Chapters 1–2: Introduction to AI

DIT410/TIN173 Artificial Intelligence

Peter Ljunglöf

(inspired by slides by Poole & Mackworth, Russell & Norvig, et al)

22 March, 2016

Outline

Outline

What is intelligence?

"It is not my aim to surprise or shock you – but the simplest way I can summarize is to say that there are now in the world machines that can think, that learn, and that create. Moreover, their ability to do these things is going to increase rapidly until – in a visible future – the range of problems they can handle will be coextensive with the range to which human mind has been applied."

Herbert A Simon

What is intelligence?

"It is not my aim to surprise or shock you – but the simplest way I can summarize is to say that there are now in the world machines that can think, that learn, and that create. Moreover, their ability to do these things is going to increase rapidly until – in a visible future – the range of problems they can handle will be coextensive with the range to which human mind has been applied."

Herbert A Simon (1957)

Strong and Weak AI

Weak AI – acting intelligently

- the belief that machines can be made to act as if they are intelligent

Strong AI – being intelligent

- the belief that those machines are actually thinking

Most AI researchers don't care

- "the question of whether *machines can think*...
  ...is about as relevant as whether *submarines can swim*."
  (Edsger W Dijkstra, 1984)

Weak AI

- Weak AI is a category that is flexible, as soon as we understand how an AI-program works, it appears less "intelligent".
- And as soon as a part of AI is successful, it becomes an own research area! E.g., large parts of advanced search, parts of language understanding, parts of machine learning and probabilistic learning etc.
- And AI is left with the remaining hard-to-solve problems!

What is an AI system?

Do we want a system that...

What is an AI system?

Do we want a system that...

- thinks like a human?
  - cognitive neuroscience / cognitive modelling
  - AGI = artificial general intelligence

What is an AI system?

Do we want a system that...

- thinks like a human?
  - cognitive neuroscience / cognitive modelling
  - AGI = artificial general intelligence
- acts like a human?
  - the Turing test

What is an AI system?

Do we want a system that...

- thinks like a human?
  - cognitive neuroscience / cognitive modelling
  - AGI = artificial general intelligence
- acts like a human?
  - the Turing test
- thinks rationally?
  - "laws of thought"
  - from Aristotle's syllogism to modern day theorem provers

What is an AI system?

Do we want a system that...

- thinks like a human?
  - cognitive neuroscience / cognitive modelling
  - AGI = artificial general intelligence
- acts like a human?
  - the Turing test
- thinks rationally?
  - "laws of thought"
  - from Aristotle's syllogism to modern day theorem provers
- acts rationally?
  - "rational agents"
  - maximise goal achievement, given available information

Outline

A brief history of AI

1943	McCulloch & Pitts: Boolean circuit model of brain
1950	Alan Turing's "Computing Machinery and Intelligence"
1951	Marvin Minsky develops a neural network machine
1950s	Early AI programs: e.g., Samuel's checkers program, Gelernter's Geometry Engine, Newell & Simon's Logic Theorist and General Problem Solver
1956	Dartmouth meeting: "Artificial Intelligence" adopted
1965	Robinson's complete algorithm for logical reasoning
1966	Joseph Weizenbaum creates Eliza
1969	Minsky&Papert show limitations of the perceptron Neural network research almost disappears
1971	Terry Winograd's Shrdlu dialogue system
1972	Alain Colmerauer invents Prolog programming language

A brief history of AI

1976	MYCIN, an expert system for disease diagnosis
1980s	Era of expert systems
1990s	Neural networks, probability theory, AI agents
1993	RoboCup initiative to build soccer-playing robots
1997	IBM Deep Blue beats the World Chess Champion
2003	Very large datasets: genomic sequences
2007	Very large datasets: WAC (web as corpus)
2011	IBM Watson wins Jeopardy
2012	US state of Nevada permits driverless cars
2014	"Deep learning": recommendation systems, image tagging, board games, speech translation, pattern recognition
2016	Google AlphaGo beats the World Go Champion

Outline

Course overview

Course website: http://chalmersgu-ai-course.github.io/

Teachers: Peter Ljunglöf, Pablo Buiras, John J. Camilleri, Jonatan Kilhamn, Prasanth Kolachina, Irene Lobo Valbuena

Student representatives: Emmanuel Batis (MPIDE), Mathias Bylund (MPIDE), William Dahlberg (MPALG), Jimmy Eliasson Malmer (TKITE), Deepak Kiran (MPSYS)
+ Additional volunteers!? Especially from GU!

