

CSCI 400 Artificial Intelligence
David Keil, Framingham State University

Topic 1: Cognition and computation

1. Cognition
2. The rational-agent approach to AI
3. Models of computation

Readings: Ch. 1-2

Inquiry

- Is the general problem for any cognitive system to give *timely responses to percepts from an environment*?
- Can intelligence be summarized as *reasoning; ontogenetic learning (single agent); phylogenetic learning (evolutionary); and sociogenetic learning (distributed)*?

Objectives

- 1a. Explain what *hard computational problems* means formally and in practice
- 1b. Describe the *computational-representational understanding of mind*
- 1c. Distinguish classes of problem environment
- 1d. Describe neural and transition-system models of computation
- 1e. Describe a reflex agent within the rational-agent model of AI

1. Cognition

- *Cognitive science*: “the interdisciplinary study of mind and intelligence, embracing philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology” (ix)
- *Founders of cognitive science*:
 - John McCarthy, theorist of AI based on logic
 - Allen Newell, Herbert Simon, theorists
 - Marvin Minsky, AI theorist
 - Noam Chomsky, theorist of linguistics
 - George Miller, researcher in human memory

Cognitive theories

- A *theory* postulates representational structures and operations on them
- A *computational model* interprets structures and operations by analogy that have programs with data structures and algorithms
- *Programs* that implement this model can help test and refine it
- *Central hypothesis*: “Thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures”

Criteria for evaluation of theories

- Three kinds of problem solving are to be accounted for by theories of representation: planning, decision making, explanation
- Must account for *learning, language*
- Must help understand *human* cognition
- Neurological plausibility
- Practical applicability

Approaches to cognitive science

- *Logic*: concerned with *form* of reasoning
- *Rules*: condition-action pairs representing long-term memory
- *Concepts*: representations of typical entities or situations, with instances, frames, schemas, scripts
- *Analogies*: relying on past situations for guidance in new ones
- *Images*: Useful for storing, processing virtual or spatial information
- *Connections*: parallel distributed processing, parallel constraint satisfaction, on model of the brain

Theories of mind

- *Dualism*: mind and body are separate
- *Materialism* (monism): mind is not a different substance from body
- *Functionalism*: mental states may be either brain states or based on other physical systems
- *Idealism* (no longer widely held): what is mental is all or dominates what is material

Computational-Representational Understanding of Mind

- CRUM: logic, rules, concepts, analogies, images, connections
- The mind's *computation* capacity operates on its *representation* of the world
- Possible limitations of CRUM: weakness or failure addressing brain, emotions, consciousness, body, world, dynamic systems, social aspects of intelligence

Computation and representation

- *Definition of computational system*: One whose physical states can be seen as representations of states of another system (symbols); and whose state transitions can be seen as operations on representations
- *Representation*: “a structure that stands for something by virtue of relations such as similarity, causal history, and connections with other representations”

Concepts

- *Definition*: “representations of typical entities or situations,” e.g.,. frames, schemas, scripts; with *slots*
- *Example*: a *course* is a kind of process with an instructor, room, meeting time, title; with subcategories *lecture*, *seminar*
- Concepts are related to other concepts in a situation, such that an activation is spread from one to another
- Inferences about a situation are made by use of inheritance and containment relationships

Analogic reasoning

- *Definition*: “dealing with a new situation by adapting a similar familiar situation”
- In AI, this is *case-based reasoning*
- Beliefs about causal relations give power to analogic reasoning
- Typical application: a *target* problem and a heuristic *source* problem are identified

Diverse theories of representation

- Six approaches rather than one may be appropriate for current state of cognitive science research
- Formal logic and rules systems have conciseness and independence as representations
- Concepts, analogies, and images bundle information into structures
- A unified theory of representation would have to bring together sensory richness, organizing power, and verbal expressiveness

Consciousness

- “Some of the neurological and computational elements of a theory of consciousness are starting to appear”
- A theory of consciousness must describe mechanisms that explain it
- *Mechanism*: “a group of components that have properties and relations to each other that produce regular changes in those properties and relations”
- Evidence of consciousness is associated with life, biological activity, and specifically brain processes – neural, electrical, chemical

Emotions and cognition

- The view is growing that emotions are part of decision making, including rational decisions
- Emotions provide *appraisal* or problem-solving situations, enabling *focus* and *action*
- Empathy and bodily reactions are part of emotional apparatus
- Emotional associations seem to be part of representations of concepts
- *Definition*: “an emotion is a pattern of activation in a population of neurons with connections to both inferential and sensory brain areas”

