cis20.2
design and implementation of software applications II
spring 2008
session # III.1
intelligent systems 1

topics:

- fundamental issues in intelligent systems
- what is artificial intelligence (AI)?
- different views of AI

So, what is AI?

- AI is both science and engineering:
  - the *science* of understanding intelligent entities—of developing theories which attempt to explain and predict the nature of such entities;
  - the *engineering* of intelligent entities.
- four views of AI
  1. AI as *acting humanly*
     - as typified by the Turing test
  2. AI as *thinking humanly*
     - cognitive science.
  3. AI as *thinking rationally*
     - as typified by logical approaches.
  4. AI as *acting rationally*
     - the intelligent agent approach.

acting humanly

- A problem that has greatly troubled AI researchers: *when can we count a machine as being intelligent*?
- Most famous response due to Alan Turing, British mathematician and computing pioneer—**The Turing Test**:

  Human interrogates entity via teletype for 5 minutes. If, after 5 minutes, human cannot tell whether entity is human or machine, then the entity must be counted as intelligent.
- No program has yet passed Turing test! (Annual Loebner competition & prize.)
- A program that succeeded would need to be capable of:
  - natural language understanding & generation;
  - knowledge representation;
  - learning;
  - automated reasoning.
- Note no *visual* or *aural* component to basic Turing test—augmented test involves video & audio feed to entity.

Searle's Chinese Room (another test)

- you have a computer in a room that reads Chinese characters as input, follows a program and outputs (other) Chinese characters
- suppose this computer does this so well that it passes the Turing test (convinces a human Chinese speaker that it is talking to another human Chinese speaker)
- *does the computer understand Chinese*?
- suppose Searle is in the room, and he uses a dictionary to translate the input characters from Chinese to English; he then constructs his answer to the question, translates that back into Chinese and delivers the output—*does Searle understand Chinese*?
- of course not
- and this is Searle's argument: the computer doesn't understand it either, because alll it is doing is translating words (symbols) from one language (representation) to another

thinking humanly

- Try to understand how the mind works—how do we think?
- Two possible routes to find answers:
  - by *introspection* — we figure it out ourselves!
  - by *experiment* — draw upon techniques of psychology to conduct controlled experiments. ("Rat in a box"!)
- The discipline of *cognitive science*: particularly influential in *vision*, *natural language processing*, and *learning*.

thinking rationally

- Trying to understand how we *actually* think is one route to AI—but how about how we *should* think.
- Use *logic* to capture the *laws of rational thought* as *symbols*.
- *Reasoning* involves shifting symbols according to well-defined rules (like algebra).
- Result is *idealised* reasoning.
- Logicist approach theoretically attractive.

- Lots of problems:
  - *transduction* — how to map the environment to symbolic representation;
  - *representation* — how to represent real world phenomena (time, space, ...) symbolically;
  - *reasoning* — how to do symbolic manipulation
  - *tractability* — so it can be done by real computers within our life time
- We are still a long way from solving these problems.
- In general logic-based approaches are unpopular in AI at the moment.

acting rationally

- Acting rationally = acting to achieve one's goals, given one's beliefs.
- An *agent* is a system that perceives and acts; intelligent agent is one that acts rationally w.r.t. the goals we delegate to it.
- Emphasis shifts from designing *theoretically best* decision making procedure to best decision making procedure possible in circumstances.
- Logic may be used in the service of finding the best action — not an end in itself.

- Achieving *perfect* rationality — making the best decision theoretically possible — is not usually possible, due to *limited resources*:
  - limited time;
  - limited computational power;
  - limited memory;
  - limited or uncertain information about environment.
- The trick is to *do the best with what you've got*!
- This is easier than doing perfectly, but still tough.

differing views of AI

- *connectionist*:
  - intelligence is the result of emergent behaviors that are the result of a network of interconnected simple units
  - neural networks, perceptrons
  - Rumelhart and McClelland (1970-80's)
  - McCulloch and and Pitts (Princeton, 1940's)
- logic:
  - intelligence is enacted by resolving and unifying logic statements; "deduction"
  - early rules, that later became the basis for symbolic AI
  - McCarthy (Stanford, 1950's and onward)
- symbolic:
  - intelligence is represented in a computer by explicit structures that symbolize different bits of knowledge and activities; explicit rules control intelligent behavior
  - Newell and Simon (CMU, 1950's and onward)