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

Research Statement

My research interests span across sequential decision making and machine learning with ethical consideration such as fairness and privacy. In sequential decision making, I am generally interested in devising machine learning algorithms with strong mathematical foundations which can work with incomplete forms of feedback often encountered in real-life scenarios. Secondly, I am also interested in interdisciplinary research with the aim of incorporating human aspects like fairness and privacy in machine learning. 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 incomplete feedback. Close the gap 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 privacy in realistic settings. Moreover, use findings from social science literature to drive forward the research on societal issues in machine learning.**

In the next couple of years 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.

Private learning and learning via relative feedback

Here we formulated sequential learning via multi-armed bandits which is a standard mathematical model for decision-making with uncertainty. i) We introduced a variant called *corrupt bandits* [15, 17] for the task of privacy protection in online recommendation systems. We proved a general lower bound and proposed a near-optimal frequentist algorithm and a near-optimal Bayesian algorithm. ii) I also worked on *dueling bandits* which is motivated from information retrieval systems where users provide implicit relative feedback. In [18], we proved a general lower bound and proposed a nearly-optimal algorithm for dueling bandits.

Future work: i) In our work on corrupt bandits, we considered stationary stochastic environment. Extensions to non-stationary and non-stochastic environments are worth exploring to make it more suitable to real-world applications. ii) The literature of adversarial dueling bandits relies on the existence of *Condorcet winner*, (an item being preferred when compared with any other item) or a *Borda winner* (an item with the highest Borda score - probability that it is preferred over another item chosen uniformly at random). But, in the literature of stochastic dueling bandits, other notions have been explored. This gap is of practical interest too as information retrieval datasets with other notions of preference arise in practice.

Sequential learning with partial feedback

Many problems such as dynamic pricing, learning with expect advice, label efficient prediction (as well as dueling bandits and corrupt bandits) can be formulated using partial monitoring which is a framework for sequential learning with partial feedback. In [19], we proved that, in terms of performance bounds, generic partial monitoring algorithms are suboptimal as compared to problem-specific algorithms.

Future work: 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.

Fairness in machine learning considering unequal access to resources and social conditioning

In [14], we provided theoretical as well as empirical critiques of fairness notions in the machine learning literature via their corresponding notions from the social sciences literature and explain 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.

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 [20].

Future work: I would like to work on adding curiosity-driven exploration to reinforcement learning algorithms leading to provably fair solutions for sequential decision-making problems.

Autonomous exploration and navigation

The problem of autonomous exploration can be formulated as learning to reliably navigate an unknown environment formalized by a Markov decision process (MDP) with no external rewards [22].

Future work: For autonomous exploration in MDPs, I would like to propose a generic meta-algorithm which can use any reinforcement learning algorithm with sub-linear regret as a black-box and turn it into an exploration algorithm. A preliminary paper achieving this goal is currently under review at AAAI-2023, but it is likely that the conversion bounds proved there can be improved.

Learning and exploration in non-stationary environments

In contrast to the previous related work, we achieved (nearly) optimal performance guarantees without knowing the number of changes in [11, 12] for non-stationary stochastic bandits. In [10], we considered reinforcement learning in non-stationary MDPs and proposed an algorithm with suitable performance guarantees, which were the first of their kind in the general reinforcement learning setting. In [4], we considered autonomous exploration in non-stationary MDPs and devised a meta-algorithm which can use any algorithm for the stationary variant of the problem as a subroutine. For the proposed meta-algorithm, we proved sample complexity bounds in terms of the number of changes in the environment.

Future work: I would like to work on proving better sample complexity bounds for autonomous exploration in non-stationary MDPs using the approach given in [23]. An initial attempt was recently made by a group of MSc students under my supervision.

Risk-averse learning and learning from batched/delayed feedback

In [9], we considered risk-averse bandits which can be considered as a bandit version of the classic gambler's ruin game. In [7, 8], we considered bandits with batched feedback and provided a policy-agnostic analysis and proved upper and lower bounds for the performance. We also provided experimental results on real data acquired from our industrial partner.

Future work: In many scenarios, feedback about a decision is delayed and may also arrive via partial rewards that are observed with different delays. To solve such problems, I would like to propose a solution to bandits formulation with generalized temporally-partitioned rewards. A preliminary work in this direction was recently completed by a group MSc students under my supervision [1].

References

References [1] to [19] can be found in my CV.

[20] Bernadette Bucher, Siddharth Singh, Clélia de Mutalier, Kostas Daniilidis, and Vijay Balasubramanian. Curiosity increases equality in competitive resource allocation. In the workshop for Bridging AI and Cognitive Science at ICLR, 2020.

[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, 2016.

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

[23] 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.