Course book(s): Russell & Norvig (2002/10/14)
Poole & Mackworth (2010) – http://artint.info

Obligatory course moments

Group work: form a group (24 March)

Group work: written essay

- Write a 6-page essay about AI (23 April)
- Review two other essays (7 May)
- Revise your essay according to the reviews you got (21 May)
- Present it to the other students (24–25 May)

Group work: Shrdlite programming project

- Two intermediate submissions: A^* planner (30 April), and natural language interpreter (14 May)
- Implement extensions for higher grades (28 May)

Individual written examination

- *Peer-corrected* exam (3 May) + traditional re-exam (1 June)

Individual self-evaluation

- Grade your own and your team members' performance (28 May)

Additional course moments

Group supervision

- Mondays and Thursdays (mostly) during weeks 15–21
- You book a fixed supervision slot on Doodle
- Supervision is compulsory for all group members!
  - it is part of your examination

Lectures

- Tuesday and Friday (mostly), 10:00 (mostly), weeks 12, 15, 16, 17

Exercise sessions

- Tuesday and Friday (mostly), 08:00 (mostly), weeks 12, 15, 16, 17

Grading

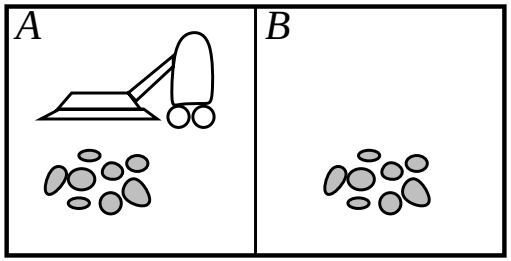
- Higher grade than pass/3/G only depends on the group work!
- You collect up to 10 bonus points:
  - The essay can give 0–3 points
  - Your reviews can give 0–1 point
  - Shrdlite can give 0–6 points (every extension gives 1–3 points)
  - Your individual bonus points can be more or less than the group's

GU grades: G/pass (0–5 points), VG/distinction (6–10 points)

Chalmers grades: 3/pass (0–2 points), 4 (3–6 points), 5 (7–10 points)

Outline

A vacuum-cleaner agent



Percepts: location and contents, e.g., [*A*, *Dirty*]

Actions: *Left*, *Right*, *Suck*, *NoOp*

A simple agent function is:

- If the current square is dirty, then suck; otherwise, move to the other square.

How do we know if this is a *good* agent function?

- What is the *best* function? — Is there one?
- Who decides this?

Rationality

Fixed *performance measure* evaluates the environment sequence

- one point per square cleaned up in time T?
- one point per clean square per time step, minus one per move?
- penalize for $> k$ dirty squares?

A rational agent chooses any action that

- maximizes the expected value of the performance measure
- given the percept sequence to date

Rational $\neq$ omniscient — percepts may not supply all relevant information

Rational $\neq$ clairvoyant — action outcomes may not be as expected

Hence, rational $\neq$ successful

PEAS

To design a rational agent, we must specify the task environment, which consists of the following four things:

- Performance measure
- Environment
- Actuators
- Sensors

PEAS – autonomous car

Example: The task environment for an autonomous car:

- Performance measure
  - getting to the right place, following traffic laws, minimising fuel consumption/time, maximising safety, ...
- Environment
  - roads, other traffic, pedestrians, road signs, passengers, ...
- Actuators
  - steering, accelerator, brake, signals, loudspeaker, ...
- Sensors
  - cameras, sonar, speedometer, GPS, odometer, microphone, ...

Enviroment types / dimensions of complexity

Dimension	**Possible values**
Observable?	full vs. partial
Deterministic?	deterministic vs. stochastic
Episodic?	episodic vs. sequential
Static?	static vs. dynamic (semidynamic)
Discrete?	discrete vs. continuous
Number of agents	single vs. multiple (competetive/cooperative)

The environment type largely determines the agent design

Environment types, examples

	Chess (w. clock)	Poker	Driving	Image recognition
Observable?				
Deterministic?				
Episodic?				
Static?				
Discrete?				
N:o agents				

Environment types, examples

	Chess (w. clock)	Poker	Driving	Image recognition
Observable?	fully			
Deterministic?	determ.			
Episodic?	sequential			
Static?	semi			
Discrete?	discrete			
N:o agents	multiple (compet.)			

