Fundamentals of Artificial Intelligence

Alan Smaill

Sep 25, 2008

School of informatics

Course Aims

- Necessary background for symbolic AI
- Based on formal techniques, rather than Psychology
- Bringing everyone up to a common starting point
- Alternative approaches and debates addressed

Course Organisation

- Lectures weeks 1–5,7–10
- Tutorials from week 3
- 2 Practical exercises
- Exam at end of first semester.

Sources of info

- Module web page – slides, handouts and links will appear here.
- The course text is

  Artificial Intelligence: A Modern Approach, Russell and Norvig, 2nd edition, 2003, Prentice Hall (£44.99).

- Web site for R&N:

  `aima.cs.berkeley.edu`

- Newsgroup: `eduni.inf.course.fai`
  I will monitor the newsgroup.
- email myself with any queries during the course

Course Topics

- Rational agents and agent architectures
- Algorithm and problem complexity
- Search Spaces and Algorithms
- Heuristic Search
- Logic as a representation language
- Logical semantics and deduction
- First-order logic

Course Topics ctd

- Proof search
- Definite clause logic and the logic of Logic Program
- Constraint Satisfaction Problems and algorithms
- Alternative Approaches
- Current philosophical debates in AI

What is Intelligence, anyway?

There are plenty of possible answers to this. Here are some definitions from the "Penguin dictionary of Psychology":

Intelligence:
1 The relating activity of mind; *insight* as understood by the Gestalt psychologists; in its lowest terms, intelligence is present where the individual, animal, or human being is aware, however dimly, of the relevance of his behaviour to an objective.
2 The capacity to meet novel situations, or to learn to do so, by new adaptive responses.
3 The ability to perform tests or tasks, involving the grasping of relationships, the degree of intelligence being proportional to the complexity, or the abstractness, or both, of the relationships.

Intelligence: awareness, or performance?

Notice that the first definition talks about (conscious) *awareness*. This makes it very hard to know when it applies.

The other definitions are more in terms of *performance*: the second could apply to simple adaptive devices.

For the third definition, we can ask whether *grasping* involves *conscious* awareness of relationships.

We should also ask where the abstractness, and difficulty of the tasks resides – in the tasks, or relative to the knowledge and skills of the thinker? (It takes less intelligence to solve the puzzle the second time.)

What is AI?

Alan Turing in a famous 1950 paper tackled the question:

Can machines think?

See "Computing Machinery and Intelligence", Mind, 53, 433–460 and in many anthologies.

He proposed we should replace this with the question

Can machines behave like humans?

so that there is a possibility of *testing*:
look at the behaviour, not at what is going on to produce the behaviour.

Turing

- Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes
- Anticipated all major arguments against AI in following 50 years
- Suggested major components of AI: knowledge, reasoning, language understanding, learning

The *Turing Test* is still influential, but surely not in itself sufficient to characterise AI.

Loebner

There is an annual competition based on the idea of the Turing test; see

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

The big prize is still to be won, for a machine judged to be indistinguishable from a human on the basis of a teletype interface; the best program each year wins a small prize.

2005's best program was judged on the basis of conversations like the following.


```
PROGRAM: Hello, my name is George and I am the human.
JUDGE: Hello, I'm John.
PROGRAM: Where are you from?
JUDGE: I live on Martha's Vineyard.
PROGRAM: I'm going to Mars and am going to smash your
         face with a baseball bat!
JUDGE: Well, I do hope you're the program, because if you're human. . .
       Well, let's jus say that's a little rude.
PROGRAM: I'm not bores.
JUDGE: You're not bores? As in, an augur or something? A drill bit?
PROGRAM: Just a bit.
JUDGE: So, where are you from? ANd how old are you, etc?
PROGRAM: I'm 39!
JUDGE: And you reside. . .?
PROGRAM: No, I live in the netherlands.
JUDGE: Which part? Town or country? Also, do you grow flowers?
```

```
PROGRAM: Lots. I'm about to plant tomatoes.
JUDGE: A little late in the season for that.
PROGRAM: No, a little animal.
JUDGE: OK, enough for now. I'll go chat with the other entity.
PROGRAM: Ok then it was nice to share your hair brush.
JUDGE: What do mean by that? How have you shared my hair brush?
PROGRAM: Good.
JUDGE: Good?
PROGRAM: Yes.
JUDGE: This has become a very zen-like conversation.
PROGRAM: And worse yet, I gave it to charity.
JUDGE: What? WHat did you give to charity? This conversation?
PROGRAM: Er.. 1 million Dollars?
JUDGE: Which charity?
PROGRAM: Are we talking about the simpsons or unreal?
JUDGE: I'm not sure what the hell we're talking about, actually.
```


