Modular Value Function Factorization in Multi-Agent Reinforcement Learning

Oliver Järnefelt [1] Carlo D'Eramo [1,2]

[1]Technische Universität Darmstadt [2]Hessian.AI

TL;DR

Leverage the **modularity** of the embedded coordination graph by formulating the total utility as a sum of **subteam mixings**

- Each constituent is trained via a **local mixing** function, which needs to only consider a subset of agent utilities to form the total value estimate.
- We suggest finding the **closest disjoint approximation** of non-divisible graphs via graph partitioning
- We evaluate partitioning graphs with a novel **value-based partitioning distance** measure

Value factorization

Decompose joint value function

- Value decomposition network [1]:

$$Q(s, \mathbf{u}) = \sum_{k=1}^{K} Q_k(s, u) \qquad (1)$$

- (Weighted) QMIX, QPLEX [2,3,4]
- QTRAN [5]:

$$Q(s, \mathbf{u}) = \sum_{k \in \mathcal{K}} Q(s, u_k) + V(s) \qquad (2)$$

where
$V(s) = \max Q(s, \mathbf{u}) - \sum_{k=1}^{K} Q_k(s, u_k)$

Modular value function factorization

- We consider problems where:
  - coordination graph structure is **available**
  - mutually independent subpopulations of agents (**modular**)
- *Known Exact Structure Leveraging (KESL)*
  - Leverage the notion of *Disconnected Hyper-Coordination Graph*
  - $\mathcal{V}$: set of agents
  - $\mathcal{E}$: set of hyper-edges
  - $\mathcal{C}_i = \langle \mathcal{V}_i, \mathcal{E}_i \rangle$
  - Disconnectivity property: $\mathcal{V}_i \cap \mathcal{V}_j \neq \emptyset \iff i = j$
  - A disconnected hyper-coordination graph is:

$$\mathcal{H} = \cup_{i=1}^{C} \mathcal{C}_i \qquad (3)$$

  - Mixing agents in each component using QTRAN
  - The total utility is:

$$Q(s, \mathbf{u}) = \sum_{k=1}^{K} Q_k(s, \mathbf{u}) \cdot \mathcal{K}_k \qquad (4)$$

  where $\mathcal{K}_k$ assigns each agent to its own component
- *Known Approximate Structure Leveraging (KASL)*
  - For non-divisible coordination graphs, we find the **closest** partitioning $\mathcal{P}$ and then we use KESL
  - Novel distance measure between graphs

$$d(\mathcal{H}, \mathcal{P}) = \min_{\mathcal{H}' \in \mathcal{P}} \|Q^* - Q^*_{\mathcal{H}'}\|_\infty \qquad (5)$$

  i.e., the minimum estimation error achievable by any joint utility function in the family of factorizations determined by $\mathcal{P}$
  - Distance measure based on **utility functions**!
  - **Upper bound:** Given a hypergraph $\mathcal{H} = \langle \mathcal{V}, \mathcal{E} \rangle$, let $\mathcal{E}_c \subseteq \mathcal{E}$ denote the set of cut hyperedges by partitioning $\mathcal{P}$ that introduces new hyperedges $\mathcal{E}_n = \{e' \mid \exists\, e \in \mathcal{E}_c \text{ s.t. } e' \subset e\}$. Further, given a hyperedge pair $\langle e, e' \rangle$, with $e \in \mathcal{E}_n$ and $e' \in \mathcal{E}_c$, we denote $\mathcal{V}_+ = \{v \mid v \in e',\ e' \subset e\}$ and $\mathcal{V}_- = \{v \mid v \in e',\ e \setminus e'\}$ and $\tilde{q}_{e,+} = \frac{1}{|\mathcal{U}_-|} \sum_{\mathbf{u}'_- \in \mathcal{U}_-} q_e(\cdot, \mathbf{u}'_-)$. Then, the minimum estimation error achievable by any utility function within the family of factorizations determined by $\mathcal{P}$ is upper bounded by (proof omitted for space limitations):

$$\min_{\mathcal{H}' \in \mathcal{P}'} \|Q^*_{\text{jt}} - Q^*_{\text{jt},\mathcal{H}'}\|_\infty \leq \sum_{e \in \mathcal{E}_c} \sum_{e' \in \mathcal{E}_n, e' \subset e} \|q_e - \tilde{q}_{e,+}\|_\infty.$$

