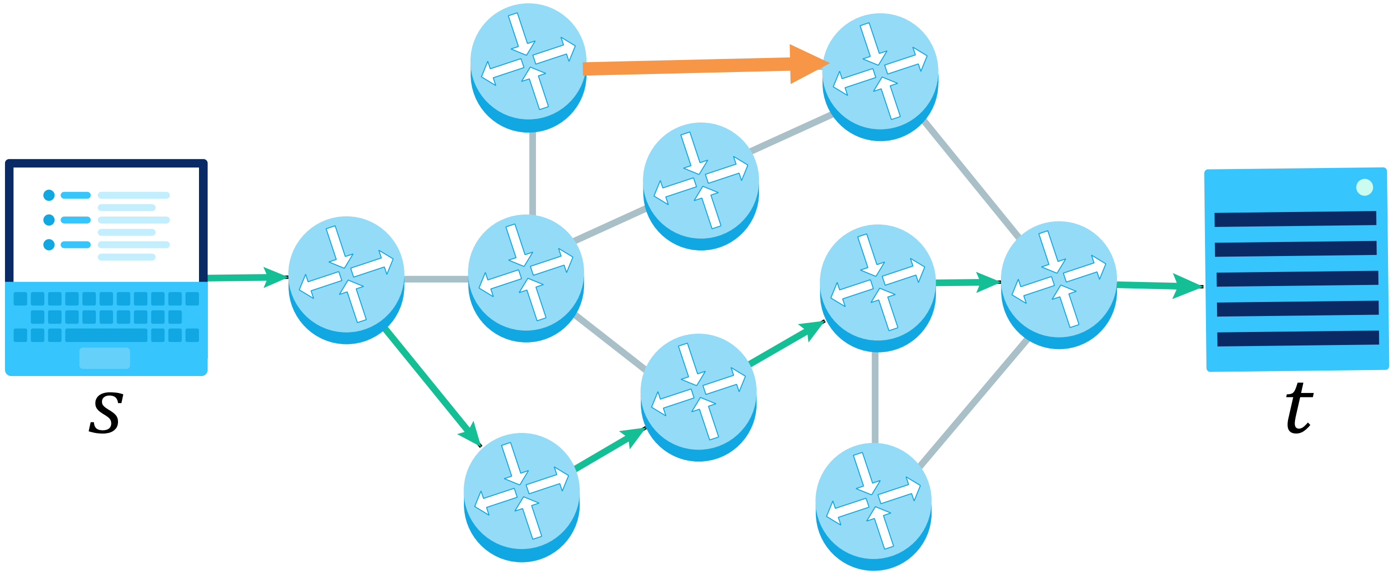
Thompson Sampling for (Combinatorial) Pure Exploration

