Chapter 1: Intelligent Agents

Intelligent Agents

Overview:

- Intelligent Agents
- Dimensions of complexity
- Sample systems
- Representations and Solutions

Artificial Intelligence

- Artificial Intelligence is the synthesis and analysis of computational agents that act intelligently.
- An agent is something that acts in an environment.
- An agent acts intelligently if:
  - its actions are appropriate for its goals and circumstances
  - it is flexible to changing environments and goals
  - it learns from experience
  - it makes appropriate choices given perceptual and computational limitations

Goals of Artificial Intelligence

- <mark>Scientific goal:</mark> to understand the principles that make intelligent behavior possible in natural or artificial systems.
  - analyze natural and artificial agents
  - formulate and test hypotheses about what it takes to construct intelligent agents
  - design, build, and experiment with computational systems that perform tasks that require intelligence
- <mark>Engineering goal:</mark> design useful, intelligent artifacts.
- Analogy between studying flying machines and thinking machines.

Agents acting in an environment
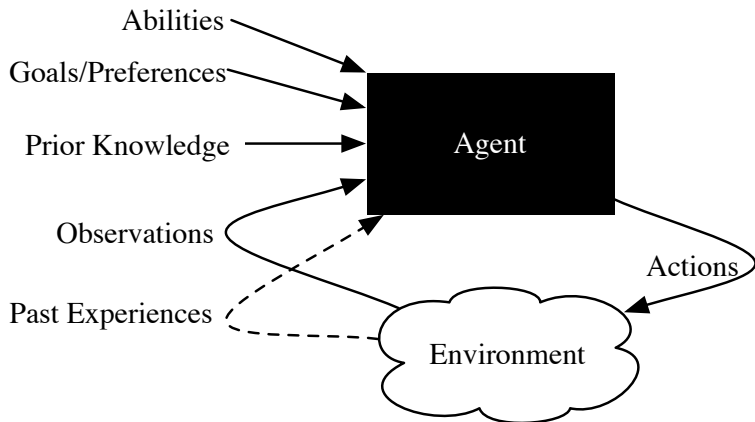
Examples of Agents

- **Organisations** Microsoft, Al Qaeda, Government of Canada, UBC, CS Dept,...
- **People** teachers, physicians, stock traders, engineers, researchers, travel agents, farmers, waiters...
- **Computers/devices** thermostats, user interfaces, airplane controllers, network controllers, games, advising systems, tutoring systems, diagnostic assistants, robots, Google car, Mars rover...
- **Animals** dogs, mice, birds, insects, worms, bacteria...

Inputs to an agent

- Abilities — the set of things it can do
- Goals/Preferences — what it wants, its desires, its values,...
- Prior Knowledge — what it comes into being knowing, what it doesn't get from experience,...
- History of observations (percepts, stimuli) of the environment
  - (current) observations — what it observes now
 - past experiences — what it has observed in the past

Example agent: robot

- **abilities:** movement, grippers, speech, facial expressions,. . .
- **goals:** deliver food, rescue people, score goals, explore,. . .
- **prior knowledge:** what is important feature, categories of objects, what a sensor tell us,. . .
- **observations:** vision, sonar, sound, speech recognition, gesture recognition,. . .
- **past experiences:** effect of steering, slipperiness, how people move,. . .

Example agent: teacher

- **abilities:** present new concept, drill, give test, explain concept,. . .
- **goals:** particular knowledge, skills, inquisitiveness, social skills,. . .
- **prior knowledge:** subject material, teaching strategies,. . .
- **observations:** test results, facial expressions, errors, focus,. . .
- **past experiences:** prior test results, effects of teaching strategies, . . .

