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

Research Statement

Contents

Research Vision

Decision making problems of sequential nature are used to model many important applications such as healthcare, finance, education, etc. In many of these real-world scenarios, the feedback available for such sequential decision making problems does not conform to the standard feedback assumed in the literature. Furthermore, the machine learning algorithms used to solve such problems are increasingly being used to make crucial decisions that affect human lives. Thus, as scientists, 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 non-standard feedback and machine learning with considerations such as fairness and privacy. The overarching vision of my research is based on the following two major themes:

i) **Devise optimal algorithms for sequential decision making with practically motivated forms of non-standard feedback.**

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

In both these directions, problems with non-stationary environments are of key interest. See some of my significant publications at the following link.

In the first major theme of my research, we devised provably optimal reinforcement learning algorithms for sequential decision making problems with non-standard feedback (such as preference feedback, delayed feedback, sparse feedback, noisy feedback, batch feedback, etc.) in stationary and non-stationary environments. We proposed frequentist as well as Bayesian algorithms, proved optimal theoretical performance guarantees, and conducted experiments on real-world datasets which show that our proposed algorithms work well in practice.

For fairness in machine learning, we provided theoretical and empirical critiques of the existing literature by juxtaposing it with the corresponding work in the social sciences literature. We suggested two notions of distributive justice which address some of the critiques. Furthermore, we studied the applicability of the existing work in fair machine learning in societal application domains and proposed guiding principles for further research and implementation.

For privacy in machine learning, we devised algorithms providing a stronger notion of privacy called *local differential privacy* which improved on the state-of-the-art and proved best-in-class performance bounds for the utility of our algorithms. We also provided an optimal mechanism using which a user can achieve the desired trade-off between privacy and utility.

Research themes such as human-AI interaction, explainable AI, biologically inspired autonomous learning and constrained/safe learning also form a part of my research agenda. The application domains of my work include application areas including healthcare, finance, employment, education, supply chain management, recommender systems, and management of electrical distribution networks.

Below, I expand on some of the salient topics in my research.

Sequential Decision Making with Non-standard Feedback

For sparse feedback, the bottleneck for data-efficient learning is typically the task of exploration as exploration cannot be guided via feedback i.e. the exploration must be autonomous. In [2, 7], we proposed a generic method

that can convert any efficient reinforcement learning algorithm into a corresponding exploration algorithm and proved strong performance guarantees for the devised algorithms. The generality of these algorithms and their ability to use any appropriate subroutine makes them efficiently applicable in various environments irrespective of the underlying structure.

In tasks that involve obtaining feedback from people, it is more practical to elicit preference-based feedback. However, most of the work in the literature on reinforcement learning assumes absolute feedback. We addressed this gap in the literature in [24] by providing a near-optimal algorithm for single-state reinforcement learning from preference feedback and a corresponding general lower bound. Experiments on real-world information retrieval datasets show that our proposed algorithm outperforms state-of-the-art benchmark algorithms.

Another form of non-standard feedback is batched feedback, which is often used in real-world scenarios e.g., personalized marketing campaigns by telecommunications companies. In [12, 13], we considered reinforcement learning with batched feedback. We provided a policy-agnostic analysis and proved lower and upper bounds on the performance. We also provided experimental results on real data acquired from our industrial partner KPN.

As for data-efficient learning from incomplete feedback, many significant problems such as dynamic pricing, learning with expert advice, and label-efficient prediction can be formulated using partial monitoring which is a generic framework for sequential learning with incomplete feedback. In [25], we proved that in terms of performance bounds, generic partial monitoring algorithms are suboptimal compared to problem-specific algorithms. Moreover, recently in [6], we proposed an efficient personalized recommendation device for cardiac rehabilitees.

Future work: I would like to devise a generic partial monitoring algorithm with an optimal bound on the performance measure applicable to all the partial monitoring problems. This generic algorithm would be efficiently applicable to many practically relevant problems.

For our work on autonomous exploration with sparse feedback, I would like to find connections with biologically inspired autonomous learning. Here the agent first goes via an extended developmental period in which it learns reusable skills autonomously and then uses those skills to achieve general goals.

In many practical scenarios (e.g., websites providing online courses, recommendation systems for music playlists), 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 devise an optimal algorithm for reinforcement learning problems with generalized temporally partitioned rewards (see our initial work in [5]).

