Some notes and solutions to Russell and Norvig's Artificial Intelligence: A Modern Approach (AIMA, 3rd edition)

Peter Danenberg

12 October 2011

Contents

1 DONE 1.1

CLOSED: *2011-10-10 Mon 03:03*

Intelligence A spontaneous faculty for associating impressions (more general than ideas); synthesizing abstractions from disparate stimuli; deducing conclusions from abstractions.

Intelligence is an emergent property of simples like e.g. neurons.

Artificial intelligence Mechanism for performing association, abstraction, deduction which appears to be spontaneous; may also be an emergent property of bit-pushing.

Agent Self-contained, autonomous input-processing mechanism.

Rationality The appropriate application of or *ratio*; this includes the mechanical process of deduction, as well as an ill-defined notion of common-sense.

Logical reasoning The mechanical aspect of rationality.

2 DONE 1.2

CLOSED: *2011-10-10 Mon 03:03*

The Mathematical Objection (3) still holds up: the halting problem; on the other hands, humans are also susceptible to the halting problem, aren't they? If one falls towards the humanity side of the humanity-rationality AI-axis, this deficit is reducible.

Lady Lovelace's Objection (6) is interesting: it denies *ex nihilo*; are genetic algorithms a counter-example?

The Argument from Informality of Behaviour (8) could be solved by fuzzy dispatch.

A modern refutation might be that there are not enough graduate students to make a satisfactory ontology of world-knowledge; thank the gods, then, for mechanical turks and unsupervised learning!

We came pretty damn close to 30% in the 2008 Loebner prize; why not double it to 60% in 2058? Despite Moore's law, let's say that AI proceeds linearly.

2.1 Turing, Computing Machinery and Intelligence

- Can machines think?

- It is A's object in the game to try and cause C to make the wrong identification.

    - Didn't realize there was an adversarial element to the Turing test.

- What will happen when a machine takes the part of A in this game?

- . . . drawing a fairly sharp line between the physical and the intellectual capacities of man.

    - A reasonable dualism

- May not machines carry out something which ought to be described as thinking but which is very different from what a man does?

- – The humanity/rationality plane of AI?

- Imitation game

  - – Simulacrum sufficeth

- It is probably possible to rear a complete individual from a single cell of the skin (say) of a man . . . but we would not be inclined to regard it as a case of "constructing a thinking machine".

- Digital computer:

  1. Store
  2. Executive unit
  3. Control

- It is not normally possible to determine from observing a machine whether it has a random element, for a similar effect can be produced by such devices as making the choices depend on the digits of the decimal for π.

- Discrete state machines: strictly speaking there are no such machines. Everything really moves continuously.

- This is reminiscent of Laplace's view that from the complete state of the universe at one moment of time, as described by the positions and velocities of all particles, it should be possible to predict all future states.

- This special property of digital computers, that they can mimic any discrete state machine, is described by saying that they are universal machines.

- "Are there imaginable digital computers which would do well in the imitation game?" → "Are there discrete state machines which would do well?"

- I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than $70 right identification after five minutes of questioning.$

  - – Russell/Norvig, 12: storage units: $10^1 5$
  - – Loebner prize:

    Elbot of Artificial Solutions won the 2008 Loebner Prize bronze award, for most human-like artificial conversational entity, through fooling three of the twelve judges who interrogated it (in the human-parallel comparisons) into believing it was human. This is coming very close to the 30% traditionally required to consider that a program has actually passed the Turing test.

* From a judge:

> He predicted that by the end of the century, computers would have a 30 per cent chance of being mistaken for a human being in five minutes of text-based conversation.

I thought this was mistaken (should be 70), but it is indeed correct.

– In other words, a damn-good guess.

- Conjectures are of great importance since they suggest useful lines of research.

- We might expect that He would only exercise this power in conjunction with a mutation which provided the elephant with an appropriately improved brain to minister to the needs of this soul.

- We like to believe that Man is in some subtle way superior to the rest of creation.

- "The consequences of machines thinking would be too dreadful." I do not think that this argument sufficiently substantial to require refutation. Consolation would would be more appropriate: perhaps this should be sought the transmigration of souls.

- There are limitations to the powers of discrete-state machines. The best known of these results is known as Gödel's theorem, and shows that in any sufficiently powerful logical system statements can be formulated which can neither be proved nor disproved within the system, unless possibly the system itself is inconsistent.

