DM²: Decentralized Multi-Agent Reinforcement Learning via Distribution Matching

Caroline Wang
The University of Texas at Austin
caroline.l.wang@utexas.edu

Ishan Durugkar
The University of Texas at Austin
ishand@cs.utexas.edu

Elad Liebman
SparkCognition Research
eliebman@sparkcognition.com

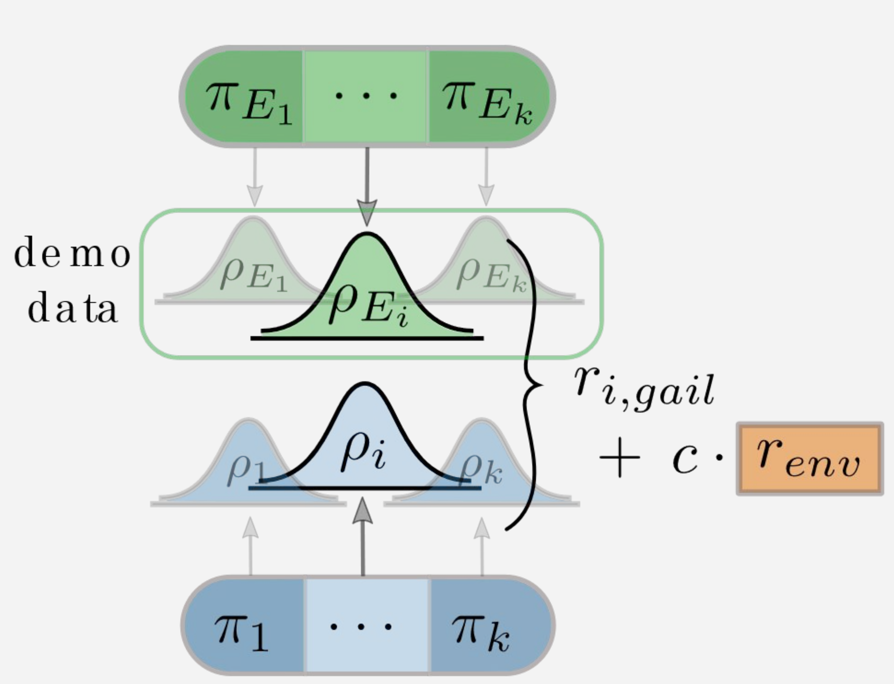
Peter Stone
The University of Texas at Austin
Sony AI; pstone@cs.utexas.edu

Overview

- Studies problem of **decentralized,** cooperative multi-agent learning **without explicit communication**
- Independent agent updates induce a nonstationary environment for other agents
- Proposes **DM2**, a decentralized MARL algorithm that performs distribution matching against expert demonstrations to facilitate coordination
- Theoretical analysis shows that…
  - Individual distribution matching against coordinated expert demonstrations improves a lower bound on a joint imitation learning objective, leading to **convergence**
  - **Expert policies are a Nash equilibrium** for mixed task and distribution matching reward
- Experimental validation on StarCraft shows that the combined imitation and task reward **improve on a fully decentralized** baseline

DM²

- Each agent independently learns from a mixed reward that consists of the environment task reward and a distribution matching reward from GAIL [1]



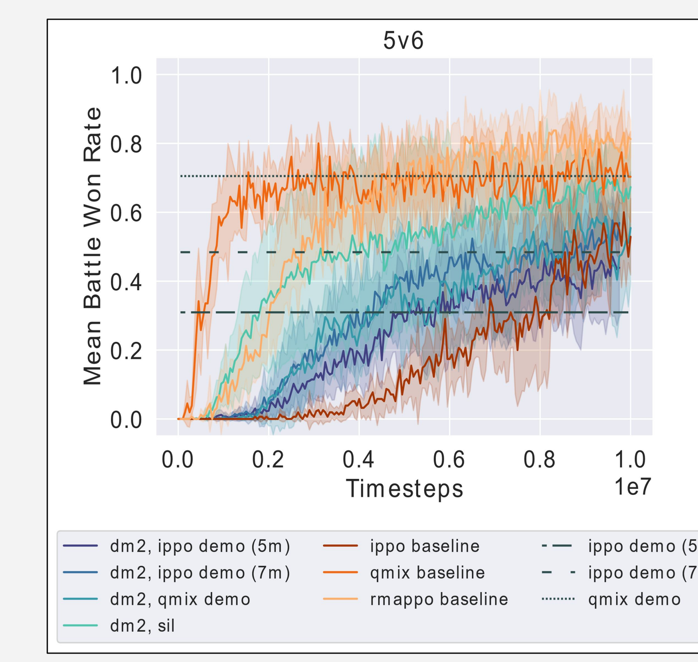
- Agent and expert policies trained with IPPO [2]
- Expert demonstrations (state-only trajectories) are **compatible** (sampled concurrently from **co-trained** experts)