The proposed research output of the above research themes could be applicable in various domains including healthcare, supply chain management, finance, efficient navigation, rescue robots in unknown environments, network management for electrical distribution networks, and microgrid management.

Fairness in Machine Learning

In [19], we argued that the problem of fairness in 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 and it continues to be used as reading material for related courses at a number of universities [26, 27]. We provided theoretical as well as empirical critiques of the fairness notions used 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 the human development paradigm by the United Nations) which address some of these critiques.

Fairness in reinforcement learning is a rapidly growing emerging field. We recently wrote an extensive survey on this topic [2] and I am going to co-deliver a tutorial on this topic at the 15th Asian Conference on Machine Learning (ACML 2023).

Future work: In collaboration with social scientists, I would like to work toward incorporating the suggested notions from our work in [19] (and other notions from social sciences literature) into concrete fair-machine learning formulations.

Curiosity-driven reinforcement learning has been shown to increase equality in competitive resource allocation [28]. I would also like to work on adding curiosity-driven exploration to reinforcement learning algorithms leading to provably fair solutions for sequential decision making problems. Currently, I am supervising a master's student on this topic and our work has led to a joint publication [4].

The proposed research output of this overall research theme could be useful in high-impact domains including healthcare, finance, education, credit, employment, personalized AI, etc.

Privacy in Machine Learning

In [21, 23], we considered the problem of providing local differential privacy in single-state reinforcement learning. We proved a general lower bound and proposed a near-optimal frequentist algorithm and a near-optimal Bayesian algorithm. We also provided an optimal mechanism using which a user can achieve the desired trade-off between privacy and utility. These articles 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. Our experimental results show that our proposed method could be useful in recommender systems. Recently in [10], we extended the problem setting to consider non-stationary environments and proposed an optimal mechanism to achieve the desired trade-off between privacy and utility.

Future work: Considering privacy-preserving learning in the general multi-state reinforcement learning setting is also a prospective research direction. Continuing this line of work, I am writing a research proposal on the topic of locally differentially private algorithms for non-stationary sequential learning. The proposed research output of the above research theme could be practically applicable in various domains including recommender systems and healthcare (for example, see a survey on using differential privacy in medical data analysis [29]).

Sequential Decision Making in Non-stationary Environments

In [16, 17], we achieved optimal performance guarantees without knowing the number of changes (in contrast to the previous related work) for non-stationary single-state reinforcement learning. Our performance guarantees were the first optimal guarantees for an algorithm that is not tuned with respect to the number of changes in the environment. We also showed that our algorithm is optimal for abrupt as well as gradual changes in the environment. In [15], we considered reinforcement learning in non-stationary environments and proposed an algorithm with suitable performance guarantees, which were the first of their kind for general reinforcement learning.

In [2], we considered autonomous exploration in non-stationary environments and devised a meta-algorithm that 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. Recently in [10], we devised an optimal frequentist algorithm using which a user can achieve the desired trade-off between privacy and utility for sequential decision making in non-stationary stochastic environments. In this work, the algorithm takes the number of changes in the environment as an input.

Future work: Proving better sample complexity bounds for autonomous exploration in non-stationary environments using the approach given in [30] would be a significant contribution to the literature.

For our work on privacy-preserving sequential decision making in non-stationary stochastic environments, there are two prospective lines of work. Firstly, it would be worthwhile to devise an optimal algorithm that does not need to know the number of changes in the environment in advance. I believe the approach from our earlier work in [16,17] would be useful here. Secondly, an optimal Bayesian algorithm for this problem would be a significant and practically relevant research direction as Bayesian algorithms have shown better empirical performance in similar problems. It would also be interesting to devise a Bayesian approach for general reinforcement learning in non-stationary environments to complement the frequentist approach we provided in [15].

The proposed output of this overall research theme could be useful in domains in which the assumption of stationarity is sometimes unrealistic e.g., recommender systems, supply chain management, etc.

Safe and Constrained Reinforcement Learning

In a wide range of modern applications of reinforcement learning, it is not sufficient for the agent to only maximize a scalar reward. Additionally, they must satisfy various constraints such as computation budget and safety that are critical in real-world problems.

Our work on reinforcement learning with batched feedback [12, 13] is motivated by computation budget constraints in recommender systems. In [14], we considered risk-averse reinforcement learning and proposed a suitable performance measure. In [1], we devised a Bayesian algorithm for reinforcement learning in constrained Markov decision processes. We proved near-optimal Bayesian regret bound for our algorithm while our

experimental results showed that it outperforms the state-of-the-art for the considered benchmarks.