Agents acting in an environment
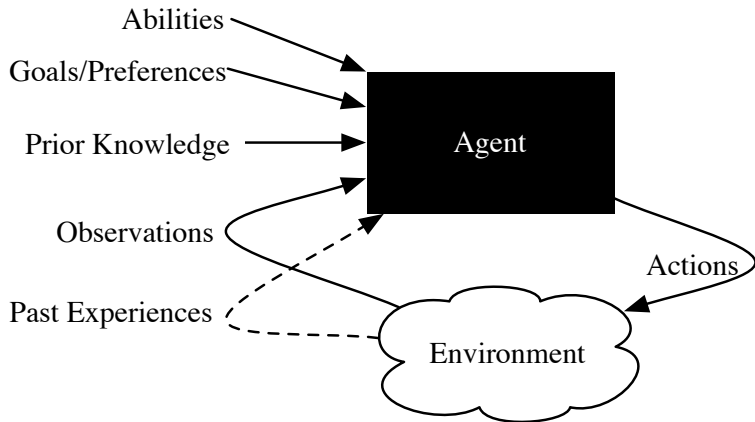
Dimensions

- Research proceeds by making simplifying assumptions, and gradually reducing them.
- Each simplifying assumption gives a dimension of complexity
  - ▸ multiple values in a dimension: from simple to complex
  - ▸ simplifying assumptions can be relaxed in various combinations

Dimensions of Complexity

- Flat or modular or hierarchical
- Explicit states or features or individuals and relations
- Static or finite stage or indefinite stage or infinite stage
- Fully observable or partially observable
- Deterministic or stochastic dynamics
- Goals or complex preferences
- Single-agent or multiple agents
- Knowledge is given or knowledge is learned from experience
- Perfect rationality or bounded rationality

Modularity

- Model at one level of abstraction: flat
- Model with interacting modules that can be understood separately: modular
- Model with modules that are (recursively) decomposed into modules: hierarchical
- Example: Planning a trip from here to see the Mona Lisa in Paris.
- Flat representations are adequate for simple systems.
- Complex biological systems, computer systems, organizations are all hierarchical
- A flat description is either continuous or discrete. Hierarchical reasoning is often a hybrid of continuous and discrete.

Succinctness and Expressiveness

Much of modern AI is about finding compact representations and exploiting the compactness for computational gains.
A agent can reason in terms of:

- Explicit states — a state is one way the world could be

- Features or propositions.
  - States can be described using features.
  - 30 binary features can represent $2^{30} = 1,073,741,824$ states.

- Individuals and relations
  - There is a feature for each relationship on each tuple of individuals.
  - Often an agent can reason without knowing the individuals or when there are infinitely many individuals.

Planning horizon

...how far the agent looks into the future when deciding what to do.

- **Static:** world does not change
- **Finite stage:** agent reasons about a fixed finite number of time steps
- **Indefinite stage:** agent reasons about a finite, but not predetermined, number of time steps
- **Infinite stage:** the agent plans for going on forever (process oriented)

Uncertainty

There are two dimensions for uncertainty

- uncertain dynamics
- uncertain perception (sensor information and its interpretation)

In each dimension an agent can have

- No uncertainty: the agent knows which world is true
- Disjunctive uncertainty: there is a set of worlds that are possible
- Probabilistic uncertainty: a probability distribution over the worlds.

Why Probability?

- Agents need to act even if they are uncertain.
- Predictions are needed to decide what to do:
  - definitive predictions: you will be run over tomorrow
  - disjunctions: be careful or you will be run over
  - point probabilities: probability you will be run over tomorrow is 0.002 if you are careful and 0.05 if you are not careful
  - probability ranges: you will be run over with probability in range [0.001,0.34]
- Acting is gambling: agents who don't use probabilities will lose to those who do.
- Probabilities can be learned from data and prior knowledge.

Uncertain dynamics

If an agent knew the initial state and its action, could it predict the resulting state?
The dynamics can be:

- Deterministic : the resulting state is determined from the action and the state
- Stochastic : there is uncertainty about the resulting state.

Perceptual Uncertainty

Whether an agent can determine the state from its observations:

- **Fully-observable**: the agent can observe the state of the world.
- **Partially-observable**: there can be a number states that are possible given the agent's observations.

