Aalto University
School of Electrical Engineering

ELEC-E8125 Reinforcement learning
Overview

Ville Kyrki

17.9.2019

Today

- Introduction to planning in sequential problems
- Overview of course contents

Let’s talk about planning

- Name planning problems from your daily life

Let’s talk about planning

- Name planning problems from your daily life

- Design a plan to solve your problem

Let’s talk about planning

- Name planning problems from your daily life
- Design a plan to solve your problem
- What is a plan?

Planning and surprises

- Does your plan allow for surprises or unknowns?
  - Raise of hands

Planning and surprises

- Does your plan allow for surprises or unknowns?
  - Raise of hands

- Discuss in groups: How would you modify the plans to allow surprises?

Planning and surprises

- Does your plan allow for surprises or unknowns?
  - Raise of hands

- Discuss in groups (10 min): How would you modify the plans to allow surprises?

- Plan can be conditional on current observation

  Policy from observation to action

Information needs

- In groups: Are there cases when current observation is not sufficient to make decisions? If yes, when does that happen?

Information needs

- In groups: Are there cases when current observation is not sufficient to make decisions? If yes, when does that happen?

- Sometimes history of observations is needed.
- Information used for decision can be abstracted as *state.*

- Discussion: Give examples of state for different problems.

Plan as policy

- Let’s consider that everything can be observed at time of each decision.
- Plan is then a policy function from state to action.

Plan as policy

- Let’s consider that everything can be observed at time of each decision.
- Plan is then a policy function from state to action.

- In groups: Can all plans (purposeful decision strategies) be represented like this?

Plan as policy

- Let’s consider that everything can be observed at time of each decision.
- Plan is then a policy function from state to action.

- In groups: Can all plans (purposeful decision strategies) be represented like this?
  - Many can, but sometimes it’s useful to be random (e.g. games)

Success

- How can you define success in planning?

Success

- How can you define success in planning?

- Reaching a particular state
- Making particular state transitions

Success

- How can you define success in planning?

- Reaching a particular state
- Making particular state transitions

- Are all plans that reach a goal equally good?

Success

- How can you define success in planning?

- Reaching a particular state
- Making particular state transitions

- Are all plans that reach a goal equally good?
- Give an example of a good and a bad plan

Objective(s)

- How can you formulate goal(s) in planning to take into account plan quality?

Objective(s)

- How can you formulate goal(s) in planning to take into account plan quality?
- Immediate reward vs cumulative return

Objective(s)

- How can you formulate goal(s) in planning to take into account plan quality?
- Immediate reward vs cumulative return
- Design rewards for your own problem.

Evaluating policy quality

- Assuming that
  - we have a policy,
  - know the associated reward function,
  - the system can be tested,

  how can the quality of the policy be evaluated?

Planning as optimization

- Planning (sequential decision making) can be understood as *optimization of a policy with respect to expected return.*

Planning as optimization

- Planning (sequential decision making) can be understood as *optimization of a policy with respect to expected return.*

- To automatically solve such problems, which information is needed? Where can the information come from?

Information for planning

- Effects of actions in different states
  - Which state I may end up to if I do X now?
- Rewards of state-action pairs
  - What’s the reward if I now do X?

