Federated $\mathcal{X}$-armed Bandit

Wenjie Li[1*], **Qifan Song**[1], **Jean Honorio**[2], **Guang Lin** [3]

[1]Department of Statistics, Purdue University
[2]School of Computing and Information Systems, The University of Melbourne
[3]Departments of Mathematics and School of Mechanical Engineering, Purdue University
li3549@purdue.edu, qfsong@purdue.edu, jhonorio@unimelb.edu.au, guanglin@purdue.edu

Abstract

This work establishes the first framework of federated $\mathcal{X}$-armed bandit, where different clients face heterogeneous local objective functions defined on the same domain and are required to collaboratively figure out the global optimum. We propose the first federated algorithm for such problems, named `Fed-PNE`. By utilizing the topological structure of the global objective inside the hierarchical partitioning and the weak smoothness property, our algorithm achieves sublinear cumulative regret with respect to both the number of clients and the evaluation budget. Meanwhile, it only requires logarithmic communications between the central server and clients, protecting the client privacy. Experimental results on synthetic functions and real datasets validate the advantages of `Fed-PNE` over various centralized and federated baseline algorithms.

Introduction

Federated bandit is a newly-developed bandit problem that incorporates federated learning with sequential decision making (McMahan et al. 2017; Shi and Shen 2021a). Unlike the traditional bandit models where the exploration-exploitation tradeoff is the only major concern, federated bandit problem also takes account of the modern concerns of data heterogeneity and privacy protection towards trustworthy machine learning. In particular, in the federated learning paradigm, the data available to each client could be drawn from non-i.i.d distributions, making collaborations between the clients necessary to make valid inferences for the aggregated global model. However, due to user privacy concerns and the large communication cost, such collaborations across the clients must be restricted and avoid direct transmissions of the local data. To make correct decisions in the future, the clients have to utilize the limited communications from each other and coordinate exploration and exploitation correspondingly.

To the best of our knowledge, existing results of federated bandits, such as Dubey and Pentland (2020); Huang et al. (2021); Shi and Shen (2021a); Shi, Shen, and Yang (2021b); Xu, Xie, and Lui (2021); Huang et al. (2023), focus on either the case where the number of arms is finite (multi-armed bandit), or the case where the expected reward is a linear function of the chosen arm (linear contextual bandit). However, for problems such as dynamic pricing (Chen and Gallego 2022) and hyper-parameter optimization (Shang, Kaufmann, and Valko 2019), the available arms are often defined on a domain $\mathcal{X}$ with infinite or even uncountable cardinality, and the reward function is usually non-linear with respect to the metric employed by the domain $\mathcal{X}$. These problems challenge the applications of existing federated bandit algorithms to more complicated real-world problems. Two applications (Figure 1) that motivate our study of federated $\mathcal{X}$-armed bandit are given below.

- **Federated medicine dosage recommendation.** For the dosage recommendation of a newly-invented medicine/vaccine (in terms of volume or weight), the clinical trials could be conducted at multiple hospitals (clients). To protect patients' privacy, hospitals cannot directly share the treatment result of each trial (reward). Moreover, because of the demographic difference among the patient groups, the best dosage obtained at each hospital (i.e., the optimal of local objectives) could be different from the optimal recommended dosage for entire population of the state (i.e., the optimal of the global objective). Researchers needs to collaboratively find the global optimal dosage by exploring and exploiting the local data.

- **Federated hyper-parameter optimization.** An important application of automating machine learning workflows with minimal human intervention consists of hyper-parameter optimization for ML models, e.g., learning rate, neural network architecture, etc. Many modern data are collected by mobile devices (clients). The model performance (reward) of each hyper-parameter setting could be different for each mobile device (i.e., local objectives) due to user heterogeneity. To fully utilize the whole dataset (i.e., global objective) for hyperparameter optimization such that the obtained auto-ML model can work seamlessly for diverse scenarios, the central server need to coordinate the local search properly without violating the regulations of consumer data privacy.

In the aforementioned examples, the reward objectives are defined on a domain $\mathcal{X}$, which can often be formatted as

[*]This work was done before Wenjie joined Amazon.

Bandit algorithms	Average Regret	Commun.cost	Conf.	Heterogeneity
HCT	$\widetilde{\mathcal{O}}\left(T^{\frac{d+1}{d+2}}\right)$	N.A.	N.A.	✗
BLiN	$\widetilde{\mathcal{O}}\left(T^{\frac{d+1}{d+2}}\right)$	N.A.	✓	✗
Centralized	$\widetilde{\mathcal{O}}\left(M^{-\frac{1}{d+2}}T^{\frac{d+1}{d+2}}\right)$	$\mathcal{O}(MT)$	✗	✓
Fed-PNE (This work)	$\widetilde{\mathcal{O}}\left(M^{-\frac{1}{d+2}}T^{\frac{d+1}{d+2}}\right)$	$\widetilde{\mathcal{O}}(M\log T \vee MT^{\frac{d}{d+2}})$	✓	✓

Table 1: Comparison of the (client-wise) regret upper bounds, the communication cost for sufficiently large T and the other properties. Columns: "Commun. cost" refer to communication cost. "Conf." refers to whether the raw rewards of one client are kept confidential from the other clients and only statistical summary is shared. "Heterogeneity" refers to whether the client functions are different/the same. Rows: HCT is a single-client $\mathcal{X}$-armed bandit algorithm. BLiN is a batched-$\mathcal{X}$-armed bandit algorithm. Centralized results are adapted from the centralized algorithms such as HOO (Bubeck et al. 2011) and HCT (Azar, Lazaric, and Brunskill 2014) by assuming that the server makes all the decisions with access to all client-wise information. Notation: M denotes the number of clients; T denotes the budget (time horizon) and d denotes the near-optimality dimension in Assumption 3.

a region of $\mathbb{R}^d$ and has infinite cardinality. Moreover, the objectives (both local and global ones) are highly nonlinear mapping with respect to the arm chosen due to the complex nature of the problem. Therefore, the basic assumptions of federated multi-armed bandit or federated linear contextual bandit algorithms are violated, and thus the existing federated bandit algorithms cannot apply or perform well on such problems.

