Breakthrough applications of artificial intelligence in Texas Hold'em poker

Fangze Liu

School of Computer Science and Technology, Donghua University, Shanghai, China

Abstract. The breakthrough progress of artificial intelligence in the field of incomplete information games has promoted the deep integration of game theory and machine learning technology with Texas Hold'em as a typical scenario. From a multi-dimensional technical perspective, this article systematically reviews the research progress of artificial intelligence in the field of Texas Hold'em, covering theoretical frameworks, core methods and technological breakthroughs, and looks forward to future development directions. The research focuses on three core technical directions: counterfactual regret minimization (CFR) and its improved algorithms, deep reinforcement learning (DRL) and swarm intelligence optimization. At the same time, this article points out the current challenges such as high computational complexity and insufficient dynamic adaptability, and proposes to optimize computing power constraints through lightweight models, meta-learning and quantum computing, and develop collaborative game theory and cross-domain migration frameworks. Progress in this field has also promoted the migration and application of core algorithms to more complex scenarios, providing a general paradigm for intelligent reasoning in uncertain environments.

1 Introduction

Texas Hold'em is a game of incomplete information. Unlike traditional games of complete information (such as chess and go), players in Texas Hold'em cannot know the private card information of their opponents, and the game is affected by the random card dealing mechanism, which makes the formulation of strategies more complicated. Players not only need to infer the possible strategies of their opponents based on the public cards and their betting behaviors, but also need to make decisions in a highly uncertain environment. This situation has a natural connection with the fields of business negotiations, financial transactions, medical diagnosis, etc. in reality. Therefore, Texas Hold'em has become an ideal research object in interdisciplinary fields such as game theory, decision theory and machine learning. In this kind of incomplete information environment, researchers have promoted the development of incomplete information game theory by overcoming core problems such as game tree search and counterfactual reasoning. In the unlimited bet variant of Texas Hold'em, the complexity of the decision node reaches 10160, which is far

221380113@mail.dhu.edu.cn

beyond the computational difficulty of traditional chess games. This puts higher requirements on the computational efficiency and strategy generalization ability of the algorithm. Through these interdisciplinary technological innovations, the research on Texas Hold'em has not only deepened the application of artificial intelligence in complex decision-making, but also provided transferable theories and methods for solving intelligent decision-making problems in a wider range of uncertain environments.

In 2007, the Zinkevich team at the University of Alberta proposed the CFR algorithm, ushering in the era of equilibrium solutions driven by game theory. In 2015, the CFR+ proposed by the Bowling team increased the convergence speed by 3 times through positive regret accumulation and linear weighted strategy [1]. The rise of deep reinforcement learning has opened up a new dimension for Texas Hold'em research. In 2001, Dahl implemented the reinforcement learning Q-learning algorithm in a simplified Texas Hold'em game. In 2012, Passos et al. combined the reinforcement learning method with a specific opponent model in Texas Hold'em. In 2016, Yakovenko et al. at Yahoo Labs implemented a Texas Hold'em game strategy based on human experience and deep learning algorithms [2]. In the same year, AlphaGo developed by the DeepMind team overturned traditional game cognition and successfully stimulated the exploration of data-driven methods. In 2017, DeepStack integrated CFR with deep neural networks and used sparse lookup tables to reduce computational complexity, becoming the first AI to defeat professional no-limit Texas Hold'em players. Libratus, launched by Carnegie Mellon University's Brown team in 2017, uses nested subgame solving technology to surpass human professional players at a high cost with the support of supercomputers [3]. In 2019, Pluribus defeated top human players at a six-person table for the first time through the Blueprint strategy library and online subgame balancing technology, demonstrating a breakthrough in multi-player game strategy [4]. Meta-learning and lightweight architecture have become the core directions of technological breakthroughs. The two-stage adaptive framework proposed by Liu Zhiqing's team at Beijing University of Posts and Telecommunications compressed the response speed to milliseconds through offline pre-training and online strategy distillation; the AlphaHoldem program of Xing Junliang's team at the Chinese Academy of Sciences has a decision-making speed that is more than 1,000 times faster than DeepStack [5]. With the rise of large language models, Texas Hold'em AI has ushered in a new revolution: the Suspicion Agent of the University of Tokyo combines GPT-4 with high-level theory of mind, and uses natural language understanding and generation capabilities to analyze opponent strategies; The open-source framework Poker large language models (LLM), which combines LLM, adopts a modular architecture to support multi-player matches, and uses generative thinking chains and post-reflection mechanisms, resulting in a 42% improvement in AI bluffing accuracy after training on 100,000 games; the commercial system ALPHAX Alpha Core achieves multi-modal expansion, using unstructured data such as voice intonation and betting time to construct a 42-dimensional psychological feature of the opponent's portrait, and the accuracy of identifying cheating behavior in WSOP online events is as high as 98%. These achievements promote the deep integration of large language models and game theory, and lead the research on incomplete information games into a new stage of "cognitive game".

