Federated Multi-armed Bandit via Uncoordinated Exploration

Rensselaer

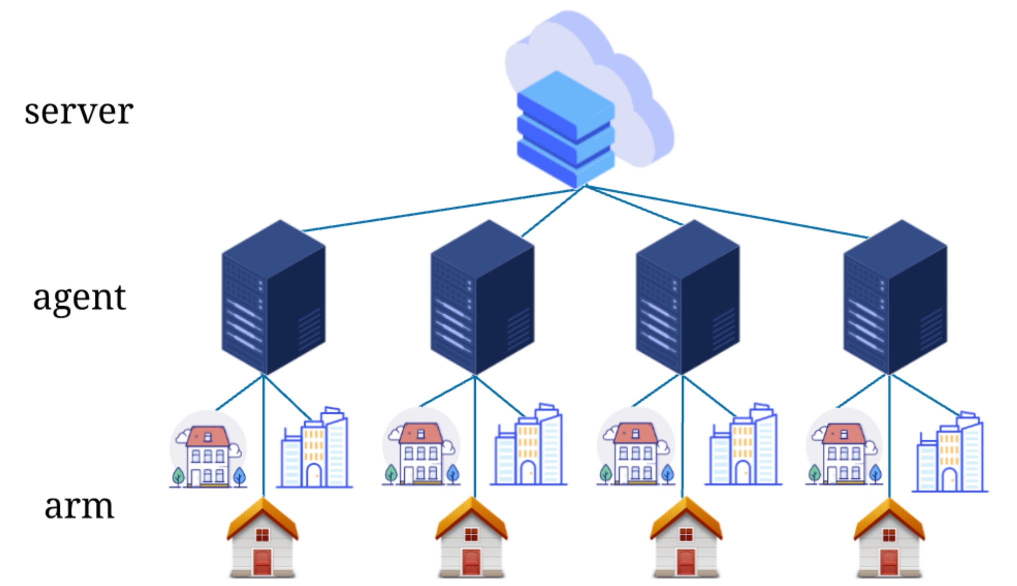
Paper #5203

Zirui Yan, Quan Xiao, Tianyi Chen, Ali Tajer

Electrical, Computer, and Systems Engineering Department, Rensselaer Polytechnic Institute

Motivation

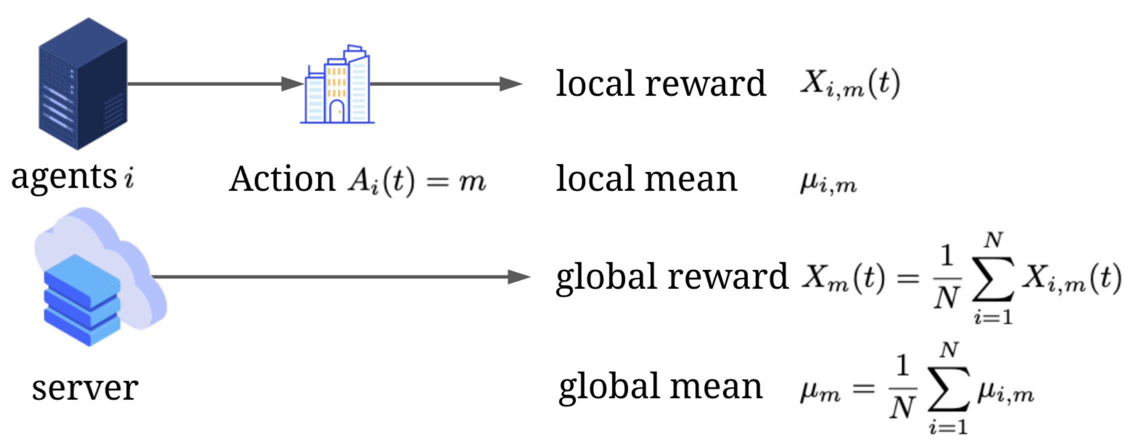
Federated multi-armed bandit (FMAB) setting: N agents and M arms



Why federated setting?

