Near-Optimal Collaborative Learning in Bandits

Clémence Réda, Sattar Vakili, Emilie Kaufmann

In a collaborative setting in bandits, when the optimal arm for each agent maximizes some *global* reward computed across agents, communication between agents often becomes necessary. **How to guarantee sample-efficiency in that case while limiting communication? What about regret guarantees?**

Introduction of the Weighted Collaborative Model

Collaborative learning is a general machine learning paradigm in which a group of **M** agents collectively train a learning algorithm. Making personalized decisions for each agent [1] leads to the twist that **each agent** *m* **should play the optimal arm**, among **K** ones, **in a mixed model** $\mu'_m \in \mathbb{R}^K$ **which is obtained as a combination of her local model** $\mu_m \in \mathbb{R}^K$ **with the local models of other agents** $(\mu_n)_{n\neq m}$

Weighted collaborative model.
$W=(w_{n,m})_{n,m} \in [0,1]^{M \times M}$: weight matrix quantifying similarities between agents. **Expected mixed reward** for arm *k* in agent *m* is

$$\mu'_{k,m} := \sum_{n \leq M} w_{n,m} \mu_{k,n}$$

Assuming that the bandits are Gaussian, that is, the observed reward of selected arm π^t at time *t* in *m* $r_{\pi,m}$ is $r_{\pi,m} = \mu_{\pi,m} + \varepsilon$, and $\varepsilon \sim \mathcal{N}(0,1)$

This setting encompasses several prior works [1-3], including best arm identification (**M**=1). The goal is to exploit the information from **W** to decrease sample complexity/regret, with little cross-agent communication about observed rewards.