Coordination graphs

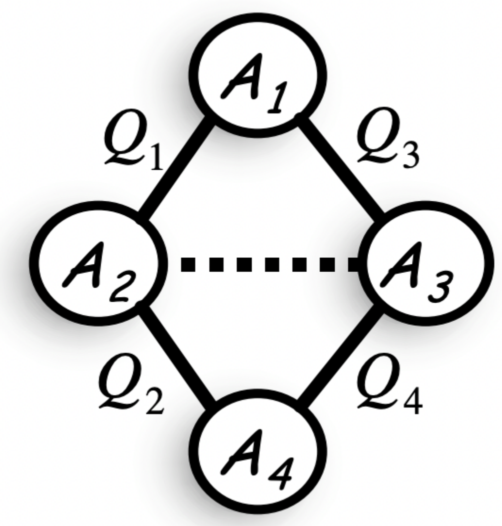


Figure 1. Coordination graph [6]

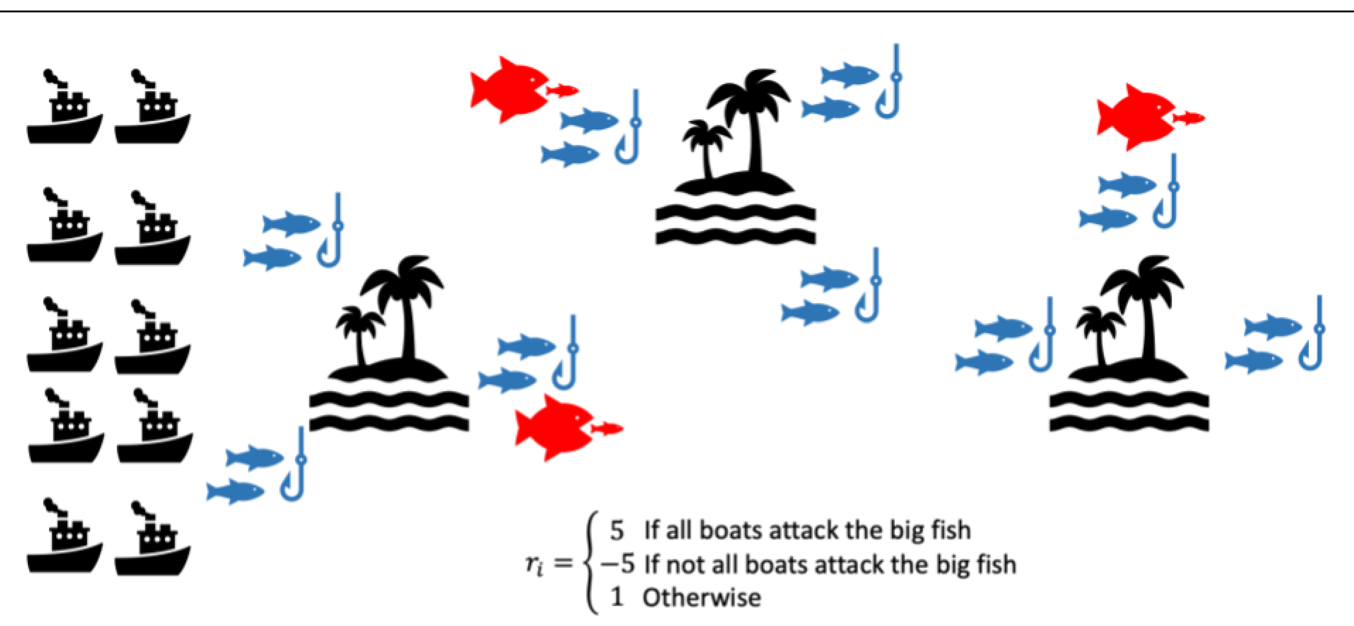
$$Q(s, \mathbf{u}) = Q_1(s, u_1, u_2) + Q_2(s, u_2, u_4) + Q_3(s, u_1, u_3) + Q_4(s, u_3, u_4) \qquad (6)$$

