A Comprehensive Review of Reinforcement Learning for Multi-Robot Task Allocation in Logistics

Zhai Shuo
School of Information, Beijing Wuzi University, Beijing, China

Abstract—This review explores the application of reinforcement learning in multi-robot task allocation within the logistics field. With the advancement of technology, intelligent robots are increasingly being utilized across various industries, particularly in complex tasks and diverse environments where multi-robot systems exhibit significant advantages over single-robot systems. Task allocation is a critical component in multi-robot systems. This paper introduces the necessity of multi-robot task allocation, its applications in different domains, and the challenges faced. Traditional algorithms such as linear programming, heuristic search, and swarm intelligence each have their strengths and weaknesses, but they show limitations in dynamic and complex environments. Reinforcement learning, due to its self-learning capability and interaction with the environment, has become a research hotspot. Through continuous exploration and feedback adjustment, reinforcement learning algorithms in multi-robot task allocation gradually approach optimal solutions, demonstrating great potential in this field. Future research in reinforcement learning should focus on lifelong learning algorithms to enhance the efficiency of multi-robot task allocation systems in complex and dynamic logistics and warehousing environments.

Keywords—*Multi-Robot Systems, Task Allocation, Deep Reinforcement Learning*

I. INTRODUCTION

With the rapid development of the social economy, the use of intelligent robots is gradually permeating various aspects of people's lives, bringing significant changes to our daily life and work. As technology advances, intelligent robots are increasingly applied across various fields, driving future development. From automated production lines in manufacturing to precision surgery in the medical field, and to efficient delivery and intelligent warehousing in logistics, these applications of intelligent robots are diversely transforming our lives.

A. The Emergence of Multi-Robot Systems

As tasks become increasingly complex and varied, single robots are finding it challenging to cope with more intricate engineering projects, leading to the emergence of multi-robot systems. The application of multi-robot systems is spearheading new trends in intelligence and automation. Multi-robot systems are no longer isolated individual robots but are coordinated teams that can collectively accomplish more complex and diverse tasks. This trend is evident across various industries, from manufacturing and emergency response to agricultural production and space exploration[1].

B. The Necessity of Multi-Robot Task Allocation

The application potential of robot systems is limitless, with task allocation being a critical component. Multi-robot task allocation refers to assigning multiple robots or drones to specific tasks or sets of tasks in an optimal or suboptimal manner. This task allocation can be static (where the attributes of tasks and robots do not change over time) or dynamic (where the attributes of tasks and robots may change over time). The goals of multi-robot task allocation typically include maximizing task completion efficiency, minimizing overall costs, minimizing task completion time, or achieving other specific performance metrics. In different industries, tasks can be highly varied[2].

C. Multi-Robot Task Allocation in Various Industries

In the industrial manufacturing sector, multi-robot task allocation has become critical for enhancing production efficiency and quality. Teams of robots in factories can work collaboratively to automatically complete tasks such as assembly, processing, and inspection, thereby reducing human errors and labor intensity while increasing production line efficiency. This not only saves time and resources but also creates more competitive advantages for enterprises[3].

In the logistics and warehousing sector, multi-robot task allocation is improving supply chain management and order processing. Robots in automated warehouses can efficiently transport and store goods, automatically selecting the optimal paths according to order requirements, reducing traffic congestion and delays. This translates to faster delivery speeds, lower operational costs, and more efficient logistics services[4][5].

Additionally, multi-robot task allocation is being applied in search and rescue, medical surgery, agriculture, and environmental monitoring. In emergency rescues, drones can work alongside ground robots to quickly search disaster areas, locate trapped individuals, and provide rescuers with critical information about the disaster zone[6]. In agriculture and agricultural automation, multi-robots can be used for tasks such as planting, harvesting, and weeding[7].

