TOWARDS FAIRNESS CONSTRAINED RESTLESS MULTI-ARMED BANDITS: A CASE STUDY OF MATERNAL AND CHILD CARE DOMAIN

Anonymous authors
Paper under double-blind review

ABSTRACT

Restless multi-armed bandits (RMABs) are widely used for resource allocation in dynamic environments, but they typically do not consider fairness implications. This paper introduces a fairness-aware approach for offline RMABs. We propose a Kullback-Leibler (KL) divergence-based fairness metric to quantify the discrepancy between the selected and the overall population. This is incorporated as a regularizer into the soft whittle index optimization. We evaluate our fairness-aware algorithm on a real-world RMAB dataset where initial results suggest that our approach can potentially improve fairness while preserving solution quality.

1 INTRODUCTION AND RELATED WORK

Restless multi-armed bandits (RMABs) have emerged as a powerful framework for optimizing resource allocation in dynamic environments with applications like machine maintenance (Abbou & Makis, 2019), anti-poaching (Qian et al., 2016), healthcare (Verma et al., 2023; Deo et al., 2013). Here, we study fairness in the context of offline RMAB problems with unknown transition dynamics but with given correlated arm features. Such problems are solved by learning the relation between arm features and transition dynamics to predict the dynamics to proceed to planning.

Wang et al. (2023) proposed a decision focused learning (DFL) solution for approaching this problem using a soft whittle index solution for planning which outperformed other baselines. However, the resulting allocation has no considerations of fairness. We propose Kullback–Leibler (KL) divergence between the total population versus the selected on the basis of their distribution across a sensitive feature(s) to capture the disparity between these two distributions as a measure of fairness. We incorporate this metric as a regularizer in the optimization in soft whittle index policy. This regularization produces a fairness constrained optimisation in soft topk selection in soft whittle index policy, thereby, making RMAB fair.

There have been efforts towards incorporating fairness in RMAB but they are not extendable to decision focused learning. Zehlike et al. (2017) presented a fair top-k algorithm which could potentially be incorporated post whittle index computation but since this version of top-k is not differentiable, it cannot be extended to contexts such as Wang et al. (2023). Killian et al. (2023) proposed a group level equitable objective optimisation for RMAB but this method provides guarantee on computed valuation, not action on a group of arms. Li & Varakantham (2022b)'s work on soft fairness fights starvation of arms but does not provide guarantee on the arms (groups) either. Li & Varakantham (2022a) and Herlihy et al. (2023) mandate a minimum probability for all groups but this does not balance the representation among the selected arms.

With integrated fairness in RMAB, we experimented on a real-world RMAB dataset pertaining to the scheduling of service calls for improving maternal and child health, as managed by ARMMAN (ARMMAN, 2008). Fairness here was assessed based on the ARMMAN recognized sensitive feature of income. We compared our fairness-aware algorithm to a baseline approach in importance sampling-based evaluation. While the baseline method achieved slightly superior performance, it exhibited degraded fairness. Conversely, our fairness-aware approach achieved lower performance but with substantially improved fairness. These early assessments present a promising trade-off between the two objectives that can be maneuvered per the system requirement.

2 METHOD

Figure 1: System Overview

Restless Multi-Armed Bandits (RMAB): The RMAB model here has N independent arms where each arm is a 2-action Markov Decision Process (MDP) defined as {S, A, R, P}. S is the state space, A is the action space, R is the reward function $R : S \times A \times S \rightarrow R$, P is the transition probability $P(s, a, s')$ where $(s, s') \in S, a \in A$. The policy function $\Pi : S \rightarrow A$ maps states to action. We adopt the Whittle solution approach for solving the RMAB followed by top K selection (figure:1). The whittle index is the infimum amount of extra reward that would make the planner indifferent among the actions. We employ the soft whittle index by (Wang et al., 2023) using soft top k selection instead of the top k algorithm, making the approach deployable in a decision focused learning setup.
KL Divergence: We introduce KL divergence as a regularizer in the optimization in soft top k algorithm. DFL requires a differentiable pathway for backpropagation. KL divergence is compatible to this requirement due to its inherent differentiability. The soft top k selection is built on optimal transport formulation (refer (Xie et al., 2020) for details) to which we add a regularizer here:

$$\Gamma^{*,\epsilon,\lambda} = \arg\min_{\Gamma \geq 0} < C, \Gamma > +\epsilon * H(\Gamma) + \lambda * KLD \quad \text{s.t.} \quad \Gamma \mathbf{1}_m = \mu, \Gamma^T \mathbf{1}_n = v$$

where $KLD =$ KL Divergence$([u_i]_i, [v_i]_i)$, u_i refers to overall population's proportion in category i and v_i refers to selected population's proportion in category i of sensitive feature. C is the measure of need (cost) quantified as whittle index and Γ signifies whether a resource is allocated.

