RESEARCH CENTRE

Inria Centre at the University of Lille

IN PARTNERSHIP WITH:

CNRS, Université de Lille

2023

ACTIVITY REPORT

Project-Team SCOOL

Sequential decision making under uncertainty problem

IN COLLABORATION WITH: Centre de Recherche en Informatique, Signal et Automatique de Lille

DOMAIN

Applied Mathematics, Computation and Simulation

THEME

Optimization, machine learning and statistical methods

Inria

Contents

Project-Team SCOOL 1

1 Team members, visitors, external collaborators 2

2 Overall objectives 3

3 Research program 4

4 Application domains 4

5 Social and environmental responsibility 5

6 Highlights of the year 5

7 New software, platforms, open data 5
7.1 New software . . . 5
7.1.1 rlberry . . . 5
7.1.2 gym-DSSAT . . . 6
7.1.3 Weight Trajectory Predictor : algorithm . . . 6
7.1.4 Adastop . . . 7
7.1.5 average-reward-reinforcement-learning . . . 7
7.1.6 FarmGym . . . 7

8 New results 8
8.1 Bandits and RL theory . . . 8
8.2 Bandits and RL under Real-life constraints . . . 10
8.3 Bandits and RL for real-life: Deep RL and Applications . . . 12
8.4 Other . . . 13
8.4.1 Methodology . . . 13
8.4.2 Digital health . . . 13
8.4.3 Sailboat digital twin . . . 14
8.4.4 Computer-assisted mathematics . . . 15
8.4.5 Interpretability . . . 15
8.4.6 Privacy . . . 16

9 Bilateral contracts and grants with industry 16
9.1 Bilateral contracts with industry . . . 16

10 Partnerships and cooperations 16
10.1 International initiatives . . . 16
10.1.1 Inria associate team not involved in an IIL or an international program . . . 16
10.1.2 Participation in other International Programs . . . 17
10.2 International research visitors . . . 18
10.2.1 Visits of international scientists . . . 18
10.2.2 Visits to international teams . . . 19
10.3 European initiatives . . . 20
10.3.1 Other european programs/initiatives . . . 20
10.4 National initiatives . . . 20
10.4.1 ANR projects . . . 20
10.4.2 PEPR projects . . . 21
10.4.3 Inria challenge . . . 22
10.4.4 Other projects in France . . . 22
10.5 Regional initiatives . . . 23

11 Dissemination **23**
11.1 Promoting scientific activities . . . 23
11.1.1 Scientific events: organisation . . . 23
11.1.2 Scientific events: selection . . . 23
11.1.3 Journal . . . 24
11.1.4 Invited talks . . . 24
11.1.5 Tutorials . . . 25
11.1.6 Scientific expertise . . . 25
11.1.7 Research administration . . . 25
11.2 Teaching - Supervision - Juries . . . 25
11.2.1 Teaching . . . 25
11.2.2 Supervision . . . 26
11.2.3 Juries . . . 26
11.3 Popularization . . . 26
11.3.1 Articles and contents . . . 26
11.3.2 Education . . . 27
11.3.3 Interventions . . . 27

12 Scientific production **27**
12.1 Major publications . . . 27
12.2 Publications of the year . . . 28
12.3 Cited publications . . . 31

Project-Team SCOOL

Creation of the Project-Team: 2020 November 01

Keywords

Computer sciences and digital sciences

- A3. – Data and knowledge
- A3.1. – Data
- A3.1.1. – Modeling, representation
- A3.1.4. – Uncertain data
- A3.1.11. – Structured data
- A3.3. – Data and knowledge analysis
- A3.3.1. – On-line analytical processing
- A3.3.2. – Data mining
- A3.3.3. – Big data analysis
- A3.4. – Machine learning and statistics
- A3.4.1. – Supervised learning
- A3.4.2. – Unsupervised learning
- A3.4.3. – Reinforcement learning
- A3.4.4. – Optimization and learning
- A3.4.5. – Bayesian methods
- A3.4.6. – Neural networks
- A3.4.8. – Deep learning
- A3.5.2. – Recommendation systems
- A5.1. – Human-Computer Interaction
- A5.10.7. – Learning
- A8.6. – Information theory
- A8.11. – Game Theory
- A9. – Artificial intelligence
- A9.2. – Machine learning
- A9.3. – Signal analysis
- A9.4. – Natural language processing
- A9.7. – AI algorithmics

Other research topics and application domains

- B2. – Health
- B3.1. – Sustainable development
- B3.5. – Agronomy
- B9.5. – Sciences
- B9.5.6. – Data science

1 Team members, visitors, external collaborators

Research Scientists

- Riad Akrour [INRIA, ISFP]
- Debabrota Basu [INRIA, ISFP]
- Rémy Degenne [INRIA, ISFP]
- Emilie Kaufmann [CNRS, Researcher, HDR]
- Odalric-Ambrym Maillard [INRIA, Researcher, HDR]
- Timothée Mathieu [INRIA, Researcher, from Oct 2023]

Faculty Member

- Philippe Preux [Team leader, UNIV LILLE, Professor Delegation, HDR]

Post-Doctoral Fellows

- Tuan Dam Quang Tuan [INRIA, Post-Doctoral Fellow]
- Riccardo Della Vecchia [INRIA, until Oct 2023]
- Timothée Mathieu [INRIA, Post-Doctoral Fellow, until Sep 2023]
- Alena Shilova [INRIA, Post-Doctoral Fellow]
- Eduardo Vasconcellos [FFU - NITEROI, Post-Doctoral Fellow, from Jul 2023 until Sep 2023]

PhD Students

- Ayoub Ajarra [INRIA, from Feb 2023]
- Achraf Azize [UNIV LILLE]
- Mickaël Basson [LILLY FRANCE, CIFRE]
- Yann Berthelot [Saint-Gobain Research, CIFRE, from May 2023]
- Nathan Grinsztajn [LIX, until Mar 2023]
- Marc Jourdan [UNIV LILLE]
- Anthony Kobanda [UBISOFT, from Apr 2023]
- Hector Kohler [UNIV LILLE]
- Penanklihi Cyrille Kone [INRIA]
- Matheus Medeiros Centa [UNIV LILLE]
- Thomas Meunier [INRIA, from Oct 2023]
- Reda Ouhamma [UNIV LILLE, until Mar 2023]
- Fabien Pesquerel [ENS PARIS, until Oct 2023]
- Patrick Saux [INRIA]
- Adrienne Tuynman [ENS PARIS-SACLAY, from Oct 2023]
- Sumit Vashishtha [UNIV LILLE]

Technical Staff

- Hernan David Carvajal Bastidas [INRIA, Engineer]
- Brahim Driss [INRIA, Engineer, from Mar 2023]
- Reda Ouhamma [INRIA, from Apr 2023 until May 2023]
- Waris Radji [INRIA, from Oct 2023]
- Tomy Soumphonphakdy [INRIA, Engineer, until Sep 2023]
- Julien Teigny [INRIA, Engineer]

Interns and Apprentices

- Priyank Agrawal [INRIA, Intern, from Aug 2023 until Sep 2023]
- Thomas Delliaux [INRIA, Intern, from Apr 2023 until Sep 2023]
- Thomas Michel [ENS PARIS-SACLAY, Intern, from Feb 2023 until Jul 2023]
- Waris Radji [INRIA, Intern, from Feb 2023 until Aug 2023]
- Etienne Renoult [CRISTAL, Intern, from Apr 2023 until Jul 2023]
- Adrienne Tuynman [ENS PARIS-SACLAY, Intern, from Apr 2023 until Aug 2023]

Administrative Assistant

- Amélie Supervielle [INRIA]

Visiting Scientists

- André Paulo Dantas De Araujo [FFU - NITEROI, from Oct 2023]
- André Paulo Dantas De Araujo [INRIA, from May 2023 until Sep 2023]
- Junya Honda [UNIV KYOTO, until Aug 2023]
- Eduardo Vasconcellos [FFU - NITEROI, until Jun 2023]

2 Overall objectives

Scool is a machine learning (ML) research group. Scool's research focuses on the study of the sequential decision making under uncertainty problem (SDMUP). In particular, we consider bandit problems [50] and the reinforcement learning (RL) problem [49]. In a simplified way, RL considers the problem of learning an optimal policy in a Markov Decision Problem (MDP) [47]; when the set of states collapses to a single state, this is known as the bandit problem which focuses on the exploration/exploitation problem.

Bandit and RL problems are interesting to study on their own; both types of problems share a number of fundamental issues (convergence analysis, sample complexity, representation, safety, *etc.*); both problems have real applications, different though closely related; the fact that while solving an RL problem, one faces an exploration/exploitation problem and has to solve a bandit problem in each state connects the two types of problems very intimately.

In our work, we also consider settings going beyond the Markovian assumption, in particular non-stationary settings, which represents a challenge common to bandits and RL. A distinctive aspect of the SDMUP with regards to the rest of the field of ML is that the learning problem takes place within a closed-loop interaction between a learning agent and its environment. This feedback loop makes our field of research very different from the two other sub-fields of ML, supervised and unsupervised learning, even when they are defined in an incremental setting. Hence, SDMUP combines ML with control: the

learner is not passive, the learner acts on its environment, and learns from the consequences of these interactions; hence, the learner can act in order to obtain information from the environment. Naturally, the optimal control community is getting more and more interested by RL (see e.g. [48]).

We wish to go on, studying applied questions and developing theory to come up with sound approaches to the practical resolution of SDMUP tasks, and guide their resolution. Non-stationary environments are a particularly interesting setting; we are studying this setting and developing new tools to approach it in a sound way, in order to have algorithms to detect environment changes as fast as possible, and as reliably as possible, adapt to them, and prove their behavior, in terms of their performance, measured with the regret for instance. We mostly consider non parametric statistical models, that is models in which the number of parameters is not fixed (a parameter may be of any type: a scalar, a vector, a function, *etc.*), so that the model can adapt along learning, and to its changing environment; this also lets the algorithm learn a representation that fits its environment.

3 Research program

Our research is mostly dealing with bandit problems, and reinforcement learning problems. We investigate each thread separately and also in combination, since the management of the exploration/exploitation trade-off is a major issue in reinforcement learning.

On bandit problems, we focus on:

- structured bandits
- bandits for planning (in particular for Monte Carlo Tree Search (MCTS))
- non stationary bandits

Regarding reinforcement learning, we focus on:

- modeling issues, and dealing with the discrepancy between the model and the task to solve
- learning and using the structure of a Markov decision problem, and of the learned policy
- generalization in reinforcement learning
- reinforcement learning in non stationary environments

Beyond these objectives, we put a particular emphasis on the study of non-stationary environments. Another area of great concern is the combination of symbolic methods with numerical methods, be it to provide knowledge to the learning algorithm to improve its learning curve, or to better understand what the algorithm has learned and explain its behavior, or to rely on causality rather than on mere correlation.

We also put a particular emphasis on real applications and how to deal with their constraints: lack of a simulator, difficulty to have a realistic model of the problem, small amount of data, dealing with risks, availability of expert knowledge on the task.

4 Application domains

Scool has 2 main topics of application:

- health
- sustainable development

In each of these two domains, we put forward the investigation and the application of the idea of sequential decision making under uncertainty. Though supervised and non supervised learning have already been studied and applied extensively, sequential decision making remains far less studied; bandits have already been used in many applications of e-commerce (e.g. for computational advertising and recommendation systems). However, in applications where human beings may be severely impacted,

bandits and reinforcement learning have not been studied much; moreover, these applications come along with a scarcity of data, and the non availability of a simulator, which prevents heavy computational simulations to come up with safe automatic decision making.

