Review-32

Deep Reinforcement Learning from Self-Play in Imperfect-Information Games

Computation of Nash Equilibrium in Imperfect Information Games is a challenging task which often requires domain knowledge. To this end, the work presents Neural Fictitious Self Play (NFSP) which is a scalable method for learning approximate Nash Equilibria. NFSP combines fictitious play using supervised learning with Deep Reinforcement Learning (DRL). To combine these strategies, NFSP leverages reservoir sampling and anticipatory dynamics which enable scalable learning. NFSP demonstrates improved performance on Leduce poker when compared to prior methods. Furthermore, NFSP depicts performance comparable to state-of-the-art superhuman algorithms on Texas Hold'em Poker.

Prior work in imperfect information games has focussed on finding Nash Equilibria based on domain specific knowledge which is not suitable with the increasing degree of task variability. This is addressed using Full-Width Extensive-Form Fictitious Self Play (XFP) which allows Fictitious players to update their strategies in Extensive form. The work builds on this insight and combines fictitious self play with deep reinforcement learning using neural networks as nonlinear function approximators. NFSP trains an agent that approximates best response strategies using off-policy Q-learning. An alternate network is used to imitate the best response behavior using supervised learning. The resulting policy at execution time is a combination of fictitious and self play strategies which denote the approximate best response of agent upon convergence, hence yielding the approximate Nash Equilibria. The NFSP framework leverages two novel techniques to stabilise scalability of agent. Firstly, the agent's supervised learning component is provided with memory called reservoir to which experiences are added when agent uses its approximate best response strategies. Secondly, NFSP makes use of anticipatory dynamics which consists of best responses to a short prediction of opponent's average strategies. Provisioning anticipatory dynamics provides the agent with its own experience of best response behavior which is otherwise found absent during approximation.

NFSP demosntrates improved stability on Leduce Poker in comparison to DQN. Ablations of NFSP depict suitability of reservoir sampling and anticipatory dynamics. Additionally, NFSP presents comparable performance to state-of-the-art superhuman algorithms from the 2014 ACPC competition. While NFSP demonsrtates promise of DRL in extensive form games, its experimental setup could be further improved. The setup does not demonstrate scalability of NFSP to high dimensional spaces such as learning from pixels or multiple decks. Lastly, it would be interesting to compare NFSP to baseline human performance which could demonstrate its social decision-making capabilities.

The work aptly combines nonlinear function approximation with fictitious self play to improve stability in approximation of Nash Equilibria. NFSP presents two new directions for future work. Firstly, the framework may be leveraged to improve learning in high dimensional spaces such as pixel inputs and continuous action spaces. Secondly, large-scale comparison of NFSP to human performance would serve as a suitable baseline for extensions to Game Theoretic formulations of DRL.

The work aims to tackle stability in approximation of Nash Equilibria occuring in imperfect information games. To this end, the work presents NFSP which combines fictitious self play with DRL using neural networks as nonlinear function approximators. NFSP provides a reservoir of best response experiences to the agent along with anticipatory dynamics. Empirical evaluation of NFSP on Leduce Poker and Texas Hold'em Poker depicts stability and improved performance respectively.