References

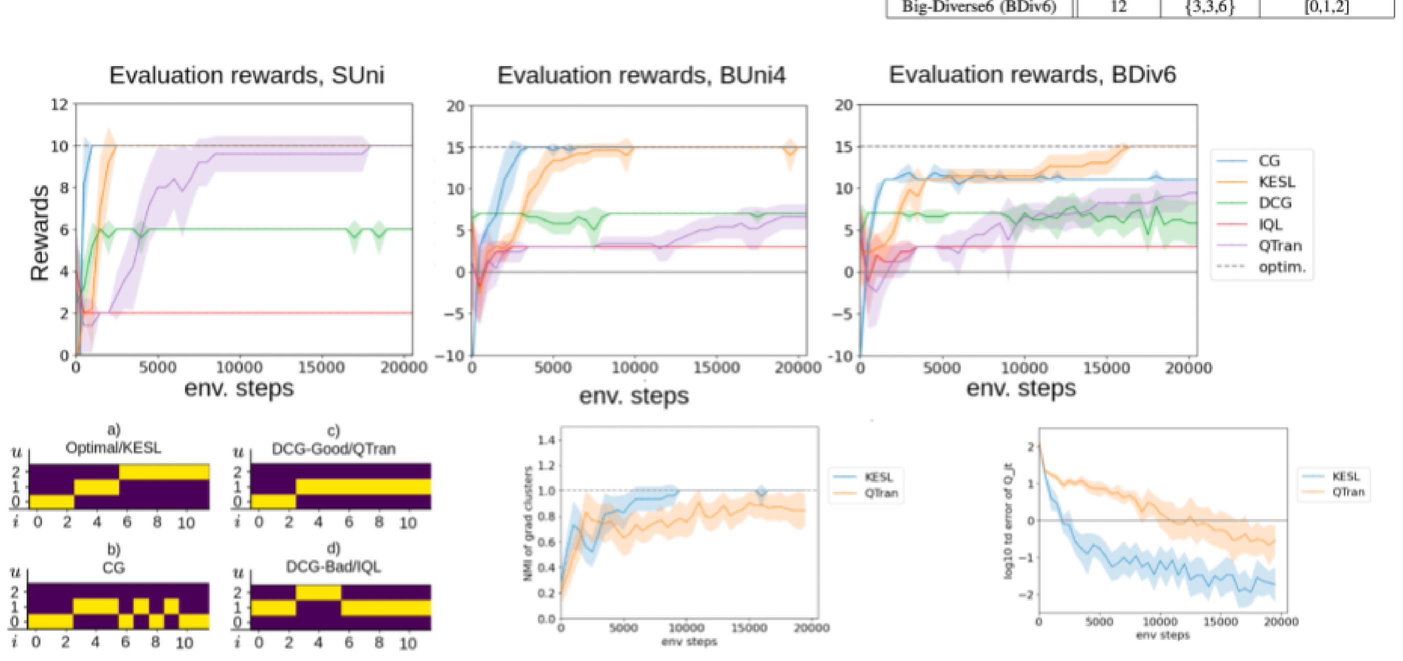
1. Sunehag, Peter, et al. "Value-decomposition networks for cooperative multi-agent learning". arXiv preprint arXiv:1706.05296 (2017).
2. Rashid, Tabish, et al. "Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning". ICML (2018).
3. Rashid, Tabish, et al. "Weighted qmix: Expanding monotonic value function factorisation for deep multi-agent reinforcement learning". NeurIPS (2020).
4. Wang, Jianhao, et al. "Qplex: Duplex dueling multi-agent q-learning". arXiv preprint arXiv:2008.01062 (2020).
5. Son, Kyunghwan, et al. "Qtran: Learning to factorize with transformation for cooperative multi-agent reinforcement learning". ICML (2019).
6. Guestrin, Carlos, Daphne Koller, and Ronald Parr. "Multiagent planning with factored MDPs". NeurIPS (2001).

Empirical results

Deadliest catch



Name	# agents	Fleet sizes	$\bar{u}$ for each fleet
Small-Uniform (SUni)	6	{3,3}	[0,1]
Big-Uniform4 (BUni4)	12	{4,4,4}	[0,1,2]
Big-Diverse6 (BDiv6)	12	{3,3,6}	[0,1,2]



Generalized firefighting