Environment types, examples

	Chess (w. clock)	Poker	Driving	Image recognition
Observable?	fully	partially		
Deterministic?	determ.	stochastic		
Episodic?	sequential	sequential		
Static?	semi	static		
Discrete?	discrete	discrete		
N:o agents	multiple (compet.)	multiple (compet.)		

Environment types, examples

	Chess (w. clock)	Poker	Driving	Image recognition
Observable?	fully	partially	partially	
Deterministic?	determ.	stochastic	stochastic	
Episodic?	sequential	sequential	sequential	
Static?	semi	static	dynamic	
Discrete?	discrete	discrete	continuous	
N:o agents	multiple (compet.)	multiple (compet.)	multiple (cooper.)	

Environment types, examples

	Chess (w. clock)	**Poker**	**Driving**	**Image recognition**
Observable?	fully	partially	partially	fully
Deterministic?	determ.	stochastic	stochastic	determ.
Episodic?	sequential	sequential	sequential	episodic
Static?	semi	static	dynamic	static
Discrete?	discrete	discrete	continuous	disc./cont.
N:o agents	multiple (compet.)	multiple (compet.)	multiple (cooper.)	single

Environment types, examples

	Chess (w. clock)	Poker	Driving	Image recognition
Observable?	fully	partially	partially	fully
Deterministic?	determ.	stochastic	stochastic	determ.
Episodic?	sequential	sequential	sequential	episodic
Static?	semi	static	dynamic	static
Discrete?	discrete	discrete	continuous	disc./cont.
N:o agents	multiple (compet.)	multiple (compet.)	multiple (cooper.)	single

The real world is (of course) partially observable, stochastic, sequential, dynamic, continuous, multi-agent

Defining a Solution

- Given an informal description of a problem, what is a solution?
- Typically, much is left unspecified, but the unspecified parts can't be filled in arbitrarily.
- Much work in AI is motivated by *common-sense reasoning*. The computer needs to make common-sense conclusions about the unstated assumptions.

Quality of Solutions

- Does it matter if the answer is wrong or answers are missing?

Classes of solutions:

- An optimal solution is a best solution according some measure of solution quality.
- A satisficing solution is one that is good enough, according to some description of which solutions are adequate.
- An approximately optimal solution is one whose measure of quality is close to the best theoretically possible.
- A probable solution is one that is likely to be a solution.

Types of agents

- Simple reflex agent
  - selects actions based on *current percept* – ignores history
- Model-based reflex agent
  - maintains an *internal state* that depends on the percept history
- Goal-based agent
  - has a *goal* that describes situations that are desirable
- Utility-based agent
  - has a *utility function* that measures the performance
- Learning agent
  - any of the above agents can be a learning agent
  - learning can be *online* or *offline*

Outline

Is AI possible?

There are different opinions...

Is AI possible?

There are different opinions...

- ...some are slightly positive:

Is AI possible?

There are different opinions...

- ...some are slightly positive:
  - "every [...] feature of intelligence can be so precisely described that a machine can be made to simulate it" (McCarthy et al, 1955)

Is AI possible?

There are different opinions...

- ...some are slightly positive:
  - "every [...] feature of intelligence can be so precisely described that a machine can be made to simulate it" (McCarthy et al, 1955)
- ...and some lean towards the negative:

Is AI possible?

There are different opinions...

- ...some are slightly positive:
  - "every [...] feature of intelligence can be so precisely described that a machine can be made to simulate it" (McCarthy et al, 1955)
- ...and some lean towards the negative:
  - "AI [...] stands not even a ghost of a chance of producing durable results" (Sayre, 1993)

Is AI possible?

There are different opinions...

- ...some are slightly positive:
  - "every [...] feature of intelligence can be so precisely described that a machine can be made to simulate it" (McCarthy et al, 1955)
- ...and some lean towards the negative:
  - "AI [...] stands not even a ghost of a chance of producing durable results" (Sayre, 1993)

It's all in the definitions:

- what do we mean by "thinking" and "intelligence"?

Computing Machinery and Intelligence

The most important paper in AI, of all times:

- (and I'm not the only one who thinks that...)
- "Computing Machinery and Intelligence" (Turing, 1950)
  - introduced the "imitation game" (Turing test)
  - discussed objections against intelligent machines, including almost every objection that has been raised since then
  - it's also easy to read...
    ...so you really have to read it!