In 2022, in health, we investigate patient follow-up with Prof. F. Pattou's research group (CHU Lille, Inserm, Université de Lille) in project B4H. This effort comes along with investigating how we may use medical data available locally at CHU Lille, and also the national social security data. We also investigate drug repurposing with Prof. A. Delahaye-Duriez (Inserm, Université de Paris) in project Repos. We also study catheter control by way of reinforcement learning with Inria Lille group Defrost, and company Robocath (Rouen).

Regarding sustainable development, we have a set of projects and collaborations regarding agriculture and gardening. With Cirad and CGIAR, we investigate how one may recommend agricultural practices to farmers in developing countries. Through an associate team with Bihar Agriculture University (India), we investigate data collection. Inria exploratory action SR4SG concerns recommender systems at the level of individual gardens.

There are two important aspects that are amply shared by these two application fields. First, we consider that data collection is an active task: we do not passively observe and record data, we design methods and algorithms to search for useful data. This idea is exploited in most of these works oriented towards applications. Second, many of these projects include a careful management of risks for human beings. We have to take decisions taking care of their consequences on human beings, on eco-systems and life more generally.

5 Social and environmental responsibility

Sustainable development is a major field of research and application of Scool. We investigate what machine learning can bring to sustainable development, identifiying challenges and obstacles, and studying how to overcome them.

Let us mention here:

- sustainable agriculture in developing countries;
- sustainable gardening.

More details can be found in Section 4.

6 Highlights of the year

- Publication of our work in collaboration with CHU Lille/Inserm in *The Lancet Digital Health* journal [14].

7 New software, platforms, open data

7.1 New software

7.1.1 rlberry

Keywords: Reinforcement learning, Simulation, Artificial intelligence

Functional Description: rlberry is a reinforcement learning (RL) library in Python for research and education. The library provides implementations of several RL agents for you to use as a starting point or as baselines, provides a set of benchmark environments, very useful to debug and challenge your algorithms, handles all random seeds for you, ensuring reproducibility of your results, and is fully compatible with several commonly used RL libraries like OpenAI gym and Stable Baselines.

URL: https://github.com/rlberry-py/rlberry

Contact: Timothee Mathieu

7.1.2 gym-DSSAT

Keywords: Reinforcement learning, Crop management, Sequential decision making under uncertainty, Mechanistic modeling

Functional Description: gym-DSSAT let you (learn to) manage a crop parcel, from seed selection, to daily activity in the field, to harvesting.

URL: https://gitlab.inria.fr/rgautron/gym_dssat_pdi

Contact: Romain Gautron

Partners: CIRAD, Cgiar

7.1.3 Weight Trajectory Predictor : algorithm

Name: Weight Trajectory Predictor : algorithm

Keywords: Medical applications, Machine learning

Scientific Description: We performed a retrospective study of clinical data collected prospectively on patients with up to five years postoperative follow-up (ABOS cohort, CHU Lille) and trained a supervised model to predict the relative total weight loss ("%TWL") of a patient 1, 3, 12, 24 and 60 months after surgery. This model consists in a decision tree, written in python, taking as input a selected subset of preoperative attributes (weight, height, type of intervention, age, presence or absence of type 2 diabetes or impaired glucose tolerance, diabetes duration, smoking habits) and returns an estimation of %TWL as well as a prediction interval based on the interquartile range of %TWL observed on similar patients. The predictions of this tool have been validated both internally and externally (on French and Dutch cohorts).

Functional Description: The "Weight Tracjectory Predictor" algorithm is part of a larger project, whose goal is to leverage artificial intelligence techniques to improve patient care. This code is the product of a collaboration between Inria SCOOL and the UMR 1190-EGID team of the CHU Lille. It aims to predict the weight loss trajectory of a patient following bariatric surgery (treatment of severe obesity) from a set of preoperative characteristics.

We performed a retrospective study of clinical data collected prospectively on patients with up to five years postoperative follow-up (ABOS cohort, CHU Lille) and trained a supervised model to predict the relative total weight loss ("%TWL") of a patient 1, 3, 12, 24 and 60 months after surgery. This model consists in a decision tree, written in python, taking as input a selected subset of preoperative attributes (weight, height, type of intervention, age, presence or absence of type 2 diabetes or impaired glucose tolerance, diabetes duration, smoking habits) and returns an estimation of %TWL as well as a prediction interval based on the interquartile range of %TWL observed on similar patients. The predictions of this tool have been validated both internally and externally (on French and Dutch cohorts).

The goal of this software is to improve patient follow-up after bariatric surgery: - during preoperative visits, by providing clinicians with a quantitative tool to inform the patient regarding potential weight loss outcome. - during postoperative control visits, by comparing the predicted and realized weight trajectories, which may facilitate early detection of complications.

This software component will be embedded in a web app for ease of use.

Release Contributions: Initial version

URL: https://bariatric-weight-trajectory-prediction.univ-lille.fr/

Contact: Julien Teigny

Participants: Pierre Bauvin, Francois Pattou, Philippe Preux, Violeta Raverdy, Patrick Saux, Tomy Soumphonphakdy, Julien Teigny, Hélène Verkindt

Partner: CHU de Lille

7.1.4 Adastop

Keywords: Hypothesis testing, Reinforcement learning, Reproducibility

Functional Description: This package contains the AdaStop algorithm. AdaStop implements a statistical test to adaptively choose the number of runs of stochastic algorithms necessary to compare these algorithms and be able to rank them with a theoretically controlled family-wise error rate. One particular application for which AdaStop was created is to compare Reinforcement Learning algorithms. Please note, that what we call here an algorithm is really a certain implementation of an algorithm.

URL: https://github.com/TimotheeMathieu/adastop

Contact: Timothee Mathieu

7.1.5 average-reward-reinforcement-learning

Keywords: Mutli-armed bandits, Reinforcement learning, Python

Functional Description: Library of RL and Bandit algorithms.

URL: https://gitlab.inria.fr/omaillar/average-reward-reinforcement-learning

Contact: Odalric-Ambrym Maillard

Participant: Odalric-Ambrym Maillard

7.1.6 FarmGym

Name: Farming Environment Gym factory for Reinforcement Learning

Keywords: Reinforcement learning, Simulator, Agroecology

Functional Description: Farming Environment Gym factory for Reinforcement Learning

Release Contributions: This version is an entire rewriting by Odlaric-Ambrym Maillard of the prototype V1 created by Thomas Carta. Version V2 features modular creation of farms, specifications of various entities and monitoring facilities. Recent additions include slight adjustment of the entities dynamics. It also features unit tests and an automatic generation of base policies done by Brahim Driss.

News of the Year: Nov 2022: Déploiement de FarmGym sur gitlab, auparavant en développement interne par Odalric-Ambrym Maillard (2021-2022). Nov-Dec 2022: Organisation de compétition interne par Timothée Mathieu.

Printemps 2023: Arrivée de Brahim Driss sur le projet, pour mettre en place tests unitaires/fonctionnels et assister Odalric-Ambrym Maillard dans le développement de fonctionnalité.

URL: https://github.com/farm-gym/

Publication: hal-03960683

Contact: Odalric-Ambrym Maillard

Participants: Odalric-Ambrym Maillard, Brahim Driss, Timothee Mathieu

Partner: Inria

8 New results

We organize our research results in a set of categories. The main categories are: bandit problems, reinforcement learning problems, and applications.

Participants: all Scool members.

8.1 Bandits and RL theory

An ϵ-Best-Arm Identification Algorithm for Fixed-Confidence and Beyond, [25]

We propose EB-TC ϵ , a novel sampling rule for ϵ-best arm identification in stochastic bandits. It is the first instance of Top Two algorithm analyzed for approximate best arm identification. EB-TC ϵ is an anytime sampling rule that can therefore be employed without modification for fixed confidence or fixed budget identification (without prior knowledge of the budget). We provide three types of theoretical guarantees for EB-TC ϵ. First, we prove bounds on its expected sample complexity in the fixed confidence setting, notably showing its asymptotic optimality in combination with an adaptive tuning of its exploration parameter. We complement these findings with upper bounds on its probability of error at any time and for any error parameter, which further yield upper bounds on its simple regret at any time. Finally, we show through numerical simulations that EB-TC ϵ performs favorably compared to existing algorithms, in different settings.

Optimistic PAC Reinforcement Learning: the Instance-Dependent View, [32]

Optimistic algorithms have been extensively studied for regret minimization in episodic tabular Markov Decision Processes (MDPs), both from a minimax and an instance-dependent view. However, for the PAC RL problem, where the goal is to identify a near-optimal policy with high probability, little is known about their instance-dependent sample complexity. A negative result of Wagenmaker et al. (2022) suggests that optimistic sampling rules cannot be used to attain the (still elusive) optimal instance-dependent sample complexity. On the positive side, we provide the first instance-dependent bound for an optimistic algorithm for PAC RL, BPI-UCRL, for which only minimax guarantees were available (Kaufmann et al., 2021). While our bound features some minimal visitation probabilities, it also features a refined notion of sub-optimality gap compared to the value gaps that appear in prior work. Moreover, in MDPs with deterministic transitions, we show that BPI-UCRL is actually near instance-optimal (up to a factor of the horizon). On the technical side, our analysis is very simple thanks to a new "target trick" of independent interest. We complement these findings with a novel hardness result explaining why the instance-dependent complexity of PAC RL cannot be easily related to that of regret minimization, unlike in the minimax regime.

Bilinear Exponential Family of MDPs: Frequentist Regret Bound with Tractable Exploration & Planning, [30]

We study the problem of episodic reinforcement learning in continuous stateaction spaces with unknown rewards and transitions. Specifically, we consider the setting where the rewards and transitions are modeled using parametric bilinear exponential families. We propose an algorithm, BEF-RLSVI, that a) uses penalized maximum likelihood estimators to learn the unknown parameters, b) injects a calibrated Gaussian noise in the parameter of rewards to ensure exploration, and c) leverages linearity of the exponential family with respect to an underlying RKHS to perform tractable planning. We further provide a frequentist regret analysis of BEF-RLSVI that yields an upper bound of $\tilde{O}((d^3H^3K)^{1/2})$, where d is the dimension of the parameters, H is the episode length, and K is the number of episodes. Our analysis improves the existing bounds for the bilinear exponential family of MDPs by $\sqrt{H}$ and removes the handcrafted clipping deployed in existing RLSVI-type algorithms. Our regret bound is order-optimal with respect to H and K.

Bregman Deviations of Generic Exponential Families, [20]

We revisit the method of mixture technique, also known as the Laplace method, to study the concentration phenomenon in generic exponential families. Combining the properties of Bregman divergence associated with log-partition function of the family with the method of mixtures for super-martingales,

we establish a generic bound controlling the Bregman divergence between the parameter of the family and a finite sample estimate of the parameter. Our bound is time-uniform and makes appear a quantity extending the classical information gain to exponential families, which we call the Bregman information gain. For the practitioner, we instantiate this novel bound to several classical families, e.g., Gaussian, Bernoulli, Exponential, Weibull, Pareto, Poisson and Chi-square yielding explicit forms of the confidence sets and the Bregman information gain. We further numerically compare the resulting confidence bounds to state-of-the-art alternatives for time-uniform concentration and show that this novel method yields competitive results. Finally, we highlight the benefit of our concentration bounds on some illustrative applications.

