Proceedings of Machine Learning Research vol 178:1–1, 2022 35th Annual Conference on Learning Theory

Multi-Agent Learning for Iterative Dominance Elimination: Formal Barriers and New Algorithms

Jibang Wu JW7JB@VIRGINIA.EDU
University of Virginia

Haifeng Xu HX4AD@VIRGINIA.EDU
University of Chicago[*]

Fan Yao FY4BC@VIRGINIA.EDU
University of Virginia

Editors: Po-Ling Loh and Maxim Raginsky

Abstract

Dominated actions are natural (and perhaps the simplest possible) multi-agent generalizations of *sub-optimal* actions as in standard single-agent decision making. Thus similar to standard bandit learning, a basic learning question in multi-agent systems is whether agents can learn to efficiently eliminate *all* iteratively dominated actions in an unknown game if they can only observe noisy bandit feedback about the payoff of their played actions. Surprisingly, despite a seemingly simple task, we show a quite negative result; that is, standard no regret algorithms — including the entire family of Dual Averaging algorithms — provably take *exponentially* many rounds to eliminate all iteratively dominated actions. Moreover, algorithms with the stronger no swap regret also suffer similar exponential inefficiency. To overcome these barriers, we develop a new algorithm that adjusts `Exp3` with `D`iminishing `H`istorical rewards (termed `Exp3-DH`); `Exp3-DH` gradually "forgets" history at carefully tailored rates. We prove that when all agents run `Exp3-DH` (a.k.a., *self-play* in multi-agent learning), all iteratively dominated actions can be eliminated within polynomially many rounds. Our experimental results further demonstrate the efficiency of `Exp3-DH`, and that state-of-the-art bandit algorithms, even those developed specifically for learning in games, fail to eliminate all iteratively dominated actions efficiently.[1]

Keywords: Iterative Dominance Elimination, Dual Averaging, Diminishing Historical Rewards

Acknowledgments

Haifeng Xu is supported by an NSF grant CCF-2132506.

[*] This work was done while the author was at University of Virginia.

1. Extended abstract. Full version appears at [2111.05486, v2]