Learning and Planning Under Uncertainty for Conservation Decisions

Lily Xu

Harvard University
lily_xu@g.harvard.edu

Abstract

My research focuses on new techniques in machine learning and game theory *to optimally allocate our scarce resources in multi-agent settings to maximize environmental sustainability*. Drawing scientific questions from my close partnership with conservation organizations, I have advanced new lines of research in learning and planning under uncertainty, inspired by the low-data, noisy, and dynamic settings faced by rangers on the frontlines of protected areas.

Environmental sustainability is a pressing global issue, threatening our food security and livelihoods. However, positive externalities are not properly accounted for in our economy, tipping the scales *away* from sustainability. Rhino horn fetches an estimated US$60,000 per kilogram on the black market, while many protected areas lack sufficient funds to even purchase boots for their rangers. Unfortunately, we have very limited resources to protect wildlife and forests; our challenge is to make best use of these resources.

My primary line of research focuses on helping rangers plan patrols to protect endangered animals from snaring. Viewed algorithmically, this problem is one of *optimizing* limited resources. The objective is to maximize the expected number of snares that rangers detect, so that we can remove those snares and prevent wildlife from getting caught. Unfortunately, the data we have available are biased and incomplete, necessitating *online learning under uncertainty*. Furthermore, we expect that poachers might eventually learn rangers' behavior and adapt their strategy accordingly, making need for *sequential planning*. These real-world problems, and the computational solutions I have developed, demonstrate that environmental domains such as wildlife conservation offer a range of fundamental new research challenges in robust planning and data-driven optimization.

Furthermore, my research bridges research and practice: algorithms I developed have been deployed in field tests in Cambodia and are being scaled globally to 1,000 parks through integration with SMART, the leading software for protected area management. My work has been executed in close partnership with conservation organizations, including WWF and Wildlife Conservation Society (WCS). I have worked directly with rangers on the frontline, spoken extensively with conservation managers and biodiversity experts, and traveled to Cambodia and Belize to meet with park rangers and go on patrol. Through my work, I hope to showcase that tackling realistic problems in socially relevant domains offers ample substance for impactful AI research. Beyond my research, I am also committed to growing and empowering the AI for social good community, through my service co-organizing MD4SG and other initiatives.

Key Research Questions My research agenda begins with identifying *algorithmic assumptions* and *prerequisites* that do not hold in the real world. I mold projects by focusing on those that pose the most significant barrier to real-life implementation. The assumptions often come in the form of *assuming ground-truth knowledge* of the environment which may not be known. For example, recognizing the need for sequential planning in patrol planning but lack of information about environmental dynamics, I developed the first framework to learn robust reinforcement learning policies under the minimax regret criterion, producing greater performance improvement than the maximin reward objective that is standard in reinforcement learning. In general, real-world limitations lay bare the greatest weakness of new artificial intelligence techniques; understanding constraints in realistic problems ought to be an essential component of determining what the research community focuses on.

Online Learning for Data-Scarce Settings

Unfortunately, many parks are only beginning patrols or have not explored vast regions of the protected area. For these data-scarce parks, we focus on learning in an online fashion, balancing visiting known hotspots with exploring new regions to improve our predictive model (Xu et al. 2021a). We consider multi-armed bandits, but traditional bandit approaches unfortunately compromise short-term performance for long-term optimality, resulting in animals poached and forests destroyed. To speed up performance, we leverage smoothness in the reward function and decomposability of actions, enabling us to reduce our uncertainty and tighten existing guarantees that are proven lower bounds for Lipschitz-continuity alone (Kleinberg et al. 2019). *With this approach, we transcend the proven lower regret bound of existing algorithms and generalize combinatorial bandits to continuous spaces*. On top of achieving the-

oretical no-regret, our algorithm, LIZARD, achieves much stronger short-term performance empirically, increasing its usefulness — particularly in high-stakes environments such as poaching prevention where we cannot compromise short-term performance, which would compromise animal lives.