Future work: It could be worthwhile to use this technique for the reward-free constrained reinforcement learning paradigm introduced in [31]. Our goal here will be to devise reward-agnostic safe policies (i.e. satisfying all the constraints) and then use them for particular tasks as dictated by the given reward functions.

I would also like to explore if the feasibility guarantees provided in [1] could be used in other problems in constrained reinforcement learning. In particular, I would like to work on extending our theoretical analysis to the frequentist regret bound by incorporating existing methods such as [32] or [33].

Research Proposal : Fair Sequential Decision Making with Non-Standard Feedback

Firstly, this is a significant research direction as non-standard feedback is seen in sequential decision making problems in high-impact domains such as healthcare, finance, education, etc. However, despite fairness being a legal requirement, unbiased/unfair behavior has been documented by machine learning algorithms in many domains. Secondly, there is a gap in the literature in this direction as can be seen in our extensive survey [2]. Moreover, in addition to being an impactful research direction, it presents significant technical challenges that cannot be effectively solved by an incremental approach combining the existing work on sequential decision making with non-standard feedback and fair machine learning (see Appendix A for further explanation). Instead, an innovative and structured approach is necessary, as I outline below.

- **Task 1** : Feasibility of fairness definitions – This will be done via theoretical analysis and empirical studies. This is a very important task because we need to understand if the target fairness definitions are feasible for the problems that we consider. Defining bias based on multiple sensitive attributes would form a major component of this task.
- **Task 2** : Characterizing strictness of performance guarantees – Our survey [2] shows a lack of discussion about the strictness of performance guarantees in the literature even for standard feedback. This makes it difficult for us to judge if the devised solutions are optimal. The output of this task will be target performance guarantees.
- **Task 3** : Efficient algorithms with optimal performance guarantees – In this task, we will devise algorithms for both stationary and non-stationary environments, and frequentist as well as Bayesian algorithms.
- **Task 4** : Fair learning while protecting user privacy – Examining the privacy-fairness-utility trade-off would form a key component of this task.
- **Task 5** : Explainability of devised solutions – Investigating the explainability of devised solutions to stakeholders such as domain-experts, decision-makers, users, etc.
- **Task 6** : Verifying the devised solutions on practical applications such as healthcare, finance, employment, personalized AI, etc.

Research Proposal : Personalized and Fair Healthcare using Reinforcement Learning

Here, we propose to develop a system consisting of a wearable, a robot, and suitable algorithms for pain detection and management for people with restricted abilities to self-report such as children, persons with intellectual disabilities, and older adults.

Firstly while there has been some progress in measuring the physiological correlates of pain, estimating the perceived pain and the emotional burden related to it remains challenging, particularly with patients who cannot verbalize their experience.

Secondly, it has been shown sexes do not feel pain the same way [34]. Furthermore, research has also shown that biases with respect to gender and sex are exhibited and perpetuated while using machine learning for healthcare. It has been seen that the design of the majority of machine learning algorithms ignores the sex and gender dimension and its contribution to health differences among individuals [35]. In particular, machine learning tools for healthcare that use unisex biochemical thresholds may disadvantage female patients [36]. Thus, we aim to make personalized pain prediction taking such biases into consideration. Reinforcement learning is a feasible solution for this goal as it has been used successfully in personalized learning tools.

This proposed project will first gather physiological and behavioral data based on an already developed prototype of a wearable connected to a robot that engages in context-appropriate interactions with users. A suite of machine learning algorithms will be developed to learn to estimate the presence, severity, and perception of pain. The output of this project will be a step toward the development of a robot-based therapy for pain management.

References

References [1] to [25] can be found in my CV.

[26] Wei Wei and James Landay. Lectures notes for fair, accountable, and transparent (facct) deep learning. Stanford University, 2020. URL https://hci.stanford.edu/courses/cs335/2020/sp/lec1.pdf.

[27] Jesse Hoey. Lectures notes for cognitive science. University of Waterloo, 2022. URL https://cs.uwaterloo.ca/~jhoey/teaching/cogsci600/schedule.html#.