Active Coverage for PAC Reinforcement Learning, [29]

Collecting and leveraging data with good coverage properties plays a crucial role in different aspects of reinforcement learning (RL), including reward-free exploration and offline learning. However, the notion of "good coverage" really depends on the application at hand, as data suitable for one context may not be so for another. In this paper, we formalize the problem of active coverage in episodic Markov decision processes (MDPs), where the goal is to interact with the environment so as to fulfill given sampling requirements. This framework is sufficiently flexible to specify any desired coverage property, making it applicable to any problem that involves online exploration. Our main contribution is an instance-dependent lower bound on the sample complexity of active coverage and a simple game-theoretic algorithm, COVGAME, that nearly matches it. We then show that COVGAME can be used as a building block to solve different PAC RL tasks. In particular, we obtain a simple algorithm for PAC reward-free exploration with an instance-dependent sample complexity that, in certain MDPs which are "easy to explore", is lower than the minimax one. By further coupling this exploration algorithm with a new technique to do implicit eliminations in policy space, we obtain a computationally-efficient algorithm for best-policy identification whose instance-dependent sample complexity scales with gaps between policy values.

On the Existence of a Complexity in Fixed Budget Bandit Identification, [22]

In fixed budget bandit identification, an algorithm sequentially observes samples from several distributions up to a given final time. It then answers a query about the set of distributions. A good algorithm will have a small probability of error. While that probability decreases exponentially with the final time, the best attainable rate is not known precisely for most identification tasks. We show that if a fixed budget task admits a complexity, defined as a lower bound on the probability of error which is attained by the same algorithm on all bandit problems, then that complexity is determined by the best non-adaptive sampling procedure for that problem. We show that there is no such complexity for several fixed budget identification tasks including Bernoulli best arm identification with two arms: there is no single algorithm that attains everywhere the best possible rate.

Non-Asymptotic Analysis of a UCB-based Top Two Algorithm, [24]

A Top Two sampling rule for bandit identification is a method which selects the next arm to sample from among two candidate arms, a leader and a challenger. Due to their simplicity and good empirical performance, they have received increased attention in recent years. For fixed-confidence best arm identification, theoretical guarantees for Top Two methods have only been obtained in the asymptotic regime, when the error level vanishes. We derive the first non-asymptotic upper bound on the expected sample complexity of a Top Two algorithm holding for any error level. Our analysis highlights sufficient properties for a regret minimization algorithm to be used as leader. They are satisfied by the UCB algorithm and our proposed UCB-based Top Two algorithm enjoys simultaneously non-asymptotic guarantees and competitive empirical performance.

Fast Asymptotically Optimal Algorithms for Non-Parametric Stochastic Bandits, [17]

We consider the problem of regret minimization in non-parametric stochastic bandits. When the rewards are known to be bounded from above, there exists asymptotically optimal algorithms, with asymptotic regret depending on an infimum of Kullback-Leibler divergences (KL). These algorithms are computationally expensive and require storing all past rewards, thus simpler but non-optimal algorithms are often used instead. We introduce several methods to approximate the infimum KL which reduce drastically the computational and memory costs of existing optimal algorithms, while keeping their regret guaranties. We apply our findings to design new variants of the MED and IMED algorithms, and demonstrate their interest with extensive numerical simulations.

8.2 Bandits and RL under Real-life constraints

CRIMED: Lower and Upper Bounds on Regret for Bandits with Unbounded Stochastic Corruption, [40]

We investigate the regret-minimisation problem in a multi-armed bandit setting with arbitrary corruptions. Similar to the classical setup, the agent receives rewards generated independently from the distribution of the arm chosen at each time. However, these rewards are not directly observed. Instead, with a fixed $\varepsilon \in (0, \frac{1}{2})$, the agent observes a sample from the chosen arm's distribution with probability $1-\varepsilon$, or from an arbitrary corruption distribution with probability ε. Importantly, we impose no assumptions on these corruption distributions, which can be unbounded. In this setting, accommodating potentially unbounded corruptions, we establish a problem-dependent lower bound on regret for a given family of arm distributions. We introduce CRIMED, an asymptotically-optimal algorithm that achieves the exact lower bound on regret for bandits with Gaussian distributions with known variance. Additionally, we provide a finite-sample analysis of CRIMED's regret performance. Notably, CRIMED can effectively handle corruptions with ε values as high as $\frac{1}{2}$. Furthermore, we develop a tight concentration result for medians in the presence of arbitrary corruptions, even with ε values up to $\frac{1}{2}$, which may be of independent interest. We also discuss an extension of the algorithm for handling misspecification in Gaussian model.

Risk-aware linear bandits with convex loss, [31]

In decision-making problems such as the multi-armed bandit, an agent learns sequentially by optimizing a certain feedback. While the mean reward criterion has been extensively studied, other measures that reflect an aversion to adverse outcomes, such as mean-variance or conditional value-at-risk (CVaR), can be of interest for critical applications (healthcare, agriculture). Algorithms have been proposed for such risk-aware measures under bandit feedback without contextual information. In this work, we study contextual bandits where such risk measures can be elicited as linear functions of the contexts through the minimization of a convex loss. A typical example that fits within this framework is the expectile measure, which is obtained as the solution of an asymmetric least-square problem. Using the method of mixtures for supermartingales, we derive confidence sequences for the estimation of such risk measures. We then propose an optimistic UCB algorithm to learn optimal risk-aware actions, with regret guarantees similar to those of generalized linear bandits. This approach requires solving a convex problem at each round of the algorithm, which we can relax by allowing only approximated solution obtained by online gradient descent, at the cost of slightly higher regret. We conclude by evaluating the resulting algorithms on numerical experiments.

On the Complexity of Differentially Private Best-Arm Identification with Fixed Confidence, [16]

Best Arm Identification (BAI) problems are progressively used for data-sensitive applications, such as designing adaptive clinical trials, tuning hyper-parameters, and conducting user studies to name a few. Motivated by the data privacy concerns invoked by these applications, we study the problem of BAI with fixed confidence under ϵ-global Differential Privacy (DP). First, to quantify the cost of privacy, we derive a lower bound on the sample complexity of any δ-correct BAI algorithm satisfying ϵ-global DP. Our lower bound suggests the existence of two privacy regimes depending on the privacy budget ϵ. In the high-privacy regime (small ϵ), the hardness depends on a coupled effect of privacy and a novel information-theoretic quantity, called the Total Variation Characteristic Time. In the low-privacy regime (large ϵ), the sample complexity lower bound reduces to the classical non-private lower bound. Second, we propose AdaP-TT, an ϵ-global DP variant of the Top Two algorithm. AdaP-TT runs in arm-dependent adaptive episodes and adds Laplace noise to ensure a good privacy-utility trade-off. We derive an asymptotic upper bound on the sample complexity of AdaP-TT that matches with the lower bound up to multiplicative constants in the high-privacy regime. Finally, we provide an experimental analysis of AdaP-TT that validates our theoretical results.

Interactive and Concentrated Differential Privacy for Bandits, [15]

Bandits play a crucial role in interactive learning schemes and modern recommender systems. However, these systems often rely on sensitive user data, making privacy a critical concern. This paper investigates privacy in bandits with a trusted centralized decision-maker through the lens of interactive Differential Privacy (DP). While bandits under pure ϵ-global DP have been well-studied, we contribute to the understanding of bandits under zero Concentrated DP (zCDP). We provide minimax and problem-dependent lower bounds on regret for finite-armed and linear bandits, which quantify the

cost of ρ-global zCDP in these settings. These lower bounds reveal two hardness regimes based on the privacy budget ρ and suggest that ρ-global zCDP incurs less regret than pure ϵ-global DP. We propose two ρ-global zCDP bandit algorithms, AdaC-UCB and AdaC-GOPE, for finite-armed and linear bandits respectively. Both algorithms use a common recipe of Gaussian mechanism and adaptive episodes. We analyze the regret of these algorithms to show that AdaC-UCB achieves the problem-dependent regret lower bound up to multiplicative constants, while AdaC-GOPE achieves the minimax regret lower bound up to poly-logarithmic factors. Finally, we provide experimental validation of our theoretical results under different settings.

Dealing with Unknown Variances in Best-Arm Identification, [26]

The problem of identifying the best arm among a collection of items having Gaussian rewards distribution is well understood when the variances are known. Despite its practical relevance for many applications, few works studied it for unknown variances. In this paper we introduce and analyze two approaches to deal with unknown variances, either by plugging in the empirical variance or by adapting the transportation costs. In order to calibrate our two stopping rules, we derive new time-uniform concentration inequalities, which are of independent interest. Then, we illustrate the theoretical and empirical performances of our two sampling rule wrappers on Track-and-Stop and on a Top Two algorithm. Moreover, by quantifying the impact on the sample complexity of not knowing the variances, we reveal that it is rather small.

Soft Action Priors: Towards Robust Policy Transfer, [19]

Despite success in many challenging problems, reinforcement learning (RL) is still confronted with sample inefficiency, which can be mitigated by introducing prior knowledge to agents. However, many transfer techniques in reinforcement learning make the limiting assumption that the teacher is an expert. In this paper, we use the action prior from the Reinforcement Learning as Inference framework (Levine 2018)-that is, a distribution over actions at each state which resembles a teacher policy, rather than a Bayesian prior-to recover state-of-the-art policy distillation techniques. Then, we propose a class of adaptive methods that can robustly exploit action priors by combining reward shaping and auxiliary regularization losses. In contrast to prior work, we develop algorithms for leveraging suboptimal action priors that may nevertheless impart valuable knowledge-which we call soft action priors. The proposed algorithms adapt by adjusting the strength of teacher feedback according to an estimate of the teacher's usefulness in each state. We perform tabular experiments, which show that the proposed methods achieve state-of-the-art performance, surpassing it when learning from suboptimal priors. Finally, we demonstrate the robustness of the adaptive algorithms in continuous action deep RL problems, in which adaptive algorithms considerably improved stability when compared to existing policy distillation methods.

Pure Exploration in Bandits with Linear Constraints, [18]

We address the problem of identifying the optimal policy with a fixed confidence level in a multi-armed bandit setup, when the arms are subject to linear constraints. Unlike the standard best-arm identification problem which is well studied, the optimal policy in this case may not be deterministic and could mix between several arms. This changes the geometry of the problem which we characterize via an information-theoretic lower bound. We introduce two asymptotically optimal algorithms for this setting, one based on the Track-and-Stop method and the other based on a game-theoretic approach. Both these algorithms try to track an optimal allocation based on the lower bound and computed by a weighted projection onto the boundary of a normal cone. Finally, we provide empirical results that validate our bounds and visualize how constraints change the hardness of the problem.

Online Instrumental Variable Regression: Regret Analysis and Bandit Feedback, [41]

The independence of noise and covariates is a standard assumption in online linear regression with unbounded noise and linear bandit literature. This assumption and the following analysis are invalid in the case of endogeneity, i.e., when the noise and covariates are correlated. In this paper, we study the online setting of Instrumental Variable (IV) regression, which is widely used in economics to identify the underlying model from an endogenous dataset. Specifically, we upper bound the identification and oracle regrets of the popular Two-Stage Least Squares (2SLS) approach to IV regression but in the online setting. Our analysis shows that Online 2SLS (O2SLS) achieves $\mathcal{O}(d^2 \log^2 T)$ identification and $\mathcal{O}(\gamma\sqrt{dT\log T})$ oracle regret after T interactions, where d is the dimension of covariates and γ is the bias due to endogeneity.Then, we leverage O2SLS as an oracle to design OFUL-IV, a linear bandit algorithm.

