UTAustinVilla: A Self-Play Framework to Learn Team-Level Strategy

Caroline Wang [*,1], William Macke [*,1], Patrick MacAlpine [2], Peter Stone [1,2]

*Equal contribution.
[1]The University of Texas at Austin [2]Sony AI

Multi-agent deep reinforcement learning has seen great success in both cooperative and competitive games [6, 7, 1]. Unlike many prior games studied by the AI community, the game of soccer is a mixed cooperative-competitive game that is two-player zero-sum on the team level, yet fully cooperative at the agent level. Prior work [4] integrated an algorithm for cooperative multi-agent reinforcement learning (MARL) into a self-play framework, enabling learning complex coordination strategies from scratch in Google Football [3]. [1]

Inspired by these successes, UTAustinVilla developed a self-play learning framework for the RoboCup 3DSimulation Fat Proxy environment, enabling learning complex team strategies via multi-agent deep reinforcement learning. The Fat Proxy environment provides two high level actions of `Dash` and `Kick`, where each action has three parameters that must be specified by the agent. Using the high-level actions, the agents must cooperate as a team to play a soccer game. The self-play learning framework is as follows. First, an 'ego' team of 11 agents is initialized to be random policies. An opponent pool is initialized with a similar team of random agents; it may optionally contain opponent teams sourced from prior RoboCup Fat Proxy challenges. An episode consists of a game played between the ego team and an opponent team drawn from the opponent pool. The reward function is cooperative—meaning that all agents on the ego team receive the same reward—and consists of terms based on the goal difference and distance of the ball to the opponent goal. The agents on the ego team are updated independently under the cooperative reward, with a variant of Proximal Policy Optimization (PPO) designed to handle the parameterized action space [5, 2]. At the end of a single self-play iteration, the current ego team is added to the opponent pool. This constitutes a form of fictitious self-play.

References

[1] Michael Bowling and Manuela Veloso. "Rational and convergent learning in stochastic games". In: *International joint conference on artificial intelligence*. Vol. 17. 1. Citeseer. 2001, pp. 1021–1026.

[2] Zhou Fan et al. "Hybrid Actor-Critic Reinforcement Learning in Parameterized Action Space". In: *International Joint Conference on Artificial Intelligence*. 2019.

[3] Karol Kurach et al. "Google Research Football: A Novel Reinforcement Learning Environment". In: *34th Conference on Artificial Intelligence (AAAI)*. 2020.

[4] Fanqi Lin et al. "TiZero: Mastering Multi-Agent Football with Curriculum Learning and Self-Play". In: *22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 2023.

[5] John Schulman et al. "Proximal Policy Optimization Algorithms". In: *ArXiv* abs/1707.06347 (2017).

[6] David Silver et al. "Mastering the game of Go with deep neural networks and tree search". In: *nature* 529.7587 (2016), pp. 484–489.

[7] Oriol Vinyals et al. "Grandmaster level in StarCraft II using multi-agent reinforcement learning". In: *Nature* 575.7782 (2019), pp. 350–354.

[1]Google Football is an alternative physics-based soccer benchmark. Unlike RoboCup 3dSim, agents use pre-specified high-level actions such as walking, kicking, and sprinting, and play against a fixed, provided enemy team. The focus of the benchmark is enabling cooperative MARL research.