D. Challenges in Multi-Robot Task Allocation

Multi-robot task allocation is a complex issue because it requires consideration of the characteristics of the robots, the nature of the tasks, environmental conditions, and various constraints. In practical applications, the cost of tasks is affected by the accuracy and timeliness of the multi-robot task allocation algorithms. Furthermore, the efficiency of production and the safety of personnel can also be impacted. Therefore, multi-robot task allocation algorithms have high research value and broad application prospects. In recent years, with the continuous development and improvement of robotics technology and intelligence, many experts and scholars at home and abroad have conducted extensive research on this issue[8]. They have been discovering new algorithm models and optimizing task allocation strategies to continuously improve the collaborative capabilities of multi-robot systems.

Multi-robot task allocation problems are classified as NP-hard problems. In complex and dynamic environments, finding the optimal task allocation solution requires searching through a vast solution space, which is difficult to accomplish within a limited time frame. This characteristic makes multi-robot task allocation a challenging research area. To address this issue, experts and scholars worldwide have proposed various algorithms. The contract net protocol simulates the bidding mechanism in a market economy to achieve task allocation among robots[9]; the fireworks algorithm draws inspiration from the randomness of fireworks explosions to search for optimal solutions in the solution space[10]; distributed algorithms enhance the response speed of multi-robot systems through decentralized computation and decision-making[11]; the responsibility domain algorithm simplifies the task allocation process by assigning specific responsibility areas to each robot[12]; and genetic algorithms simulate the mechanisms of genetic inheritance and mutation in biological evolution to continuously optimize solutions through iterations[13].

However, these algorithms have certain limitations in solving robot task allocation problems. Specifically, genetic algorithms often result in local optimal solutions due to insufficient exploration of the solution space.

E. Reinforcement Learning for Multi-Robot Task Allocation

Reinforcement Learning (RL) is an important research method in machine learning and has been successfully applied to various fields of multi-robot task allocation. By continuously exploring the environment and using feedback from the environment to update, reinforcement learning algorithms promote convergence to optimal solutions[14][15]. Compared to other algorithms, reinforcement learning excels in multi-robot task allocation due to its ability to interact with the environment and adapt to complex and dynamic settings. Previous research has shown that reinforcement learning has already achieved significant results in the field of multi-robot task allocation[16][17].

In summary, the application of reinforcement learning in multi-robot warehousing scheduling not only enhances efficiency and reduces costs but also improves the system's intelligence and adaptability. This enables it to effectively cope with complex and dynamic warehousing environments, meeting the evolving demands of logistics. Consequently, reinforcement learning becomes an indispensable tool in the warehousing industry.

II. MULTI-ROBOT TASK ALLOCATION METHODS

In recent years, the rapid development of science and technology has driven many researchers, both domestically and internationally, to delve into the field of multi-robot task allocation. This research direction has evolved from initial theoretical exploration to practical applications in the real world, presenting a diverse landscape. Meanwhile, the continuous advancements and innovations in swarm intelligence algorithms, evolutionary algorithms, and machine learning algorithms have introduced new development directions for multi-robot task allocation algorithms. The multi-robot task allocation problem (MRTA) can be specifically divided into the following aspects: linear programming-based MRTA methods, heuristic search-based MRTA methods, swarm intelligence-based MRTA methods, and reinforcement learning-based MRTA methods.

A. Linear Programming-Based MRTA Algorithms

Linear programming-based MRTA algorithms are classic approaches to handling combinatorial optimization problems. These methods abstract complex real-world problems into matrix problems, providing an effective pathway for solving various combinatorial optimization challenges. By using linear programming algorithms, we can transform the complexity of multi-robot task planning into a form of mathematical computation. In this traditional domain, classic solutions include the Hungarian algorithm and mixed-integer linear programming (MILP) algorithms. The study in [18] uses a mixed-integer linear programming formula to optimize human-multi-robot task allocation problems. By minimizing the total execution time, this approach optimizes task allocation, task quality, and the workload of both humans and robots, allowing for online monitoring and task redistribution. This further extends to solve human-multi-robot task allocation problems in the multi-robot task allocation domain. Smriti Chopra from the Georgia Institute of Technology's School of Electrical and Computer Engineering proposed a Hungarian algorithm to address multi-robot task allocation problems. In this algorithm, each robot runs a local program to handle specific sub-steps of the Hungarian algorithm and exchanges solution estimates with neighboring robots. Through a finite number of local computations and communications, all robots eventually converge to a common optimal task allocation scheme. This algorithm can achieve convergence within a limited time or within a finite number of communication rounds during synchronization [19].

