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]:
  $$Q(s, u) = \sum_{k=1}^{K} Q_k(s, u)$$  (1)
- (Weighted) QMIX, QPLEX [2,3,4]
- QTRAN [5]:
  $$Q(s, u) = \sum_{k \in K} Q_k(s, u_k) + V(s)$$  (2)

where

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

We consider problems where:
- coordination graph structure is available
- mutually independent subpopulations of agents (modular)

- Known Exact Structure Leverageing (KESL)
  - Leverage the notion of Disconnected Hyper-Coordination Graph
  - $\mathcal{V}$: set of agents
  - $E^h$: set of hyper-edges
  - $Q_i = (V_i, E^h)$
  - Disconnected property: $V_i \cap V_j = \emptyset \iff i = j$
  - A disconnected hyper-coordination graph is:
    $${\mathcal{H}} = \bigcup_{i=1}^{n} C_i$$  (3)
  - Mixing agents in each component using QTRAN

- Known Approximate Structure Leverageing (KASL)
  - For non-divisible coordination graphs, we find the closest partitioning $P$ and then use KESL
  - Novel distance measure between graphs
    $$d(\mathcal{H}, P) = \min_{P \in \mathcal{P}} \| Q - Q_H \|_{\infty}$$  (5)
  - i.e., the minimum estimation error achievable by any joint utility function in the family of factorizations determined by $P$
  - Distance measure based on utility functions!
  - Upper bound: Given a hypergraph $\mathcal{H} = (V, E)$, let $E^h \subseteq E$ denote the set of cut hyperedges by partitioning $P$ that introduces new hyperedges $E_h^c = \{ e^c | \exists e' \in E, e \neq e' \}$. Further, given a hyperedge pair $(e, e')$ with $e \in E_h$ and $e' \in E$, we define $E_h^c = \{ e' \in E | e' \not\subset e \} = \{ e' \in V | \exists e^c \in E_h^c : e \not\subset e^c \}$ and $\mathcal{V}_e = \{ v | v \in E \}$ and $\mathcal{V}_e^c = \{ v | v \notin E \}$.
  - Then, the minimum estimation error achievable by any utility function within the family of factorizations determined by $P$ is upper bounded by (proof omitted for space limitations):
    $$\min_{\mathcal{P} \in \mathcal{P}} \| Q_H - Q_H \|_{\infty} \leq \sum_{e' \in E} \sum_{e \in E_h^c} \| \mu_e - \mu_e \|_{\infty}$$

Empirical results

Deadliest catch

Generalized firefighting

$$r = \min(\text{Agentens, } 1) + \max(\text{Agentens} - 1, 0)/2$$

References