→ Finish AI overview
   → Some material from Ch. 26
→ Intelligent agents (Ch. 2)
Notes


- Piazza: https://piazza.com/ucsb/winter2019/cs165a

- No Discussion Session this week
- No TA Office Hour this week
Stop me and ask questions!

• Do your classmates a favor:
  – If you are confused about something, there must be someone else who have the same question

• Do me a favor
  – It’s my first time teaching this course.
  – I’m not a native speaker.
  – I can calibrate the pace of the course accordingly to optimize your learning.

• “The only silly question is the one that you wanted to but never asked!” --- Unknown source
How to Learn AI/ML/DS
This class

1. What are the objectives of AI?
   - Can Machines Think?
   - Does it matter?
   - Should AI replicate Human Intelligence?

2. Formally setting up the problem
   - Intelligence Agents
   - Task environment
   - Model vs. reality
AI = “A” + “I”

• Artificial
  – As in “artificial flowers” or “artificial light”?

• Intelligence
  – What is intelligence?
    ♦ The capacity to acquire and apply knowledge
    ♦ The faculty of thought and reason
    ♦ Symbol manipulation, grounded in perception of the world
    ♦ The computational part of the ability to achieve goals in the world
  – What makes someone more/less intelligent than another?
  – Are {monkeys, ants, trees, babies, chess programs} intelligent?
  – How can we know if a machine is intelligent?

Turing Test (Alan Turing, 1950), a.k.a. The Imitation Game
The "standard interpretation" of the Turing Test, in which player C, the interrogator, is given the task of trying to determine which player – A or B – is a computer and which is a human. The interrogator is limited to using the responses to written questions to make the determination. (wiki)
"On the Internet, nobody knows you're a dog."
Can machines think? Strong vs Weak AI.

• “Strong AI”
  - Makes the bold claim that computers can be made to think on a level (at least) equal to humans
  - One version: The Physical Symbol System Hypothesis
    ♦ Takes physical patterns (symbols), combining them into structures (expressions) and manipulating them (using processes) to produce new expressions.
    ♦ A physical symbol system has the necessary and sufficient means for general intelligent action
    ♦ Intelligence = symbol manipulation (perhaps grounded in perception and action)

• “Weak AI”
  - Some “thinking-like” features can be added to computers to make them more useful tools
  - Examples: expert systems, speech recognition, natural language understanding….
Philosophical and ethical implications

• Is “Strong AI” possible?
• If so (or even if not)…
  – Should we be worried? Is this technology a threat?
  – Is it okay to shut down an intelligent machine?
  – When will it happen? (Will we know?)
  – Will they keep us around? (Kurzweil, Moravec)
  – Might we become too dependent on technology?
  – Terrorism, privacy
  – Technological singularity (Vinge, Good)
  – Moral robots
• Main categories of objections to Strong AI
  – Nonsensical (Searle)
  – Impossible (Penrose)
  – Unethical, immoral, dangerous (Weizenbaum)
  – Failed (Wall Street)
Objections to “Thinking machines”

• Theological objection
• “Heads in the sand” objection
• Mathematical objection: Goedel’s incompleteness theorem
• The argument from consciousness
• Arguments from various disabilities
  – “Be kind, resourceful, beautiful, friendly, have initiative, have a sense of humour, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream, make some one fall in love with it, learn from experience, use words properly, be the subject of its own thought, have as much diversity of behaviour as a man, do something really new.”
• Lady Lovelace’s objection
• Argument from continuity in the nervous system
• The argument from informality of behavior
  – Qualification problem
• The argument from ESP
Ladylove Lace: Creativity of AI!
The Chinese Room

John Searle, 1980a, 1980b, 1990b

**The Chinese Room argument.** Imagine that a man who does not speak Chinese sits in a room and is passed Chinese symbols through a slot in the door. To him, the symbols are just so many squiggles and squiggles. But he reads an English-language rule book that tells him how to manipulate the symbols and which ones to send back out. To the Chinese speakers outside, whoever (or whatever) is in the room is carrying on an intelligent conversation. But the man in the Chinese Room does not understand Chinese; he is merely manipulating symbols according to a rule book. He is instantiating a formal program, which passes the Turing test for intelligence, but nevertheless he does not understand Chinese. This shows that instantiation of a formal program is not enough to produce semantic understanding or intentionality. 

