Deep Reinforcement Learning Research and Applications

Yujiang Qiu

Royal Grammar School Guildford, Nanjing, Jiangsu, China

Abstract. This paper provides an in-depth exploration of Deep Reinforcement Learning (DRL), a powerful combination of reinforcement learning and deep learning techniques, enabling machines to autonomously learn optimal policies through interaction with complex environments. The fundamental concepts of Reinforcement Learning (RL) and Deep Learning are introduced, detailing their core theories and how their integration leads to the advancement of DRL. The paper highlights the significance of DRL in various applications, particularly focusing on path planning as a practical case study, demonstrating its ability to solve high-dimensional and dynamic decision-making problems. Additionally, the current limitations of DRL, including challenges in scalability, sample efficiency, and interpretability, are examined, along with potential solutions and future directions to address these barriers. The conclusion reflects on the growing importance of DRL in a wide range of fields and discusses its potential for future research and real-world implementations. With ongoing advancements, DRL is poised to revolutionize a variety of industries, presenting new opportunities for innovation and efficiency in problem-solving

1 Introduction

Deep Reinforcement Learning (DRL) represents a groundbreaking approach that combines the strengths of Reinforcement Learning (RL) and Deep Learning (DL) to tackle complex decision-making problems. As artificial intelligence (AI) continues to evolve, DRL has emerged as one of the most promising areas of research, demonstrating significant potential across a wide range of applications, including robotics, autonomous vehicles, and game playing. Unlike traditional machine learning techniques, which rely on supervised learning from labeled data, DRL empowers systems to learn optimal behaviors through trial and error, interacting directly with their environment and receiving feedback in the form of rewards or penalties.

The integration of deep learning techniques with reinforcement learning algorithms has enabled the development of models capable of handling high-dimensional, continuous spaces and complex tasks, which were previously intractable. This fusion has made DRL particularly well-suited for tasks that require both perception (e.g., image processing) and

a135635@correo.umm.edu.mx

decision-making, such as path planning in autonomous systems or strategy optimization in dynamic environments.

This paper aims to provide a comprehensive overview of DRL, covering its foundational theories, key advancements. Additionally, the limitations and challenges associated with DRL, including scalability, sample efficiency, and interpretability, are discussed, along with potential directions for future development. By exploring these aspects, the paper seeks to highlight the transformative potential of DRL and its implications for both academic research and real-world applications.

2. Basic Theory

2.1 Reinforcement Learning

Reinforcement learning is a machine learning method that learns strategies by interacting with the environment. The core idea is that agents influence the environment by performing actions and receiving rewards from the environment. The goal of an intelligent agent is to maximize cumulative rewards.

The basic framework of reinforcement learning includes state, action, reward, strategy, and value function, where state refers to the current state of the environment, action refers to the actions

taken by the agent in a certain state, reward refers to the feedback obtained by the agent from the environment after executing the action, strategy refers to the rules for the agent to choose the

action in a certain state, and value function refers to evaluating the long-term value of a state or action. The classic algorithms of reinforcement learning include Q-learning and SARSA to learn optimal strategies by updating Q-values.

2.2 Deep Learning

Deep learning is a subfield of machine learning based on Artificial Neural Networks (ANNs). The core idea is to automatically extract features from data through multi-layer nonlinear transformations (i.e. "deep" networks).

The key components of deep learning include the input layer, hidden layer, and output layer. The input layer refers to the network layer that receives raw data (such as images, text, etc.), the hidden layer refers to the network layer that extracts features through nonlinear activation functions (such as ReLU), and the output layer refers to the network layer that outputs prediction results (such as classification, regression, etc.). The training process of deep learning typically includes Forward propagation and backpropagation optimize network parameters through gradient descent.

2.3 Deep Reinforcement Learning

Deep reinforcement learning is a combination of reinforcement learning and deep learning, using deep neural networks to approximate the value function or strategy function in reinforcement learning. The core idea is to process high-dimensional inputs (such as images and videos) through deep learning and optimize decision strategies through reinforcement learning.

Common algorithms include Deep Q-network (DQN) deep neural network, experience replay, and target network. Deep neural network refers to the model architecture used to approximate Q-value function, experience replay refers to the mechanism of breaking data correlation by storing and randomly sampling historical experience, and target network refers to an auxiliary network independent of the main network for stable Q-value target calculation. DQN has achieved significant results in Atari games through these key technologies.