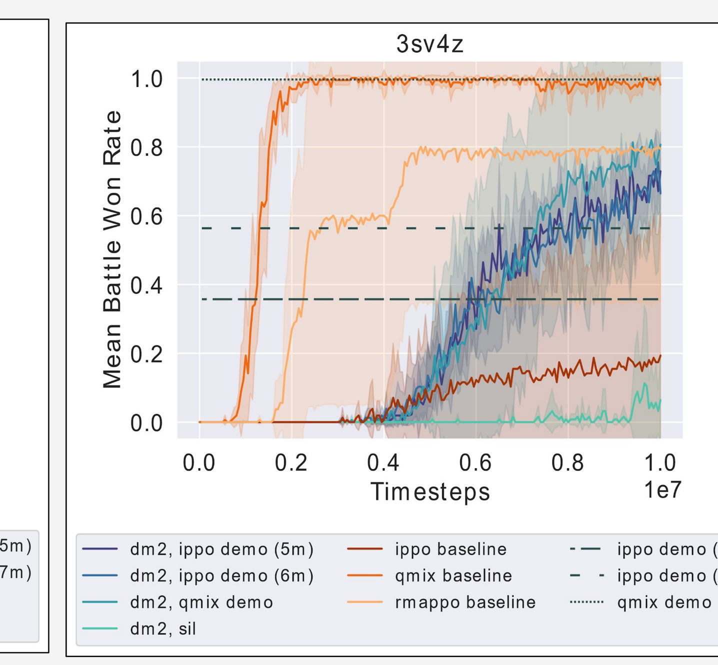
[1] Ho and Ermon, NeurIPS 2016.
[2] Yu et al., ArXiv 2021.

DM² allows a **team** of RL agents to learn a cooperative task by **independently imitating** corresponding demonstrations from an **expert team.**

Experimental Results

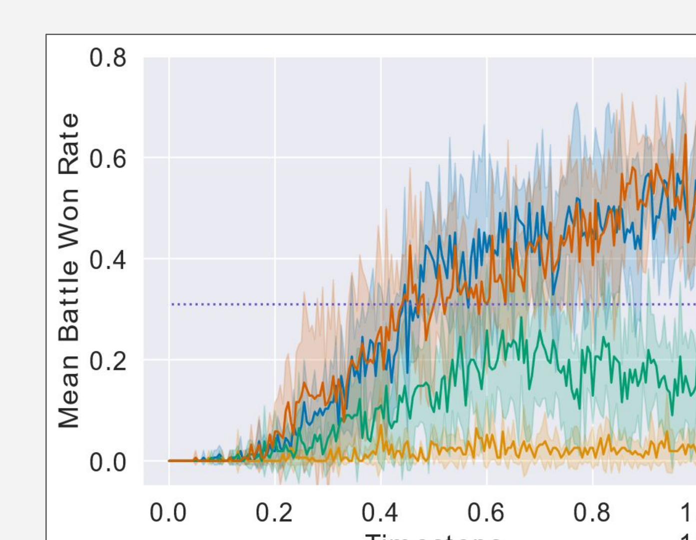
- StarCraft II benchmark [3]
- Expert demonstrations sampled from trained IPPO or QMIX policies
- CTDE Baselines: QMIX [4], RMAPPO [2]
- Distribution Matching Baseline: DM² w/SIL [5]

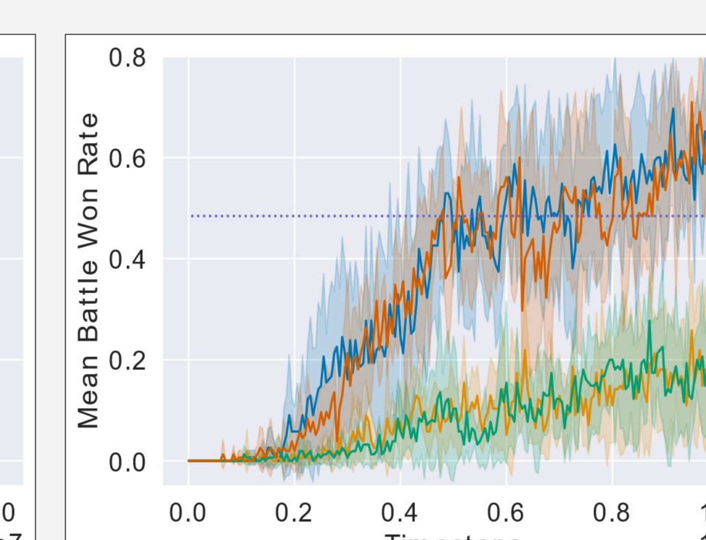




[3] Samvelyan et al., AAMAS 2019.
[4] Rashid et al., ICML 2018.
[5] Oh et al., ICML 2018.

Which Demonstrations Work?





legend	concurrent	nonconcurrent
joint	DM²	✔
not joint	✘	✘