Siwei Wang, Jun Zhu
Tsinghua University

Combinatorial Pure Exploration

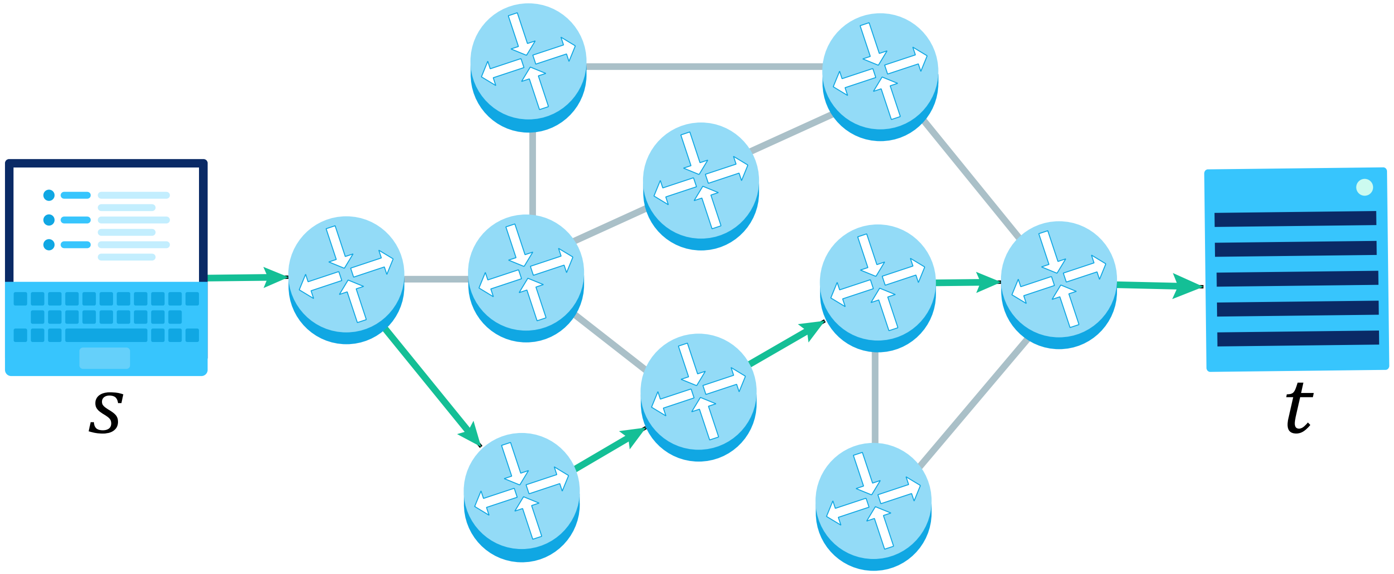
- A concrete example: routing system
  - At each time step, the system sends a message through one **edge**, and then receives a random delay
  - The system wants to look for a **path** from s to t with the minimum expected delay
  - Goal
    - High-probability $(1-\delta)$ correctness guarantee
    - Low complexity



Combinatorial Pure Exploration

- A concrete example: routing system
  - Get random observations from **single** arms
  - Look for the optimal **set of arms** (following a special combinatorial structure)

- Other applications
  - Social networks
  - Online advertisements
  - Crowdsourcing systems



Prior Works

- Pure exploration in classic MAB
  - [Even-Dar et al., 2006], [Audibert et al., 2010], [Kalyanakrishnan et al., 2012], [Kaufmann & Kalyanakrishnan, 2013], [Russo, 2016], [Shang et al, 2020]

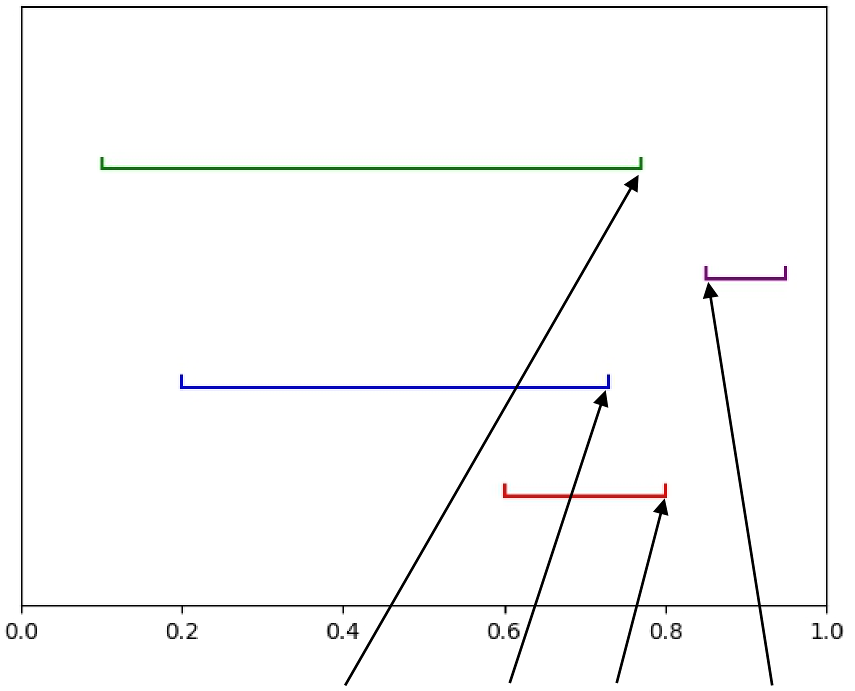
- Combinatorial pure exploration

  [Chen et al., 2014], [Gabillon et al., 2016], [Chen et al., 2017], [Jourdan et al., 2021]