Under the classical setting where centralized data is immediately available, $\mathcal{X}$-armed bandit algorithms such as HOO and HCT have been proposed to find the optimal arm inside the domain $\mathcal{X}$ (Bubeck et al. 2011; Azar, Lazaric, and Brunskill 2014). However, these algorithms cannot be trivially adapted to the task of finding the global optimum when there are multiple clients and limit communications. The local objectives could have very different landscapes across the clients due to the non-i.i.d local datasets, and no efficient communication method has been established between $\mathcal{X}$-armed bandit algorithms that run on the local data sets. In this work, we propose a new federated algorithm where all the clients collaboratively learn the best solution to the global $\mathcal{X}$-armed bandit model on average, while few communications (in terms of the amount and the frequency) are required so that the privacy of each client is preserved.

We highlight our major contributions as follows.

- **Federated $\mathcal{X}$-armed bandit.** We establish the first framework of the federated $\mathcal{X}$-armed bandit problem, which naturally connects the $\mathcal{X}$-armed bandit problem with the characteristics of federated learning. The new framework introduces many new challenges to $\mathcal{X}$-armed bandit including (1) potential severe heterogeneity among the local objectives due to non-i.i.d local data sets, (2) the non-accessibility of the global objective for all local clients or the central server, and (3) the restriction of communications between the server and the clients.
- **New algorithm with desirable regret.** We propose a new algorithm for the federated $\mathcal{X}$-armed bandit problem named Fed-PNE. Inspired by the heuristic of arm elimination in multi-armed bandits (Lattimore and Szepesvári 2020), the new algorithm performs *hierarchical node elimination* in the domain $\mathcal{X}$. More importantly, it incorporates efficient communications between the server and the clients to transmit information while protecting client-privacy. We establish the sublinear cumulative regret upper bound of the proposed algorithm as well as the bound of the communication cost. Theoretically, we prove that Fed-PNE utilizes the advantage of federation and at the same time has high communication efficiency. We also provide a regret lower bound analysis to justify the tightness of our upper bound. Theoretical comparisons of our regret bounds with existing bounds are provided in Table 1.
- **Empirical results.** By examining the empirical performance of our Fed-PNE algorithm on both synthetic functions and real datasets, we verify the correctness of our theoretical results. We show the advantages of Fed-PNE over centralized $\mathcal{X}$-armed and kernelized bandit algorithm, and federated neural and multi-armed bandit algorithm. The empirical results exhibit the usefulness of our algorithm in real-life applications.

Preliminaries

We first introduce the preliminary concepts and notations used in this paper. For a real number $a \in \mathbb{R}$, we use $\lceil a \rceil$ to represent the smallest integer larger than a. For an integer $N \in \mathbb{N}$, we use $[N]$ to represent the set of integers $\{1, 2, \cdots, N\}$. For a set A, $|A|$ denotes the number of elements in A. We use $\widetilde{\mathcal{O}}(\cdot)$ to hide the logarithmic terms in big-$\mathcal{O}$ notations, i.e., for two functions $a(n), b(n)$, $a(n) = \widetilde{\mathcal{O}}(b(n))$ represents that $a(n)/b(n) \leq \log^k(n), \forall n > 0$ for some $k > 0$.

Problem Formulation and Performance Measure

Let $\mathcal{X}$ be a measurable space of arms. We model the problem as a federated $\mathcal{X}$-armed bandit setting where a total of M clients respectively have the access to M different *local*

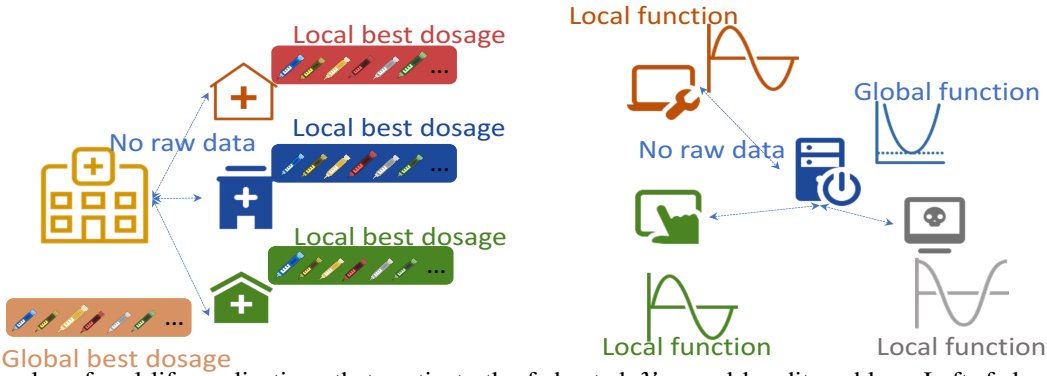


Figure 1: Examples of real-life applications that motivate the federated $\mathcal{X}$-armed bandit problem. Left: federated medicine dosage recommendation. Right: Federated hyper-parameter optimization.

objectives $f_m(x) : \mathcal{X} \mapsto \mathbb{R}$, which could be non-convex, non-differentiable and even non-continuous. Given a limited number of rounds T, each client $\in [M]$ chooses a point $x_{m,t} \in \mathcal{X}$ at each round $t \in [T]$ and observes a noisy feedback $r_{m,t} \in [0, 1]$ defined as $r_{m,t} := f_m(x_{m,t}) + \epsilon_{m,t}$, where $\epsilon_{m,t}$ is a zero-mean and bounded random noise independent from previous observations or other clients' observations. The goal of the clients is to find the point that maximizes the *global objective* $f(x)$, which is defined to be the average of the local objectives, i.e.,

$$f(x) := \frac{1}{M} \sum_{m=1}^{M} f_m(x).$$

However, the global objective is not accessible by any client. The only information that the clients have access to is: (1) noisy evaluations of their own local objective functions $f_m(x)$, and (2) communications between themselves and the central server. For the global objective, we assume that there is at least one global maximizer $x^* \in \mathcal{X}$ such that $f(x^*) = \sup_{x \in \mathcal{X}} f(x) = f^*$. Given the sequence of the points chosen by the clients $\{x_{m,t}\}_{m=1,t=1}^{M,T}$, the performance of the clients is measured by the expectation of the *cumulative regret*, defined as

$$\mathbb{E}\left[R(T)\right] := \mathbb{E}\left[\sum_{t=1}^{T} \sum_{m=1}^{M} \left(f^* - f(x_{m,t})\right)\right].$$

