IDALAB

INTELLIGENT DATA ANALYTICS SALZBURG

PARIS
LODRON
UNIVERSITÄT
SALZBURG

Simon Hirländer

Tutorial RL4AA

World state

True dynamics

Problem design - capture the right thing

- Rarely the observation o is the state s, the world state is, but often we assume it is certainty equivalence!

- POMDP $\Rightarrow$ MDPs!

14

MDP

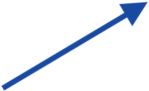
- Solve an SDM problem: Information→Decision→Information→Decision→…

- Generally stochastic!

- Consequently we build a feedback system not planing too far in the future:

- Define a **state** $s_t = h_t(o_t, a_{t-1}, o_{t-1}, a_{t-2}, o_{t-2} \ldots)$, as a function holding **sufficient statistics** until time step t for a decision - (example pong)

- Decision based on s_t via: $a_t = \pi_t(s_t)$ - the policy - optimise an expected aggregate of future rewards



o_{t-1}

o_{t-2}

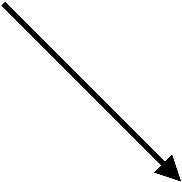
•

o_t

Internal representation

$\hat{=}$

$\bullet \, S_t$

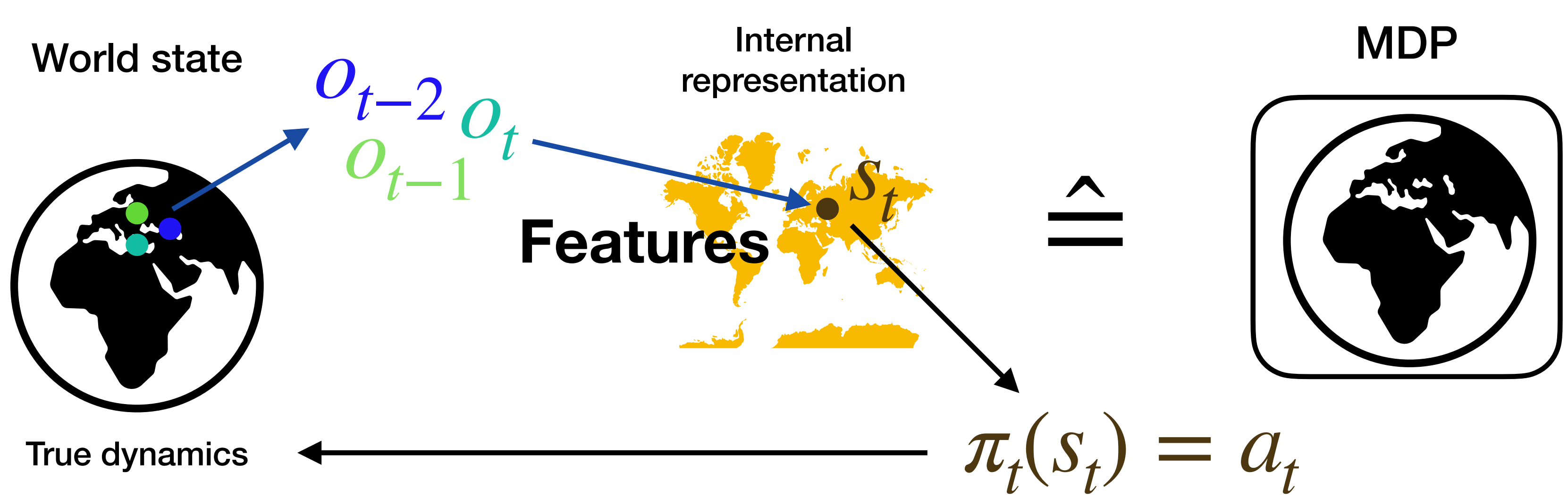


$$\pi_t(s_t) = a_t$$

Features

Problem design - capture the right thing

- Solve an SDM problem: Information→Decision→Information→Decision→…
- Generally stochastic!
- Consequently we build a feedback system not planing too far in the future:
  - Define a **state** $s_t = h_t(o_t, a_{t-1}, o_{t-1}, a_{t-2}, o_{t-2} \ldots)$, as a function holding **sufficient statistics** until time step t for a decision - (example pong)
  - Decision based on s_t via: $a_t = \pi_t(s_t)$ - the policy - optimise an expected aggregate of future rewards



- Rarely the observation o is the state s, the world state is, but often we assume it is certainty equivalence!
- POMDP $\Rightarrow$ MDPs!

How bad is it?

- Linear POMDP: believe state - $O_t = h_t(S_t, A_t, W_t)$
  - ➡ Static output feedback is NP hard (linear in O_t and dynamics)
  - ➡ General POMDPs are PSPACE hard
- There are ways out - separation principle:
  - ➡ Filtering $\hat{s}_t = f(\{o_t\})$ - prediction problem
  - ➡ Action based on <u>certainty equivalence</u>
  - ➡ Optimal filtering - if dynamics are linear and noise is Gaussian - Kalman filtering - general belief propagation - LQG
  - ➡ Kalman filtered state - <u>optimal in estimation and control</u>
  - ➡ Estimate state with prediction $S_t = h(\tau_t)$, τ_t are time lags