Reinforcement learning problem

- Determine policy

$$u = \pi(x)$$

such that expected cumulative return is maximized

$$\pi^* = arg\,max_\pi E[R]$$

$$R = \sum_t r_t$$

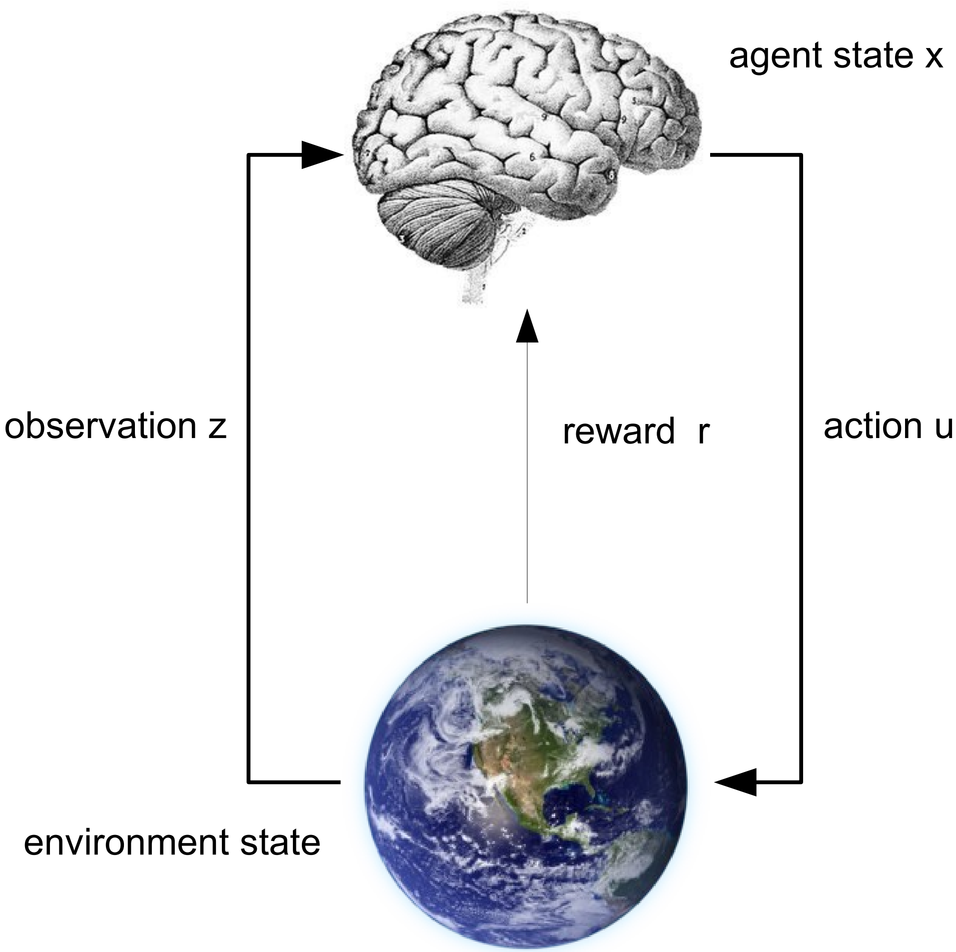
Why is RL hard?

- Effects of actions (state dynamics)
  - need to be learned
  - are often stochastic
- Rewards
  - (may) need to be learned
  - may be delayed ("sparse rewards")
  - may be difficult to choose/formulate
- Trade-off between learning (*exploration*) and maximizing rewards (*exploitation*)

Summary so far

- Can you
  - explain what is reinforcement learning
  - define a problem as a reinforcement learning problem
  - explain why reinforcement learning is difficult

