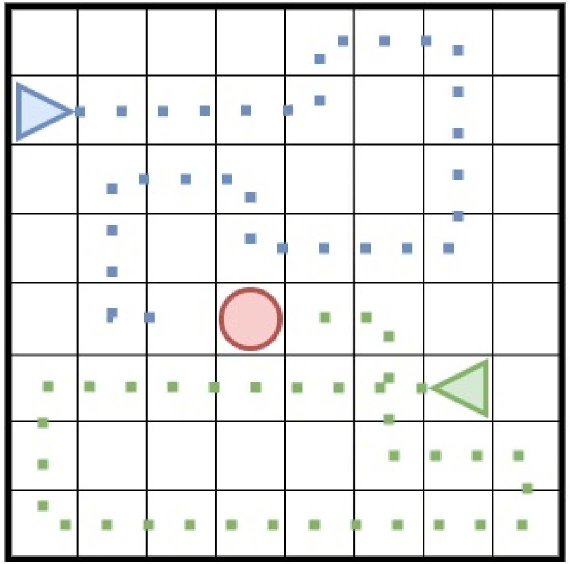
Ensemble Value Functions for Efficient Exploration in Multi-Agent Reinforcement Learning

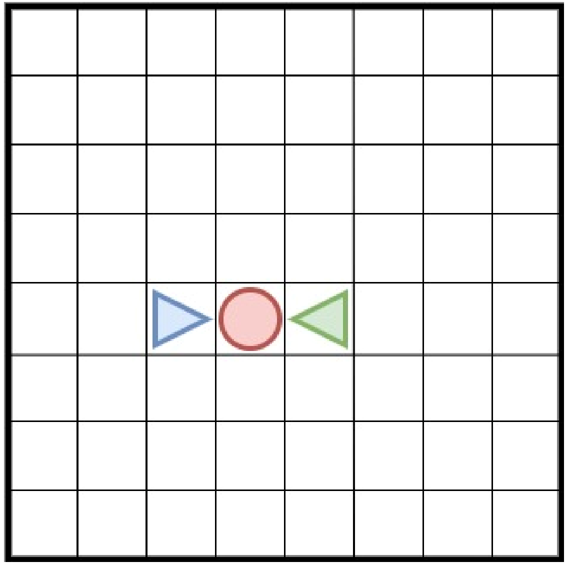
Lukas Schäfer, Oliver Slumbers, Stephen McAleer, Yali Du, Stefano V. Albrecht, David Mguni

International Conference on Autonomous Agents and Multiagent Systems
Adaptive and Learning Agents Workshop

Motivational Problem



Individual exploration of movement

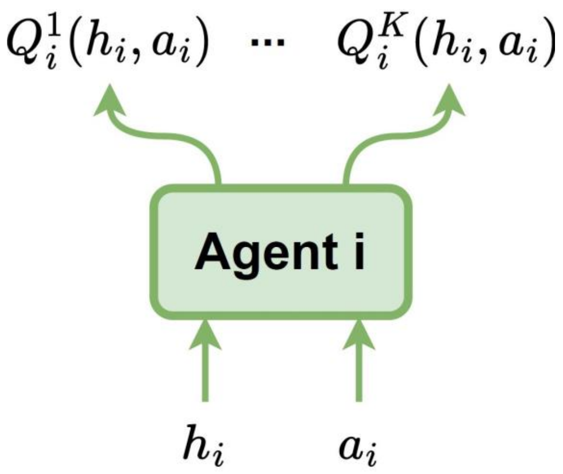


Joint exploration of cooperation

Ensemble Value Functions for Multi-Agent Exploration (EMAX)