Social and distributed cognition

- *Social*: how people use reasoning about the mental activities of others
- *Distributed*: study of thinking as occurring through cooperation
- Emotional contagion and empathy are examples of distributed cognition
- *Culture*: a system of agreed-upon guidelines and meanings

Mind (H. Simon)

- “A system that produces thought, viewed at a relatively high level of aggregation,” e.g., far above the level of neurons or processor operations
- *Primitives of mind*: symbols, structures of symbols, and symbol-manipulation processes
- *Hardware of mind*: protoplasm, silicon, other
- *Thesis*: Programs exist that process symbols similarly to the human mind. These programs don't just *simulate* thinking they *think*

The human mind

- *Heuristics* are used to reduce size of search
- Experts acquire an indexed body of about 50K “chunks” of knowledge in longterm memory, which are accessed by recognition of the index value
- Short-term memory stores about seven chunks; attention is required, making the symbolic thought process *serial*
- Sensory organs and processing are highly *parallel*

Intentionality and semantics

- Human thinking has
 - Understanding of meaning
 - Goals
- Some computer programs demonstrably have this intentionality (Simon)
- *Semantic intention* may or may not be accompanied by *consciousness*
- *Denotation* is correspondence between symbols and meaning; humans and machines have this capacity, according to Simon

Representation of knowledge

- *Forms*
 - Raster-based pictures
 - Nodes and edges graph diagrams
 - Symbolic, e.g., propositional logic
- Computers and humans use these representations
- Representation may be *informationally* equivalent, but not *computationally* (e.g., Roman numerals are informationally equivalent to Arabic ones, but not computationally)

2. The rational-agent approach to AI

- *Agent*: “anything that can be viewed as perceiving its environment through sensors and acting on that environment through actuators”
- *Rational agent*: “one that behaves as well as possible”
- *Agent function*: maps from *sequence* of percepts to actions
- *Alternative conception*: An agent has a *policy*, which is a mapping of *perceived states of the environment* to actions

Agent performance measures

- Are criteria for success of actions
- Best guide is state of environment, not action of agent
- *Rationality* maximizes *expected* long-term reward
- *Information gathering* is often a useful action
- *Learning* (changing policy) can be useful

Environments

Dimensions of classification, where the second in each pair below is more difficult:

- Full vs. partial observability
- Deterministic vs. stochastic
- Episodic vs. sequential (persistent)
- Static vs. dynamic
- Discrete vs. continuous
- Single-agent vs. multi-agent

Examples

- *Chess board* is fully observable, chess game is deterministic, sequential, dynamic, discrete
- *Stock market* is partially observable, stochastic, sequential, dynamic, discrete
- *Real world* is continuous, partially observable, stochastic, sequential, dynamic, multi-agent

Agent types

- *Simple reflex*: ignores past percepts, acts on current one. Effective only in fully observable environment
- *Model-based reflex*: maintains internal model of environment, has rules to map model states to actions
- *Goal-based*: augments agent with a representation of goal states, enabling flexibility
- *Utility-based*: goal set is replaced with performance measures that enable comparison of possible states for *utility* (reward-generating features)
- *Learning*: self modification by agent to use feedback to improve performance

Multi-agent emergent AI

- Intelligence is a property of a society of agents
- Agents are autonomous, situated in environment, interactive, cooperative
- Intelligence emerges from interaction in a structured collection of agents
- **Agent**: “an element of a society that can perceive (often limited) aspects of its environment and affect that environment either directly or through cooperation with other agents”

Intelligence as a social outcome

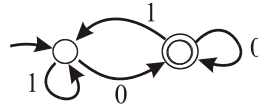
- H. Simon associated intelligence with interaction; ants with simple behavior in complex environment show collective intelligence
- Humans are products of culture and environment
- Rationalist good-old-fashioned AI (GOF AI) saw intelligence as a set of logical propositions about the world
- Existentialist-phenomenologist view is that the role of intelligence is to enable survival in the world

3. Models of computation

- To serve the idea that intelligence is computation, we may define some *models* of algorithmic computation
- A *transition system* is a labeled digraph, where vertices denote *state* (memory), edges denote transitions and labels
- *Components*: alphabet (set of symbols); state set (a state can be anything); transition function or relation (rules for going from one state to another)

Some models

- *Finite automata*: set of states is finite
- *Pushdown automata*: finite state set is augmented by unbounded stack
- *Turing machines*: state set is augmented by read/write tape and tape head
- *Random access machines*: memory cells are addressable as in Von Neumann architecture
- *Parallel and connectionist models*: supporting notion of concurrency



Algorithms and computability

- Methods, procedures, and some high-level-language “functions” *implement* algorithms
- We distinguish functions from the algorithms that *compute* them
- Some functions are *uncomputable*
- All computable functions are *recursively definable*

