Convergence Rates for Localized Actor-Critic in Networked Markov Potential Games

Zhaoyi Zhou[1], Zaiwei Chen[2], Yiheng Lin[2], Adam Wierman[2]

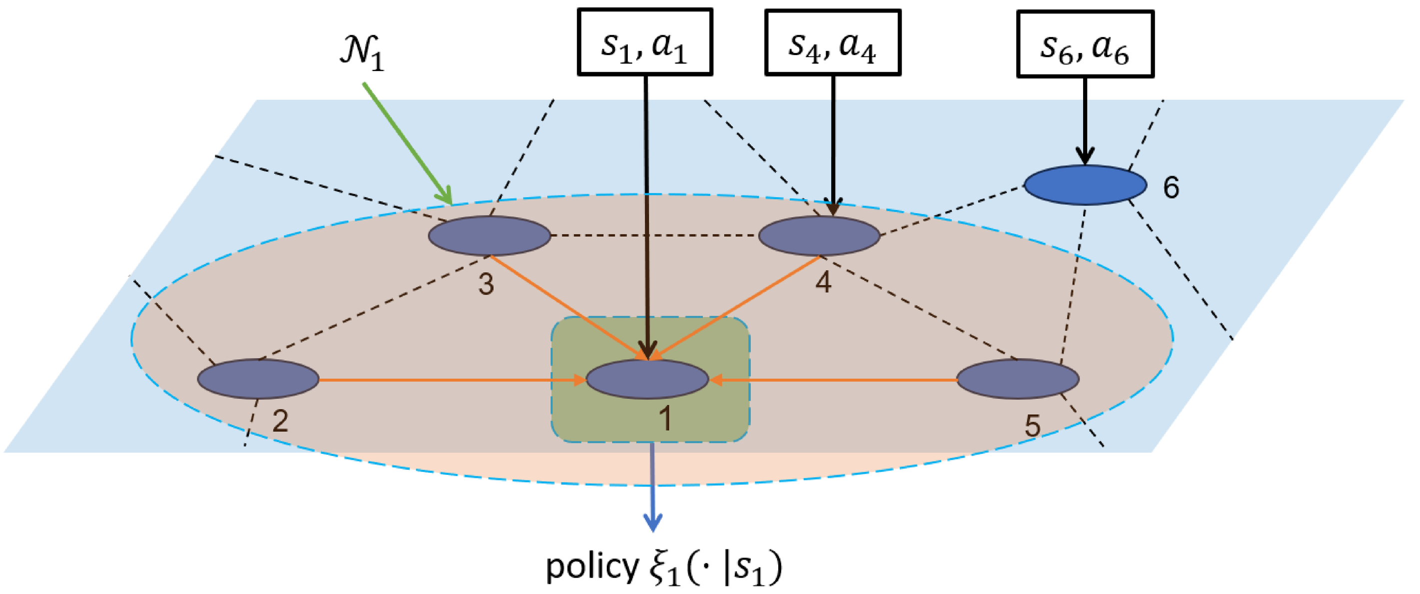
[1] Tsinghua University [2] California Institute of Technology

39th Conference on Uncertainty in Artificial Intelligence

Setting

We consider n agents associated with an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ ("communication network"). Each agent i has its local state $s_i \in \mathcal{S}_i$ and local action $a_i \in \mathcal{A}_i$. Global state/action space can be decomposed as

$$\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_n, \text{ and } \mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times \cdots \times \mathcal{A}_n.$$



The next local state depends on local action and neighboring local states:

$$\mathcal{P}(s(t+1) \mid s(t), a(t)) = \prod_{i=1}^{n} \mathcal{P}_i(s_i(t+1) \mid s_{\mathcal{N}_i}(t), a_i(t)).$$

Local reward depends on states and actions of agents within κ_r-graph distance, i.e., $r_i(s, a) = r_i(s_{\mathcal{N}_i^{\kappa_r}}, a_{\mathcal{N}_i^{\kappa_r}}), \forall i \in \mathcal{N}$.