OFUL-IV can tackle endogeneity and achieves $\mathcal{O}(d\sqrt{T}\log T)$ regret.For datasets with endogeneity, we experimentally show the efficiency of OFUL-IV in terms of estimation error and regret.

8.3 Bandits and RL for real-life: Deep RL and Applications

Farm-gym: A modular reinforcement learning platform for stochastic agronomic games, [36]

We introduce Farm-gym, an open-source farming environment written in Python, that models sequential decisionmaking in farms using Reinforcement Learning (RL). Farm-gym conceptualizes a farm as a dynamical system with many interacting entities. Leveraging a modular design, it enables us to instantiate from very simple to highly complicated environments. Contrasting many available gym environments, Farm-gym features intrinsically stochastic games, using stochastic growth models and weather data. Further, it enables to create farm games in a modular way, activating or not the entities (e.g. weeds, pests, pollinators), and yielding non-trivial coupled dynamics. Finally, every game can be customized with .yaml files for rewards, feasible actions, and initial/end-game conditions. We illustrate some interesting features on simple farms. We also showcase the challenges posed by Farm-gym to the deep RL algorithms, in order to stimulate studies in the RL community.

Learning crop management by reinforcement: gym-DSSAT, [35]

We introduce gym-DSSAT, a gym environment for crop management tasks, that is easy to use for training Reinforcement Learning (RL) agents. gym-DSSAT is based on DSSAT, a state-of-the-art mechanistic crop growth simulator. We modify DSSAT so that an external software agent can interact with it to control the actions performed in a crop field during a growing season. The RL environment provides predefined decision problems without having to manipulate the complex crop simulator. We report encouraging preliminary results on a use case of nitrogen fertilization for maize. This work opens up opportunities to explore new sustainable crop management strategies with RL, and provides RL researchers with an original set of challenging tasks to investigate.

Adaptive Algorithms for Relaxed Pareto Set Identification, [28]

In this paper we revisit the fixed-confidence identification of the Pareto optimal set in a multi-objective multi-armed bandit model. As the sample complexity to identify the exact Pareto set can be very large, a relaxation allowing to output some additional near-optimal arms has been studied. In this work we also tackle alternative relaxations that allow instead to identify a relevant subset of the Pareto set. Notably, we propose a single sampling strategy, called Adaptive Pareto Exploration, that can be used in conjunction with different stopping rules to take into account different relaxations of the Pareto Set Identification problem. We analyze the sample complexity of these different combinations, quantifying in particular the reduction in sample complexity that occurs when one seeks to identify at most k Pareto optimal arms. We showcase the good practical performance of Adaptive Pareto Exploration on a real-world scenario, in which we adaptively explore several vaccination strategies against Covid-19 in order to find the optimal ones when multiple immunogenicity criteria are taken into account.

Reinforcement Learning in the Wild with Maximum Likelihood-based Model Transfer, [42]

In this paper, we study the problem of transferring the available Markov Decision Process (MDP) models to learn and plan efficiently in an unknown but similar MDP. We refer to it as *Model Transfer Reinforcement Learning (MTRL)* problem. First, we formulate MTRL for discrete MDPs and Linear Quadratic Regulators (LQRs) with continuous state actions. Then, we propose a generic two-stage algorithm, MLEMTRL, to address the MTRL problem in discrete and continuous settings. In the first stage, MLEMTRL uses a *constrained Maximum Likelihood Estimation (MLE)*-based approach to estimate the target MDP model using a set of known MDP models. In the second stage, using the estimated target MDP model, MLEMTRL deploys a model-based planning algorithm appropriate for the MDP class. Theoretically, we prove worst-case regret bounds for MLEMTRL both in realisable and non-realisable settings. We empirically demonstrate that MLEMTRL allows faster learning in new MDPs than learning from scratch and achieves near-optimal performance depending on the similarity of the available MDPs and the target MDP.

8.4 Other

8.4.1 Methodology

AdaStop: sequential testing for efficient and reliable comparisons of Deep RL Agents, [45]

The reproducibility of many experimental results in Deep Reinforcement Learning (RL) is under question. To solve this reproducibility crisis, we propose a theoretically sound methodology to compare multiple Deep RL algorithms. The performance of one execution of a Deep RL algorithm is random so that independent executions are needed to assess it precisely. When comparing several RL algorithms, a major question is how many executions must be made and how can we assure that the results of such a comparison is theoretically sound. Researchers in Deep RL often use less than 5 independent executions to compare algorithms: we claim that this is not enough in general. Moreover, when comparing several algorithms at once, the error of each comparison accumulates and must be taken into account with a multiple tests procedure to preserve low error guarantees. To address this problem in a statistically sound way, we introduce AdaStop, a new statistical test based on multiple group sequential tests. When comparing algorithms, AdaStop adapts the number of executions to stop as early as possible while ensuring that we have enough information to distinguish algorithms that perform better than the others in a statistical significant way. We prove both theoretically and empirically that AdaStop has a low probability of making an error (Family-Wise Error). Finally, we illustrate the effectiveness of AdaStop in multiple use-cases, including toy examples and difficult cases such as Mujoco environments.

8.4.2 Digital health

Impact of Robotic Assistance on Complications in Bariatric Surgery at Expert Laparoscopic Surgery Centers: A Retrospective Comparative Study With Propensity Score, [12]

Objective: To investigate the way robotic assistance affected rate of complications in bariatric surgery at expert robotic and laparoscopic surgery facilities. Background: While the benefits of robotic assistance were established at the beginning of surgical training, there is limited data on the robot's influence on experienced bariatric laparoscopic surgeons. Methods: We conducted a retrospective study using the BRO clinical database (2008–2022) collecting data of patients operated on in expert centers. We compared the serious complication rate (defined as a Clavien score≥3) in patients undergoing metabolic bariatric surgery with or without robotic assistance. We used a directed acyclic graph to identify the variables adjustment set used in a multivariable linear regression, and a propensity score matching to calculate the average treatment effect (ATE) of robotic assistance. Results: The study included 35,043 patients [24,428 sleeve gastrectomy (SG); 10,452 Roux-en-Y gastric bypass (RYGB); 163 single anastomosis duodenal-ileal bypass with sleeve gastrectomy (SADI-S)], with 938 operated on with robotic assistance (801 SG; 134 RYGB; 3 SADI-S), among 142 centers. Overall, we found no benefit of robotic assistance regarding the risk of complications (average treatment effect=-0.05, P =0.794), with no difference in the RYGB+SADI group (P =0.322) but a negative trend in the SG group (more complications, P =0.060). Length of hospital stay was decreased in the robot group (3.7 ± 11.1 vs 4.0 ± 9.0 days, $P < 0.001$). Conclusions: Robotic assistance reduced the length of stay but did not statistically significantly reduce postoperative complications (Clavien score≥ 3) following either GBP or SG. A tendency toward an elevated risk of complications following SG requires more supporting studies.

Development and validation of an interpretable machine learning-based calculator for predicting 5-year weight trajectories after bariatric surgery: a multinational retrospective cohort SOPHIA study, [14]

Background Weight loss trajectories after bariatric surgery vary widely between individuals, and predicting weight loss before the operation remains challenging. We aimed to develop a model using machine learning to provide individual preoperative prediction of 5-year weight loss trajectories after surgery. Methods In this multinational retrospective observational study we enrolled adult participants (aged ≥18 years) from ten prospective cohorts (including ABOS [NCT01129297], BAREVAL [NCT02310178], the Swedish Obese Subjects study, and a large cohort from the Dutch Obesity Clinic [Nederlandse Obesitas Kliniek]) and two randomised trials (SleevePass [NCT00793143] and SM-BOSS [NCT00356213]) in Europe, the Americas, and Asia, with a 5 year followup after Roux-en-Y gastric bypass, sleeve gastrectomy, or gastric band. Patients with a previous history of bariatric surgery or large delays between scheduled and

actual visits were excluded. The training cohort comprised patients from two centres in France (ABOS and BAREVAL). The primary outcome was BMI at 5 years. A model was developed using least absolute shrinkage and selection operator to select variables and the classification and regression trees algorithm to build interpretable regression trees. The performances of the model were assessed through the median absolute deviation (MAD) and root mean squared error (RMSE) of BMI. Findings10 231 patients from 12 centres in ten countries were included in the analysis, corresponding to 30 602 patient-years. Among participants in all 12 cohorts, 7701 (75.3%) were female, 2530 (24.7%) were male. Among 434 baseline attributes available in the training cohort, seven variables were selected: height, weight, intervention type, age, diabetes status, diabetes duration, and smoking status. At 5 years, across external testing cohorts the overall mean MAD BMI was 2.8 kg/m^2 (95% CI 2.6-3.0) and mean RMSE BMI was 4.7 kg/m^2 (4.4-5.0), and the mean difference between predicted and observed BMI was -0.3 kg/m^2 (SD 4.7). This model is incorporated in an easy to use and interpretable web-based prediction tool to help inform clinical decision before surgery. InterpretationWe developed a machine learning-based model, which is internationally validated, for predicting individual 5-year weight loss trajectories after three common bariatric interventions.

Elbow trauma in children: development and evaluation of radiological artificial intelligence models, [13]

Rationale and Objectives: To develop a model using artificial intelligence (A.I.) able to detect post-traumatic injuries on pediatric elbow X-rays then to evaluate its performances in silico and its impact on radiologists' interpretation in clinical practice. Material and Methods: A total of 1956 pediatric elbow radiographs performed following a trauma were retrospectively collected from 935 patients aged between 0 and 18 years. Deep convolutional neural networks were trained on these X-rays. The two best models were selected then evaluated on an external test set involving 120 patients, whose X-rays were performed on a different radiological equipment in another time period. Eight radiologists interpreted this external test set without then with the help of the A.I. models. Results: Two models stood out: model 1 had an accuracy of 95.8% and an AUROC of 0.983 and model 2 had an accuracy of 90.5% and an AUROC of 0.975. On the external test set, model 1 kept a good accuracy of 82.5% and AUROC of 0.916 while model 2 had a loss of accuracy down to 69.2% and of AUROC to 0.793. Model 1 significantly improved radiologist's sensitivity (0.82 to 0.88, P = 0.016) and accuracy (0.86 to 0.88, P = 0,047) while model 2 significantly decreased specificity of readers (0.86 to 0.83, P = 0.031). Conclusion: End-to-end development of a deep learning model to assess post-traumatic injuries on elbow Xray in children was feasible and showed that models with close metrics in silico can unpredictably lead radiologists to either improve or lower their performances in clinical settings.

8.4.3 Sailboat digital twin

Reinforcement-learning robotic sailboats: simulator and preliminary results, [33]

This work focuses on the main challenges and problems in developing a virtual oceanic environment reproducing real experiments using Unmanned Surface Vehicles (USV) digital twins. We introduce the key features for building virtual worlds, considering using Reinforcement Learning (RL) agents for autonomous navigation and control. With this in mind, the main problems concern the definition of the simulation equations (physics and mathematics), their effective implementation, and how to include strategies for simulated control and perception (sensors) to be used with RL. We present the modeling, implementation steps, and challenges required to create a functional digital twin based on a real robotic sailing vessel. The application is immediate for developing navigation algorithms based on RL to be applied on real boats.

