Rational Agents

16 October 2020 13:46

Inputs

in the context of a self-driving car

- Abilities – e.g. steering, braking
- Goals – safety, timeliness
- Prior knowledge – what signs mean
- Stimuli – vision, lasers
- Past experience – effects of steering, friction of surfaces, how people move

Rational Action

- Goals can be defined as a performance measure, defining a numerical value for a given environment history
- Rational action is whichever action maximises the expected value of the performance measure given the percept sequence to date - I.e. doing the right thing
- Previous perceptions are typically important

Rational Agents are not omniscient – they don't know the actual outcome of their actions. They just do the best they can, given the current percepts

Actions that were expected to give a good return but failed to, can still be considered rational

Dimensions of Complexity

We can view the design space for AI as being defined by a set of dimensions of complexity. These dimensions define a **design space** for AI; different points in this space are obtained by varying the values on each dimension.

1. Modularity

Flat – one level of abstraction – adequate for simple systems. Continuous or discrete.
Modular – interacting modules that can be understood separately
Hierarchical – agent has modules that are recursively decomposed into modules – complex computers, biological systems etc

2. Planning horizon

Statis – world does not change
Finite – agent reasons about a fixed number of steps into the future
Indefinite – the agent thinks about a finite number of steps but we do not predetermine the number of steps
Infinite – the agent has to keep planning forever – process oriented

3. Representation

- Modern AI is about finding compact representations and exploiting the compactness for computational gain
- Explicitly – e.g. a chess board, one way the world could be
- Features of propositions – states can be described using features 30 binary features can represent 2^30 possible 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 man individuals
- When describing a complex world, the features can depend on **relations** and **individuals**. What

we call an *individual* could also be called a **thing**, an **object** or an **entity**. A relation on a single individual is a **property**. There is a feature for each possible relationship among the individuals.

E.g. With a light switch s_2
Instead of the feature, it could use the relation $position(s2, up)$. This relation enables the agent to reason about all switches or for an agent to have general knowledge about switches that can be used when the agent encounters a switch.

By reasoning in terms of relations and individuals, an agent can reason about whole classes of individuals without ever enumerating the features or propositions, let alone the states. An agent may have to reason about infinite sets of individuals, such as the set of all numbers or the set of all sentences. To reason about an unbounded or infinite number of individuals, an agent cannot reason in terms of states or features; it must reason at the relational level.

4. Computational Limits

Perfect rationality: the agent can determine the best course of action, without taking into account its limited computational resources
Bounded rationality – we have to make good decisions based on limited resources e.g. memory

To take into account bounded rationality, an agent must decide whether it should act or reason for longer. This is challenging because an agent typically does not know how much better off it would be if it only spent a little bit more time reasoning. Moreover, the time spent thinking about whether it should reason may detract from actually reasoning about the domain.

5. Learning from experience

The model is specified a priori

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

Usually some mix of prior knowledge is used – nature vs nurture

6. Uncertainty

Sensing and effect

In some cases, an agent can observe the state of the world directly. For example, in some board games or on a factory floor, an agent may know exactly the state of the world. In many other cases, it may only have some noisy perception of the state and the best it can do is to have a probability distribution over the set of possible states based on what it perceives. For example, given a patient's symptoms, a medical doctor may not actually know which disease a patient has and may have only a probability distribution over the diseases the patient may have.

The **sensing uncertainty dimension** concerns whether the agent can determine the state from the stimuli

In each dimension an agent can have

- No uncertainty – e.g. you will run out of power
- Disjunctive uncertainty – there is a set of states that are possible e.g. charge for 30 minutes, or you will run out of power
- Probabilistic uncertainty - Agents need to act even if they are uncertain. We need to predict what might happen in order to decide what to do - e.g. probability you will run out of power is 0.01 if you charge for 30 minutes and 0.8 otherwise

Probabilities can be learned from data and prior knowledge
Acting is gambling – if you don't use probabilities you will lose to agents that do

Sensing Uncertainty

- Fully observable – the agent can observe the state of the world
- Partially-observable – there can be a number of states possible given what the agent can perceive

Effect uncertainty

If an agent knew the initial state and its action, could it predict the resulting state?
Deterministic: the resulting state is determined from the action and the state
Stochastic – there is only a probability distribution over the resulting states.