2 Imperfect Information Game

Game theory is a mathematical theory system that studies the behavioral choices of rational decision makers in strategic interactions. Its core goal is to characterize the stability of strategy combinations by constructing Nash equilibrium solutions, that is, in the stable state formed by the strategies of all participants, any unilateral deviation cannot improve their own benefits [6]. The Nash equilibrium strategy is a strategy combination composed of the

best response of each player, satisfying the mathematical conditions, as shown in Equation 1:

$$\forall i, u_i(\pi_i^*, \pi_{-i}^*) \geqslant \max_{\pi_i'} u_i(\pi_i', \pi_{-i}^*) \quad (1)$$

In this important branch of imperfect information games, participants use "information sets", such as private hands, public card sequences, and opponent betting behavior in Texas Hold'em, to represent their limited knowledge of the situation, and require strategies to remain consistent within the information set to cope with information asymmetry. The equilibrium strategy π of this type of game needs to meet dual characteristics: it must not only build a probability model of the opponent's strategy range (Hand Range), but also balance bluffing and value betting in the mixed strategy space. In view of the computational complexity obstacles of strict Nash equilibrium, actual systems often use approximate Nash equilibrium that allows ε errors and meets the mathematical conditions, as shown in Equation 2

$$\forall i, u_i(\pi_i^*, \pi_{-i}^*) + \epsilon \geqslant \max_{\pi_i'} u_i(\pi_i', \pi_{-i}^*) \quad (2)$$

While retaining the essence of strategy stability, this model achieves computability of complex decision-making scenarios by relaxing constraints, providing an analytical framework that combines theoretical rigor and engineering feasibility for practical systems such as Texas Hold'em AI.

3 Counterfactual Regret Minimization

3.1 CFR Basics

Counterfactual regret minimization is a method used to solve the Nash equilibrium problem in game theory [7], and is particularly suitable for two-player zero-sum extended incomplete information game scenarios such as Texas Hold'em. In order to solve the computational difficulties caused by the exponential growth of the game tree size in extended games, CFR introduces the concept of "counterfactual regret" to decompose the overall strategy optimization into local regret minimization on each information set. First, the "regret value" is defined as the difference between the benefit of a certain action and the benefits of other optional actions in the information set. Starting from the initial strategy, the player evaluates the potential benefit difference of each action based on historical results and the current strategy after each round of decision-making. The strategy is adjusted through the "regret matching" mechanism so that the probability of action selection is proportional to the positive regret value. The actions with higher historical accumulated regret (i.e., better potential benefits) are given priority, and the strategy deviation is gradually corrected. In each iteration of the CFR algorithm, as shown in Equation 3-6.
Regret value updates the policy:

$$\sigma_p^t(I,a) = \begin{cases} \frac{R^t(I,a)^+}{\sum_{b\in A(I)} R^t(I,b)^+} & \sum_{b\in A(I)} R^t(I,b)^+ > 0 \\ \frac{1}{|A(I)|} & otherwise \end{cases} \quad (3)$$

The new policy updates the average policy:

$$\bar{\sigma}_p^t(I,a) = \frac{1}{t}\sum_{t'=1}^{t} \pi_p^{\sigma^t}(I)\sigma_p^{t'}(I,a) \quad (4)$$

The new policy calculates the virtual value:

$$v_p^{\sigma^t}(I,a) = \sum_{h\epsilon I\cdot a} v_p^{\sigma^t}(h) \quad (5)$$

The virtual value updates the regret value:

$$R^t(I,a) = R^{t-1}(I,a) + v_p^{\sigma^t}(I,a) - \sum_{a\in A(I)} \sigma^t(I,a) v_p^{\sigma^t}(I,a) \quad (6)$$

This iterative learning mechanism not only takes into account the uncertainty of player decision-making, but also avoids the traditional method of fully traversing the game tree. By dynamically updating the expected return and strategy weights, it eventually converges to an approximate Nash equilibrium in theory, providing a feasible framework that takes into account both accuracy and efficiency for solving strategies for large-scale complex games. Figure 1 shows a basic CFR algorithm iteration process.

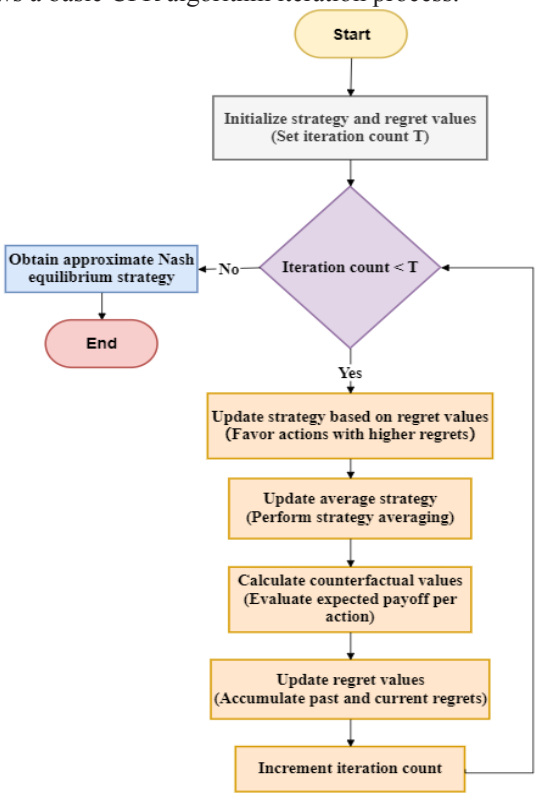

Figure 1 CFR algorithm flow chart (Picture credit: Original)

