Report: MARL Patrolling with Obstacles and Penalty Mechanisms

Introduction

Multi-Agent Reinforcement Learning (MARL) involves the training of multiple agents that interact within a shared environment. In patrolling tasks, MARL simulates scenarios where predators aim to catch prey within a defined grid. The environment is often enhanced with obstacles to add realism and complexity. However, in the original implementation of this project, obstacles were static and did not impose penalties for collisions, limiting their impact on agent learning and performance.

This project introduces a penalty mechanism for obstacle collisions, evaluates its effect on performance, and compares the outcomes for two configurations:

- **3 Predators and 1 Prey (3v1)**
- **2 Predators and 2 Prey (2v2)**

Simulations were conducted for 10,000 and 20,000 episodes, and the results were analyzed through three metrics:

1. **Loss**
2. **Returns**
3. **Number of Collisions**

Objectives

1. Enhance the obstacle functionality by introducing collision penalties.
2. Examine the performance impact on predator-prey dynamics in obstacle-rich environments.
3. Conduct experiments under different configurations (3v1 and 2v2) and episode durations.
4. Interpret and evaluate the resulting patterns in losses, returns, and collisions.

Experimental Setup

Environment

- **Grid Dimensions**: A bounded grid environment with random obstacle placement.
- **Agents**:
    - Predators: Aim to capture prey.
    - Prey: Attempt to evade predators while navigating the grid.
- **Obstacles**:
    - Impose movement constraints.
    - Collision penalty added to the reward function.

Configurations

1. **3 Predators vs. 1 Prey (3v1)**:
    - Higher predator cooperation is required to capture the single prey.
2. **2 Predators vs. 2 Prey (2v2)**:
    - Balances predator-prey dynamics with increased complexity due to equal agent numbers.

Training

- **Framework**: Deep Q-Learning with penalty-based rewards.
- **Episodes**: Simulations were run for 10,000 and 20,000 episodes.

Results

The experiments yielded data for **loss**, **returns**, and **number of collisions** across configurations and episode durations. Visualizations for each metric are provided below.

Metric 1: Loss

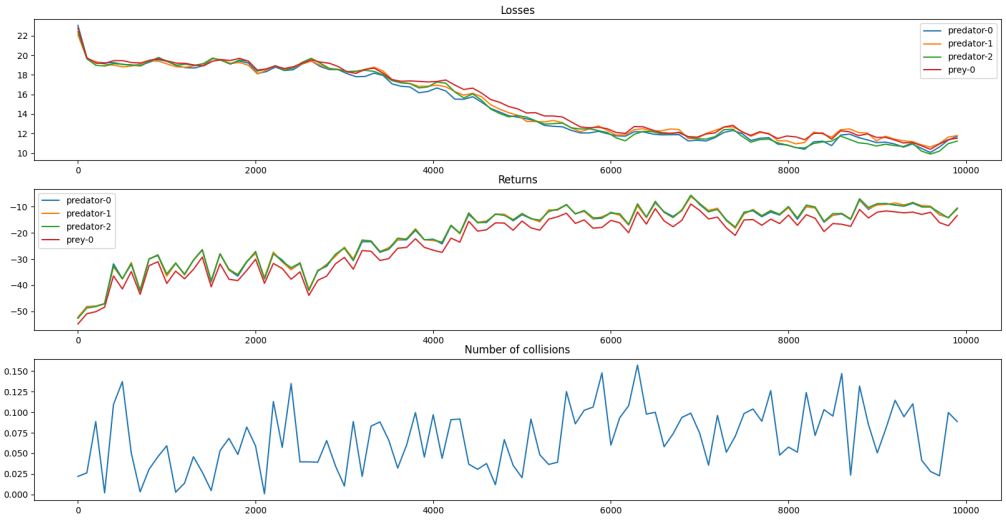
- **Observation**:
    - Losses decreased over time as agents learned optimal policies.
    - Initial instability was observed, reflecting exploration phases.
- **Comparison**:
    - Loss converged faster in the 3v1 configuration, likely due to simpler prey dynamics.

