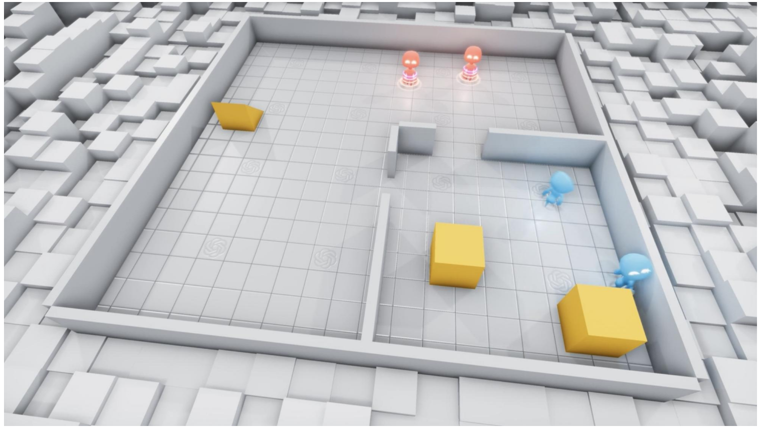
Non-Linear Coordination Graphs

Yipeng Kang*, Tonghan Wang*, Qianlan Yang*, Xiaoran Wu,
Chongjie Zhang

清華大學 交叉信息研究院
Tsinghua University Institute for Interdisciplinary Information Sciences

Multi-Agent Reinforcement Learning



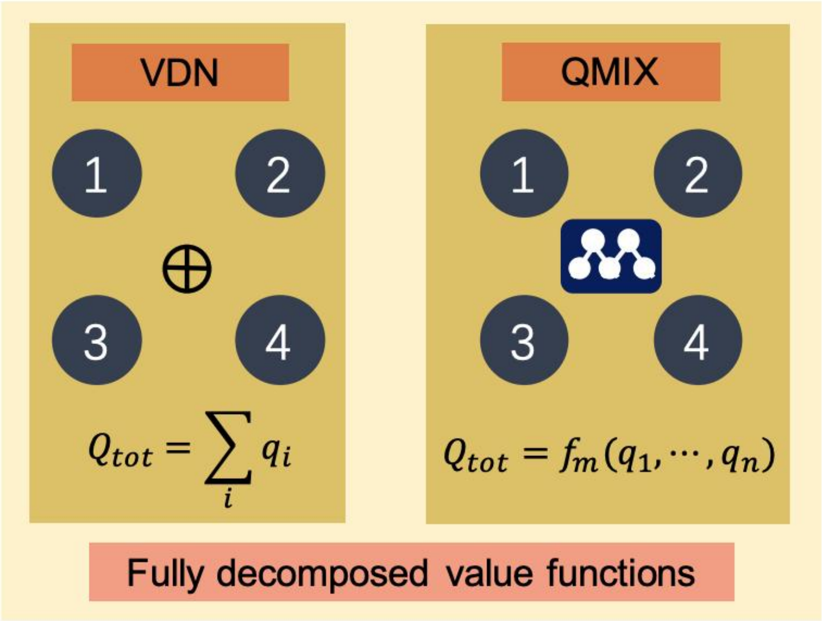
Emergent Tool Use

StarCraft II

One core problem for MARL

- Estimating Q_{tot}
  - Why challenging?
    - Large action-observation space
      - Require high representational capacity for Q-networks
    - Selecting greedy action: $O(A^n)$
      - Exponential complexity: A is the number of action, n is the number of agent

Previous method: fully decomposition



- VDN->QMIX->QPLEX
  - Global maximizer of Q_{tot} can be obtained locally.
- Problem:
  - Miscoordination
  - Relative Overgeneralization

Previous method: coordination graphs

- $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ that represents a higher order decomposition:

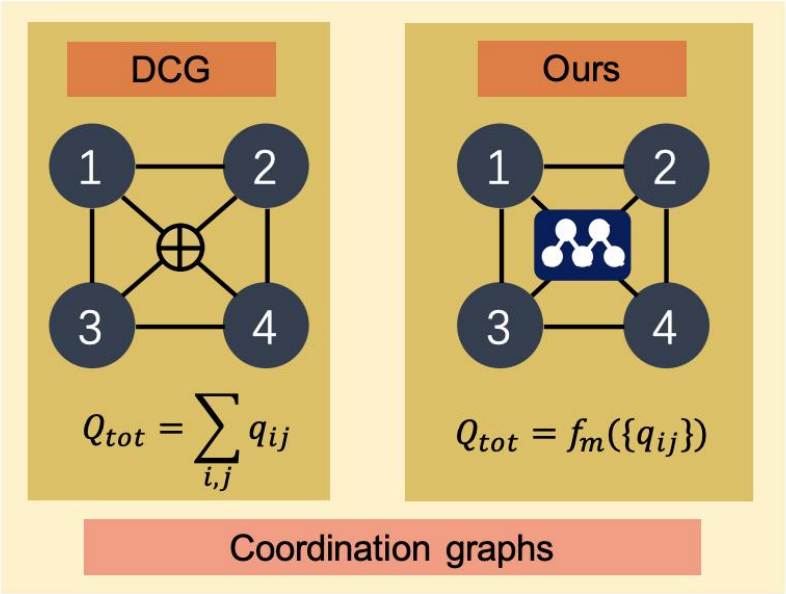
$$Q_{tot}(\boldsymbol{\tau}, \boldsymbol{a}) = \frac{1}{|\mathcal{V}|} \sum_i q_i(\tau_i, a_i) + \frac{1}{|\mathcal{E}|} \sum_{\{i,j\} \in \mathcal{E}} q_{ij}(\boldsymbol{\tau}_{ij}, \boldsymbol{a}_{ij})$$

- Incorporating *pairwise* payoff functions

- Problem:
  - Linear decomposition, limited representational capacity

Our work

- Extends CGs to non-linear value decomposition



Major challenge

- Recall one challenge of multi-agent Q
  - How to select greedy actions?
  - Conventional CGs use Max-Sum (message passing), but is only applicable to linear cases.
  - How to calculate for non-linear CGs?