3.2 CFR Optimized Variants

Monte Carlo CFR (MCCFR) combines CFR with the Monte Carlo (MC) method [8]. In each round, only part of the information set is accessed and updated, and the estimated sampled virtual value is used to minimize the regret. The "time for space" strategy solves the defect of the CFR algorithm that needs to traverse and update the game tree under limited computing resources. Noam Brown proposed a pruning-based CFR algorithm [9], which accelerates its iteration process and makes it more suitable for handling large-scale

games. The CFR+ algorithm improves efficiency through three key optimizations: only accumulating positive regret values to avoid invalid oscillations of low-return actions, using a linear weighted average strategy to enhance the impact of later iterations, and introducing an alternating update mechanism to reduce the amount of calculation. Discounted CFR (DCFR) and Linear CFR (LCFR) further optimize the convergence performance. By adjusting the regret value accumulation method and the strategy update rules, the algorithm is continuously improved in efficiency and stability.

DeepCFR introduces deep learning technology into the CFR framework [10], innovatively using neural networks to approximate CFR behavior, avoiding the traditional CFR's complete traversal of the entire game tree. Its core consists of a regret value network and an average policy network. The value network is responsible for estimating the regret value of the information set to guide policy optimization, while the policy network learns and outputs the probability distribution of each action in the current state. Figure 2 shows the network structure of DeepCFR.

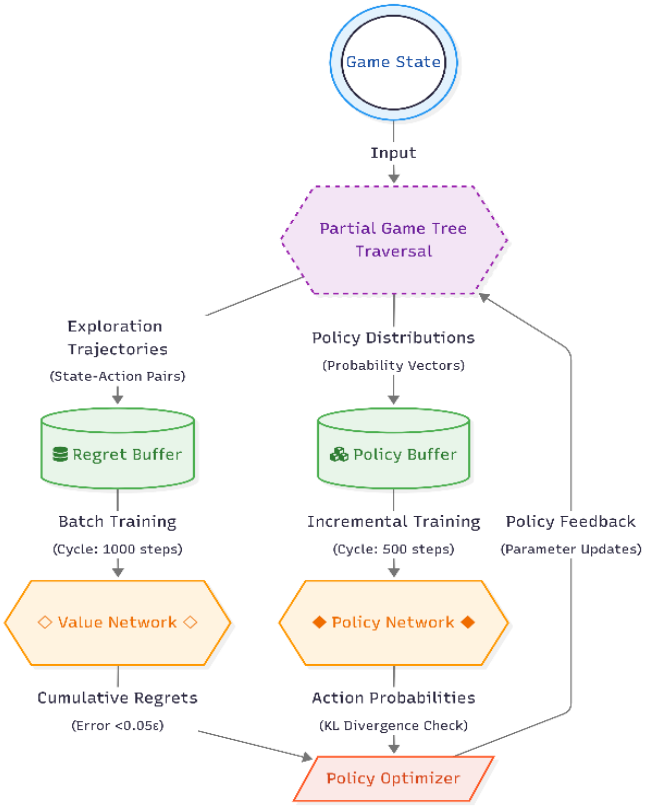

Figure 2 DeepCFR algorithm training framework (Picture credit: Original)

Double Neural CFR (DNCFR) adopts a dual neural network architecture and optimizes the sampling and training process [11]. Simplified Deep CFR (SD-CFR) improves performance by simplifying the policy update process [12]. It directly calculates the average policy based on the regret value network, eliminating the step of training the average policy network separately, reducing approximation errors and accelerating convergence. Steinberger proposed Dream, a model-free algorithm based on deep learning, based on DeepCFR, combined with variance reduction, result sampling and other techniques. It is still one of the powerful baseline algorithms in the field of imperfect information [13]. McAleer proposed ESCHER to directly train the historical value function

to estimate the cumulative regret, and use the estimated value as a baseline to reduce the valuation variance. On the basis of Dream, it further improves the convergence speed and stability [14]. Multi-Head DeepCFR (MHDeepCFR) optimizes the feature representation ability of the policy network by introducing a multi-head attention mechanism to cope with the decision-making needs of complex feature combinations in imperfect information games [15]. The Fig.3 below shows the network structure of DeepCFR combined with multi-head attention.

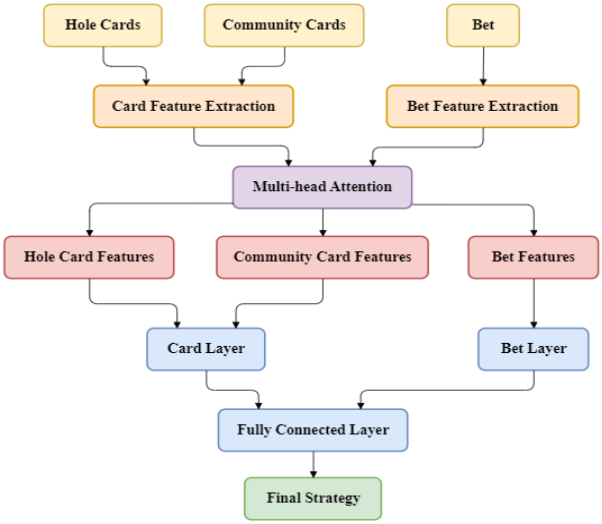

Figure 3 Flowchart of the DeepCFR algorithm incorporating mutual information (Picture credit: Original)