3 Field of Reference

3.1 Game

In the gaming field, it uses convolutional neural networks to process game graphics, combined with experience replaypools and target network mechanisms, and successfully demonstrates excellent decision-making ability in complex strategy based games such as Starcraft 2 and Dota 2. This algorithm can directly learn gameplay and develop tactical strategies from the original game graphics, reaching the level of professional players in multiple dimensions such as hero control in MOBA games and resource management in RTS games. It provides key technical support for the development of modern electronic sports artificial intelligence and also promotes industrial applications such as game intelligence testing and balance optimization [1, 2].

Luo Furong et al. proposed a reinforcement learning-based multi-agent game method for the game of Three Kingdoms (SGS-MAPG). In the 2v2 game scenario of the game of Three Kingdoms, this method models multiple cooperating agents through the idea of policy gradient, solving the problem of team cooperation and confrontation in the multi-agent environment. Experimental results show that this method enables the agents to acquire good learning and decision-making capabilities after training, and improves the final reward of the team, which is at least 12% higher than the winning rate of the basic algorithm [3].

Liu Yuan proposed an automatic checkers playing system based on deep reinforcement learning. The system combines the strategy value network and the Monte Carlo tree search algorithm to optimize the agent's strategy. Compared with traditional methods, it has stronger adaptability and generalization ability, and can better cope with the uncertainty and high competitiveness in checkers. The method improves the training efficiency and agent performance by inputting the chessboard state into a deep neural network and outputting the chess piece's movement strategy or action value. In addition, the system is trained by playing against itself, and the experimental results show that it has superior performance and good universality [4].

Wan Jie proposed a 3D shooting game control algorithm based on deep reinforcement learning. In order to deal with the problems of sparse and delayed rewards and poor policy convergence in 3D shooting games, this paper designed and implemented a 3D tank shooting game based on Unity, and proposed a game control model based on the proximal policy optimization algorithm (PPO). The model designs a fine-grained reward function through a reward shaping method, and uses an entropy regularization method to adjust the objective function to enhance the exploration of the strategy and avoid early convergence. Experimental results show that the control model performs well in terms of policy convergence effect and learning strategy, and has better performance than the original model and other classic deep reinforcement learning algorithms [5].

3.2 Path Planning

Deep Q-network (DQN) has demonstrated strong decision-making capabilities in the field of path planning. Its core principle is to fit the optimal Q-value function through deep neural networks, combine experience replay mechanisms to store and reuse historical path exploration experience, and use an independent target network to stabilize the training process. In dynamic obstacle scenes, DQN can directly learn the optimal path strategy from sensor inputs (such as LiDAR data), accumulate rewards through continuous trial and error optimization (such as shortest path, successful obstacle avoidance), and ultimately achieve more adaptive real-time path planning than traditional A *, RRT and other algorithms.

Experiments have shown that in tasks such as robot navigation and drone trajectory planning, DQN can not only handle high-dimensional state spaces, but also cope with some observable environmental changes. Its end-to-end learning approach avoids the limitations of artificial feature design and provides a new paradigm for intelligent path decision-making in dynamic uncertain environments [6, 7, 8].

Yi-Bo Zhang and Bing-Peng Gao proposed an AUV path planning algorithm based on deep reinforcement learning, aiming to solve the path planning problem of autonomous underwater vehicles (AUVs) in marine environments, especially when considering the influence of ocean currents. This method improves the D3QN path planning algorithm. First, the state sequence containing ocean current information is extracted through LSTM to capture the change of state over time and the influence of ocean currents on AUV movement; secondly, a comprehensive reward function is designed to consider factors such as path length, obstacle avoidance, number of steps, and using ocean currents to reduce energy consumption; finally, N-step TD is used to update network parameters in order to learn longer-term planning and exploration. Experimental results show that the proposed algorithm shows efficient path planning capabilities in path planning tasks in terms of obstacle avoidance and using ocean currents [9].

