Mini Football

ECS427 - Multi-Agent Reinforcement Learning | Prof. Sujit P B | Fall 2024-25

Sattwik Kumar Sahu
21241

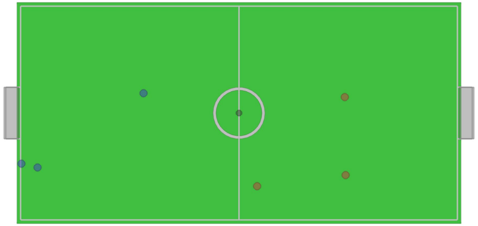
Rugved Upaddhye
21294

Problem Statement

- To develop and optimize a multi-agent reinforcement learning (MARL) policy for playing **mini football**.
- Challenges
  - **Cooperative & Competitive Strategies:** To make the agents *cooperatively* and *competitively* learn strategies to **score goals, dribble, pass, defence**, and **adapting to the actions of the opponent**.
  - **Individual and Collective Learning:** Each agent should learn to **control the ball individually** and also **coordinate with teammates** to *keep possession of the ball* with its own team.
  - **Continuous State Space:** Agents must learn from an **infinite possibilities** of *positions, velocities, and orientations* on the field. It is **difficult to map** each combination of state variables to a particular optimal action.
- We aim to create **autonomous players (agents)** that can learn **complex behaviours** such as teamwork, strategy, and optimal decision-making in dynamic, rapidly changing environments.

Setup

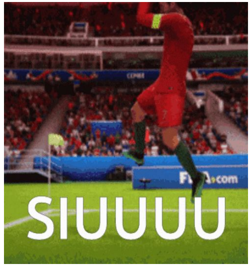
Environment



- 3v3 mini football match
- Team that scores goal wins
- Episode ends if ball out of pitch

Goal

Score a GOAL



Agent

Dribble	Shoot	Pass	Move
Move with the ball toward the goal.	Shoot the ball toward the goal.	Shoot the ball toward a teammate.	Move without the ball to a more strategic position.

Rewards

- **Dense:** Using "attack value" from
  - Distances to opponents
  - Distance from opponent's goal
  - How clear is the path from agent to the ball, and to the opponent goal?
- **Sparse:**
  - Goal scored by own team? **+1**
  - Goal scored by opponent? **-1**

Approach

Algorithms

- Multi-Agent Deep Deterministic Policy Gradient (MADDPG)

Software

- **Language:** `Python`
- **Training:** `torchrl`
- **Benchmarking:** `BenchMARL` (uses `torchrl`)
- **Environment:** Vectorized Multi-Agent Simulator (VMAS), *included with* `torchrl`

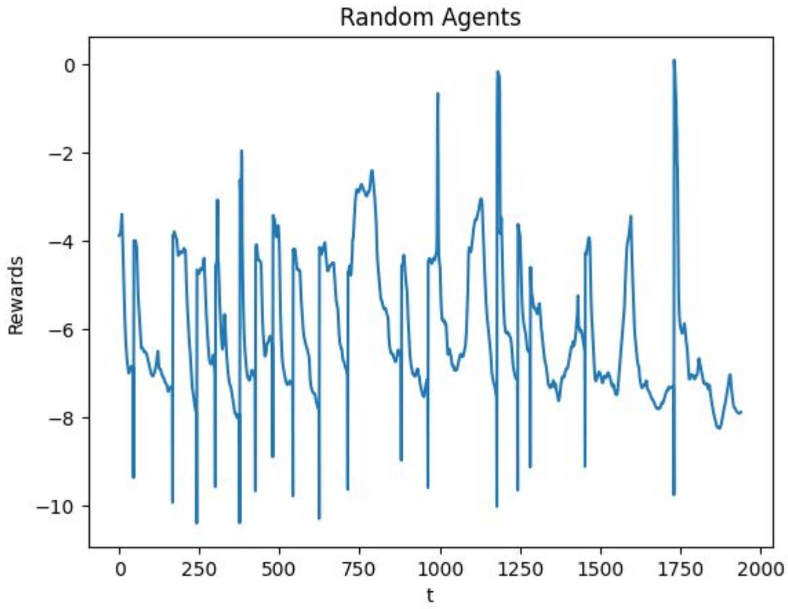
Methodology

Network Architecture

- Actor Network:
  - Input: (s, a)
  - 3 hidden layer of size 64, 256 with ReLU activation.
  - Output: Q(s,a)

