# Interactive Reinforcement Learning

Peter Wüppen, Alvin Fazrie

Seminar Intelligent Robotics

19.02.2016

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

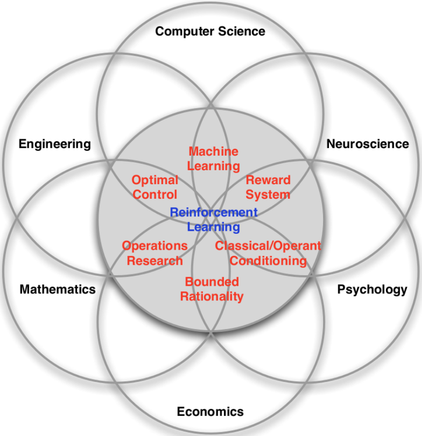
informatik die zukunft

# Contents

# Motivation

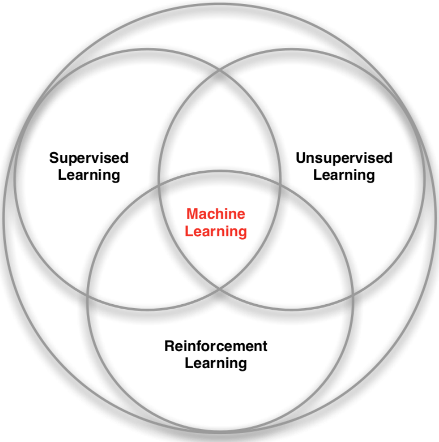
- Artificial Intelligence has been a vital part in daily life
- One of the fundamental topics in the field of robotic and machine learning is Reinforcement Learning
- To introduce the basic concepts of RL
- Make use of interactive components to improve the learning process

# Reinforcement Learning Basics




Many faces of Reinforcement Learning [Sil15]

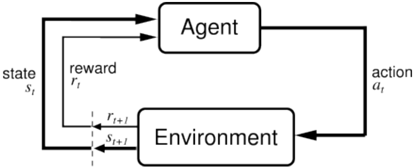
# Reinforcement Learning Basics




Branches of Reinforcement Learning [Sil15]

# Reinforcement Learning Basics

1. Observe state, $s_t$
2. Decide on an action, $a_t$
3. Perform action
4. Get the reward
5. Observe new state, $s_{t+1}$
6. Update policy based on the given reward
7. Repeat




Reinforcement Learning Diagram [SB98]

# Q Value function

The total reward estimation in the current state (action-value pairs):

$$Q^{\pi}(s,a) = E_{\pi}\{\sum_{k=0}^{\infty} \gamma^{k} r_{t+k+1} | s_t = s, a_t = a\} \quad (1)$$

Bellman Equation:

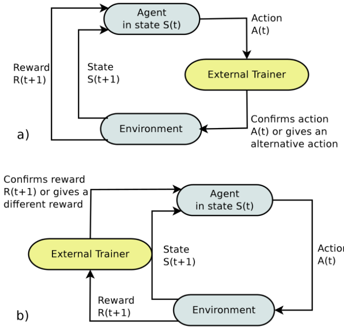
$$Q^{*}(s,a) = \sum_{s'} P(s'|s,a)[r(s,a,s') + \gamma \max_{a'} Q(s',a')] \quad (2)$$

# Interactive Reinforcement Learning

- Addition to regular RL
- Learning agent receives help from a teacher
- goal: faster learning, better final policy
- variety of different approaches

# Interactive Reinforcement Learning

Reinforcement Learning with Interactive Feedback.




Approaches to interaction between a robotic agent and an external trainer [CTM$^+$15]

# Agent Advising

- Analysis on the domain of teaching video games by Taylor et al. in [TCF$^+$14]
- Goal: finding appropriate advice strategies for agents to use when teaching humans
- Specific attention to advice budgets
- Introduction of multiple strategies for advice distribution

# Agent Advising: Budgeting strategies

- Early Advising
- Importance Advising
- Mistake Correcting
- Predictive Advising

# Early Advising

Lets the teaching agent spend its budget as soon as possible

**Algorithm 1** Early Advising

**procedure** EarlyAdvising $(\pi, n)$.
    **for** each student state $s$ **do**
        **if** $n > 0$ **then**
            $n \leftarrow n - 1$
            Advice $\pi(s)$
        **end if**
    **end for**
**end procedure**

# Importance Advising

Restrict advice to states perceived as important

**Algorithm 2** Importance Advising

**procedure** ImportanceAdvising $(\pi, n, t)$.<br>
    **for** each student state $s$ **do**<br>
        **if** $n > 0$ and $I(s) \geq t$ **then**<br>
            $n \leftarrow n - 1$<br>
            Advice $\pi(s)$<br>
        **end if**<br>
    **end for**<br>
**end procedure**

