



SPRING 2020

RENSSELAER POLYTECHNIC INSTITUTE

DEPARTMENT OF MATHEMATICAL SCIENCES COLLOQUIUM

Optimal Learning for Structured Bandits

Abstract:

"We study structured multi-armed bandits, which is the problem of online decision-making under uncertainty in the presence of structural information. In this problem, the decision-maker needs to discover the best course of action despite observing only uncertain rewards over time. The decision-maker is aware of certain structural information regarding the reward distributions and would like to minimize his regret by exploiting this information, where the regret is its performance difference against a benchmark policy which knows the best action ahead of time. In the absence of structural information, the classical UCB and Thomson sampling algorithms are well known to suffer only minimal regret. Neither algorithm is, however, capable of exploiting structural information which is commonly available in practice. We propose a novel learning algorithm which we call "DUSA" whose worst-case regret matches the information-theoretic regret lower bound up to a constant factor and can handle a wide range of structural information. Our algorithm DUSA solves a dual/robust counterpart of regret lower bound at the empirical reward distribution and follows the suggestion made by the dual problem. Our proposed algorithm is the first computationally viable learning policy for structured bandit problems that suffers asymptotic minimal regret.

Bart Van Parys (MIT)

Monday, February 10, 2020

Amos Eaton 216 @ 4-5pm

Refreshments served 3:30-4pm Amos Eaton 4th Floor Lounge