The multi-head attention mechanism is embedded in the policy network, enabling the model to dynamically capture the correlation between state features, such as the potential connection between hand combinations and opponent behaviors in poker, thereby optimizing feature weight distribution and enhancing the accuracy of strategy generation. Mutual Information DeepCFR (MIDeepCFR) focuses on the unknown decision exploration problem in policy optimization. By introducing mutual information as a regularization term, it balances the diversity of policies and the correlation of states. The mutual information term quantifies the dependency between states and actions. Regularization terms are added to policy network training to balance exploration and utilization, encourage attempts at unexplored actions, and improve unknown decisions. By maximizing mutual information, the policy network is encouraged to explore decision paths that have not been fully evaluated, effectively balancing exploration and utilization. As shown in Fig. 4.

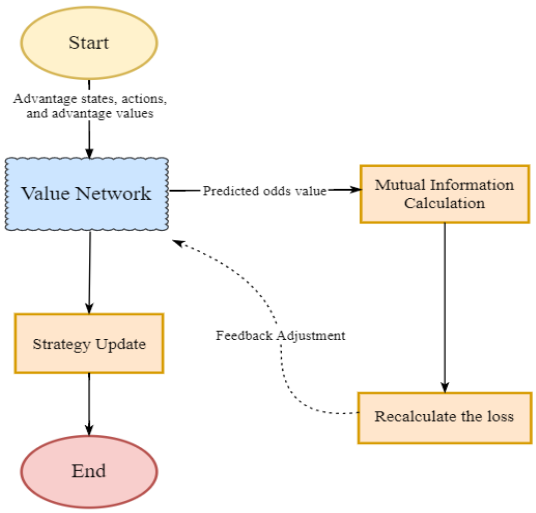

Figure 4 Flowchart of DeepCFR algorithm incorporating mutual information (Picture credit: Original)

4 Deep Reinforcement Learning

The core logic of reinforcement learning is to learn the mapping strategy from high-dimensional states to actions through trial and error interaction between the agent and the environment, so as to maximize the long-term cumulative return. The reinforcement learning modeling of Texas Hold'em requires the definition of a state space containing hand cards, public cards, and historical betting information, as well as action spaces such as folding, calling, and raising. The reward mechanism is designed based on the final hand strength and betting results. Actor-Critic is a hybrid architecture that combines Policy Gradient and Value Function in reinforcement learning. It separates strategy optimization from value evaluation, and improves learning efficiency and stability through the collaboration of two independent modules. The basic Actor-Critic structure is shown in Fig. 5.

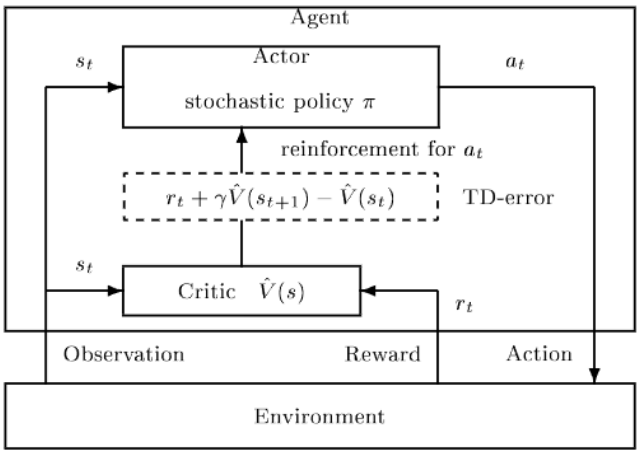

Figure 5 Actor-Critic algorithm framework diagram (Picture credit: Original)

Advantage Actor-Critic (A2C) introduces an advantage function to replace the original return, reduces the policy gradient variance by subtracting the state value baseline, and

adopts a synchronous update method to allow multiple parallel environments to uniformly collect data for update to avoid asynchronous noise. Asynchronous Advantage Actor-Critic (A3C) uses asynchronous parallel training to allow multiple agents to execute strategies in parallel in independent environments and update the global model, and improves computational efficiency through asynchronous updates, making it more suitable for distributed training. Soft Actor-Critic (SAC) improves policy diversity through entropy regularization to avoid targeted exploitation by opponents. Proximal Policy Optimization (PPO) limits the policy update range through importance sampling and pruning, optimizes the advantage function estimation in combination with Generalized Advantage Estimation (GAE), and uses batch normalization, hierarchical training and hybrid exploration strategies to enhance adaptability to complex environments. Deep Deterministic Policy Gradient (DDPG) optimizes the continuous action space, directly outputs deterministic actions to avoid discretization losses, suppresses Q value overestimation through dual critic networks (TD3), delays updating actors and improves convergence stability through priority experience replay, introduces architectural innovations such as Long Short-Term Memory (LSTM), residual connections and target networks, and enhances state reasoning capabilities in some observable environments. The Fig.6 shows the network structure of DDPG.

