Artificial Intelligence I

Dr Sean Holden

Computer Laboratory, Room FC06

Telephone extension 63725

Email: sbh11@cl.cam.ac.uk

http://www.cl.cam.ac.uk/users/sbh11/

Introduction: what's AI for?

Homo Sapiens = "Man the wise"

What is the purpose of Artificial Intelligence (AI)?

- To *understand intelligence*. This aim is shared by philosophy and psychology.
- To *make* intelligent systems. More exclusively the realm of CS.
- To understand *ourselves*.
- To make and sell cool stuff!

Our brain is small and slow (?). So why is it so good?

Introduction: what's the character of the field?

In many ways this is a young field (1956).

- This means we can actually *do* things!
- Also, we know what we're trying to do is *possible*.

Philosophy has addressed such problems for at least 2000 years.

- *Can* we do AI?
- *Should* we do AI?
- Is AI *impossible*?

Arguably, philosophy has had relatively little success.

Perhaps the most important open problem the world has left?

Why so little popular coverage?

Introduction: what's happened since 1956?

Computers have taken us from theory to practice.

The simple ability to *try things out* has led to huge advances in a relatively short time.

- Perception (vision, speech processing...)
- Logical reasoning (prolog, expert systems...)
- Playing games (chess, backgammon, go...)
- Diagnosis of illness (in various contexts...)
- Theorem proving (assorted mathematical results...)
- Literature and music (automated writing and composition...)

The nature of the pursuit

What is AI?

Well, it depends on who you ask...

We can find many definitions and a rough categorisation can be made depending on whether we are interested in:

- the way in which a system *acts* or the way in which it *thinks*, and;
- whether we want it to do this in a *human* way or a *rational* way.

Here, the word *rational* has a special meaning: it means doing the *correct* thing in given circumstances.

Acting like a human

Alan Turing proposed what is now known as the *Turing Test*.

- A human judge is allowed to interact with an AI program via a terminal.
- This is the *only* method of interaction.
- If the judge can't decide whether the interaction is produced by a machine or another human then the program passes the test.

In the *unrestricted* Turing test the AI program may also have a camera attached, so that objects can be shown to it, and so on.

Acting like a human

The Turing test is informative, and (very!) hard to pass.

- It requires many abilities that seem necessary for AI, such as learning. BUT: a human child would probably not pass the test!
- Sometimes an AI system needs human-like acting abilities—for example *expert systems* often have to produce explanations—but *not always*.

See the Loebner Prize:

`http://www.loebner.net/Prizef/loebner-prize.html`

Thinking like a human

There is always the possibility that a machine *acting* like a human does not actually *think*.

The *cognitive modelling* approach to AI has tried to:

- deduce *how humans think*—for example by *introspection* or *psychological experiments*—and,
- copy the process by mimicking it within a program.

An early example of this approach is the *General Problem Solver* produced by Newell and Simon in 1961. They were concerned with whether or not the program reasoned in the same manner that a human did.

Computer Science + Psychology = *Cognitive Science*

Thinking rationally: the “laws of thought”

The idea that intelligence reduces to *rational thinking* is a very old one. Aristotle first tried to model thought this way through *syllogisms*.

The general field of *logic* made major progress in the 19th and 20th centuries, allowing it to be applied to AI.

- we can *represent* and *reason* about many different things;
- The *logicist* approach to AI.

Unfortunately there are obstacles to any naive application of logic. It is hard to:

- represent *commonsense knowledge*;
- deal with *uncertainty*;
- reason without being tripped up by *computational complexity*.

Acting rationally

Basing AI on the idea of *acting rationally* means attempting to design systems that act to *achieve their goals* given their *beliefs*.

What might be needed?

- To make *good decisions* in many *different situations* we need to *represent* and *reason* with *knowledge*.
- We need to deal with *natural language*.
- We need to be able to *plan*.
- We need *vision*.
- We need *learning*.

Acting rationally

The idea of acting rationally has several advantages:

- the concepts of *action*, *goal* and *belief* can be defined precisely making the field suitable for scientific study, whereas;
- dealing with humans involves a system that is still changing and adapted to a very specific environment.

Also, all of the things needed to pass a Turing test seem necessary for rational acting. Finally, logic can form *part* of acting rationally. BUT:

- sometimes it's necessary to act when there's no logical course of action;
- sometimes inference is unnecessary (reflex actions).

Foundations I: philosophy (428 B.C. to present)

Socrates

↓

Plato (born 428 B.C.)

↓

Aristotle

Socrates wanted to know whether there was an algorithm (!) for "piety", prompting Plato to consider the rules governing rational thought.

This led to the *syllogisms*.

Foundations I: philosophy (428 B.C. to present)

The possibility of reasoning being done *mechanically*: Ramon Lull's *concept wheels* (approx. 1315).