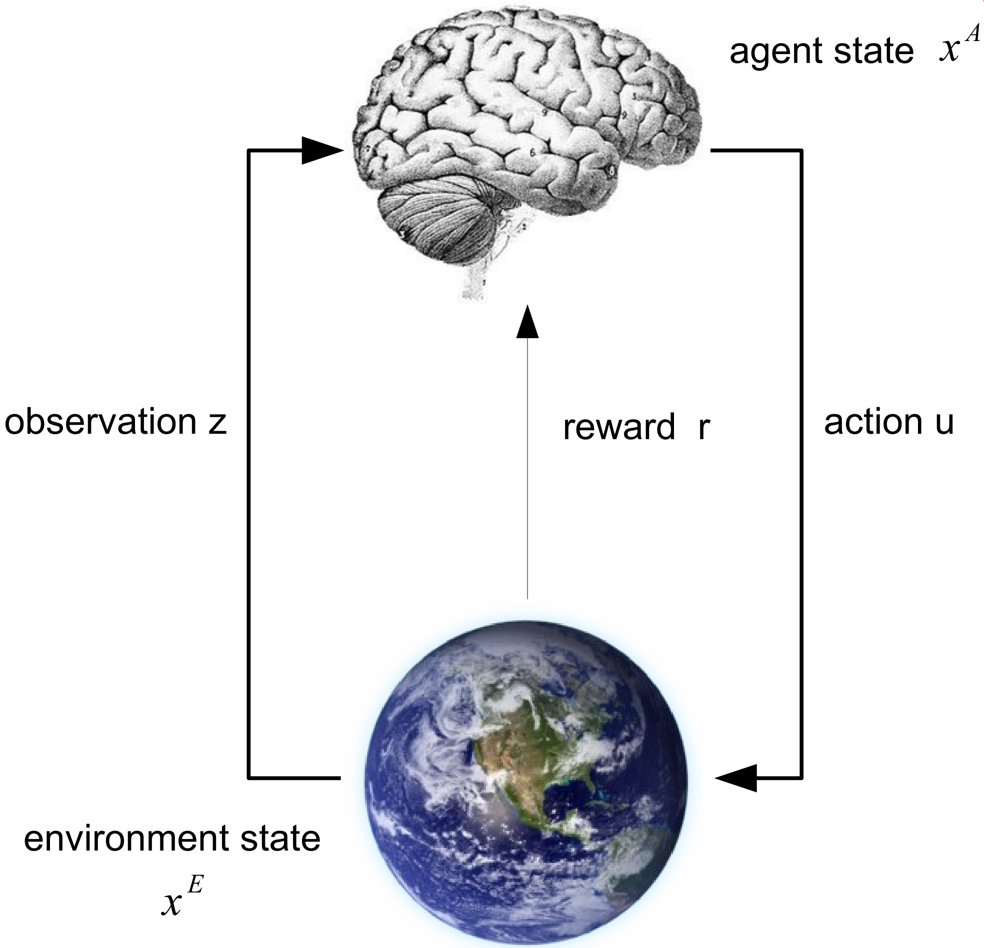
Setting



Task
Choose a sequence of actions that maximizes cumulative reward.

Can you explain what does Markovianity mean?

Markov decision process



MDP

Environment observable

$o = x^E = x^A$

Defined by dynamics

$P(x_{t+1}|x_t, u_t)$

And reward function

$r_t = r(x_{t+1}, x_t)$

Solution e.g.

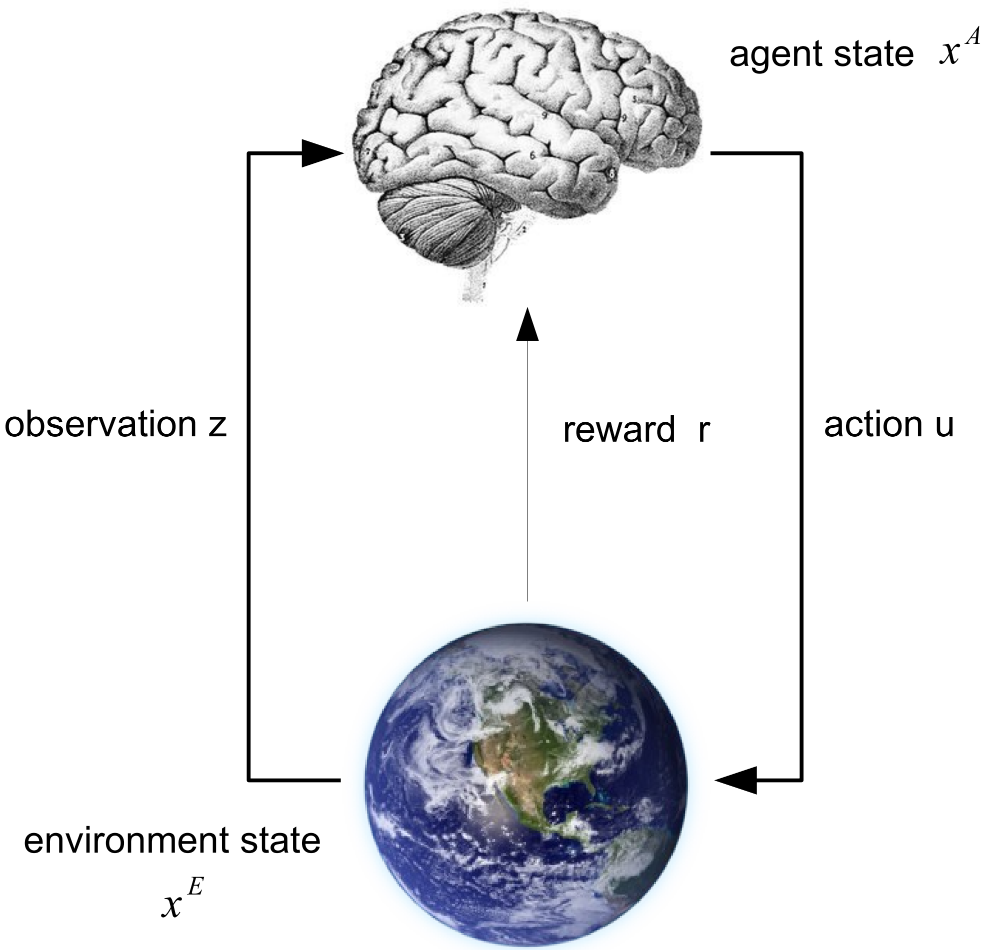
$u^*_{1,\dots,T} = max_{u_1,\dots,u_T} \sum_{t=1}^{T} r_t$

Represented as policy

$u = \pi(x^A)$

Can you see parallels with control?