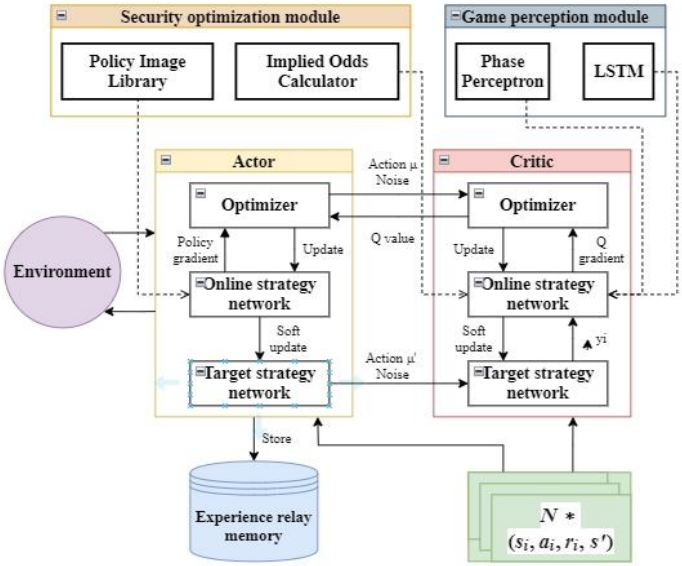

Figure 6 DDPG network structure diagram (Picture credit: Original)

Deep Q-Learning (DQN) combines deep learning with reinforcement learning, introduces a target network and experience replay mechanism, avoids policy oscillation, and breaks temporal correlation. The improved version Double DQN reduces overestimation bias by separating action selection and value evaluation, while Dueling DQN further optimizes convergence stability by decomposing state value and action advantage. The Politex algorithm revolutionizes the policy iteration paradigm and optimizes the policy update trajectory by constraining the upper bound of the regret value. In the direction of game theory integration, multi-dimensional innovations have emerged: the LONR algorithm implants the regret-free mechanism into the Q-learning framework and proves its convergence in the Markov Decision Process (MDP) [16]. The Neural Regret Minimization Dynamics (NeuRD) algorithm creatively integrates the imitator dynamics and policy gradient of evolutionary game theory, and uses the network logit value to replace the

policy output to eliminate gradient bias [17]. The Exploitability Descent (ED) algorithm introduces the asymmetric optimization idea of CFR-BR and constructs a policy search space based on the best response hypothesis; for large-scale information set scenarios, the Adaptive Regret Minimization for Action Coordination (ARMAC) algorithm achieves valuation optimization through policy mirror backup and off-policy learning. These technological innovations jointly build an algorithm system that has both environmental adaptability and theoretical convergence, promoting the development of Texas Hold'em AI.DDPG network structure diagram.

5 Swarm Intelligence Optimization Algorithm

Swarm intelligence optimization algorithms simulate the behavior of groups in nature and build a computational framework for distributed search and adaptive evolution. Their application in complex optimization tasks has shown a significant expansion trend [18]. As a typical representative in this field, the particle swarm optimization algorithm (PSO) has become a core tool for solving high-dimensional nonlinear optimization problems with its simple principles, few parameters, and easy implementation. Figure 7 shows a simple PSO algorithm flow.

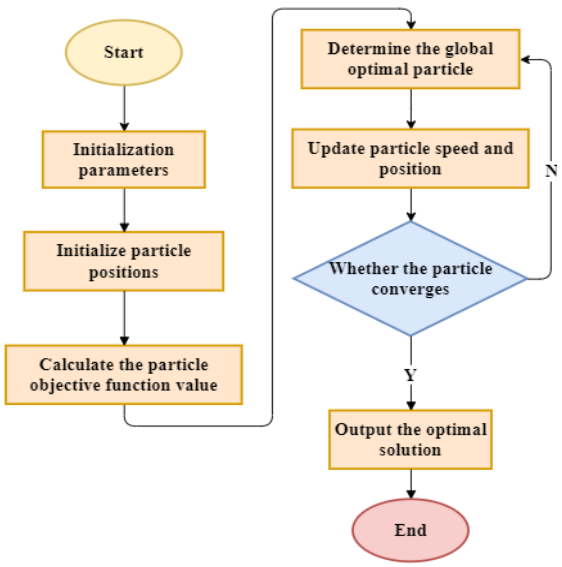

Figure 7 PSO algorithm flow chart (Picture credit: Original)

In the Texas Hold'em game scenario, the distributed exploration characteristics and dynamic environment adaptability of swarm intelligence algorithms show unique value. By simulating the multi-agent collaboration mechanism, it efficiently searches the strategy space without traversing the game tree, breaking through the computational bottleneck of traditional methods such as CFR, and providing key technical support for Texas Hold'em AI to quickly generate equilibrium strategies in real-time confrontation. In the application of PSO, poker strategies are modeled as particles in a multidimensional space, and each particle represents a strategy configuration, such as betting ratio or bluffing frequency. By dynamically adjusting the inertia weight and acceleration coefficient through adaptive PSO, the convergence speed of the algorithm in multi-player scenarios is significantly improved. For example, adaptive PSO can quickly respond to changes in opponent strategies, optimize decision parameters in the pre-flop and flop stages, and ensure that AI remains competitive