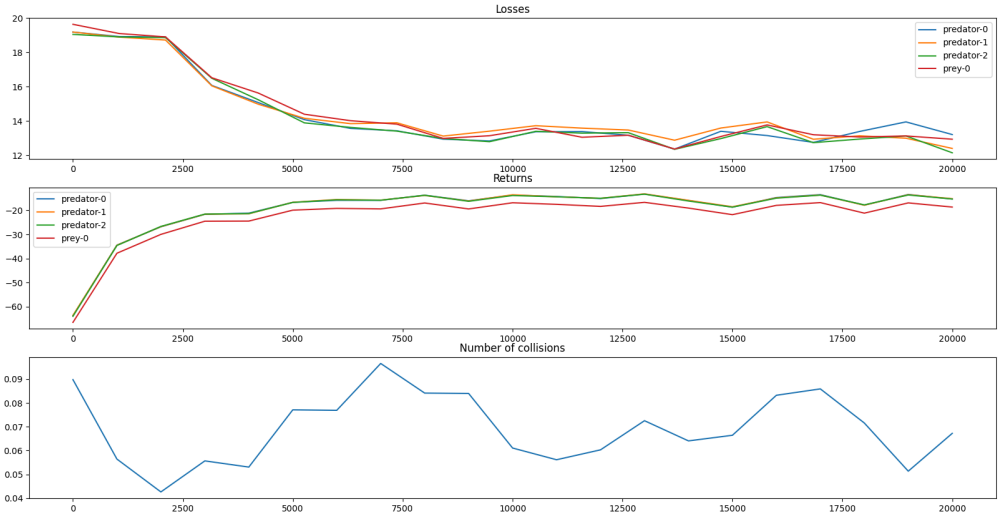
Metric 2: Returns

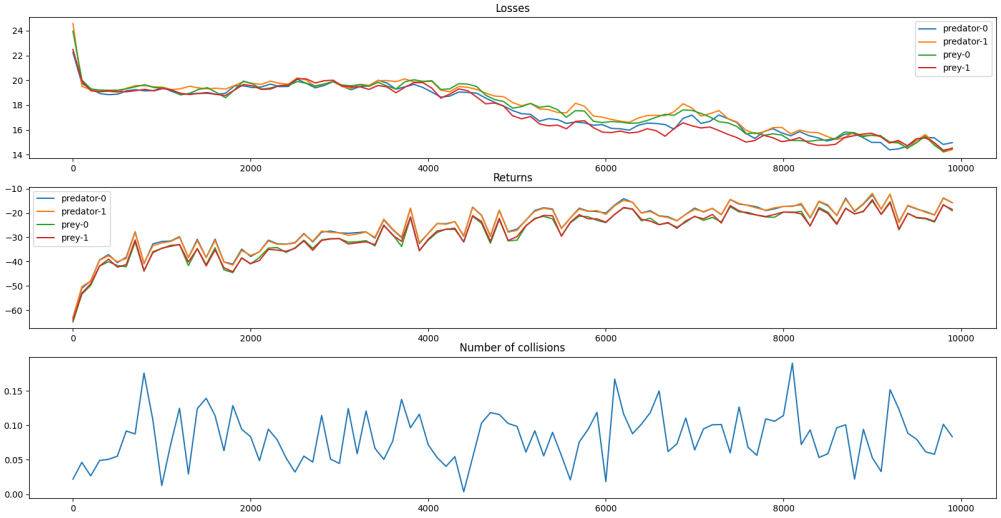
- **Observation**:
  - Returns increased with episodes, indicating improved agent coordination and performance.
  - Prey returns remained consistently lower due to the penalty mechanism.
- **Comparison**:
  - In 2v2, returns plateaued slower, suggesting that prey could evade predators longer due to increased complexity.

Metric 3: Number of Collisions

- **Observation**:
  - Collisions peaked early during training as agents explored the environment.
  - A gradual decline followed, demonstrating learned obstacle avoidance.
- **Comparison**:
  - 3v1 resulted in fewer collisions due to fewer agents in the grid.
  - 2v2 saw higher initial collisions, stabilizing with longer training.







Analysis and Interpretation

Impact of Obstacles

- Adding penalties for obstacle collisions increased the complexity of decision-making for agents.
- Predators adapted to avoid obstacles while pursuing prey, improving learning dynamics.

Comparison Across Configurations

- 3v1 was more stable and faster to converge due to reduced agent interactions.
- 2v2 showcased more complex behaviors but required longer training to stabilize.

Training Duration

- Prolonged training (20,000 episodes) showed marginal improvements in returns and stability, suggesting diminishing returns with extended learning.

Conclusions

1. **Obstacle Penalties**: Successfully added realism and improved learning outcomes by penalizing undesirable behaviors.
2. **Configurations**: The 3v1 scenario exhibited faster learning, while 2v2 demonstrated richer dynamics at the cost of complexity.
3. **Training Insights**: Prolonged training improved agent behavior but introduced diminishing returns after a certain threshold.

Future Work

1. **Dynamic Obstacles**: Introduce moving obstacles to further challenge agent adaptability.
2. **Advanced Reward Shaping**: Incorporate rewards for cooperative behaviors among predators.
3. **Real-World Simulations**: Extend to robotic agents for real-world patrolling tasks.