Robust Sequential Planning for Uncertain Environments

In some parks, poachers are sophisticated and thus adapt to ranger patrols over time. The dynamic necessitates sequential planning, but sequential planning techniques such as reinforcement learning (RL) require a precise simulator of the environment, which we lack. Thus, our goal is to plan robust patrols under environment uncertainty. We focus on robust sequential planning following the minimax regret criterion, overcoming the shortcoming of standard maximin reward which leads to overly conservative policies (Xu et al. 2021b). To address the minimax regret objective, we pose the problem as a zero-sum game between an agent and nature, leveraging the double oracle approach to optimize policies within the continuous strategy spaces. Our algorithm, MIRROR, is a general RL framework to plan against worst-case instantiations of the uncertain environment. We prove convergence to an ϵ–optimal strategy in a finite number of iterations, overcoming the difficulty of continuous state and action spaces, and empirically evaluate our algorithm on real poaching data. *MIRROR is the first approach for calculating minimax regret–optimal policies using RL.*

Uncertain environment dynamics is pervasive in many other critical real-world problems, including healthcare. We have extended this work to consider robustness in restless bandits, a problem of constrained resource allocation with a number of evolving and couple Markov decision processes (Killian et al. 2022).

Ranked Prioritization Among Groups

Existing approaches in bandit-based allocation typically focus on utilitarian reward maximization, ignoring the effects of allocation on various subgroups. When there are multiple groups present with varying distribution, it becomes critical to ensure that we allocate more resources to more vulnerable groups. In the conservation setting, as our ultimate goal is biodiversity conservation, we recognize that some conservation needs are of higher priority: we ought to focus more effort on conserving tigers and elephants than deer and wild pig. We thus consider sequential planning under uncertainty while ensuring that we prioritize the most vulnerable groups.

To fairly prioritize these important groups, we propose a novel bandit objective that considers meritocratic fairness under ranked prioritizations over groups, enabling the designer to tune the degree to which they prioritize fairness (Xu et al. 2022). Key challenges are (i) combinatorial actions, (ii) the need to learn in an online fashion, and (iii) joint prioritization of fairness and reward. In addressing this problem, *we uncover an important general result for combinatorial bandits*, which addresses the broadly applicable setting where the individual rewards are *weighted* in the overall objective. We provide a fast, optimal algorithm for this general problem of selecting a combinatorial action at each timestep to simultaneously maximize reward and fairness, building off the principle of optimism in the face of uncertainty.

Future Work: Continuing to Close Gaps Between Research and Practice

To transfer research advances effectively to the ground, there remain significant barriers to deploying sequential decision-making approaches such as bandits and RL.

Strategic Agents I have considered increasingly complex environment dynamics, beginning with a stochastic setting, moving to online learning, then considering a reinforcement learning setting. I will consider more complex settings with strategic agents who compete by planning actions with imperfect information. My recent paper focused on solving Stackelberg games with multiple self-interested followers who are unable to communicate (Wang et al. 2022). These game theoretic interactions are relevant in wildlife conservation (and beyond) as poachers may strategically respond to rangers as well as local informants or larger bodies.

Robustness and Fairness I am interested in exploring the intersection of robustness and fairness. Under the minimax regret criterion, the greater our uncertainty interval is for one demographic, the more likely we are to unjustly prioritize that group over others—a natural outcome of hedging our bets to protect against worst-case regret. We thus need to balance between robustness and fairness, recognizing that the more imbalanced our uncertainties are for varying groups, the poorer we may perform in terms of fairness.

Moving forward, I will continue work to identify cracks within the theoretical foundation of machine learning in order to most effectively apply these decision-making approaches to urgent problems. Throughout these challenges, my ultimate goal is to help us make more informed decisions on the best strategic interventions across critical, low-resource domains such as conservation.

References

Killian, J. A.; Xu, L.; Biswas, A.; and Tambe, M. 2022. Robust Restless Bandits: Tackling Interval Uncertainty with Deep Reinforcement Learning. *UAI*.

Kleinberg, R.; et al. 2019. Bandits and experts in metric spaces. *J. of the ACM*.

Wang, K.; Xu, L.; Perrault, A.; Reiter, M. K.; and Tambe, M. 2022. Coordinating Followers to Reach Better Equilibria: End-to-End Gradient Descent for Stackelberg Games. In *AAAI*.

Xu, L.; Biswas, A.; Fang, F.; and Tambe, M. 2022. Ranked Group Prioritization in Combinatorial Bandit Allocation. In *IJCAI*.

Xu, L.; Bondi, E.; Fang, F.; Perrault, A.; Wang, K.; and Tambe, M. 2021a. Dual-Mandate Patrols: Multi-Armed Bandits for Green Security. In *AAAI*.

Xu, L.; Perrault, A.; Fang, F.; Chen, H.; and Tambe, M. 2021b. Robust Reinforcement Learning Under Minimax Regret for Green Security. In *UAI*.