Challenges

- The UCB approach
  - Compute the confidence interval for the expected reward of each arm
    - $\left[\hat{\mu}_i - \sqrt{\frac{\log 1/\delta}{2N_i}}, \hat{\mu}_i + \sqrt{\frac{\log 1/\delta}{2N_i}}\right]$
  - Claim one arm is optimal only if its lower confidence bound is larger than the upper confidence bounds of the others

$$\Pr\left[|\hat{\mu}_i - \mu_i| \geq \Delta\right] \leq 2\exp(-2N_i\Delta^2)$$



upper/lower confidence bounds

Challenges

- Challenge of the UCB approach
  - To guarantee efficiency, we use the sum of upper (lower) confidence bounds in an arm set S to be the upper (lower) confidence bound of S

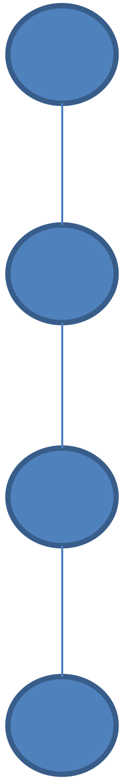
$$\hat{\mu}_1 - \sqrt{\frac{\log 1/\delta}{2N_1}} \qquad \hat{\mu}_1 + \sqrt{\frac{\log 1/\delta}{2N_1}}$$

$$+ \qquad\qquad +$$

$$\hat{\mu}_2 - \sqrt{\frac{\log 1/\delta}{2N_2}} \qquad \hat{\mu}_2 + \sqrt{\frac{\log 1/\delta}{2N_2}}$$

$$+ \qquad\qquad +$$

$$\hat{\mu}_3 - \sqrt{\frac{\log 1/\delta}{2N_3}} \qquad \hat{\mu}_3 + \sqrt{\frac{\log 1/\delta}{2N_3}}$$



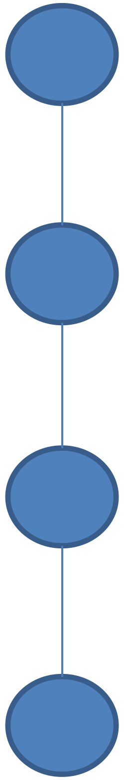
$$\Downarrow \qquad\qquad \Downarrow$$

$$\sum_{i\in S}\hat{\mu}_i \pm \sum_{i\in S}\sqrt{\frac{\log 1/\delta}{2N_i}}$$

Challenges

- Challenge of the UCB approach
  - To guarantee efficiency, we use the sum of upper (lower) confidence bounds in an arm set S to be the upper (lower) confidence bound of S
  - It is not tight when the observations are independent
    - The tight confidence interval is

$$\hat{\mu}_1 - \sqrt{\frac{\log 1/\delta}{2N_1}} + \hat{\mu}_2 - \sqrt{\frac{\log 1/\delta}{2N_2}} + \hat{\mu}_3 - \sqrt{\frac{\log 1/\delta}{2N_3}}$$



$$\hat{\mu}_1 + \sqrt{\frac{\log 1/\delta}{2N_1}} + \hat{\mu}_2 + \sqrt{\frac{\log 1/\delta}{2N_2}} + \hat{\mu}_3 + \sqrt{\frac{\log 1/\delta}{2N_3}}$$

$$\Downarrow$$

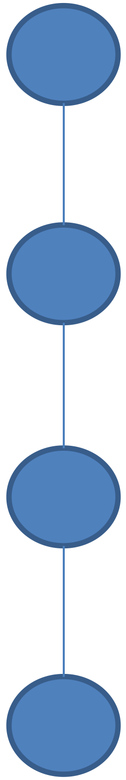
$$\sum_{i \in S} \hat{\mu}_i \pm \sqrt{\sum_{i \in S} \frac{\log 1/\delta}{2N_i}}$$

$$\sum_{i \in S} \hat{\mu}_i \pm \sum_{i \in S} \sqrt{\frac{\log 1/\delta}{2N_i}}$$