7. Preference

What is the agent trying to achieve?

- Achieve goal is a goal – this can be a complex logical formula
- Complex preference – maybe involve trade-offs between desiderate, perhaps at different times
- Ordinal – the order is the only thing that matters
- Cardinal – counts matter e.g. we want exactly 0 crashes

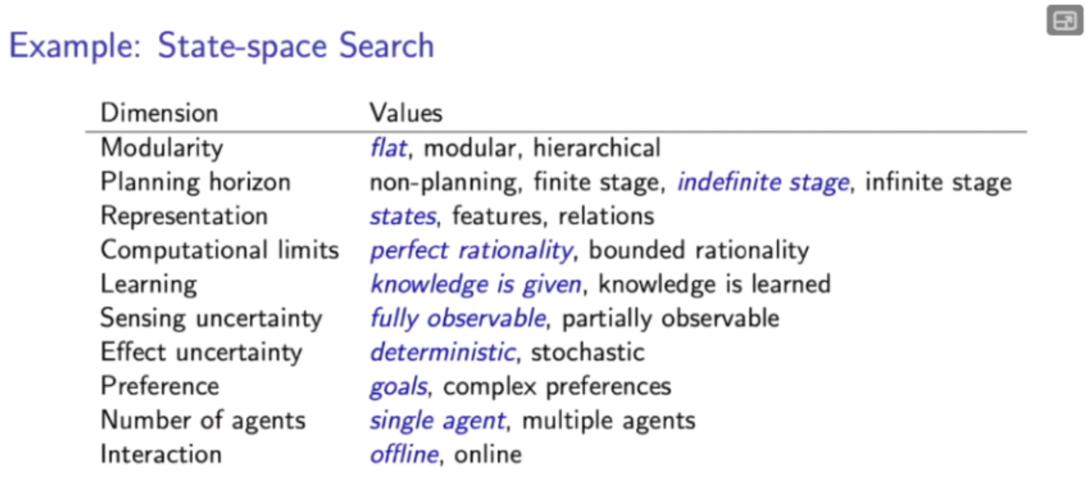
8. Number of agents

Are there multiple agents?
Single agent – any other agents are part of the environment
Multiple agent reasoning – an agent reasons strategically about the reasoning of other agents

9. Interaction

When does the agent reason to determine what to do?

- Online – while interacting
- Offline – before acting



Example: State-space Search

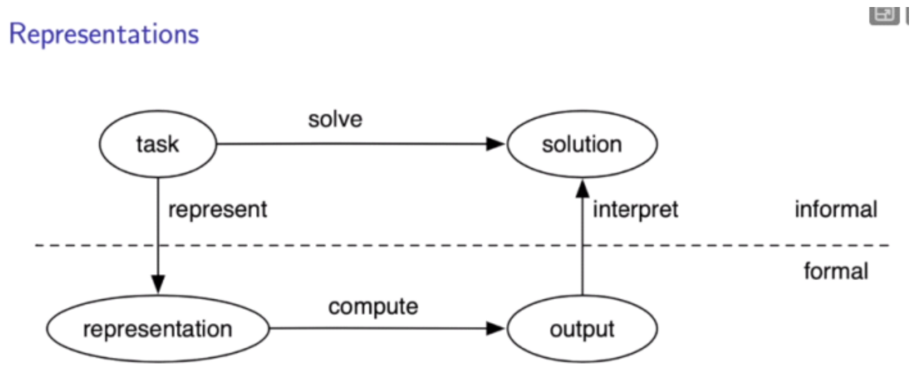
Dimension	Values
Modularity	*flat*, modular, hierarchical
Planning horizon	non-planning, finite stage, *indefinite stage*, infinite stage
Representation	*states*, features, relations
Computational limits	*perfect rationality*, bounded rationality
Learning	*knowledge is given*, knowledge is learned
Sensing uncertainty	*fully observable*, partially observable
Effect uncertainty	*deterministic*, stochastic
Preference	*goals*, complex preferences
Number of agents	*single agent*, multiple agents
Interaction	*offline*, online

