Mastering Leduc Hold’em: A Comparative Study of CFR in Leduc Hold’em

Presenter: Alex Chandler, Jesse Judd

04-28-2024

Motivation

- Imperfect information games model real-world strategic interactions
    - Spans card games, game theory
    - Requires reasoning about opponents
- Leduc Hold'em: A simplified variant of No-Limit Texas Hold'em
    - Preserves imperfect information and betting
    - Reduces complexity for efficient computation
    - Enables testing new algorithms and strategies

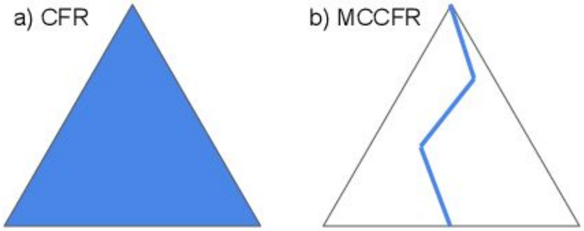
Image Source - StarryAI 2023

Previous Work

- Counterfactual Regret Minimization (CFR) (Zinkevich et al., 2008)
  - Iterative algorithm that minimizes regret
  - Converge to Nash equilibrium in two-player zero-sum games
  - Many Variants (CFR+, MCCFR, DeepCFR)

- DeepStack (Moravčík et al., 2017)
  - Deep Learning + CFR
  - Defeat professionals in heads-up no-limit Texas Hold'em

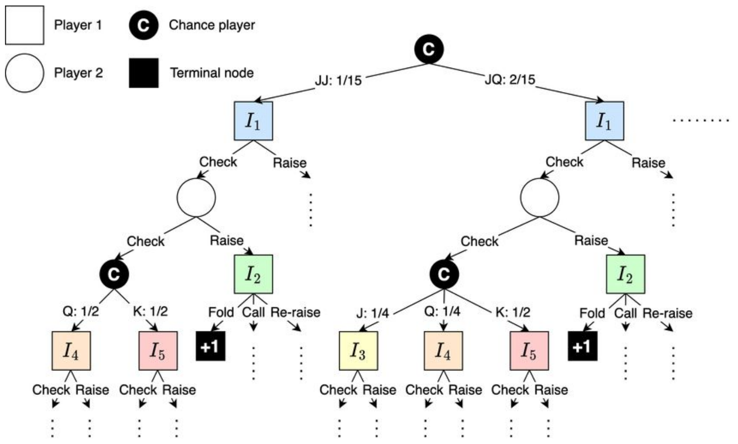
- Libratus and Pluribus (Brown & Sandholm, 2018), (Brown & Sandholm, 2019)
  - CFR and detailed abstractions
  - Defeat professionals in **multiplayer** no-limit Texas Hold'em tournaments



(Source: Schmid et al., 2018)

Leduc Hold’em: An Abstraction of NLTH

- Two-player card game
- Iterative betting -> Sequential MDP
- Actions
  - Call
  - Raise
  - Fold
  - Check
- States
  - Your card
  - Public card
  - Your chip count
  - Opponent chip count



(Habara et al. 2023)

State Representation: A 1D Binary Array of One-Hot Encoded Values

State $\leftarrow$ (Your Card $\in \{J, Q, K\}$,

Public Card $\in \emptyset$ if in first round, $\in \{J, Q, K\}$ in second round

Your Chip Count $\in \{0, 1, \ldots, 14\}$,

Opponent’s Chip Count $\in \{0, 1, \ldots, 14\}$)

Understanding Counterfactual Regret Minimization (CFR)