Note: For more on Turing tests, see Map 2. For more on formal programs and instantiation, see the “Is the brain a computer?” arguments on Map 1, the “Can functional states generate consciousness?” arguments on Map 6, and sidebar, “Formal Systems: An Overview” on Map 7.

© MacroVU Press
The Chinese-Room Argument

- It’s possible to pass Turing Test, yet not (really) think

I (native Chinese speaker)

\[ \text{story + questions} \]
\[ \text{(in Chinese)} \]
\[ \text{responses} \]
\[ \text{(in fluent Chinese)} \]

H (who can’t understand Ch.)

(Eng.) program for manipulating [Ch.] “squiggles”
The Chinese Room Argument (Searle)

- Computer programs are formal, syntactic entities.
- Minds have mental contents, or semantics.
- Syntax by itself is not sufficient for semantics.
- Brains cause minds.
The textbook view of this problem.

- “Strong AI” vs. “Weak AI” remains unsettled, but it’s outcome bears little significance.

- Focus on using AI solve problems.
- And pay attention to ethics and social impacts.
Definitions of AI

• Thinking humanly
  – Cognitive science

• Acting humanly
  – Turing test

• Thinking rationally
  – Logic

• Acting rationally
  – The approach adopted here
## Definitions of AI

<table>
<thead>
<tr>
<th>Thought processes and reasoning</th>
<th>Human</th>
<th>Ideal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systems that think like humans</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systems that act like humans</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systems that think rationally</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systems that act rationally</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Human/Biological Intelligence

• Thinking humanly (Cognitive modeling)
  – Cognitive science
    ♦ 1960s – Information processing replaced behaviorism as the dominant view in psychology
  – Cognitive neuroscience
    ♦ Neurophysiological basis of intelligence and behavior?

• Acting humanly (Operational intelligence)
  – The Turing Test – operational test for intelligent behavior
    ♦ What does it require?
  – Required: knowledge, reasoning, language understanding, learning…
  – Problem: It is not reproducible or amenable to mathematical analysis; rather subjective
Ideal/Abstract Intelligence

• Thinking rationally (Laws of Thought)
  – Rational thought: governed by “Laws of Thought”
  – Logic approach – mathematics and philosophy

• Acting rationally (Rational agents)
  – Rational behavior: doing the right thing
    ♦ Maximize goal achievement, given the available information (knowledge + perception)
  – Can/should include reflexive behavior, not just thinking
  – General rationality vs. limited rationality
  – Basic definition of agent – something that perceives and acts
  – The view adopted here
Replicating human intelligence?

- AI doesn’t *necessarily* seek to replicate human intelligence
- Sometimes more, sometimes less…

- “Essence of X” vs. “X”
- Examples
  - Physical vs. electronic newspaper
  - Physical vs. virtual shopping
  - Birds vs. planes

“Saying Deep Blue doesn’t really think about chess is like saying an airplane doesn’t really fly because it doesn’t flap its wings.”

– Drew McDermott
How can you tell it’s AI?

• It does something that is clearly “human-like”

…or…

• Separation of
  – data/knowledge
  – operations/rules
  – control

• Has
  – a knowledge representation framework
  – problem-solving and inference methods
What is an AI Program?

- AI programs can generally be thought of as comprising three *separated* parts
  - Data / knowledge (“knowledge base”)
  - Operations / rules (“production rules”)
  - Control
    - Which rules to apply when
    - Selecting operations and keeping track of their effects
    - Typically defined by the *search strategy*

- *Data* and *Operations* should be modular and easy to modify
AI and Intelligent Agents
What's an Agent?

