Fairness and Privacy Guarantees in Federated Contextual Bandits

Extended Abstract

Sambhav Solanki
IIITH
Hyderabad, India
sambhav.solanki@research.iiit.ac.in

Sujit Gujar
IIITH
Hyderabad, India
sujit.gujar@iiit.ac.in

Shweta Jain
IIT Ropar
Ropar, India
shwetajain@iitrpr.ac.in

ABSTRACT

This paper studies the contextual multi-armed bandit problem with fairness and privacy guarantees in a federated setting. It proposes a collaborative algorithm, Fed-FairX-LinUCB that achieves sub-linear fairness regret and can be adapted to ensure differential privacy. The key challenge is designing a communication protocol that balances privacy and regret. The proposed protocol achieves both sub-linear fairness regret and effective use of privacy budget. Experiments validates the efficacy of both Fed-FairX-LinUCB and its private counterpart, Priv-FairX-LinUCB

KEYWORDS

Multi-Armed Bandits; Fairness; Differential Privacy; Federated Learning

ACM Reference Format:
Sambhav Solanki, Sujit Gujar, and Shweta Jain. 2024. Fairness and Privacy Guarantees in Federated Contextual Bandits: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

The *bandit* problem [2] deals with the trade-off between exploration and exploitation, with applications in crowdsourcing, recommendation systems, sponsored search, smart grids etc [1, 3–6, 8–12]. This paper tackles *contextual MAB* in a federated setting. Linear contextual bandits associate contextual features with actions, modeling rewards as a linear combination of these features. Existing works primarily on linear bandits assume a single-agent scenario and aim to maximize reward. However, this work considers ensuring fairness among actions in a federated setting.

Traditional bandit approaches exhibit a "winner takes all" behaviour [15], where they consistently favor the optimal action at the expense of other actions, leading to starvation and action set reduction. We leverage *fairness of exposure* in federated contextual bandits, ensuring a proportional selection of actions based on their merits, extending it to a private setting. Our contributions are:

- **Novel communication protocol**: We propose a communication protocol enabling collaboration while achieving sub-linear fairness regret (fairness for actions) and minimizing privacy leakage.
- **Differentially private algorithm**: We extend the protocol to a differentially private algorithm, ensuring agent privacy with bounded fairness regret.
- **Theoretical and empirical validation**: We prove theoretical guarantees and demonstrate through experiments that both algorithms outperform non-collaborative learners.

The notational use is limited, and theoretical results are simplified for clarity. All the details are available in the full version [13].

2 MODEL PRELIMINARIES

This work explores federated contextual bandits with fairness and privacy guarantees, where multiple agents (m) are collaboratively learning about different actions. In each round, each agent observes a set of context vectors and chooses an action based on it. The key objective is to ensure fairness amongst all actions – offering every action a fair chance to be chosen, proportional to its potential reward. Specifically, we aim to learn a policy that selects actions with probabilities proportional to their merit. Note that the objective here is to learn the fair policy itself rather than a fairness constrained optimal-reward policy.

In a single-agent MAB setting, Wang et al. [15] proposes the FairX-LinUCB algorithm to achieve fairness of exposure as defined above. The key idea of their algorithm is to construct a confidence region around reward parameters at every round t. Subsequently, the proposed algorithm then optimistically selects parameters from the confidence region, and the selection policy is constructed using this optimistic regression estimate.

Privacy plays a crucial role in this setting. We ensure that each agent's context and reward data remain confidential. While federated learning provides a level of privacy, we further define a differential privacy constraint for our setting. It ensures that any singular change in an agent selection history should not affect the selection policy of any other agent by much.

