COS402 Artificial Intelligence
Fall, 2006

Lecture I: Introduction

David Blei
Princeton University

(many thanks to Dan Klein for these slides.)
Course Site

http://www.cs.princeton.edu/courses/archive/fall06/cos402

- Updated syllabus
- Links to optional readings
- Information about subscribing to the mailing list
- All grading and lateness policies
- Assignments (including HW #0)
Course Details

- Prerequisites:
  - 217 and 226 (not taking 217 is usually no big deal...)
- Homework and Grading
- Late Policy
- Accounts
Today

- What is AI?
- History of AI
- What can AI do?
- What is this course?
- Precisely when are the robots going to take over?
Sci Fi AI
What is AI?

The science of making machines that can:

<table>
<thead>
<tr>
<th>Think like humans</th>
<th>Think rationally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Act like humans</td>
<td>Act rationally</td>
</tr>
</tbody>
</table>
Acting like humans

- Turing (1950) “Computing Machinery and Intelligence”
  - “Can machines think?” → “Can machines behave intelligently?”
  - Operational test for intelligent behavior: the Imitation Game (later dubbed “the Turing test”)

- Predicted by 2000, a 30% chance of fooling someone for 5 min
- Anticipated major arguments against AI for the next 50 years
- Suggested major components of AI: knowledge, reasoning, language understanding, learning

- Problem—Turing test is not reproducible or amenable to mathematical analysis
Thinking like humans

- **The Cognitive Science Approach**
  - 1960’s “cognitive revolution”: information-processing psychology replaced behaviorism.

- **Scientific theories of internal activities of the brain**
  - What level of abstraction? Knowledge or circuits?
  - **Cognitive Science**— Predicting and testing behavior of human subjects (top down)
  - **Cognitive Neuroscience**— Direct identification from neurological data (bottom up)
  - Both approaches are now distinct from AI
  - Have this in common: All the available theories do not explain anything resembling human-level general intelligence

- Hence, all three fields share one principal direction!
Thinking Rationally

• The “Laws of Thought” approach
  • What does it mean to “think rationally”?  
  • Normative/prescriptive rather than descriptive

• Logicist tradition
  • Logic– Notation and rules of derivation for thoughts
  • Aristotle– What are the correct thought processes?

• Problems
  • Not all intelligent behavior mediated by logical deliberation
  • What is the purpose of thinking?
  • Logical systems tend to do the wrong thing in the presence of uncertainty
Acting Rationally

- Rational behavior: Doing the “right thing”
  - The right thing: that which is expected to maximize goal achievement, given the available information
  - Doesn't necessarily involve thinking, e.g., blinking
  - Thinking can be in the service of rational action
  - Entirely dependent on goals!
  - Irrational ≠ insane, irrationality is sub-optimal action
  - Rational ≠ successful
- Our focus here: Rational agents
  - Systems which make the best possible decisions given goals, evidence, and constraints.
  - In the real world, usually lots of uncertainty and complexity
  - Usually, we are only approximating rationality.
Rational Agents

• An agent is an entity that perceives and acts
• This course is about designing rational agents.
• Abstractly, an agent is a function from percept histories to actions.
• For a class of environments and tasks, we seek the agent with the best performance
• Computational limitations make perfect rationality unachievable
• We want the best program for given machine resources
AI-adjacent fields

- **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 and statistics
- **Psychology**
  - Adaptation
  - Phenomena of perception and motor control
  - Experimental techniques (psychophysics, etc.)
- **Economics:** formal theory of rational decisions
- **Linguistics:** knowledge representation, grammar
- **Neuroscience:** physical substrate for mental activity
- **Control theory:**
  - Homeostatic systems, stability
  - Simple optimal agent designs
A Brief History of AI

- 1940-1950: Early days
  - 1943: McCulloch & Pitts: Boolean circuit model of brain
  - 1950: Turing's "Computing Machinery and Intelligence"

- 1950—70: Excitement: 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 by McCarthy
  - 1965: Robinson's complete algorithm for logical reasoning

- 1970—88: Knowledge-based approaches
  - 1969—79: Early development of knowledge-based systems
  - 1980—88: Expert systems industry booms
  - 1988—93: Expert systems industry busts: "AI Winter"

- 1988—: Statistical approaches
  - Resurgence of probability, focus on uncertainty
  - General increase in technical depth
  - Agents, agents, everywhere… "AI Spring"?

- 2000—: Where are we now?
What can AI do?

Quiz: Which of the following can be done at present?

