Proceedings of CONF-CDS 2025 Symposium: Data Visualization Methods for Evaluatio
DOI: 10.54254/2755-2721/2025.PO24698

Research on Board Game Strategy Methods Based on Reinforcement Learning

Runzhou Luo[1*], Boning Yao[2], Qingwen Zhang[3]

[1]School of Software, Dalian University of Technology, Dalian, China
[2]College of Artificial Intelligence, Shenyang Normal University, Shenyang, China
[3]International Education College, Shanghai Jian Qiao University, Shanghai, China
**Corresponding Author. Email: 1919493194@mail.dlut.edu.cn*

Abstract. As a typical sequential decision-making and gaming problem, board games have the complexity of large state space and strong dynamic confrontation, however, traditional methods have many limitations in dealing with them, so they need to be based on reinforcement learning to achieve strategy optimization by virtue of data-driven. Reinforcement learning can promote the realization of AI decision-making ability from rule-dependent to data-driven leap, and show significant advantages in game AI. This paper systematically sorts out the core algorithms of reinforcement learning in board games, comparatively analyzes their technical characteristics, applicable scenarios, advantages and disadvantages, discusses the current technical bottlenecks and ethical challenges, and look forward to the future development direction. This paper concludes that reinforcement learning is effective in board games, which not only helps AIs such as AlphaGo and Libratus to surpass the human level in Go, Texas Hold'em and other scenarios, but also forms the transition from “model-dependent” to “data-driven”, From “model-dependent” to “data-driven”, and from “single-intelligence” to “multi-intelligence”, it has also formed a technological evolution vein. At the same time, reinforcement learning has been breaking through in processing high-dimensional states, complex reward functions, etc., and has shown the potential of generalization in the fields of education, healthcare, etc. [1]. This paper can provide theoretical references and practical guidance for subsequent AI research on board games, as well as a universal methodology for complex decision-making problems.

Keywords: Reinforcement learning, board games, Q-learning, DQN, PPO

1. Introduction

As a classic paradigm of human intellectual competition, board games are essentially sequential decision-making and game-versus-play problems in dynamic environments [1]. Its core challenges lie in the exponential explosion of the state space (e.g., 10^170 positions in Go), information incompleteness (e.g., hidden hands in Texas Hold'em), and strategic adversariality in multi-intelligent body interactions [1]. Traditional approaches rely on rule engines and heuristic search, e.g., the Minimax algorithm in chess searches the game tree by manually designing the evaluation

DOI: 10.54254/2755-2721/2025.PO24698

function, but its efficiency decreases exponentially with the growth of the state space, making it difficult to cope with complex scenarios such as Go, and rule-based strategies are even more incapable of dealing with the uncertainty brought by hidden information in incomplete information games due to their lack of generalization ability [1]. As the complexity of the game increases, the limitations of such methods in high-dimensional state representation, dynamic strategy adjustment, etc. become more and more significant, i.e., data-driven intelligent decision-making techniques are needed to break through the bottleneck of traditional frameworks [1].

The rise of reinforcement learning has revolutionized board game AI [1]. The technique optimizes decision-making strategies to maximize long-term cumulative rewards through trial-and-error interactions between the intelligences and the environment, and demonstrates unique advantages in dealing with high-dimensional states, dynamic rivalry, and sparse reward problems [1]. For example, in 2023, the DouZero framework proposed by the Racer AI team [2] combines deep neural networks and Monte Carlo tree search, targeting the super-large scale action space of Landlord (about 27,000 legal card combinations), and through action coding and parallel self-gaming techniques, the AI's winning rate in complex card games is significantly improved, ranking first among 344 AI models on the Botzone platform [3].In 2024, the Tencent AI Lab team proposed the DQN-IRL framework for the sparse reward problem of landlord fighting, and reconstructed the reward function through inverse reinforcement learning, which improved the AI decision-making efficiency by more than 30% [4]. In the field of incomplete information games, Tuomas Sandholm's team at Carnegie Mellon University proposed the Actor-Critic framework based on fictitious self-pairing games in 2025, which enables AI to achieve Nash equilibrium strategy approximation in multi-player Texas Hold'em poker, with a winning rate of more than 25% over the human professional players through joint optimization of the strategy function and the value function [5]. These practices not only validate the technical leap of reinforcement learning from “model-dependent” to “data-driven”, but also provide methodological support for robot control, autonomous driving and other fields through innovative techniques such as experience playback and goal networks.