Another possible measure of algorithm performance is the so-called *simple regret* which only evaluates the goodness of optimizer in the final round, i.e., $r(T) = \sum_{m=1}^{M}(f^* - f(x_{m,T}))$. This paper aligns with the standard federated bandit analysis framework and focuses on cumulative regret only (Shi and Shen 2021a; Huang et al. 2021). Moreover, as mentioned by Bubeck et al. (2011), we always have $\mathbb{E}[r(T)] \leq \mathbb{E}[R(T)]/T$ if we select the path via a cumulative regret-based policy.

Hierarchical Partitioning of the Parameter Space

Following the recent progress in centralized $\mathcal{X}$-armed bandit (e.g., Azar, Lazaric, and Brunskill 2014; Shang, Kaufmann, and Valko 2019; Bartlett, Gabillon, and Valko 2019), we utilize a pre-defined infinitely-deep hierarchical partitioning $\mathcal{P} := \{\mathcal{P}_{h,i}\}_{h,i}$ of the parameter space $\mathcal{X}$ to optimize the objective functions. The hierarchical partition discretizes the space by recursively defining the following relationship:

$$\mathcal{P}_{0,1} := \mathcal{X}, \qquad \mathcal{P}_{h,i} := \bigcup_{j=0}^{k-1} \mathcal{P}_{h+1,ki-j},$$

where k is the (maximum) number of disjoint children for one node, and for every node $\mathcal{P}_{h,i}$, (h, i) denotes the depth and the index of the node inside the partition. Each node $\mathcal{P}_{h,i}$ on depth h is partitioned into k children on depth $h+1$, while the union of all the nodes on each depth h equals the parameter set $\mathcal{X}$. The partition is chosen before the optimization process and the same partition of the space $\mathcal{X}$ is shared and used by all the M clients as the partition itself reveals no information of the reward distributions of local objectives. A simple and intuitive example is a binary equal-sized partition on the domain $\mathcal{X} = [0, 1]$, where each node on depth h has length $(0.5)^h$.

Communication Model and Privacy Concerns

Similar to the setting of federated multi-armed bandit (Shi and Shen 2021a; Huang et al. 2021), we assume that there exists a central server that coordinates the behaviors of all the different clients. The server has access to the same partition of the parameter space used by all the clients, and is able to communicate with the clients. Due to privacy concerns, the client-side algorithm should keep the reward of each evaluation confidential and the only things that can be transmitted to the server are the local statistical summary of the rewards. The clients are not allowed to communicate with each other. In accordance to McMahan et al. (2017); Shi and Shen (2021a), we assume that the server and the clients are fully synchronized. Although the clients can communicate with the server, the number of clients M could be very large and thus the communication would be very costly. We take into account such communication cost in our algorithm design and the theoretical analysis.

Definitions and Assumptions

To analyze the performance of the proposed algorithms, we use the following set of definitions and assumptions,

which are also present in the prior works on $\mathcal{X}$-armed bandit (Bubeck et al. 2011; Azar, Lazaric, and Brunskill 2014).

Assumption 1. (Dissimilarity) *The space $\mathcal{X}$ is equipped with a dissimilarity function $\ell : \mathcal{X}^2 \mapsto \mathbb{R}$ such that $\ell(x, x') \geq 0, \forall (x, x') \in \mathcal{X}^2$ and $\ell(x, x) = 0$*

Throughout this work, we assume that $\mathcal{X}$ satisfies Assumption 1. Given the dissimilarity function ℓ, the diameter of a set $\mathcal{A} \subset \mathcal{X}$ is defined as $\operatorname{diam}(\mathcal{A}) = \sup_{x,y \in \mathcal{A}} \ell(x, y)$. The open ball of radius r and with center c is then defined as $\mathcal{B}(c, r) = \{x \in \mathcal{X} : \ell(x, c) \leq r\}$. We now introduce the local smoothness assumptions.

Assumption 2. (Local Smoothness) *We assume that there exist constants $\nu_1, \nu_2 > 0$, and $0 < \rho < 1$ such that for all nodes $\mathcal{P}_{h,i}, \mathcal{P}_{h,j} \in \mathcal{P}$ on depth h,*

- $\operatorname{diam}(\mathcal{P}_{h,i}) \leq \nu_1 \rho^h$
- $\exists x^\circ_{h,i} \in \mathcal{P}_{h,i}$ *s.t.* $\mathcal{B}_{h,i} := \mathcal{B}\left(x^\circ_{h,i}, \nu_2 \rho^h\right) \subset \mathcal{P}_{h,i}$
- $\mathcal{B}_{h,i} \cap \mathcal{B}_{h,j} = \emptyset$ *for all* $1 \leq i < j \leq k^h$.
- *The global objective function f satisfies that for all $x, y \in \mathcal{X}$, we have*

$$f^* - f(y) \leq f^* - f(x) + \max\{f^* - f(x), \ell(x, y)\}$$

Remark 1. Similar to the existing works on the $\mathcal{X}$-armed bandit problem, the dissimilarity function ℓ is not an explicit input required by our `Fed-PNE` algorithm and only the smoothness constants ν_1, ρ are accessed (Bubeck et al. 2011; Azar, Lazaric, and Brunskill 2014). As mentioned by Bubeck et al. (2011); Grill et al. (2015), most regular functions satisfy Assumption 2 on the standard equal-sized partition with accessible ν_1 and ρ.

Finally, we introduce the definition of the near-optimality dimension, which measures the number of near-optimal regions and thus the difficulty of the problem (Azar, Lazaric, and Brunskill 2014).

