Exploration in RL

Prediction-based Intrinsic Exploration[a]

Marius Lindauer

Winter Term 2021

[a] based on Blog by Lilian Weng

Prediction-based Exploration Schmidhuber. 1991

- Idea: If the agent is able to predict what will happen in the future, it is already well informed
- In contrast, if the agent is not able to predict the future, it is surprised.

$$
\begin{aligned}
f : (s_t, a_t) &\mapsto s_{t+1} \\
e(s_t, a_t) &= ||f(s_t, a_t) - s_{t+1}||_2^2
\end{aligned}
$$

  - the higher the error e, the less familiar the agent is with that state / more surprised

Intelligent Adaptive Curiosity Oudeyer et al. 2007

- Memory of all observed state transitions $M = (s_t, a_t, s_{t+1})$
- Split the state space S similarly as in decision node:
  - Split only if enough states were observed
  - Variance of states in each leaf should be minimal
  - For each leaf, learn a forward dynamic predictor f
- Reward regions where we can make fast progress via decreasing error

$$r_t^i = \frac{1}{k}\sum_{i=0}^{k-1}(e_{t-i-\tau} - e_{t-i})$$

  - moving window with offset τ and moving window size k

Decay Stadie et al. 2015

- Normalize error to [0,1] by the maximal error observed so far

$$\bar{e}_t = \frac{e_t}{\max_{i \leq t} e_i}$$

- decay intrinsic reward over time

$$r_t^i = \frac{e_t(s_t, a_t)}{t \cdot C}$$

  - C being a constant