# Mistake Correcting

Only spend advice budget if the right action is not being chosen anyways

**Algorithm 3** Mistake Correcting

**procedure** MistakeCorrecting $(\pi, n, t)$.
  **for** each student state $s$ **do**
    Observe students announced action $a$
    **if** $n > 0$ and $I(s) \geq t$ and $a \neq \pi(s)$ **then**
      $n \leftarrow n - 1$
      Advice $\pi(s)$
    **end if**
  **end for**
**end procedure**

# Predictive Advising

Learn about the students behaviour and predict important wrong decisions

**Algorithm 4** Predictive Advising

**procedure** PredictiveAdvising $(\pi, n, t)$.
    **for** each student state $s$ **do**
        Predict students intended action $a$
        **if** $n > 0$ and $I(s) \geq t$ and $a \neq \pi(s)$ **then**
            $n \leftarrow n - 1$
            Advice $\pi(s)$
        **end if**
    **end for**
**end procedure**

# Human Advising

- Analysis on human teaching behaviour and its implications by Thomaz et al. in [TB06]
- Goal: Human successfully assists robot in learning a previously unknown task
- Problems arise from different representation of the problem and lack of knowledge of the learning process by a non-expert user
- Experimental results suggest combination of guidance and reward channels for maximum success

# Table-cleaning scenario by Cruz et al. [CMWW14]

- Robot is positioned in front of a table
- Positions: *left*, *right*, *home*
- Objects: *sponge*, *cup*
- Actions: *get*, *drop*, *go*, *clean*
- Initially, robot arm and sponge are at *home*, cup at *left*
- optimal (minimal) number of actions: 15

# Table-cleaning scenario

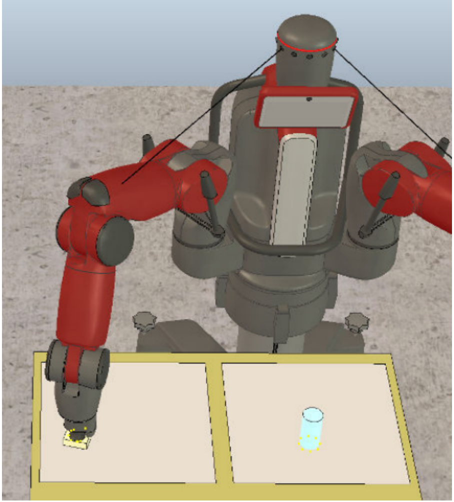


**Figure:** Simulation for the table-cleaning scenario

# Table-cleaning scenario

- 46 different, valid states
- Learning algorithm: SARSA
- Rewards: 1 for final state, -1 for failed state, -0.01 otherwise
- 30 agents trained in parallel
- Initial IRL approach: random advising