- "Will this machine every answer 'Yes' to any question?" It can be shown that the answer is either wrong or not forthcoming.

- The only way to know that a man thinks is to be that particular man. It is in fact the solipsist point of view.

- I do not wish to give the impression that I think there is no mystery about consciousness. There is, for instance, something of a paradox connected with any attempt to localise it.

- When a burnt child fears the fire and shows that he fears it by avoiding it, I should say that he was applying scientific induction.

- It would deliberately introduce mistakes in a manner calculated to confuse the interrogator.

- By observing the results of its own behaviour it can modify its own programmes so as to achieve some purpose more effectively.

- This is the assumption that as soon as a fact is presented to a mind all consequences of that fact spring into the mind simultaneously with it.

- The undistributed middle is glaring.

- I would defy anyone to learn from these replies sufficint about the programme to be able to predict any replies to untried values.

- A smallish proportion are super-critical. An idea presented to such a mind may give rise to a whole "theory" consisting of secondary, tertiary and more remote ideas.

  – Sponteneity

- These last two paragraphs should be described as "recitations tending to produce belief."

- The only satisfactory support that can be given will be that provided by waiting for the end of the century and then doing the experiment described.

- Estimates for the storage capacity of the brain vary from $10^1 0$ to $10^1 5$ binary digits.

  – Russell/Norvig (12): $10^1 3$ synapses

- At my present rate of working I produce about a thousand digits of programme a day, so that about sixty workers, working steadily through the fifty years might accomplish the job, if nothing went into the waste-paper basket.

  – Mythical man-month?

- The child-programme and the education process

- One might have a complete system of logical inference "built in". The store would be largely occupied with definitions and propositions. Certain propositions may be described as "imperatives". As soon as an imperative is classed as "well-established" the appropriate action takes place.

  – Compare McCarthy, Programs with Common Sense, regarding imperatives.

- These choices make the difference between a brilliant and a footling reasoner.

- We can only see a short distance ahead, but we can see plenty there that needs to be done.

3   TODO 1.3

Reflex actions are rational in the sense that they are the result of induction on e.g. hot objects and the scientific method (see Turing); though the acquisition may require intelligence (induction, storage), the reflex itself is not intelligent in the sense that it requires no induction: it is immediate.

4 TODO 1.4

At the very least, a program that can score better on IQ tests; the mapping from IQ-test-scores has been debated (though the correlation to income is well established), leading to the multiplication of "intelligences": emotional, spatial, synthetic, etc.

Aplysia, Eric Kandel

20,000 neurons; memory-updates/second: $10^-9; cycletime : 10^{15}$, high end;

Is memory-updates/second merely memory / cycle time? In which case: $20000 / 10^-9 =$

5 DONE 1.14

CLOSED: *2011-10-10 Mon 03:52*

1. The Japanese got this one; just a toy, though.

2. There is at least one driverless car in Cairo; it's not self-controlling, though, but rather remotely driven. Driving in clusterfuck-Cairo (like Athens) is taxing for humans, let alone AI. (Google's making political inroads in Nevada, though.) Sufficiently sensitive sensation of surrounding objects, conditions; physics; navigation; are required.

3. DARPA Grand Challenge

4. This robot fetches a sandwich.

5. Grocery IQ will order groceries; a week's worth, though?

6. Zia Mahmood got clowned once or twice; like poker, though, bridge is probabilistic and psychological.

7. TheoryMine is selling new computer-generated proofs for č15; standard objections apply.

8. The Bulhak-Larios Postmodernism Generator is funny; intentionally so?

9. Hilariously-named SHYSTER: ad-hoc expert system

10. Google Translate

11. Mechanically, but there is a human agent (telemanipulator); see this, though, where "In May 2006 the first AI doctor-conducted unassisted robotic surgery on a 34 year old male to correct heart arythmia."

6 Notes

6.1 1

- Two dimensions: thought vs. action, humanity vs. rationality.

- Physical simulation of a person is unnecessary for intelligence.

  - Mind-body dualism of Descartes?

- Cognitive science brings together computer models from AI and experimental techniques from psychology.

- Real cognitive science, however, is necessarily based on experimental investigation of actual humans.

- The standard of rationality is mathematically well defined and completely general.

- We will adopt the working hypothesis that perfect rationality is a good starting point for analysis.

- Limited rationality: acting appropriately when there is not enough time

- Materialism, which holds that the brain's operation according to the laws of physicas constitutes the mind.

