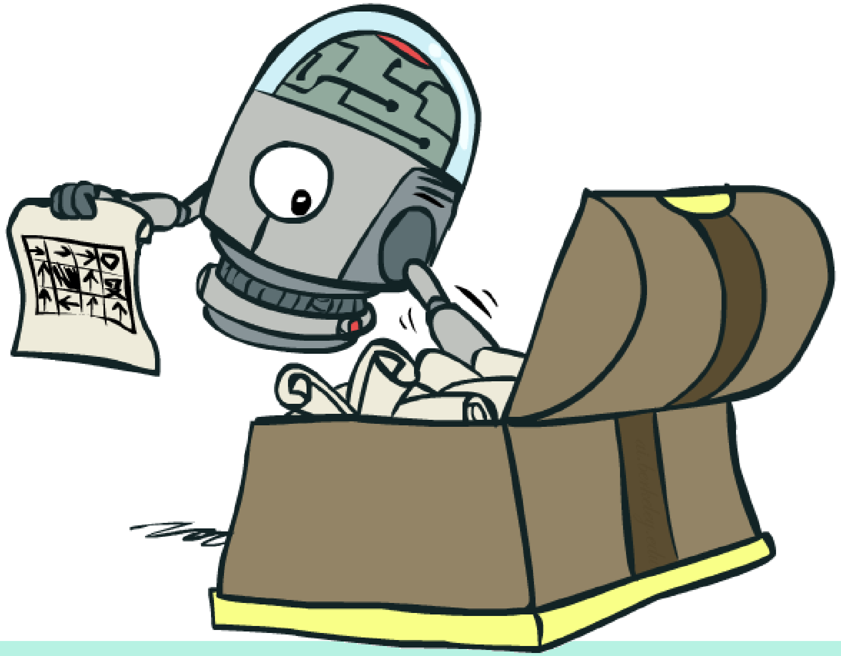
Advanced Topics in AI

Policy Search



Instructor: Prof. Dr. techn. Wolfgang Nejdl

Leibniz University Hannover

Co-financed by the Connecting Europe Facility of the European Union

[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All materials are available at http://ai.berkeley.edu.]

Policy Search

- Problem: often the feature-based policies that work well (win games, maximize utilities) aren't the ones that approximate V / Q best
  - E.g. your value functions from project 2 were probably horrible estimates of future rewards, but they still produced good decisions
  - Q-learning's priority: get Q-values close (modeling)
  - Action selection priority: get ordering of Q-values right (prediction)
  - We'll see this distinction between modeling and prediction again later in the course

- Solution: learn policies that maximize rewards, not the values that predict them
- Policy search: start with an ok solution (e.g. Q-learning) then fine-tune by hill climbing on feature weights

Co-financed by the Connecting Europe Facility of the European Union

Policy Search

- Simplest policy search:
  - Start with an initial linear value function or Q-function
  - Nudge each feature weight up and down and see if your policy is better than before