Assumption 3. (Near-optimality dimension) *Let $\epsilon = 6\nu_1 \rho^h$ and $\epsilon' = \rho^h < \epsilon$, for any subset of ϵ-optimal nodes $\mathcal{X}_\epsilon = \{x \in \mathcal{X} : f^* - f(x) \leq \epsilon\}$, there exists a constant C such that $\mathcal{N}(\mathcal{X}_\epsilon, \ell, \epsilon') \leq C(\epsilon')^{-d}$, where d is the near-optimality dimension of function f and $\mathcal{N}(\mathcal{X}_\epsilon, \ell, \epsilon')$ is the ϵ'-cover number of the set $\mathcal{X}_\epsilon$ w.r.t. the dissimilarity ℓ.*

Remark 2. Some recent progress of solving centralized $\mathcal{X}$-armed bandit problem such as Shang, Kaufmann, and Valko (2019); Bartlett, Gabillon, and Valko (2019) have proposed an even weaker version of Assumption 2, i.e., the *local smoothness without a metric* assumption. Correspondingly, they define the complexity measure named *near-optimal dimension w.r.t. the partition $\mathcal{P}$*. However, it is highly non-trivial to directly adopt this weaker local smoothness assumption in the federated $\mathcal{X}$-armed bandit problem. The limited communications and the weak assumption will lead to continual sampling in the sub-optimal regions, and thus yielding large cumulative regrets. As a pioneer work in federated $\mathcal{X}$-armed bandit, we choose to use the slightly stronger assumptions in Bubeck et al. (2011) so that theoretical guarantees of our `Fed-PNE` algorithm can be successfully established. Weakening our set of assumption while keeping the regret bound guarantee is an interesting future work direction.

Algorithm and Analysis

The federated $\mathcal{X}$-armed bandit problem encounters several challenges, the core of which is to accommodate the heterogeneity among local objectives with limited communications. Hence, how to design an efficient communication pattern and construct an unbiased estimation of the global objective while taking advantage of the large number of clients is a crucial component in algorithmic design. Moreover, since local rewards are not instantaneously observable due to communication limitation, any algorithm that "uses instant rewords of each time step to estimate the optimal region with high confidence", e.g., `HOO` (Bubeck et al. 2011) and `HCT` (Azar, Lazaric, and Brunskill 2014), cannot be directly applied to this problem. Instead, an algorithm that gradually eliminates the sub-optimal regions in phases is preferred.

In this section, we propose the new algorithm to solve the above challenges, show its uniqueness compared with prior algorithms, and provide its theoretical analysis.

The `Fed-PNE` Algorithm

We propose the new Federated-Phased-Node-Elimination (`Fed-PNE`) algorithm, which consists of one client-side algorithm (Algorithm 1) and one server-side algorithm (Algorithm 2). The `Fed-PNE` algorithm runs in dynamic phases and it utilizes the hierarchical partition to gradually find the optimum by eliminating different regions of the domain. For a node $\mathcal{P}_{h,i} \in \mathcal{P}$, since its depth h and index i uniquely identifies the node, we will use (h, i) to index the nodes in the elimination and expansion process. We use $\mathcal{K}^p$ to denote the indices of active nodes that need to be sampled in phase p and $\mathcal{E}^p$ for the indices of nodes that need to be eliminated. To obtain a reward r over a node $\mathcal{P}_{h,i}$ (i.e., pull a node), the client evaluate the local objective at some x where x is either uniformly sampled from the node as in Bubeck et al. (2011) or some pre-defined point in the node as in Azar, Lazaric, and Brunskill (2014). The regret analysis will only be slightly different because of the smoothness assumption. In the theoretical analysis, we have used the latter strategy to derive our regret bound.

Algorithm Explanation: At initialization, the server starts from the root of the partition $\mathcal{K}^1 = \{(0, 1)\}$. At the beginning of each phase $p > 0$, the server expands the exploration tree as described in Algorithm 2 and the set $\mathcal{K}^p$ until the criterion $|\mathcal{K}^p| \tau_h \geq M$ is satisfied, where the threshold number τ_h is the minimum required number of times each node on depth h needs to be pulled, defined as $\tau_h := \left\lceil \frac{c^2 \log(c_1 T/\delta)}{\nu_1^2} \rho^{-2h} \right\rceil$ where c, c_1 are two absolute constants, and δ is the confidence (details in Lemma 2). The number of times $t_{m,h,i}$ each node $\mathcal{P}_{h,i}$ has to be sampled by each client m and the phase length $|\mathcal{T}^p|$ are then computed. This unique expansion criteria and sampling scheme

Algorithm 1: `Fed-PNE: m-th client`

1: **Input:** k-nary partition $\mathcal{P}$
2: **Initialize** $p = 0$
3: **while** not reaching the time horizon T **do**
4: Update $p = p + 1$
5: Receive $\{\mathcal{P}_{h,i}, t_{m,h,i}\}_{(h,i)\in\mathcal{K}^p}$ from the server
6: **for** $\mathcal{P}_{h,i}$ with $(h,i) \in \mathcal{K}^p$ **do**
7: Pull the node for $t_{m,h,i}$ times, receive rewards $\{r_{m,h,i,t}\}_{t=1}^{t_{m,h,i}}$
8: Calculate $\widehat{\mu}_{m,h,i} = \frac{1}{t_{m,h,i}} \sum_t r_{m,h,i,t}$
9: **end for**
10: Send the estimates $\{\widehat{\mu}_{m,h,i}\}_{(h,i)\in\mathcal{K}^p}$ to the server
11: **end while**

