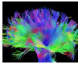
Near-Optimal Collaborative Learning in Bandits

Clémence Réda[a] · Sattar Vakili[b] · Émilie Kaufmann[c]



[a] Université Paris Cité, Inserm, NeuroDiderot, Paris
[b] MediaTek Research, UK
[c] CNRS, Univ. Lille, Inria Scool

NeurIPS 2022 – New Orleans

Fixed-Confidence Best Arm Identification (BAI) Problem

Identify with prob. $1-\delta$ the arm $k^\star \in [K]$ with highest expected reward $\arg\max_k \mu_k$ by observing as few samples as possible (*low sample complexity*)

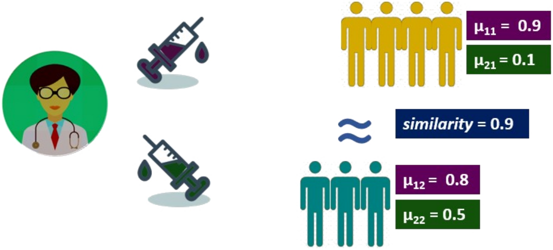
Reward from k

$r_k = \mu_k + \varepsilon$

$\varepsilon \sim \mathcal{N}(0,1)$

Fixed-Confidence BAI Problem with M populations

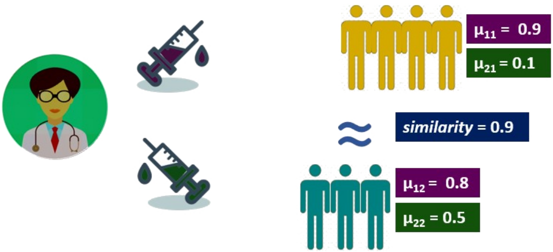
For each population m, identify with prob. $1-\delta$ the arm $k^\star_m$ with highest expected reward $\arg\max_k \mu_{k,m}$ with low sample complexity



Reward from k in m

$$r_{k,m} = \mu_{k,m} + \varepsilon$$

$$\varepsilon \sim \mathcal{N}(0,1)$$



- **Run independently M bandit algorithms on K arms**

  Oversampling due to ignoring info from similar populations

- **Run 1 bandit algorithm on $K \times M$ arms**

  High communication cost across populations

Exploit info from other populations with little communication
$\Rightarrow$ maximize *mixed* (instead of local) rewards[1]

Weighted Collaborative Model

$W = (w_{n,m})_{n,m} \in [0,1]^{M \times M}$ weight matrix on populations
Expected *mixed* reward for arm k in population m is

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

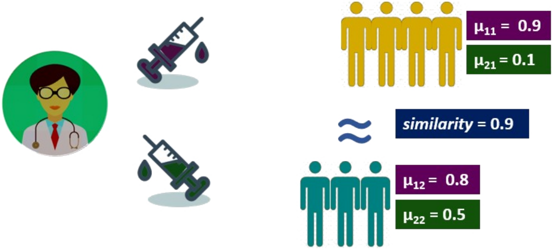
For population m, identify $k'^{\star}_m$ *s.t.* $\mu'_{k'^{\star}_m, m} = \arg\max_k \mu'_{k,m}$
with low sample complexity and with little communication

[1] *Shi, Shen, and Yang (2021). AISTATS. PMLR, pp. 2917–2925*

Lower bound on the expected sample complexity τ

For any δ-correct algorithm $\mathfrak{A}$ where $\delta \leq 0.5$ and $\forall m, w_{m,m} \neq 0$

$$\mathbb{E}_{\mathfrak{A},\mu}[\tau] \geq T^\star_W(\mu) \log\left(\frac{1}{2.4\delta}\right)$$



$$\mu = \begin{bmatrix} 0.9 & 0.8 \\ 0.1 & 0.5 \end{bmatrix}$$

$$W = \frac{1}{1.9}\begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

$$\underbrace{T^\star_W(\mu)}_{\approx 28} \ll \underbrace{T^\star_{\mathrm{Id}_M}(\mu)}_{\approx 101}$$

In our paper

- A phased algorithm based on a relaxation of the lower bound…
- … with near-optimal sample complexity and low communication cost
- New insights on the regret minimization counterpart of this problem