Robust Multi-Agent Reinforcement Learning
By Minimax Deep Deterministic Policy Gradient

Shihui Li1 Yi Wu2 Xinyue Cui3 Honghua Dong3
Fei Fang1 Stuart Russell2
1CMU 2UC Berkeley 3Tsinghua U.

Overview (TL;DR)
• Improve the robustness of deep MARL policies – agent can still work well when opponent changes
• Introduce minimax into the MARL and optimize the objective using adversarial learning techniques
• A new algorithm M3DDPG (MiniMax Multi-agent Deep Deterministic Policy Gradient) – a minimax extension of MADDPG
• GitHub: https://github.com/dadadidodi/m3ddpg

Motivation
Multi-agent reinforcement learning (MARL) is promising for building intelligent agents
• Learns complex strategies with little human supervisions: StarCraft, Dota, robotics, green-security games, etc
• Provides a natural curriculum: agents help other agents get better

Deep RL policy is brittle & tends to overfit
• The environment is non-stationary from each single agent’s perspective → unstable
• Naïve MARL converges to a brittle equilibrium
• Opponent changes strategy → perform worse!

Robustness is critical for practical AI!

Multi-Agent Adversarial Learning
Adversarial learning for optimizing $J(\theta_i)$
• Replace min by one-step gradient descent
• Similar to adversarial training, each agent performs worst perturbation from its current policy
• Advantages: fast, differentiable, natural curriculum

Minimax MARL
Idea: learning minimax policy
• We consider continuous action
• Assume opponents perform adversarially
• Naturally leads to MADDPG formulation – decentralized Q function and policy

Key point: defining minimax Q function
• $Q_{M,i}(x, a_1, ..., a_N)$: the Q value for agent $i$ at state $s$ with actions $a_1, ..., a_N$ from $N$ agents, assuming other agents are adversarial
• $Q_{M,i}(x, a_1, ..., a_N) = r(x, a_1, ..., a_N) + \gamma \min_{a_j \neq i} Q_{M,i}(x', a_1', ..., a_N')$
• $J(\theta_i) = \min_{a_j \neq i} \mathbb{E}[Q_{M,i}(x_0, a_1, ..., a_N)]$

M3DDPG
The worst case performance – compete against best response
M3DDPG performs better with the worst opponents