CentraleSupélec

Master Project Proposal
2017-2018

OPTIMAL EXPLORATION IN MULTI-ARMED BANDITS

RICHARD COMBES
E-MAIL: richard.combes@centralesupelec.fr
Centrale-Supelec - L2S

The problem of optimal exploration in structured stochastic multi-armed bandits is considered. The goal of this project is to study the finite-time behaviour of some algorithms which have rencently been proven to be asymptotically optimal, as well as design new algorithms with improved finite-time guarantees.

Keywords: Stochastic Multi-Armed Bandits, Structured Bandits, Learning, Optimization.

1 General presentation of the topic

The stochastic multi-armed bandit (MAB) is a basic and important problem in Machine Learning where a learner samples K populations (called 'arms') adaptively in order to identify the one with the largest expectation. The MAB problem is the most elementary illustration of the trade-off between exploration and exploitation which appears in most learning problems. The performance of an algorithm is quantified by its regret, which is the difference between the cumulated reward received by an oracle and the algorithm in question.

The MAB problem is surprisingly rich, and while it has been investigated for more than half a century by statisticians and more recently by computer scientists, it is still a very active reseach field, and some fundamental questions are still open. One of the main questions is how to exploit structure in order to reduce regret and learn as fast as theoretically possible.

2 Instructions

The goal of this internship will be to study several families of provably asymptotically optimal, completely generic bandit algorithms, and a family of optimization problems which characterize the regret of these algorithms. The goal is to understand the structure of these optimization problems and quantify their regularity. Achiving this goal would yield new, state of the art finite time regret upper bounds and most likely lead to the discovery of new algorithms with improved finite time behaviour.

3 Expected ability of the student

The student should be mathematically strong and interested in solving theoretical problems using sophisticated probabilistic tools. A prior knowledge of the MAB litterature 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).

4 Administrative details

The internship will take place in Centrale-Supelec in Saclay, Ile-de-France. The student will be advised by R. Combes (Centrale-Supelec).

- About Centrale-Supelec: http://www.centralesupelec.fr/wordpress/?lang=en
- About L2S: http://www.l2s.centralesupelec.fr/en/content/presentation
- About Richard Combes: http://rcombes.supelec.free.fr/

5 References

- R. Combes, S. Magureanu and A. Proutiere, Minimal Exploration in Structured Stochastic Bandits, NIPS 2017
- T. Lattimore and C. Szepesvari. The end of optimism? an asymptotic analysis of finite-armed linear bandits. AISTATS, 2016.
- S. Bubeck and N. Cesa Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. Foundations and Trends in Machine Learning, 2012
- T. L. Graves and T. L. Lai. Asymptotically efficient adaptive choice of control laws in controlled markov chains. SIAM J. Control and Optimization, , 1997