Goals or complex preferences

- **achievement goal** is a goal to achieve. This can be a complex logical formula.
- **complex preferences** may involve tradeoffs between various desiderata, perhaps at different times.
  - ▶ **ordinal** only the order matters
  - ▶ **cardinal** absolute values also matter
- **Examples:** coffee delivery robot, medical doctor

Single agent or multiple agents

- **Single agent** reasoning is where an agent assumes that any other agents are part of the environment.
- **Multiple agent** reasoning is when an agent reasons strategically about the reasoning of other agents.

Agents can have their own goals: cooperative, competitive, or goals can be independent of each other

Learning from experience

Whether the model is fully specified a priori:

- Knowledge is given.
- Knowledge is learned from data or past experience.

Perfect rationality or bounded rationality

- **Perfect rationality:** the agent can determine the best course of action, without taking into account its limited computational resources.
- **Bounded rationality:** the agent must make good decisions based on its perceptual, computational and memory limitations.

Dimensions of Complexity

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single-agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

State-space Search

- **flat** or modular or hierarchical
- **explicit states** or features or individuals and relations
- static or finite stage or **indefinite stage** or infinite stage
- **fully observable** or partially observable
- **deterministic** or stochastic dynamics
- **goals** or complex preferences
- **single agent** or multiple agents
- **knowledge is given** or knowledge is learned
- **perfect rationality** or bounded rationality

Classical Planning

- **flat** or modular or hierarchical
- explicit states or features or **individuals and relations**
- static or finite stage or **indefinite stage** or infinite stage
- **fully observable** or partially observable
- **deterministic** or stochastic dynamics
- **goals** or complex preferences
- **single agent** or multiple agents
- **knowledge is given** or knowledge is learned
- **perfect rationality** or bounded rationality

Decision Networks

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Markov Decision Processes (MDPs)

- **flat** or modular or hierarchical
- **explicit states** or features or individuals and relations
- static or finite stage or **indefinite stage or infinite stage**
- **fully observable** or partially observable
- deterministic or **stochastic** dynamics
- goals or **complex preferences**
- **single agent** or multiple agents
- **knowledge is given** or knowledge is learned
- **perfect rationality** or bounded rationality

Decision-theoretic Planning

- **flat** or modular or hierarchical
- explicit states or **features** or individuals and relations
- static or finite stage or **indefinite stage or infinite stage**
- **fully observable** or partially observable
- deterministic or **stochastic** dynamics
- goals or **complex preferences**
- **single agent** or multiple agents
- **knowledge is given** or knowledge is learned
- **perfect rationality** or bounded rationality

Reinforcement Learning

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Classical Game Theory

- **flat** or modular or hierarchical
- **explicit states** or features or individuals and relations
- **static or finite stage** or indefinite stage or infinite stage
- fully observable or **partially observable**
- deterministic or **stochastic** dynamics
- goals or **complex preferences**
- single agent or **multiple agents**
- **knowledge is given** or knowledge is learned
- **perfect rationality** or bounded rationality

Humans

- flat or modular or **hierarchical**
- explicit states or features or **individuals and relations**
- static or finite stage or **indefinite stage or infinite stage**
- fully observable or **partially observable**
- deterministic or **stochastic** dynamics
- goals or **complex preferences**
- single agent or **multiple agents**
- knowledge is given or **knowledge is learned**
- perfect rationality or **bounded rationality**

The Dimensions Interact in Complex Ways

- Partial observability makes multi-agent and indefinite horizon reasoning more complex
- Modularity interacts with uncertainty and succinctness: some levels may be fully observable, some may be partially observable
- Three values of dimensions promise to make reasoning simpler for the agent:
  - Hierarchical reasoning
  - Individuals and relations
  - Bounded rationality

Four Example Application Domains

- **Autonomous delivery robot** roams around an office environment and delivers coffee, parcels,...

