1 Background

Evolutionary algorithm and reinforcement learning have both shown great potential in optimization of agent behaviors in both single and multi-agent setting. Specifically, evolutionary algorithm has been shown to have better performance when there exists both local and global optima, compared to gradient based methods. In addition, previous result from YP [1] has shown that the population structure used in evolutionary algorithm could play an important role in affecting the performance. This leads us to the idea or goal of exploring the effect of population structure in improving the performance of evolutionary algorithms.

2 Progress in the past two weeks

2.1 Getting my feet wet in Reinforcement Learning

Bipedal walker is a classic example for reinforcement learning which could also be solved using evolutionary algorithm. This task involves a walker with two legs and four degree of freedom trying to move forward. The reward function is the distance moved in both cases. While the walker started off stumbling often and falling easily, it is able to eventually learn a human-like walking motion under training with both reinforcement learning and evolutionary algorithm.

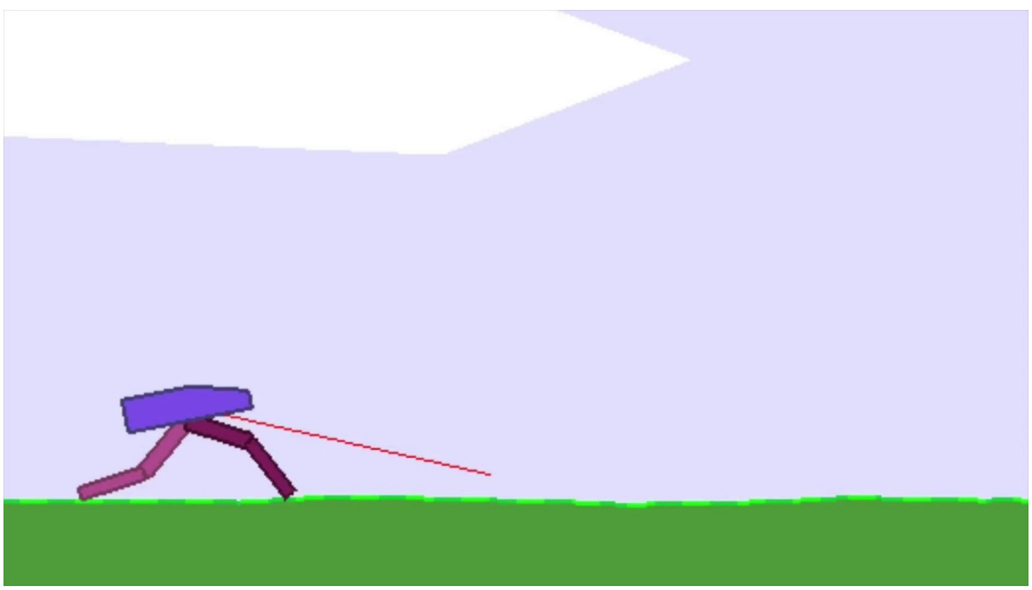

Figure 1: Simple example of a bipedal walker

During the past two weeks, I played with the simple example and got it to successfully train under both RL and evolutionary algorithm. One interesting fact is that under the same number of iterations, evolutionary algorithm seems to converge to optimal much faster than RL.

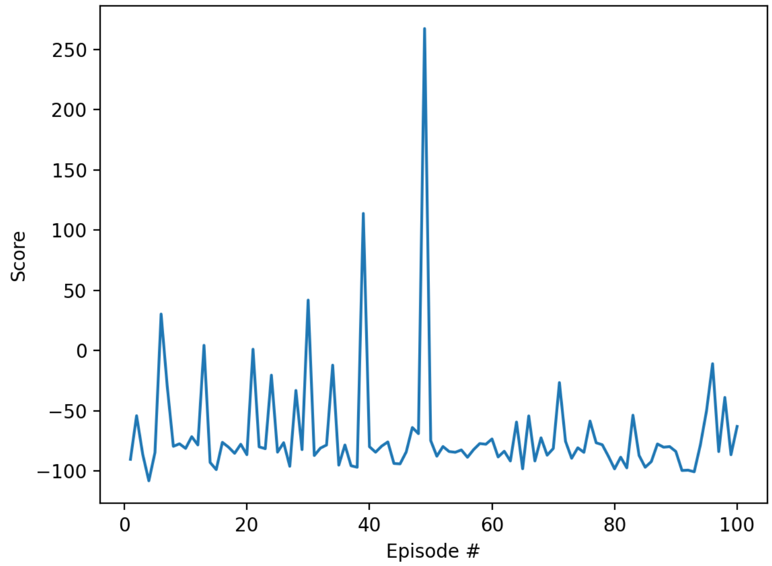

Figure 2: Score of walker under RL in the first 100 episodes

2.2 Brainstorming

The other major part of my work in the past two weeks are to continue reviewing literature and to brainstorm a way moving forward.
Most of the ideas revolve around the idea of a robot swarm trying to find the best out of n sites, with communication between individual robots to help most robot to identify the best site eventually.
Here are some of the good ideas that I have been thinking of:

