

Completing the classification of adversarial partial monitoring games

Csaba Szepesvári *

szepesva@ualberta.ca

Partial monitoring is a generalization of bandit problems where the relationship between the feedback and the loss is loosened. As such, partial monitoring offers a rich framework to study the exploration-exploitation dilemma.

In the finite, adversarial version of the problem that we consider here a learner and an adversary take actions in a sequential fashion from their respective finite action sets. A pair of actions result in a fixed loss and a fixed observation; the mapping that maps pairs of actions to the losses and observations is given to the learner ahead of time. The unknown are the actions taken by the adversary. Rather, the learner gains information of the adversaries actions by receiving the observation underlying the joint action of a round. Since what observations can be received in a round is governed by the action taken, the learner has control over what it can receive information about. The learner's goal, as usual, is to keep its total regret, the difference between the total loss suffered and the loss that could have been suffered had the had full knowledge of the adversaries actions and played the best response to these, as small possible.

The classification problem of partial monitoring games is concerned with determining how to play in a given game described by the map from joint actions to losses and observations so as to achieve at most a constant multiple of the minimax regret. The study of partial monitoring games started with the work Rusitchini in 1999, and significant advances were made by Piccolboni and Schindelhauer, Cesa-Bianchi, Lugosi and and Stoltz. Around 2010, the goal of the full classification emerged and an almost complete answer was given by the joint work of Bartok, Foster, Pal, Rakhlin and Szepesvari (2014).

This work gave an almost complete characterization of all possible partial monitoring and identified four regimes: trivial, easy, hard and impossible games. The characterization, however, left out a set of games, namely those when there is at least one actions that are

*Department of Computer Science, University of Alberta, Edmonton, AB T6G 2E8, CANADA

optimal under some play of the adversary, but the region where this happens has a dimension defect. As a result, even though these “tricky” actions are nondominated, they could be omitted if not for the information they bring in. This creates a tricky situation when designing algorithms and until now deciding whether the presence of these tricky actions can move a game from the easy to the hard category remained open. The answer was known to be “no” for stochastic games when the adversary plays a fixed stochastic strategy. In this talk, I present the solution to this problem: It turns out that a simplified and improved version of the algorithm by Foster and Rakhlin can achieve $O(T^{1/2})$ regret for these problems, thus resolving the open problem and completing the characterization of finite adversarial partial monitoring games. In the talk, I will cover the key ideas of the characterization in previous works, then motivate and explain the new ideas that led to the new algorithm and explain the proof of the new result.

The talk is based on joint work with Tor Lattimore.