Challenges

- Challenge of the UCB approach
  - To guarantee efficiency, we use the sum of upper (lower) confidence bounds in an arm set S to be the upper (lower) confidence bound of S
  - It is not tight when the observations are independent
    - The tight confidence interval is

$$\hat{\mu}_1 - \sqrt{\frac{\log 1/\delta}{2N_1}} + \hat{\mu}_2 - \sqrt{\frac{\log 1/\delta}{2N_2}} + \hat{\mu}_3 - \sqrt{\frac{\log 1/\delta}{2N_3}} \Downarrow$$



$$\hat{\mu}_1 + \sqrt{\frac{\log 1/\delta}{2N_1}} + \hat{\mu}_2 + \sqrt{\frac{\log 1/\delta}{2N_2}} + \hat{\mu}_3 + \sqrt{\frac{\log 1/\delta}{2N_3}} \Downarrow$$

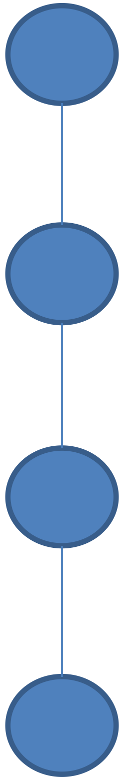
$$\sum_{i\in S} \hat{\mu}_i \pm \sqrt{\sum_{i\in S} \frac{\log 1/\delta}{2N_i}}$$

$$\sum_{i\in S} \hat{\mu}_i \pm \sum_{i\in S} \sqrt{\frac{\log 1/\delta}{2N_i}}$$

The Idea of Thompson Sampling

- Using **independent** random samples instead of the confidence bounds (for each single arm)
  - With high probability, the sum of random samples will not exceed the tight confidence interval

$$\sum_{i\in S}\hat{\mu}_i \pm \sqrt{\sum_{i\in S}\frac{\log 1/\delta}{2N_i}}$$



$$\theta_1 \sim \mathcal{N}\left(\hat{\mu}_1, \frac{1}{N_1}\right)$$

$$+$$

$$\theta_2 \sim \mathcal{N}\left(\hat{\mu}_2, \frac{1}{N_2}\right)$$

$$+$$

$$\theta_3 \sim \mathcal{N}\left(\hat{\mu}_3, \frac{1}{N_3}\right)$$

$$\Downarrow$$

$$\sum_{i\in S}\theta_i \sim \mathcal{N}\left(\sum_{i\in S}\hat{\mu}_i, \sum_{i\in S}\frac{1}{N_i}\right)$$