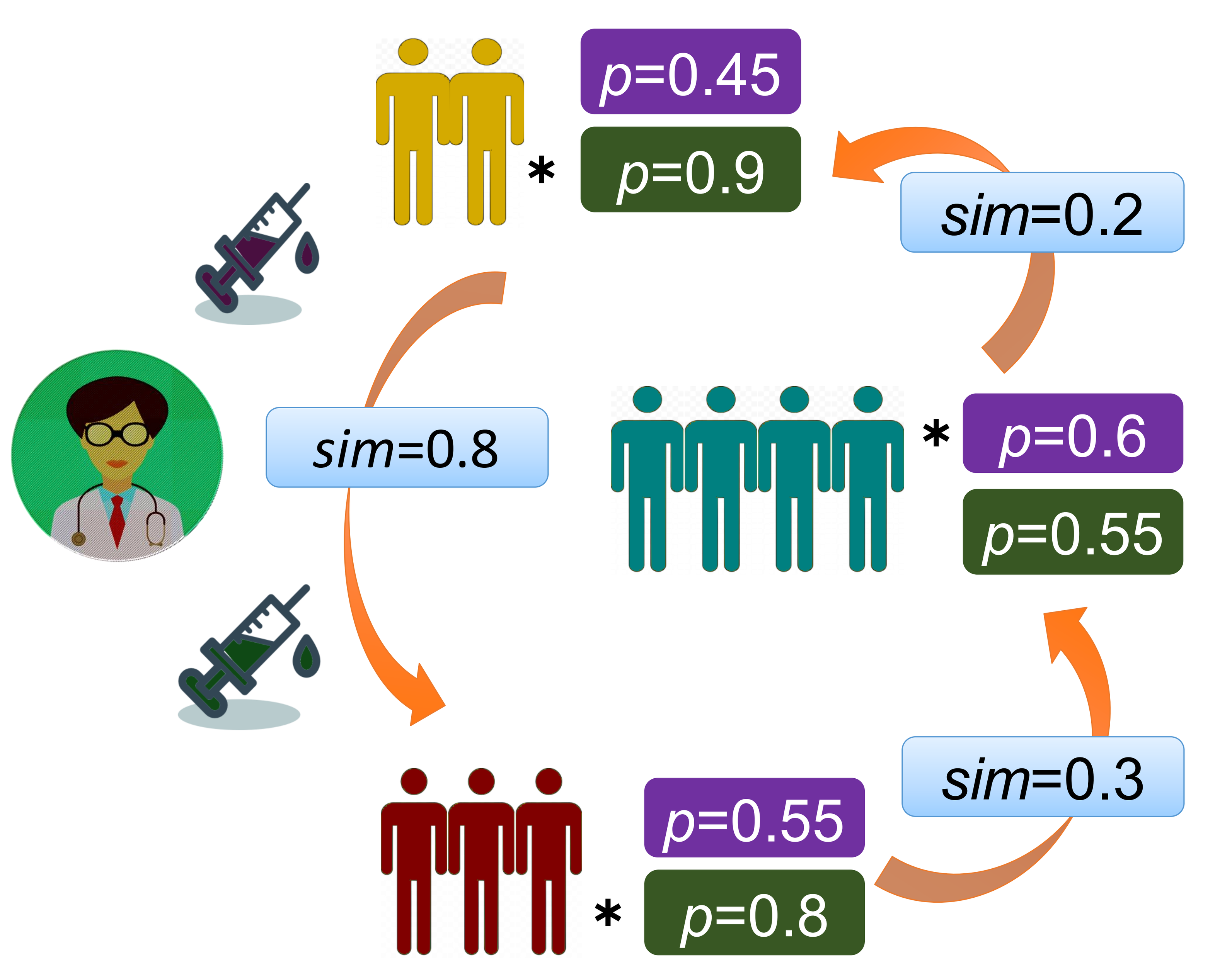

Figure 1. Collaborative setting, with **M**=3 agents/populations, **K**=2 arms/treatments. ∗ denotes optimal arms for each agent.

Collaborative Best-Arm Identification (CBAI)

Complexity of the Problem

For each agent *m*, identify with prob. $1-\delta$ the arm $\star_m :=$ **arg max**$_{k\leq K} \mu'_{k,m}$ (with highest expected reward) by observing as few rewards as possible (low *sample complexity*).

Lower bound on the sample complexity.
On instance $\mu \in \mathbb{R}^{K\times M}$ and weight matrix **W** *s.t.* $\forall m$, $w_{m,m} \neq 0$, any algorithm $\mathcal{A}$ which is correct with prob. $1-\delta$ ($\delta \leq 0.5$), and communicates at each round, samples in expectation at least $T^\star_W(\mu) \log(1/(2.4\delta))$ times, where

$$T^\star_W(\mu) := \min_{t \in (\mathbb{R}_+)^{K\times M}} \sum_{k,m} t_{k,m}$$

$$\text{s.t. } \forall m \leq M,\ \forall k \neq \star_m,\ \sum_{n \leq M} w^2_{n,m}(1/t_{k,n} + 1/t_{\star_m,n}) \leq (\mu'_{\star_m,m} - \mu'_{k,m})^2/2$$

In the instance in **Figure 1**, $T^\star_W(\mu) \approx 1{,}422$ whereas $T^\star_{Id}(\mu) \approx 3{,}368$.

Near-Optimal Algorithm for CBAI

We introduce a phased-elimination algorithm (**Algorithm 1**) to solve this problem, based on a relaxation of the lower bound problem $\tilde{P}^\star$

Relaxed lower bound problem.
For any $\Delta \in \mathbb{R}^{K\times M}$ and weight matrix **W**

$$\tilde{P}^\star(\Delta) := \min_{t \in (\mathbb{R}_+)^{K\times M}} \sum_{k,m} t_{k,m} \text{ s.t. } \forall m \leq M,\ \forall k \leq K,\ \sum_{n \leq M} w^2_{n,m}/t_{k,n} \leq (\Delta_{k,m})^2/2$$

We show that $\tilde{P}^\star(\Delta) \leq T^\star_W(\mu) \leq 2\tilde{P}^\star(\Delta)$ for $\Delta_{k,m} := \mu'_{\star_m,m} - \mu'_{k,m}$. Then:

Sample Complexity Upper Bound for W-CPE-BAI.
With prob. $1-\delta$, **W-CPE-BAI** outputs the optimal arm for each agent by sampling at most $32T^\star_W(\mu)\log_2(8/(\min_{k\leq K}\Delta_{k,m}))\log(1/\delta) + o_{\delta\to 0}(\log(1/\delta))$ times.