However, as the complexity of the problem increases, using linear programming methods to solve multi-robot task allocation problems leads to a significant increase in computation time and does not guarantee the optimal solution, thus presenting a dual challenge: computational efficiency and result quality.

B. Heuristic Search-Based MRTA Algorithms

Heuristic search-based MRTA algorithms use accumulated system experience to guide subsequent algorithm choices, effectively enhancing spatial exploration efficiency and significantly reducing algorithm convergence time. The core process of heuristic search methods typically starts from a set of randomly generated feasible solutions and then uses information from these feasible solutions to guide the iterative process of the algorithm, gradually approaching the optimal solution. Compared to linear programming methods, heuristic searches can find a near-global optimal solution at an acceptable computational cost when handling large-scale combinatorial optimization problems, and this approximate solution is usually feasible in practical applications.

The study in [20] introduced an innovative method based on simulated annealing to coordinate the attack actions of drone swarms. This method successfully overcame premature convergence issues in particle swarm optimization algorithms, achieving optimal resource allocation for the drone swarm. To further enhance the global search capability of the algorithm and quickly reach the optimal solution, researcher Li Zhenping and colleagues applied a tabu search algorithm in the distribution vehicle domain. By applying a variable neighborhood tabu search algorithm to improve the initial solution, they successfully obtained a near-optimal solution [21]. Artificial immune algorithms can continuously search for the global optimal solution without requiring advanced conditions such as target function differentiability. In [22], a method was proposed to solve resource allocation problems in variable power distribution systems using artificial immune algorithms. From a practical application perspective, heuristic search-based multi-robot task allocation methods can generally meet the needs of various application scenarios. Researchers worldwide are dedicated to finding ways to make the task allocation results obtained by heuristic searches closer to the optimal solution.

However, heuristic searches do not rely on strict mathematical formulas to guarantee finding the global optimal solution; they can only quickly discover a set of locally optimal solutions. In complex real-world environments, these locally optimal solutions are often sufficient to approximate the global optimal solution.

C. Swarm Intelligence-Based MRTA Methods

Swarm intelligence-based MRTA methods are also a type of heuristic search. The concept of swarm intelligence algorithms is inspired by various group behaviors in the biological world, simulating the information transfer methods within biological populations to find the optimal allocation schemes for the group. The study in [23] proposed a multi-UAV (FIAM) task allocation algorithm inspired by fish behavior, simulating behaviors such as fish aggregation, leader following, and forgetting, to achieve autonomous and adaptive task allocation, particularly suitable for search and rescue missions. Compared to other multi-robot task allocation methods, FIAM has the advantages of adaptability and autonomous decision-making, enabling it to execute search and rescue tasks in hostile environments. The study compared FIAM with traditional methods through simulated search and rescue missions, showing that FIAM significantly outperformed in survivor rescue rates, task completion times, and operational times.

Particle swarm optimization (PSO) algorithms guide particles towards the system's optimal solution by leveraging the best solutions within the group, social experiences, and individual experiences, offering the advantages of easy parameter settings and simple operation. In [24], researchers proposed an improved particle swarm optimization algorithm successfully applied to spatial crowdsourcing task allocation problems, thereby improving task allocation efficiency. In recent years, experts worldwide have gradually drawn inspiration from biological group behaviors to develop swarm intelligence algorithms suitable for various application scenarios. These algorithms provide innovative solutions, particularly in the field of multi-robot task allocation, compensating for the shortcomings of traditional technologies and achieving significant results in practical applications.

However, swarm intelligence algorithms still face some issues, such as premature convergence and getting stuck in local optima, which require further research and improvement.