Against this background, this paper focuses on the systematic application of reinforcement learning in board games, aiming to construct a theoretical framework for the subsequent research by sorting out the core algorithms' evolution and analyzing the technical suitability and scenario constraints. The study firstly explains the basic theories of reinforcement learning, including the core elements and Markov Decision Process, which lays the theoretical foundation for the subsequent analysis; secondly, it categorizes and analyzes the typical algorithms, covering the value function methods (e.g., Q-learning, DQN), strategy gradient methods (e.g., PPO, Actor-Critic), etc., and compares the technical characteristics, advantages and limitations of these algorithms; Then, the practical application cases of the algorithms in specific chess game scenarios of chess games are analyzed (e.g., complete information game, incomplete information game, single-intelligence and multi-intelligence confrontation); finally, the current technical bottlenecks were summarized, the future development direction was prospected, and promote the research of reinforcement learning in the field of chess and poker from the optimization of a single algorithm to the evolution of a systematic solution.

DOI: 10.54254/2755-2721/2025.PO24698

2. Grounded theory for reinforcement learning

2.1. Core elements

Reinforcement learning is a machine learning paradigm that achieves goal optimization through the interaction of an intelligent body with its environment [1]. Intelligent bodies perceive states in the environment, perform actions, and adjust their strategies according to the rewards fed back from the environment, with the goal of maximizing long-term cumulative rewards [6-7]. The core lies in learning through "trial and error", seeking a balance between exploration and exploitation, and gradually forming an optimal decision-making strategy [1]. Figure 1 illustrates the reinforcement learning process.

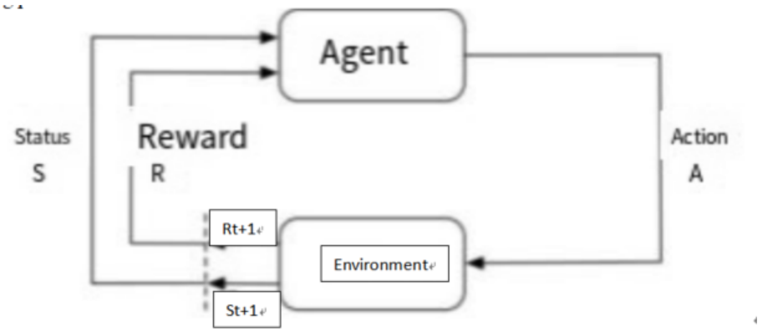

Figure 1: Reinforcement learning process

In a reinforcement learning system, the strategy serves as the core that determines the behavior of the intelligent in a given state; while the reward defines the goal of the whole reinforcement learning problem, which is the scalar value that the environment feeds back to the intelligent within each time step as the main factor influencing the strategy and guiding the intelligent to adjust its behavior. Unlike the immediacy of rewards, the value function provides a longer-term consideration for strategy optimization by expecting to predict long-term future gains. Intelligent bodies acquire states in the environment and output actions and decisions, and the environment will feed back the next state and corresponding rewards based on these decisions. The ultimate goal of intelligent bodies is to optimize their own strategies and obtain as many rewards as possible through continuous interaction with the environment[8,9].

During the interaction process, the intelligent body learns by trial and error through the cycle of "state→action→reward→new state", with the goal of maximizing the accumulated rewards and forming an optimal strategy.

