QPLEX: Duplex Dueling Multi-Agent Q-Learning

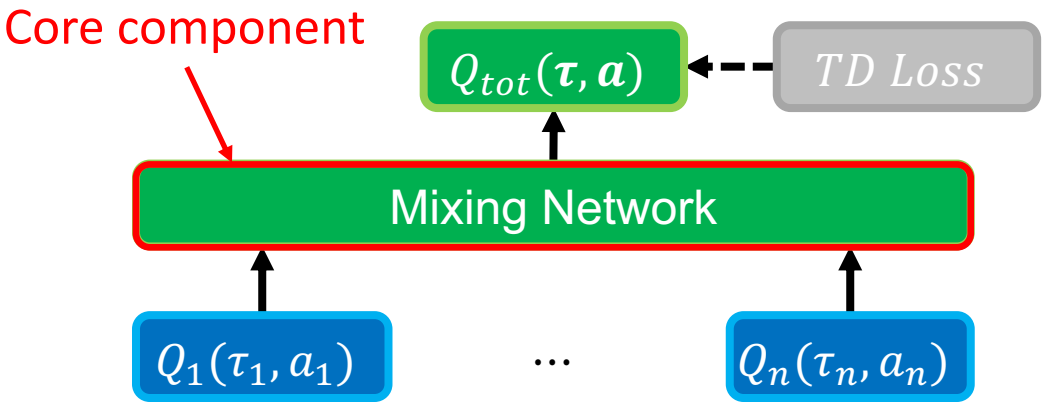
Jianhao Wang*, Zhizhou Ren*, Terry Liu, Yang Yu, Chongjie Zhang

Tsinghua University; Polixir Technologies

*equal contribution

Value Function Factorization Methods

- Paradigm: centralized training with decentralized execution

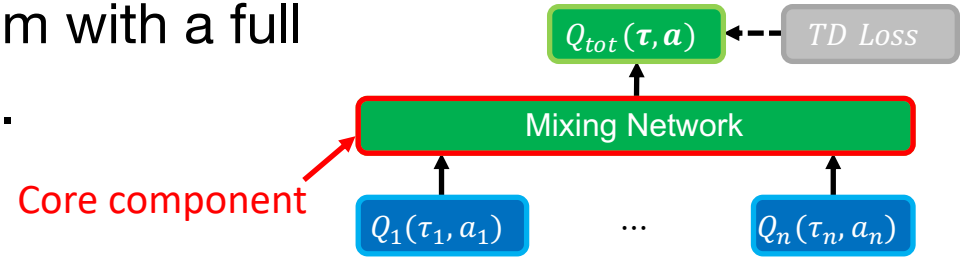


$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}\left[\left(r + \gamma V(\boldsymbol{\tau}'; \boldsymbol{\theta}^-) - Q(\boldsymbol{\tau}, \boldsymbol{a}; \boldsymbol{\theta})\right)^2\right]$$

$$V(\boldsymbol{\tau}'; \boldsymbol{\theta}^-) = \max_{\boldsymbol{a}'} Q(\boldsymbol{\tau}', \boldsymbol{a}'; \boldsymbol{\theta}^-)$$

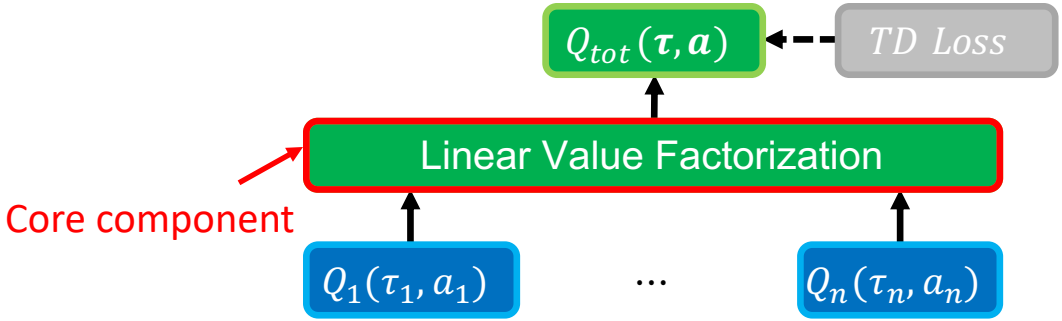
Individual-Global-Maximization (IGM) Principle

- Paradigm: centralized training with decentralized execution
- Individual-Global Maximization (IGM) Principle
  - Consistent greedy action selection between joint and individuals
  - $\underset{\boldsymbol{a}}{\operatorname{argmax}}\, Q_{tot}(\boldsymbol{\tau}, \boldsymbol{a}) = \left(\operatorname{argmax}_{a_1} Q_1(\tau_1, a_1), \ldots, \operatorname{argmax}_{a_n} Q_n(\tau_n, a_n)\right)$
  - Open problem: How to design an effective and scalable multi-agent Q-learning algorithm with a full realization of the IGM principle.