in complex games. In addition, the combination of PSO and PPO further improves the efficiency of strategy generation by optimizing hyperparameters or directly adjusting the CFR strategy table. The Best Response PSO (BR-PSO) algorithm has made multi-dimensional improvements to the defects of traditional PSO in dynamic games, and significantly improved the stability and pertinence of strategy optimization through a dual optimal solution replacement mechanism [19]. In local optimization, when random card dealing causes particle fitness to fluctuate negatively, the algorithm selects alternative solutions from the historical positive fitness particle library to avoid degradation tracking. At the global optimization level, the search direction is dynamically adjusted by calculating the center of gravity of the elite solution, weakening the interference of randomness on the global optimal solution, and accurately capturing the opponent's strategy characteristics. In addition, BR-PSO proposes the concept of "time-dimensional particles", breaking through the traditional space-dimensional population definition, making the population evolution adapt to the serial game characteristics of online competitions, and building personalized response plans through progressive optimization of strategies within a limited number of games. BR-PSO realizes real-time strategy generation based on a lightweight S-curve parameterized model, and improves computing efficiency by more than two orders of magnitude. Its technical advantages are specifically manifested in Texas Hold'em: the low-dimensional model supports opponent behavior response in seconds, and the anti-randomness mechanism improves the long-term profit stability by more than 30%. The time-dimensional particle framework constructs a feasible path from strategy initialization to targeted optimization within a hundred games, providing a new paradigm for the development of low-cost, highly adaptable poker AI.

In the discrete decision-making scenario, ant colony optimization (ACO) handles discrete decision-making problems in poker by simulating the ant colony pheromone mechanism, such as folding, calling, raising and other strategic choices. Figure 8 shows a ACO algorithm flow.

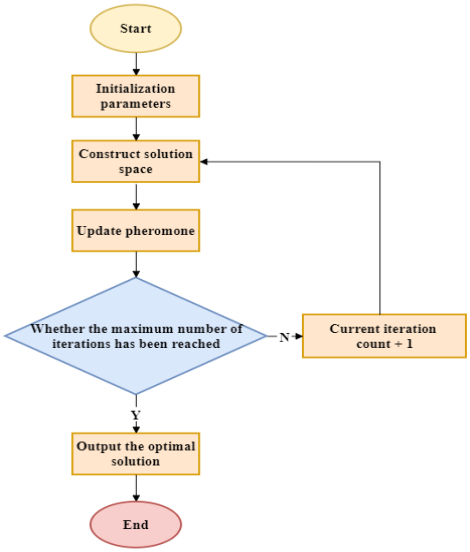

Figure 8 ACO algorithm flow chart (Picture credit: Original)

The improved ACO algorithm combines a multi-objective optimization framework to maximize expected returns while controlling risks (such as chip losses from failed bluffs) to generate more balanced strategies. When combined with the Monte Carlo CFR variant, ACO focuses on high-value action sequences (such as value bets in key rounds), reducing

the computational cost of large-scale game tree traversal, and is particularly suitable for real-time multi-player scenarios. It allocates sub-groups to different game stages such as pre-flop and post-flop for modular optimization, enabling AI to adjust strategies in real time according to the opponent's betting pattern. For example, the deep Q network optimized by PSO can accelerate equilibrium convergence in complex games, and the CFR variant optimized by ACO significantly reduces the demand for computing resources. In addition, swarm intelligence technology supports cross-variant migration of strategies. The optimized strategy can be migrated from the unlimited Texas Hold'em scenario to the limited scenario, effectively improving the versatility of the algorithm.

6 Classic Framework and Future Research

The AI Texas Hold'em system is a modular, multi-level adaptive architecture built by integrating game theory, deep reinforcement learning and real-time decision optimization technology. Its core is the game state management module that dynamically maintains the information of the game. The probability calculation engine evaluates the winning rate based on Monte Carlo simulation, analyzes the behavior pattern and predicts the strategy range in combination with the opponent modeling system, and then generates an approximate Nash equilibrium strategy through the application of the improved CFR+ algorithm through the game theory solver. The system adopts an end-to-end learning framework, optimizes the strategy network through self-game training, and integrates the risk management module to dynamically adjust the capital allocation. In the decision engine, it integrates multi-dimensional inputs to achieve real-time decision-making that balances exploration and utilization. At the key technical level, the system realizes information set abstraction through deep neural network learning compact state representation, uses Bayesian range estimation and recursive neural network to build a dynamic opponent model, and combines hierarchical parallel computing and GPU acceleration to solve the computational complexity challenges brought by high-dimensional state space. The auxiliary modules cover data management, simulation testing and continuous performance monitoring, forming a complete closed loop from strategy generation, decision execution to self-optimization, with high computing efficiency (such as AlphaHoldem decision-making in just 3 milliseconds) and strong environmental adaptability, and can achieve strategy evolution through continuous learning mechanisms. The overall AI Texas Hold'em system framework is shown in the Fig. 9.

