Count-Based vs. RND Exploration in ObstructedMaze

Giulia Luongo

1 Motivation

In many real-world tasks rewards are rare and hidden behind obstacles. In ObstructedMaze environments, for example, an agent must first find a hidden key before reaching the exit, yet receives no feedback until both steps are complete. By comparing two classic exploration bonuses in this setting, this study aims to identify which bonus more reliably drives discovery under partial observability. Count-based bonuses encourage agents to visit states that have been seen less often, whereas random network distillation (RND) rewards agents for encountering observations that their predictors find difficult to predict. By testing both methods on ObstructedMaze variants, it's possible to determine which bonus helps the agent find the key and open the door more quickly. This comparison will reveal whether state-visit counts or learned prediction errors are more effective at guiding exploration through hidden passages. It is hypothesised that RND will outperform count-based bonuses in the full maze due to its learned novelty signal coping better with aliasing, while count-based bonuses may suffice in the simpler 1Dlhb variant.

2 Related Topics

Random Network Distillation and Count-Based Exploration (Week 7), State Aliasing and Partial Observability (Week 5).

3 Idea

The performance of two DQN agents will be evaluated in the ObstructedMaze-1Dlhb-v0 and ObstructedMaze-Full-v0 environments. One agent will use count-based exploration bonuses and the other will use RND intrinsic rewards. Both agents will have the same network architecture and hyperparameters; the only difference will be in how they compute intrinsic reward. The ObstructedMaze benchmark is ideal because it combines sparse rewards (no feedback until the key is picked up and the door opened) with state aliasing caused by hidden passages, forcing the agent to explore efficiently. Learning curves will be tracked and the 'frame-to-solve' (steps until the first successful key-door run) will be computed, as well as periodic success rates, to determine whether state-visit counts or network prediction error helps the agent find the key and open the door faster under aliasing.

This is a confirmatory study. In the Full maze, where walls obscure passages, it is expected that RND's prediction-error bonus will yield a lower 'Frame-to-Solve' than count-based bonuses. In the simpler 1Dlb maze, however, it is anticipated that both methods will perform similarly.

4 Experiments

Environments & Metrics

- **Environments**: ObstructedMaze-1Dlhb-v0 and ObstructedMaze-Full-v0.
- **Metrics**:

- *Frame-to-Solve*: Number of steps until the agent first completes the key–door sequence. The two agents' median Frame-to-Solve will be compared using a Mann–Whitney U test ($\alpha = 0.05$) to see if RND is significantly faster than count-based exploration.
- *Success Rate*: Proportion of episodes solved (key collected and door opened) measured at regular frame intervals.

Experimental Scope Each exploration method (Count-Based bonus vs. RND bonus) will be run on both ObstructedMaze-1Dlhb-v0 and ObstructedMaze-Full-v0, with 5 independent random seeds per agent–environment pair. 5 seeds are chosen as a common compromise between obtaining reasonably tight confidence intervals and limiting overall training time. All runs will use identical network architectures and hyperparameters. In total, we will perform

$$2\ agents \times 2\ environments \times 5\ seeds\ =\ 20\ runs$$

To assess the effect of obstacle complexity, the performance of each agent on the simple (1Dlb) and full (Full) variants of the maze will be compared directly.

Estimated Computational Load

5 Timeline

- Read papers and review ObstructedMaze environments (1-2 days)
- Implement count-based bonus and RND modules in the DQN codebase (3 days)
- Run full training experiments for both methods on both maze variants (3-4 days)
- Analyze results, generate learning curves, compute "frame to solve" and success-rate plots, perform statistical tests (4 days)
- Write the final report and prepare poster (2 days)