Qitao Pan, Yuesheng Zhao and Yuguo Gan proposed a robot path planning method based on an improved Q-Learning algorithm. This method optimizes the problems of the traditional Q-Learning algorithm in path planning, such as slow learning speed, low exploration efficiency and long path. First, based on the grid map, the radial basis function (RBF) network is used to approximate the action value function of the Q-Learning algorithm; second, the greed factor is dynamically adjusted to balance the ratio of exploration and utilization; finally, the robot's selectable actions are expanded and improved to eight-direction exploration. Simulation results show that compared with the traditional Q-Learning algorithm, the improved algorithm can significantly shorten the optimal path length (reduced by 23.33%), reduce the number of inflection points (reduced by 63.16%), and reduce the number of training rounds (reduced by 31.22%) [10].

Yong Jin, Zhengchao Chen, and Huizhen Yang proposed a wireless charging scheduling algorithm for electric vehicles based on deep reinforcement learning. The algorithm aims to solve the problems of high cost, low efficiency, and heavy load on urban power grids encountered in the charging process of electric vehicles, especially in the application of wireless charging technology. Wireless charging of electric vehicles is achieved by laying wireless charging coils on the driving lanes, and the scheduling problem is optimized under the deep reinforcement learning framework to maximize the total remaining power under the deadline and energy constraints. Simulation experiments show that the proposed algorithm is significantly better than the traditional comparison algorithm in multiple performance indicators (such as the number of electric vehicles, the number of charging sections, and the mean deadline), and has better remaining power performance [11].

4 Current Limitations and Future Prospects

Deep Q-Network is a fundamental algorithm in deep reinforcement learning, demonstrating significant success in various applications such as robotics, gaming, and autonomous systems. However, despite its achievements, DQN still faces several challenges that limit its effectiveness in certain scenarios.

At the theoretical level, one major issue is the lack of strict convergence guarantees. Unlike traditional Q-learning, the combination of deep neural networks with reinforcement learning introduces instability, making it difficult to ensure consistent learning outcomes. This instability stems from correlated training samples and the non-stationary nature of the target values, which can cause unpredictable fluctuations in performance.

In practical applications, DQN also suffers from low exploration efficiency. The commonly used ε-greedy strategy is often insufficient in complex environments, leading to poor exploration of the action space and difficulty in discovering optimal policies. This issue is particularly problematic in tasks requiring extensive trial-and-error learning.

Additionally, function approximation errors accumulate during training, introducing biases in Q-value estimation. These errors can propagate over time, leading to overestimation or underestimation of values, ultimately affecting the algorithm’s decision-making accuracy. Addressing these limitations remains crucial for improving the stability and effectiveness of DQN in real-world applications.

On the algorithmic side, DQN's experience replay mechanism, while essential for breaking temporal correlations between training samples, also presents inherent contradictions. The mechanism of storing and randomly sampling experiences can lead to inconsistencies, which may hinder the learning process. Furthermore, in engineering applications, DQN is often limited by high sample complexity and sensitivity to hyperparameters, making it difficult to apply in real-world scenarios where computational resources and data are often constrained.

To address these challenges, future advancements should focus on several key areas.

First, establishing a convergence theoretical framework for deep Q-learning is crucial to ensuring the algorithm's stability and reliability. Without proper guarantees, the algorithm's performance can become unpredictable, especially in complex environments.

Second, improving exploration efficiency is essential. This can be achieved by developing new algorithmic structures that integrate techniques such as Bayesian inference, causal inference, and hierarchical reinforcement learning, which would allow the agent to explore the action space more effectively.

Additionally, addressing the issue of function approximation errors through more robust value function estimation methods could reduce biases and improve the overall accuracy of Q-value estimates.

Finally, developing an efficient parallel training system and embedding domain-specific knowledge into the learning process would increase the scalability and adaptability of DQN, making it better suited for real-world applications with diverse and dynamic environments.

These innovations will not only enhance DQN's efficiency, reliability, and interpretability but also maintain its end-to-end learning advantages, which are key to its versatility in complex tasks. By addressing current limitations, these advancements will ensure that DQN remains a powerful tool for real-world applications, particularly in security-critical fields, where robust and transparent decision-making is paramount. Furthermore, these developments will push the boundaries of deep reinforcement learning theory, facilitating its growth and enabling more advanced applications across various industries. With improved stability, exploration efficiency, and value estimation, DQN's potential will be fully realized, offering new possibilities for dynamic decision-making systems.