General System Architecture and COTS Prototyping of an AIoT-Enabled Sailboat for Autonomous Aquatic Ecosystem Monitoring, [11]

Unmanned vehicles keep growing attention as they facilitate innovative commercial and civil applications within the Internet of Things (IoT) realm. In this context, autonomous sailing boats are becoming important marine platforms for performing different tasks, such as surveillance, water, and environmental monitoring. Most of these tasks heavily depend on artificial intelligence (AI) technologies, such as visual navigation and path planning, and comprise the so-called Artificial Intelligence of Things (AIoT). In this paper, we propose (i) the OpenBoat, an automating system architecture for AIoTenabled sailboats with application-agnostic autonomous environment monitoring capability, and (ii) the F-Boat, a fully

functional prototype of OpenBoat built with Commercial Off-The-Shelf (COTS) components on a real sailboat. F-Boat includes low-level control strategies for autonomous path following, communication infrastructure for remote operation and cooperation with other systems, edge computing with AI accelerator, modular support for application-specific monitoring systems, and navigation aspects. F-Boat is also designed and built for robustness situations to guarantee its operation under extreme events, such as high temperatures and bad weather, through extended periods of time. We show the results of field experiments running in Guanabara Bay, an important aquatic ecosystem in Brazil, that demonstrate the functionalities of the prototype and demonstrate the AIoT capability of the proposed architecture.

8.4.4 Computer-assisted mathematics

A Formalization of Doob's Martingale Convergence Theorems in mathlib, [34]

We present the formalization of Doob's martingale convergence theorems in the mathlib library for the Lean theorem prover. These theorems give conditions under which (sub)martingales converge, almost everywhere or in L^1. In order to formalize those results, we build a definition of the conditional expectation in Banach spaces and develop the theory of stochastic processes, stopping times and martingales. As an application of the convergence theorems, we also present the formalization of Lévy's generalized Borel-Cantelli lemma. This work on martingale theory is one of the first developments of probability theory in mathlib, and it builds upon diverse parts of that library such as topology, analysis and most importantly measure theory.

8.4.5 Interpretability

Optimal Interpretability-Performance Trade-off of Classification Trees with Black-Box Reinforcement Learning, [43]

Interpretability of AI models allows for user safety checks to build trust in these models. In particular, decision trees (DTs) provide a global view on the learned model and clearly outlines the role of the features that are critical to classify a given data. However, interpretability is hindered if the DT is too large. To learn compact trees, a Reinforcement Learning (RL) framework has been recently proposed to explore the space of DTs. A given supervised classification task is modeled as a Markov decision problem (MDP) and then augmented with additional actions that gather information about the features, equivalent to building a DT. By appropriately penalizing these actions, the RL agent learns to optimally trade-off size and performance of a DT. However, to do so, this RL agent has to solve a partially observable MDP. The main contribution of this paper is to prove that it is sufficient to solve a fully observable problem to learn a DT optimizing the interpretability-performance trade-off. As such any planning or RL algorithm can be used. We demonstrate the effectiveness of this approach on a set of classical supervised classification datasets and compare our approach with other interpretability-performance optimizing methods.

"How Biased are Your Features?": Computing Fairness Influence Functions with Global Sensitivity Analysis, [23]

Fairness in machine learning has attained significant focus due to the widespread application in high-stake decision-making tasks. Unregulated machine learning classifiers can exhibit bias towards certain demographic groups in data, thus the quantification and mitigation of classifier bias is a central concern in fairness in machine learning. In this paper, we aim to quantify the influence of different features in a dataset on the bias of a classifier. To do this, we introduce the Fairness Influence Function (FIF). This function breaks down bias into its components among individual features and the intersection of multiple features. The key idea is to represent existing group fairness metrics as the difference of the scaled conditional variances in the classifier's prediction and apply a decomposition of variance according to global sensitivity analysis. To estimate FIFs, we instantiate an algorithm that applies variance decomposition of classifier's prediction following local regression. Experiments demonstrate that captures FIFs of individual feature and intersectional features, provides a better approximation of bias based on FIFs, demonstrates higher correlation of FIFs with fairness interventions, and detects changes in bias due to fairness affirmative/punitive actions in the classifier. The code is available online.

8.4.6 Privacy

From Noisy Fixed-Point Iterations to Private ADMM for Centralized and Federated Learning, [21]

We study differentially private (DP) machine learning algorithms as instances of noisy fixed-point iterations, in order to derive privacy and utility results from this well-studied framework. We show that this new perspective recovers popular private gradient-based methods like DP-SGD and provides a principled way to design and analyze new private optimization algorithms in a flexible manner. Focusing on the widely-used Alternating Directions Method of Multipliers (ADMM) method, we use our general framework to derive novel private ADMM algorithms for centralized, federated and fully decentralized learning. For these three algorithms, we establish strong privacy guarantees leveraging privacy amplification by iteration and by subsampling. Finally, we provide utility guarantees using a unified analysis that exploits a recent linear convergence result for noisy fixed-point iterations.

Marich: A Query-efficient Distributionally Equivalent Model Extraction Attack using Public Data, [27]

We study design of black-box model extraction attacks that can send minimal number of queries from a publicly available dataset to a target ML model through a predictive API with an aim to create an informative and distributionally equivalent replica of the target. First, we define distributionally equivalent and Max-Information model extraction attacks, and reduce them into a variational optimisation problem. The attacker sequentially solves this optimisation problem to select the most informative queries that simultaneously maximise the entropy and reduce the mismatch between the target and the stolen models. This leads to an active sampling-based query selection algorithm, Marich, which is model-oblivious. Then, we evaluate Marich on different text and image data sets, and different models, including CNNs and BERT. Marich extracts models that achieve $\sim 60-95\%$ of true model's accuracy and uses $\sim 1,000-8,500$ queries from the publicly available datasets, which are different from the private training datasets. Models extracted by Marich yield prediction distributions, which are $\sim 2-4\times$ closer to the target's distribution in comparison to the existing active sampling-based attacks. The extracted models also lead to $84-96\%$ accuracy under membership inference attacks. Experimental results validate that Marich is query-efficient, and capable of performing task-accurate, high-fidelity, and informative model extraction.

9 Bilateral contracts and grants with industry

Participants: Odalric-Ambrym Maillard, Philippe Preux.

9.1 Bilateral contracts with industry

- contract with Ubisoft, 2023–2026: PI: O-A. Maillard.
- contract with Lily Group, 2023–2026, PI: Ph. Preux.
- contract with Saint-Gobain Research, 2023–2026, PI: Ph. Preux.

10 Partnerships and cooperations

Participants: Debabrota Basu, Rémy Degenne, Émilie Kaufmann, Odalric-Ambrym Maillard, Philippe Preux.

10.1 International initiatives

10.1.1 Inria associate team not involved in an IIL or an international program

DC4SCM

Title: Data Collection for Smart Crop Management

Duration: 2020 → 2024

Coordinator: Philippe Preux

Partners:

- Bihar Agriculture University, India,
- Inria FUN, Lille.

Inria contact: Philippe Preux

Summary: As part of our research activities related to the application of reinforcement learning and bandits to agriculture, this associate teams aim at providing us with in-field data, and also the ability to perform in-field experiments. This sort of experiments is extremely useful to train our algorithms which have to explore, that is test new actions in the field and observe their outcome. This approach is complementary to the one we investigate with the use of the DSSAT simulator.

RELIANT

Title: Real-life bandits

Duration: 2022 → 2024

Coordinator: Junya Honda (honda@i.kyoto-u.ac.jp)

Partners:

- Kyoto University Kyoto (Japon)

Inria contact: Odalric-Ambrym Maillard

Summary: The RELIANT project is about studying applicability to the real-world of sequential decision making from a reinforcement learning (RL) and multi-armed bandit (MAB) theory standpoint. Building on over a decade of leading expertise in advancing the field of MAB and RL theory, our two teams have also developed interactions with practitioners (e.g. in healthcare, personalized medicine or agriculture) in recent projects, in the quest to bring modern bandit theory to societal applications, for real. This quest for real-world reinforcement learning, rather than working in simulated and toyish environments is actually today's main grand-challenge of the field that hinders applications to the society and industry. While MABs are acknowledged to be the most applicable building block of RL, as experts interacting with practitioners from different fields we have identify a number of key bottlenecks on which joining our efforts is expected to significantly impact the applicability of MAB to the real-world. Those as related to the typically small samples size that arise in medical applications, the complicated type of rewards distributions that arise, e.g. in agriculture, the numerous constraints (such as fairness) that should be taken into account to speed up learning and make ethical decisions, and the possible non-stationary aspects of the tasks. We suggest to connect on the mathematical level our complementary expertise on multi-armed bandit (MAB), sequential hypothesis testing (SHT) and Markov decision processes (MDP) to address these challenges and significantly advance the design of the next generation of sequential decision making algorithms for real-life applications.

10.1.2 Participation in other International Programs

International collaborators:

- C. Dimitrakakis, Professor, Université de Neuchâtel, safe reinforcement learning, Bayesian reinforcement learning, Online learning in games.
- M. Alibeigi, Researcher, Zenseact AB of Volvo, safe reinforcement learning.

- B. Ghosh, Research Scientist, ASTAR Singapore, fairness in machine learning.
- K. Meel, Associate Professor, National University of Singapore and University of Toronto, Formal methods for machine learning.
- S. Bressan, Associate Professor, National University of Singapore, Machine learning for quantum physics.

10.2 International research visitors

10.2.1 Visits of international scientists

Esteban Clua

Status: Professor

Institution of origin: Institute of Computing at Fluminense Federal University, Niterói/RJ

Country: Brazil

Dates: June 28–July 7

Context of the visit: on-going collaboration

Mobility program/type of mobility: research stay

Eduardo Vasconcellos

Status: Post-doctoral fellow

Institution of origin: Institute of Computing at Fluminense Federal University, Niterói/RJ

Country: Brazil

Dates: Jul. 2022–Sep. 2023

Context of the visit: development of a digital twin for a sailing boat; control of the digital twin with reinforcement learning

Mobility program/type of mobility: research stay

André Araujo

Status: Ph.D. student

Institution of origin: Institute of Computing at Fluminense Federal University, Niterói/RJ

Country: Brazil

Dates: May – Dec.

Context of the visit: control of the digital twin with reinforcement learning

Mobility program/type of mobility: research stay

Wouter Koolen

Status: Professor

Institution of origin: CWI Amsterdam and University of Twente

Country: Netherlands

Dates: Oct. 8–Oct. 13

Context of the visit: on-going collaboration

Mobility program/type of mobility: research stay

Junya Honda

Status: Associate Professor

Institution of origin: University of Kyoto

Country: Japan

Dates: Nov. 25–Dec. 03

Context of the visit: on-going collaboration with RELIANT team

Mobility program/type of mobility: research stay

Junpei Komiyama

Status: Assistant Professor

Institution of origin: University of Kyoto

Country: Japan

Dates: Nov. 26–Dec. 06

Context of the visit: on-going collaboration with RELIANT team

Mobility program/type of mobility: research stay

Bishwamittra Ghosh

Status: Research scientist

Institution of origin: ASTAR research institute

Country: Singapore

Dates: Apr. 26–Apr. 30

Context of the visit: on-going collaboration on fairness in machine learning

Mobility program/type of mobility: research stay

10.2.2 Visits to international teams

Debabrota Basu

Institution: Indian Statistical Institute, Kolkata

Country: India

Dates: January 2023 and June 2023

Context of the visit: on-going collaboration with Applied Computing and Statistics divisions and finalising applications for joint Indo-French collaborative team

Institution: Bihar Agricultural University, Sabour

Country: India

Dates: July 2023

Context of the visit: on-going collaboration with DC4SCM team

Institution: National University of Singapore

Country: Singapore

Dates: August 2023

Context of the visit: invited talk and finalising application for French-Singaporean collaboration

Institution: CWI Amsterdam

Country: The Netherlands

Dates: November 2023

Context of the visit: on-going collaboration with CausalXRL team

10.3 European initiatives

10.3.1 Other european programs/initiatives

Title: CausalXRL

Duration: 2021 → 2024

Coordinator: Aditya Gilra, U. Amsterdam

Partners:

- U. Amsterdam
- U. Sheffield
- U. Vienna
- Inria Scool

Inria contact: Philippe Preux

Summary: Deep reinforcement learning systems are approaching or surpassing human-level performance in specific domains, from games to decision support to continuous control, albeit in non-critical environments. Most of these systems require random exploration and state-action-value-based exploitation of the environment. However, in important real-life domains, like medical decision support or patient rehabilitation, every decision or action must be fully justified and certainly not random. We propose to develop neural networks that learn causal models of the environment relating action to effect, initially using offline data. The models will then be interfaced with reinforcement learning and decision support networks, so that every action taken online can be explained or justified based on its expected effect. The causal model can then be refined iteratively, enabling to better predict future cascading effects of any action chain. The system, subsequently termed CausalXRL, will only propose actions that can be justified on the basis of beneficial effects. When the immediate benefit is uncertain, the system will propose explorative actions that generate most-probable future benefit. CausalXRL thus supports the user in choosing actions based on specific expected outcomes, rather than as prescribed by a black box.