- **Diagnostic assistant** helps a human troubleshoot problems and suggests repairs or treatments. E.g., electrical problems, medical diagnosis.

- **Intelligent tutoring system** teaches students in some subject area.

- **Trading agent** buys goods and services on your behalf.

Domain for Delivery Robot
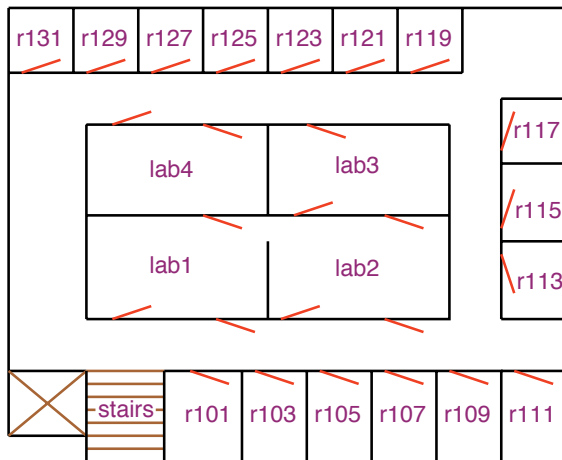
Autonomous Delivery Robot

Example inputs:

- **Abilities:** movement, speech, pickup and place objects.
- **Prior knowledge:** its capabilities, objects it may encounter, maps.
- **Past experience:** which actions are useful and when, what objects are there, how its actions affect its position.
- **Goals:** what it needs to deliver and when, tradeoffs between acting quickly and acting safely.
- **Observations:** about its environment from cameras, sonar, sound, laser range finders, or keyboards.

What does the Delivery Robot need to do?

- Determine where Craig's office is. Where coffee is...
- Find a path between locations.
- Plan how to carry out multiple tasks.
- Make default assumptions about where Craig is.
- Make tradeoffs under uncertainty: should it go near the stairs?
- Learn from experience.
- Sense the world, avoid obstacles, pickup and put down coffee.

Domain for Diagnostic Assistant

Diagnostic Assistant
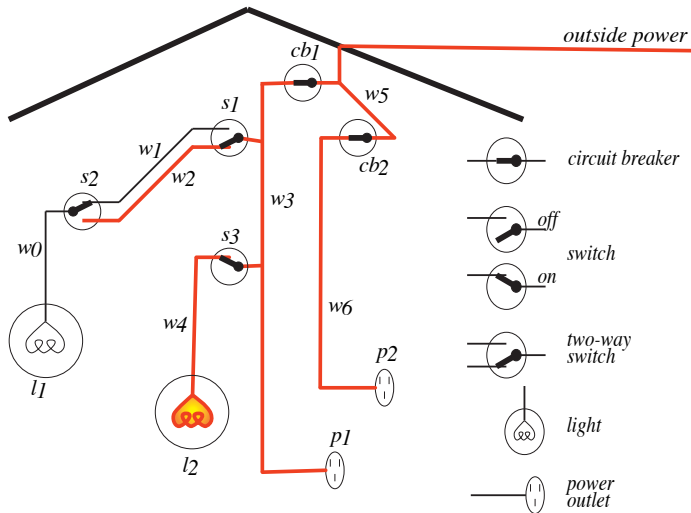
Example inputs:

- **Abilities:** recommends fixes, ask questions.
- **Prior knowledge:** how switches and lights work, how malfunctions manifest themselves, what information tests provide, the side effects of repairs.
- **Past experience:** the effects of repairs or treatments, the prevalence of faults or diseases.
- **Goals:** fixing the device and tradeoffs between fixing or replacing different components.
- **Observations:** symptoms of a device or patient.

Subtasks for the diagnostic assistant

- Derive the effects of faults and interventions.
- Search through the space of possible fault complexes.
- Explain its reasoning to the human who is using it.
- Derive possible causes for symptoms; rule out other causes.
- Plan courses of tests and treatments to address the problems.
- Reason about the uncertainties/ambiguities given symptoms.
- Trade off alternate courses of action.
- Learn what symptoms are associated with faults, the effects of treatments, and the accuracy of tests.

