Reinforcement Learning and Markovian Bandits

Bruno Gaujal
Université Grenoble Alpes,
Inria, CNRS, Grenoble INP, LIG, 38000 Grenoble
bruno.gaujal@inria.fr

Tutoriel proposé par l'axe « Décision, Modélisation, Evaluation, Incertitude » (DMEI) et l'action transverse « Données, Apprentissage Automatique et Optimisation » (DAAO) du GDR RO

Résumé :

In this talk, I will first present the main principles of model based reinforcement learning in Markov Decision Processes. I will explain the different frameworks where learning can be done and give the definition of the most popular performance measure, namely the regret. I will also present the ideas and techniques supporting the construction of no-regret algorithms for general MDPs, mainly optimism in the face of uncertainty and concentration inequalities for differences of Martingales. Several learning algorithms (UCRL2 and UCBVI), based on these principles and with near optimal regret, will be discussed.

The second part of the talk will present Markovian multi-armed bandits and the construction of an optimal policy using Gittins index. I will then show how to leverage on the optimality of Gittins index policies to design adapted reinforcement learning algorithms whose regret and time complexity scale (sub)linearly with the number of arms.

This talk is based on a joint work with Nicolas Gast and Kimang Khun.