2.2. MDP

Markov decision process, as a fundamental model of reinforcement learning, provides an important theoretical support and algorithm design framework [1].MDP describes the interaction between an intelligent and its environment in terms of a quintuple ⟨S, A, P, R, γ⟩, where the core Markovianity ensures that the state transfer relies only on the current state, a reward function R gives immediate feedback, and the discount factor γ is used to balance short-term and long-term gains, these technical features build up the unique theoretical system of MDP [1]. With standardized modeling capability, MDP supports various algorithms such as dynamic programming, Monte Carlo method and Q-learning to solve the optimal strategy, which has shown great value in practical applications, for example, the card selection process of Collective Card Game can be modeled as MDP, and the

DOI: 10.54254/2755-2721/2025.PO24698

strategy optimization can be realized using deep reinforcement learning [1]. However, MDP also has obvious limitations: high-dimensional state spaces can lead to an exponential increase in computational complexity, as in the case of Go, where the number of states far exceeds the number of atoms in the observable universe; and in real-world scenarios, strict Markovianity is often difficult to satisfy, and approximation assumptions are usually needed to be relied upon for applications [1].

3. Comparison and analysis of typical technologies

In this paper, five typical algorithms, MDP, Q-learning, DQN, PPO, and Actor-Critic, are compared and analyzed technically [1].

The core mechanism of the Q-learning algorithm is offline learning, discrete state-action space, and iterative updating of action value functions through Q-tables [1]. Its advantages include simplicity and stability, data efficiency, and applicability to small-scale discrete scenarios (e.g., chess, backgammon) [1]. However, it also suffers from certain drawbacks such as the state explosion problem (Q-table storage complexity $O(|S|\times|A|)$), the exploration strategy relies on ε-greedy, and it is difficult to converge in high-dimensional states [1].

The technological breakthrough of DQN algorithm is to approximate the Q-function into deep neural network to solve the problem of high-dimensional state (e.g., image input) representation; combined with Experience Replay to reduce the data correlation, and Target Network to stabilize the training process [1]. Its typical applications include AlphaGo evaluating the drop value by DQN, combining with MCTS to realize accurate decision-making; Landlord AI using DQN-IRL framework to deal with the sparse reward problem [3]. However, it also has some limitations such as only applicable to discrete action spaces, continuous action scenarios (e.g., Texas Hold'em betting amount) need to be discretized; target network synchronization delay may lead to training bias [10,11].

The architectural advantage of Actor-Critic is the explicit separation of the strategy function (Actor) and the value function (Critic), where Actor outputs the action probability or determines the action, and Critic evaluates the state/action value to guide the strategy optimization [1]. The strategy is updated in real time without waiting for the end of Episode and the sample efficiency is higher than Monte Carlo method [1]. It also has some practical applications such as asynchronous dominance Actor-Critic (A3C) based to improve the efficiency of parallel training for large-scale gaming scenarios (e.g., StarCraft micro-manipulation); Chess training system based on the asynchronous A3C algorithm, which provides real-time move suggestion and strategy for Alzheimer's disease patients through adaptive cognitive agent analysis [6].

The core innovation of the PPO algorithm is to limit the magnitude of policy updates through importance sampling and truncated objective function (CLIP technique) to avoid “policy collapse”, which combines stability and sample efficiency [1]. It is suitable for discrete/continuous action spaces and performs well in games with incomplete information (e.g., Texas Hold'em AI Libratus) [1]. Its technical details are the estimation of action values using the dominance function A(s,a) [1]. Its application cases include landlord AI to achieve efficient strategy learning under ultra-large action space through PPO combined with action compression coding (mapping 27,000 legal plays into continuous vectors) [4]; and PPO algorithms to directly optimize the action selection strategy through a policy network for the problem of dynamic hidden information and opponent modeling for incomplete-information card games (e.g., poker) [1].

By studying these five typical reinforcement learning algorithms, it can be found that all five algorithms apply to discrete action spaces, but among them, MDP, PPO, and Actor-Critic algorithms

DOI: 10.54254/2755-2721/2025.PO24698

are also applicable to continuous action spaces. Each algorithm has its core idea, advantageous technology, and application scenario (Table 1).