Trading Agent

Trading agent interacts with an information environment to purchase goods and services.

- It acquires a users needs, desires and preferences.
- It finds what is available.
- It purchases goods and services that fit together to fulfill your preferences.
- It is difficult because users preferences and what is available can change dynamically, and some items may be useless without other items.

Trading Agent Inputs

- **Abilities:** acquire information, make recommendations, purchase items.
- **Prior knowledge:** the ontology of what things are available, where to purchase items, how to decompose a complex item.
- **Past experience:** how long special last, how long items take to sell out, who has good deals, what your competitors do.
- **Goals:** what the person wants, their tradeoff.
- **Observations:** what items are available, prices, number in stock,

Intelligent Tutoring System

- **Abilities:** Present information, give tests
- **Prior knowledge:** subject material, primitive strategies
- **Past experience:** common errors, effects of teaching strategies
- **Goals:** the students should master subject material, gain social skills, study skills, inquisitiveness, interest
- **Observations:** test results, facial expressions, questions, what the student is concentrating on

Common Tasks of the Domains

- **Modeling the environment** Build models of the physical environment, patient, or information environment.

- **Evidential reasoning or perception** Given observations, determine what the world is like.

- **Action** Given a model of the world and a goal, determine what should be done.

- **Learning from past experiences** Learn about the specific case and the population of cases.

Representations
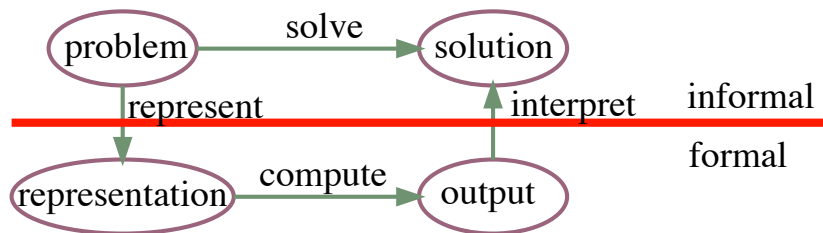
What do we want in a representation?

We want a representation to be

- rich enough to express the knowledge needed to solve the problem;
- as close to the problem as possible: compact, natural and maintainable;
- amenable to efficient computation
  - ▶ able to express features of the problem that can be exploited for computational gain
  - ▶ able to trade off accuracy and computation time and/or space
- able to be acquired from people, data and past experiences.

Defining a Solution

- Given an informal description of a problem, what is a solution?

- Typically much is left unspecified, but the unspecified parts can't be filled in arbitrarily.

- Much work in AI is motivated by common-sense reasoning. The computer needs to make common-sense conclusions about the unstated assumptions.

Quality of Solutions

- Does it matter if the answer is wrong or answers are missing?

Classes of solution:

- An optimal solution is a best solution according some measure of solution quality.

- A satisficing solution is one that is good enough, according to some description of which solutions are adequate.

- An approximately optimal solution is one whose measure of quality is close to the best theoretically possible.

- A probable solution one that is likely to be a solution.

Decisions and Outcomes

- Good decisions can have bad outcomes. Bad decisions can have good outcomes.

- Information can be valuable because it leads to better decisions: value of information.

- We can often trade off computation time and solution quality.
  An anytime algorithm can provide a solution at any time; given more time it can produce better solutions.

An agent isn't just concerned about finding the right answer, but about acquiring the appropriate information, and computing it in a timely manner.