Algorithm 2: `Fed-PNE: server`

1: **Input:** k-nary partition $\mathcal{P}$, smooth parameters ν_1, ρ
2: **Initialize** $\mathcal{K}^1 = \{(0,1)\}, h = 0, p = 0$
3: **while** not reaching the time horizon T **do**
4: $p = p + 1; h = h + 1$
5: **while** $|\mathcal{K}^p|\tau_h \le M$ or $\tau_h \le 1$ **do**
6: $\mathcal{K}^p = \{(h'+1, ki-j) \mid \forall (h',i) \in \mathcal{K}^p, j < k\}$
Renew $h = h + 1$
7: **end while**
8: Compute the number $t_{m,h,i} = \lceil \frac{\tau_h}{M} \rceil$ and the phase length $|\mathcal{T}^p| = |\mathcal{K}^p| t_{m,h,i}$
9: Broadcast the set of nodes and pulled times $\{\mathcal{P}_{h,i}, t_{m,h,i}\}_{(h,i)\in\mathcal{K}^p}$ to every client m
10: Receive local estimates $\{\widehat{\mu}_{m,h,i}\}_{m\in[M],(h,i)\in\mathcal{K}^p}$ from the clients
11: **for** every $(h,i) \in \mathcal{K}^p$ **do**
12: Calculate $\widehat{\mu}_{h,i} = \frac{1}{M} \sum_{m=1}^{M} \widehat{\mu}_{m,h,i}$
13: **end for**
14: Compute $(h^p, i^p) = \arg\max_{(h,i)\in\mathcal{K}^p} \widehat{\mu}_{h,i}$
15: Compute the elimination set $\mathcal{E}^p$
16: Compute the new set of nodes $\mathcal{K}^{p+1} = \{(h+1, ki-j) \mid (h,i) \in (\mathcal{K}^p \setminus \mathcal{E}^p), j < k\}$
17: **end while**

guarantee four important things at the same time: (1) Every client samples every node at least one time so that the global objective is explored; (2) The empirical averages in line 12 of Algorithm 2 are unbiased estimators of the global function values for every node; (3) Every node is sampled enough number of times (larger than τ_h); (4) The waste of budget due to the limitation on communication is minimized. After the broadcast in line 9, every client receives $\{\mathcal{P}_{h,i}, t_{m,h,i}\}_{(h,i)\in\mathcal{K}^p}$ from the server.

Next, the clients perform the exploration and send only the empirical reward averages $\widehat{\mu}_{m,h,i}$ back to the server, as in Algorithm 1. The server then computes the best node, denoted by $\mathcal{P}_{h^p,i^p}$, and decides the elimination set $\mathcal{E}^p$ by the following selection criteria.

$$\mathcal{E}^p := \{(h,i) \in \mathcal{K}^p \mid \widehat{\mu}_{h,i} + b_{h,i} + \nu_1 \rho^h < \widehat{\mu}_{h^p,i^p} - b_{h^p,i^p}\} \tag{1}$$

where $b_{h,i} = c\sqrt{\log(c_1 T/\delta)/T_{h,i}}$ and $T_{h,i} = M t_{m,h,i}$. In other words, for any node $\mathcal{P}_{h,i}$ such that $(h,i) \in \mathcal{E}^p$, the function value of the global objective inside the node is much worse than the function value in the best node with high probability, and thus can be safely eliminated. The server then eliminate the bad nodes and proceed to the next phase with the new set $\mathcal{K}^{p+1}$, which consists of nodes that are children of un-eliminated nodes in the previous phase, as shown in line 15-16 in Algorithm 2.

Remark 3. `Fed-PNE` is very different from the traditional Phased-Elimination (`PE`) algorithm in multi-armed bandit(Lattimore and Szepesvári 2020), though both algorithms utilize the idea of successive elimination of the sub-optimal arms/nodes. Apart from the obvious uniqueness in the algorithm design such as line 5-8, 15-16 in Algorithm 2, `Fed-PNE` also introduces the new idea of "node elimination", which is based on the hierarchical partitioning of the parameter space. Even if we treat nodes in the partition as the "arms" in multi-armed bandit, `Fed-PNE` is still unique in the following aspects:

- `Fed-PNE` utilizes the hierarchical partition and gradually eliminate nodes on deeper layers that represent smaller and smaller regions in domain $\mathcal{X}$. The nodes can not be eliminated until the algorithm reaches their layer in the partition. In other words, the eliminated nodes are different in nature, whereas in multi-armed bandit problem, the arms have equal roles and can be eliminated in any phase;
- While eliminating the sub-optimal regions, `Fed-PNE` also explores deeper in the partition and splits one node into multiple nodes, which means that the number of nodes to be sampled may increase instead of decrease as p increases. However, the number of remaining arms never increases in `PE`. This feature also brings more difficulty to the analysis of `FedPNE` because the phase length is dynamic instead of fixed;
- The elimination criteria in Eqn. (1) is carefully designed so that non-optimal nodes are gradually eliminated. The design takes account of not only the Upper-Confidence Bound (UCB) terms $b_{h,i}$ for statistical uncertainty, but also the smoothness term $\nu_1 \rho^h$, which reflects for the variation of the objective function inside one node.

Remark 4. Compared with centralized $\mathcal{X}$-armed bandit algorithms such as `HOO` and `HCT`, our algorithm is also unique in the sense that none of them can deal with the federated, heterogeneous learning setting. The collaboration scheme and the length of each phase $\mathcal{T}^p$ is carefully designed so that the communication to the server is effective. It is worth mentioning that our algorithm requires the parameters ν_1, ρ as part of the input, which measures how fast the diameter of a node shrinks in the partition. These parameters are important because they characterize the smoothness of the global objective and we need them to determine the threshold τ_h and the elimination set $\mathcal{E}^p$. This information is crucial to ensure that validity of cumulative regret analysis theorems even for

centralized $\mathcal{X}$-armed bandit problems. Most of existing $\mathcal{X}$-armed bandit algorithms, such as Bubeck et al. (2011), Azar, Lazaric, and Brunskill (2014) and Li et al. (2023b), require these parameters.

Theoretical Analysis

We provide the upper bound on the expected cumulative regret of the proposed `Fed-PNE` algorithm as follows, which exhibits our theoretical advantage over non-federated algorithms.