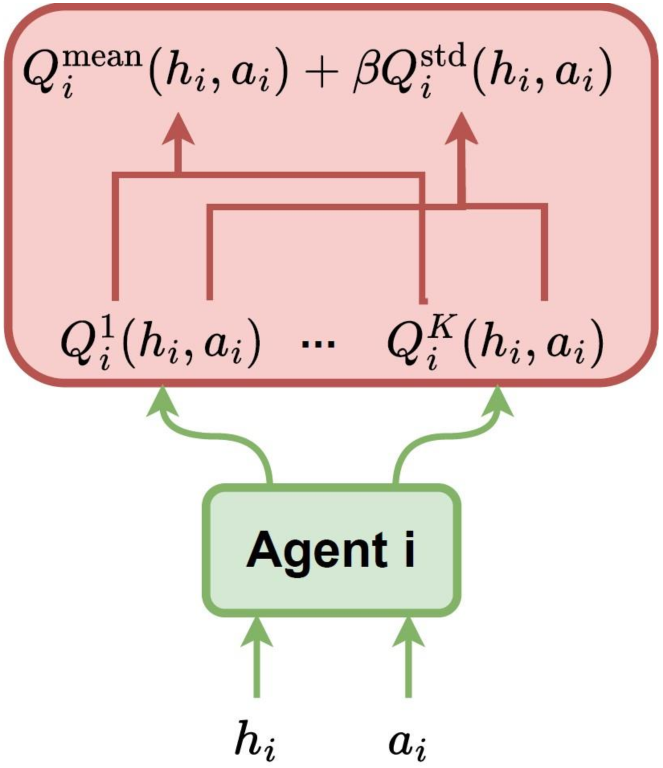
- Plug-and-play approach to extend value-based MARL algorithms

