Reinforcement Learning for Energy-Efficient Trajectory Planning of Drone Networks

Changling Li ‘22, Jiyao Chen ‘22 and Ying Li Professor
Department of Computer Science, Colby College, Waterville, ME

Background and Motivation

- Drone’s high mobility and easy deployment makes it prevalent in task executions. (Figure 1)
- **Problem**: limited battery capacity and constrained operation time.
- **Past research**: Trajectory planning using linear algorithms [1]. However, time complexity grows exponentially as the environment scales up. When the task location changes, the algorithm needs to recalculate the optimal path.
- **Solution**: Decentralized multi-agent reinforcement learning (MARL) [2].


Fig. 1. Drone executes real-life tasks

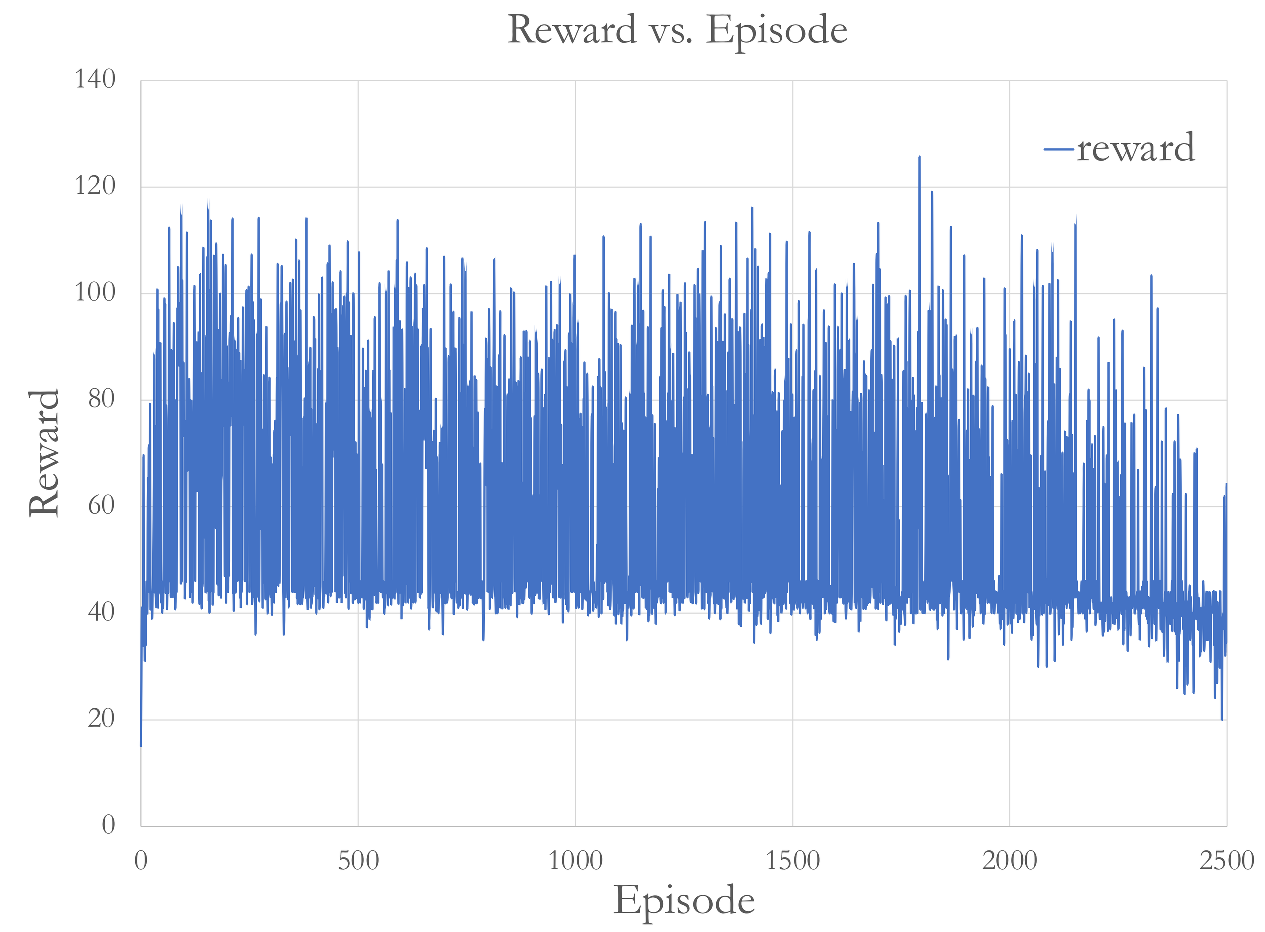

Fig. 4. Plot of shared reward per episode

Methodology

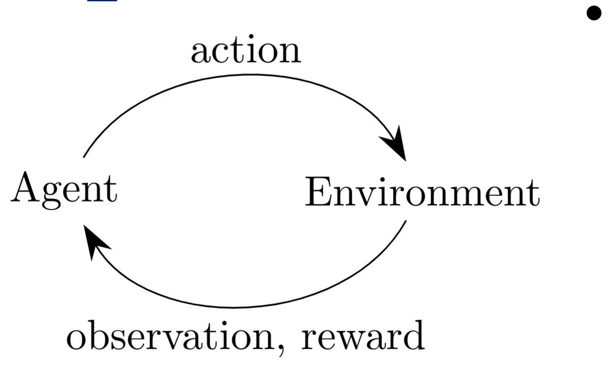

Fig. 2. The agent-environment loop.

- **Reinforcement Learning** imitates human’s learning process. Through interaction with the environment, the agents receive a reward to adjust its behavior towards successful task execution. By maximizing the cumulative reward, the agents will find the optimal solution to the problem. (Fig. 2)
