Foundations of Artificial Intelligence

3. Introduction: Rational Agents

Malte Helmert

University of Basel

March 3, 2021

3.1 Agents

3.2 Rationality

3.3 Summary

Introduction: Overview

Chapter overview: introduction
- ▶ 1. What is Artificial Intelligence?
- ▶ 2. AI Past and Present
- ▶ 3. Rational Agents
- ▶ 4. Environments and Problem Solving Methods

3.1 Agents

Heterogeneous Application Areas

AI systems are used for very different tasks:

- ▶ controlling manufacturing plants
- ▶ detecting spam emails
- ▶ intra-logistic systems in warehouses
- ▶ giving shopping advice on the Internet
- ▶ playing board games
- ▶ finding faults in logic circuits
- ▶ ...

How do we capture this diversity in a systematic framework emphasizing commonalities and differences?

common metaphor: rational agents and their environments

German: rationale Agenten, Umgebungen

Agents
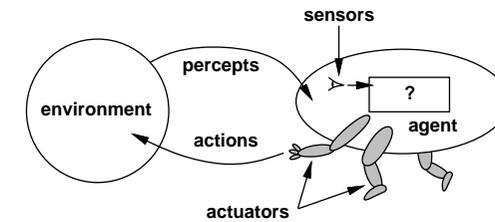
Agents
- ▶ agent functions map sequences of observations to actions:
$$f : \mathcal{P}^+ \to \mathcal{A}$$
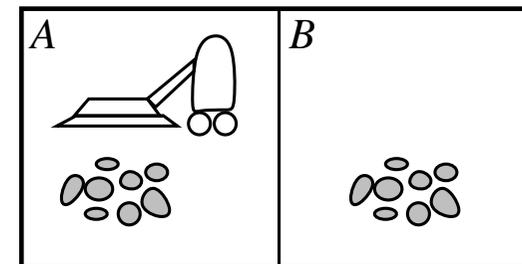- ▶ agent program: runs on physical architecture and computes f

---

Examples: human, robot, web crawler, thermostat, OS scheduler

German: Agenten, Agentenfunktion, Wahrnehmung, Aktion

Introducing: an Agent

Vacuum Domain

- ▶ observations: location and cleanness of current room:
  $\langle a, clean \rangle$, $\langle a, dirty \rangle$, $\langle b, clean \rangle$, $\langle b, dirty \rangle$
- ▶ actions: left, right, suck, wait

Vacuum Agent

a possible agent function:

observation sequence	action
⟨a, clean⟩	right
⟨a, dirty⟩	suck
⟨b, clean⟩	left
⟨b, dirty⟩	suck
⟨a, clean⟩, ⟨b, clean⟩	left
⟨a, clean⟩, ⟨b, dirty⟩	suck
. . .	. . .

Reflexive Agents

Reflexive agents compute next action only based on last observation in sequence:

- ▶ very simple model
- ▶ very restricted
- ▶ corresponds to Mealy automaton (a kind of DFA) with only 1 state
- ▶ practical examples?

German: reflexiver Agent

Example (A Reflexive Vacuum Agent)

def reflex-vacuum-agent(*location*, *status*):
    if *status* = dirty: **return** *suck*
    else if *location* = a: **return** *right*
    else if *location* = b: **return** *left*

Evaluating Agent Functions

What is the right agent function?

3.2 Rationality

Rationality

> **Rational Behavior**
> Evaluate behavior of agents with performance measure
> (related terms: utility, cost).
>
> perfect rationality:
> - always select an action maximizing
> - expected value of future performance
> - given available information (observations so far)
>
> German: Performance-Mass, Nutzen, Kosten, perfekte Rationalität

Is Our Agent Perfectly Rational?

Question: Is the reflexive vacuum agent
of the example perfectly rational?

depends on performance measure and environment!
- Do actions reliably have the desired effect?
- Do we know the initial situation?
- Can new dirt be produced while the agent is acting?

Rational Vacuum Agent

> **Example (Vacuum Agent)**
> performance measure:
> - +100 units for each cleaned cell
> - −10 units for each *suck* action
> - −1 units for each *left*/*right* action
>
> environment:
> - actions and observations reliable
> - world only changes through actions of the agent
> - all initial situations equally probable
>
> How should a perfect agent behave?

Rationality: Discussion

- perfect rationality ≠ omniscience
  - incomplete information (due to limited observations) reduces achievable utility
- perfect rationality ≠ perfect prediction of future
  - uncertain behavior of environment (e.g., stochastic action effects) reduces achievable utility
- perfect rationality is rarely achievable
  - limited computational power ⇝ bounded rationality

German: begrenzte Rationalität

3.3 Summary

Summary (1)

common metaphor for AI systems: rational agents

agent interacts with environment:

▶ sensors perceive observations about state of the environment

▶ actuators perform actions modifying the environment

▶ formally: agent function maps observation sequences to actions

▶ reflexive agent: agent function only based on last observation

Summary (2)

rational agents:

▶ try to maximize performance measure (utility)

▶ perfect rationality: achieve maximal utility in expectation given available information

▶ for "interesting" problems rarely achievable
  ⤳ bounded rationality