Applications for Reinforcement Learning in Robotics: A Comprehensive Review

Zicheng Shang*

Department of Electrical and Electronic Engineering, University of Liverpool, Liverpool, United Kingdom

*Corresponding author: Zicheng.Shang20@student.xjtlu.edu.cn

Abstract. Reinforcement learning has become a promising and effective method to solve robots' complex control problems, enabling agents to learn optimal behaviors independently in a dynamic, uncertain, and usually high-dimensional environment. This paper comprehensively reviews the latest progress in applying reinforcement learning technology in robotics, paying special attention to key areas such as path planning, dynamic obstacle avoidance, multi-robot cooperation, and human-computer interaction. By critically analyzing various methods based on reinforcement learning, this paper emphasizes the key progress in improving robot autonomy, decision-making, and adaptability. It also discusses the challenges of deploying reinforcement learning in real-world robot applications, including sample efficiency, security, scalability, and generalization to new environments. In particular, Deep Reinforcement Learning (DRL) shows great potential in enhancing the robot's ability, especially in a complex, unstructured environment. However, challenges such as high computing costs and long training time still need to be addressed. Finally, the paper suggests that future research should focus on the hybrid RL method, improve learning efficiency, combine domain knowledge, and combine RL with other advanced technologies such as computer vision, multi-agent system, and real-time feedback mechanism, to further expand its potential application and increase its influence in the field of robotics.

Keywords: Reinforcement Learning; Robotics; Path Planning; Multi-Robot Collaboration; Human-Robot Interaction.

1. Introduction

In recent years, significant progress has been made in robotics. The progress of artificial intelligence (AI) and machine learning has completely changed the design and function of autonomous systems. Robotics covers a variety of applications, from industrial automation and medical robots to self-driving cars and cooperative robots (cobot). People are constantly pushing to build autonomous systems that can interact with dynamic environments and complete complex tasks, which leads to growing interest in reinforcement learning (RL) as a tool to enhance robot decision-making [1].

RL is a machine-learning paradigm in which agents learn to make decisions by interacting with the environment. Compared with supervised learning (models learn from tagged data), RL allows agents to learn from their own experiences, take actions, and receive feedback through rewards or punishments [2]. This trial-and-error process makes RL especially suitable for tasks involving sequential decision-making, in which the agent's behavior will affect the future state, and the best solution is not immediately apparent. Unlike traditional methods, RL can only partially understand the environmental model, which is especially beneficial in dynamic, uncertain, and complex environments, such as those encountered in robotics. The application of reinforcement learning in robotics is of great significance. First, reinforcement learning enables robots to adapt to the environment and learn optimal behaviors without being explicitly programmed to understand which behavior is better. It is crucial for robots operating in unknown or dynamically changing environments to have self-learning capabilities because predefined control systems or traditional algorithms cannot provide reliable solutions in such situations[3]. In terms of path planning technology, traditional path planning algorithms such as A* or Dijkstra rely on static map slam modeling and predefined algorithms. The difference is that reinforcement learning methods train robots through reward mechanisms, enabling them to find the optimal route in a dynamic environment and adjust the real-time path according to changing conditions [4].

Reinforcement learning has achieved many significant results in robotic systems recently. Nie et al. proposed a new path planning method for automated guided vehicles based on RL. In the RL model, Nie et al. introduced sample regularization technology and combined it with adaptive learning rate [5]. In addition, Chen et al. focused on real-time path planning and dynamic obstacle avoidance, using deep reinforcement learning on the control system to enable the robot to respond to dynamic changes in the surrounding environment with minimal computing resources [6]. These latest research results show the key role of reinforcement learning in the navigation and path planning of robotic systems.

Reinforcement learning also plays a vital role in multi-robot systems. The rapid development of automated factories and assembly lines has made factories no longer satisfied with the work of a single robot. As a result, engineers have turned their needs to multi-robot systems [7]. In scenarios where multiple robots need to collaborate to complete tasks, reinforcement learning can optimize the coordination and task allocation between agents, thereby improving the efficiency of task completion. In this context, Li et al. studied the application of deep reinforcement learning in multi-robot collaboration, focusing on path planning and task allocation. Their research showed that reinforcement learning can help coordinate robots' activities in complex scenarios and effectively assign tasks to appropriate robots, which is crucial for minimizing completion time and avoiding collisions [8].

Although reinforcement learning has significantly progressed in robotics recently, it has brought some challenges. One of the main problems is the low sample efficiency of reinforcement learning algorithms. Robot models must use training data and interact with simulated environments to learn effective policies. This large amount of simulation is particularly problematic in robotics because testing in the real world is time-consuming and expensive.