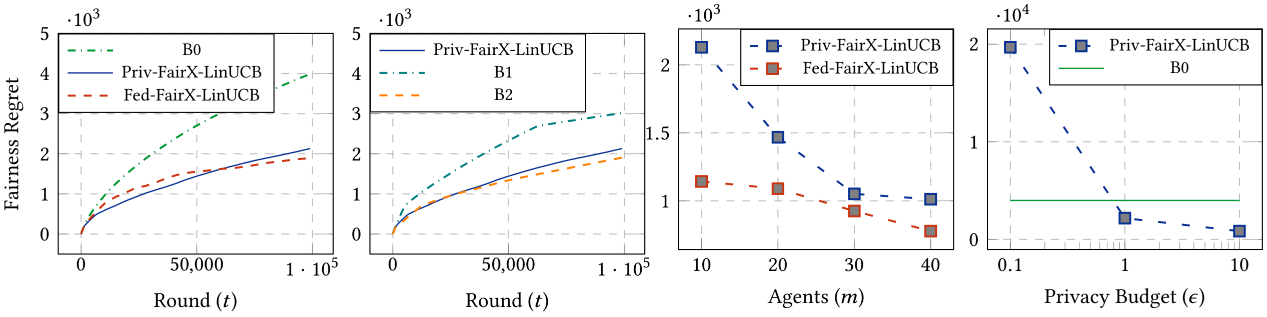
3 FED-FAIRX-LINUCB AND PRIV-FAIRX-LINUCB

The communication protocol currently used in federated bandits literature is not suitable for achieving bounded fairness regret. It is important to limit the number of communication rounds and maintain a constrained gap between communication instances in order to ensure both bounded fairness regret and scalability with private methods.

3.1 Multi-Agent Fair Contextual Bandit Algorithm

For any agent i, at round t, let the last synchronization round take place at instant t'. Then, there exist two sets of parameters. The

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).



(a) **(b)** **(c)** **(d)**

Figure 1: (a) Exp 1 : Fairness Regret vs. Rounds for single-agent baseline and proposed federated learning algorithms (m=10) (b) Exp 2 : Fairness Regret vs. Rounds for different communication protocol baselines and proposed algorithms (m=10) (c) Exp 3 : Fairness Regret trend w.r.t. number of agents (t=100,000) (d) Exp 4 : Fairness Regret trend w.r.t. privacy budget (t=100,000)

first set of parameters is the set of all observations made by all agents till round t'. Secondly, each agent has access to its own observations since the last communication round. The agents use combined parameters for estimating a linear regression estimate.

If the agents were to communicate in every round without any optimization, they could enhance their fairness regret by order of $O(1/\sqrt{m})$, where m is the number of agents. However, communicating at every round results in inefficiencies and potential privacy breaches. To address these concerns, our algorithm suggests a communication strategy limiting the order of number of total communications. In our proposed approach, we suggest that the agents communicate with increasing intervals between two consecutive communication rounds during the first few rounds. Subsequently, they communicate only after every certain fixed number of rounds have passed since last communication. Rapid communication in the initial rounds proves beneficial in practice, considering the trend in regret is sublinear in T. Concurrently, the number of communication rounds and the gap between the communication rounds remain bounded.

Theorem 1. *With high probability, Fed-FairX-LinUCB achieves a fairness regret of* $\tilde{O}\left(\sqrt{\beta}\sqrt{mTd\log(1+\frac{T}{d}) + m^2d^3\log^3(1+\frac{T}{d})}\right)$ *under mild assumptions.*

Here d is the dimension of the context vectors and β is the carefully chosen confidence interval.

3.2 Multi-Agent Fair and Private Contextual Bandit Algorithm

The key difference between the algorithm we propose for the private setting, Priv-FairX-LinUCB, and non-private settings, Fed-FairX-LinUCB, lies in the communication perturbation. In a non-private setting, we communicate exact observations about context and reward to all other agents. However, for the private setting, we must carefully add perturbation to satisfy the differential privacy constraints mentioned in section 2.

To achieve privacy, we introduce a privatized version of the synchronization process amongst the agents. We do so by using a privatizer routine, which uses a tree-based mechanism to communicate while limiting the noise addition.

Theorem 2. *With high probability, under mild assumptions, Priv-FairX-LinUCB achieves bounded fairness regret.*

4 EXPERIMENTAL ANALYSIS

Experimental Set-up. We generate a synthetic datasets. We consider context size equal to five. Additionally, we sample noise from a normal distribution centered around 0, to produce the reward observations. Similar to Wang et al. [15], we use merit function $f(\cdot) = e^{10\mu}$, which is a steep merit function. In each round, projected gradient descent is employed to solve the non-convex optimization problem.

The results presented are averaged over five runs each. The FairX-LinUCB algorithm, noted as $B0$, performs single-agent learning and allows comparison between federated learning (our proposed algorithms) and single-agent learning. The fairness regret in all experiments for FairX-LinUCB considers no communication between agents. Baseline $B1$ and $B2$ represent Priv-FairX-LinUCB with communication protocol replaced, using communication protocols as defined in [7] and [14] respectively.