- Communication could take place at two layers, a local communication layer that is based on a certain radius around the given robot, and a global layer regardless of robots' real locations in the setting.
- The case where we get an adversary in the robot swarm would be extremely interesting. There are two possible scenarios for the adversary, 1) a technical fault, which takes the form of faulty measurement and systemic error in information reported, and 2) a Byzantine failure, which means the robot could have any arbitrary unpredictable behaviors. In the former case, there are strategies that we could employ to minimize the damage of such error to the entire swarm.

2.2.1 Challenges

Our challenge remains to come up with an ideal setting that allows us to combine reinforcement learning, evolutionary algorithm and some level of population structure into one setting and assess its performance. So far, we have had some rough ideas but still need to fine-tune it and experiment.

3 Plans to move forward

Our current plan of moving forward will be using an example in swarm robotics where there could be an adversary sending false information to others. The task for the robot swarm would be to find the best site among n where $n > 3$, inspired by [2].

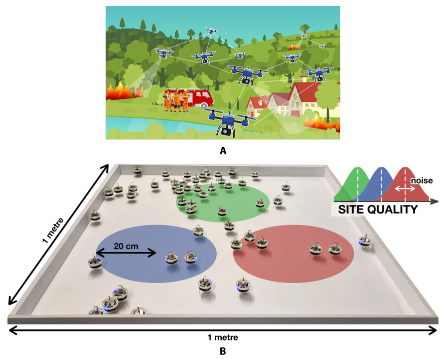

Figure 3: Figure demonstrating robot swarm exploring an enclosed space and looking for the best site

Robots that are within a certain radius would be able to communicate information they found, which may include the location and quality of the best site they have found. However, in practice, multiple factors could contribute to an incorrect information being disseminated, such as faulty sensors or adversarial deceptions and attacks.

In particular, assume with high probability only one robot in the swarm experiences technical issues which lead to systemic errors, then it is possible for the entire swarm to "learn" the error via either evolution or reinforcement learning, so the swarm could eventually reach agreement on the correct site.

Specifically, when one robot r_a, originally believing in site s_a with quality q_a being the optimal, hears from another robot r_i about a certain cite s_i with quality q_i, there is a certain probability that could be proportional to $\frac{q_a}{q_a+q_i}$ for robot r_a to switch to believe that site s_i is the optimal site. However, if the robot r_i happen to have a technical failure that causes it to broadcast the real quality q_i of site s_i to be $1 - q_i$, this could harm the performance of the swarm in achieving the task of finding the best of all n sites.

In this case, if through an evolutionary algorithm the rest of the robots are able to learn to interpret the alternative way of communication, the overall population should be able to attain a higher level of fitness, which could be reflected as a faster rate of reaching consensus about the correct site compared to a population that does nothing about individuals who keep giving misinformation.

4 Expected deliverable

Given the model set up, I believe the reward that function that best describes the performance of the swarm would be

$$reward = \frac{\text{no. of robots correct}}{\text{no. of robots in swarm}}$$

There are a few baseline models that we could compare to:

1. all robots act by themselves, no communication exist among the swarm at all
2. ideal case, when robots have some local communications, without any adversary
3. one adversary in the robot, but no strategy applied (same protocol as 2.)

Our goal would be to devise a strategy using evolutionary algorithm such that the swarm could perform well despite the presence of an adversary. We will then compare the performance of our model against the three baseline models listed above.

References

[1] Carja, O., and Kuo, Y. Evolutionary graph theory beyond pairwise interactions: higher-order network motifs shape times to fixation in structured populations.

[2] Talamali, M. S., Saha, A., Marshall, J. A., and Reina, A. When less is more: Robot swarms adapt better to changes with constrained communication. *Science Robotics 6*, 56 (2021), eabf1416.