- Leveraging data from distributed agents without sharing raw data
- Collaboratively finding a globally optimal arm

FMAB Problem



Goal: minimize the **static regret** after T rounds

$$R(T) \triangleq NT\mu_1 - \sum_{t=1}^{T}\sum_{i=1}^{N} \mathbb{E}\left[X_{A_i(t)}(t)\right]$$

Arm 1 is assumed to be the optimal arm.

What an FMBA algorithm should contain?

Personalized arm-selection decisions for each agent

- delayed information - different exploration degree

Key Challenge:

Data heterogeneity + **Exploration-exploitation dichotomy**

Prior Art

Gossip UCB (Zhu et al. 2021):

Decentralized MAB, Local information + Neighbor information

PF-UCB (Shi et al. 2021): Federated MAB, Global information

Intuition: Form local decisions

Obtain global information only intermittently

FedUCB-UE Algorithm

exploration time counter | reward estimator

Agents: $n_{i,m}(t) \triangleq \sum_{\tau=0}^{t} \mathbb{1}\{A_i(\tau) = m\}$ $\quad \widehat{X}_{i,m}(t) \triangleq \frac{1}{n_{i,m}(t)} \sum_{\tau=0}^{t} X_{i,m}(t)\mathbb{1}\{A_i(\tau)=m\}$ (1)

Server: $n_m(t) \triangleq \max_i n_{i,m}(t)$ $\quad \widehat{X}_m(t) \triangleq \frac{1}{N}\sum_{i=1}^{N} \widehat{X}_{i,m}(t)$ (2)

Initialization:

Agents $n_{i,m}(0) = 1, \widehat{X}_{i,m}(0) = X_{i,m}(0)$

Server $n_m(0) = 1, \widehat{X}_m(0) = \frac{1}{N}\sum_{i=1}^{N} \widehat{X}_{i,m}(0)$

E rounds of local exploration:

Last communication time

Compute underexplored set $S_i(t) \triangleq \{m | n_{i,m}(t-1)E < n_m(t_0)\}$

- If $S_i(t) \neq \emptyset$, agent i randomly selects $A_i(t)$ from $S_i(t)$
- Otherwise, agent i chooses the arm that maximizes $\text{UCB}_{i,m}(t)$

$$\text{UCB}_{i,m}(t) \triangleq B_{i,m}(t-1) + C_m(t-1)$$

Unbiased estimator:

$$B_{i,m}(t) \triangleq \widehat{X}_m(t_0) + \frac{1}{N}\left[\widehat{X}_{i,m}(t) - \widehat{X}_{i,m}(t_0)\right]$$

Confidence level:

$$C_m(t) \triangleq \min\left\{\sqrt{\frac{8\log(t+1)}{N}}, \sqrt{\frac{8\log(t+1)}{N(n_m(t_0)-2)}}\right\}$$

Updates the $n_{i,m}(t)$ and $\widehat{X}_{i,m}(t)$ according to (1)

Communication:

After E local rounds, agents will communicate with the server

- Each agent i transmits $n_{i,m}(t)$ and $\widehat{X}_{i,m}(t)$ to the server
- The sever calculates $n_m(t)$ and $\widehat{X}_m(t)$ according to (2)

The sever broadcasts $n_m(t)$ and $\widehat{X}_m(t)$ to all agents

Convergence Guarantee

Conjecture: Define the event $D^c_{i,m}(t) \triangleq \left\{n_{i,m}(t) < \frac{n_m(t)}{2} - 1\right\}$.