Reinforcement learning



RL
MDP with **<u>unknown</u>** Markovian dynamics
$P(x_{t+1} | x_t, u_t)$

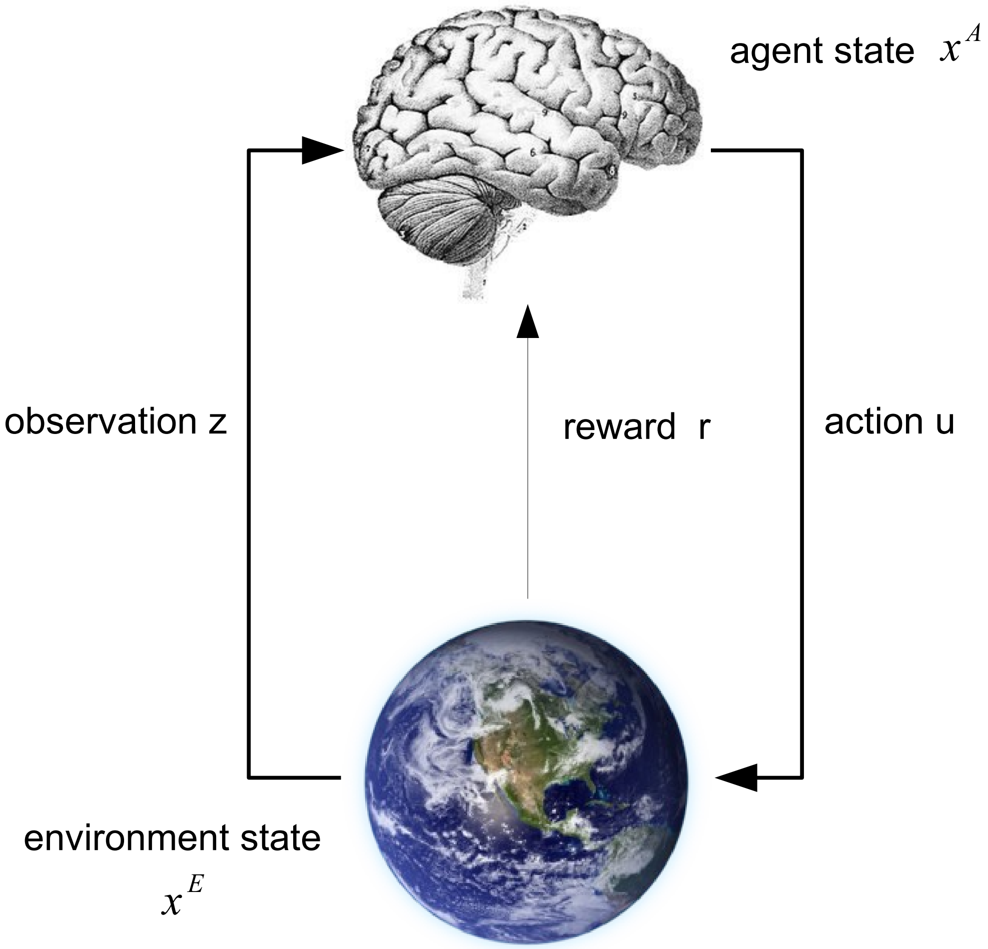
Unknown reward function
$r_t = r(x_{t+1}, x_t)$

Solution similar, e.g.
$u^*_{1,\ldots,T} = max_{u_1,\ldots,u_T} \sum_{t=1}^{T} r_t$

Learning must **<u>explore</u>** policies

How does this then relate to the control case?

Partially observable MDP (POMDP)



POMDP
Environment not directly observable

Defined by dynamics
$P(x^E_{t+1}|x^E_t, u_t)$

Reward function
$r_t = r(x_{t+1}, x_t)$

Observation model
$P(z_t|x^E_t, u_t)$

Solution similar, eg.
$u^*_{1,\dots,T} = max_{u_1,\dots,u_T} E\left[\sum_{t=1}^{T} r_t\right]$

Aalto University
School of Electrical
Engineering

How does this relate to the control case?

Course outline

- Optimal decision making with known dynamics
- Markov decision processes
- Reinforcement learning
- Partially observable Markov decision processes

Next time: Discrete planning in deterministic worlds

- Read LaValle, “Planning Algorithms”, Sections 2–2.2.2, 2.3–2.3.2 (~20 pages)
- Read Platt, “Introduction to linear quadratic regulation”, Sec. 1-3 (~5 pages)
- Complete Quiz 1