10.4 National initiatives

10.4.1 ANR projects

Scool is involved in 4 ANR projects:

- ANR Bold, headed by V. Perchet (ENS Paris-Saclay, ENSAE), local head: E. Kaufmann, 2019–2023.
- ANR JCJC FATE, PI: R. Degenne, 2023–2026
- ANR JCJC REPUBLIC, PI: D. Basu, 2023–2026
- ANR BIP-UP, partnership: Scool/Inserm (CHU de Lille), PI: Ph. Preux, 2023–2026.

10.4.2 PEPR projects

Scool is involved in 2 PEPR:

- PEPR AI: project FOUNDRY, local head: E. Kaufmann (description below);
- PEPR « Agroécologie et numérique », Pl@ntAgroEco, local head: O.-A. Maillard.

Title: FOUNDRY

Duration: July 2024 → June 2028

Coordinator: Panayotis Mertikopoulous, Polaris, Univ. Grenoble Alpes

Partners:

- POLARIS: a joint research team between the CNRS, Inria, and Univ. Grenoble Alpes.
- ENS Lyon: faculty from the pure and applied mathematics department of ENS Lyon.
- Inria FAIRPLAY: a joint team between Criteo, IP Paris (ENSAE and Ecole Polytechnique), and Inria.
- LTCI: the informations and communications laboratory of Télécom Paris.
- MILES: the machine intellligence and learning systems of the LAMSADE lab at Paris Dauphine.
- Inria Scool

Inria contact: Emilie Kaufmann

Summary: From automated hospital admission systems powered by machine learning (ML), to flexible chatbots capable of fluent conversations and self-driving cars, the wildfire spread of artificial intelligence (AI) has brought to the forefront a crucial question with far-reaching ramifications for the society at large: Can ML systems and models be relied upon to provide trustworthy output in high-stakes, mission- critical environments? These questions invariably revolve around the notion of *robustness*, an operational desideratum that has eluded the field since its nascent stages. One of the main reasons for this is the fact that ML models and systems are typically data-hungry and highly sensitive to their training input, so they tend to be brittle, narrow-scoped, and unable to adapt to situations that go beyond their training envelope. On that account, the core vision of the proposed research is that robustness cannot be achieved by blindly throwing more data and computing power to larger and larger models with exponentially growing energy requirements (and a commensurate carbon footprint to boot). Instead, our proposal intends to rethink and develop the core theoretical and methodological FOUNDations of Robustness and reliabilitY (FOUNDRY) that are needed to build and instill trust in ML-powered technologies and systems from the ground up.

Title: Pl@ntAgroEco

Duration: July 2024 → June 2028

Coordinator: Alexis Joly, Inria Zenith, and Pierre Bonnet CIRAD, AMAP.

Partners:

- INRAE
- INRIA
- IRD
- CIRAD
- Tela Botanica
- Université de Montpellier

- Université Paris-Saclay

Inria contact: Odalric-Ambrym Maillard

Summary: Agroecology necessarily involves crop diversification, but also the early detection of diseases, deficiencies and stresses (hydric, etc.), as well as better management of biodiversity. The main stumbling block is that this paradigm shift in agricultural practices requires expert skills in botany, plant pathology and ecology that are not generally available to those working in the field, such as farmers or agri-food technicians. Digital technologies, and artificial intelligence in particular, can play a crucial role in overcoming this barrier to access to knowledge.

The aim of the Pl@ntAgroEco project will be to design, experiment with and develop new high-impact agro-ecology services within the Pl@ntNet platform. This includes :

- research in AI and plant sciences ;
- agile development of new components within the platform;
- organization of participatory science programs and animation of the Pl@ntNet user community.

Ce programme de travail a pour but de produire une amélioration de la détection et reconnaissance des maladies végétales, de l'identification des niveaux infraspécifiques. Il permettra le développement d'outils d'estimation de la sévérité des symptômes, carences, stades de déclin et stress hydrique ou de caractérisation des associations d'espèces à partir d'images multi-spécimens. Il améliorera la connaissance des espèces.

Le projet Pl@ntAgroEco rassemble des forces complémentaires en matière de recherche, de développement et d'animation. S'ajouteront à l'équipe pluridisciplinaire chargée de la plateforme Pl@ntNet de nouvelles forces de recherche ayant une expertise reconnue dans les sciences participatives. Le consortium rassemblera 10 partenaires incluant des organismes de recherche, des universités, des acteurs de la société civile et des partenaires internationaux

10.4.3 Inria challenge

Scool has been involved in the Challenge HY_AIAI. In this challenge, we collaborated with L. Gallaraga, CR Inria Rennes, about the combination of statistical and symbolic approaches in machine learning.

10.4.4 Other projects in France

Scool is involved in the Regalia pilot-project.

Other collaborations:

- A. Bellet, CR, Inria Lille-Nord Europe (Équipe Magnet).
- B. De-Saporta, Université de Montpellier, piecewise-deterministic Markov processes.
- A. Garivier, Professor, ENS Lyon, PhD co-supervisor of Aymen Al-Marjani.
- R. Gautron, Post-doctoral researcher, CGIAR, agricultural practices recommendation, gym-DSSAT software.
- B. Raffin, DR Inria Grenoble, continuous-time reinforcement learning.
- L. Soulier, Associate Professor, Sorbonne Université, reinforcement learning for information retrieval.
- L. Richert, R. Thiébaut, Inria SISTM, Bordeaux, bandits for vaccine clinical trials.
- A. Tirinzoni, Meta AI, instance-dependent sample complexity of reinforcement learning.

10.5 Regional initiatives

- "Bandits for Health" project (B4H) funded by I-Site Lille. 2020–2023. PI: Ph. Preux.

  Collaboration between Scool and Prof. F. Pattou Inserm Unit 1190/CHU de Lille. We investigate how the exploitation of data may be used to improve patient follow-up.

- O.-A. Maillard and Ph. Preux are supported by an AI chair. 3/5 of this chair is funded by the Metropole Européenne de Lille, the other 2/5 by the Université de Lille and Inria, through the AI Ph.D. ANR program. 2020–2024.

  This chair is dedicated to the advancement of research on reinforcement learning.

11 Dissemination

Participants: Riad Akrour, Debabrota Basu, Rémy Degenne, Emilie Kaufmann, Odalric-Ambrym Maillard, Timothée Mathieu, Philippe Preux.

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

General chair, scientific chair

- A. Shilova co-organized a workshop at NeurIPS 2023 on "Advancing Neural Network Training (WANT): Computational Efficiency, Scalability, and Resource Optimization".

11.1.2 Scientific events: selection

Member of the conference program committees

- D. Basu: member of the PC of AAAI, IJCAI, and AAMAS.
- E. Kaufmann: member of the Senior PC of the conference Algorithmic Learning Theory (ALT) and of the Conference On Learning Theory (COLT).
- O-A. Maillard: member of the Senior PC of the conference Algorithmic Learning Theory.
- Ph. Preux: member of the Senior PC of AAAI, member of the PC of IJCAI and ECML.

Reviewer

- R. Akrour: reviewer for NeurIPS and for the "Advancing Neural Network Training (WANT)" workshop at NeurIPS
- D. Basu: reviewer for ICML, NeurIPS, AI&Stats, ICLR, EWRL, FAccT
- R. Degenne: reviewer for COLT, ALT, NeurIPS
- E. Kaufmann: reviewer for the European Workshop on Reinforcement Learning (EWRL), COLT, ALT, NeurIPS (emergency reviewer)
- O-A. Maillard: reviewer for the conferences Artificial Intelligence and Statistics (AI&Stats, 10% best reviewer award), International Conference on Machine Learning (ICML, emergency reviewer)
- T. Mathieu: reviewer for the Conference On Learning Theory (COLT), the Conference on Uncertainty in Artificial Intelligence (UAI), the International Conference on Machine Learning, the European Conference on Machine Learning (ECML), Conference on Machine Learning Theory, Algorithmic Learning Theory

- Ph. Preux: reviewer for the European Workshop on Reinforcement Learning
- TQ. Tuan: Reviewer for the European Workshop on Reinforcement Learning, CORL, AI&Stats, IEEE T-RO

11.1.3 Journal

Member of the editorial boards

- O.-A. Maillard, editorial board of Journal of Machine Learning Research.

Reviewer - reviewing activities

- D. Basu: reviewer of JMLR, TMLR, IEEE Transactions on Dependable and Secure Computing.
- R. Degenne: reviewer for the Journal of Machine Learning Research.
- E. Kaufmann: reviewer for the Journal of Machine Learning Research (3 papers).
- O-A. Maillard: reviewer for the journals Statistics and Computing, the Annals of Statistics, Journal of Machine Learning Research.
- T. Mathieu: reviewer for the Journal of American Statistical Association, the Latin American Journal of Probability and Mathematical Statistics, Les Annales de l'Institut Henri Poincaré, the Journal of Machine Learning Research.

11.1.4 Invited talks

- D. Basu: invited speaker, ACMU seminars, Indian Statistical Institute (Kolkata, India).
- D. Basu: invited speaker, Journée Security at CRIStAL (Lille, France).
- D. Basu: invited speaker, JSPS Japan-Singapore Joint Seminars (National Institute of Informatics, Japan).
- D. Basu: invited speaker, Descartes seminars, CNRS@CREATE, National University of Singapore (Singapore).
- R. Degenne: invited talk, Probability and Statistics seminar, University of Freiburg (Freiburg, Germany).
- R. Degenne: invited speaker, Lean for the Curious Mathematician 2023 (Düsseldorf, Germany)
- R. Degenne: invited speaker, Workshop on Bandits and Statistical Tests (Potsdam, Germany).
- E. Kaufmann: invited talk at the Machine Learning Theory Bootcamp in CWI (Amsterdam).
- E. Kaufmann: invited speaker at the European Workshop on Reinforcement Learning (Brussels).
- E. Kaufmann: invited talk at the Statistics seminar of University Paris-Saclay (Orsay).
- E. Kaufmann: invited speaker at the Workshop on Bandits and Statistical Tests (Potsdam).
- O.-A. Maillard, invited speaker, PMSMA: Processus markoviens, semi-markoviens et leurs applications
- T. Mathieu, invited speaker, Conference on Statistical estimation.
- Ph. Preux, invited speaker, Machine Learning in Poland (ML in PL).
- Ph. Preux, keynote speaker, Inria-Chile Scientific days.