This paper comprehensively reviews the application of reinforcement learning in robotics, especially in path planning, task allocation, multi-robot collaboration, human-robot interaction, and dynamic obstacle avoidance. The purpose is to summarize and explore the recent progress in this field, focusing on the latest developments and breakthroughs and, in this way, identify the remaining challenges and potential future directions. In the following sections, this paper will discuss the different reinforcement learning techniques used in robotic systems, analyze their impact on the field, and discuss future challenges and opportunities.

2. Methods

2.1. Reinforcement Learning for Path Planning

This section reviews different approaches to using reinforcement learning algorithms to optimize path planning and dynamic obstacle avoidance tasks, especially in AGV systems. Reinforcement learning effectively optimizes the path planning process by training intelligent agents to make effective decisions in dynamic and complex environments and ensuring that each choice is the best choice through a reward mechanism [9]. This section focuses on the improved PPO algorithm proposed by Nie et al. The algorithm they proposed introduces two parameters, sample regularization (SR) and adaptive learning rate (ALR), which can overcome the problems of insufficient exploration and slow convergence faced by traditional reinforcement learning methods in AGV path planning.

2.1.1 Algorithm Foundation

The algorithm proposed by Nie et al. is based on the Markov decision process (MDP) framework, which defines the relationship between the state of the environment, action selection, and reward feedback.

A Markov decision process consists of the following four elements: State, Action, Transition Probability and Reward. The characteristics of MDP are following:

The current state st fully describes the historical information of the system. The future state is only related to the current state and action, and has nothing to do with the past state and action：

$$P(s_{t+1} \mid s_t, a_t, s_{t-1}, a_{t-1}, \dots) = P(s_{t+1} \mid s_t, a_t) \tag{1}$$

Maximize cumulative rewards (rewards):

$$Gt = \sum \gamma^k R(s_{t+k}, a_{t+k}, s_{t+k+1}) \tag{2}$$

In MDP, the agent selects actions and adjusts its strategy based on feedback (rewards) from the environment to influence the state of the environment. The core goal of reinforcement learning is to maximize the cumulative reward to guide the agent to adopt the optimal strategy. Common RL methods include policy gradient method (PG) and actor-critic algorithm (AC).

2.1.2 Design of the New Algorithm

Nie et al. proposed an improved PPO algorithm. The PPO algorithm is an important algorithm in reinforcement learning. Its principle is to optimize the strategy by maximizing the objective function. The algorithm uses a clipped objective function to make the update more stable, so that it can efficiently handle continuous action spaces and large-scale problems. Based on this theory, Nie et al. combined sample regularization (SR) and adaptive learning rate (ALR), which solved the problems of insufficient exploration ability and slow convergence in AGV path planning.

2.2. Multi-Robot Collaboration Using RL

Multi-robot systems mainly involve multiple robots working together to achieve the target requirements in a coordinated manner, such as covering an area or collaboratively transporting large items. A seemingly simple system requires a very high level of intelligence to implement, and the system needs to assign tasks based on the position of each robot. As a machine learning theory, reinforcement learning can help effectively coordinate these robots and optimize task allocation, communication, and movement. In recent years, Li et al. proposed a DRL-based multi-robot collaborative path planning and task allocation method. The authors demonstrated that their method allows multiple robots to work efficiently without collisions while minimizing the task completion time.

On the other hand, Le et al. took a different approach and applied DRL to the complete coverage planning of reconfigurable robots. Their research focused on reconfigurable robots based on multi-rhombuses, which can adjust their shapes according to the environment to ensure entire area coverage when working in complex environments.

3. Result and Analysis

In this section, the article mainly analyzes the results of different methods under two reinforcement learning applications and analyzes

3.1. Path Planning and Dynamic Obstacle Avoidance

Nie et al. adopted the PPO algorithm in reinforcement learning. They combined it with sample regularization and adaptive learning rate to enhance the stability and efficiency of the path planning model.

3.1.1 Model Review

Proximal Policy Optimization: PPO is a policy gradient method for optimizing policies in reinforcement learning. The algorithm updates the policy by maximizing a clipped objective to avoid excessive updates, thereby improving the stability of learning. Compared with the traditional TRPO (Trust Region Policy Optimization) method, PPO has higher computational efficiency and can maintain stable performance in more complex environments.

Sample Regularization: SR enhances the exploration ability of the strategy by adjusting the probability distribution of samples. In reinforcement learning, the balance between exploration and exploitation is particularly important. Over-reliance on exploitation may cause the model to converge to the local optimal solution too early. Sample regularization avoids this problem by increasing the diversity of samples, thereby improving the performance of the agent in complex environments.