- Play a decent game of table tennis? ✔
- Drive safely along a curving mountain road? ✔
- Drive safely along Nassau? ✗
- Buy a week's worth of groceries on the web? ✔
- Buy a week's worth of groceries at Wild Oats? ✗
- Discover and prove a new mathematical theorem? ?
- Converse successfully with another person for an hour? ✗
- Detect positive or negative bias in a movie review? ?
- Unload a dishwasher and put everything away? ✗
- Translate spoken English into spoken Swedish in real time? ✔
- Write an intentionally funny story? ✗
Logic

- Logical systems
  - Theorem provers
  - NASA fault diagnosis
  - Question answering

- Methods:
  - Deduction systems
  - Constraint satisfaction
  - Satisfiability solvers (huge advances here!)
Natural Language Processing

- **Speech technologies**
  - Automatic speech recognition (ASR)
  - Text-to-speech synthesis (TTS)
  - Dialog systems

- **Language processing technologies**
  - Machine translation
  - Information extraction
  - Information retrieval, question answering
  - Document organization, extracting themes
  - Text classification, spam filtering, etc…
## Learned Topics from a Corpus

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>human</td>
<td>evolution</td>
<td>says</td>
<td>disease</td>
<td>computer</td>
</tr>
<tr>
<td>genome</td>
<td>evolutionary</td>
<td>researchers</td>
<td>host</td>
<td>models</td>
</tr>
<tr>
<td>dna</td>
<td>species</td>
<td>colleagues</td>
<td>bacteria</td>
<td>information</td>
</tr>
<tr>
<td>genetic</td>
<td>organisms</td>
<td>team</td>
<td>diseases</td>
<td>data</td>
</tr>
<tr>
<td>genes</td>
<td>life</td>
<td>just</td>
<td>resistance</td>
<td>computers</td>
</tr>
<tr>
<td>sequence</td>
<td>origin</td>
<td>like</td>
<td>bacterial</td>
<td>system</td>
</tr>
<tr>
<td>gene</td>
<td>biology</td>
<td>new</td>
<td>new</td>
<td>network</td>
</tr>
<tr>
<td>molecular</td>
<td>groups</td>
<td>work</td>
<td>strains</td>
<td>systems</td>
</tr>
<tr>
<td>sequencing</td>
<td>phylogenetic</td>
<td>years</td>
<td>control</td>
<td>model</td>
</tr>
<tr>
<td>map</td>
<td>living</td>
<td>called</td>
<td>infectious</td>
<td>parallel</td>
</tr>
<tr>
<td>information</td>
<td>diversity</td>
<td>dont</td>
<td>malaria</td>
<td>methods</td>
</tr>
<tr>
<td>genetics</td>
<td>group</td>
<td>say</td>
<td>parasite</td>
<td>networks</td>
</tr>
<tr>
<td>mapping</td>
<td>new</td>
<td>get</td>
<td>parasites</td>
<td>software</td>
</tr>
<tr>
<td>project</td>
<td>two</td>
<td>see</td>
<td>united</td>
<td>new</td>
</tr>
<tr>
<td>sequences</td>
<td>common</td>
<td>university</td>
<td>tuberculosis</td>
<td>simulations</td>
</tr>
</tbody>
</table>
Vision (perception)

True caption
market people
Corr–LDA
people market pattern textile display

True caption
scotland water
Corr–LDA
scotland water flowers hills tree
Robotics

- Robotics
  - Part mech. eng.
  - Part AI
  - Reality much harder than simulations!

- Technologies
  - Vehicles
  - Rescue
  - Soccer!
  - Lots of automation…

- In this class:
  - We ignore mechanical aspects
  - Methods for planning
  - Methods for control
  - Rescue
  - Soccer!
  - Lots of automation…
Game Playing

- May, '97: Deep Blue vs. Kasparov
  - First match won against world-champion
  - ``Intelligent creative'' play
  - 200 million board positions per second!
  - Humans understood 99.9 of Deep Blue's moves
  - Can do about the same now with a big PC cluster

- Open question:
  - How does human cognition deal with the search space explosion of chess?
  - Or: How can humans compete with computers at all?

- 1996: Kasparov Beats Deep Blue
  “I could feel–I could smell–a new kind of intelligence across the table.”

- 1997: Deep Blue Beats Kasparov
  “Deep Blue hasn't proven anything.”
• Many applications of AI are decision making
  • Scheduling, e.g., airline routing, military
  • Route planning, e.g., mapquest
  • Medical diagnosis, e.g., Pathfinder system
  • Automated help desks
  • Fraud detection

Q: How do you get to Carnegie Hall?  
A: Practice!
Course Topics

• Search and Logic ("Classical" AI)
  • Heuristic search
  • First order and propositional logic

• Reasoning with Uncertainty
  • Bayesian networks
  • Statistical learning
  • Reinforcement learning

• Applications
  • Natural language
  • Vision
  • Robotics
  • Games