Time flies like an arrow”
    - assume the computer can recognise all the words
    - how many different interpretations are there?

Understanding human speech

- Understanding is different to recognition:
  - “Time flies like an arrow”
    - assume the computer can recognise all the words
    - how many different interpretations are there? [ambiguity]
      - 1. time passes quickly like an arrow?
      - 2. command: time the flies the way an arrow times the flies
      - 3. command: only time those flies which are like an arrow
      - 4. “time-flies” are fond of arrows

Understanding human speech

- only 1. (in the previous slide) makes sense,
  - but how could a computer work this out?
  - clearly humans use both implicit common sense knowledge as well as knowledge about language in communication

- Conclusion: NO, a lot of what we say is beyond the capabilities of a current computer program to understand at present

Can computers learn and adapt?

- Learning and Adaptation
  - consider a computer application learning to drive on a motorway
  - we could teach it a lot of rules about what to do
  - or we could let it drive and steer it back on course when something goes wrong
    - systems like this are indeed under development, however there are still a series of issues to resolve, including safety and security concerns.

https://www.nytimes.com/2019/09/26/business/autonomous-cars-sensors.html

Can computers learn and adapt?

- **machine learning** allows computers to learn to do things without explicit programming, for example to self-improve their performance
- many successful applications

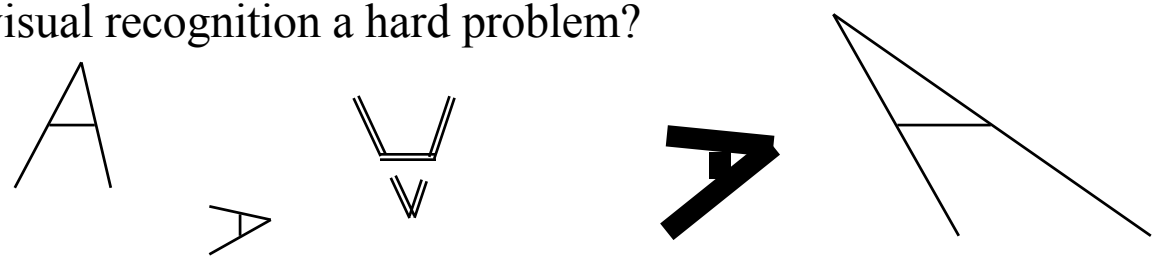
  Consider prediction/forecasting/recommendation algorithms' related applications

- Conclusion: YES, computers can learn and adapt, when presented with information in the appropriate way

Can computers see?

- Recognition v. Understanding (similar to Speech)
  - Recognition and Understanding of Objects in a scene
    - look around this room
    - you can effortlessly recognise objects
    - human brain can map 2d visual image to 3d “map”

- Why is visual recognition a hard problem?



- Conclusion:
  - mostly NO: computers can only “see” certain types of objects under limited circumstances
  - YES for certain constrained problems (e.g. face recognition)

Can computers plan and make optimal decisions?

- Intelligence
  - involves solving problems and making decisions and plans
  - e.g. you want to take a holiday in Spain
    - you need to decide on dates, flights, Brexit…
    - you need to get to the airport, etc.
    - involves a sequence of decisions, plans, and actions

- What makes planning hard?
  - the world is not predictable:
    - your flight is canceled or there's traffic on the M25
  - there is potentially a huge number of details to consider
    - do you consider all flights? all dates? Alternative means of transport

    - AI systems are successful in constrained planning problems

Example: PODS at Bristol and at Heathrow

- Driverless cars (PODS) operate at Heathrow terminal 5 since 2011, connecting the car park to terminal 5.

You can see their journey in the video below.

- https://www.youtube.com/watch?v=F5Knmgr2Ge8

Can computers plan and make optimal decisions?