Theoretical Results

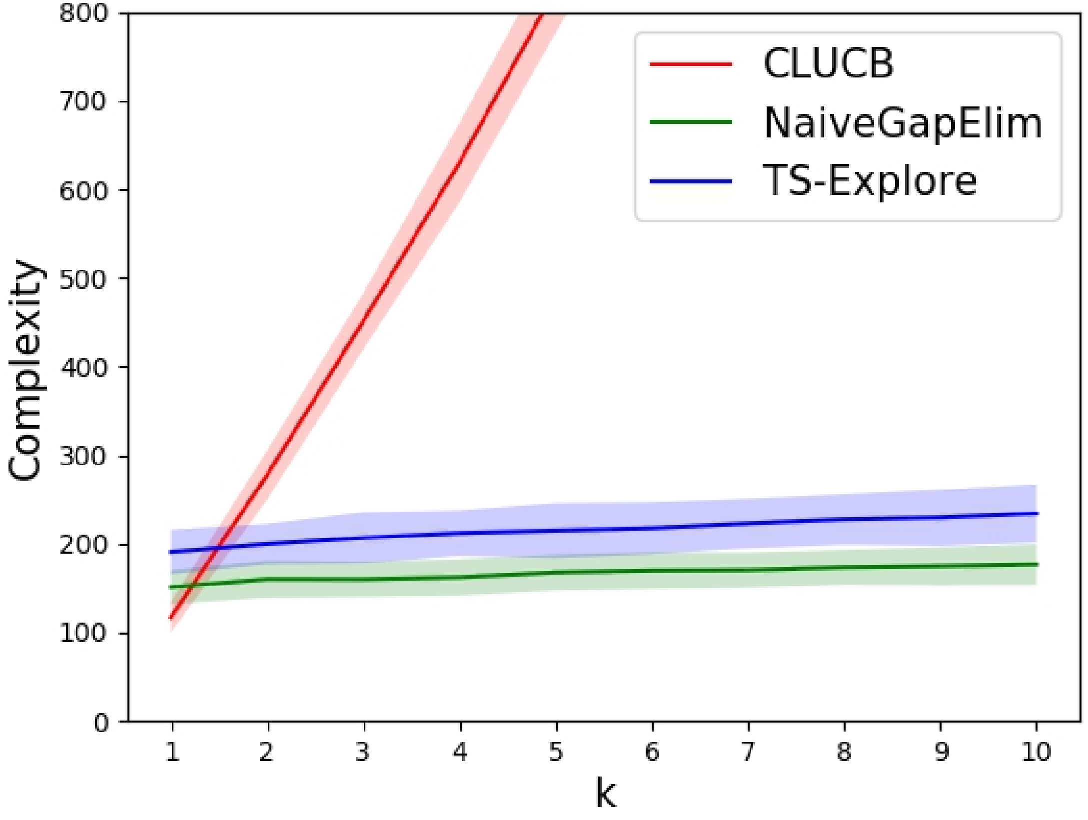
- TS-Explore

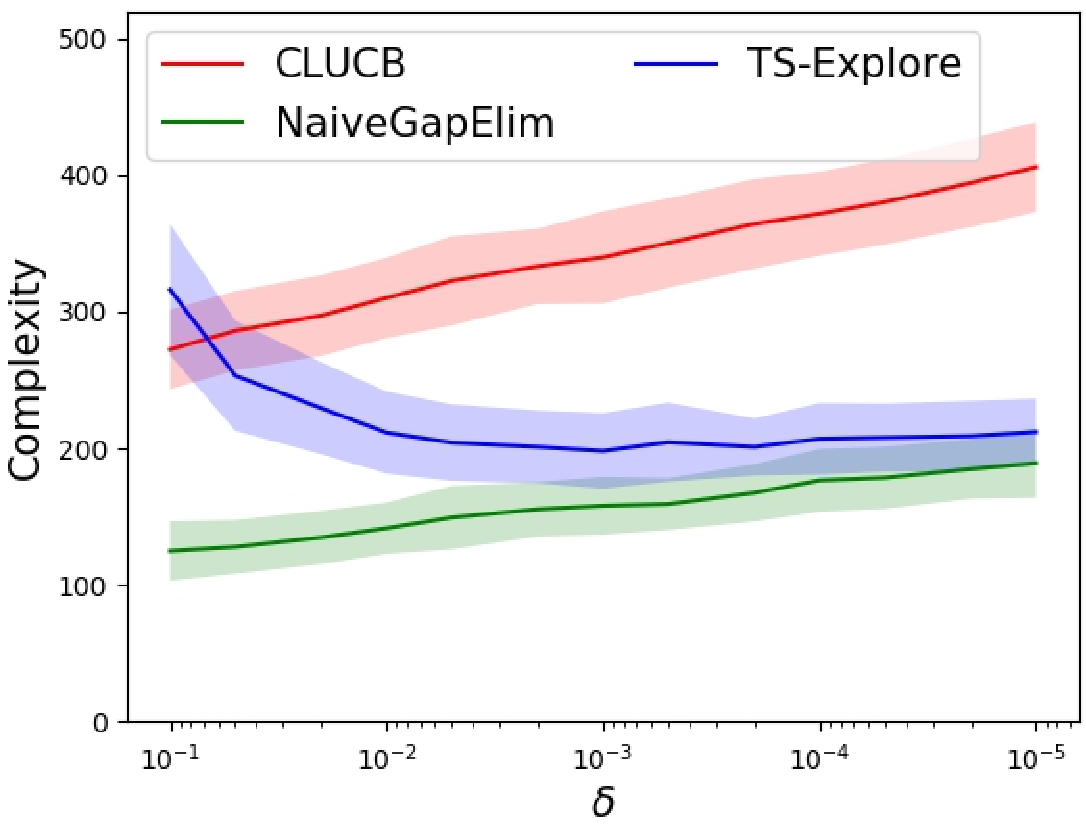
Algorithm	Efficient	Complexity
CLUCB	Yes	$O(mk^2 \log 1/\delta)$
TS-Explore	Yes	$O(mk \log 1/\delta)$[1]
NaiveGapElim	No	Optimal

  - Efficient
  - Reduce one factor of k in the complexity bound

[1] There exist many cases such that this bound is also optimal

Experimental Results





Thank you！