Pratik Gajane | pratik.gajane@gmail.com | +31-633-313-415 | pratikgajane.github.io

Research Statement

Decision-making problems of sequential nature, where decisions made in the past may have an impact on future, are used in many important applications such as college admission, hiring, criminal justice, healthcare etc. In many real-world scenarios, the feedback available for such sequential decision making problems is incomplete or partial. Furthermore, the machine learning algorithms used to solve such problems are increasingly being used to make crucial decisions that affect human lives. Thus, it behooves us to construct machine learning solutions taking human issues like fairness and privacy into consideration. Motivated by these two themes, my research interests span across sequential decision making with partial feedback and machine learning with ethical considerations such as fairness and privacy. In general, I am interested in devising machine learning algorithms with strong mathematical foundations for problems encountered in real-life scenarios. The overarching vision of my research is as follows :

i) **Investigate the gaps in the respective lower and upper bounds for the performance of algorithms in various sequential decision making problems with partial feedback. Close the gap in the respective bounds by either improving the lower bound or by devising a method for which the upper bound matches the known lower bound.**

ii) **Devise machine learning algorithms with optimal mechanisms to guarantee a desired level of privacy and/or fairness in realistic settings.**

In the next couple of years, I envision myself as an worldwide expert on these two themes. Below, I expand on some of the salient related topics in my research.

Locally Differentially Private Reinforcement Learning and Learning from Preference Feedback

In [16, 18], we considered the problem of providing local differential privacy in stateless reinforcement learning. We proved a general lower bound and proposed near-optimal frequentist and Bayesian algorithms. This work advanced the state of the art as most of the previous related work in the literature of reinforcement learning focused on global differential privacy, which is a milder privacy notion. I also worked on reinforcement learning from preference feedback motivated from the observation that it is more practical to elicit pairwise preferences in tasks that involve human feedback. For this problem, we proved a general lower bound and proposed a near-optimal algorithm in [19].

Future work: For our work on locally differentially private stateless reinforcement learning, extensions to non-stationary and non-stochastic environments are worth exploring to make it more suitable to real-world applications. Moreover, a tighter lower bound for the fixed-budget variant of this problem (following the work on an analogous problem in [21]) would close the gap between the respective upper and lower bound and thus it is an attractive future direction for me. Considering privacy-preserving learning in the general reinforcement learning setting is also a prospective future work. For reinforcement learning from preference feedback, the current literature relies on the existence of particular notions of preference e.g.; Condorcet winner or Borda winner. However, other notions of preference are found in real-world datasets, hence formulating a solution for reinforcement learning with practical notions of preference is an attractive future direction.

Autonomous Exploration via Intrinsic Motivation

The problem of autonomous exploration has been studied under various names including intrinsic motivation, intrinsic reward and curiosity-driven learning. Conceptually, it can also be formulated as learning to reliably navigate an unknown environment with no external rewards or sparse rewards [22].

Future work: For autonomous exploration in an unknown environment, a generic meta-algorithm, which can use any efficient reinforcement learning algorithm as a black-box and turn it into an exploration algorithm, would be a significant contribution to the literature as it will formally verify the intuition that any effective reinforcement learning algorithm needs to explore its environment. This work will also provide theoretical backing for ideas that have been shown to preform well empirically in Go-Explore [23]. Our work proposing a generic meta-algorithm for autonomous exploration is already completed, but it is likely that the proved conversion bounds can be improved.

Fairness in Machine Learning

In [14], we argued that the problem of fairness with machine learning cannot be addressed without considering social issues such as unequal access to resources and social conditioning. This was one of the inaugural works arguing for an interdisciplinary approach to fairness-aware machine learning research. We provided theoretical as well as empirical critiques of fairness notions in the machine learning literature and explained how these critiques limit the suitability of the fairness formalizations to certain domains. Furthermore, we suggested two notions of distributive justice (one of which has been used in the foundations of human development paradigm by the United Nations) which address some of these critiques.

