

RL: Deep

Prioritized Replay

Marius Lindauer



Leibniz
Universität
Hannover

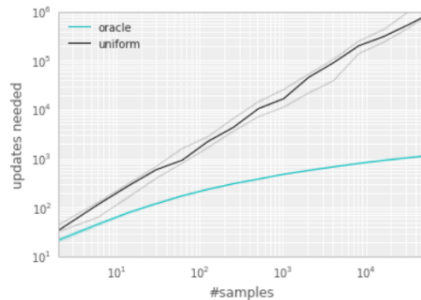
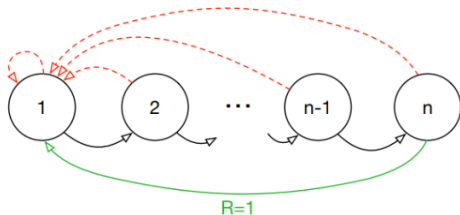


Winter Term 2021

Impact of Replay?

- ▶ In tabular TD-learning, **order** of replaying updates could help speed learning
- ▶ Repeating some updates seem to better propagate info than others
- ▶ Systematic ways to prioritize updates?

Potential Impact of Ordering Episodic Replay Updates [Schaul et al. 16]



- ▶ Oracle: picks (s, a, r, s') tuple to replay that will minimize global loss
- ▶ Exponential improvement in convergence
 - ▶ Number of updates needed to converge
- ▶ Oracle is not a practical method but illustrates impact of ordering

Prioritized Experience Replay [Schaul et al. 16]

- ▶ Let i be the index of the i -th tuple of experience (s_i, a_i, r_i, s_{i+1})
- ▶ Sample tuples for update using priority function
- ▶ Priority of a tuple i is proportional to DQN error

$$p_i = |r + \gamma \max_{a' \in A} Q(s_{i+1}, a'; \vec{w}^-) - Q(s_i, a; \vec{w})|$$

- ▶ Update p_i every update. p_i for new tuples is set to maximum value
- ▶ One method: proportional (stochastic prioritization)

$$P(i) = \frac{p_i^\beta}{\sum_k p_k^\beta}$$

- ▶ $\beta = 0$ yields random selections