5 Conclusions

Deep Reinforcement Learning, which integrates deep learning's perception with reinforcement learning's decision-making, has shown exceptional capabilities in tackling complex tasks. At its core, DRL leverages neural networks to handle high-dimensional state spaces, with key algorithms such as Deep Q-Network utilizing techniques like experience replay and target networks to stabilize training. This combination of deep learning and reinforcement learning allows DRL to excel across a wide range of applications, from robotics to gaming and autonomous systems.

Despite its successes, DRL faces several challenges that hinder its theoretical and practical development. On the theoretical side, there are gaps such as the lack of proven convergence guarantees, and the accumulation of errors in value approximation, which can undermine the effectiveness of the learning process.

Algorithmically, DRL suffers from low sample efficiency, meaning it requires large amounts of data to learn effectively, and it is highly sensitive to hyperparameters, which complicates tuning and limits its scalability. Additionally, engineering constraints like high computational costs present significant challenges for deploying DRL models in real-world applications.

To address these issues, future research should focus on several key areas. First, there is a need to establish mathematical frameworks that provide convergence guarantees, ensuring that DRL algorithms can reliably converge to optimal solutions. Second, integrating techniques such as Bayesian inference or hierarchical reinforcement learning could enhance exploration efficiency, helping the algorithm make better decisions with fewer data points. Lastly, optimizing DRL models for efficiency—by focusing on lightweight architectures and developing parallel training systems—will make them more practical for real-world deployment, especially in resource-constrained environments.

By addressing these challenges, DRL has the potential to drive advancements in secure, interpretable AI applications, particularly in safety-critical fields such as autonomous systems and healthcare. This progress could pave the way for broader adoption of intelligent decision-making technologies, transforming industries and enabling more reliable, autonomous systems.

References

1. B. Tian, L. Lüyao, C. Li, et al., "Game intelligent guidance algorithm based on deep reinforcement learning," J. Jilin Univ. Sci. Ed., vol. 63, no. 1, pp. 91–98, 2025. DOI: 10.13413/j.cnki.jdxblxb.2023555.
2. H. Bian, "Research on game lineup character recommendation method based on deep reinforcement learning," M.S. thesis, Qilu University of Technology, 2024. DOI: 10.27278/d.cnki.gsdqc.2024.000083.
3. F. Luo, Y. Wang, J. Qin, et al., "A reinforcement learning-based multi-agent game method for the Three Kingdoms Kill," Comput. Simulation, vol. 41, no. 7, pp. 484–490, 2024.
4. Y. Liu, "Research on automatic international checkers playing system based on deep reinforcement learning," M.S. thesis, North University of China, 2024. DOI: 10.27470/d.cnki.ghbgc.2024.001438.
5. J. Wan, "Research on 3D shooting game control algorithm based on deep reinforcement learning," M.S. thesis, Hubei University, 2024. DOI: 10.27130/d.cnki.ghubu.2024.001789.

6. L. Dong, S. Xiao, S. Niu, et al., "Cooperative path planning under constraints based on deep reinforcement learning," Control Decision, vol. 1, pp. 1–9, Feb. 22, 2025. DOI: 10.13195/j.kzyjc.2024.1019.
7. Y. Jin, Z. Chen, and H. Yang, "Path planning for electric vehicles aimed at wireless charging based on deep reinforcement learning," Autom. Appl., vol. 66, no. 2, pp. 72–75, 2025. DOI: 10.19769/j.zdhy.2025.02.019.
8. Y. Hu, Y. Qu, W. Li, et al., "A Pareto Front searching algorithm based on reinforcement learning for constrained multiobjective optimization," Inf. Sci., vol. 705, p. 121985, 2025.
9. Y. Zhang and B. Gao, "AUV path planning research based on deep reinforcement learning," J. Northeast Normal Univ. (Nat. Sci. Ed.), vol. 57, no. 1, pp. 53–62, 2025. DOI: 10.16163/j.cnki.dslkxb202312260002.
10. Q. Pan, Y. Zhao, and Y. Gan, "Robot path planning based on improved Q-learning algorithm," Internet of Things Technol., vol. 15, no. 3, pp. 82–86, 2025. DOI: 10.16667/j.issn.2095-1302.2025.03.021.
11. Y. Jin, Z. Chen, and H. Yang, "Path planning for electric vehicles aimed at wireless charging based on deep reinforcement learning," Autom. Appl., vol. 66, no. 2, pp. 72–75, 2025. DOI: 10.19769/j.zdhy.2025.02.019.