Table 1: Comparison of typical reinforcement learning algorithms

Algorithms	Core Ideas	Action Space	Dominant Technologies	Typical Scenarios	Key Challenges
MDP	State Transfer Modeling	Discrete/ Continuous	Markovianity	Grounded Theory Framework	High-Dimensional Computational Complexity
Q-learning	Value function iteration	Discrete	Offline learning	Small-scale games	State explosion
DQN	Depth-valued function approximation	Discrete	Experience playback, goal networks	Image-input games (Go, Landlord)	Insufficient continuous action processing
PPO	Proximal policy optimization	Discrete/ continuous	Importance sampling, truncation techniques	Complex incomplete information games	Structural complexity
Actor-Critic	Joint optimization of strategy and value function	Discrete/ continuous	Real-time update, variance reduction	Multi-intelligent body confrontation	Overfitting risk

By comparing the process of five typical algorithms in the study, their evolution can be summariz, from the basic modeling of MDP to the discrete scenario solving of Q-learning, to the breakthrough of high-dimensional state limitation by DQN, and finally, the generalization of continuous action space and complex game environment by the Actor-Critic framework and the PPO algorithm, which reflects the technology iteration from “model-dependent” to “data-driven”, “single-intelligence body” to “multi-intelligence body”. This reflects the technology iteration from “model-dependent” to “data-driven” and from “single-intelligence” to “multi-intelligence”.

4. Challenges and future directions

The application of reinforcement learning in board games faces many challenges, with technical bottlenecks being the top priority [1]. First, board games usually involve high-dimensional state and action spaces, such as mahjong's pool combinations and Texas Hold'em's hidden information [1]. This high-dimensional nature makes traditional reinforcement learning algorithms face great challenges in terms of storage and computational resources. For example, Go AI needs to rely on Monte Carlo tree search pruning with neural network dimensionality reduction, while poker AI needs to abstract card power evaluation to reduce state dimensionality [5]. Second, board games often need to deal with collaborative and adversarial relationships between multiple intelligences [1]. In games such as Fight the Landlord, the intelligences need to deal with teammate collaboration and opponent modeling at the same time, and traditional single-intelligence algorithms are difficult to cope with dynamic changes in strategies [1]. Current research approximates the Nash equilibrium by fictional self-pairing games, but real-time strategy coordination still needs a breakthrough [1]. In addition, reward signals in board games are usually sparse and delayed [1]. In many complex games, rewards are only given at the end of the game (e.g., win/loss determination), and the intelligences

DOI: 10.54254/2755-2721/2025.PO24698

cannot obtain useful feedback in time to adjust their behaviors. This property especially poses a serious challenge to value function-based reinforcement learning methods, as the estimation of cumulative rewards is prone to bias. [1]

In the application of AI in board games, the ethical and social impacts should not be ignored. On the one hand, AI can easily crush human players in the game by virtue of its powerful computational ability, which may destroy the entertainment and competitive fairness of the game. Therefore, how to design “restricted AI”, such as simulating human decision-making delays, so that AI and human players can play in a relatively fair environment, has become an important issue that needs to be studied urgently. On the other hand, the existence of AI also brings the risk of cheating, some human players may use the camera to recognize the card face, and use AI to generate the optimal strategy in real time, which has a serious impact on the fairness of the tournament, so it is necessary to develop the anti-cheating technology in order to maintain the order of the board game tournaments [1].

There are several valuable future directions for the application of reinforcement learning in board games. Enhancing generalization capabilities, developing cross-game general algorithms, and reducing reliance on domain-specific knowledge is one of them [1].AlphaZero fully demonstrated the great potential of transfer learning in this regard by simultaneously proficiently playing Go, Xiangqi, and Shogi against itself [5]. Multi-modal fusion will also become an important trend. By combining visual recognition of card faces, using natural language processing to analyze opponents' strategies, and combining it with reinforcement learning, it is expected to create smarter interactive systems, such as AI coaches that can parse player operations in real time and provide strategy suggestions [1]. In addition, the concept of human-computer integration design is gradually emerging, and the focus of research has shifted from “AI defeating humans” to “AI assisting humans”, such as the chess training system designed for Alzheimer's disease patients, which can dynamically adjust the difficulty with the help of AI to help patients improve their cognitive rehabilitation [6]. Chess training system designed for Alzheimer's disease patients can dynamically adjust the difficulty with the help of AI to help patients improve the effect of cognitive rehabilitation [6].