- We employed **Deep Q-learning Network** (**DQN**) [3], a reinforcement learning model, in this project. Each agent has its own network, and they interact with the environment and other drones with shared reward.
- We created an environment to reflect our project as shown in Fig. 3. Each task has a non-binary length.
- To capture the relation between the environment and drones’ action and state, we formulated our **reward functions**:

$$R(s,a) = T_t, \text{ if not all tasks are finished,}$$
$$R(s,a) = T_t - B + \eta, \text{ otherwise.}$$

where T_t denotes the number of tasks executed at time step t, B represents the number of drones depleted, and η refers to energy efficiency of task execution.

- We aim to train the model over 5000 episodes to gain stable results.

Fig. 5. Plots of losses of selected 3 agent per episode. The episodes are divided into two plots to better reveal the details.

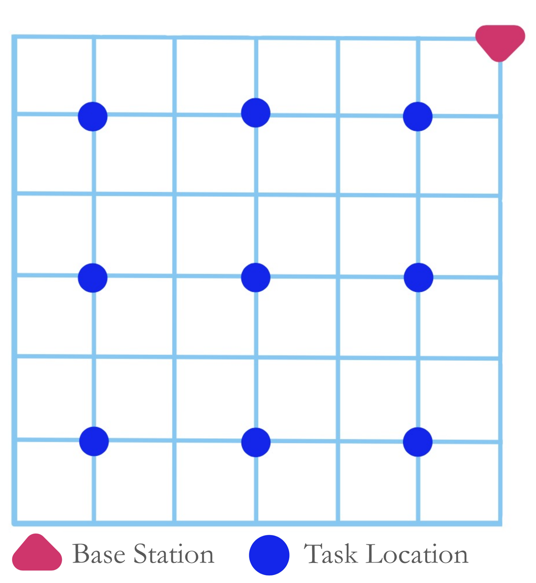

Fig. 3. The environment has 9 tasks and 1 base station (BS). All drones launch from BS and learn to execute the tasks at right locations.

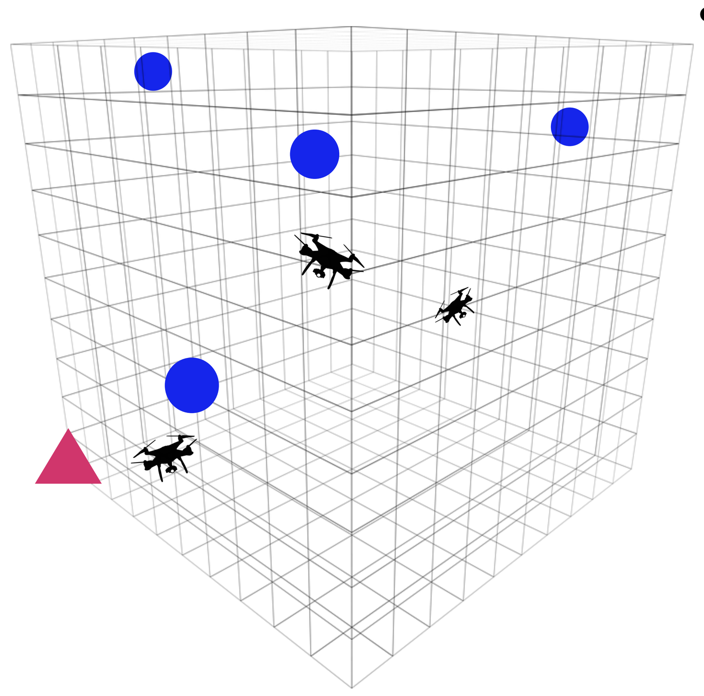
Fig. 8. The schematic diagram of 3D environment

