Texas Hold'em Win Rate Estimation and Decision-Making Based on Monte Carlo Simulation and Machine Learning

Zhenhao Ding, He Zong, Jingyan Li

Dec 12, 2025

Abstract

This report proposes a hybrid decision-making framework for Texas Hold'em that integrates Monte Carlo simulation with modern machine learning techniques. We utilize Monte Carlo methods to estimate win rates across various game stages, providing a robust mathematical foundation for strategic actions. To address the computational bottlenecks associated with traditional hand evaluation functions, a Convolutional Neural Network (CNN) is implemented to classify hand strengths. This model achieves an accuracy of 98.75% and increases calculation efficiency by 59 times. Furthermore, a poker decision agent is trained using the Proximal Policy Optimization (PPO) algorithm and self-play, enabling dynamic strategy adjustments based on real-time win rates, opponent behavior, and pot conditions.

1 Problem Description

1.1 Background

Texas Hold'em is an incomplete information game centered on probability analysis and strategic decision-making. Players must infer their win probability based on private hole cards and shared community cards while navigating four distinct betting rounds: Pre-flop, Flop, Turn, and River.

1.2 Hand Rankings

The strength of a hand is determined by the best five-card combination chosen from seven available cards (two private and five shared). These rankings range from a High Card to a Royal Flush, dictating the ultimate winner in a showdown.

2 Decision Elements

The decision process in Texas Hold'em is multi-faceted and depends on the following elements:

- **Context:** Decisions are made at specific stages influenced by card quality, seat position, pot size, and observed opponent actions.
- **Objectives:** The primary goal is to maximize long-term profit through strategic betting and loss mitigation.
- **Alternatives:** Available actions include Raise, Call, Check, and Fold.
- **Uncertainties:** Critical unknowns include future community cards, opponent hole cards, and the unpredictable nature of opponent reactions.

3 Methodology

The proposed approach utilizes Monte Carlo simulation at each decision point for win rate estimation. To optimize performance, a CNN model is implemented to accelerate the evaluation process within the simulation loop. Finally, a PPO-based reinforcement learning agent learns optimal betting strategies through extensive self-play.

4 Model Construction and Analysis

4.1 Monte Carlo Win Rate Estimation

The simulation algorithm estimates the win rate by randomly assigning opponent hands and completing the community deck over millions of iterations. The probability of winning is calculated as:

$$P(\text{win}) = \frac{\text{win\_count}}{\text{iterations}} \tag{1}$$

In a simulation of 5 million Pre-flop hands, pocket Aces (A, A) were identified as the strongest starting hand with the highest overall win rate.

4.2 CNN Optimization for Classification

Hand evaluation is treated as a classification task to overcome the speed limitations of traditional logic.

- **Dataset Construction:** A dataset of 40,000 samples was generated, balanced across all hand classes, and supplemented with complex cases to ensure model robustness.
- **Architecture:** The CNN utilizes two convolutional layers, two pooling layers, and two fully connected layers. Dropout and ReLU activation are used to enhance learning and prevent overfitting.
- **Performance:** The CNN achieved 98.75% accuracy. Crucially, it reduced the processing time for 100,000 hands from 21.85 seconds to just 0.37 seconds.

4.3 PPO Decision Agent

The agent is trained via reinforcement learning in a self-play environment. The state space includes Monte Carlo win rates, opponent decisions, and chip counts. The PPO algorithm ensures stable policy updates, allowing the agent to explore complex strategies without deviating into unstable behaviors.

5 Quantitative Analysis

The agent was further refined using a framework based on Counterfactual Regret Minimization (CFR). Performance metrics over 1,000 training iterations are summarized below:

Table 1: Training Performance Metrics

Hands Trained	Win Rate (%)	Chip Change	Regret (CFR)	EV Difference (%)	Quality (1-10)
100	52.5%	+150	0.052	2.3%	5
500	57.8%	+330	0.012	1.0%	9
1000	60.4%	+460	0.005	0.2%	10

The consistent reduction in regret and strategy deviation indicates that the model successfully approaches a Nash equilibrium.

6 Conclusion

This research demonstrates that combining Monte Carlo simulations with CNN-based acceleration provides a powerful and feasible framework for complex game decisions. While the agent shows high technical proficiency, it remains vulnerable to human bluffing and psychological tactics. Future work will focus on integrating behavioral data and deeper game theory models to enhance decision precision in live play.