Bachelor's Proposal: Sequential Decision Making with Delayed and Sparse Rewards

Valdemar H. Lorenzen (valor21)

Problem

In reinforcement learning, obtaining algorithms with good sample efficiency while maintaining high asymptotic performance is crucial. This is especially challenging in environments with delayed and sparse rewards, where traditional reward design can be expensive or infeasible. Examples include drug design or image synthesis, where intermediate steps cannot be easily evaluated. This project aims to explore methods for sequential decision-making in such environments, focusing on sparse reward techniques.

Data

We will evaluate the success of our algorithms using illustrative simulation environments commonly used in research on discrete Markov Decision Processes, such as:

- n-chain
- GridWorld

Methods

We will investigate sparse reward techniques, including:

- GFlowNets[1]
- Deep exploration networks (BEN)[2]

Evaluation

To evaluate our algorithms, we will use the simulation environments mentioned in the Data section.

Distribution of work

The following is an estimate of the percentage-wise distribution of workload for this project:

- Literature Survey: 20%
- Implementation of proof of concept: 25%
- Iterative hypothesis refinement: 40%
- Report writing: 15%

Bibliography

[1] E. Bengio, M. Jain, M. Korablyov, D. Precup, and Y. Bengio, "Flow Network based Generative Models for Non-Iterative Diverse Candidate Generation." [Online]. Available: https://arxiv.org/abs/2106.04399

[2] M. Fellows, B. Kaplowitz, C. S. de Witt, and S. Whiteson, "Bayesian Exploration Networks." [Online]. Available: https://arxiv.org/abs/2308.13049