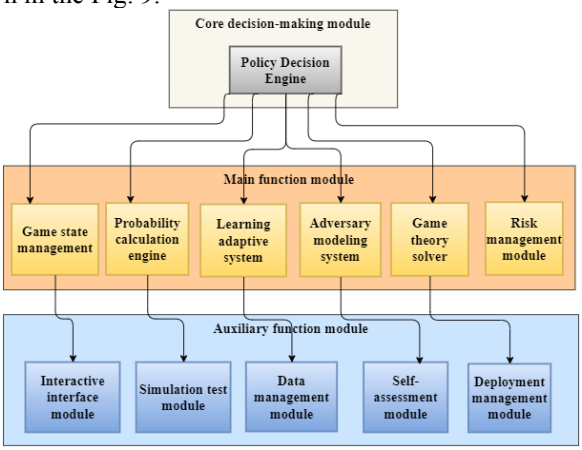

Figure 9 AI Texas Hold'em system framework diagram (Picture credit: Original)

With the continuous progress of artificial intelligence technology and the in-depth development of incomplete information game research, a series of Texas Hold'em AI systems with their own characteristics have been launched on the basis of the theoretical and technical accumulation of the above core framework. They continue to explore new technical paths and application boundaries, and promote the development of this field to a higher level. While inheriting the core ideas of the framework, these systems combine the technical advantages and research priorities of different periods to make innovative breakthroughs, forming rich technical practice results. The Fig. 10-13 below show several classic system architectures.

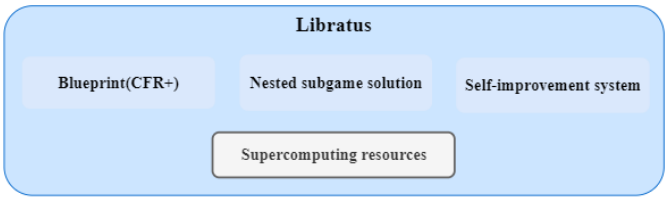

Figure 10 Libratus system framework (Picture credit: Original)

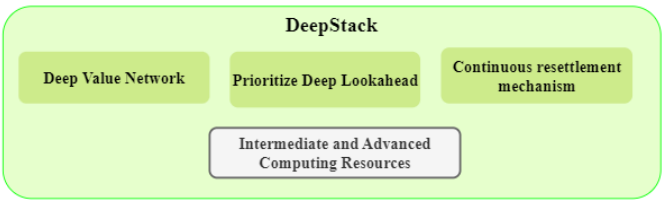

Figure 11 DeepStack system framework (Picture credit: Original)

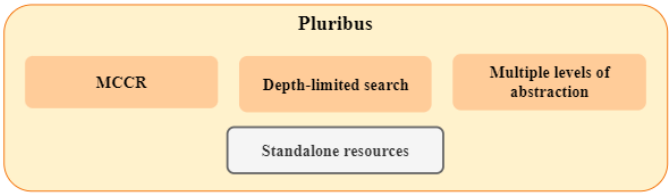

Figure 12 Pluribus system framework (Picture credit: Original)

Figure 13 AlphaHoldem system framework (Picture credit: Original)

The development of Texas Hold'em AI is constrained by the ultra-high state space and complex game environment caused by imperfect information. Traditional algorithms such as counterfactual regret minimization have obvious bottlenecks such as the sharp increase in computing costs caused by traversing the game tree, difficulty in implementing real-time scenarios, insufficient adaptation to dynamic strategies, and limited generalization capabilities. Dependence on supercomputers, delayed response to sudden changes in opponent styles under the assumption of fixed strategies, the complexity dilemma of multiplayer games, and "irrational" strategy response defects restrict its further development. To this end, researchers have proposed multiple innovative solutions to the

problem of dynamic strategy optimization: construct a lightweight and efficient two-stage adaptive framework, realize the real-time strategy generation of new opponents through offline multi-style opponent training and online dynamic modeling; design a three-level module general framework based on meta-game theory, including strategy identification, space reconstruction and utilization, to systematically solve the strategy matching problem of dynamic confrontation; use the swarm neural evolution method to capture significant behavioral characteristics of the opponent network to drive the real-time decision optimization of the pre-trained network, and balance strategy diversity and convergence efficiency; develop an end-to-end self-game reinforcement learning architecture, realize the autonomous evolution of strategies through the game between intelligent agents and historical strategies, and avoid the high cost of traditional game tree search; optimize the structure of deep reinforcement learning models, reduce the output layer dimension to reduce the computational complexity of large state space, and improve the algorithm operation efficiency. These breakthroughs from adaptive frameworks, dynamic modeling, evolutionary optimization, self-game mechanisms to lightweight models provide a path to solve core challenges such as incomplete information and strategy space explosion, and promote the migration of results to real-world scenarios such as financial risk control and multi-agent collaboration. In the future, the team will focus on lightweighting and meta-learning to achieve few-sample adaptation in the short term, and explore quantum computing to optimize game tree search in the long term. At the same time, the team will develop adaptive strategies that integrate online learning and dynamic opponent modeling, as well as CFR variants that support collaborative reasoning. Its game reasoning technology has demonstrated migration value in fields such as finance and healthcare, and is expected to expand from game scenarios to broader areas of intelligent decision-making, achieving a deep integration of theory and application.

