COMPUTATIONAL COMPLEXITY IN MULTI-ARMED BANDITS

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

The goal of this project is to solve the important question of computational complexity in multi-armed bandits, and identify the trade-off between computational complexity and asymptotic regret, which is important both in theory and for the practical use of bandit algorithms in applications.

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.

In several MAB problems with a paticular structure, it occurs that known algorithms with low regret usually have high computational complexity, so that for large problems, one has to rely on weaker algorithms which can run in reasonable time. A natural question is therefore: what is the inherent computational complexity of multi-armed bandits and what are the bandit problems that can be solved efficiently ?

2 Instructions

The goal of this internship is to study several structured MAB problems (e.g. linear bandits, combinatorial bandits) and to characterize when it is possible to achieve both low regret and low computational complexity (say polynomial in the problem size) using a single algorithm. A side objective would be to understand the structure of certain optimization problems which characterize the lowest regret achievable by any algorithms (algorithms matching this bound are usually called asymptotically optimal algorithms).

3 Expected ability of the student

The student should be mathematically strong and interested in solving theoretical problems using probability, statistics and optimization. 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

- R. Combes, S. Talebi, A. Proutiere and M. Lelarge, Combinatorial Bandits Revisited, NIPS 2015

- 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