CSci 543
Advanced Artificial Intelligence

Fall 2017

Lecture: Dr. M. Eunjin Kim

Course Overview

- Intelligent Agents
- Problem Solving by Searching and Game-Playing
- Logical Systems: Knowledge and Reasoning
- Planning System
- Uncertainty
- Machine Learning
- Probabilistic Reasoning
- Decision Making

Chapter 1

Introduction

(in AIMA 3)

The field of Artificial Intelligence

- **Goal:** To build *autonomous intelligent entities* which *think* and *act rationally* as well as understand how human think.
- computes with *human-level intelligence*
- A huge variety of subfields:
  - General-purpose areas: perception(vision), knowledge representation, logical (automated) reasoning, (machine) learning, acting(robotics), communication(natural language processing), etc.
  - Specific tasks: game-playing, mathematical theorem proving, medical disease diagnosis, robot soccer play, etc.

What is AI?

- The automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning …
  -- Bellman, 1978
  The exciting new effort to make computers *think*… machines with *minds*, in the full & literal sense.
  -- Haugeland, 1985
- The study of how to make computers do things at which, at the moment, *people* are better.
  -- Rich + Knight, 1991
  The art of creating machines that *perform functions* that require intelligence when performed by *people*. -- Kurzweil, 1990
- The study of mental faculties through the use of computational models.
  -- Charniak + McDermott, 1985
  The study of the computations that make it possible to *perceive*, *reason*, and *act*. – Winston, 1992
- Computational *Intelligence* is the study of the design of intelligent *agents*. -- Poole et.al 1998
  AI … is concerned with *intelligent behavior* in artifacts.
  -- Nilsson, 1998.

5

What is AI? .. continued

- Systems that think like *humans*
- Systems that act like *humans*

It measure success in terms of fidelity to *human* performance

- Systems that think *rationally*
- *Systems that act rationally*.

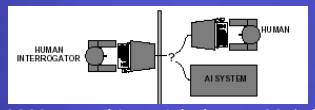
It measure success against an ideal concept of *intelligence*, i.e. *rationality*.

6

Acting Humanly: Turing Test

Turing(1950) " Computing machinery and intelligence"

- The test was designed to provide an operational definition of intelligence.
  -- *"Can machines think?" → "Can machines behave intelligently?"*
- A test of an intelligent machine which is indistinguishable from human.
- Operational test for intelligent behavior: the Limitation Game



- Predicted 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 areas of AI: knowledge representation, automated reasoning, natural language processing, machine learning + computer vision, robotics

Problem: Turing test is not reproducible, constructive or amenable to mathematical analysis

7

Thinking Humanly: Cognitive Science

- Needs *Theory of Human mind* ⇒ express the theory as a computer program.
- Cognitive Science:
  - 1960s "cognitive revolution": information-processing psychology replaced prevailing orthodox of behaviorism
  - computer models from AI + experimental techniques from Psychology
  - Aim: Construct precise/testable theories of the workings of the human mind.
- Requires scientific theories of internal activities of the brain
  - What level of abstraction? "Knowledge" or "circuits"?
  - How to validate? Requires
    1) Predicting and testing behavior 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

8

Thinking Rationally: *Laws of Thought*

- Normative (or prescriptive) rather than descriptive
- Aristotle: what are correct arguments/thought process?
  - ***Syllogism*** *: Socrates is a man. All men are mortal; therefore, Socrates is mortal, i.e.* $((S \to M) \wedge (M \to T)) \to (S \to T)$.
- Several Greek schools developed various forms of logic:
  notation and rules of derivation for thought;
  may or may not have proceeded to the idea of mechanization.
- Direct line through mathematics and philosophy to modern AI.
  - Formal logic (19^{th} –early 20^{th} C.) : syllogism ⇒ formal logic
  - Formal logic provided a precise notation for statements about all kinds of things and their relations.
  - Logician tradition: Build an intelligent systems through programs which take a description of a problem in logical notation and find the solution to the problem, if one exists.
- Aim: Build ***Computational Reasoning System*** through the ***Formal Logic***.
- Problems:
  1) Not easy to take informal knowledge/state in the logical notation when knowledge is uncertain.
  2) Big difference b/t solving a problem in *principle* and solving it in *practice*.

9

Acting Rationally

- Rational behavior: doing the *right thing.*
- The right thing is what is expected to *maximize goal achievement* given the available information.
  ⇒ *Rational Agents*
- rational reasoning (thinking) ⟶ rational action (behavior),
  however,
  rational reasoning ←/— rational action.
  - e.g.) blinking reflex
- Aristotle (Nicomachean Ethics);
  Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good
- Advantage:
  - More general: correct inference ∈ mechanisms for achieving rationality.
  - More amenable to scientific development.