7 Conclusion

The research on Texas Hold'em algorithms represents the pinnacle of the integration of artificial intelligence and game theory, demonstrating the huge potential for technological breakthroughs and theoretical innovations. From early explorations based on CFR to Libratus and Pluribus surpassing top human players in no-limit Texas Hold'em, the performance of algorithms in imperfect information games has reached a superhuman level. Its core contribution lies in the development of efficient strategy search technologies, such as deep reinforcement learning and opponent modeling, which effectively solve the computational difficulties of high-dimensional decision spaces. AlphaHoldem achieves millisecond-level decision-making through lightweight design, further marking a leap in the practicality of algorithms. These advances not only enhance AI's reasoning ability in dynamic adversarial environments, but also provide a general framework for dealing with uncertainty. Its impact goes far beyond the competitive field, promoting breakthroughs in multi-agent systems, especially strategy optimization in scenarios with asymmetric information, laying the foundation for applications such as autonomous driving and robot collaboration; in game theory, it verifies the computability of Nash equilibrium in complex environments and enriches the theory of mixed strategies. Cross-domain applications are equally far-reaching. For example, in quantitative investment, risk management can be optimized and the accuracy of return prediction can be improved by 10%-15%, highlighting the value of Texas Hold'em AI as a general intelligent testing ground. Looking ahead, algorithm efficiency can be further improved through lightweight models and meta-learning to reduce computing power dependence and enhance generalization capabilities; collaborative game research should focus on dynamic alliances and human-machine hybrid strategies to simulate real interaction scenarios; cross-domain migration requires the

establishment of a standardized framework to systematically apply game reasoning to social issues such as resource allocation and negotiation optimization. Collaboration in the open source community will accelerate innovation, while paying attention to the ethical boundaries of AI competition ensures technical fairness. The continued evolution of Texas Hold'em AI will drive artificial intelligence towards higher intelligence and open up new possibilities in the theory and practice of uncertain decision-making.

References

1. Bowling, M., Burch, N., Johanson, M., et al.: 'Heads-up limit hold'em poker is solved', Science, 2015, 347(6218):145–149
2. Yakovenko, N., Cao, L., Raffel, C., et al.: 'Poker-CNN: A pattern learning strategy for making draws and bets in poker games using convolutional networks', Proceedings of the AAAI Conference on Artificial Intelligence, 2016, 30(1)
3. Brown, N., Sandholm, T.: 'Superhuman AI for heads-up no-limit poker: Libratus beats top professionals', Science, 2018, 359(6374): 418–424
4. Brown, N., Sandholm, T.: 'Superhuman AI for multiplayer poker', Science, 2019, 365(6456): 885–890
5. Zhao, E., Yan, R., Li, J., et al.: 'AlphaHoldem: High-Performance Artificial Intelligence for Heads-Up No-Limit Poker via End-to-End Reinforcement Learning', Proceedings of the AAAI Conference on Artificial Intelligence. 2022, 36(4):4689–4697
6. Huale, L., Wang, X., Jia, F., et al.: 'A Survey of Nash Equilibrium Strategy Solving Based on CFR', Archives of Computational Methods in Engineering, 2020, 28: 2749–2760
7. Musah, I., Boah, D. K., Seidu, B.: 'A comprehensive review of solution methods and techniques for solving games in game theory', Journal of Game Theory, 2020, 9(2): 25–31
8. Brown, N., Sandholm, T.: 'Superhuman AI for Multiplayer Poker', Science, 2019, 365(6456): 885–890
9. Brown, N., Sandholm, T.: 'Regret-based Pruning in Extensive-form Games', Advances in Neural Information Processing Systems. Toronto, Ontario, Canada: AAAI Press, 2015: 1972–1980
10. Brown, N., Lerer, A., Gross, S., et al.: 'Deep counterfactual regret minimization', Proceedings of the 36th International Conference on Machine Learning (ICML). Long Beach, USA, 2019: 793–802
11. Li, H., Hu, K., Zhang, S., et al.: 'Double Neural Counterfactual Regret Minimization', Proceedings of the 8th International Conference on Learning Representations (ICLR), Addis Ababa, Ethiopia, 2020
12. Steinberger, E.: 'Single deep counterfactual regret minimization', arXiv preprint arXiv:1901.07621, 2019
13. Steinberger, E., Ler, A., Brown, N.: 'Dream: Deep regret minimization with advantage baselines and model-free learning', arXiv preprint arXiv:2006.10410, 2020
14. Mcaleer, S., Farina, G., Lanctot, M., et al.: 'Escher: Eschewing importance sampling in games by computing a history value function to estimate regret', arXiv preprint arXiv:2206.04122, 2022.

15. Xue, K. C.: 'Research on machine gaming with incomplete information based on CFR algorithm', Harbin University of Science and Technology, 2024
16. Kash, I. A., Sullins, M., Hofmann, K.: 'Combining no-regret and Q-learning', Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Auckland, New Zealand, 2020: 593–601
17. Hennes, D., Morrill, D., Omidshafiei, S., et al.: 'Neural replicator dynamics: Multiagent learning via hedging policy gradients', Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Auckland, New Zealand, 2020: 492–501
18. Tang, J., Liu, G., Pan, Q., T.: 'A review on representative swarm intelligence algorithms for solving optimization problems: Applications and trends', IEEE/CAA Journal of Automatica Sinica, 2021, 8(10): 1627-1643
19. Hu, Z., Chen, S., Yuan, W., Li, P., Chen, J.: 'Online opponent exploitation in Texas Hold'em based on particle swarm optimization'. Control and Decision, 2024, 39(5): 1687–1696