Foundations of Artificial Intelligence

AIMA Chapter 1 (after Russell and Norvig)

Outline

◇ Administration

◇ What is AI?

- the understanding and building of intelligent entitites

◇ Foundations

◇ A brief history

◇ The state of the art

What is AI?

"[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)	"The study of mental faculties through the use of computational models" (Charniak+McDermott, 1985)
"The study of how to make computers do things at which, at the moment, people are better" (Rich+Knight, 1991)	"The branch of computer science that is concerned with the automation of intelligent behavior" (Luger+Stubblefield, 1993)

Views of AI fall into four categories:

Thinking humanly	Thinking rationally
Acting humanly	Acting rationally

Examining these, we will plump for acting rationally (sort of)

Acting humanly: The Turing test

Turing (1950) "Computing machinery and intelligence":

◊ "Can machines think?" $\longrightarrow$ "Can machines behave intelligently?"

◊ Operational test for intelligent behavior: the Imitation Game

◊ Predicted that 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 components of AI: knowledge, reasoning, language understanding, learning

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

Loebner Prize: "The first formal instantiation of a Turing test"

- http://www.loebner.net/Prizef/loebner-prize.html

Thinking humanly: Cognitive Science

1960s "cognitive revolution": information-processing psychology replaced prevailing orthodoxy of behaviorism

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 and thus direction: *the available theroeis do not explain (or engender) anything resembling human-level general intelligence*

Thinking rationally: Laws of Thought

Normative (or prescriptive) rather than descriptive

Aristotle: what are correct arguments/thought processes?

Several Greek schools developed various forms of logic:
    notation and rules of derivation for thoughts;
may or may not have proceeded to the idea of mechanization

Direct line through mathematics and philosophy to modern AI

Problems:
1) Not all intelligent behavior is mediated by logical deliberation
2) What is the purpose of thinking? What thoughts should I have?
3) Computationally intractable
4) Expressive inadequacies

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 reflex—but thinking should be in the service of rational action

Aristotle (Nicomachean Ethics):

> *Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good*

Beyond programs: autonomy, perception, persistence, adaptive, goal adoption

Rational agents

An agent is an entity that perceives and acts

Abstractly, an agent is a function from percept histories to actions:

$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance

Caveat: computational limitations make perfect rationality unachievable
$\rightarrow$ design best program for given machine resources

Course: Rational agents

This course is about designing rational agents

Thinking rationally is only one way of acting rationally

Rationality is more clearly defined and general than human-centered approaches

AI Prehistory and History

Many disciplines (philosophy, mathematics, economics, psychology, linguistics, computer engineering, control theory, neuroscience, and more) have contributed ideas, viewpoints, and techniques to AI

The history of AI has had cycles of success, misplaced optimism, and resulting retrenchments; cycles of new creativity and systematic refinement of best approaches

Potted history of AI

1943-1955: Gestation

1956: Birth

1952-1969: Great Expectations

1966-1973: Reality

1969-1979: Knowledge is Power

1980-present: AI and Industry

1986-present: The Return of Neural Networks

1987-present: AI Becomes a Science

1995-present: Intelligent Agents

AI becomes a science, Intelligent Agents

1987– Rapid increase in technical depth
Build on existing theories
Base claims on theorems and/or experments rather than intuition
Real world applications rather than toy examples
Replication of experiments with data and code repositories
Less isolationism

1995– Whole agents rather than fragments
Situated movement
Internet environments ("bots")

State of the art

Autonomous planning and scheduling (NASA)

Game playing (Deep Blue)

Automomous control (minivan steering)

Diagnosis (medicine)

Logistics planning (Gulf war)

Robotics (surgery)

Language understanding and generation (translation, dialogue)

Problem solving (crossword puzzles)

Summary

Different people think of AI differently (thinking or behavior, humans or ideal)

We will adopt the Rational Action view: an Intelligent Agent takes the best possible action in a situation

AI has its roots in many disciplines

The history of AI has had various cycles

Currently: greater use of the scientific method, progress in both theory and practice, integration within subfields and across disciplines