Various other attempts at mechanical calculators.

Mind as a *physical system*: Rene Descartes (1596-1650).

- is *mind* distinct from *matter*?
- what is *free will*?

Dualism: part of our mind—-the *soul* or *spirit*— is set apart from the rest of nature.

Foundations I: philosophy (428 B.C. to present)

The opposing position of *materialism* was taken up by Wilhelm Leibnitz (1646-1716).

He attempted to build a machine to perform mental operations but failed as his logic was too weak.

There is an intermediate position: mind is physical but unknowable.

Foundations I: philosophy (428 B.C. to present)

If mind is physical where does *knowledge* come from?

Francis Bacon (1561-1626): *empiricism*.

↓

John Locke (1632-1704): “Nothing is in the understanding, which was not first in the senses”.

In *A Treatise of Human Nature*, David Hume (1711-1776) introduced the concept of *induction*: we obtain rules by repeated exposure.

This was developed by Bertrand Russel (1872-1970): *observation sentences* are connected to *sensory inputs*, and all knowledge is characterised by logical theories connected to these. *Logical positivism*.

The *nature* of the connection between theories and sentences is the subject of Rudolf Carnap and Carl Hempel’s *confirmation theory*.

Foundations I: philosophy (428 B.C. to present)

Finally: what is the connection between *knowledge* and *action*? How are actions *justified*?

Aristotle: don't concentrate on the *end* but the *means*.

If to achieve the end you need to achieve something intermediate, consider how to achieve that, and so on.

This approach was implemented in Newell and Simon's 1972 GPS.

Foundations II: mathematics (800 to present)

Philosophers have had some great ideas, but to be *scientific* about AI three areas of mathematics are needed: computation, logic, and probability.

Logic:

- To the likes of Aristotle, a philosophical rather than mathematical pursuit.
- George Boole (1815-1864) made it into mathematics.
- Gottlob Frege (1848-1925) founded all the essential parts of first-order logic.
- Alfred Tarski (1902-1983) founded the theory of reference: what is the relationship between *real* objects and those in logic.

Foundations II: mathematics (800 to present)

Computation:

- Concept of an algorithm: Arab mathematician *al-Khowarazmi*.
- What are the limits of algorithms? David Hilbert's (1862-1943) *entscheidungsproblem*.
- Solved by Turing, who (with others) formulated precisely what an algorithm *is*.
- Ultimately, this has lead to the idea of *intractability*.
- Kurt Godel (1906-1978): theorems on completeness and incompleteness.

Foundations II: mathematics (800 to present)

Probability:

- Gerolamo Cardano (1501-1576): gambling outcomes.
- Further developed by Fermat, Pascal, Bernoulli, Laplace...
- Bernoulli (1654-1705) in particular proposed probability as a measure of *degree of belief*.
- Bayes (1702-1761) showed how to update a degree of belief when new evidence is available.
- Probability forms the basis for the modern treatment of uncertainty.
- The decision theory of Von Neumann and Morgenstern (1944) combines uncertainty with action.

Foundations III: psychology (1879 to present)

Modern psychology (arguably) began with the study of the human visual system performed by Hermann von Helmholtz (1821-1894).

The first *experimental psychology* lab was founded by his student Wilhelm Wundt (1832-1920) at the University of Leipzig.

- The lab conducted careful, controlled experiments on human subjects.
- The idea was for the subject to perform some task and *introspect* about their thought processes.

Other labs followed this lead. BUT: a strange—and fatal—effect appeared.

For each lab, the introspections of the subjects turned out to conform to the preferred theories of the lab!

Foundations III: psychology (1879 to present)

The main response to this effect was *behaviourism*, founded by John Watson (1878-1958) and Edward Lee Thorndike (1874-1949).

- They regarded evidence based on introspection as fundamentally unreliable, so...
- ...they simply rejected all theories based on any form of mental process.
- They considered only *objective* measures of *stimulus* and *response*.

Learnt a LOT of interesting things about rats and pigeons!

Foundations III: psychology (1879 to present)

The (arguably somewhat more sophisticated) view of the brain as an *information processing device*—the view of cognitive psychology—was steamrollered by behaviourism until Kenneth Craik's *The Nature of Explanation* (1943).

The idea that concepts such as reasoning, beliefs, goals *etc* are important is re-stated.

Critically: the system contains a model of the world and of the way its actions affect the world.

Foundations III: psychology (1879 to present)

stimuli converted to internal representation

$\downarrow$

cognitive processes manipulate internal representations

$\downarrow$

internal representations converted into actions

Foundations IV: computer engineering (1940 to present)

To have AI, you need a means of *implementing* the intelligence. Computers are (at present) the only devices in the race! (Although quantum computation is looking interesting...)