Adaptive Learning Rate: The ALR mechanism dynamically adjusts the learning rate according to the changes in the current strategy. In reinforcement learning, a learning rate that is too large will cause the strategy to update too quickly, which will cause training instability or oscillation; while a learning rate that is too small may lead to slow convergence. By combining KL divergence and Fisher information matrix, the ALR mechanism can intelligently adjust the learning rate so that the model can converge quickly and stably in a dynamic environment.

3.1.2 Experimental results

Nie et al. experimentally verified that the improved PPO algorithm has higher completion and accuracy in dynamic obstacle avoidance tasks. The experimental results show that the PPO model combined with sample regularization and adaptive learning rate can continue to avoid obstacles and complete tasks faster when facing dynamic obstacles and changing environments, and its path also follows the optimal principle. Compared with traditional methods, the task completion time and path planning success rate have been significantly improved, which also shows that the PPO model can better adapt to the dynamic changes of obstacles and avoid collisions.

In complex dynamic environments, the optimized PPO model can avoid obstacles by adjusting the path in real time and can find the optimal path in a shorter time. In addition, the performance of this method in different scenarios also shows good stability and efficiency, especially in changing dynamic obstacle environments, showing a high task completion rate. For the model itself, the optimized new model improves learning efficiency and convergence speed while ensuring model stability and adaptability. The experimental results verify the advantages of this method in dynamic obstacle avoidance and path planning, and demonstrate its potential in practical applications.

3.2. Multi-Robot Systems

In a multi-robot system (MRS), reinforcement learning can significantly improve the efficiency of inter-robot collaboration and better achieve task allocation. Li et al. and Le et al. showed that applying reinforcement learning methods can effectively manage task allocation in multi-robot systems, thereby reducing task completion time and improving resource utilization. Le et al. proposed the DRL-MPC-GNNs model, demonstrating its unique advantages and performance in multi-robot collaboration.

3.2.1 Model analysis and advantages

Application of reinforcement learning: First, Li's paper's deep reinforcement learning component allows each robot to interact with the environment and learn the optimal decision-making strategy in this interactive training. Traditional multi-robot task allocation methods usually rely on rules or optimization algorithms, which means that these methods have environmental limitations, cannot perform well in complex or dynamic environments, and lack flexibility and adaptability. In contrast, deep reinforcement learning can significantly improve the efficiency of task completion by continuously allowing the robot system to learn and adapt and more efficiently adjust between task requirements and environmental changes.

Introduction of model predictive control (MPC): The MPC component can predict the system's future state in real time, provide feedback on the current behavior based on the expected state, and generate control decisions to ensure the robot can consider environmental changes and task complexity when performing tasks. In a multi-robot system, MPC can handle operational constraints during path planning and task allocation, ensuring that the robot's movement and execution are within the physical constraints while maximizing efficiency.

Cooperative enhancement of graph neural networks (GNNs): Robot coordination and information exchange are crucial in a multi-robot system. The GNN method can capture the complex interaction information between robots and optimize the collaborative efficiency of the entire robot network by aggregating the information of adjacent nodes. This method allows the robot system to make corresponding decisions based on its status and the impact of other robots on task completion, thereby achieving more efficient team collaboration.

Comprehensive advantages: Integrating DRL, MPC and GNNs together, the DRL-MPC-GNNs model effectively combines the advantages of policy learning, real-time control, and inter-robot collaboration, solving the problem of multi-robot collaboration that traditional single methods cannot handle at the same time. challenge. Through reinforcement learning, the model can achieve dynamic task adaptation and autonomous decision-making; through model predictive control, it can make real-time path adjustments in complex environments; through graph neural networks, it can enhance information sharing and collaboration between robots and ensure the coordination of the entire system. and efficiency.

3.2.2 Experimental results and analysis

In experiments, the DRL-MPC-GNNs model was validated on multiple datasets and compared with existing traditional methods such as rule-based optimization methods. Experimental results show that the new model shows significant advantages in terms of task completion time, path planning accuracy and task allocation efficiency.

Reduced task completion time: Through DRL's autonomous learning and MPC's real-time optimization, the robot can quickly adapt in a dynamic environment and complete tasks efficiently. Compared with traditional methods, the task completion time is reduced by approximately 12%.

Improved path planning accuracy: The introduction of MPC makes the robot's path planning in a dynamic environment more accurate, avoiding path conflicts and non-optimal paths in traditional methods, and the path planning accuracy increases by about 15%.

Improved task allocation efficiency: GNNs improve the efficiency of task allocation by optimizing collaboration between robots. When multiple robots collaborate, GNNs can ensure that information is efficiently shared among robots, thereby reducing conflicts and delays in task allocation and improving task allocation efficiency by more than 10%.