III. REINFORCEMENT LEARNING-BASED MRTA METHODS

A. Overview of Reinforcement Learning

Reinforcement Learning (RL) is a machine learning method aimed at optimizing an agent's decision-making behavior through interactions with the environment. Its core idea stems from behavioral psychology, particularly operant conditioning theory, which simulates how organisms adapt to their environment and obtain rewards through trial-and-error learning. The goal of reinforcement learning is to continuously explore and utilize environmental feedback to find the optimal policy that maximizes cumulative rewards.

In reinforcement learning, the agent selects an action at each time step (t) based on the current state . After executing the action, the environment transitions to the next state and provides the agent with a reward . The agent's task is to continuously explore and learn to find a policy ( ) that maximizes the cumulative reward over the long term. This process can be described through the following main components:

State (State): A specific description of the environment, usually represented as a feature vector.

Action (Action): Operations that the agent can take in the environment.

Reward (Reward): Feedback signal from the environment for the agent's action, usually a scalar value used to evaluate the quality of the action.

Policy (Policy): The rule by which the agent selects actions based on states, represented as , the probability of selecting action ( ) in state ( ).

Value Function (Value Function): Used to assess the quality of a state or state-action pair. Common value functions include the state value function and the action value function ( ).

The primary goal of reinforcement learning is to learn the optimal policy ( ) that allows the agent to maximize the expected cumulative reward from any initial state:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$$

where is the discount factor , used to measure the importance of future rewards.

In recent years, with the rapid development of computing power and deep learning technology, reinforcement learning has achieved remarkable results in various fields. Particularly, Deep Reinforcement Learning (DRL), which combines deep neural networks with reinforcement learning algorithms, has enabled efficient learning in complex environments. DRL has demonstrated strong performance and potential in numerous application scenarios, including game AI, autonomous driving, and robot control.

In the domain of multi-robot task allocation, the strength of reinforcement learning lies in its robust self-learning and adaptability, which enables the continuous optimization of task allocation strategies in dynamic and uncertain environments. Through ongoing interactions with the environment, reinforcement learning algorithms can effectively address varying task requirements and environmental conditions, progressively converging to the global optimal solution.

The reinforcement learning-based MRTA (Multi-Robot Task Allocation) approach represents a significant research direction in the field, primarily utilizing real-time interactions and error-detection mechanisms between agents and the environment to achieve the maximum cumulative reward, thereby training the algorithm's optimal action decision sequence.

To delve deeper into how reinforcement learning can be applied in multi-robot task allocation, it is essential first to gain a thorough understanding of reinforcement learning and to explore the extensive research and discussions conducted by scholars both domestically and internationally. Reinforcement learning is a type of machine learning methodology designed to enable intelligent systems to learn and improve their behavior through interaction with the environment, aiming to achieve specific goals or maximize cumulative rewards. Inspired by theories of behavioral psychology, this field seeks to emulate how humans and animals acquire new knowledge and skills through trial-and-error learning. A key feature of reinforcement learning is the agent's interaction with the environment, where actions are taken and then environmental feedback is observed, typically in the form of reward signals. The objective for the agent is to find an optimal strategy, a series of actions, to maximize cumulative rewards. To achieve this goal, reinforcement learning algorithms commonly use value functions to assess the value of actions in different states, guiding the agent's decision-making process.

B. Development of Multi-Agent Deep Reinforcement Learning

Over time, reinforcement learning has achieved notable success across various fields, and multi-agent deep reinforcement learning has rapidly evolved, addressing complex real-world problems. However, as noted in the literature [25], current multi-agent deep reinforcement learning still faces several challenges. One major issue is complexity; training a large number of agents is inherently difficult. The presence of each agent in the environment adds complexity to the learning problem. As the number of agents increases, the scalability of individual learning models decreases due to the computational demands posed by the potential combinations.

Since breakthroughs in deep learning methodologies, the field has undergone rapid transformation, gradually making previously unsolvable problems manageable. Studies [26][27] introduced a cooperative deep reinforcement learning task allocation method based on deep Q-networks. This method employs deep reinforcement learning algorithms, enabling collaborative agents to act autonomously and learn how to communicate with other nearby agents to allocate tasks and share resources. Through this learning capability, agents can conveniently reason, devise appropriate strategies, and make sound decisions.