Result

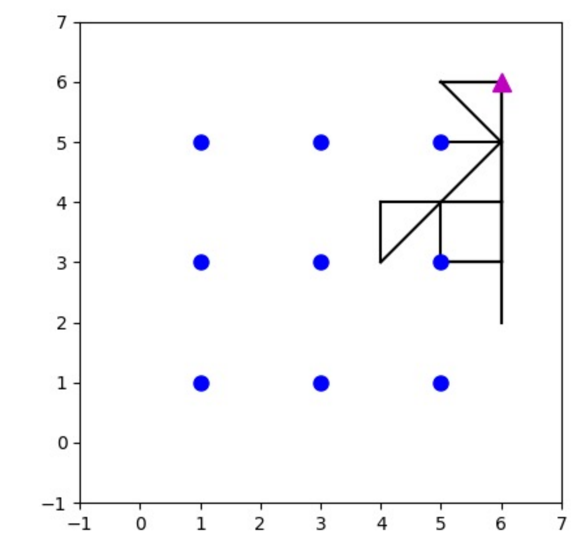
Fig. 6. The trajectory of one of the drones

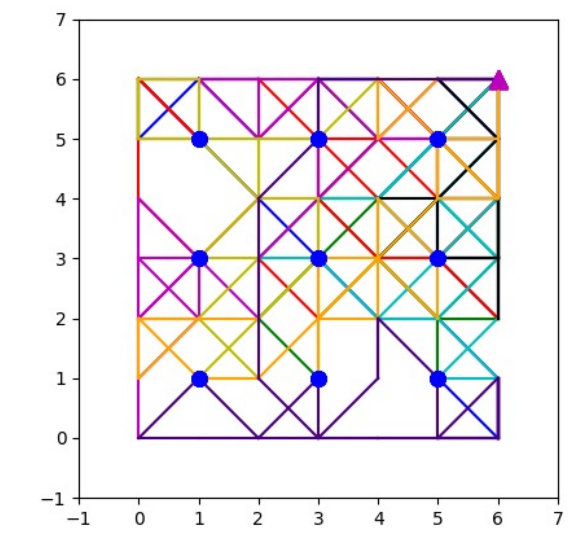
Fig. 7. The trajectories of all the drones. Each color represents a different drone.

- Due to time limitation, we can only present partial results and draw a few analysis from the current data over 2500 episodes.
- As shown in the plot of reward vs episode (Fig. 4), the best reward shows a slight tendency to increase over episodes. The current best reward is 125.69 with corresponding steps of 93 to finish the tasks. One of the drones’ trajectory is plotted on coordinate shown in Fig. 6 and all drones trajectories are shown in Fig. 7 where each color represents a different drone. The trajectories show that all tasks are executed.
- As shown in Fig. 5 , the losses of 3 selected agents decrease over episodes. This reveals that the current trajectory approaches the current optimal trajectory.

Conclusion & Future Work

- The current best trajectory predicted by the DQN model is still far from the optimal result gained from linear algorithm. However, it is promising to employ multi-agent DQN model for energy-efficient trajectory planning of drone networks as the best reward continues to increase.
- More episodes of training are necessary to determine whether the performance is stable.
- Current result is based on one set of hyperparameter values. We plan to explore other possible combination of values to study the relationship between the parameters and the performance.
- In addition, we will evaluate the performance of multi-agent reinforcement learning in trajectory planning with different numbers of agents and tasks.
- We also plan to train the DQN model with random tasks for each episode to check how well it performs. We will also use a 3D environment to check the performance of drones in a more realistic environment setting.

Reference & Acknowledgement

[1] Li, Ying, and Chunchao Liang. "Energy-aware Trajectory Planning Model for Mission-oriented Drone Networks." *2021 IEEE International Systems Conference (SysCon)*. IEEE, 2021.
[2] Li, Ang, et al. "Deep Reinforcement Learning in Continuous Multi Agent Environments." http://www.rohansawhney.io/multi-agent-rl.pdf
[3] Sawhney, Rohan. "Multi-agent Reinforcement Learning." https://github.com/rohan-sawhney/multi-agent-rl