- Conclusion: to some extent, but human real-world planning and decision-making is still beyond the capabilities of modern computers
  - exception: well-defined, constrained problems

Intelligenet systems in your every day life

- Post Office
  - automatic address recognition and sorting of mail
- Banks
  - automatic check readers, signature verification systems
  - automated loan application classification
- Customer Service
  - automatic voice recognition

Intelligent systems in your every day life

- The Web
  - Identifying your age, gender, location, from your Web surfing
  - Automated fraud detection
- Digital Cameras
  - Automated face detection and focusing
- Computer Games
  - Intelligent characters/agents

Artificial Intelligence and ethics

- Artificial Intelligence has had a significant impact on our lives
  - Although we have not (yet) been able to develop systems with true intelligence, even systems with limited AI, such as machine learning, have changed our world
  - From the systems that learn what to recommend to us on ecommerce websites to smart speakers in our homes to systems that learn our political preferences from our social media use among others.
- Because of this impact on our lives, AI poses questions about how these systems should be used, what they should be used for, who should be in control of them and so on.
  - Some of this impact is useful and beneficial, but some may be harmful.
- These are ethical questions

Ethical questions…

- Ethics is the study of what behaviour we consider acceptable or permissible, and what behaviour is not. Ethics is often a question of what rules we should adopt.
- For example, “don’t bribe people to gain an advantage” is a rule that says bribing people is not acceptable.
- When it comes to AI, we might consider this kind of rule:
  - “Don’t gather information on users to learn their preferences without properly informing them about what you’re learning about them and how it could be used”
  - ”Don’t classify people in ways that those people might consider unacceptable”
  - “Don’t manipulate people into voting in a particular way by changing what appears on their social media feed”

Harms to people

- Fundamental to ethical questions are concerns about harm to people
- Can AI systems harm people?
- By “harm”, we don’t just mean “physical harm”. We also mean things like violations of rights. In ethics, an individual is harmed if their rights are violated, not just if someone hurts them physically or emotionally.
- Some important rights:
  - Right of autonomy – freedom to make your own decisions about your life, free of outside pressures and influences. Also includes not being used as simply a means for someone else to gain an advantage
  - Right of privacy – to be able to conduct your life and your affairs without other people knowing what you are doing or thinking

Key ethical issues for AI

- There are many areas in which AI poses ethical questions. Here are two examples:
- Manipulation of behaviour
  - This is not a new ethical question, but AI makes it much easier for our behaviour to be manipulated. It harms our right to autonomy to be manipulated.
- Bias
  - Bias is also something that is not new. People are biased. That is, they might behave towards other people in a way that is fundamentally unfair.
  - AI systems can exhibit bias in a way that is difficult to detect.

Manipulation of behaviour

- People are vulnerable to “nudges” or deception.
- For example, packaging a detergent in a green bottle may lead some people to think it is environmentally-friendly. The colour is a “nudge” that alters behaviour.
- AI can be used to identify specific groups of people. These groups can then be “nudged”.
- For example, consider the Facebook-Cambridge Analytica scandal, where millions of users’ data were collected, without their consent and used for political advertising. The users were successfully manipulated. Is this ethically acceptable?
- If we answer “no”, can we specify the rule that says it is not acceptable in a way that still allows identifying groups to give them messages that are beneficial to them, such as services that really help their lives?

Bias

- When we exhibit bias, we behave in ways towards others which are/might be unfair. For example, deciding not to hire someone for a job for a reason that has nothing to do with their ability to do it.

- AI is often used for predictive purposes. For example, would someone be good at this job? Is this person likely to cause trouble?

- Bias in AI systems has two different possible sources:
  - The bias (conscious or unconscious bias) of the people developing the system. For example, an AI system may be trained on a particular dataset – the items chosen for inclusion in the dataset may reflect human bias. Someone who dislikes people from a certain part of the UK may decide to use samples of people speaking with an accent from that area to train a system to spot troublemakers.

  - Bias in the system – for example, statistical bias. A dataset used to train an AI system for one thing may in fact represent a different thing. For example, a dataset of people in a company who have been promoted that is then used for an AI system that decides who to recruit. These are two different issues: promotion and recruitment. If the company have a historic issue with lack of diversity for example in only promoting white men, then this bias will affect recruitment of women, BME employees etc.

Summary

- AI systems hold the promise of greatly improving our lives
- But their “intelligence”, however limited, that allows them to operate without human involvement and management raises fundamental ethical issues because the same power that promises benefits can also threaten harm
- Hence we need rules to govern their development, use and effect. We need rules to say what is acceptable when it comes to AI systems. In other words, we need to apply ethics to AI systems.

Thank you!