Curriculum Reinforcement Learning

Self-paced Learning

Marius Lindauer

Winter Term 2021

Self-Paced Learning

- We can base the curriculum on the agent's performance instead of on instance features
- There are different kinds of performance markers:

Self-Paced Learning

- We can base the curriculum on the agent's performance instead of on instance features
- There are different kinds of performance markers:
  - Evaluation reward
  - Goals reached
  - Confidence in policy
  - Internal value function

Self-Paced Deep Reinforcement Learning [Klink et al. 2020]

- Goal: enable agent to solve specific very hard instances (common in robotics)
- Idea: define (easy) start distribution and (hard) goal distribution of instances, then slowly shift towards the goal
- Speed of the shifting depends on the agents state evaluations $V(s)$ as a progress measure

Curriculum Generation through Value Disagreement Sampling [Zhang et al. 2020]

- Idea: train an ensemble of agents and use the differences in their Q-values as a measure of uncertainty
- Goals are proposed at a point where the ensemble starts to disagree, assuming that agreement means all agents have learned a good policy on this instance already
- Note: currently this method is an extension of HER and thus only works if the transition function stays the same between instances.