Collaborative Regret Minimization

For each agent *m*, the objective is to minimize the following regret at horizon **T** $\quad R_\mu(T) := \mathbf{E}\left[\sum_{m,t} \mu'_{\star_m,m} - \mu_{\pi^t,m}\right]$

Lower bound on the regret.
On instance $\mu \in \mathbb{R}^{K\times M}$ and weight matrix **W,** any uniformly efficient algorithm in which all agents communicate after each round satisfies

$$\liminf_{T\to\infty} R_\mu(T)/\log(T) \geq C^\star_W(\mu),$$

where

$$C^\star_W(\mu) := \min_{c \in (\mathbb{R}^+)^{K\times M}} \sum_{m,k\neq\star_m} c_{k,m}\Delta_{k,m}$$

$$\text{s.t. } \forall m \leq M,\ \forall k \leq K,\ \sum_{n \text{ s.t. } k\neq\star_n} w^2_{n,m}/c_{k,n} \leq (\Delta_{k,m})^2/2$$

This lower bound proves the conjecture in [1] with weaker assumptions.

Initialize $r \leftarrow 0$, $\forall k,m$, $\tilde{\Delta}_{k,m}(0) \leftarrow 1$, $n_{k,m}(0) \leftarrow 1$, $\forall m$, $B_m(0) \leftarrow [K]$
Draw each arm k by each agent m once
repeat
  # Central server
  $B(r) \leftarrow \bigcup_{m\in[M]} B_m(r)$
  Compute $t(r) \leftarrow \tilde{\mathcal{P}}^\star\left(\left(\sqrt{2}\tilde{\Delta}_{k,m}(r)\right)_{k,m}\right)$
  For any $k \in [K]$, compute

$$(d_{k,m}(r))_{m\in[M]} \leftarrow \arg\min_{d\in\mathbb{N}^M} \sum_m d_m \text{ s.t. } \forall m \in [M], \frac{n_{k,m}(r-1)+d_m}{\beta_\delta(n_{k,\cdot}(r-1)+d)} \geq t_{k,m}(r)$$

  Send to each agent m $(d_{k,m}(r))_{k,m}$ and $d_{\max} := \max_{n\in[M]} \sum_{k\in[K]} d_{k,n}(r)$

  # Agent m
  Sample arm $k \in B(r)$ $d_{k,m}(r)$ times, so that $n_{k,m}(r) = n_{k,m}(r-1) + d_{k,m}(r)$
  Remain idle for $d_{\max} - \sum_{k\in[K]} d_{k,m}(r)$ rounds
  Send to the server empirical mean $\hat{\mu}_{k,m}(r) := \sum_{s\leq n_{k,m}(r)} X_{k,m}(s)/n_{k,m}(r)$ for any $k \in [K]$

  # Central server
  Compute the empirical mixed means $(\hat{\mu}'_{k,m}(r))_{k,m}$ based on $(\hat{\mu}_{k,m}(r))_{k,m}$ and W
  // Update set of candidate best arms for each user
  for $m = 1$ **to** M **do**

$$B_m(r+1) \leftarrow \left\{k \in B_m(r) \mid \hat{\mu}'_{k,m}(r) + \Omega_{k,m}(r) \geq \max_{j\in B_m(r)} \left(\hat{\mu}'_{j,m}(r) - \Omega_{j,m}(r)\right)\right\}$$

  end for
  // Update the gap estimates
  For all k,m, $\tilde{\Delta}_{k,m}(r+1) \leftarrow \tilde{\Delta}_{k,m}(r) \times (1/2)^{\mathbb{1}(k\in B_m(r+1) \wedge |B_m(r+1)|>1)}$
  $r \leftarrow r+1$
until $\forall m \in [M], |B_m(r)| \leq 1$
Output: $\{k \in B_m(r) : m \in [M]\}$

Algorithm 1. Weighted Collaborative Phased Elimination for Best Arm Identification (**W-CPE-BAI**) algorithm for CBAI.

Discussion

- The strategy to build a near-optimal algorithm for CBAI has the potential to be extended to other identification problems.
- A possible subsequent work would add privacy-preserving features to these algorithms [4].

References

[1] Shi, Shen & Yang. "Federated multi-armed bandits with personalization." *International Conference on Artificial Intelligence and Statistics.* PMLR, 2021.

[2] Tao, Zhang & Zhou. "Collaborative learning with limited interaction: Tight bounds for distributed exploration in multi-armed bandits." *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS).* IEEE, 2019.

[3] Russac, Yoan, et al. "A/B/n Testing with Control in the Presence of Subpopulations." *Advances in Neural Information Processing Systems* 34 (2021): 25100-25110.

[4] Dubey & Pentland. "Differentially-private federated linear bandits." *Advances in Neural Information Processing Systems* 33 (2020): 6003-6014.

SCAN ME