Ivey: A Reinforcement Learning Poker Engine

Aaron Jin, Ryan Cheng

Department of Computer Science, Stanford University

aaronjin@stanford.edu
rcheng07@stanford.edu

Stanford
Computer Science

Overview

This study develops poker agents using reinforcement learning (RL), focusing on preflop decision-making in heads-up No-Limit Texas Hold'em. Beginning with a push-fold framework, we expanded to include sophisticated betting options and implemented a Q-learning based system that generates strategies through self-play. Our single-agent learning paradigm, where one agent learns both positions, outperforms traditional dual-agent approaches and develops sophisticated preflop strategies that approach theoretical optimality.

Data Architecture & Features

- Training data generated through **iterative self-play** with **Q-table** structure.
- State space includes **169** unique **preflop hands**: 13 pairs, 78 suited, and 78 offsuit combinations.
- Binary position encoding (first/second to act).
- Action history tracks complete sequence of betting decisions in current hand.
- Training batch consists of **100K hands per epoch**, multiple decision points per hand.
- Pre-computed equity lookup tables stored for quick hand strength evaluation.
- Stack and pot sizes normalized to big blinds.
- Legal action masking to prevent invalid moves.
- Position-specific exploration rates optimize asymmetric learning patterns.

References

Noam Brown, Adam Lerer, Sam Gross, and Tuomas Sandholm. 2019. Deep counterfactual regret minimization. In *Proceedings of the 36th International Conference on Machine Learning*.

Sergiu Hart and Andreu Mas-Colell. 2000. A simple adaptive procedure leading to correlated equilibrium.

Michael Bowling Martin Zinkevich, Michael Johanson and Carmelo Piccione. 2007. Regret minimization in games with incomplete information. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Michael Johanson Michael Bowling, Neil Burch and Oskari Tammelin. 2007. Heads-up limit hold'em poker is solved. In *Communications of the ACM, Volume 60, Issue 11*.

Michael Bowling Viliam Lisý, Marc Lanctot. 2015. Online monte carlo counterfactual regret minimization for search in imperfect information games. In *AAMAS*.

Results

Decision Charts

Figure 1: Agent 1 vs. Random

Figure 2: Agent 1 First-to-Act

Figure 3: Agent 2 vs. Raise

Figure 4: Agent 2 vs. Call

Figure 5: Agent 1 vs. Raise

Figure 6: Agent 2 vs. Pot-Sized Raise

Methods & Experiments

$$Q(s,a) \leftarrow Q(s,a) + \alpha[R + \gamma \max Q(s',a') - Q(s,a)]$$

- Initial **dual-agent paradigm**: 2 agents with independent Q-tables competing directly, with synchronized episode updates.
- **Single-agent paradigm**: 1 agent learns both positions, with strategy copying between positions every 100,000 hands → Better convergence, reduced variance.
- Implemented a more sophisticated **batch Q-learning** system that aggregates rewards for state-action pairs + updates Q-values in batches. Improved learning stability, cached state transitions, + improved use of gathered experience data.

Discussion

- Figure 1: Agent exploits random play with aggressive raising range.
- Figure 2: Conservative strategy with premium pairs for trapping and mixed strategy for medium hands.
- Figure 3: Highly polarized pattern; strong hands call, marginal hands fold.
- Figure 4: Position-aware aggression with looser raising thresholds when acting last.
- Figure 5: Strictest strategy when out of position, only strong hands continue.
- Figure 6: Complex mixed strategy; strong hands raise, strongest hands all-in, with calculated bluffs.
- ← Graphs: Average BB/hand converges rapidly vs. fixed agents, stable converge against dynamic opponents.

Conclusion & Future Work

- Q-learning successfully learns sophisticated poker concepts through self-play.
- Primary challenge for post-flop extension: State space explosion from 169 preflop combinations to 1.7M+ flop combinations.
- For the future: Neural network integration via Deep Q-learning to combat/manage greater complexity while maintaining efficiency for training.