- Each agent trains an ensemble of value functions

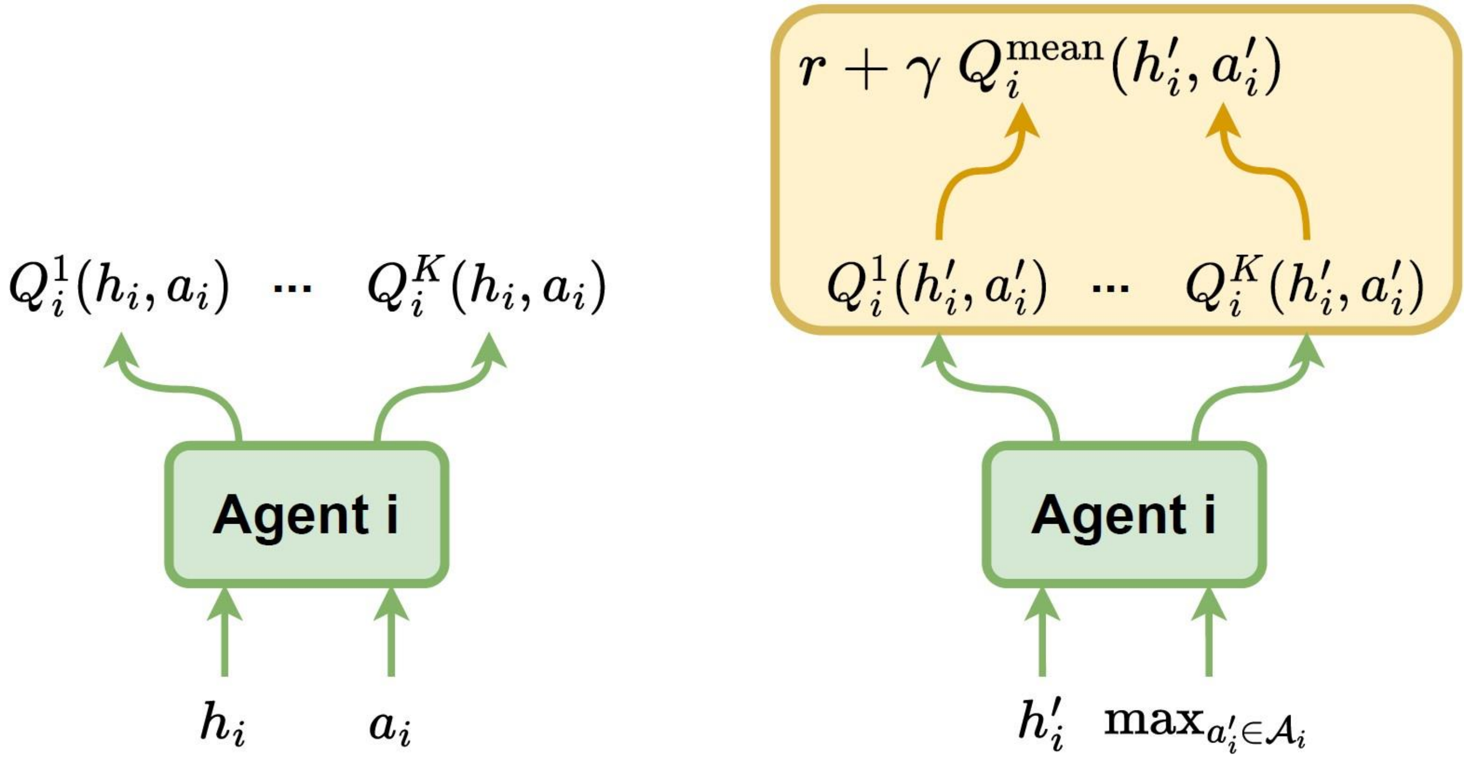


EMAX – Exploration Policy

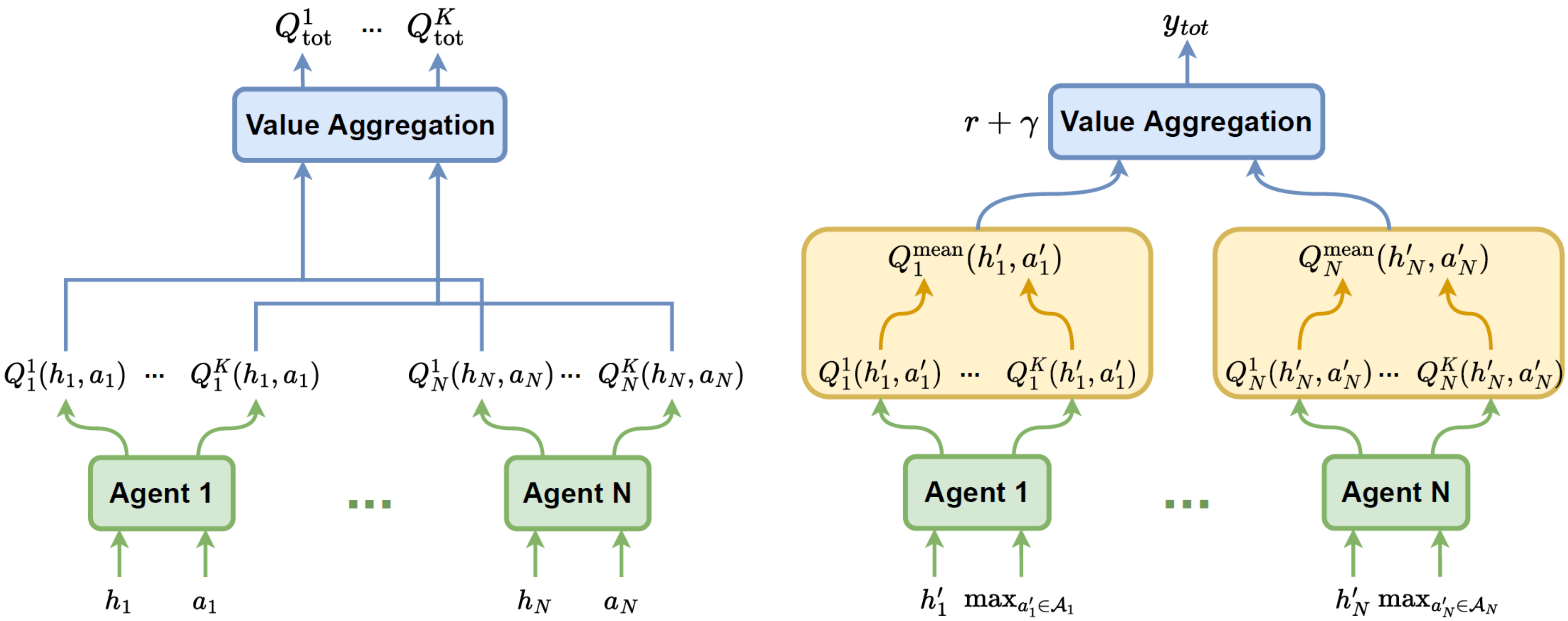
- Disagreement of value estimates is large for states which require coordination
- Use disagreement in UCB exploration policy to guide exploration



