Project Proposal
COMS4998: Deep Learning

Teammates: Natania Wolansky (nrw2107),
Guy Aridor (ga2449), Jisha Jacob (jsj2130)

Imperfect Information Games: Modeling Responses to Irrational Behaviour using Deep Learning

Problem Statement and Goal of Project

Modeling strategies in imperfect-information games, such as Heads-Up No-Limits (HUNL) Poker and Leduc Holdem, is difficult because the full state (the cards of the opponent) is unknown. There has been recent breakthrough work (see [6], [1]) in searching for Nash-equilibrium strategies in this space. Yet, previous work has not yet exploited the average player's tendency to deviate from rationality in the context of games [5], exhibiting changes in risk-aversion or risk-affinity over the course of play [4].

The goal of the project, therefore, is to see whether introducing irrational behavior patterns when recursively reasoning via self-learning improves the AI agent's ability to outperform average players over the Nash equilibrium strategy in the game of HUNL Leduc Holdem.

Previous Work

To overcome the challenge of states lacking unique values in imperfect-information games, two notable approaches have been explored. Prior to the approach of multi-valued states taken by Noam et. al [1], AI DeepStack used the technique of modifying the definition of state into a joint belief state across all players [6], which proved inferior to the top two HUNL players. Extending the approach of [1], Schmid et. al re-implemented the DeepStack AI agent for HUNL Leduc Holdem. Separately, recursive reasoning approaches to solve imperfect-information games have been explored since 2008, with counterfactual regret minimization being the most notable technique exploiting the degree of incomplete information in an extensive game and employing self-play to do recursive reasoning.

Data/Method and Evaluation

Our project builds off the approach of Noam et al[1]. Following their approach, we consider multi-valued states to determine optimal strategies in the game of Leduc. Using a Fall 2017 COMS 4995 project [3] as starter code, we will train a neural network via self-learning to estimate the value of each player based on their hand. We will utilize the data on observed behavior from [2] to train a second neural network, which estimates value based on self-play but with a modified self exhibiting occasionally irrational behavior as an opponent. Finally, we will train a third neural network, which estimates value based on supervised learning of historical data.

To evaluate our bots against average players (instead of rational poker bots), we will match our rational AI, our semi-rational AI and our historical AI against a holdout set of real Leduc Holdem games found in [2]. We will measure exploitability of each AI agent's technique as an objective indicator of performance and then, compare relative performance of the bots in different scenarios – against bad, good, and great players – to try to identify any domains in which the semi-rational or historical training gives an edge.

References

[1] Noam Brown, Tuomas Sandholm, and Brandon Amos. Depth-limited solving for imperfect-information games. *arXiv preprint arXiv:1805.08195*, 2018.

[2] Computer Poker Research Group, University of Alberta. Michael Maurer's IRC Poker Database. http://poker.cs.ualberta.ca/irc$_{p}oker_{d}atabase.html, 2018. Online; accessed 25 October.$

[3] Echlin, Porubin and Dunn. Coms 4995 Final Project - Fall 2017. https://github.com/mp3242/coms4995-finalproj, 2018. Online; accessed 25 October.

[4] David Eil and Jaimie W. Lien. Staying ahead and getting even: Risk attitudes of experienced poker players. *Games and Economic Behavior*, 87, 2014.

[5] Jacob K Goeree and Charles A Holt. Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review*, 91(5):1402–1422, 2001.

[6] Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisỳ, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337):508–513, 2017.