Research internship proposal
Centre Inria de l'Université de Lille
Team project Scool – Spring 2024

"Multi-armed bandits for soil-regeneration of agrosystems: Theory and application."

Keywords: Multi-armed bandits, Sequential statistics, Societal challenge.
Supervision: The intern will be advised by Odalric-Ambrym Maillard from Inria team-project Scool.
Place: This internship will be primarily held at the research center Inria Lille – Nord Europe, 40 avenue Halley, 59650 Villeneuve d'Ascq, France, in the Inria team-project Scool (previously known as SequeL).

Context Multi-armed bandit theory has witnessed tremendous progress over the last decade, yielding algorithms achieving strong learning guarantees (regret minimization, best-arm identification) in increasingly challenging context involving sequential decision making in uncertain environment. In particular, provably optimal strategies such as KL-UCB [Cappé et al., 2013], TS [Korda et al., 2013] or IMED [Honda and Takemura, 2015] have been shown to achieve strong optimality in parametric context, while other strategies such as NPTS from [Riou and Honda, 2020] or SDA [Baudry et al., 2020, Baudry et al., 2021b] obtained non-parameteric optimality, enabling application of multi-armed bandit to a large range of applications when reward distributions are not easily modeled with classical families. Further, progress has been made to handle risk-averse objective rather than simply obtaining guarantees in expectation [Baudry et al., 2021a], or when reward feedback is available in infrequent batches rather than immediately [Gautron et al., 2022]. Last, the extension of classical bandits to structured bandits [Magureanu, 2016, Saber, 2022], including e.g. contextual bandits enable novel algorithms in recommender systems that can learn to provide decisions (actions) personalized to a given context in an efficient way [Abbasi-Yadkori et al., 2011], see also [Kirschner and Krause, 2019], and even to Markove decision processes [Pesquerel and Maillard, 2022]. Despite this, applying multi-armed bandit in a real-life application comes with many additional challenges, including availability of data, compliance [Della Penna et al., 2016], or external risks.

In this internship, we consider a use-case in agriculture, where data is easier to access than e.g. in health care, due to less stringent privacy issues. More specifically, we consider an application of contextual multi-armed bandits to assist experimentation of promising agricultural practices to regenerate soil fertility in Madagascar agrosystems. There are important food-security risks in the region, some due to bad practical practices, and improving soil fertility has became of utmost societal importance. Soil analysts from Laboratoire des RadioIsotopes at Madagascar manage to design a tool able to estimate fertility level in a large portion of the island, and classify farms in four categories. Some practices are identified as promising, depending on context, to improve upon the base practice used by farmers, however with some uncertainty. One challenge is to recommend while testing, these practices to farmers in order to ensure compliance from farmers and effectively improving soil fertility.

Proposal The objective of this internship is to assist experimentation, by adapting the algorithms from the literature in bandits to this application domain, taking into account risks, delays, and specifying group-sequential hypothesis testing strategies (how many arms, contexts, which phases, etc). The goal is not directly to apply bandit strategies directly on farmers, but to produce a preliminary study enabling such an application, considering as well ethical concerns, before effectively starting real experiments (that will take a few years to complete). Though applied, the task requires a solid mathematical understanding of the algorithms at hand, including their range of application, being able to formalize some of the challenges identified, and excellent programming skills. To be successful, this internship also requires excellent communication skills, due to the interaction with researchers both within Inria Scool team and researchers in Madagascar. No prior knowledge of agrosystems is required for this internship. The internship may lead to a follow-up Ph.D. on real-life applicability of bandit and RL strategies.

Host institution and supervision The student will be hosted at Centre Inria de l'Université de Lille, in the Scool team. Scool (Sequential COntinual and Online Learning) is an Inria team-project. It was created on November 1st, 2020 as the follow-up of the team SequeL. In a nutshell, the research topic of Scool is the study of the sequential decision making problem under uncertainty. Most of our activities are related to either bandit problems, or reinforcement learning problems. Through collaborations, we are working on their application in various fields including health, agriculture and ecology, sustainable development. More information, please visit `https://team.inria.fr/scool/projects`.

Odalric-Ambrym Maillard is a permanent researcher at Inria. He has worked for over a decade on advancing the theoretical foundations of reinforcement learning, using a combination of tools from statistics, optimization and control, in order to build more efficient algorithms able to provide decision making in uncertain environments. He was PI of several projects, including ANR-JCJC project BADASS (BAnDits Against non-Stationarity and Structure), Inria Action Exploratoire SR4SG (Sequential Recommendation for Sustainable Gardening) and Inria-Japan Associate team RELIANT (Reliable Bandit strategies). His goal is to push forward key fundamental and applied questions related to the grand-challenge of making reinforcement learning applicable in real-life societal applications.

References

[Abbasi-Yadkori et al., 2011] Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24.

[Baudry et al., 2021a] Baudry, D., Gautron, R., Kaufmann, E., and Maillard, O. (2021a). Optimal thompson sampling strategies for support-aware cvar bandits. In *International Conference on Machine Learning*, pages 716–726. PMLR.

[Baudry et al., 2020] Baudry, D., Kaufmann, E., and Maillard, O.-A. (2020). Sub-sampling for efficient non-parametric bandit exploration. *Advances in Neural Information Processing Systems*, 33:5468–5478.

[Baudry et al., 2021b] Baudry, D., Saux, P., and Maillard, O.-A. (2021b). From optimality to robustness: Adaptive re-sampling strategies in stochastic bandits. *Advances in Neural Information Processing Systems*, 34:14029–14041.

[Cappé et al., 2013] Cappé, O., Garivier, A., Maillard, O.-A., Munos, R., and Stoltz, G. (2013). Kullback-leibler upper confidence bounds for optimal sequential allocation. *The Annals of Statistics*, pages 1516–1541.

[Della Penna et al., 2016] Della Penna, N., Reid, M. D., and Balduzzi, D. (2016). Compliance-aware bandits. *arXiv preprint arXiv:1602.02852*.

[Gautron et al., 2022] Gautron, R., Baudry, D., Adam, M., Falconnier, G. N., and Corbeels, M. (2022). Towards an efficient and risk aware strategy for guiding farmers in identifying best crop management. *arXiv preprint arXiv:2210.04537*.

[Honda and Takemura, 2015] Honda, J. and Takemura, A. (2015). Non-asymptotic analysis of a new bandit algorithm for semi-bounded rewards. *J. Mach. Learn. Res.*, 16:3721–3756.

[Kirschner and Krause, 2019] Kirschner, J. and Krause, A. (2019). Stochastic bandits with context distributions. *Advances in Neural Information Processing Systems*, 32.

[Korda et al., 2013] Korda, N., Kaufmann, E., and Munos, R. (2013). Thompson sampling for 1-dimensional exponential family bandits. *Advances in neural information processing systems*, 26.

[Magureanu, 2016] Magureanu, S. (2016). *Structured Stochastic Bandits*. PhD thesis, KTH Royal Institute of Technology.

[Pesquerel and Maillard, 2022] Pesquerel, F. and Maillard, O.-A. (2022). Imed-rl: Regret optimal learning of ergodic markov decision processes. *Advances in Neural Information Processing Systems*, 35:26363–26374.

[Riou and Honda, 2020] Riou, C. and Honda, J. (2020). Bandit algorithms based on thompson sampling for bounded reward distributions. In *Algorithmic Learning Theory*, pages 777–826. PMLR.

[Saber, 2022] Saber, H. (2022). *Structure Adaptation in Bandit Theory*. PhD thesis, Université de Lille.