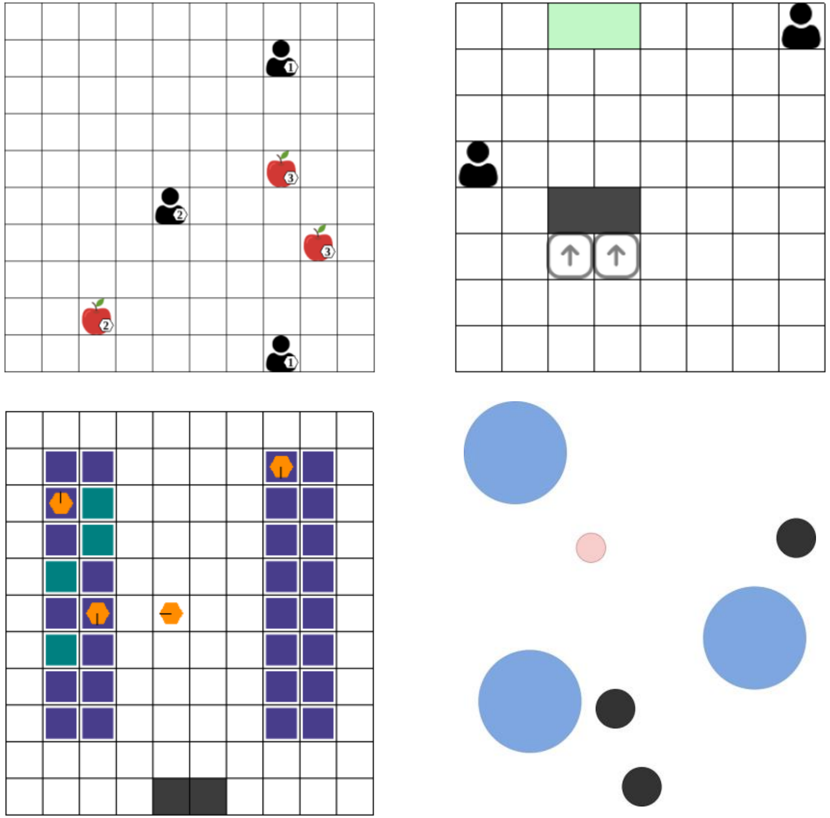
EMAX – Independent Robust Target Estimates



EMAX - Robust Target Estimates with Value Decomposition



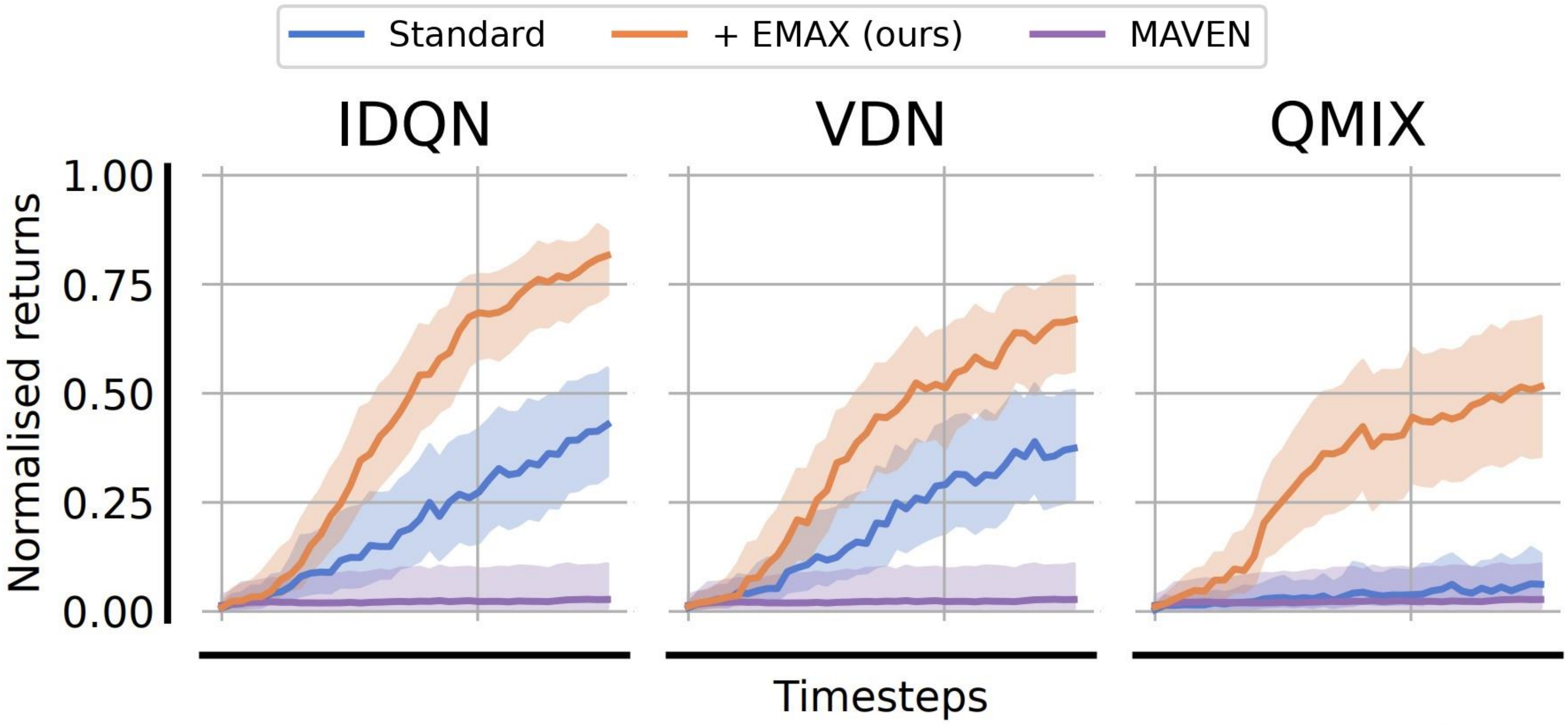
Evaluation with Deep Value-Based MARL Algorithms