Future work: I would like to work toward incorporating the suggested notions from our work in [14] (and other notions from social sciences literature) into concrete fair-machine learning formulations. As we argued in [14], the suitability of these prospective formulations (unlike the current formulations) to domains in which natural endowments or social endowments or both impede positive outcomes for individuals makes the open problem of formulating them worthwhile.

Sequential Learning with Partial Feedback

Partial monitoring is a broad framework for formulating any problem which can be expressed as sequential learning with partial feedback. In [20], we proved that, in terms of performance bounds, generic partial monitoring algorithms are suboptimal compared to problem-specific algorithms.

Future work: An optimal universally applicable partial monitoring algorithm will be a significant contribution to the literature as many notable problems such as dynamic pricing, learning with expert advice and label efficient prediction can be formulated using partial monitoring. Thus, I would like to investigate if it is possible to have a generic partial monitoring algorithm with an optimal bound on the performance measure applicable to all partial monitoring problems.

Reinforcement Learning and Exploration in Non-stationary Environments

In contrast to previous related work, we achieved (nearly) optimal performance guarantees without knowing the number of changes in [11, 12] for non-stationary stateless reinforcement learning. In [10], we considered general reinforcement learning in non-stationary environments and proposed an algorithm with suitable performance guarantees, which were the first of their kind in general reinforcement learning setting. In [4], we devised a meta-algorithm for autonomous exploration in non-stationary environments and proved its sample complexity bound in terms of the number of changes.

Future work: Proving better sample complexity bounds for autonomous exploration in non-stationary environments using the approach given in [24] is a prospective future work. An initial attempt was recently made by a group of MSc students under my supervision.

Curiosity-driven Fairness in Reinforcement Learning

Fairness in reinforcement learning is a rapidly-growing emerging field (see a preliminary version of our extensive survey [2]). Curiosity-driven learning (also studied under the name of autonomous exploration in our work [4]) has been shown to increase equality in competitive resource allocation [25].

Future work: Following the promising results shown by [25], the goal here is to add appropriate curiosity-driven exploration to reinforcement learning algorithms leading to improved solutions for fairness-aware sequential decision making problems. Currently, I am supervising a MSc student on this topic.

Safe Learning and Learning from Batched/Delayed Feedback

In [9], we considered safe (i.e., risk averse) reinforcement learning and proposed a suitable performance measure. In [7, 8], we considered reinforcement learning with batched feedback, provided a policy-agnostic analysis and proved lower and upper bounds on the performance.

Future work: In some real-world applications, feedback about a decision is delayed and may arrive via partial rewards that are observed with different delays. Almost all the existing work in the reinforcement learning literature will be ineffective in these practically relevant scenarios. This motivates me to work on an effective reinforcement learning algorithm for generalized temporally-partitioned rewards applicable in such scenarios. Preliminary work in this direction was recently completed by a group MSc students under my supervision [1].

References

References [1] to [20] can be found in my CV.

[21] Alexandra Carpentier and Andrea Locatelli. Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Proceedings of the 29th Annual Conference on Learning Theory*, pages 590–604, 2016.

[22] Shiau Hong Lim and Peter Auer. Autonomous exploration for navigating in mdps. In *Proceedings of the 25th Annual Conference on Learning Theory*, pages 40.1–40.24, 2012.

[23] Adrien Ecoffet, Joost Huizinga, Joel Lehman, Kenneth O. Stanley, and Jeff Clune. First return, then explore. *Nature*, 590(7847):580–586, Feb 2021. ISSN 1476-4687. doi: 10.1038/s41586-020-03157-9. URL `https://doi.org/10.1038/s41586-020-03157-9`.

[24] Jean Tarbouriech, Matteo Pirotta, Michal Valko, and Alessandro Lazaric. Improved sample complexity for incremental autonomous exploration in mdps. In *Advances in Neural Information Processing Systems*, 2020.

[25] Bernadette Bucher, Siddharth Singh, Clélia de Mutalier, Kostas Daniilidis, and Vijay Balasubramanian. Curiosity increases equality in competitive resource allocation. 2020.