10

Rational agents

- An agent is an entity that *perceives* and *acts.*
- *Rational agent*: one that acts so as to achieve one's goals at the *best*, given one's belief.
- Aim of AI: To study / design rational agents
- Abstractly, an agent is a function from percept histories to actions:
  $f: P^* \to A$
- For any given class of environments and tasks, we seek the agent (or class of agents) with the *best performance*
- Warning:
  computational limitations make perfect rationality unachievable
  → design best program for given machine resources.

11

AI Prehistory: Foundation

- **Philosophy (b.c. 428 -)**
  - Q: Can formal rules be used to draw valid conclusions?
  - Q: How does the mental mind arise from a physical brain?
  - Q: Where does knowledge come from?
  - Q: How does knowledge lead to action?

  logic, methods of reasoning,
  mind as physical system,
  foundations of learning, language, rationality
- **Mathematics (c. 800 –)**
  - Q: What are the formal rules to draw valid conclusions?
  - Q: What can be computed?
  - Q: How do we reason with uncertain information?

  formal theory of *logic*, proof algorithms,
  computation, (un)decidability, (in)tractability, *probability*
- **Psychology (1879 -)**
  - Q: How do humans and animals think and act?

  adaptation
  phenomena of perception and motor control
  experimental techniques (psychophysics, etc.)

12

AI Prehistory: continued

- **Economics (1776 -)**
  - Q: How should we make decisions so as to maximize payoff?
  - Q: How should we do this when others may not go along?
  - Q: How should we do this when the payoff may be far in the future?

    formal theory of rational decisions
- **Linguistics (1957 -)**
  - Q: How to does language relate to thought?

    knowledge representation, grammar
- **Neuroscience (1861 -)**
  - Q: How do brains process information?

    plastic physical substrate for mental activity
- **Computer engineering (1940 -)**
  - Q: How can we build an efficient computer?

    the invention of calculating devices

    the invention/evolution of operational, programmable, electronic computers: Robinson, Z-3, ABC, ENIAC, etc.
- **Control theory & Cybernetics (1948 -)**
  - Q: How can artifacts operate under their own control?

    homeostatic systems, stability

    simple optimal agent design

Potted history of AI

- 1943 McCulloch & Pitts: Boolean circuit model of brain. - Neural Network
- 1950 Turing's "Computing Machinery and Intelligence"
- 1950s Early AI programs, including Samuel's *Checkers* program, Newell & Simon's *Logic Theorist*, *General Problem Solver(GPS)*, Gelernter's *Geometry Theorem Prover*.
- 1956 Dartmouth meeting: "*Artificial Intelligence*" adopted. – J. McCarthy
- 1952-69 Early enthusiasm, great expectation!
- 1965 Robinson's complete algorithm for logical reasoning: *Resolution*
- 1966-74 AI discovers computational complexity

  Neural network research almost disappears
- 1969-79 Early development of *Knowledge-Based Systems: DENDRAL, MYCIN*
- 1980-88 *Expert systems* industry booms
- 1985-95 *Neural networks* return to popularity: *Connectionist models*
- 1988- Resurgence of *Probability*: general increase in technical depth
- "Nouvelle AI": *Fuzzy logic, Evolutionary Algorithms* ∈ *Soft Computing* - ***Computational Intelligence***
- 1995- *Agents theory*, *Swarm Intelligence*

Potted history of AI in Game

- Checkers (1994)

  Chinook ended 40-year-reign of human world champion Marion Tinsley in 1994. Used a precomputed endgame database defining perfect play for all positions involving 8 or fewer pieces on the board, a total of 444 billion positions.
- Chess (1997)

  Deep Blue defeated human world champion Garry Kasparov in a six-game match in 1997. Deep Blue searches 200 million positions per second, uses very sophisticated evaluation, and undisclosed methods for extending some lines of search up to 40 ply. Recently, HYDRA, RYBKA(2008-2009).
- Othello (1997)

  LOGISTELLO defeated the human world champion in 1997.

  Human champions refuse to compete against computers, who are too good.
- Jeopardy (2011)

  IBM's *Watson* competed against former winners Brad Rutter and Ken Jennings and defeated them.
- Alpha-Go (March, 2016)

  Alpha-Go developed by Google DeepMind played the board game Go against the human champion Lee Sedol in 5-game matches: 4-1(win-loss). In 2017, it also beat Ke Jie, the world #1 player of 2017 in 3-game matches, then retired.

The Emergence of Intelligent Agents

- The whole complete agent problem
- Aim: Understand the workings of agents embedded in real environments with continuous sensory inputs.
- Internet: the most important environments for intelligent agents

  - common in web-based applications.

  - internet tools: search engines, recommender systems, web site construction systems.
- Consequence of building complete agents:

  1. Realization that previously isolated subfields of AI might need to be recognized when their results are to be tied together.

     E.g.) due to the unreliable information from the sensory systems, reasoning & planning systems of AI should be able to handle uncertainty.

  2. The agent perspective is that AI has been drawn into much closer contact with other fields, e.g.) control theory, economics, etc.