BANDITS WITHOUT REGRETS: THE POWER OF ADAPTIVE ADVERSARIES

Nicolò Cesa-Bianchi

Dipartimento di Informatica
Università degli Studi di Milano, Italy
nicolo.cesa-bianchi@unimi.it

ABSTRACT

Prediction with expert advice is an abstract, game-theoretic model of sequential decision-making in which an adversary repeatedly assigns losses to actions that are selected by a randomized decision-maker (or player). The type of feedback received by the player in each step depends on the specific setting in which the game is played. In the expert setting, the player observes the loss currently associated to every action; in the bandit setting, instead, only the loss of the selected action is revealed to the player. Online services, such as recommendation systems, are naturally studied as bandit problems because of the conflicting needs of focusing on users' interests and exploring new options. In both bandit and expert settings, the player's performance is evaluated in terms of regret, measuring how much better the player could have performed if the best action had been chosen at each step. In this talk, we start by introducing two basic player algorithms, HEDGE and EXP3, along with their regret analysis. Then, we introduce a new notion of regret, also known as policy regret, which better captures the adversary's adaptiveness to the player's behavior. In a setting where losses are allowed to drift, we characterize the power of adaptive adversaries with bounded memories and switching costs. In particular, we show that with switching costs, the attainable rate with bandit feedback is $\widetilde{\Theta}(T^{2/3})$. Interestingly, this rate is significantly worse than the $\Theta(\sqrt{T})$ rate attainable with switching costs in the expert case. Via a novel reduction from experts to bandits, we also show that a bounded memory adversary can force $\widetilde{\Theta}(T^{2/3})$ regret even in the expert case, proving that switching costs are easier to control than bounded memory adversaries. Our lower bounds rely on a new stochastic adversary strategy that generates loss processes with strong dependencies.

Part of this talk is based on [1] co-authored by Ofer Dekel (Microsoft Research, USA) and Ohad Shamir (Weizmann Institute, Israel)

1. REFERENCES

[1] N. Cesa-Bianchi, O. Dekel, and O. Shamir, "Online learning with switching costs and other adaptive adversaries," in *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, 2013, To appear.