In general, the DRL-MPC-GNNs model combines the advantages of reinforcement learning, model predictive control and graph neural networks, and can effectively improve the collaboration efficiency, task allocation capabilities and path planning accuracy of multi-robot systems. Through experimental verification, the model shows great potential in practical applications and provides new solutions for the optimization of multi-robot collaboration tasks.

4. Challenges and Future Directions

Although research in recent years has shown that reinforcement learning has shown great potential in the field of robotics, this does not mean that it is mature enough for widespread application. The application of reinforcement learning in the field of robotics still faces some challenges and difficulties. One of the main challenges is that the sample efficiency of reinforcement learning algorithms is relatively low, which means that the model must interact with the environment many times before it can converge to the optimal solution. Long-term interactions make the training process very time-consuming and require a lot of computing resources. To address this problem, people need to establish an open-source database to meet the data required for model training, and explore technologies such as transfer learning and imitation learning to use prior knowledge to accelerate the learning process.

Another major challenge is that reinforcement learning algorithms lack sufficient scalability in multi-robot systems. As the number of robots increases, the computational difficulty of multi-robot systems does not increase gradually, but exponentially, which means that the model cannot cope with overly complex task allocation and coordination. Therefore, it is particularly important to develop more scalable reinforcement learning methods to manage the growing robot team.

In addition, safety and explainability are also essential reference factors for reinforcement learning applications. Engineers must ensure that robots trained by reinforcement learning are in safe operation and do not cause any harm to humans, especially in human-intensive environments [10]. In this regard, future research should explore combining reinforcement learning with formal verification techniques to ensure safety. In addition, improving the interpretability of RL models can help diagnose and understand robot behavior, and only by making it understandable to more people can people have more trust in robotic systems in daily life.

5. Conclusion

Reinforcement learning has become a powerful tool for achieving advanced robotic behaviors and multi-machine system collaboration, especially in complex and dynamic environments that are difficult to handle with traditional methods. This paper discusses various applications of RL in robotics, focusing on areas such as navigation path planning, dynamic obstacle avoidance, and multi-robot collaboration. The research of many papers shows the great potential of RL in enhancing the adaptability, efficiency, and safety of robotic systems.

Although some applications of reinforcement learning have become mature, there are still many challenges related to sample efficiency, scalability, and safety. How to deal with and solve these challenges has become the focus of the development of reinforcement learning in the future, which requires the combination of new algorithms, hybrid methods, and interdisciplinary collaboration. As RL methods continue to develop, this paper expects that their combination with robotics will create more intelligent, autonomous, and collaborative robots that can operate in complex and dynamic environments. For the application of RL, future research should focus on improving learning efficiency, ensuring safe operation, and expanding the applicability of RL to a more diverse range of robotic tasks.

References

[1] Hao J et al., Exploration in Deep Reinforcement Learning: From Single-Agent to Multi-Agent Domain, arXiv preprint arXiv:2109.06686, Feb. 2023.

[2] Li Y, Deep reinforcement learning: An overview, arXiv preprint arXiv:1701.07274, 2017.

[3] Placed J A, Castellanos J A, A deep reinforcement learning approach for active SLAM, Applied Sciences, 2020. https://doi.org/10.3390/app10093364

[4] Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, Riedmiller M, Playing Atari with deep reinforcement learning, 2013.

[5] Nie J, Zhang G, Lu X, Wang H, Sheng C, Sun L, Reinforcement learning method based on sample regularization and adaptive learning rate for AGV path planning, Neurocomputing, 2024, 128820. https://doi.org/10.1016/j.neucom.2024.128820

[6] Chen P, Pei J, Lu W, Li M, A deep reinforcement learning-based method for real-time path planning and dynamic obstacle avoidance, Neurocomputing, 2022, 497: 64–75. https://doi.org/10.1016/j.neucom.2022.05.006

[7] Madridano A, Al-Kaff A, Martín D, De La Escalera A, Trajectory planning for multi-robot systems: Methods and applications, Expert Syst. Appl., 2021, 173: 114660.

[8] Li Z, Shi N, Zhao L, Zhang M, Deep reinforcement learning path planning and task allocation for multi-robot collaboration, Alexandria Engineering Journal, 2024, 109: 408–423. https://doi.org/10.1016/j.aej.2024.08.102

[9] Ye F, Zhang S, Wang P, Chan C Y, A survey of deep reinforcement learning algorithms for motion planning and control of autonomous vehicles, in 2021 IEEE Intelligent Vehicles Symposium (IV), Nagoya, Japan, 2021: 1073–1080. https://doi.org/10.1109/IV48863.2021.9575880

[10] Sarıkaya B S, Bahtiyar Ş, A survey on security of UAV and deep reinforcement learning, Ad Hoc Networks, 2024, 164: 103642. https://doi.org/10.1016/j.adhoc.2024.103642