- Critic Network:
  - Input: (s, a)
  - 1 hidden layer of size 64 with ReLU activation.
  - Output: Q(s,a)

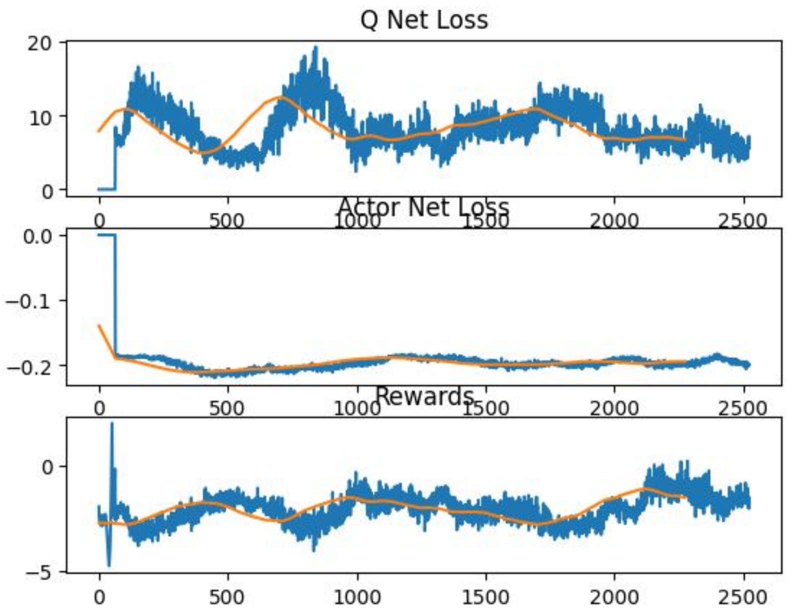
Experiment 1: Random (Baseline)



Remarks:
Mean Reward = -6.01

The agents moved haphazardly and even went out of the pitch.

Experiments (MADDPG)

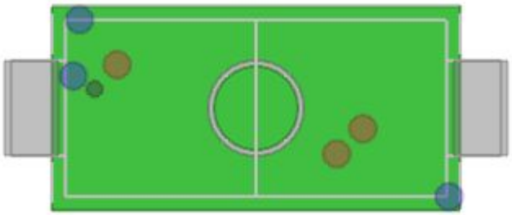



```
n_epochs = 20
batch_size = 32
should_update = False
min_buffer_items = 64
buffer_size = 256
n_updates = 128
gamma = 0.99
steps = 200


noise = 0.2
noise_decay = 5e-5


actor_lr = np.exp(-3)
q_lr = np.exp(-3)


tau = 0.005
```

Remarks:

Mean Reward: -2.12

Blue agents learned that the team benefits if there is a *goalkeeper*



Rewards

$$R_{\mathrm{BA}} = -\ln\left[\min_{1\le i\le n_{\mathrm{agents}}}[\mathbf{p_i} - \mathbf{p}_{\mathrm{ball}}] + \exp(-R_0)\right]$$

(Penalty based on the distance between ball & agent)

$$R_{\mathrm{half}} = \tanh(5 \cdot x_{\mathrm{ball}})$$

(Reward based on in which half the ball is)

$$R_{\mathrm{goal}} = \begin{cases} 10, & \text{if } \texttt{BLUE} \text{ has scored} \\ -10, & \text{if } \texttt{RED} \text{ has scored} \end{cases}$$

(Reward based on goal scored)

$$R_{\mathrm{border}} = \frac{1}{1 + \exp\left[\frac{|y_{\mathrm{agents}}| - r \cdot W_{\mathrm{pitch}}}{a}\right]}$$

(Penalty for going near border/corners)

$$\boxed{\begin{aligned} R_{\mathrm{sparse}} &= R_{\mathrm{BA}} \\ R_{\mathrm{dense}} &= 2R_{\mathrm{half}} + 3R_{\mathrm{goal}} + 5R_{\mathrm{border}} \end{aligned}} \longrightarrow \boxed{R = \rho R_{\mathrm{dense}} + (1-\rho) R_{\mathrm{sparse}}}$$

Observations

1. In early episodes, blue agents are not wise whereas red agents know how to play. Therefore they goal. Defending them gives positive reward to our agents. So blue agents learn to defend, whereas their attacking learning is very poor.
   Thus, over some iterations, Blue agents become better defenders.
2. In earlier experiments, it was observed that the agents tend to go towards the borders and corners of the pitch. The border penalty was added to mitigate this issue.

Thank You