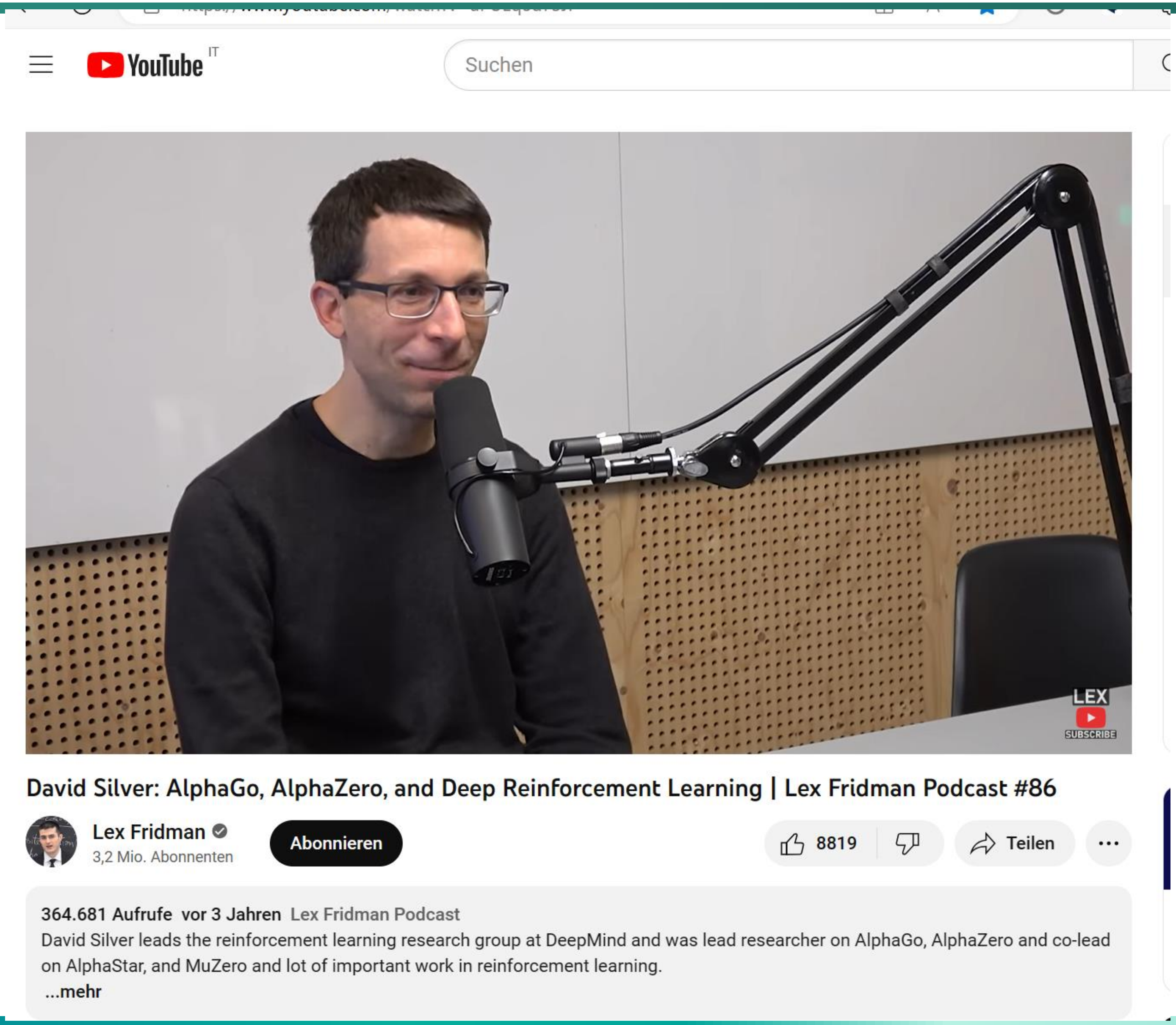
- Problems:
  - How do we tell the policy got better?
  - Need to run many sample episodes!
  - If there are a lot of features, this can be impractical

- Better methods exploit lookahead structure, sample wisely, change multiple parameters…

Co-financed by the Connecting Europe Facility of the European Union

AlphaGo, AlphaZero, and RL

[Schulman, Moritz, Levine, Jordan, Abbeel, ICLR 2016]



David Silver: AlphaGo, AlphaZero, and Deep Reinforcement Learning | Lex Fridman Podcast #86

Lex Fridman
3,2 Mio. Abonnenten

364.681 Aufrufe vor 3 Jahren Lex Fridman Podcast
David Silver leads the reinforcement learning research group at DeepMind and was lead researcher on AlphaGo, AlphaZero and co-lead on AlphaStar, and MuZero and lot of important work in reinforcement learning.
...mehr

Co-financed by the Connecting Europe Facility of the European Union

MDPs and RL

Known MDP: Offline Solution

Goal	Technique
Compute V^*, Q^*, π^*	Value / policy iteration
Evaluate a fixed policy π	Policy evaluation

Unknown MDP: Model-Based

Goal	Technique
Compute V^*, Q^*, π^*	VI/PI on approx. MDP
Evaluate a fixed policy π	PE on approx. MDP