What is AI then?

The answer can be more or less modest . . .

- Systems that think like humans
  . . . *machines with minds*, in the full and literal sense
- Systems that act like humans
  . . . machines that perform functions that require intelligence when performed by people
- Systems that think rationally
  . . . systems that make it possible to perceive, reason and act
- Systems that act rationally
  . . . intelligent behaviour in artefacts

Cognitive Science

1960s "cognitive revolution": information-processing psychology replaced prevailing orthodoxy of **behaviourism**.

Requires scientific theories of *internal* activities of the brain:

- What level of abstraction? "**Knowledge**" or "**circuits**"?
- How to validate? Requires
  1) Predicting and testing behaviour of human subjects (top-down)
  or 2) Direct identification from neurological data (bottom-up)

Both approaches (roughly, **Cognitive Science** and **Cognitive Neuroscience**) are now distinct from AI.

Both share with AI the following characteristic: the available theories do not explain (or engender) anything resembling human-level general intelligence.

Acting rationally

The right thing: that which is expected to maximise goal achievement, given the available information.

Doesn't necessarily involve thinking—e.g., blinking reflex—but thinking should be in the service of rational action.

There is plenty of evidence that humans often act irrationally, so systems built on these principles are not expected to be psychologically plausible.

Where agents combine there are ethical issues:

> "Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good."
>
> Aristotle, Nicomachean Ethics

Rational agents

An agent is an entity that perceives and acts (here, often by receiving and sending messages).

We are interested in rational agents.

Abstractly, an agent is a function from percept histories to actions: i.e. the agent will compute for a history of perceptions $(p_0, p_1, p_2, \dots)$ an action a, taking into account the agent's own goals.

For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance.

Caveat: *computational limitations make perfect rationality unachievable* so we design best program for given machine resources.

AI prehistory

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
Psychology	adaptation
	phenomena of perception and motor control
	experimental techniques (psychophysics, etc.)
Economics	formal theory of rational decisions
Linguistics	knowledge representation, grammar
Neuroscience	plastic physical substrate for mental activity
Control theory	homeostatic systems, stability
	simple optimal agent designs

1943	McCulloch & Pitts: Boolean circuit model of brain
1950	Turing's "Computing Machinery and Intelligence"
1952–69	Look, Ma, no hands!
1950s	Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
1956	Dartmouth meeting: "Artificial Intelligence" adopted
1965	Robinson's complete algorithm for logical reasoning
1966–74	AI discovers computational complexity Neural network research almost disappears
1969–79	Early development of knowledge-based systems
1980–88	Expert systems industry booms
1988–93	Expert systems industry busts: "AI Winter"
1985–95	Neural networks return to popularity
1988–	Resurgence of probability; general increase in technical depth "Nouvelle AI": ALife, GAs, soft computing
1995–	Agents agents everywhere . . .

Currently . . .

- Intelligent agents –
  complex single agents, or agent systems from light-weight individual agents
- Distributed intelligence –
  the internet and prevasive computing makes the question of dealing with and understanding distributed information and reasoning more and more important
- Semantic Web –
  an ambitious proposal to allow all sorts of agents to communicate using *descriptions* rather than *pointers*

Today

- Course organisation
- What is intelligence, what is AI?
- First idea of *rational agent*
- A spot of history