Theorem 1. *Suppose that $f(x)$ satisfies Assumption 2, and d is the near-optimality dimension of the global objective f as defined in Assumption 3. Setting $\delta = 1/M$, we have the following upper bound on the expected cumulative regret of the `Fed-PNE` algorithm.*

$$\mathbb{E}[R(T)] \leq C_1 M^{1-\frac{1}{2\log_k \rho}} + C_2 M^{\frac{d+1}{d+2}} T^{\frac{d+1}{d+2}} (\log(MT))^{\frac{1}{d+2}}$$

where C_1 and C_2 are two absolute constants that do not depend on M and T. Moreover, the number of communication rounds of `Fed-PNE` scales as $\widetilde{\mathcal{O}}(M \log T)$

Remark 5. The proof of the above theorem and the exact values of the two constants are relegated to the Appendix. Theorem 1 displays a desirable regret upper bound for the `Fed-PNE` algorithm because the first term on the right-hand side only depends on M and it is a cost due to federation across all the clients. When T is sufficiently large compared with M [1], the second term dominates the bound and it depends sub-linearly on both the number of rounds T and the number of agents M, which means that the algorithm converges to the optimum of the global objective. Moreover, the average cumulative regret of each client is of order $\widetilde{\mathcal{O}}\left(M^{-\frac{1}{d+2}} T^{\frac{d+1}{d+2}}\right)$, which represents that increasing the number of clients helps reducing the regret of each client, and thus validates the effectiveness of federation. Compared with the regret of centralized $\mathcal{X}$-armed bandit algorithms, i.e., $\widetilde{\mathcal{O}}\left(T^{\frac{d+1}{d+2}}\right)$ (Bubeck et al. 2011; Azar, Lazaric, and Brunskill 2014), the average regret bound of our algorithm is smaller when M is large, which means that our algorithm is faster.

Remark 6. When T is relatively small, the first term in Theorem 2 dominates the regret bound, yielding a super-linear dependence w.r.t. M (but no dependence on T). Such a rate is mainly due to the lack of information and thus (potentially) inefficient sampling in the early stage, especially when there are too many clients. For example, when we explore the shallow layers, i.e., h is small, in the partition at the beginning of the search, the total number of pulls of the node $\mathcal{P}_{h,i}$, i.e., $T_{h,i} = M t_{m,h,i}$, could be much larger than the required threshold τ_h.

Remark 7. (Communication Rounds and Information) Moreover, the number of communication ***rounds*** in Theorem 1 only depends logarithmically on the time horizon T, showing that there are no frequent communications between the server and the clients during the federated learning process. Moreover, only the mean rewards are shared instead of all the rewards. Therefore, our algorithm successfully protects data confidentiality to certain extent and saves the communication cost. Similar dependence is observed in prior federated bandit works (Shi and Shen 2021a) (Huang et al. 2021).

It's also worth mentioning that since the number of nodes $|\mathcal{K}^p|$ could increase when we increase the phase number p, a better measure of the communication cost is the amount of ***information*** communicated instead of the number of ***rounds***. In this measure, the communication cost depends on the near-optimality dimension d (Assumption 3). If $d = 0$, it is easy to show that the communicated information is also of logarithmic order $\widetilde{\mathcal{O}}(M \log T)$. As mentioned by prior research, $d = 0$ is the most commonly observed case for blackbox objectives (Bubeck et al. 2011; Valko, Carpentier, and Munos 2013). However, when $d > 0$, the communicated information could be as large as $\mathcal{O}(MT^{\frac{d}{d+2}})$ because both $|\mathcal{K}^p|$ and τ_{h^p} can exponentially increase when we increase p. In the Appendix, we show that such dependence on T is unfortunately unavoidable by any algorithm that has the same regret rate as `Fed-PNE`, and thus our cost is already optimal.

Remark 8. (Privacy) The privacy guarantee in the main text refers to the limited communications between the server and the clients as in Shi and Shen (2021a); Shi, Shen, and Yang (2021b); Huang et al. (2021), instead of the quantitative privacy measures such as differential privacy (DP) (Dwork et al. 2010). However, since `Fed-PNE` only requires communications of the average rewards in very few rounds, it would be easy to guarantee differential privacy by adding Laplacian/Gaussian noise to the rewards in the `Fed-PNE` algorithm. In the Appendix, we prove our claim by presenting the differentially-private version of our algorithm (`DP-Fed-PNE`) and its analysis.

Regret Lower Bound

To show the tightness of the regret bound in Theorem 1, we provide the following lower bound.

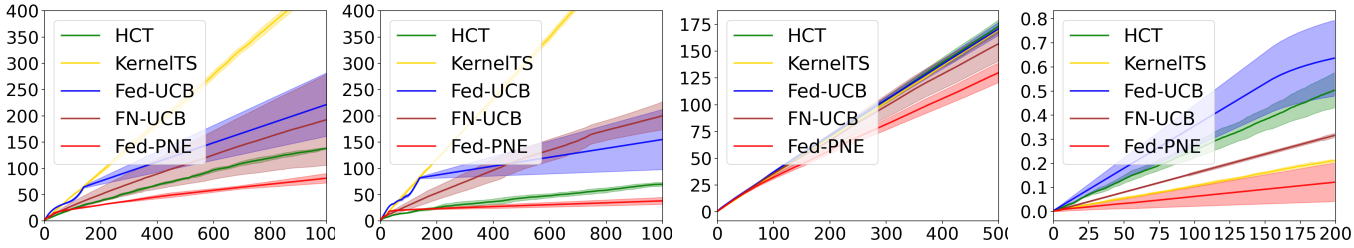
Theorem 2. *There exists an instance of the federated $\mathcal{X}$-armed bandit problem satisfying Assumptions 2 and 3 such that the expected cumulative regret of any multi-client algorithm is lower bounded as $\mathbb{E}[R(T)] = \Omega(M^{\frac{d+1}{d+2}} T^{\frac{d+1}{d+2}})$.*

Remark 9. The proof of the above theorem is provided in the Appendix. Theorem 2 essentially claims an $\Omega(M^{\frac{d+1}{d+2}} T^{\frac{d+1}{d+2}})$ regret lower bound for the M-client, T-round federated $\mathcal{X}$-armed bandit problem, even if we allow instantaneous and unlimited number of communications between the clients and the server, i.e., the clients and the server can communicate in every round about the reward of any $x_{m,t}$ they choose. Therefore, the regret upper bound in Theorem 1 is asymptotically unimprovable if we ignore the logarithmic term $\mathcal{O}(\log(MT)^{\frac{1}{d+2}})$.