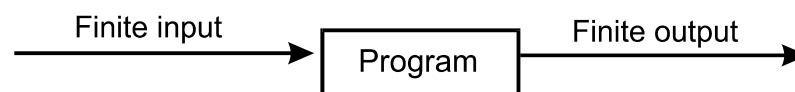
Interaction vs. algorithms

- New paradigm: computing is not just *algorithms*, but also *interaction*
- Problems are not *functions*, but *services* and *missions*, with Quality-of-Service constraints
- *Examples of interaction*: agent computation in dynamic environments; reinforcement learning; networked computing
- Is interaction *more powerful* than algorithmic computing?
- Are *multi-agent systems* more powerful than sequential interaction?

Algorithms

Algorithmic computation (Knuth):

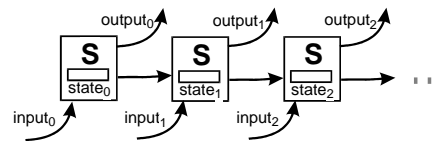
The effective transformation of a finite, pre-specified input, to a finite output, in a finite number of steps.



- Algorithms *compute functions*
- A system that executes an algorithm is *closed*
- Algorithms are equivalent to Turing-machine computation

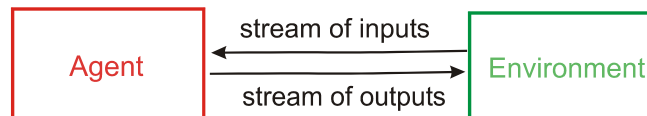
Interactive computation

- Feature of computing today: Computation as an ongoing *service*, not assumed to terminate
- Must solve problems whose inputs cannot be completely specified a priori
- Dynamic input and output *during* computation
- *Persistence of state* between interaction steps
- *Environment* is an active partner in computation



Sequential interaction

(Synchronous) sequential interactive computation:
Interaction involving two participants, at least one of which is a finite computing agent (machine, device).



- Characterized by a single interaction stream of input alternating with output
- If one participant is an agent, the other is its *environment*
- Interaction may involve changes of state

The brain as a computational system

- In the *brain*, computation is interaction of groups of neurons that result in changes in firing patterns, where firing patterns represent aspects of the world
- The brain is not one connectionist network, but a system of subnetworks that interact
- Neurons may represent concepts, as firing rates – certain neurons or groups of neurons firing many times per second may represent presence of an instance of a concept
- Neurons may also encode features of the world by spike train patterns – e.g., 0101 vs. 1001

The brain

- Part of nervous system that connects to every part of the body
- Also communicates via hormones in bloodstream
- The basis for connectionist models of computation and intelligence (see subtopic 5.2)
- The brain *adapts* by modifying its internal connections

Some brain components and functions

- *Memory* consists of changes in weights of synapses
- *Cerebral cortex* supports executive function and personality
- *Glial cells* and chemicals supplant neuron firing in brain function
- *Amygdala* (reptile part of brain) supports fight/flight responses

Neurons

- Brain is composed of neuron cells, each with up to thousands of connections to other neurons, communicating by electrochemical impulses
- Neurons' inputs are *dendrites*, outputs are *axons*, I/O contacts between neurons are *synapses*
- Animal neurons are of similar complexity
- Bees: 1M neurons, ants 250K, snails 20K, nematode worms 300, humans 100G
- Neurons *fire* (emit impulse) when inputs' sum exceeds a threshold at synapse

Connectionist models

- Concepts may be represented as nodes with positive or negative one or two-way excitation relation [pic., p. 114]
- Feedforward networks propagate inputs through hidden layer, to output layer [pic, p. 115]
- *Concept* “can be seen as a pattern of activation of nodes in a distributed network”
- Representations in connectionist systems consist of many units linked by excitatory and inhibitory connections

Concepts

algorithm	consciousness	rational agent
analogic reasoning	distributed cognition	reflex agent
analogy	dynamic system	representation
brain	emotion	representational
case-based reasoning	episodic environment	structure
chaos	finite automaton	satisfiability
Church-Turing thesis	goal-based agent	social cognition
cognition	image	tractable
cognitive science	inheritance	transition system
computational	intractable	Turing machine
complexity	logic	uncomputable
computational model	mind	undecidable
CRUM	neuron	utility-based agent
concept	parallel distributed	validity
	processing	

References

George Luger. *Artificial Intelligence*. Addison Wesley, 2005.

Stuart Russell and Peter Norvig. *AI: A Modern Approach, 2nd ed.* Prentice Hall, 2003.

Paul Thagard. *Mind: Introduction to Cognitive Science, 2nd ed.* MIT press, 2005.