

Learning and Planning Under Uncertainty for Green Security

Lily Xu
Harvard University
lily_xu@g.harvard.edu

Abstract

Green security concerns the protection of the world’s wildlife, forests, and fisheries from poaching, illegal logging, and illegal fishing. Unfortunately, conservation efforts in green security domains are constrained by the limited availability of defenders, who must patrol vast areas to protect from attackers. Artificial intelligence (AI) techniques have been developed for green security and other security settings, such as US Coast Guard patrols and airport screenings, but effective deployment of AI in these settings requires learning adversarial behavior and planning in complex environments where the true dynamics may be unknown. My research develops novel techniques in machine learning and game theory to enable the effective development and deployment of AI in these resource-constrained settings. Notably, my work has spanned the pipeline from learning in a supervised setting, planning in stochastic environments, sequential planning in uncertain environments, and deployment in the real world. The overarching goal is to optimally allocate scarce resources under uncertainty for environmental conservation.

1 Overview of Contributions

My work consists of technical research motivated by environmental conservation. The research that will comprise my dissertation is aimed at eliminating obstacles to deploying existing AI techniques through close interaction with the domain to identify and understand these barriers to deployment. A primary objective of mine is to demonstrate that real-world challenges uncover interesting and fruitful technical challenges that advance the state-of-the-art in AI.

1.1 Data-Rich Settings and Field Tests

The illegal wildlife trade is a multi-billion dollar industry pushing countless species to the brink of extinction. Profit-driven poachers will enter protected areas and place snares to trap animals. To prevent poaching, rangers conduct patrols to remove snares. Unfortunately, poachers have the upper hand; wire snares are extremely cheap to make, so poachers can easily blanket the ground with snares. Our collaborators



Figure 1: Rangers searching for snares (right) near a waterhole (left) in Srepok Wildlife Sanctuary in Cambodia.

at Srepok Wildlife Sanctuary in Cambodia have noted that poachers are prolific: park managers estimate that there are four snares for every one deer. In contrast, Srepok has only 72 rangers to patrol its 3,750 km². Given this resource imbalance, it is critical to efficiently plan ranger patrols.

In well-established parks, rangers often have long records of past patrols which can be leveraged with supervised learning to predict poaching hotspots. To address bias and uncertainty in the data, we focus on quantifying uncertainty in machine learning models, which we then employ during planning [Xu *et al.*, 2020a]. Specifically, we leverage uncertainty from Gaussian processes in an ensemble method designed for class imbalance. We then incorporate the uncertainty of each prediction into a mixed integer linear program to determine the optimal patrol strategy, using a scaling parameter to set our threshold for risk.

We evaluated our predictive models on the ground in Srepok. During a one-month field test in Srepok Wildlife Sanctuary in Cambodia, *rangers detected and removed over 1,000 snares — five times that of an average month*. These tests demonstrate that our algorithms perform well not just in the digital tests that we simulate but also on the ground. Based on this success, we are making our predictive algorithms broadly available by integrating our research advances into established software. We have teamed up with leading conservation organizations to integrate our machine learning model into the SMART platform, the leading software for protected area management. This integration democratizes AI and enables us to contribute more broadly to the global effort to protect wildlife from poaching, bringing our work to over 800 protected areas across 55 countries that are using SMART.

1.2 Data-Scarce Settings

For data-rich parks, we use historical information to generate predictions, with the assumption that this data offers a sufficiently accurate picture of attacker behavior from which we can plan patrols. 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, simultaneously exploring regions of the park to improve our predictive model [Xu *et al.*, 2021a].

We seek to plan *dual-mandate patrols* to simultaneously detect illegal activities and collect valuable data to improve our predictive model and achieve higher long-term reward. We therefore explicitly consider uncertainty in the learning phase by integrating learning and planning. We use a multi-armed bandit formulation, where each action represents a patrol strategy, to balance exploration of infrequently visited regions and exploitation of known hotspots. However, traditional bandit approaches compromise short-term performance for long-term optimality, resulting in animals poached and forests destroyed. Additionally, the combinatorial problem setting poses a challenge as each action requires planning a patrol route, which selects a set of regions to patrol.