Assume that for all non-optimal arms m, we have

$$q_{i,m} = \sum_{t=1}^{\infty} \mathbb{P}\{D^c_{i,m}(t)\} < +\infty \ .$$

Intuition: Agents' actions tend to achieve consensus

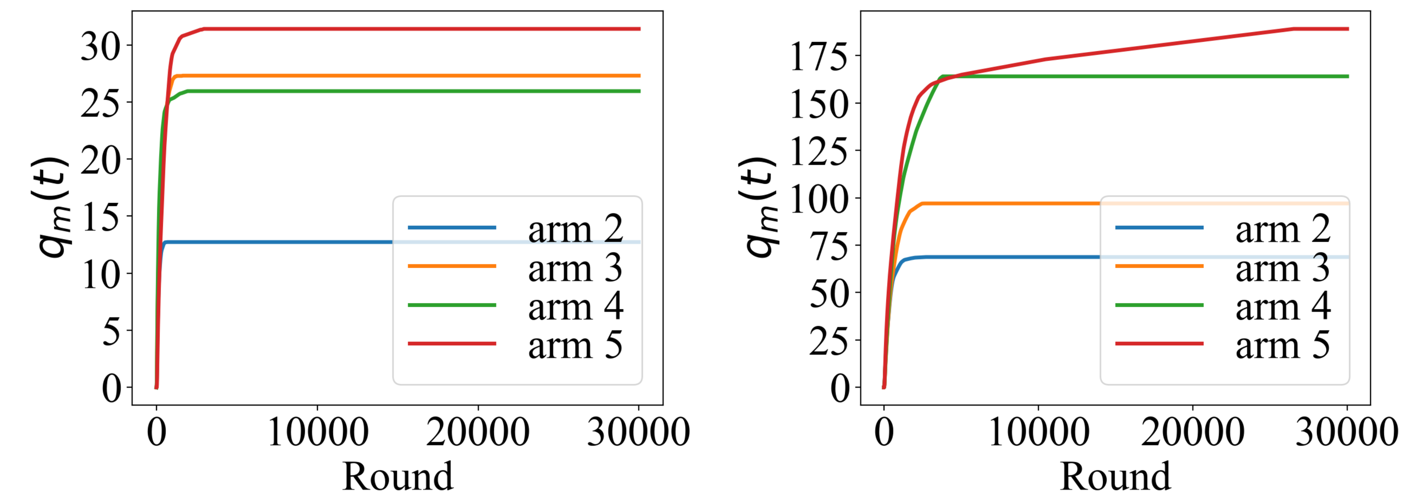
Theorem: When Conjecture holds, and $E \geq M$, the regret bound for the FedUCB-UE algorithm satisfies

$$R(T) = \mathcal{O}(\log T) \ .$$

Experiments

Verification of the conjecture:

FMAB setting with $N = 20$ and $M = 5$ and $\mu_{i,m} \sim \mathcal{N}\left(\frac{5-m}{100}, 1\right)$.

