Finite-Time Analysis of Distributed TD(0) On Multi-Agent Systems

Thinh T. Doan

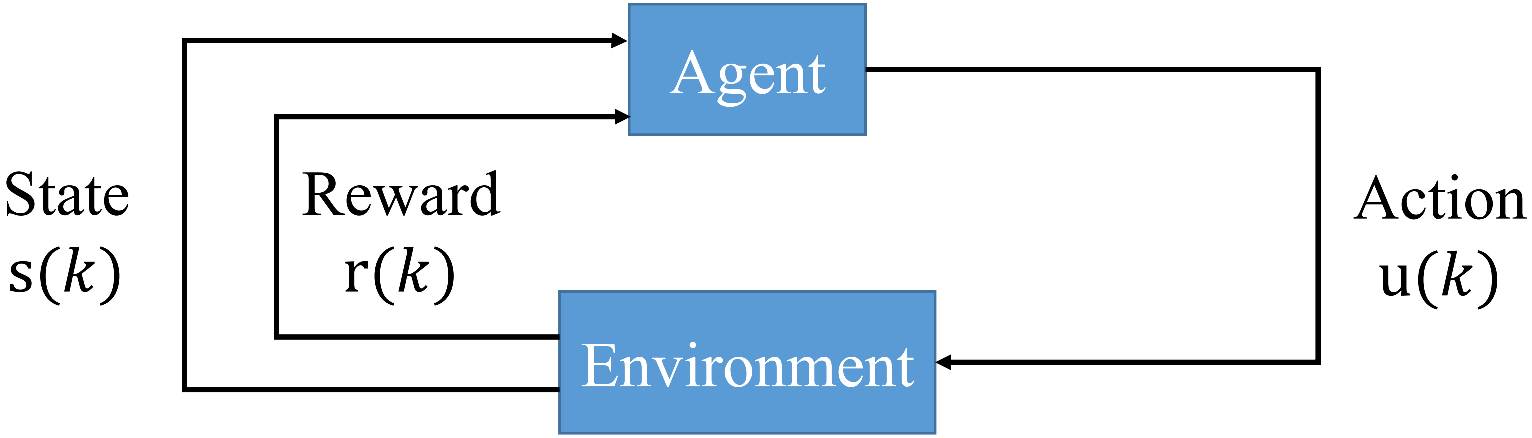
School of Electrical and Computer Engineering

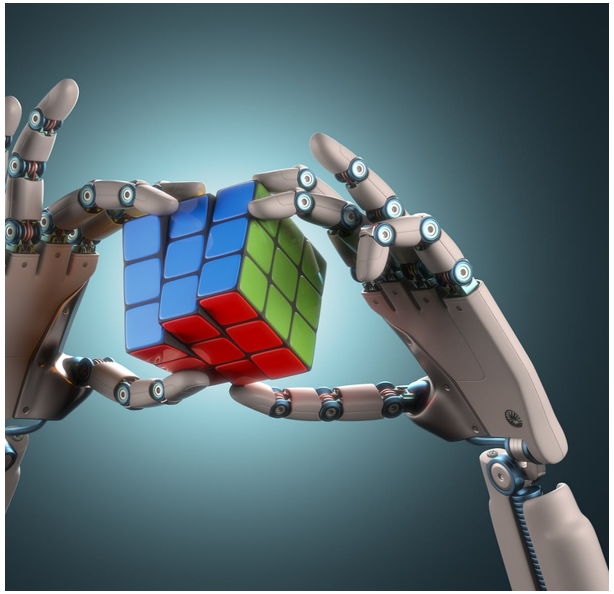
School of Industrial and System Engineering