AI has had a major effect on computer science:

- time sharing
- interactive interpreters
- linked lists
- storage management
- some fundamental ideas in object-oriented programming
- an so on...

When AI has a success, the ideas in question tend to stop being called AI!

Foundations V: linguistics (1957 to present)

B. F. Skinner's *Verbal Behaviour* (1951) set out the approach to *language* developed by the behaviourists.

It was reviewed by Noam Chomsky, author of *Syntactic Structures*:

- He showed that the behaviourists could not explain how we understand or produce sentences that we have *not previously heard*.
- Chomsky's own theory—based on syntactic models as old as the Indian linguist Panini (350 B.C.), did not suffer in this way.
- Chomsky's own theory was also formal, and could be programmed.

Foundations V: linguistics (1957 to present)

This overall problem is considerably harder than was realised in 1957.

It requires knowledge representation, and the fields have informed one another.

“Time flies like an arrow”

“Fruit flies like a banana”

Foundations VI: economics (1776 to present)

How should I act, perhaps in the presence of adversaries, to obtain something nice in the future?

- Adam Smith: *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776).
- When we say "something nice," how can the "degree of niceness" be measured?

  This leads to the idea of *utility* as a mathematical concept.

  Developed by Leon Walras (1834-1910), Frank Ramsey (1931) and John Von Neumann and Oskar Morgenstern (1944).

Foundations VI: economics (1776 to present)

- For *large* economies:

  Probability theory + utility theory = decision theory

- *Game theory* is more applicable to *small* economies.

  In some games it turns out to be *rational* to act (apparently) randomly.

- Dealing with *future* gains resulting from a sequence of actions: operations research and *Markov decision processes*, the latter due to Richard Bellman (1957).

Unfortunately it is computationally hard to act rationally.

Herbert Simon (1916-2001) won the Nobel Prize for Economics in 1978 for his work demonstrating that *satisficing* is a better way of describing the actual behaviour of humans.

Foundations VII: neuroscience (1861 to present)

Nasty bumps on the head

$\downarrow$

We know that the brain has something to do with consciousness

Experiments by Paul Broca (1824-1880) led to the understanding that localised regions have different tasks.

Around that time the presence of *neurons* was understood *but* there were still major problems.

For example, even now there is no complete understanding of how our brains store a single memory!

More recently: EEG, MRI and the study of single cells.

Foundations VIII: cybernetics and control theory (1948 to present)

Ktesibios of Alexandria (250 BC)

The first machine to be able to modify its own behaviour was a water clock containing a mechanism for controlling the flow of water.

- James Watt (1736-1819): governor for steam engines
- Cornelius Drebbel (1572-1633): thermostat
- Control theory as a mathematical subject: Norbert Wiener (1894-1964) and others.

This presented another challenge to behaviourism.

Foundations VIII: cybernetics and control theory (1948 to present)

Interesting behaviour caused by a *control system* minimising *error*

error = difference between *goal* and *current situation*

More recently, we have seen *stochastic optimal control* dealing with the maximisation over time of an *objective function*.

This is connected directly to AI, but the latter moves away from *linear*, *continuous* scenarios.

What's in this course?

This course introduces some of the fundamental areas that make up modern AI:

- An outline of the background to the subject.
- An introduction to the idea of an agent.
- Solving problems in an intelligent way by search.
- Playing games.
- Knowledge representation, and reasoning.
- Learning.
- Planning.

If time, a little philosophy. (A crash course on how to survive at parties!)

What's *not* in this course?

- Nothing is said about the classical AI programming languages Prolog and Lisp.
- A great deal of all the areas on the last slide!
- Perception: vision, hearing and speech processing, touch (force sensing, knowing where your limbs are, knowing when something is bad), taste, smell.
- Natural language processing
- Acting on and in the world: robotics (effectors, locomotion, manipulation), control engineering, mechanical engineering, navigation.
- Genetic algorithms.
- Fuzzy logic.
- Uncertainty and much further probabilistic material.

Text books and prerequisites

The course is based on the relevant parts of:

Artificial Intelligence: A Modern Approach, Second Edition (2003).
Stuart Russell and Peter Norvig, Prentice Hall International Editions.

The prerequisites for the course are:

- A little logic.
- Algorithms and data structures.
- Discrete and continuous mathematics.
- Basic computational complexity.

Interesting things on the web

A few interesting web starting points:

- General resource page for machine learning:

  `www.aic.nrl.navy.mil/~aha/research/machine-learning.html`

- The Cyc project:

  `www.cyc.com`

- Human-like robots:

  `www.ai.mit.edu/projects/humanoid-robotics-group/`

- Sony robots:

  `www.aibo.com`

- NEC “PaPeRo”:

  `www.incx.nec.co.jp/robot`