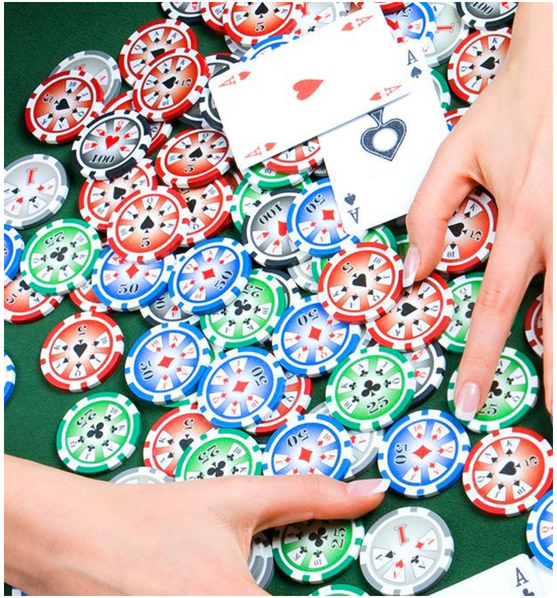
- Iterative self-play algorithm for solving imperfect information games
  - Aims to minimize regret, the difference between actual and optimal payoffs
  - Regret minimization leads to learning Nash equilibrium strategies
- Operates on the concept of counterfactual value
  - Represents the expected value of reaching a state if a player tries to reach it
  - Used to compute regrets for each action and update strategies
- Employs self-play and strategy iteration
  - Agents play against their own previous strategies
  - Strategies are updated to minimize counterfactual regret

Agents of Focus

CFR	Sarsa	Other
CFR RLCard	Sarsa	Raise Call
CFR Vanilla	Expected Sarsa	Random
CFR+	Sarsa(λ)	Q-Learning
Deep CFR	Sarsa(λ) Variant	REINFORCE
MCCFR-Outcome	N-step Sarsa	
MCCFR-External	N-step Sarsa Variant	
MCCFR-Average	True Online Sarsa(λ)	
	Deep True Online Sarsa (λ)	

Metrics

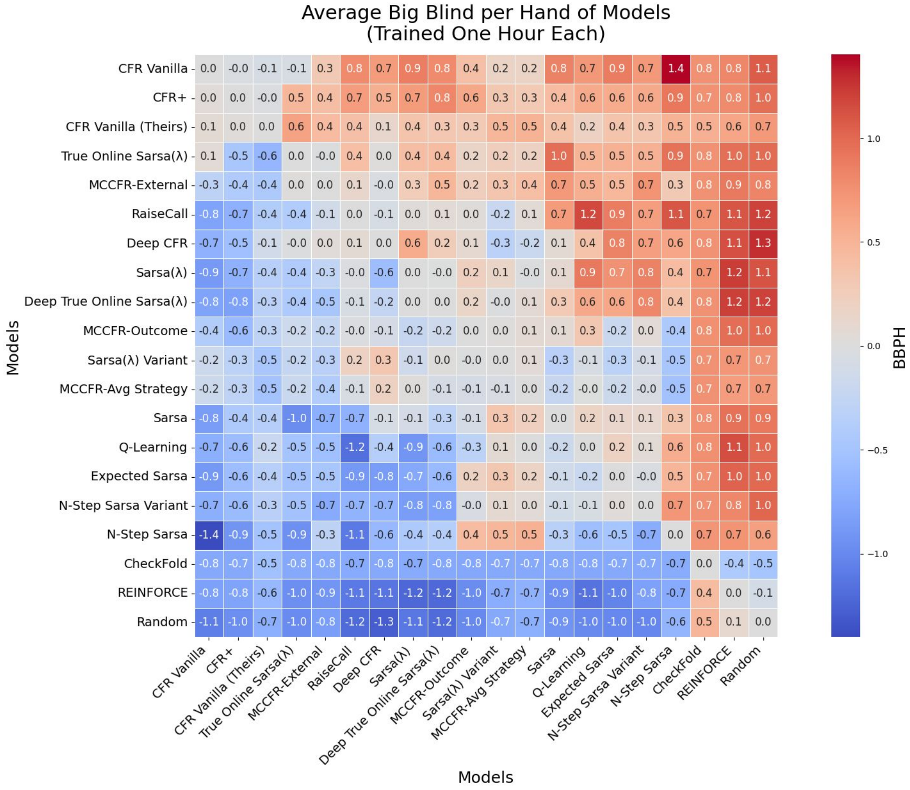
- Elo rating
  - Relative skill level between agent players
  - Accounts for strength of opponents faced
- Big Blinds Per Hand (BBPH)
  - Quantifies average profit or loss per hand
  - Expressed in terms of big blinds, a standard poker unit