"An intelligent agent is an entity capable of combining cognition, perception and action in behaving autonomously, purposively and flexibly in some environment."

- Possible properties of agents:
  - Agents are **autonomous** – they act on behalf of the user
  - Agents can **adapt** to changes in the environment
  - Agents don't only act **reactively**, but sometimes also **proactively**
  - Agents have **social ability** – they communicate with the user, the system, and other agents as required
  - Agents also **cooperate** with other agents to carry out more complex tasks than they themselves can handle
  - Agents **migrate** from one system to another to access remote resources or even to meet other agents
Our view of AI

• AIMA view – AI is building intelligent (rational) agents
  – Principles of rational agents, and
  – Models/components for constructing them

• Rational = “Does the right thing” in a particular situation
  – Maximize expected performance (not actual performance)

• So a rational agent does the “right” thing (at least tries to)
  – Maximizes the likelihood of success, given its information
  – How is “the right thing” chosen?
    ♦ Possible actions (from which to choose)
    ♦ Percept sequence (current and past)
    ♦ Knowledge (static or modifiable)
    ♦ Performance measure (wrt goals – defines success)
Our model of an agent

- An agent **perceives** its environment, **reasons** about its goals, and **acts** upon the environment
  - Abstractly, a function from percept histories to actions
    \[ f : P^* \rightarrow A \]

- Main components of an agent
  - Perception (sensors)
  - Reasoning/cognition
  - Action (actuators)

- Supported by
  - knowledge representation, search, inference, planning, uncertainty, learning, communication….
Our view of AI (cont.)

• So this course is about designing rational agents
  – Constructing $f$
  – For a given class of environments and tasks, we seek the agent (or class of agents) with the “best” performance
  – Note: Computational limitations make complete rationality unachievable in most cases

• In practice, we will focus on problem-solving techniques (ways of constructing $f$), not agents per se
 Ideal Rational Agent

• In other words…

“For each possible percept sequence, an ideal rational agent should do whatever action is expected to maximize its performance measure, on the basis of the evidence provided by the percept sequence and whatever built-in knowledge the agent has.”

Note that:

Rational ≠ Omniscient
Rational ≠ Clairvoyant
Rational ≠ Successful
Describing the Task Environment

- ** PEAS – Performance measure, Environment, Actuators, Sensors 
  - Goals may be explicit or implicit (built into performance measure)

- Not limited to physical agents (robots)
  - Any AI program
The Vacuum World

Performance measure, Environment, Actuators, Sensors
The Vacuum World

- Performance (P)
  - Keep world clean
  - Possible performance measures

- Environment (E)
  - Location
  - Cleanliness

- Three actions (A)
  - Move right
  - Move left
  - Remove dirt

- Sensed information (percepts) of environment (S)
  - Two locations
    - Left
    - Right
  - Two states
    - Dirty
    - Clean
## PEAS Descriptions

<table>
<thead>
<tr>
<th>Agent Type</th>
<th>P</th>
<th>E</th>
<th>A</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical diagnosis system</td>
<td>Healthy patient, minimize costs</td>
<td>Patient, hospital</td>
<td>Questions, tests, treatments</td>
<td>Symptoms, findings, patient’s answers</td>
</tr>
<tr>
<td>Satellite image analysis system</td>
<td>Correct categorization</td>
<td>Images from orbiting satellite</td>
<td>Print a categorization of scene</td>
<td>Pixels of varying intensity, color</td>
</tr>
<tr>
<td>Part-picking robot</td>
<td>Place parts in correct bins</td>
<td>Conveyor belt with parts</td>
<td>Pick up parts and sort into bins</td>
<td>Pixels of varying intensity</td>
</tr>
<tr>
<td>Refinery controller</td>
<td>Maximize purity, yield, safety</td>
<td>Refinery</td>
<td>Open, close valves; adjust temperature</td>
<td>Temperature, pressure readings</td>
</tr>
<tr>
<td>Interactive English tutor</td>
<td>Maximize student’s score on test</td>
<td>Set of students</td>
<td>Print exercises, suggestions, corrections</td>
<td>Typed words</td>
</tr>
</tbody>
</table>
Environments