# Results: Random Advising

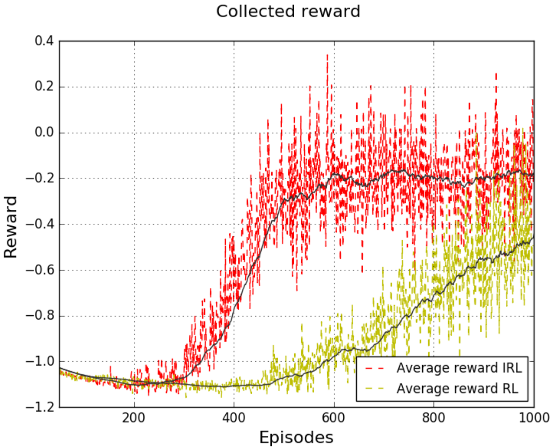



**Figure:** Acquired Reward (Random Advising)

# Advice strategy

Instead of random advising, we implemented and tested the performance of

- Early Advising
- Importance Advising
- Mistake Correction

while keeping the total amount of advice roughly the same

# Results: Early Advising

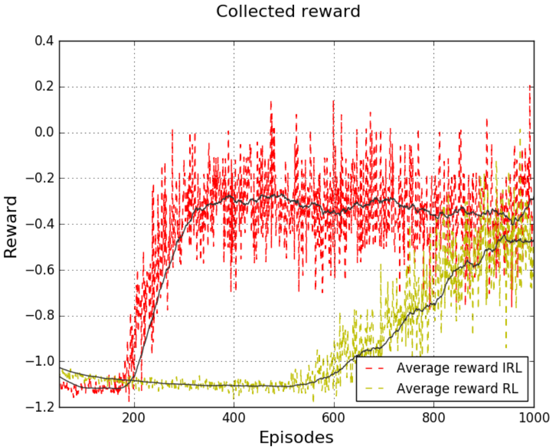



**Figure:** Acquired Reward (Early Advising)

# Results: Importance Advising

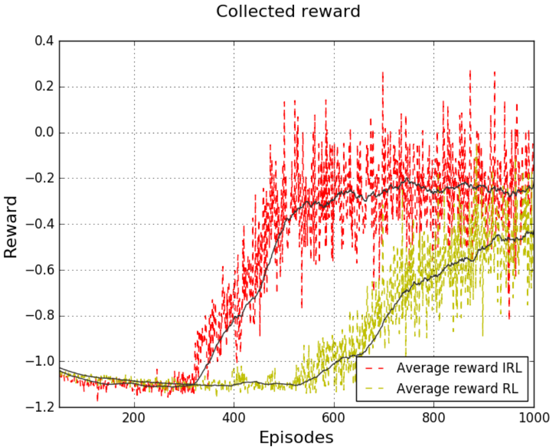



**Figure:** Acquired Reward (Importance Advising)

# Results: Mistake correction


**Figure:** Acquired Reward (Mistake correction)

# Results: Mistake correction

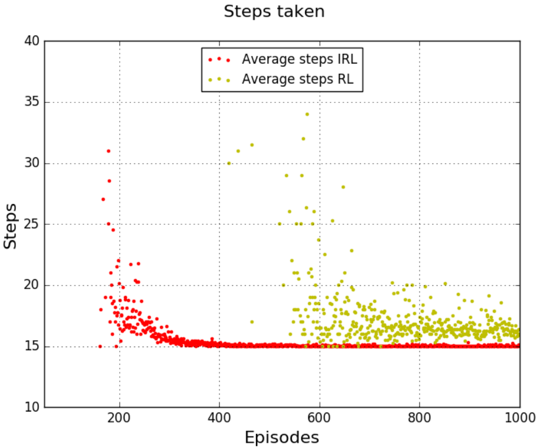



**Figure:** Actions taken (Mistake correction)

# Results: Comparison

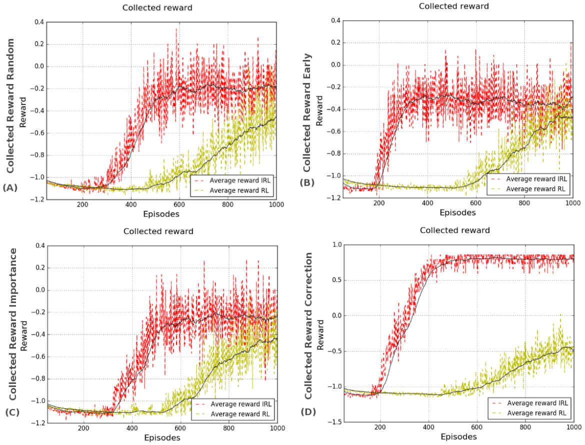



**Figure:** Comparison of the 4 tested strategies

# Conclusion

- Interactive approaches can improve the speed and quality of RL significantly
- Plenty of potential use-cases, e.g. in the field of HCI
- In case of limited advice, the choice of advice strategy is crucial for success
- Tradeoff between complexity and performance

# Bibliography I

T Baier-Lowenstein and Jianwei Zhang.
Learning to grasp everyday objects using reinforcement-learning with automatic value cut-off.
In *Intelligent Robots and Systems*, page 1551–1556. IEEE, 2007.

Francisco Cruz, Sven Magg, Cornelius Weber, and Stefan Wermter.
Improving reinforcement learning with interactive feedback and affordances.
In *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014 Joint IEEE International Conferences on*, pages 165–170. IEEE, 2014.

# Bibliography II


Francisco Cruz, Johannes Twiefel, Sven Magg, Cornelius Weber, and Stefan Wermter.
Interactive reinforcement learning through speech guidance in a domestic scenario.
In *Neural Networks (IJCNN), 2015 International Joint Conference on*, pages 1–8. IEEE, 2015.

W Bradley Knox and Peter Stone.
Reinforcement learning from human reward: Discounting in episodic tasks.
In *RO-MAN, 2012 IEEE*, pages 878–885. IEEE, 2012.

Richard S. Sutton and Andrew G. Barto.
*Introduction to Reinforcement Learning*.
MIT Press, Cambridge, MA, USA, 1st edition, 1998.

# Bibliography III


David Silver.
Introduction to reinforcement learning.
2015.

Andrea L. Thomaz and Cynthia Breazeal.
Reinforcement learning with human teachers: Evidence of feedback and guidance with implications for learning performance.
In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1*, AAAI'06, pages 1000–1005. AAAI Press, 2006.

# Bibliography IV

Matthew E. Taylor, Nicholas Carboni, Anestis Fachantidis, Ioannis Vlahavas, and Lisa Torrey.
Reinforcement learning agents providing advice in complex video games.
*Connect. Sci*, 26(1):45–63, January 2014.

Andrea Lockerd Thomaz, Guy Hoffman, and Cynthia Breazeal.

Real-time interactive reinforcement learning for robots.
In *AAAI 2005 workshop on human comprehensible machine learning*, 2005.