Results

Agent	Action Percentage				ELO
	Call	Raise	Fold	Check	
CFR^{+}	26.9	25.8	20.8	26.5	1531.2 ± 20.3
CFR Vanilla	26.8	23.9	19.8	29.5	1501.8 ± 20.3
CFR Vanilla (Theirs)	27.6	22.7	19.0	30.7	1499.8 ± 20.3
Sarsa(λ) Variant	39.7	23.4	18.5	18.5	1382.3 ± 20.5
MCCFR-Avg Strategy	25.3	24.7	25.2	24.7	1356.0 ± 20.6
MCCFR-External	26.9	26.7	20.3	26.2	1337.5 ± 20.3
True Online Sarsa(λ)	27.2	28.6	19.6	24.7	1305.0 ± 20.3
Sarsa	24.9	32.7	20.6	21.8	1302.3 ± 20.4
Q-Learning	23.8	36.7	19.4	20.1	1286.3 ± 20.3
Expected Sarsa	23.8	36.3	20.1	19.8	1286.2 ± 20.4
Deep True Online Sarsa(λ)	23.9	39.0	18.5	18.5	1286.0 ± 20.5
N-Step Sarsa Variant	26.1	31.2	19.9	22.7	1280.2 ± 20.2
MCCFR-Outcome	25.9	26.6	21.1	26.3	1284.6 ± 20.3
Deep CFR	25.9	24.6	24.9	24.6	1285.0 ± 20.4
Sarsa(λ)	23.9	39.0	18.5	18.5	1274.3 ± 20.5
Raise Call	27.4	62.7	0.0	0.0	1271.1 ± 20.5
REINFORCE	25.4	25.9	24.6	24.0	1259.4 ± 20.2
N-Step Sarsa	24.3	33.2	22.8	19.7	1250.6 ± 20.3
Random	25.3	24.8	25.2	24.7	1247.6 ± 20.5

Table 1: Average percentages across agents, where each agent played 8,000 hands of poker against each other. Elo numbers are shown with 95% confidence intervals. We see that the top performing agents utilize a balanced betting strategy.



Bluffing CFR Agent

Our CFR agent (top) displays its ability to vary its strategy - luckily resulting in a win!

(GUI Source: Zha et al., 2020a)

Summary

- Implemented and trained 16 RL agents for Leduc Hold'em
  - CFR methods outperform other RL approaches
  - CFR+ outperforming all agents
- Betting Strategies for Top Agents
  - Employ a balanced betting strategy
  - Capable of Bluffing
- Future work
  - Extend agents to No Limit Texas Holdem
  - Incorporate betting history to state space

Sources

- Brown, N., & Sandholm, T. (2018). Superhuman AI for heads-up no-limit poker: Libratus beats top professionals.
- Brown, N., & Sandholm, T. (2019). Superhuman AI for multiplayer poker.
- Moravčík, M., Schmid, M., Burch, N., Lisý, V., Morrill, D., Bard, N., ... & Bowling, M. (2017). Deepstack: Expert-level artificial intelligence in heads-up no-limit poker
- Sutton, R. S., & Barto, A. G. (1998). Reinforcement Learning: An Introduction.
- Tammelin, O. (2014). Solving large imperfect information games using CFR+.
- Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning.
- Zha (2019). RLCard: A platform for reinforcement learning in card games. https://github.com/datamllab/rlcard
- Zinkevich, M., Johanson, M., Bowling, M., & Piccione, C. (2008). Regret minimization in games with incomplete information