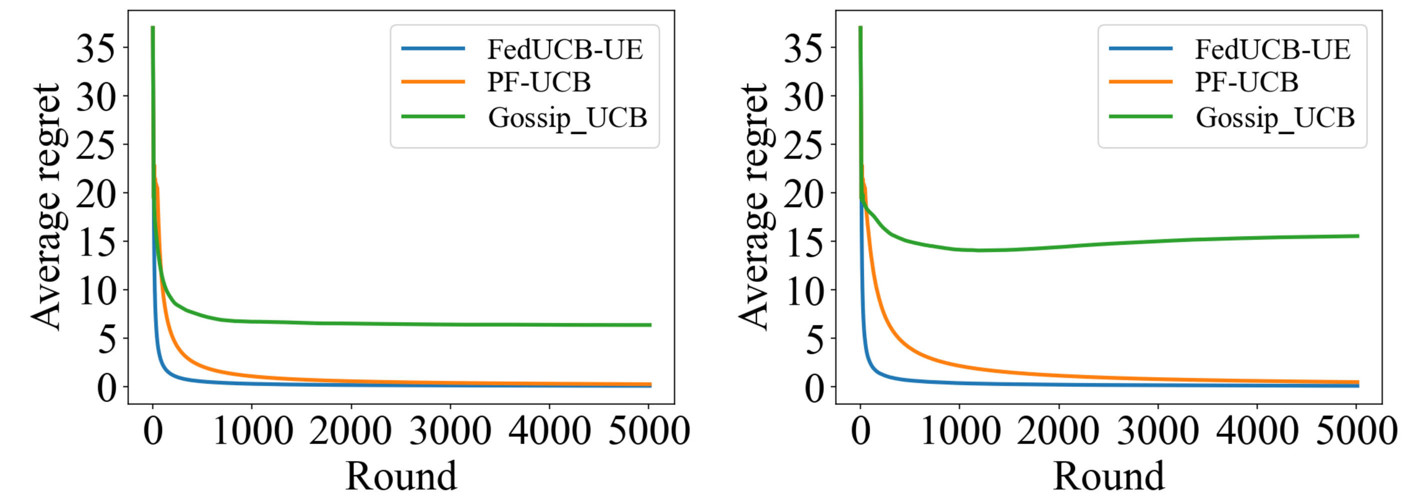


a) $E = 10$ b) $E = 20$

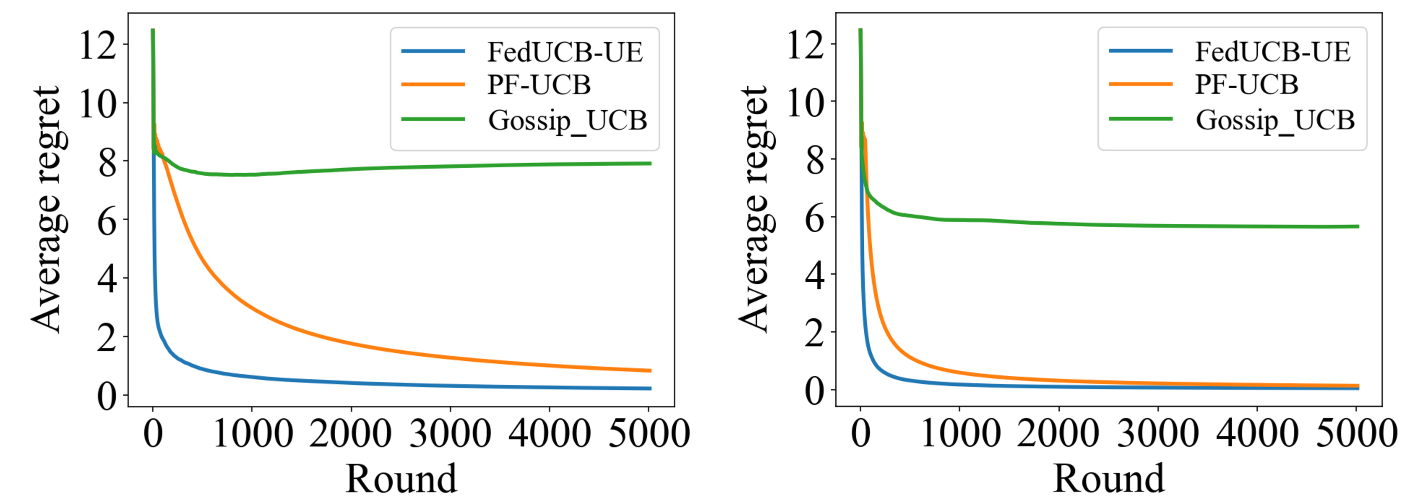
Comparison with prior art:

FMAB setting with $N = 20$, $M = 10$ and $E = 10$

Model 1: $\mu_{i,m} \sim \mathcal{N}\left(\frac{20-m}{5}, 1\right)$; **Model 2:** $\mu_{i,m} \sim \mathcal{N}\left(\frac{20-m}{20}, 1\right)$.



a) Model 1: theoretical confidence level b) Model 1: optimally-tuned confidence level



a) Model 2: theoretical confidence level b) Model 2: optimally-tuned confidence level

Conclusion

- FedUCB-UE has two features:
  - Agents form local decisions
  - Agents obtain global information only intermittently
- FedUCB-UE achieves the optimal $\mathcal{O}(\log T)$ regret bound
- FedUCB-UE outperforms the state-of-the-art algorithms