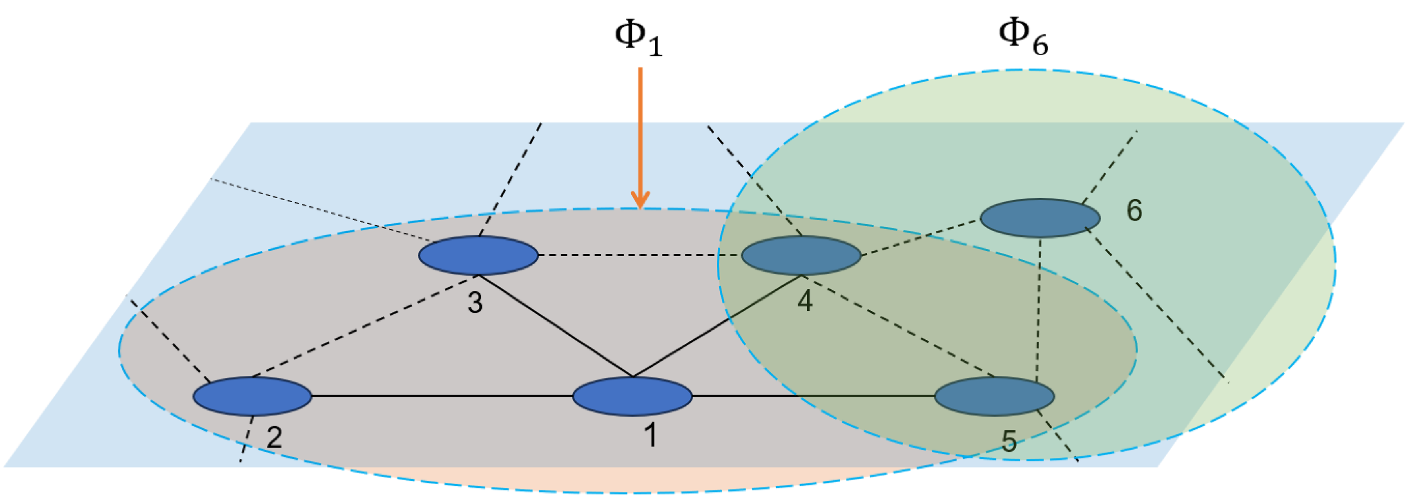
Each agent i adopts a localized policy $\xi_i(\cdot|s_i)$, which maps local state space $\mathcal{S}_i$ to $\Delta(\mathcal{A}_i)$ using softmax parameterization. Every agent i tries to maximize its own expected γ-discounted return, denoted by $J_i(\xi)$.

Networked Markov Potential Game (NMPG)

Relax MPG by restricting the impact of potential to "nearby" agents.

Def 1 (NMPG). *A multi-agent Markov game is called a κ_G-NMPG if there exists a set of* ***local*** *potential functions $\{\Phi_i\}_{i \in \mathcal{N}}$, such that Φ_i tracks policy deviation of any agent j* ***within κ_G -graph-distance of i***:

$$J_j(\xi_j', \xi_{-j}) - J_j(\xi_j, \xi_{-j}) = \Phi_i(\xi_j', \xi_{-j}) - \Phi_i(\xi_j, \xi_{-j}), \forall i, \forall j \in N_i^{\kappa_G}, \forall \xi_j, \xi_{-j}, \xi_j'.$$



Our Contributions:

(1) **Networked** Markov potential games;

(2) **Localized** Actor-Critic;

(3) **Finite-Sample** Analysis.

Take a picture to download the full paper

Localized Actor-Critic

Hyperparameter κ_c controls trade-off between communication distance and truncation accuracy.

Algorithm 1 Localized Actor-Critic

Let $\theta_i(0) = 0$, $\Delta_i^0(m) = 0$ for all i and m. Only κ_c-hop neighborhood
for $m = 0, 1, 2, \cdots, M-1$ **do**
Critic
  Construct ϵ-exploration policy $\hat{\xi}$ from $\xi^{\theta(m)}$.
  Use $\hat{\xi}$ to collect samples for K iterations;
  Each agent i records $\boxed{\{(s_{N_i^{\kappa_c}}(k), a_i(k), r_i(k))\}}_{0 \le k \le K-1}$.
  Each agent i estimates κ_c-**truncated** averaged Q-function by **linear function approximation**, with weight vector w_i^m.
Actor
  for $t = 0, 1, \cdots, T-1$ **do**
    Use $\xi^{\theta(m)}$ to collect samples for H iterations;
    Each agent i records $\boxed{\{(s^t_{N_i^{\kappa_c}}(k), a_i^t(k)))\}}_{0 \le k \le H-1}$.
    Every agent i uses its own record and w_i^m to obtain $\Delta_i^T(m) \approx \nabla_i J_i(\theta(m))$.
    Independent Policy Gradient: Every agent i performs $\theta_i(m+1) = \theta_i(m) + \beta \Delta_i^T(m)$.
  end for
end for