11.1.5 Tutorials

- D. Basu gave a tutorial on Auditing Bias of Machine Learning Algorithms: Tools and Overview at IJCAI, 2023.
- D. Basu gave a tutorial on Artificial Intelligence for Agricultural Sciences at Bihar Agricultural University, Sabour.
- É. Kaufmann gave a class on bandits at the Reinforcement Learning Summer Scool, Barcelona.
- O.-A. Maillard gave a class on reinforcement learning at Ecole Polytechnique International Executive Master (May) and Executive Master (October)
- Ph. Preux gave a class on reinforcement learning at the INSA-Rouen Summer Scool.

11.1.6 Scientific expertise

Ph. Preux is:

- a member of the IRD CSS 5 (data science and models),
- a member of the scientific committee of the PEPR « agro-écologie et numérique »,
- a member of the scientific committee on ethics of the Health Data warehouse of the CHU de Lille.

11.1.7 Research administration

Ph. Preux is the scientific head of the CornelIA CPER project.

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

- R. Akrour: « Option découverte: Machine Learning », L3 in Computer Science, Université de Lille
- R. Akrour: « Perception et motricité 2 », L2 MIASHS, Université de Lille
- R. Akrour: « Perception et motricité 1 », L1 MIASHS, Université de Lille
- D. Basu: Sequential Decision Making, M2 in Data Science, Centrale Lille and Université de Lille
- D. Basu: Research Reading Group, M2 in Data Science, Centrale Lille and Université de Lille
- D. Basu: Advanced Machine Learning and Decision Making, Centrale Lille
- R. Degenne: Sequential learning, M2 MVA, ENS Paris-Saclay
- R. Degenne: Sequential learning, Centrale Lille
- R. Degenne: Sciences des données 3, L3 MIASHS, Université de Lille
- E. Kaufmann: Sequential Decision Making (24h), M2 Data Science, Ecole Centrale Lille.
- O.-A. Maillard: Statistical Reinforcement Learning (48h), MAP/INF641, Master Artificial Intelligence and advanced Visual Computing, École Polytechnique.
- Ph. Preux: « Prise de décision séquentielle dans l'incertain », M2 in Computer Science, Université de Lille.
- Ph. Preux: « Apprentissage par renforcement », M2 in Computer Science, Université de Lille.
- Ph. Preux: « Science des données II », L3 MIASHS, Université de Lille.
- Ph. Preux: « IA et apprentissage automatique », DU IA & Santé, Université de Lille.

11.2.2 Supervision

- D. Basu and O.-A. Maillard supervised Thomas Michel's M2 internship.
- D. Basu supervised Priyank Agarwal's PhD internship.
- D. Basu supervised Udvas Das's masters thesis.
- E. Kaufmann and R.Degenne co-supervised the M2 internship of Adrienne Tuynman.
- O.-A. Maillard supervised Waris Radji's M2 internship.
- A. Shilova, Ph. Preux and B. Raffin (Inria Grenoble) co-supervised Thomas Delliaux's M2 internship.

11.2.3 Juries

D. Basu was a member of the Ph.D. defense committee for:

- Paolo Recchia, Université Côte d'Azur

E. Kaufmann was a member of the Ph.D. defense committees for:

- Chloé Rouyer, University of Copenhagen (reviewer), January
- Hannes Errikson, Chalmers University, September
- Kaito Ariu, KTH, November
- Aymen Al-Marjani, ENS Lyon (supervisor), December
- Otmane Sahki, Criteo/ENSAE (reviewer), December

E. Kaufmann was a reviewer of the HDR defense of Raphaël Feraud (Orange Labs, Université Paris-Saclay) who defended in Feburary.

O.-A. Maillard was a member of the Ph.D. defense committees for:

- Réda Ouhamma, Université de Lille (supervisor), April.
- Antoine Barrier, Ecole normale supérieure de Lyon (reviewer), July.
- Fabien Pesquerel, Université de Lille (supervisor), December.

Ph. Preux was a member of the Ph.D. defense committees for:

- Hélène Plisnier, VUB (reviewer), March.
- Nathan Grinsztajn, Université de Lille (supervisor), June.
- Hannes Eriksson, Chalmers University, September.
- Meyssa Zouambi, Université de Lille (chair), December.
- Étienne Ménager, Université de Lille (chair), December.

11.3 Popularization

11.3.1 Articles and contents

O.-A. Maillard was interviewed for an article for Inria Demain, un compagnon d'aide à la décision pour l'agriculture ?

11.3.2 Education

E. Kaufmann did 4 one hour sessions of "CHICHE: Un Scientifique, Une Classe" in the Lycée Kernanec, Marcq-en-Barœul.

Ph. Preux made a presentation on AI for a group of pupils at Collège M. Yourcenar, April, Marchiennes.

11.3.3 Interventions

E. Kaufmann gave a presentation at the "Rencontre des Jeunes Mathématiciennes et Informaticiennes" (RJMI) organized at Inria. A. Tuynman (PhD student in Scool) organized a problem session at RJMI for a group a high school female students.

O.-A. Maillard gave a presentation at table ronde "Decision making in a uncertain environment" organized for all master students of graduate programs of Université de Lille, December.

Ph. Preux:

- was part of the Merlin project of Université de Lille that led to "The big investigation on AI" TV program produced by « L'esprit Sorcier TV » channel.
- was part of the scientific committee of the « Forum des Sciences » in Villeneuve d'Ascq regarding the season on AI.
- gave a talk on AI to the ASAP, Université de Lille: "A short history of AI".
- gave a talk during the Robotik exhibition in Orchies: "AI: from myths to reality".
- participated to the TV program "We tell you more" on AI on the Wéo channel.
- made a presentation about AI during a meeting of the Comité d'Orientation Stratégique et Pédagogique (COSP) of the INSPE HdF regarding how AI impacts teaching and how future (primary, elementary, high) school teachers should be trained on AI.

12 Scientific production

12.1 Major publications

[1] L. Besson and E. Kaufmann. 'Multi-Player Bandits Revisited'. In: *Algorithmic Learning Theory*. Mehryar Mohri and Karthik Sridharan. Lanzarote, Spain, Apr. 2018. URL: https://hal.inria.fr/hal-01629733.

[2] G. Dulac-Arnold, L. Denoyer, P. Preux and P. Gallinari. 'Sequential approaches for learning datum-wise sparse representations'. In: *Machine Learning* 89.1-2 (1st Oct. 2012), pp. 87–122. DOI: 10.1007/s10994-012-5306-7. URL: https://hal.inria.fr/hal-00747724.

[3] Y. Flet-Berliac and P. Preux. 'Only Relevant Information Matters: Filtering Out Noisy Samples to Boost RL'. In: *IJCAI 2020 - International Joint Conference on Artificial Intelligence*. Yokohama, Japan, July 2020. DOI: 10.24963/ijcai.2020/376. URL: https://hal.inria.fr/hal-02091547.

[4] A. Garivier and E. Kaufmann. 'Optimal Best Arm Identification with Fixed Confidence'. In: *29th Annual Conference on Learning Theory (COLT)*. Vol. 49. JMLR Workshop and Conference Proceedings. New York, United States, June 2016. URL: https://hal.archives-ouvertes.fr/hal-01273838.

[5] B. Ghosh, D. Basu and K. S. Meel. 'Justicia: A Stochastic SAT Approach to Formally Verify Fairness'. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Conference on Artificial Intelligence. Vol. 35. Proceedings of the AAAI Conference on Artificial Intelligence 9. Virtual, Canada, Feb. 2021, pp. 7554–7563. URL: https://hal.science/hal-03445831.

[6] E. Kaufmann and W. M. Koolen. 'Monte-Carlo Tree Search by Best Arm Identification'. In: *NIPS 2017 - 31st Annual Conference on Neural Information Processing Systems*. Advances in Neural Information Processing Systems. Long Beach, United States, Dec. 2017, pp. 1–23. URL: https://hal.archives-ouvertes.fr/hal-01535907.

[7] O.-A. Maillard. 'Boundary Crossing Probabilities for General Exponential Families'. In: *Mathematical Methods of Statistics* 27 (2018). URL: https://hal.archives-ouvertes.fr/hal-01737150.

[8] O.-A. Maillard, H. Bourel and M. S. Talebi. 'Tightening Exploration in Upper Confidence Reinforcement Learning'. In: *International Conference on Machine Learning*. Vienna, Austria, July 2020. URL: https://hal.archives-ouvertes.fr/hal-03000664.

[9] O. Nicol, J. Mary and P. Preux. 'Improving offline evaluation of contextual bandit algorithms via bootstrapping techniques'. In: *International Conference on Machine Learning*. Ed. by E. Xing and T. Jebara. Vol. 32. Journal of Machine Learning Research, Workshop and Conference Proceedings; Proceedings of The 31st International Conference on Machine Learning. Beijing, China, June 2014. URL: https://hal.inria.fr/hal-00990840.

[10] F. Pesquerel and O.-A. Maillard. 'IMED-RL: Regret optimal learning of ergodic Markov decision processes'. In: NeurIPS 2022 - Thirty-sixth Conference on Neural Information Processing Systems. Thirty-sixth Conference on Neural Information Processing Systems. New-Orleans, United States, 28th Nov. 2022. URL: https://hal.science/hal-03825423.

12.2 Publications of the year

International journals

[11] A. P. D. Araújo, D. Daniel, R. Guerra, D. Brandão, E. C. Vasconcellos, A. Negreiros, E. Clua, L. Goncalves and P. Preux. 'General System Architecture and COTS Prototyping of an AIoT-Enabled Sailboat for Autonomous Aquatic Ecosystem Monitoring'. In: *IEEE Internet of Things Journal* (2023). DOI: 10.1109/JIOT.2023.3324525. URL: https://hal.science/hal-04355027.

[12] R. Caiazzo, P. Bauvin, C. Marciniak, P. Saux, G. Jacqmin, R. Arnoux, S. Benchetrit, J. Dargent, J.-M. Chevallier, V. Frering, J. Gugenheim, D. Lechaux, S. Msika, A. Sterkers, P. Topart, G. Baud and F. Pattou. 'Impact of Robotic Assistance on Complications in Bariatric Surgery at Expert Laparoscopic Surgery Centers: A Retrospective Comparative Study With Propensity Score'. In: *Annals of Surgery* 278.4 (7th Sept. 2023), pp. 489–496. DOI: 10.1097/SLA.0000000000005969. URL: https://hal.science/hal-04198805.

[13] C. Rozwag, F. Valentini, A. Cotten, X. Demondion, P. Preux and T. Jacques. 'Elbow trauma in children: development and evaluation of radiological artificial intelligence models'. In: *Research in Diagnostic and Interventional Imaging* 6 (29th Apr. 2023). DOI: 10.1016/j.redii.2023.100029. URL: https://hal.science/hal-04244410.

[14] P. Saux, P. Bauvin, V. Raverdy, J. Teigny, H. Verkindt, T. Soumphonphakdy, M. Debert, A. Jacobs, D. Jacobs, V. Monpellier, P. C. Lee, C. H. Lim, J. C. Andersson-Assarsson, L. Carlsson, P.-A. Svensson, F. Galtier, G. Dezfoulian, M. Moldovanu, S. Andrieux, J. Couster, M. Lepage, E. Lembo, O. Verrastro, M. Robert, P. Salminen, G. Mingrone, R. Peterli, R. V. Cohen, C. Zerrweck, D. Nocca, C. W. Le Roux, R. Caiazzo, P. Preux and F. Pattou. 'Development and validation of an interpretable machine learning-based calculator for predicting 5-year weight trajectories after bariatric surgery: a multinational retrospective cohort SOPHIA study'. In: *The Lancet Digital Health* (29th Aug. 2023). DOI: 10.1016/S2589-7500(23)00135-8. URL: https://hal.science/hal-04192198.