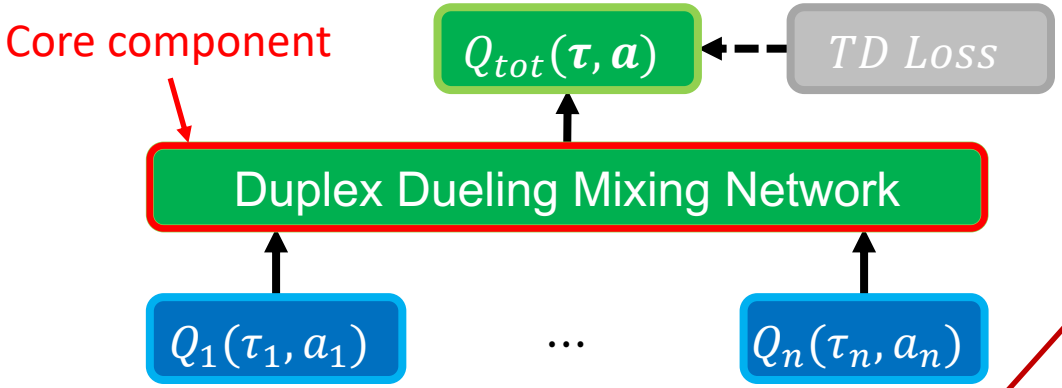
Previous Work: Value Decomposition Networks (VDN)

- VDN: $Q_{tot}(\boldsymbol{\tau}, \boldsymbol{a}) = \sum_i Q_i(\tau_i, a_i)$
- Sufficient for IGM constraint
  - $\underset{\boldsymbol{a}}{\text{argmax}}\, Q_{tot}(\boldsymbol{\tau}, \boldsymbol{a}) = \left(\text{argmax}_{a_1} Q_1(\tau_1, a_1), \ldots, \text{argmax}_{a_n} Q_n(\tau_n, a_n)\right)$
- Excellent scalability
- Cons:
  - Not necessary for IGM
  - Limited expressiveness



QPLEX: Duplex Dueling Multi-Agent Q-Learning

- Paradigm: centralized training with decentralized execution

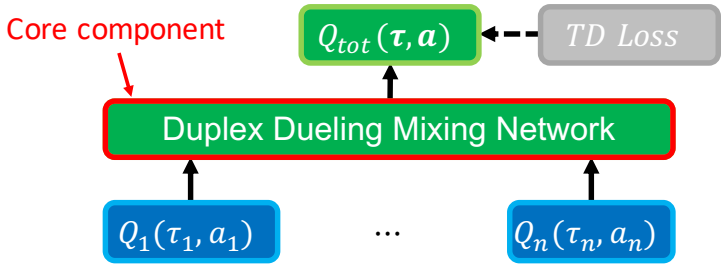


Linear Value Factorization

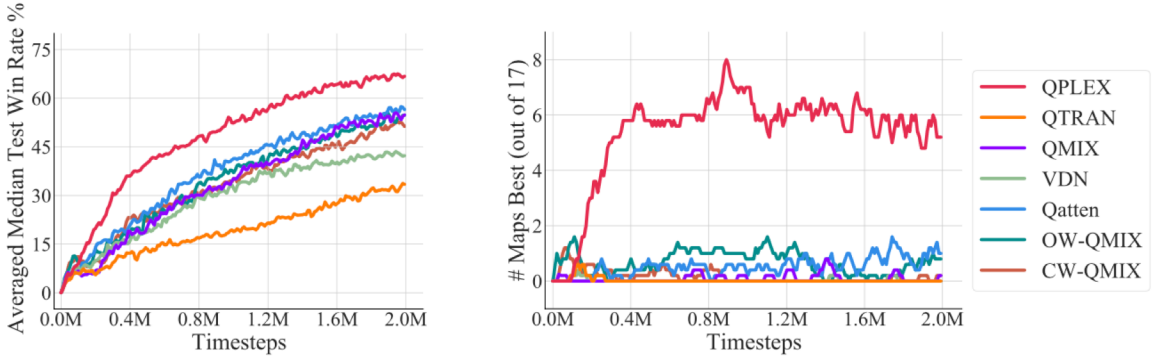
- QPLEX : $Q_{tot}(\boldsymbol{\tau}, \boldsymbol{a}) = \sum_i Q_i(\boldsymbol{\tau}, a_i) + \sum_{i=1}^{n} (\lambda_i(\boldsymbol{\tau}, \boldsymbol{a}) - 1)\, A_i(\boldsymbol{\tau}, a_i)$
  - Strong representation capacity
  - Easily realized and learned by neural networks

QPLEX: Duplex Dueling Multi-Agent Q-Learning

Theorem: *The joint action-value function class that QPLEX can realize is equivalent to what is induced by the IGM principle.*



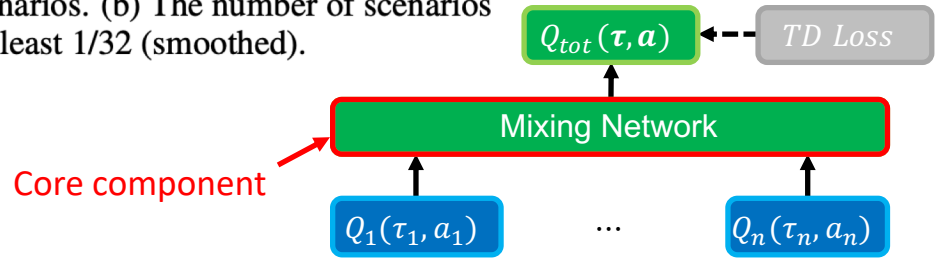
StarCraft II Benchmark: Online Learning



(a) Averaged test win rate

(b) # Maps best out of 17 scenarios

Figure 4: (a) The median test win %, averaged across all 17 scenarios. (b) The number of scenarios in which the algorithms' median test win % is the highest by at least 1/32 (smoothed).



Thanks for your listening