- Logical positivism

- Carnap, The Logical Structures of the World, was probably the first theory of mind as a computational process.

- Intelligence requires action as well as reasoning.

- Actions are justified by a logical connection between goals and knowledge of the action's outcome.

- Regression planning system

- The leap to a formal science required a level of mathematical formalization: logic, computation, probability.

- The world is an extremely large problem instance.

- Models based on satisficing—making decisions that are "good enough"—gave a better description of actual human behavior.

- Searle: brains cause minds.

- Behaviorism

- "A cognitive theory should be like a computer program."

- Intelligence and an artifact

- Parallelism—a curious convergence with the properties of the brain.

- The state of a neuron was conceived of as "factually equivalent to a proposition which proposed its adequate stimulus." McCulloch and Pitts (1943)

  - Neural events and the relations among them can be treated by means of propositional logic.
  - For any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes.
  - For every net behaving under one assumption, there exists another net which behaves under the other and gives the same results.

- Perhaps "computational rationality" would have been more precise and less threatening, but "AI" stuck.

- AI from the start embraced the idea of duplicating human faculties such as creativity.

- John McCarthy referred to this period as the "Look, Ma, no hands!" era.

- "A physical symbol system has the necessary and sufficient means for general intelligent action."

- 1958 . . . McCarthy define Lisp, which was to become the dominant AI programming language for the next 30 years.

- It is useful to have a formal, explicit representation of the world and its workings and to be able to maniplutae that representation with deductive processes.

- McCarthy, Programs with Common Sense

  - In this program the procedures will be described as much as possible in the language itself and, in particular, the heuristics are all so described.
  - If one wants a machine to be able to discover an abstraction, it seems most likely that the machine must be able to represent this abstraction in some relatively simple way.
  - The improving mechanism should be improvable.
  - Must have or evolve concepts of partial success.
    * Something about ~1995 that made for a cute blog.
  - For example, to mest people, the number 3812 is not an object: they have nothing to say about it except what can be deduced from its structure. On the other hand, to most Americans the number 1776 is an object because they have filed somewhere the fact that it represents the year when the American Revolution started.

- One might conjecture that division in man between conscious and unconscious thought occurs at the boundary between stimulus-response heuristics which do not have to be reasoned about but only obeyed, and the others which have to serve as premises in deductions.

- Machine evolution (genetic algorithms): Friedberg, 1958, 1959.

    - Friedberg. 1958. A learning machine Part 1. IBM Journal of Research and Development, 2, 2–13.

        * From and intent, to be sure, are related quite discontinuously in the compact, economical programs that programmers wrte.

    - Friedberg, Dunham, North. 1959. A learning machine, Part 2. IBM Journal of Research and Development, 3, 282–287.

- Failure to come to grips with the "combinatorial explosion"

- The new back-propagation learning algorithms for multilayer netwrks that were to cause an enormous resurgence in neural-net research in the late 1980s were actually discovered first in 1969.

- Bruce Buchanan: a philosopher turned computer scientist

- DENDRAL was the first successful knowledge-intensive system (expert system).

- AI Winter

- Parallel Distributed Processing (Rumelhart, McClelland. 1986)

- Connectionist models: competitors to symbols models and logicist approach

- Ones that act rationally according to the laws of decision theory and do not try to imitate the thought steps of human experts

- Control theory deals with designing devices that act optimally on the basis of feedback from the environment.

6.2 Lectures

6.2.1 1

- AI: mapping from sensors to actuators

    - Voice, child-like engagement

- Fully vs. partially observable

- Deterministic vs. stochastic

9

- Discrete vs. continuous

- Benign vs. adversarial

- Uncertainty management

6.2.2 2

- Initial state

- $\texttt{actions}(state) \rightarrow a_1, a_2, a_3, \ldots$

- $\texttt{result}(state, action) \rightarrow state'$

- $\texttt{goal-test}(state) \rightarrow T|F$

- $\texttt{path-cost}(state \xrightarrow{action} state \xrightarrow{action} state) \rightarrow n$

- $\texttt{step-cost}(state, action, state') \rightarrow n$

- Navigate the state space by applying actions

- Separate state into three parts: ends of paths (frontier); explored and unexplored regions.

- Step-cost

- Tree-search

  - Family-resemlance; difference: which path to look at first.

- Depth-first search: shortest-first search

-