Joint work with S. T. Maguluri and J. Romberg – GaTech

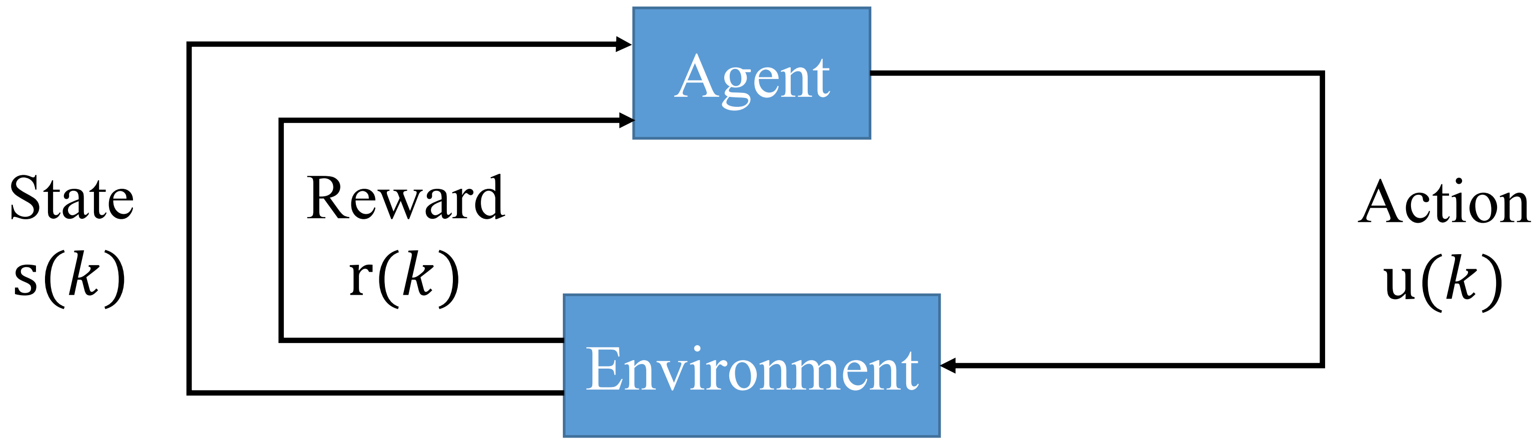
ICML, Long Beach, CA, 2019

Reinforcement Learning





Policy Evaluation Problems



- **Problem**: given a stationary policy μ, find discounted accumulative reward $J^* : S \to \mathbb{R}$

$$J^*(i) \triangleq \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k \mathcal{R}((s(k), s'(k)) \,|\, s(0) = i\right],$$

where $(s(k), r(k), s'(k))$ is the observation at time k

- Bellman equation

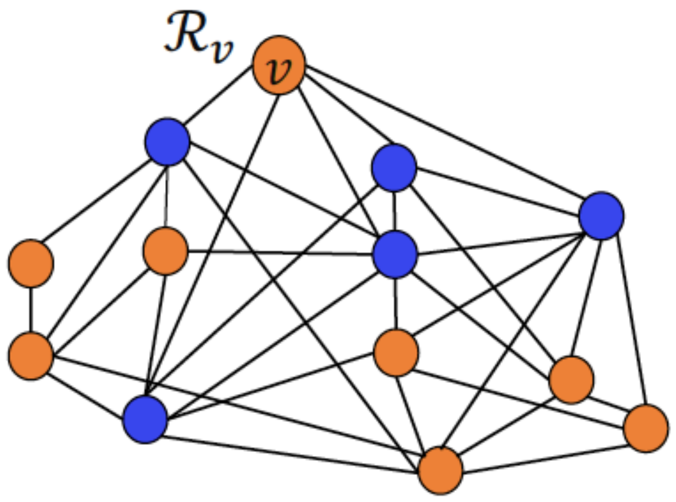
$$J^*(i) \triangleq \sum_{j=1}^{n} p_{ij}\left[\mathcal{R}(i,j) + \gamma J^*(j)\right]$$

Multi-Agent Systems

Multi-Agent Reinforcement Learning

- A network of N agents communicating through $\mathcal{G}(k) = (\mathcal{V}, \mathcal{E}(k))$
- Agent v applies a stationary policy $\mu_v : S \rightarrow U_v$
- Agent v receives an instantaneous reward $\mathcal{R}_v$
- **Goal**: collaborate to find discounted reward $J^* : S \rightarrow \mathbb{R}$

$$J^*(i) \triangleq \mathbb{E}\left[\frac{1}{N}\sum_{k=0}^{\infty}\gamma^k \sum_{v\in\mathcal{V}} \mathcal{R}_v(s(k), s'(k)) \mid s(0) = i\right]$$