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
Three values of dimensions promise to make reasoning simpler for the agent

- Hierarchical reasoning
- Individuals and relations
- Bounded rationality

Desirable Properties for a Representation


Representations

- determine what constitutes a solution
- represent the task in a way a computer can reason about
- use the computer to compute an output, which is answers presented to a user or actions to be carried out in the environment, and
- interpret the output as a solution to the task.

We have a task, that is represented in the AI system in some way.
The computation is done on this representation of the task, and then the output is relayed back out into the world.

A **representation** of some piece of knowledge is the particular data structures used to encode the knowledge so it can be reasoned with. A **knowledge base** is the representation of all of the knowledge that is stored by an agent.

We want our representation to have a few characteristics:

- Rich in data to express the knowledge needed to solve the problem
- As close to the problem as possible: compact, natural and maintainable
- Amenable to efficient computation – expresses features of the problem that can be exploited for computational gain - Able to trade off accuracy and computation time/space
- Able to be acquired from people, data, and past experiences.

Defining a Solution

- Given an informal description of a solution, 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 unstated assumptions

Quality of Solutions

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

- An optimal solution is a best solution according to some measure of solution quality
- A satisficing solution is one that is good enough according to some description of solutions that are adequate
- An approximately optimal solution is one whose measure of quality is close to the best theoretically possible. E.g. get within 10% of the optimal solution. This is sometimes still just as hard as getting the optimal solution, though.
- A probably solution is one that is likely to be a solution

Decision and Outcomes

- Good and bad decisions can have good and bad 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 in any time, but makes better decisions with more time

Agents are concerned not just about finding the right answer, but finding the information that will allow them to find the right answer

Choosing a representation
We need to represent a problem to solve it on a computer

Physical symbol system hypothesis
A symbol is a physical pattern that can be manipulated
A symbol system allows you to create/modify/delete symbols

The hypothesis states that 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 entities – biological and computational

- Knowledge level is about the external world – what the agent knows and what its goals are
- Symbol level is about what symbols the agent uses to implement the knowledge level – it is a level of description of an agent in terms of what reasoning it is doing

Mapping from Problem to Representation

- What level of abstraction of a problem to represent?
- What individuals and relations in the world to represent?
- How can an agent represent 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

- High-level is easier to understand for a human
- a low-level description can be more accurate and more predictive, we lose details when we abstract away details in high level abstractions
- You may not know the information needed for a low-level description

A delivery robot can model the environment at a high level of abstraction in terms of rooms, corridors, doors, and obstacles, ignoring distances, its size, the steering angles needed, the slippage of the wheels, the weight of parcels, the details of obstacles, the political situation in Canada, and virtually everything else. The robot could model the environment at lower levels of abstraction by taking some of these details into account. Some of these details may be irrelevant for the successful implementation of the robot, but some may be crucial for the robot to succeed. For example, in some situations the size of the robot and the steering angles may be crucial for not getting stuck around a particular corner. In other situations, if the robot stays close to the centre of the corridor, it may not need to model its width or the steering angles.

Although no level of description is more important than any other, we conjecture that you do not have to emulate every level of a human to build an AI agent but rather you can emulate the higher levels and build them on the foundation of modern computers. This conjecture is part of what AI studies.

It is sometimes possible to use multiple levels of abstraction

Reasoning and Acting

Reasoning determines what action an agent should do

1. Design time reasoning – done by the designer of the agent
2. Offline computation – done by the agent before it has to act

3. Online computation – computation done by an agent receiving information and acting