5 CONCLUSION

This work studies federating contextual bandits with fairness goals. It proposes Fed-FairX-LinUCB a novel algorithm achieving sublinear fairness regret (compared to linear in non-federated settings). Notably, Fed-FairX-LinUCB also has a differentially private counterpart, Priv-FairX-LinUCB, with bounded fairness regret. Experiments validate significant improvements of Priv-FairX-LinUCB over non-federated methods while preserving privacy. We believe that alternate objectives, such as fairness, are essential for the practical adoption of the bandit framework in many use cases and that this work helps pave the way for other exciting works.

ACKNOWLEDGEMENT

The authors acknowledge financial support from the Department of Science & Technology, India, with grant number CRG/2022/004980.

REFERENCES

[1] Kumar Abhishek, Shweta Jain, and Sujit Gujar. 2020. Designing Truthful Contextual Multi-Armed Bandits based Sponsored Search Auctions. *arXiv preprint arXiv:2002.11349* (2020).

[2] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. 2002. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning* (2002).

[3] Sanjay Chandlekar, Shweta Jain, and Sujit Gujar. 2023. A Novel Demand Response Model and Method for Peak Reduction in Smart Grids - PowerTAC. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, 19th-25th August 2023, Macao, SAR, China*. ijcai.org, 3497–3504.

[4] Sanjay Chandlekar, Easwar Subramanian, and Sujit Gujar. 2023. Multi-armed Bandit Based Tariff Generation Strategy for Multi-agent Smart Grid Systems. In *International Workshop on Engineering Multi-Agent Systems*. Springer, 113–129.

[5] Debojit Das, Shweta Jain, and Sujit Gujar. 2022. Budgeted Combinatorial Multi-Armed Bandits. In *21st International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2022, Auckland, New Zealand, May 9-13, 2022*. 345–353.

[6] Ayush Deva, Kumar Abhishek, and Sujit Gujar. 2021. A Multi-Arm Bandit Approach To Subset Selection Under Constraints. In *AAMAS '21: 20th International Conference on Autonomous Agents and Multiagent Systems, Virtual Event, United Kingdom, May 3-7, 2021*. ACM, 1492–1494.

[7] Abhimanyu Dubey and AlexSandy' Pentland. 2020. Differentially-private federated linear bandits. *Advances in Neural Information Processing Systems* (2020).

[8] Shweta Jain and Sujit Gujar. 2020. A Multiarmed Bandit Based Incentive Mechanism for a Subset Selection of Customers for Demand Response in Smart Grids. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2046–2053.

[9] Shweta Jain, Sujit Gujar, Satyanath Bhat, Onno Zoeter, and Yadati Narahari. 2018. A quality assuring, cost optimal multi-armed bandit mechanism for expertsourcing. *Artificial Intelligence* 254 (2018), 44–63.

[10] Shweta Jain, Balakrishnan Narayanaswamy, and Y. Narahari. 2014. A Multiarmed Bandit Incentive Mechanism for Crowdsourcing Demand Response in Smart Grids. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*, Carla E. Brodley and Peter Stone (Eds.). AAAI Press, 721–727.

[11] Akash Das Sharma, Sujit Gujar, and Y Narahari. 2012. Truthful multi-armed bandit mechanisms for multi-slot sponsored search auctions. *Current Science* (2012), 1064–1077.

[12] Akansha Singh, P. Meghana Reddy, Shweta Jain, and Sujit Gujar. 2021. Designing Bounded Min-Knapsack Bandits Algorithm for Sustainable Demand Response. In *PRICAI 2021: Trends in Artificial Intelligence - 18th Pacific Rim International Conference on Artificial Intelligence, PRICAI 2021, Hanoi, Vietnam, November 8-12, 2021, Proceedings, Part I*, Vol. 13031. Springer, 3–17.

[13] Sambhav Solanki, Shweta Jain, and Sujit Gujar. 2024. Fairness and Privacy Guarantees in Federated Contextual Bandits. arXiv:2402.03531 [cs.LG]

[14] Sambhav Solanki, Samhita Kanaparthy, Sankarshan Damle, and Sujit Gujar. 2022. Differentially Private Federated Combinatorial Bandits with Constraints. *arXiv preprint arXiv:2206.13192* (2022).

[15] Lequn Wang, Yiwei Bai, Wen Sun, and Thorsten Joachims. 2021. Fairness of exposure in stochastic bandits. In *International Conference on Machine Learning*.