Tim van Erven

An Improved Algorithm for Adversarial Linear Contextual Bandits via Reduction

I will present an efficient algorithm for linear contextual bandits with adversarial losses and stochastic action sets. Our approach reduces this setting to misspecification-robust adversarial linear bandits with fixed action sets. Without knowledge of the context distribution or access to a context simulator, the algorithm achieves tilde{O}(d^2 sqrt{T}) regret and runs in poly(d,C,T) time, where d is the feature dimension, C is the number of linear constraints defining the action set in each round, and T is number of rounds. This resolves the open question by Liu et al. (2023) on whether one can obtain poly(d)sqrt{T} regret in polynomial time independent of the number of actions. For the important class of combinatorial bandits with adversarial losses and stochastic action sets, our algorithm is the first to achieve poly(d)\sqrt{T} regret in polynomial time, while no prior algorithm achieves even o(T) regret in polynomial time to our knowledge. When a simulator is available, the regret bound can be improved to tilde{O}(d\sqrt{L*}), where L* is the cumulative loss of the best policy.

This is joint work with Jack Mayo, Julia Olkhovskaya and Chen-Yu Wei.