Turing's objections to AI [1–3]

Turing's objections to AI [1–3]

(1) The Theological Objection

- "Thinking is a function of man's immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine can think."

Turing's objections to AI [1–3]

(1) The Theological Objection

- "Thinking is a function of man's immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine can think."

(2) The "Heads in the Sand" Objection

- "The consequences of machines thinking would be too dreadful. Let us hope and believe that they cannot do so."

Turing's objections to AI [1–3]

(1) The Theological Objection

- "Thinking is a function of man's immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine can think."

(2) The "Heads in the Sand" Objection

- "The consequences of machines thinking would be too dreadful. Let us hope and believe that they cannot do so."

(3) The Mathematical Objection

- Based on Gödel's incompleteness theorem.

Turing's objections to AI [4–5]

Turing's objections to AI [4–5]

(4) The Argument from Consciousness

- "No mechanism could feel [...] pleasure at its successes, grief when its valves fuse, [...], be angry or depressed when it cannot get what it wants."

Turing's objections to AI [4–5]

(4) The Argument from Consciousness

- "No mechanism could feel [...] pleasure at its successes, grief when its valves fuse, [...], be angry or depressed when it cannot get what it wants."

(5) Arguments from Various Disabilities

- "you can make machines do all the things you have mentioned but you will never be able to make one to do X."
- where X can... "be kind, resourceful, beautiful, friendly, [...], have a sense of humour, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream, [...], use words properly, be the subject of its own thought, [...], do something really new."

Turing's objections to AI [6–8]

Turing's objections to AI [6–8]

(6) Lady Lovelace's Objection

- "The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform"

Turing's objections to AI [6–8]

(6) Lady Lovelace's Objection

- "The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform"

(7) Argument from Continuity in the Nervous System

- "one cannot expect to be able to mimic the behaviour of the nervous system with a discrete-state system."

Turing's objections to AI [6–8]

(6) Lady Lovelace's Objection

- "The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform"

(7) Argument from Continuity in the Nervous System

- "one cannot expect to be able to mimic the behaviour of the nervous system with a discrete-state system."

(8) The Argument from Informality of Behaviour

- "if each man had a definite set of rules of conduct by which he regulated his life he would be no better than a machine. But there are no such rules, so men cannot be machines."

The final objection [9]

The final objection [9]

(9) The Argument from Extrasensory Perception

- this was the strongest argument according to Turing...
- "the statistical evidence [...] is overwhelming"
- "Let us play the imitation game, using as witnesses a man who is good as a telepathic receiver, and a digital computer. The interrogator can ask such questions as 'What suit does the card in my right hand belong to?' The man by telepathy or clairvoyance gives the right answer 130 times out of 400 cards. The machine can only guess at random, and perhaps gets 104 right, so the interrogator makes the right identification."

Strong AI: Brain replacement

The brain replacement experiment

- by Searle (1980) and Moravec (1988)
- suppose we gradually replace each neuron in your head with an electronic copy
  - what will happen to your mind, your consciousness?
  - Searle argues that you will gradually feel dislocated from your body
  - Moravec argues you won't notice anything

Strong AI: The Chinese room

The Chinese room experiment (Searle, 1980)

- an English-speaking person takes input and generates answers in Chinese
  - he/she has a rule book, and stacks of paper
  - the person gets input, follows the rules and produces output
- i.e., the person is the CPU, the rule book is the program and the papers is the storage device

Does the system understand Chinese?

The technological singularity

Will AI lead to superintelligence?

- "...ever accelerating progress of technology and changes in the mode of human life, which gives the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue"
  (von Neumann, mid-1950s)
- "We will successfully reverse-engineer the human brain by the mid-2020s. By the end of that decade, computers will be capable of human-level intelligence." (Kurzweil, 2011)
- "There is not the slightest reason to believe in a coming singularity." (Pinker, 2008)

Ethical issues of AI

Possible risks

- AI might be used towards undesirable ends
  - e.g., surveillance by speech recognition, detection of "terrorist phrases"
- AI might result in a loss of accountability
  - what's the legal status of a self-driving car?
  - or a medical expert system?
- AI might mean the end of the human race
  - what if the new superintelligent race won't obey Asimov's robot laws?