Experiments

We empirically evaluate the proposed `Fed-PNE` algorithm on both synthetic functions and real-world datasets.

[1] Specifically, when $T^{d+1} > M^{1-(d+2)/(2\log_k \rho)}$ is satisfied.



(a) Garland (b) DoubleSine (c) Landmine (d) COVID

Figure 2: Cumulative regret of different algorithms over the number of rounds on the synthetic functions and the real-life datasets. Unlimited communications are allowed for centralized algorithms.

We compare `Fed-PNE` with centralized $\mathcal{X}$-armed bandit algorithm `HCT` (Azar, Lazaric, and Brunskill 2014), centralized kernelized bandit algorithm `KernelTS` (Chowdhury and Gopalan 2017), federated multi-armed bandit algorithm `Fed1-UCB` (Shi and Shen 2021a), and federated neural bandit algorithm `FN-UCB` (Dai et al. 2023). Additional details of algorithm implementations and more comparisons against other blackbox optimization algorithms such as Bayesian Optimization and Batched Bayesian Optimization algorithms, are provided in the Appendix.

Remark 10. For the federated algorithms (`Fed-PNE`, `Fed1-UCB`, `FN-UCB`), we plot the average cumulative regret per client against the rounds. For the centralized algorithms (`HCT`, `KernelTS`), we plot the cumulative regret on the global objective of each task against the number of evaluations. Such a comparison is fair in terms of overall computation resource, since the global objective itself is not directly accessible, and we can view one evaluation of global objective as the result of instant public communications of all local objective evaluations in one round. For all the curves presented in this section (and the numerical results in the appendix), they are averaged over 10 independent runs with shaded area standing for the 1 standard deviation.

Synthetic Dataset. We evaluate the algorithms on two synthetic functions that are commonly used in $\mathcal{X}$-armed bandit problem, which are the Garland function and the DoubleSine function, both defined on $\mathcal{X} = [0, 1]$. These two functions are well-known for their large number of local optimums. The randomly perturbed versions of these two functions are used as the local objective while the averages of the local objectives are used as the global objective. The average cumulative regret of different algorithms are provided in Figure 2(a) and 2(b). As can be observed in the figures, `Fed-PNE` has the smallest cumulative regret.

Landmine Detection. We federatedly tune the hyperparameters of machine learning models fitted on the Landmine dataset (Liu, Liao, and Carin 2007), where the features of different locations on multiple landmine fields extracted from radar images are used to detect the landmines. Following the setting of Dai, Low, and Jaillet (2020), each client only has the access to the data of one random field, and trains a support vector machine with the RBF kernel parameter chosen from [0.01, 10] and the L_2 regularization parameter chosen from $[10^{-4}, 10]$. The local objectives and the global objective are the AUC-ROC scores on the local landmine field and all the landmine fields respectively. The average cumulative regret of different algorithms are provided in Figure 2(c). As can be observed in the figures, our algorithm achieved smallest cumulative regret and thus the best performance.

COVID-19 Vaccine Dosage Optimization. In combat to the pandemic, we optimize the vaccine dosage in epidemiological models of COVID-19 to find the best fractional dosage for the overall population following Wiecek et al. (2022). Using fractional dosage of the vaccines will make them less effective, but at the same time more people get the chance of vaccination and thus can possibly accelerate the process of herd immunity. In our experimental setting, the local objectives are the final infectious rate of different countries/regions. Different countries have different parameters such as population size and the number of ICU units, and thus make the objectives heterogeneous. The results are shown in Figure 2(d). Our algorithm also achieves the fastest convergence.

Discussions and Conclusions

In this work, we establish the framework of federated $\mathcal{X}$-armed bandit problem and propose the first algorithm for such problems. The proposed `Fed-PNE` algorithm utilizes the intrinsic structure of the global objective inside the hierarchical partitioning and achieves desirable regret bounds in terms of both the number of clients and the evaluation budget. Meanwhile it requires only logarithmic communications between the server and the clients, protecting the privacy of the clients. Both theoretical analysis and the experimental results show the advantage of `Fed-PNE` over centralized algorithms and prior federated multi-armed bandit algorithms. Many interesting future directions can be explored based on the framework proposed in this work. For example, other summary statistics of the client-wise data can potentially accelerate the proposed algorithm, such as the usage of empirical variance in Li et al. (2023b). Moreover, the current algorithm still needs a the weak lipschitzness assumption. Whether the weakest assumption in the literature of $\mathcal{X}$-armed bandit, i.e., the *local smooth without a metric* assumption proposed by Grill et al. (2015) can be used to prove similar regret guarantees remains challenging.

Acknowledgements

Jean Honorio gratefully acknowledges the support of the National Science Foundation (DMS-2134209). Guang Lin gratefully acknowledges the support of the National Science Foundation (DMS-2053746, DMS-2134209, and ECCS-2328241), and U.S. Department of Energy (DOE) Office of Science Advanced Scientific Computing Research program DE-SC0021142, DE-SC0023161, and the Uncertainty Quantification for Multifidelity Operator Learning (MOLUcQ) project (Project No. 81739)

References

Azar, M. G.; Lazaric, A.; and Brunskill, E. 2014. Online stochastic optimization under correlated bandit feedback. In *International Conference on Machine Learning*, 1557–1565. PMLR.

Bartlett, P. L.; Gabillon, V.; and Valko, M. 2019. A simple parameter-free and adaptive approach to optimization under a minimal local smoothness assumption. In *30th International Conference on Algorithmic Learning Theory*.

Bastani, H.; and Bayati, M. 2020. Online decision making with high-dimensional covariates. *Operations Research*, 68(1): 276–294.