Truncated averaged Q-function:

- Can be approximated using information in κ_c-hop neighborhood.
- Truncation error decays exponentially with κ_c.

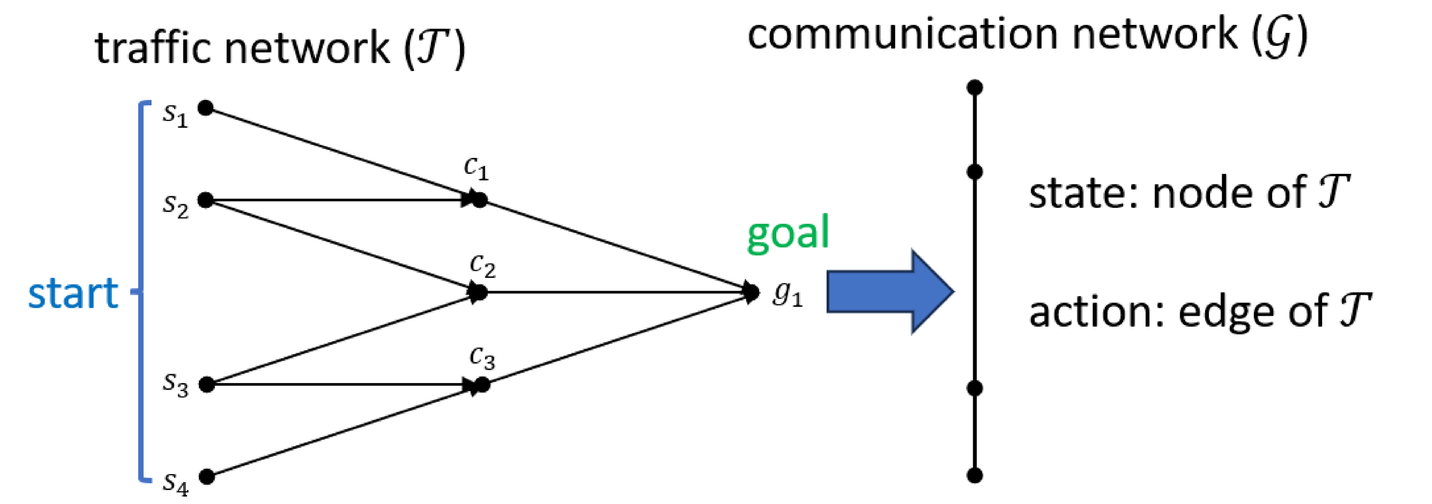
Averaged Q-function		κ_c-truncated averaged Q-function		κ_c-truncated averaged Q-function with linear function approximation
$\overline{Q}_i(s, a_i)$	$\Rightarrow$	$\overline{Q}_i^{\kappa_c}(s_{N_i^{\kappa_c}}, a_i)$	$\Rightarrow$	$\hat{Q}_i^{\kappa_c}(s_{N_i^{\kappa_c}}, a_i; w_i)$
	Use κ_c-hop information		Further reduce parameter dimension	

Main Results

We prove that Localized Actor-Critic has $\tilde{\mathcal{O}}(\bar{\epsilon}^{-4})$ sample complexity.

Thm 1 (Informal). *With $\tilde{\mathcal{O}}(\bar{\epsilon}^{-4})$ samples, the expectation of averaged Nash regret is upper bounded by $\bar{\epsilon}$ plus following additional errors:*
1) Localization error, decaying exponentially with hyperparameter κ_c;
2) Exploration error, depending on exploration coefficient ϵ;
3) Function approximation error.

Example: Markov Congestion Game



Each agent tries to commute from its start node to its destination. Two agents share an edge in the communication network iff they can reach the same node in traffic network.