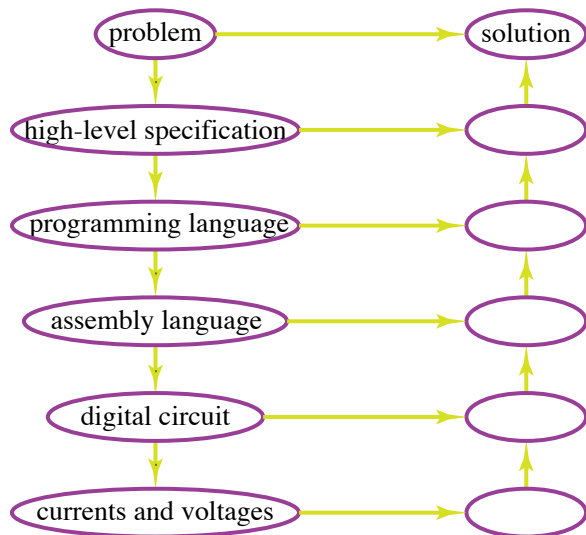# Solution quality and computation time
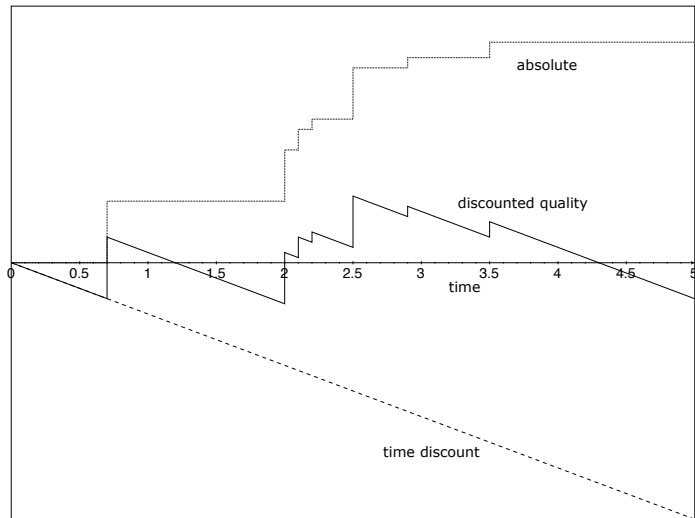
Choosing a Representation Language

We need to represent a problem to solve it on a computer.

$$\left[\begin{array}{l} \text{problem} \\ \quad \rightarrow \text{ specification of problem} \\ \qquad \rightarrow \text{ appropriate computation} \end{array}\right]$$

Example representation languages: Machine Language, C++, Java, Prolog, English

Hierarchy of representations

Physical symbol system hypothesis

- A **symbol** is a meaningful physical pattern that can be manipulated.
- A **symbol system** creates, copies, modifies and destroys symbols.

Physical symbol system hypothesis:

- A physical symbol system has the necessary and sufficient means for general intelligent action.

Knowledge & Symbol Levels

Two levels of abstraction seem to be common among biological and computational entities:

- The <mark>knowledge level</mark> is in terms of what an agent knows and what its goals are.
- The <mark>symbol level</mark> is a level of description of an agent in terms of what reasoning it is doing.

The knowledge level is about the external world to the agent. The symbol level is about what symbols an agent uses to implement the knowledge level.

Mapping from Problem to Representation

What level of abstraction of the problem to represent?
What individuals and relations in the world to represent?
How can an agent represent the knowledge to ensure that
the representation is natural, modular, and maintainable?
How can an agent acquire the information from data,
sensing, experience, or other agents?

Choosing a level of abstraction

- A high-level description is easier for a human to specify and understand.
- A low-level description can be more accurate and more predictive. High-level descriptions abstract away details that may be important for actually solving the problem.
- The lower the level, the more difficult it is to reason with.
- You may not know the information needed for a low-level description.

It is sometimes possible to use multiple levels of abstraction.

Reasoning and acting

Reasoning is the computation required to determine what an agent should do.

- Design time reasoning and computation is carried out by the designer the agent.

- Offline computation is the computation done by the agent before it has to act.
  Background knowledge and data ⇝ knowledge base .

- Online computation is the computation that's done by an agent between receiving information and acting.