Foundations of Artificial Intelligence

3. Introduction: Rational Agents

Malte Helmert

University of Basel

March 5, 2018

Introduction: Overview

Chapter overview: introduction

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

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?

Heterogeneous Application Areas

AI systems are used for very different tasks:

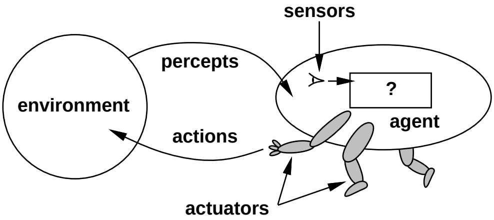
- controlling manufacturing plants
- detecting spam emails
- intra-logistic systems in warehouses
- giving shopping advice on the Internet
- playing board games
- finding faults in logic circuits
- ...

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

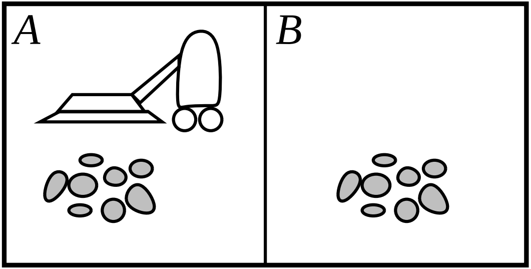
- agent functions map sequences of observations to actions:
$$f : \mathcal{P}^+ \to \mathcal{A}$$
- agent program: runs on physical architecture and computes f

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: ⟨a, clean⟩, ⟨a, dirty⟩, ⟨b, clean⟩, ⟨b, dirty⟩
- actions: left, right, suck, wait

Vacuum Agent

a possible agent function:

observation sequence	action
$\langle \mathsf{a}, \mathsf{clean} \rangle$	*right*
$\langle \mathsf{a}, \mathsf{dirty} \rangle$	*suck*
$\langle \mathsf{b}, \mathsf{clean} \rangle$	*left*
$\langle \mathsf{b}, \mathsf{dirty} \rangle$	*suck*
$\langle \mathsf{a}, \mathsf{clean} \rangle,\ \langle \mathsf{b}, \mathsf{clean} \rangle$	*left*
$\langle \mathsf{a}, \mathsf{clean} \rangle,\ \langle \mathsf{b}, \mathsf{dirty} \rangle$	*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?

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?

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

German: begrenzte Rationalität

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
  $\rightsquigarrow$ bounded rationality