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

[29] WeiKang Liu, Yanchun Zhang, Hong Yang, and Qinxue Meng. A survey on differential privacy for medical data analysis. *Annals of Data Science*, Jun 2023. ISSN 2198-5812. doi: 10.1007/s40745-023-00475-3.

[30] 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.

[31] Sobhan Miryoosefi and Chi Jin. A simple reward-free approach to constrained reinforcement learning. In *Proceedings of the 39th International Conference on Machine Learning*, Proceedings of Machine Learning Research, pages 15666–15698, 2022.

[32] Shipra Agrawal and Randy Jia. Optimistic posterior sampling for reinforcement learning: worst-case regret bounds. In *Advances in Neural Information Processing Systems*, 2017.

[33] Daniil Tiapkin, Denis Belomestny, Daniele Calandriello, Eric Moulines, Remi Munos, Alexey Naumov, Mark Rowland, Michal Valko, and Pierre MENARD. Optimistic posterior sampling for reinforcement learning with few samples and tight guarantees. In *Advances in Neural Information Processing Systems*, 2022.

[34] Amber Dance. Why the sexes don't feel pain the same way. *Nature*, 567(7749):448–450, March 2019. doi: 10.1038/d41586-019-00895-. URL https://ideas.repec.org/a/nat/nature/v567y2019i7749d10.1038_d41586-019-00895-3.html.

[35] Davide Cirillo, Silvina Catuara-Solarz, Czuee Morey, Emre Guney, Laia Subirats, Simona Mellino, Annalisa Gigante, Alfonso Valencia, María José Rementeria, Antonella Santuccione Chadha, and Nikolaos Mavridis. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *npj Digital Medicine*, 3(1):81, Jun 2020. ISSN 2398-6352. doi: 10.1038/s41746-020-0288-5. URL https://doi.org/10.1038/s41746-020-0288-5.

[36] Isabel Straw and Honghan Wu. Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction. *BMJ Health & Care Informatics*, 29(1), 2022. doi: 10.1136/bmjhci-2021-100457. URL https://informatics.bmj.com/content/29/1/e100457.

[37] Matthew Joseph, Michael Kearns, Jamie H Morgenstern, and Aaron Roth. Fairness in learning: Classic and contextual bandits. In *Advances in Neural Information Processing Systems*, 2016.

[38] Matthew Joseph, Michael J. Kearns, Jamie Morgenstern, Seth Neel, and Aaron Roth. Fair algorithms for infinite and contextual bandits. *CoRR*, abs/1610.09559, 2016. URL http://arxiv.org/abs/1610.09559.

[39] Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. Fairness in reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1617–1626, 2017.

[40] Arun Verma, Zhongxiang Dai, and Bryan Kian Hsiang Low. Bayesian optimization under stochastic delayed feedback. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 22145–22167. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/verma22a.html.

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

A Ineffectiveness of Combined Approach for Fair Sequential Decision Making with Non-Standard Feedback

To see why an incremental approach combining the existing work on sequential decision making with non-standard feedback and fair machine learning will not lead to effective solutions, let us see some examples.

- Meritocratic fairness and preference feedback – Meritocratic fairness is a popular fairness notion used in the literature [37, 38, 39] and it can be defined as follows.

  Definition 1 *Any individual who is currently more qualified than another individual should always have at least as good a chance of positive outcome/treatment as a less qualified individual.*

  Typically, the algorithms proposed in the existing literature use the expected absolute feedback as a proxy for the qualifications referred to in the above definition. However, if the problem formulation considers preference feedback, such absolute feedback is not available to the algorithm. Therefore the existing algorithmic solutions are incapable of providing meritocratic fairness for sequential decision making problems with preference feedback. Going forward, we will need to investigate if the notion of meritocratic fairness is even feasible for problems with preference feedback. Task 1, as outlined in the research proposal above, will scrutinize such concerns.

- Bayesian fair algorithm for delayed feedback – When fairness considerations are not part of the problem formulation, a typical Bayesian solution for handling delayed feedback involves assuming a fictitious minimum feedback value for a decision till the actual feedback is received [40]. Due to this approach, the posterior mean around that decision does not increase which leads to better exploration, which in turn leads to an efficient solution. However, when fairness considerations are part of the problem formulation, the above approach can be considered unfair toward decisions with longer delays. As a result, alternative approaches are required for devising Bayesian fair algorithms for sequential decision making problems with delayed feedback. Task 3, as outlined in the research proposal above, comprises of devising algorithms for such problem formulations.