Tuomas Sandholm
(Carnegie Mellon University, Strategy Robot, Optimized Markets, Strategic Machine)

What if we don't know the game? Finding and Certifying (Near-)Optimal Strategies in Black-Box Extensive-Form Imperfect-Information Games

---

In many settings -- for example war games, strategy video games, and financial simulations -- the game is given to us only as a black-box simulator in which we can play it, rather than via explicitly declared rules. While there have been impressive demonstrations in such settings using deep reinforcement learning (e.g., in StarCraft II and DOTA 2), prior techniques have not offered bounds on the game-theoretic exploitability of the computed strategies. In a NeurIPS-20 paper, we introduce an approach that shows that it is possible to provide such bounds without exploring the entire game. We introduce a notion of a certificate of an extensive-form (approximate) Nash equilibrium. For verifying a certificate, we give an algorithm that runs in time linear in the size of the certificate rather than the size of the whole game. In zero-sum games, we further show that an optimal certificate -- given the exploration so far -- can be computed with any game-solving algorithm (e.g., LP, CFR, or EGT). However, unlike in the cases of normal form or perfect information, we show that certain families of extensive-form games do not have small approximate certificates, even after making extremely nice assumptions on the structure of the game. Despite this difficulty, we find experimentally that very small certificates, even exact ones, often exist in large and even in infinite games. Overall, our approach enables one to try one's favorite exploration strategies while offering exploitability guarantees, thereby decoupling the exploration strategy from the equilibrium-finding process. Our first cut at this guaranteed black-box approach assumed that the black box could sample or expand arbitrary nodes of the game tree at any time, and that a series of exact game solves can be conducted to compute the certificate. In a AAAI-21 paper, we relax both of those assumptions. We show that high-probability certificates can be obtained with a black box that can do nothing more than play through games, using only a regret minimizer as a subroutine. As a bonus, we obtain an equilibrium-finding algorithm with $\tilde{O}(1/\sqrt{T})$ convergence rate in the extensive-form game setting that does not rely on a sampling strategy with lower-bounded reach probabilities (which MCCFR assumes). We demonstrate experimentally that our methods provide strong exploitability guarantees while exploring only a small portion of the game.

This talk is joint work with my PhD student Brian Hu Zhang. It covers results from our NeurIPS-20 paper, our AAAI-21 paper, and our newer results on this topic.