Bubeck, S.; Munos, R.; Stoltz, G.; and Szepesvári, C. 2011. χ-Armed Bandits. *Journal of Machine Learning Research*, 12(46): 1655–1695.

Chen, N.; and Gallego, G. 2022. A Primal–Dual Learning Algorithm for Personalized Dynamic Pricing with an Inventory Constraint. *Mathematics of Operations Research*.

Chowdhury, S. R.; and Gopalan, A. 2017. On Kernelized Multi-armed Bandits. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, 844–853. PMLR.

Dai, Z.; Low, B. K. H.; and Jaillet, P. 2020. Federated Bayesian Optimization via Thompson Sampling. In *Advances in Neural Information Processing Systems*, volume 33, 9687–9699. Curran Associates, Inc.

Dai, Z.; Low, B. K. H.; and Jaillet, P. 2021. Differentially Private Federated Bayesian Optimization with Distributed Exploration. In *Advances in Neural Information Processing Systems*, volume 34, 9125–9139. Curran Associates, Inc.

Dai, Z.; Shu, Y.; Verma, A.; Fan, F. X.; Low, B. K. H.; and Jaillet, P. 2023. Federated Neural Bandits. In *The Eleventh International Conference on Learning Representations*.

Dubey, A.; and Pentland, A. S. 2020. Differentially-Private Federated Linear Bandits. In *Advances in Neural Information Processing Systems*, volume 33, 6003–6014. Curran Associates, Inc.

Dwork, C.; Naor, M.; Pitassi, T.; and Rothblum, G. 2010. Differential Privacy Under Continual Observation. In *STOC '10: Proceedings of the 42nd ACM symposium on Theory of computing*, 715–724. ACM.

Feng, Y.; Huang, z.; and Wang, T. 2022. Lipschitz Bandits with Batched Feedback. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 19836–19848. Curran Associates, Inc.

Frazier, P. I. 2018. A Tutorial on Bayesian Optimization.

Grill, J.-B.; Valko, M.; Munos, R.; and Munos, R. 2015. Black-box optimization of noisy functions with unknown smoothness. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Huang, R.; Wu, W.; Yang, J.; and Shen, C. 2021. Federated Linear Contextual Bandits. In *Advances in Neural Information Processing Systems*.

Huang, R.; Zhang, H.; Melis, L.; Shen, M.; Hejazinia, M.; and Yang, J. 2023. Federated Linear Contextual Bandits with User-Level Differential Privacy. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org.

Jones, D.; Perttunen, C.; and Stuckman, B. 1993. Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications*, 79(1): 157–181.

Khodak, M.; Tu, R.; Li, T.; Li, L.; Balcan, M.-F. F.; Smith, V.; and Talwalkar, A. 2021. Federated Hyperparameter Tuning: Challenges, Baselines, and Connections to Weight-Sharing. In *Advances in Neural Information Processing Systems*, volume 34, 19184–19197. Curran Associates, Inc.

Kleinberg, R.; Slivkins, A.; and Upfal, E. 2008. Multi-Armed Bandits in Metric Spaces. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, 681–690. New York, NY, USA: Association for Computing Machinery. ISBN 9781605580470.

Lattimore, T.; and Szepesvári, C. 2020. *Bandit algorithms*. Cambridge University Press.

Li, T.; Song, L.; and Fragouli, C. 2020. Federated recommendation system via differential privacy. In *2020 IEEE International Symposium on Information Theory (ISIT)*, 2592–2597. IEEE.

Li, W.; Li, H.; Honorio, J.; and Song, Q. 2023a. PyXAB – A Python Library for $\mathcal{X}$-Armed Bandit and Online Blackbox Optimization Algorithms.

Li, W.; Wang, C.-H.; Cheng, G.; and Song, Q. 2023b. Optimum-statistical Collaboration Towards General and Efficient Black-box Optimization. *Transactions on Machine Learning Research*.

Liu, Q.; Liao, X.; and Carin, L. 2007. Semi-Supervised Multitask Learning. In *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc.

McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.

Munos, R. 2011. Optimistic Optimization of a Deterministic Function without the Knowledge of its Smoothness. In *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc.

Shang, X.; Kaufmann, E.; and Valko, M. 2019. General parallel optimization a without metric. In *Algorithmic Learning Theory*, 762–788.

Shariff, R.; and Sheffet, O. 2018. Differentially Private Contextual Linear Bandits. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.

Shi, C.; and Shen, C. 2021a. Federated Multi-Armed Bandits. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(11): 9603–9611.

Shi, C.; Shen, C.; and Yang, J. 2021b. Federated Multi-armed Bandits with Personalization. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, 2917–2925. PMLR.

Tossou, A. C. Y.; and Dimitrakakis, C. 2016. Algorithms for Differentially Private Multi-Armed Bandits. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, 2087–2093. AAAI Press.

Valko, M.; Carpentier, A.; and Munos, R. 2013. Stochastic Simultaneous Optimistic Optimization. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, 19–27. PMLR.

Wang, C.-H.; Li, W.; Cheng, G.; and Lin, G. 2022. Federated Online Sparse Decision Making.

Wang, Z.; Gehring, C.; Kohli, P.; and Jegelka, S. 2018. Batched Large-scale Bayesian Optimization in High-dimensional Spaces. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, 745–754. PMLR.

Wiecek, W.; Ahuja, A.; Chaudhuri, E.; Kremer, M.; Gomes, A. S.; Snyder, C. M.; Tabarrok, A.; and Tan, B. J. 2022. Testing fractional doses of COVID-19 vaccines. *Proceedings of the National Academy of Sciences*, 119(8): e2116932119.

Xu, X.; Xie, H.; and Lui, J. C. S. 2021. Generalized Contextual Bandits With Latent Features: Algorithms and Applications. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8): 4763–4775.

Zhu, Z.; Zhu, J.; Liu, J.; and Liu, Y. 2021. Federated bandit: A gossiping approach. In *Abstract Proceedings of the 2021 ACM SIGMETRICS/International Conference on Measurement and Modeling of Computer Systems*, 3–4.