MARL Algorithms

- IDQN, VDN, QMIX
- MAVEN (exploration-focused extension of QMIX)
- IDQN, VDN, QMIX + EMAX

Evaluation Results: Aggregated



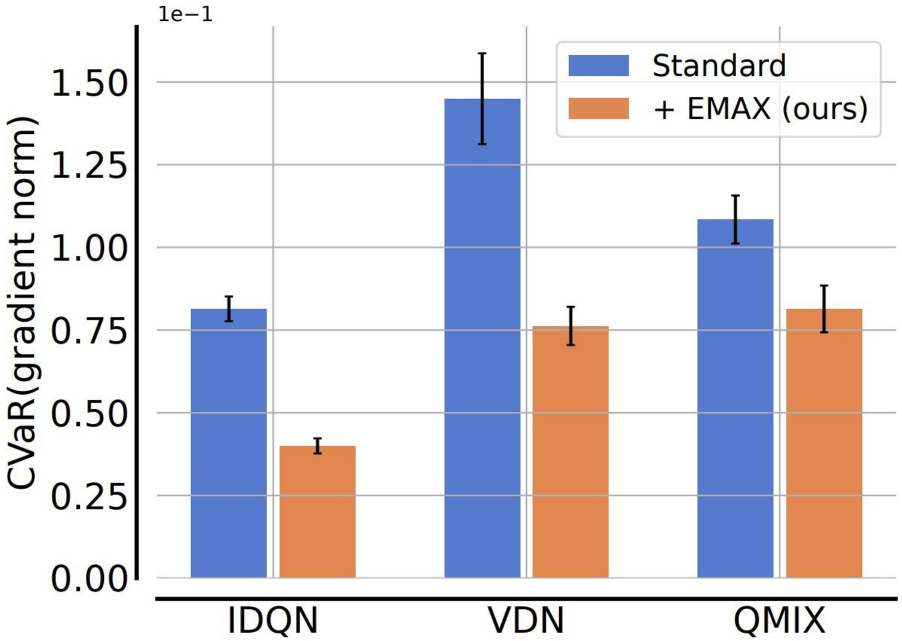
Analysis: Training Stability

Do ensemble target values stabilise the optimisation of trained value functions?

→ Inspect stability of gradients:

$$CVaR(\nabla') = \mathbb{E}[\,\nabla' \mid \nabla' \geq VaR_{95\%}(\nabla')\,]$$

$$\nabla'_t = |\nabla_{t+1}| - |\nabla_t|$$



Ensemble Value Functions for Efficient Exploration in Multi-Agent Reinforcement Learning

https://arxiv.org/abs/2302.03439

Contributions:

1. Train ensembles of value functions to guide exploration using uncertainty of value estimates and compute more robust target estimates
2. EMAX is plug-and-play and can significantly improve training stability and sample efficiency of value-based MARL algorithm