Our approach

- Mixing network that composes the payoffs as Q_tot: ReLU-series activation functions induce *piece-wise linear* functions.

- A quick idea:
  - Max-Sum on each piece.
  - Problematic:
    - Given a linear region P_i and the piece ρ_i, run Max-Sum may get a solution located in $R_{j \neq i}$. (***The shifted solutions***)

How to solve this problem?

- The maximum of all local solutions.
  - Why does this work?
  - We first show that a shifted solution cannot be optimal

Lemma 1. *Denote affine function pieces and their cells of a fully-connected feedforward mixing network with LeakyReLU activation as* $\mathcal{P}_{all} = \{\rho_j\}_1^{2^m}$ *and* $\{P_j\}_1^{2^m}$. *For* $\mathbf{q}$ *in the cell of the* r*th piece,* P_r*, and* $\forall \rho_s \in \mathcal{P}_{all}$*, we have* $\rho_r(\mathbf{q}) \geq \rho_s(\mathbf{q})$.

This means

- Global optimal solutions do not have this problem.
- Moreover,

$$\max_{\boldsymbol{q}} f_m(\boldsymbol{q}) = \max_{\boldsymbol{q}} \max_{\rho} \rho(\boldsymbol{q}) = \max_{\rho} \max_{\boldsymbol{q}} \rho(\boldsymbol{q})$$

- Indicating that the maximum of local optima is the global optimum.

How many pieces need enumerating?

- Width of the hidden layer: m
  - When m is small: 2^m

    Enumerating slope configuration
  - When m is large:

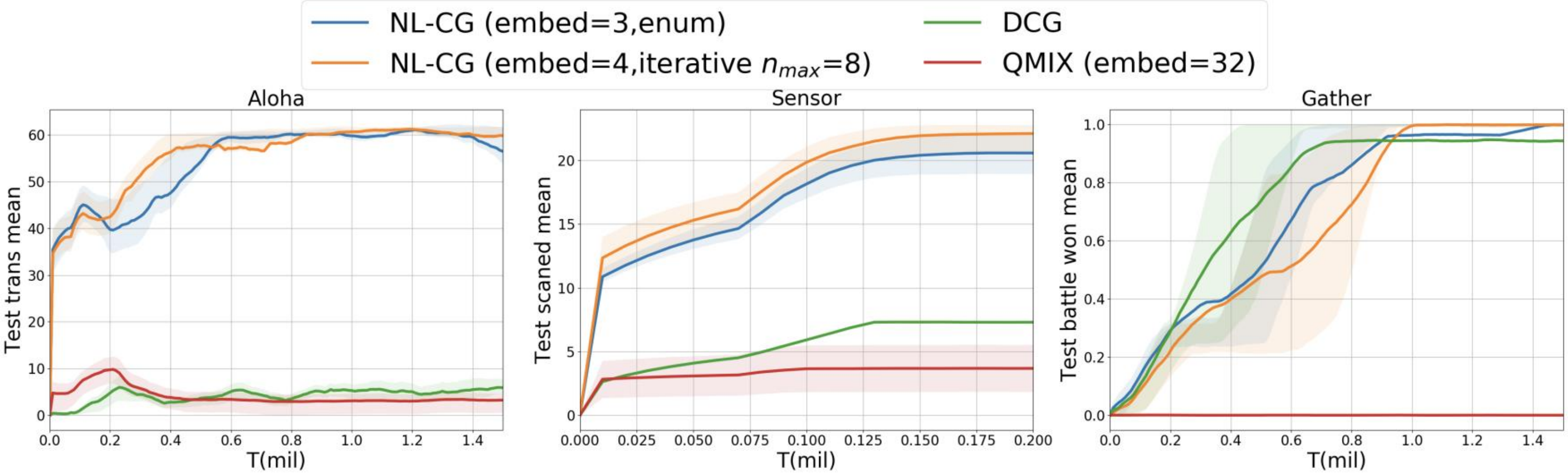
$$n_{m,d} = \sum_{i=0}^{d} \binom{m}{d-i}$$

How to reduce time complexity?

- Based on Lemma 1, we give an iterative method
  - Step 1: Randomly select a piece ρ_j
  - Step 2: Run Max-Sum, get a solution x_j
  - Step 3: Calculate the real piece ρ_{real}
  - Step 4: If $\rho_{real} = \rho_j$, return x_j; Otherwise, move to ρ_{real} and go to Step 2.

- This algorithm guarantees a local optimum:
  - Monotonically increasing & finite inputs

Performance on the MACO benchmark



Thanks for your listening

Machine Intelligence Group

清華大學 交叉信息研究院

Tsinghua University Institute for Interdisciplinary Information Sciences