Linear Function Approximation

- Low-dimensional approximation $\tilde{J}$ of J^*, parameterized by $\theta \in \mathbb{R}^K$

$$\tilde{J}(i,\theta) = \sum_{\ell=1}^{K} \theta_\ell \phi_\ell(i)$$

where $\phi_\ell : \mathcal{S} \to \mathbb{R}$ is the set of $K \ll n$ feature vectors known by the nodes

- **Goal**: each node finds θ^* s.t. $\tilde{J}$ is the best approximation of J^*

Distributed Temporal Difference Learning

- Low-dimensional approximation $\tilde{J}$ of J^*, parameterized by $\theta \in \mathbb{R}^K$

$$\tilde{J}(i, \theta) = \sum_{\ell=1}^{K} \theta_\ell \phi_\ell(i)$$

where $\phi_\ell : \mathcal{S} \to \mathbb{R}$ is the set of $K \ll n$ feature vectors known by the nodes

- **Goal**: each node finds θ^* s.t. $\tilde{J}$ is the best approximation of J^*

- **Distributed TD(0)**: each node maintains θ_v, an estimate of θ^*

- Observe one sample $(s(k), r_v(k), s'(k))$ and sequentially update θ_v

$$\theta_v(k+1) = \sum_{u \in \mathcal{N}_v(k)} a_{vu}(k)\theta_u(k) + \alpha(k) d_v(k) \phi(s(k))$$

where the TD $d_v(k)$ is given

$$d_v(k) = r_v(k) + \theta_v(k)^T \Big(\gamma\phi(s'(k)) - \phi(s(k))\Big)$$

Main Results

Mirror the results of distributed SGD for solving convex optimization

Finite-Time Analysis

1. If $\alpha(k) = \alpha > 0$ then

$$\left\|\tilde{J}(\hat{\theta}_v(k)) - \tilde{J}(\theta^*)\right\|_D^2 \leq \frac{C_0}{1-\gamma}\frac{1}{k+1} + \frac{C_1\alpha}{1-\gamma}$$

2. If $\alpha(k) = \frac{1}{\sqrt{k+1}}$ then

$$\left\|\tilde{J}(\hat{\theta}_v(k)) - \tilde{J}(\theta^*)\right\|_D^2 \leq \mathcal{O}\left(\frac{C_2 \ln(k+1)}{(1-\gamma)\sqrt{k+1}}\right)$$

$$||J||_D^2 = J^T D J$$

$$\hat{\theta}_v(k) = \frac{\sum_{t=0}^{k} \alpha(t)\theta_v(t)}{\sum_{t=0}^{k} \alpha(t)}$$

Finite-Time Analysis

- Recall that θ^* satisfies

$$A\theta^* = \frac{1}{N}\sum_{v\in\mathcal{V}} b_v$$

  where

$$A = \mathbb{E}_\pi\left[\phi(s)\Big(\phi(s) - \gamma\phi(s')\Big)^T\right] \quad \text{and} \quad b_v = \mathbb{E}_\pi\left[r_v\phi(s)\right]$$

- Let $\sigma_{\min}$ and σ_{max} be the smallest and largest singular value of A

Finite-Time Analysis

1. If $\alpha(k) = \alpha \in \left(0, \frac{1}{\sigma_{min}}\right)$ then

$$\mathbb{E}\left[\|\hat{\theta}_v(k) - \theta^*\|^2\right] \leq C_3\rho^k + \frac{C_4\sigma_{\max}\alpha}{1-\rho}$$

2. If $\alpha(k) = \frac{\alpha_0}{k+1}$ where $\alpha_0 > \frac{1}{\sigma_{min}}$ then

$$\mathbb{E}\left[\|\hat{\theta}_v(k) - \theta^*\|^2\right] \leq \mathcal{O}\left(\frac{C_5\sigma_{\max}}{\sigma_{\min}(1-\delta)}\frac{\ln(k+1)}{k+1}\right)$$

$$\rho = \max\{1 - \sigma_{\min}\alpha\,,\ \delta\} \in (0,1)$$

$$\hat{\theta}_v(k) = \frac{\sum_{t=0}^{k}\alpha(t)\theta_v(t)}{\sum_{t=0}^{k}\alpha(t)}$$

Georgia Tech

Thinh T. Doan

thinhdoan@gatech.edu

Thank you for your attention!