Unknown MDP: Model-Free

Goal	Technique
Compute V^*, Q^*, π^*	Q-learning
Evaluate a fixed policy π	Value Learning

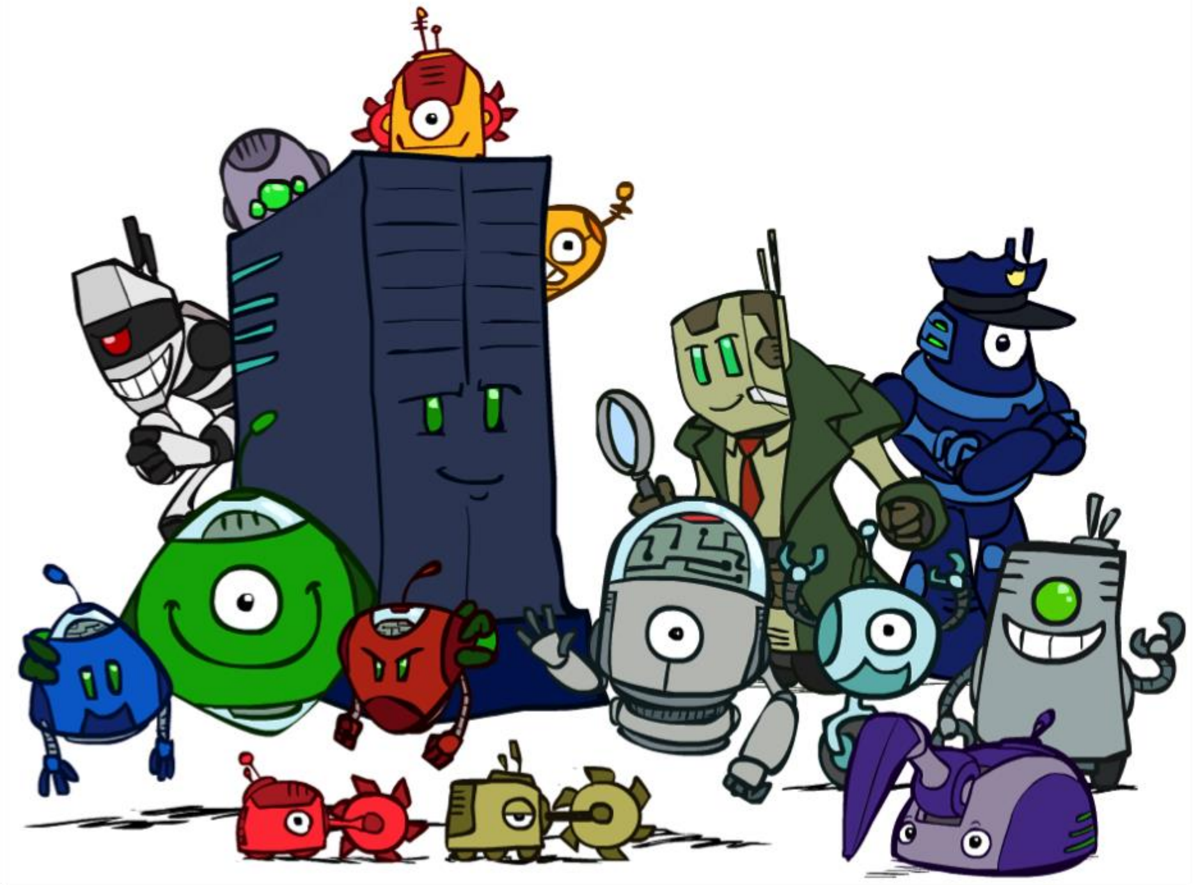
use features to generalize

Co-financed by the Connecting Europe Facility of the European Union

Conclusion

- We’re done with Search and Planning!

- We’ve seen how AI methods can solve problems in:
    - Search
    - Constraint Satisfaction Problems
    - Games
    - Markov Decision Problems
    - Reinforcement Learning



Co-financed by the Connecting Europe
Facility of the European Union