PROJECT PROPOSAL

TEAM SPACE FORTRESS

Project Topic

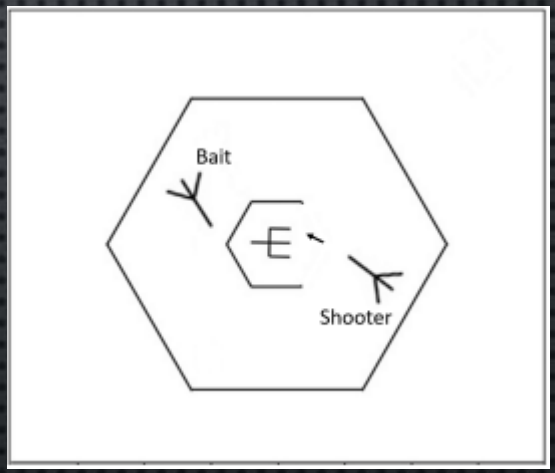
- Team Space Fortress – Multi-Agent Reinforcement Learning
- Yikang Gui, Gang Su.

BACKGROUND

- Space fortress or the revised space fortress game is a cognitive game that was originally developed by cognitive psychologists at the University of Illinois, as a tool to study learning and training strategies . Notably, it is one of the few cognitive training tools that has shown transfer of training to real-world performance.

BACKGROUND

- In our project, we will use team space fortress as our environment. The different between space fortress and team space fortress is that in team space fortress, there are two players instead of just a single player. There are two roles in the game, bait and shooter.



Related work (no more than 1 slide)

- Our project is a multi-agent cooperative reinforcement learning project. There are some related work, such as MADDPG, JAL, LOLA.
- MADDPG: an adaptation of actor-critic methods that considers action policies of other agents and able to successfully learn policies that require complex multi-agent coordination
- LOLA: a method in which each agent shapes the anticipated learning of the other agents in the environment

Problem definition (no more than 1 slide)

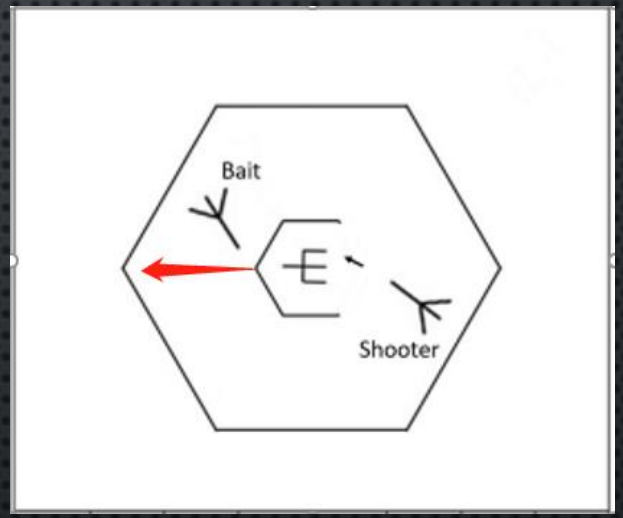
- Multi-Agent Reinforcement learning problem. Specifically, two agents in the environment.
- Try to maximize the reward in each time.
- Two players: bait and shooter
- Shooter can move and shoot the shell
- Bait can only move
- Fortress will keep attacking the spaceship that enters the area first
- When the fortress is attacking, the back of the fortress is vulnerable, and the shell can destroy the fortress.
- Goal: destroy the fortress as much as possible in a given time
- Avoid colliding with the boundary and the fortress

Approach (no more than 4 slides)

- From our preliminary experiments, our baseline model will be two independent reinforcement learning agent, which means one for bait and the other for shooter.
- The baseline model will use PPO to train the agent.
- However, one of the downside of PPO is the exploration and due to the original reward structure is very sparse. The PPO behaves poorly on the original reward function.
- It failed to converge to optimal policy or even suboptimal policy.

APPROACH

- Therefore, we will introduce reward shaping to the original reward function.
- For example, we know that what the bait should do is that try to make sure the fortress aiming itself and move as slowly as possible to stabilize the fortress such that the shooter can shoot the fortress easily.
- Thus, the reward shaping for bait could be a smooth function that the bait will get higher reward as it aligns with the shooting angle of the fortress. The red arrow in the figure is the maxima in the reward function.



APPROACH

- In terms of the shooter, since it the symmetrical to the bait, it can use the negative reward function from the bait. What's more, adding another bonus of destroying the fortress will be enough for shooter to have a good result.

APPROACH

- However, above techniques deal with the environment by individual reinforcement learning agent. Therefore, we want to try some multi-agent reinforcement learning algorithm as well. We have read some papers, such as MAPPO, MADDPG. And we want to implement those algorithm on our environment as well to see whether there is any improvement by using multi-agent algorithms.
- We prefer to use centralized critic and decentralized actor because we want different types of bait and shooter and they can cooperate.

EVALUATION OR DEMONSTRATION PLAN (NO MORE THAN 1 SLIDE)

- THE PERFORMANCE IS EVALUATED BY THE TOTAL REWARD FOR AN EPISODE.
- THE REWARD IS INFLUENCED BY THE NUMBER OF KILLED FORTRESS OR THE TIME TO KILL THE FORTRESS.
- WE WILL HAVE A GRAPHICAL VIDEO FOR EACH EPISODE TO HAVE A BETTER UNDERSTANDING OF WHAT AGENTS ARE DOING.

DELIVERABLES (NO MORE THAN 1 SLIDE)

- We will write a report for this project.
- If everything goes right, we will have a good pair of agents that can cooperate smoothly in the environment.
- Also, we can play as bait or shooter and the agent will play the other role.

Responsibilities of team members (no more than 1 slide)

- Build Environment: Yikang
- Baseline model(independent agent): Yikang and Gang
- Advanced model(Multi-agent algo) : Yikang and Gang
- Report: Gang