Experiment: We experimented with ARMMAN data[1] with income as the sensitive attribute. The beneficiaries (arms) are modeled as 2-state 2-action MDP, states as Engaging and Non-Engaging based on engagement with ARMMAN in a week. The action is active if a live service call is provided, otherwise passive. Given state s, action a, the reward is the engagement state $R(s, a) = s$ where the objective is to maximize expected reward over all beneficiaries in the long term, under the constraint of maximum K live service calls allowed due to limited budget of the NGO. We measure the solution quality with importance sampling as described in Wang et al. (2023). We experimented with λ to threshold fair regularization in optimisation, and found that without much decrease in performance, there was decrease in KL divergence, thereby, increasing fairness as presented:

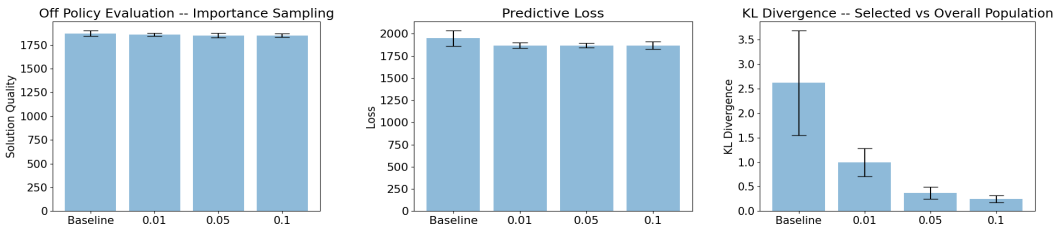

Figure 2: x-axis: λ. y-axis: solution quality, predictive loss, and KL divergence moving left to right.

3 CONCLUSION

This paper moves towards fairness in offline RMAB problems with unknown transition dynamics incorporating fairness constraints using Kullback-Leibler divergence. Our approach remains compatible with the Decision Focused Learning solution for RMAB. Early-stage experiments on a real-world Restless Multi-Armed Bandit dataset demonstrate that our fairness-aware algorithm can be potentially harnessed to improve fairness with minimal impact on solution quality.

[1]More details about the dataset can be found in subsection 'Real Dataset' in Wang et al. (2023)

URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2023 Tiny Papers Track.

REFERENCES

Abderrahmane Abbou and Viliam Makis. Group maintenance: A restless bandits approach. *INFORMS Journal on Computing*, 31, 06 2019. doi: 10.1287/ijoc.2018.0863.

ARMMAN. About armman. https://armman.org/about-us, 2008. Accessed: 2022-08-12.

Sarang Deo, Seyed Iravani, Tingting Jiang, Karen Smilowitz, and Stephen Samuelson. Improving health outcomes through better capacity allocation in a community-based chronic care model. *Operations Research*, 61, 12 2013. doi: 10.1287/opre.2013.1214.

Christine Herlihy, Aviva Prins, Aravind Srinivasan, and John P. Dickerson. Planning to fairly allocate: Probabilistic fairness in the restless bandit setting. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23. ACM, August 2023. doi: 10.1145/3580305.3599467. URL http://dx.doi.org/10.1145/3580305.3599467.

Jackson A. Killian, Manish Jain, Yugang Jia, Jonathan Amar, Erich Huang, and Milind Tambe. Equitable restless multi-armed bandits: A general framework inspired by digital health, 2023.

Dexun Li and Pradeep Varakantham. Efficient resource allocation with fairness constraints in restless multi-armed bandits, 2022a.

Dexun Li and Pradeep Varakantham. Towards soft fairness in restless multi-armed bandits, 2022b.

Yundi Qian, Chao Zhang, Bhaskar Krishnamachari, and Milind Tambe. Restless poachers: Handling exploration-exploitation tradeoffs in security domains. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, AAMAS '16, pp. 123–131, Richland, SC, 2016. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9781450342391.

Shresth Verma, Gargi Singh, Aditya Mate, Paritosh Verma, Sruthi Gorantla, Neha Madhiwalla, Aparna Hegde, Divy Thakkar, Manish Jain, Milind Tambe, and Aparna Taneja. Expanding impact of mobile health programs: Saheli for maternal and child care. *AI Magazine*, 10 2023. doi: 10.1002/aaai.12126.

Kai Wang, Shresth Verma, Aditya Mate, Sanket Shah, Aparna Taneja, Neha Madhiwalla, Aparna Hegde, and Milind Tambe. Scalable decision-focused learning in restless multi-armed bandits with application to maternal and child health, 2023.

Yujia Xie, Hanjun Dai, Minshuo Chen, Bo Dai, Tuo Zhao, Hongyuan Zha, Wei Wei, and Tomas Pfister. Differentiable top-k operator with optimal transport, 2020.

Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. Fa*ir: A fair top-k ranking algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17. ACM, November 2017. doi: 10.1145/3132847.3132938. URL http://dx.doi.org/10.1145/3132847.3132938.