The Fire Squad: Multi-Agent Reinforcement Learning for Collaborative Wildfire Suppression

Abdul Rahman Hussain Siddique (ah244)
Anirudh Ramesh (aramesh8)
Yashwanth Pulimi (ypulimi)

Fall 2025

1 Short Summary

Wildfire suppression is a complex, spatiotemporal challenge where the environment changes dynamically based on wind, terrain, and fuel. Current single-agent reinforcement learning (RL) approaches often struggle to cover large grid worlds effectively. This project proposes extending the FirecastRL environment to a Multi-Agent Reinforcement Learning (MARL) setting. The primary research question is whether a **Cooperative** reward structure (penalizing global fire spread) induces better emergent behaviors and suppression rates compared to a **Greedy/Independent** reward structure (maximizing individual extinguishment counts). This is an important problem as it applies MARL to disaster response, where resource coordination is critical for minimizing damage.

2 Objectives

The main goal is to build a "Fire Squad" of helicopters that learn to coordinate their actions to contain wildfires more efficiently than independent agents. The specific objectives are:

1. To modify the existing FirecastRL environment (Gymnasium) to support multiple simultaneous agents ($N = 3$) with vectorized observation and action spaces.
2. To implement a baseline "Greedy" agent using Proximal Policy Optimization (PPO), where agents are rewarded solely for their own water drops.
3. To implement a "Cooperative" agent where the reward function includes a global penalty for the total number of burning cells, incentivizing containment over individual scoring.
4. To quantitatively compare the two strategies based on the total percentage of forest saved and the time required to extinguish the fire.

3 Methodology

This project falls under the Environment Extension and Algorithm Implementation category. We will simulate a fleet of helitack agents fighting a fire on a 160×240 grid.

3.1 MDP Formulation

We model the multi-agent suppression as a centralized MDP:

- **State ($s \in S$):** A global grid representing ignition times and fire status, combined with a vector of coordinates for all N helicopters.
- **Action ($a \in A$):** A MultiDiscrete vector $[a_1, a_2, ...a_N]$ where each agent can move (Up, Down, Left, Right) or Drop Water.
- **Reward ($r \in R$):**
  - *Greedy Mode:* $r = \sum$(Individual Extinguishments) − Wasted Water.
  - *Cooperative Mode:* $r = \sum$(Individual Extinguishments) $- \alpha \times$ (Total Burning Cells).

3.2 Algorithms

We will utilize **Proximal Policy Optimization (PPO)** from the Stable-Baselines3 library. We will employ a Centralized Training, Decentralized Execution (CTDE) paradigm where the policy observes the full state during training to coordinate the agents.

4 Evaluation

We will evaluate the models using a custom evaluation loop over 10 seeded episodes to ensure statistical significance.

- **Metric 1: Final Burnt Area.** The total count of cells destroyed by the fire. We hypothesize the Cooperative agent will result in a lower burnt area.
- **Metric 2: Suppression Speed.** The number of timesteps required to terminate the episode (0 active fires).

5 Environment

We will use the **FirecastRL** environment, a physics-informed wildfire simulation compatible with Gymnasium. The environment simulates fire spread using Rothermel's surface fire spread model, incorporating real-world data for elevation and land cover.