– Master Project Proposal–

REINFORCEMENT LEARNING WITH PREDICTIVE STATE REPRESENTATIONS

O-A. MAILLARD
SequeL, Inria Lille – Nord de France
E-MAIL: odalric.maillard@inria.fr

Keywords: Reinforcement Learning, Predictive State Representations, Hankel matrices, Spectral methods.

Introduction Reinforcement Learning is a field of research that models the problem of a learning agent interacting with a partially known dynamical system, that plays action based on past observations and receives reward that the agent tries to maximize. Markov Decision Process is a standard way of modeling a reinforcement learning problem when the observations are states (sufficient statistics of the past observations). However, in many situations, states are not directly observables and should be reconstructed from empirical data. Combining this approach with the reinforcement learning objective is a notoriously hard challenge.

A promising model to handle the notion of states are the Predictive State Representations (PSR), based on the fact that the the bi-infinite matrix that associates to each prefix and suffix sequences of observations the probability of observing their concatenation, called the Hankel matrix, is generally of finite rank equal related to the number of states of the system. Spectral methods can then be applied in order to build a representation space approximately capturing the states. Recent progress in the literature shows that it is possible to estimate the Hankel matrix of the system from a single trajectory of observations, with provable finite-time learning guarantees. It is then natural to ask whether we can do reinforcement learning with a single trajectory of observations in the framework of PSR, with theoretical learning guarantees.

Answering these questions may trigger significant progress in the theoretical and practical understanding of reinforcement learning, and lead to a major progress in the field.

Goal In this project, we want to better understand how to combine the estimation of the Hankel matrix from a single trajectory of observations of finite length with the reinforcement learning goal to minimize the regret accumulated over time.

Main tasks The goal of this Master project is first to provide a bibliographic overview of the notion of Predictive State Representation, Hankel matrices and undiscounted reinforcement learning. The next step will be to investigate the proposed combination from a theoretical perspective. These ways should be theoretically-grounded as much as possible and supported with key numerical experiments.

Other information The student should be mathematically strong and interested in solving theoretical problems using probability, statistics and optimization. A prior knowledge of the MDP literature would be a great addition. While the main goal of this internship is to solve a theoretical problem, the student should be able to run some simple numerical experiments to assess the practical performance of the algorithms (in the programming language of his/her choice).

SequeL is an Inria research team based in Lille and specialized in all aspects of sequential decision making, with a rich scientific activity. This research internship proposal is part of a national research project funded by the ANR that focuses on handling non-stationarity and structure in multi-armed bandits.

References

[1] Balle, B., Maillard, O. A. Spectral Learning from a Single Trajectory under Finite-State Policies. In *International conference on machine learning*, 2017.

[2] Azizzadenesheli, Kamyar, Alessandro Lazaric, and Animashree Anandkumar. Reinforcement learning of POMDPs using spectral methods. In *Conference on Learning Theory*, 2016.