- Properties of environments
  - Fully vs. partially observable
  - Deterministic vs. stochastic
  - Episodic vs. sequential
  - Static vs. dynamic
  - Discrete vs. continuous
  - Single agent vs. multiagent

- The environment types largely determine the agent design
- The real world is partially observable, stochastic, sequential, hostile, dynamic, and continuous
  - Bummer…
Generic Agent Program

• Implementing $f: P^* \rightarrow A$ ...or... $f(P^*) = A$
  – Lookup table?
  – Learning?

Knowledge, past percepts, past actions

function SKELETON-AGENT(percept) returns action
  static: memory, the agent’s memory of the world

  memory ← UPDATE-MEMORY(memory, percept)
  action ← CHOOSE-BEST-ACTION(memory)
  memory ← UPDATE-MEMORY(memory, action)
  return action

e.g.,
Table-Driven-Agent
Add percept to percepts
LUT [percepts, table]
NOP
Basic types of agent programs

- Simple reflex agent
- Model-based reflex agent
- Goal-based agent
- Utility-based agent
- Learning agent
Simple Reflex Agent

- Input/output associations
- Condition-action rule: “If-then” rule (*production rule*)
  - If *condition* then *action* (if in a certain state, do this)
  - If *antecedent* then *consequent*
Simple Reflex Agent

function SIMPLE-REFLEX-AGENT(percept) returns action
static: rules, a set of condition-action rules

state ← INTERPRET-INPUT(percept)
rule ← RULE-MATCH(state, rules)
action ← RULE-ACTION[rule]
return action

• Simple state-based agent – Classify the current percept into a known state, then apply the rule for that state
Model-Based Reflex Agent

- Internal state – keeps track of the world, models the world
Model-Based Reflex Agent

function REFLEX-AGENT-WITH-STATE(\textit{percept}) \textbf{returns} \textit{action}

\textbf{static}: \textit{state}, a description of the current world state
\textit{rules}, a set of condition-action rules

\textit{state} \leftarrow \textsc{Update-State}(\textit{state}, \textit{percept})
\textit{rule} \leftarrow \textsc{Rule-Match}(\textit{state}, \textit{rules})
\textit{action} \leftarrow \textsc{Rule-Action}[\textit{rule}]
\textit{state} \leftarrow \textsc{Update-State}(\textit{state}, \textit{action})
\textbf{return} \textit{action}

- State-based agent – Given the current state, classify the current percept into a known state, then apply the rule for that state
Goal-Based Agent

- Goal: immediate, or long sequence of actions?
  - Search and planning – finding action sequences that achieve the agent’s goals
Utility-Based Agent

- Utility function: Specifies degree of usefulness (happiness)
  - Maps a state onto a real number
- Conflicting goals; ordering of goals

\[ U = \frac{1}{\text{cost}} \]
Learning Agent

![Diagram of Learning Agent]

- Performance standard
- Critic
- Sensors
- Feedback
- Learning element
- Changes
- Knowledge
- Problem generator
- Actuators
- Environment

43
When to use which type of agent?

- Depends on the problem (environment)
  - Stochastic/deterministic/stateful/adversarial …
- Depends the amount of data available
  - Often we need to learn how the world behaves
- Depends on the dimensionality of your observations

Solving the right problem approximately vs Solving an approximation of the problem exactly
All models are wrong, but some models are useful

– George Box
(1919 - )
Next two lectures: reasoning with uncertainty

- Probability
- Statistics
- Graphical models / BayesNets