

Assignment 2: Policy Gradients with Intrinsic Exploration in Gymnasium

CS395T: Foundations of Machine Learning for Systems Researchers
Assignment 2: Policy Gradients with Intrinsic Exploration in Gymnasium

Objective

The goal of this assignment is to implement a policy gradient algorithm with intrinsic motivation to solve a sparse-reward reinforcement learning problem. You will use **Gymnasium** and target a hard exploration environment such as *Montezuma's Revenge*. The focus will be on integrating intrinsic reward mechanisms to improve exploration efficiency.

Task Description

1. Baseline Method Implementation

Implement a policy gradient method such as PPO or A2C. Do not use libraries such as StableBaselines-3 that provide existing implementations.

2. Intrinsic Motivation Module

- Add one of the following as an intrinsic reward signal:
 - Random Network Distillation (RND)
 - Intrinsic Curiosity Module (ICM)
- Log extrinsic and intrinsic rewards.

3. Comparative Evaluation

Compare your agent's performance against a baseline agent without intrinsic reward. Try **at least two environments** in Gymnasium: an easier task of your choice, and Montezuma's Revenge.

Deliverables

1. **Code:** Code implementing a baseline method and an intrinsic reward module such as RND or ICM. You can use any platform you prefer.
2. **Plots:** Learning curves showing:
 - Total reward over training timesteps, compared to the baseline agent without intrinsic rewards.
 - Extrinsic vs. intrinsic rewards over training timesteps.
3. **Report (2–3 pages):**
 - Experimental setup: algorithm details, hyperparameters, hardware, environment information.
 - Performance comparison with baseline agents, including plots.
 - Observations about training stability, exploration, and limitations.

Grading Breakdown

- Correct baseline method implementation: 30%.
- Correct intrinsic reward implementation 30%.
- Comparative evaluation and plots: 20%.
- Report quality and clarity: 20%.