Combining RND-DQN with Prioritized Experience Replay leads to better performance in MiniGrid DoorKey Environment

Clara Schindler, Sarah Secci

1 Motivation

In week 7 we combined RND with DQN and trained it on MiniGrid. The rewards showed that the training was not as effective as we have hoped. We believe that this is due to the agent looking for ways to get the intrinsic reward for exploration from RND rather than revisiting states that lead further towards the goal. To counteract that we want to implement Prioritized Experience Replay (PER) (week 4, level 3) to only prioritize visiting new states in the beginning of the training.

2 Related Topics

- DQN (week 4, level 1)
- Prioritized Experience Replay (week 4, level 3)
- RND (week 7, level 1)

3 Idea

Combining the RND-DQN from the exercise in week 7 with a PER.

4 Experiments

Test RND-DQN with PER against RND-DQN without PER and compare which setup performs better in the MiniGrid DoorKey environment.

Environments & Metrics We want to train on the MiniGrid DoorKey environment because we think that exploring new states is important to solve this environment. The agent needs to pick up a key in order to open the door that blocks the way towards the goal. We want to measure how successfully the agents learn by the returned extrinsic rewards over 500e3 - 1e6 steps.

Experimental Scope

- setups: RND-DQN vs RND-DQN with PER
- seeds: 10-20
- hyperparameters: network architecture (DQN and RND), learning rate, ϵ

Question: What if the best hyperparameters for RND-DQN with PER are not the same as the best hyperparameters for the baseline?

Estimated Computational Load To keep the compuptational load small we want to train a 5x5 (or 6x6) MiniGrid. Our estimate computational time is 15-20 hours.

Project for Reinforcement Learning lecture

5 Timeline

- Research: 2 days
- Implementation: 2 days
- Experiments: 3-4 days
- Analysis: 3-4 days
- Reporting: 2 days