To speed up performance, we leverage smoothness in the reward function and decomposability of actions. Specifically, we assume the reward is Lipschitz-continuous and can be tracked individually across targets. *With this approach, we transcend the proven lower regret bound of Lipschitz bandits and generalize combinatorial bandits to continuous spaces.* On top of achieving theoretical no-regret, we also demonstrate that our algorithm, LIZARD, achieves better short-term performance empirically, increasing the usefulness of this approach in practice — particularly in high-stakes environments where we cannot compromise short-term reward.

1.3 Robust Sequential Planning

So far, my work has focused on learning models of poacher behavior based on historical patterns, assuming stochasticity. However, poachers’ behaviors are determined not only by their historical preferences, but also their reactions to ranger patrols. An open question, both in computer science and in conservation biology, has been: How do poachers react to ranger patrols? Some past work consider a game-theoretic adversary based on an SUQR model, but this model is not justified with real-world data. For the first time, we demonstrate a measurable deterrence effect in real-world poaching data [Xu *et al.*, 2020b]. These findings highlight the need to take a *sequential decision-making* approach to patrol planning, rather than one-shot as in past work. Furthermore, predictions of poacher behavior are imperfect, highlighting the need to consider *robustness* in our patrol plans. Thus, we propose a robust patrol-planning approach using reinforcement learning [Xu *et al.*, 2021b].

Building off our deterrence findings, we now focus on patrol planning using our learned adversary model. We consider the sequential decision-making problem of planning an optimal strategy for the defender, where the defender’s actions in one timestep influences the potential reward in the future where the adversary reacts based not on game-theoretical best response but rather on delayed feedback. However, we do not

have perfect knowledge of the poachers preferences, so we must determine policies that are robust to uncertainty.

My proposed algorithm, MIRROR, formulates the robust planning problem as a zero-sum game between an agent, who learns an optimal defender policy, and nature, who selects worst-case environment parameter values. We consider optimizing for minimax regret, which tends to favor strategies that are less overly conservative than the traditional notion measure of maximin reward. MIRROR uses reinforcement learning-based oracles for the agent and nature to enable learning sequential policies in a large, continuous state and action space. Minimax regret is a challenging objective to solve for, as the performance is assessed relative to the optimal strategy under nature’s parameter selection, which creates a constantly changing goalpost. The difficulty of this task is evidenced by the fact that our work is the first to learn minimax regret-optimal policies using reinforcement learning.

2 Future Work

AI offers considerable promise for environmental conservation, but successful deployment of AI methods requires thoughtful consideration of challenges including uncertainty, resource constraints, and short time horizons. Focusing on environmental challenges open the door to general-purpose algorithms for faster learning and more robust planning. Issues I intend to address next concern heterogeneous defender resources, execution uncertainty in actions, and more realistic game-theoretic adversaries.

Looking at the field of AI for Social Good more broadly, I have also spent time thinking about how AI researchers can more proactively and effectively bring about social good, specifically to adopt a participatory approach which focuses on elevating the capabilities of those who are most marginalized [Bondi *et al.*, 2021]. From an application standpoint, I intend to focus on problems related to environmental justice in future work.

References

- [Bondi *et al.*, 2021] Elizabeth Bondi, Lily Xu, Diana Acosta-Navas, and Jackson A. Killian. Envisioning communities: A participatory approach towards AI for social good. In *AIES-21*, 2021.
- [Xu *et al.*, 2020a] Lily Xu, Shahrzad Gholami, Sara Mc Carthy, et al. Stay ahead of poachers: Illegal wildlife poaching prediction and patrol planning under uncertainty with field test evaluations. In *ICDE-20*, 2020.
- [Xu *et al.*, 2020b] Lily Xu, Andrew Perrault, et al. Game theory on the ground: The effect of increased patrols on deterring poachers. *CRCs AI4SG 2020 Workshop*, 2020.
- [Xu *et al.*, 2021a] Lily Xu, Elizabeth Bondi, Fei Fang, Andrew Perrault, Kai Wang, and Milind Tambe. Dual-mandate patrols: Multi-armed bandits for green security. In *AAAI-21*, 2021.
- [Xu *et al.*, 2021b] Lily Xu, Andrew Perrault, Fei Fang, Haipeng Chen, and Milind Tambe. Robust reinforcement learning under minimax regret for green security. In *UAI-21*, 2021.