Exploration in RL

Intrinsic Exploration[a]

Marius Lindauer

Leibniz Universität Hannover

Winter Term 2021

---

[a] based on Blog by Lilian Weng

The Hard Exploration Problem

- hard-exploration problems:
  - very sparse rewards
  - or even deceptive rewards

The Hard Exploration Problem

- hard-exploration problems:
    - very sparse rewards
    - or even deceptive rewards
- Examples
    - Montezuma's Revenge (Atari): long sequence of steps needed to figure out that "key" is needed to open "door"
    - Noisy-TV problem:
        - Assumption: Agent gets explicit reward for seeking novel experience
        - Agent discovers TV that only shows random images
        - Agent will watch TV forever (without solving the real task)!

Intrinsic Rewards as Exploration Bonus

- Augment reward by reward external reward r^e and intrinsic reward r^i

$$r_t = r_t^e + \beta r_t^i$$

- Inspired by intrinsic motivation in psychology
  - Children are driven by curiosity which helps to learn
  - Intrinsic rewards could be correlated with curiosity, surprise, familiarity of the state and more
- Two main ideas for RL
  - Discovery of novel states
  - Improvement of the agent's knowledge about the environment

Count-based Exploration

- What does it mean that the agent is surprised that it discovered something new?

$\leadsto$ Measure whether the state is novel or appeared often

- Count how many times a state was encountered and assign bonus to rarely encountered states
  - Count-based exploration
  - $N_n(s)$: number of visits of state s in the sequence $s_{1:n}$
  - Problem: Most $N(s)$ will be zero for non-trivial environments

Counting by Density Model Bellemare et al. 2016

- Use a density model to approximate the frequency of state visits
- $p_n(s) = p(s \mid s_{1:n})$ is the probability of the $(n+1)$-th state being s
  - empirically: $p_n(s) = N_n(s)/n$
- $p'_n(s) = p(s \mid s_{1:n}s)$: probability assigned by the density model to s after observing a new occurrence of s

$$p_n(s) = \frac{\hat{N}_n(s)}{\hat{n}} \leq \frac{\hat{N}_n(s)+1}{\hat{n}+1} = p'_n(s)$$

  - where $\hat{N}_n(s)$ is a pseudo-count function and $\hat{n}$ a pseudo-count total which regulates the density function.
  - learning-positive of density function is required since visiting s again ($p'_n(s)$) should increase probability

Count-based Intrinsic Bonus

- Common choice [Strehl and Littmann. 2008]

$$r_t^i = N(s_t, a_t)^{-1/2}$$

- For pseudo-count based exploration, very similar:

$$r_t^i = (\hat{N}_n(s_t, a_t) + 0.01)^{-1/2}$$