International peer-reviewed conferences

[15] A. Azize and D. Basu. 'Interactive and Concentrated Differential Privacy for Bandits'. In: EWRL 2023 – European Workshop on Reinforcement Learning. Brussels (Belgium), Belgium, Sept. 2023. URL: https://hal.science/hal-04215685.

[16] A. Azize, M. Jourdan, A. A. Marjani and D. Basu. 'On the Complexity of Differentially Private Best-Arm Identification with Fixed Confidence'. In: NeurIPS 2023 – Conference on Neural Information Processing Systems. New Orleans (US), United States, Dec. 2023. URL: https://hal.science/hal-04215474.

[17] D. Baudry, F. Pesquerel, R. Degenne and O.-A. Maillard. 'Fast Asymptotically Optimal Algorithms for Non-Parametric Stochastic Bandits'. In: *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*. NeurIPS 2023 - Thirty-seventh Conference on Neural Information Processing Systems. New Orleans (Louisiana), United States, 10th Dec. 2023. URL: https://inria.hal.science/hal-04337742.

[18] E. Carlsson, D. Basu, F. D. Johansson and D. Dubhashi. 'Pure Exploration in Bandits with Linear Constraints'. In: EWRL 2023 – European Workshop on Reinforcement Learning. Brussels, Belgium, Sept. 2023. URL: https://hal.science/hal-04203235.

[19] M. Centa and P. Preux. 'Soft Action Priors: Towards Robust Policy Transfer'. In: AAAI 2023 - Thirty-Seventh AAAI Conference on Artificial Intelligence. Washington DC, United States, 7th Feb. 2023. URL: https://inria.hal.science/hal-03976459.

[20] S. R. Chowdhury, P. Saux, O.-A. Maillard and A. Gopalan. 'Bregman Deviations of Generic Exponential Families'. In: Conference On Learning Theory (COLT). Bangalore, India, 12th July 2023. URL: https://hal.science/hal-04161043.

[21] E. Cyffers, A. Bellet and D. Basu. 'From Noisy Fixed-Point Iterations to Private ADMM for Centralized and Federated Learning'. In: Proceedings of the 40th International Conference on Machine Learning (ICML). Honolulu, United States, July 2023. URL: https://hal.science/hal-04260417.

[22] R. Degenne. 'On the Existence of a Complexity in Fixed Budget Bandit Identification'. In: *Proceedings of Machine Learning Research*. Thirty Sixth Conference on Learning Theory. Vol. 195. Bengaluru (Bangalore), India, 30th June 2023. URL: https://inria.hal.science/hal-04337726.

[23] B. Ghosh, D. Basu and K. Meel. '"How Biased are Your Features?": Computing Fairness Influence Functions with Global Sensitivity Analysis'. In: FAccT '23: the 2023 ACM Conference on Fairness, Accountability, and Transparency. Chicago IL, United States: ACM, 12th June 2023, pp. 138–148. DOI: 10.1145/3593013.3593983. URL: https://hal.science/hal-03770346.

[24] M. Jourdan and R. Degenne. 'Non-Asymptotic Analysis of a UCB-based Top Two Algorithm'. In: *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*. Thirty-seventh Conference on Neural Information Processing Systems. New Orleans (Louisiana), United States, 10th Dec. 2023. URL: https://inria.hal.science/hal-03830958.

[25] M. Jourdan, R. Degenne and E. Kaufmann. 'An ε-Best-Arm Identification Algorithm for Fixed-Confidence and Beyond'. In: Advances in Neural Information Processing Systems (NeurIPS). New Orleans, United States, 10th Dec. 2023. URL: https://hal.science/hal-04306214.

[26] M. Jourdan, R. Degenne and E. Kaufmann. 'Dealing with Unknown Variances in Best-Arm Identification'. In: *Proceedings of Machine Learning Research (PMLR)*. Algorithmic Learning Theory (ALT). Singapore (SG), Singapore, 20th Feb. 2023. URL: https://hal.science/hal-04306221.

[27] P. Karmakar and D. Basu. 'Marich: A Query-efficient Distributionally Equivalent Model Extraction Attack using Public Data'. In: Advances in Neural Information Processing Systems (NeurIPS). New orleans, USA, United States, Dec. 2023. URL: https://hal.science/hal-04260442.

[28] C. Kone, E. Kaufmann and L. Richert. 'Adaptive Algorithms for Relaxed Pareto Set Identification'. In: NeurIPS 2023 - 37th Conference on Neural Information Processing Systems. La Nouvelle Orléans, LA, United States, 10th Dec. 2023. URL: https://hal.science/hal-04306210.

[29] A. Al-Marjani, A. Tirinzoni and E. Kaufmann. 'Active Coverage for PAC Reinforcement Learning'. In: *Proceedings of Machine Learning Research (PMLR)*. Conference on Learning Theory 2023. Bangalore, India, 2023. URL: https://hal.science/hal-04215441.

[30] R. Ouhamma, D. Basu and O.-A. Maillard. 'Bilinear Exponential Family of MDPs: Frequentist Regret Bound with Tractable Exploration & Planning'. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 37. Proceedings of the AAAI Conference on Artificial Intelligence, Volume 3 8. Washignton DC, United States, June 2023, pp. 9336–9344. DOI: 10.1609/aaai.v37i8.26119. URL: https://hal.science/hal-03790997.

[31] P. Saux and O.-A. Maillard. 'Risk-aware linear bandits with convex loss'. In: International Conference on Artificial Intelligence and Statistics (AISTATS). Vol. PMLR. 206. Valencia, Spain, 25th Apr. 2023. URL: https://hal.science/hal-04044440.

[32] A. Tirinzoni, A. Al-Marjani and E. Kaufmann. 'Optimistic PAC Reinforcement Learning: the Instance-Dependent View'. In: *Proceedings of Machine Learning Research (PMLR)*. Algorithmic Learning Theory (ALT). Singapore (SG), Singapore, 20th Feb. 2023. URL: https://hal.science/hal-04306228.

[33] E. C. Vasconcellos, R. M. Sampaio, A. P. D. Araújo, E. W. Gonzales Clua, P. Preux, R. Guerra, L. M. G. Gonçalves, L. Martí, H. Lira and N. Sanchez-Pi. 'Reinforcement-learning robotic sailboats: simulator and preliminary results'. In: NeurIPS 2023 Workshop on Robot Learning Workshop: Pretraining, Fine-Tuning, and Generalization with Large Scale Models. New Orelans, United States, 11th Dec. 2023. URL: https://inria.hal.science/hal-04395990.

[34] K. Ying and R. Degenne. 'A Formalization of Doob's Martingale Convergence Theorems in mathlib'. In: *Proceedings of the 12th ACM SIGPLAN International Conference on Certified Programs and Proofs*. 12th ACM SIGPLAN International Conference on Certified Programs and Proofs. Boston (Massachusetts), United States, 11th Jan. 2023. DOI: 10.1145/3573105.3575675. URL: https://inria.hal.science/hal-04337785.

Conferences without proceedings

[35] R. Gautron, E. J. Padrón, P. Preux, J. Bigot, O.-A. Maillard, G. Hoogenboom and J. Teigny. 'Learning crop management by reinforcement: gym-DSSAT'. In: AIAFS 2023 - 2nd AAAI Workshop on AI for Agriculture and Food Systems. Washignton DC, United States, 13th Feb. 2023. URL: https://inria.hal.science/hal-03976393.

[36] O.-A. Maillard, T. Mathieu and D. Basu. 'Farm-gym: A modular reinforcement learning platform for stochastic agronomic games'. In: AIAFS 2023 - Artificial Intelligence for Agriculture and Food Systems. Wahington DC, United States, 14th Feb. 2023. URL: https://inria.hal.science/hal-03960683.

Doctoral dissertations and habilitation theses

[37] N. Grinsztajn. 'Reinforcement learning for combinatorial optimization : leveraging uncertainty, structure and priors'. Université de Lille, 15th June 2023. URL: https://theses.hal.science/tel-04353766.

[38] R. Ouhamma. 'Toward realistic reinforcement learning'. Université de Lille, 14th Apr. 2023. URL: https://theses.hal.science/tel-04324714.

[39] F. Pesquerel. 'Information per unit of interaction in stochastic sequential decision making'. Université de Lille, 4th Dec. 2023. URL: https://hal.science/tel-04501905.

Reports & preprints

[40] S. Agrawal, T. Mathieu, D. Basu and O.-A. Maillard. *CRIMED: Lower and Upper Bounds on Regret for Bandits with Unbounded Stochastic Corruption*. 28th Sept. 2023. URL: https://hal.science/hal-04260464.

[41] R. Della Vecchia and D. Basu. *Online Instrumental Variable Regression: Regret Analysis and Bandit Feedback*. 20th Feb. 2023. URL: https://hal.science/hal-03831210.

[42] H. Eriksson, D. Basu, T. Tram, M. Alibeigi and C. Dimitrakakis. *Reinforcement Learning in the Wild with Maximum Likelihood-based Model Transfer*. 18th Feb. 2023. URL: https://hal.science/hal-04260795.

[43] H. Kohler, R. Akrour and P. Preux. *Optimal Interpretability-Performance Trade-off of Classification Trees with Black-Box Reinforcement Learning*. RR-9503. Inria Lille Nord Europe - Laboratoire CRIStAL - Université de Lille, Apr. 2023. URL: https://hal.science/hal-04060986.

[44] A. Al-Marjani, A. Tirinzoni and E. Kaufmann. *Towards Instance-Optimality in Online PAC Reinforcement Learning*. 23rd Oct. 2023. URL: https://hal.science/hal-04270888.

[45] T. Mathieu, R. Della Vecchia, A. Shilova, M. Centa de Medeiros, H. Kohler, O.-A. Maillard and P. Preux. *AdaStop: sequential testing for efficient and reliable comparisons of Deep RL Agents*. RR-9513. Inria Lille Nord Europe - Laboratoire CRIStAL - Université de Lille, June 2023. URL: https://inria.hal.science/hal-04132861.

[46] A. Shilova, T. Delliaux, P. Preux and B. Raffin. *Learning HJB Viscosity Solutions with PINNs for Continuous-Time Reinforcement Learning*. RR-9541. Inria Lille - Nord Europe, CRIStAL - Centre de Recherche en Informatique, Signal et Automatique de Lille - UMR 9189; Univ. Lille, CNRS, Centrale Lille, Inria UMR 9189 - CRIStAL,INRIA Lille Nord Europe, Villeneuve d'Ascq, France; Univ. Grenoble Alps, CNRS, Inria, Grenoble INP, LIG, 38000 Grenoble, France, 7th Feb. 2024, pp. 1–30. URL: https://inria.hal.science/hal-04445160.

12.3 Cited publications

[47] M. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 1994.

[48] B. Recht. 'A Tour of Reinforcement Learning: The View from Continuous Control'. arxiv preprint 1806.09460. 2018.

[49] R. Sutton and A. Barto. *Reinforcement Learning: an Introduction*. 2nd ed. http://incompleteideas.net/book/the-book-2nd.html. MIT Press, 2018.

[50] C. Szepesvári and T. Lattimore. *Bandit Algorithms*. Cambridge University press, 2019.