5. Conclusion

This paper focuses on the application of reinforcement learning in board games, and explores the advantages and limitations of reinforcement learning in actual board games by combining the evolution of core algorithms, comparing technical features and analyzing actual cases. The application of reinforcement learning in board games is not only a competition at the level of algorithms, but also a competition of ideas for complex decision-making problems. From the basic framework of Q-learning to the stable optimization of PPO, it always focuses on “how to efficiently deal with high-dimensional, dynamic and adversarial environments”. Although there are still challenges in terms of computational complexity and multi-intelligence synergy, chess AI has already migrated to the direction of education and healthcare, and future reinforcement learning algorithms should also break through the algorithm generalization, real-time, high-dimensional state compression, multi-intelligence synergistic optimization and other key technological bottlenecks, so as to promote the development of reinforcement learning from “special intelligence” to “general intelligence”. “General Intelligence”, with the solution of sequential decision-making problems to solve financial risk control, automatic driving and other scenarios. The research in this paper reveals the theoretical value of reinforcement learning in the intersection of game theory and cognitive science, and provides inspiration for the transformation of artificial intelligence technology to “assisting human beings”.

DOI: 10.54254/2755-2721/2025.PO24698

Authors contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Ronaldo e Silva Vieira, Anderson Rocha Tavares, Luiz Chaimowicz, Exploring reinforcement learning approaches for drafting in collectible card games, Entertainment Computing, Volume 44, 2023, 100526, ISSN 1875-9521, https: //doi.org/10.1016/j.entcom.2022.100526.

[2] Zha D , Xie J , Ma W , et al.DouZero: Mastering DouDizhu with Self-Play Deep Reinforcement Learning [J]. 2021.DOI: 10.48550/arXiv.2106.06135.

[3] Kong, Y., Shi, H., Wu, X. et al. Application of DQN-IRL Framework in Doudizhu's Sparse Reward. Neural Process Lett 55, 9467–9482 (2023). https: //doi.org/10.1007/s11063-023-11209-0

[4] Luo and T. -P. Tan, "RARSMSDou: Master the Game of DouDiZhu With Deep Reinforcement Learning Algorithms, " in IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 8, no. 1, pp. 427-439, Feb. 2024, doi: 10.1109/TETCI.2023.3303251.

[5] Heinrich J , Silver D .Deep Reinforcement Learning from Self-Play in Imperfect-Information Games [J]. 2016.DOI: 10.48550/arXiv.1603.01121.

[6] J. M and R. Surendran, "Chess Game to Improve the Mental Ability of Alzheimer's Patients using A3C, " 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, 2023, pp. 1-6, doi: 10.1109/ICECCT56650.2023.10179809.

[7] Zhechao Wang, Qiming Fu, Jianping Chen, Quan Liu, You Lu, Hongjie Wu, Fuyuan Hu, LinFa-Q: Accurate Q-learning with linear function approximation, Neurocomputing, Volume 611, 2025, 128654, ISSN 0925-2312, https: //doi.org/10.1016/j.neucom.2024.128654.

[8] Wang Z , Fu Q , Lu Y , et al.LinFa-Q: Accurate Q-learning with linear function approximation [J].Neurocomputing, 2025(Jan.1): 611.DOI: 10.1016/j.neucom.2024.128654.

[9] Clark T , Towers M , Evers C , et al.Beyond The Rainbow: High Performance Deep Reinforcement Learning On A Desktop PC [J]. 2024.

[10] Wu L , Wu Q , Zhong H , et al.Mastering "Gongzhu" with Self-play Deep Reinforcement Learning [C]//International Conference on Cognitive Systems and Signal Processing.Springer, Singapore, 2023.DOI: 10.1007/978-981-99-0617-8_11.

[11] Liu Z .Learning Explainable Policy For Playing Blackjack Using Deep Reinforcement Learning (Reinforcement Learning) [J]. 2021.