C. Application of Reinforcement Learning in Logistics and Warehousing

A study [28] addressing the problem of multi-robot task allocation in warehouse settings proposed a solution based on reinforcement learning, the RTAW. This research framed the MRTA problem within the warehouse as a reinforcement learning challenge, designing the state space, action space, and reward functions as part of a Markov decision process.

Another piece of literature [29] introduced a new multi-robot task allocation and decentralized navigation solution, the DC-MRTA. By employing a dual-layer computational approach, this method integrates low-level collision-free decentralized navigation with high-level reinforcement learning-based task allocation, transforming the decentralized multi-robot task allocation challenge into a Markov decision process tackled through reinforcement learning strategies. This approach can be integrated with decentralized multi-agent path finding (MAPF) methods.

D. Deep Reinforcement Learning Algorithms

The literature [30] introduced a Shared Experience Actor-Critic (SEAC) deep reinforcement learning algorithm, which updates the actor and critic parameters of agents by combining gradients calculated from the agents' experiences and weighing these gradients with experiences from other agents. This shared experience approach allows for more effective learning among agents. Agents can learn from each other's experiences without needing to have the same reward functions. Ultimately, SEAC outperformed independent learning, shared strategy training, and advanced Multi-Agent Reinforcement Learning (MARL) algorithms in four different environments [31].

The Q-learning algorithm is a classical reinforcement learning method designed to maximize cumulative rewards by addressing Markov decision processes, thereby generating the best decision sequence. As the field of reinforcement learning has evolved, Q-learning has been widely applied in high-tech areas, including task scheduling, system resource allocation, and gaming [32]. Q-learning has demonstrated exceptional performance in solving multi-robot task allocation problems, despite these problems differing from traditional task planning issues [33]. To successfully extend the Q-learning algorithm from single-robot to multi-robot issues, literature [34] proposed a new method called Team-Q. The core idea of this algorithm is to transform individual robot action selection into collaborative action selection among multiple robots, thus enabling the Q-learning algorithm to effectively solve multi-robot problems. Furthermore, literature [35] introduced the Distributed-Q algorithm to address the challenges of vast state-action spaces in multi-robot collaborative task allocation. This method focuses on handling states and actions in a distributed manner, enhancing the efficiency and scalability of multi-robot systems.

E. Methods of Hierarchical Reinforcement Learning

Implementing Hierarchical Reinforcement Learning (HRL) primarily involves several different algorithmic frameworks. The first is the options-based HRL algorithm, which decomposes tasks into manageable subtasks by defining options that include policies, termination conditions, and initial conditions, allowing for selection and switching among different options [36]. The second approach is based on the hierarchical abstract machine HRL algorithm, which uses a hierarchical abstract machine to structure tasks at different levels, achieving the gradual refinement and execution of tasks through decision-making and learning across these levels [37]. The third type is the value function decomposition HRL algorithm, which decomposes the global value function into local value functions, allowing each subtask to learn and optimize independently within the local environment, thus effectively executing the overall task [38]. The fourth type is the end-to-end HRL algorithm, which learns directly from raw inputs to final outputs through end-to-end training, thereby implementing layered learning and execution of tasks [39]. Literature [40] introduced a hierarchically stable Multi-Agent Deep Reinforcement Learning (MADRL) algorithm. In the hierarchical learning component, a two-layer policy model is used to reduce the complexity of the solution space and train these two interlinked strategies using an interleaved learning paradigm.

However, in traditional reinforcement learning (RL) settings, it is usually assumed that agents operate in a static environment, meaning the dynamics and rewards of the environment are fixed. Yet, in more realistic scenarios, this stability assumption seldom holds.

CONCLUSION

In summary, multi-robot task allocation as a complex and significant research area has attracted considerable attention and study from scholars. From traditional linear programming methods to heuristic searches, swarm intelligence algorithms, and the recently popularized reinforcement learning approaches, each algorithm has its unique strengths and limitations. Linear programming excels in small-scale problems but becomes computationally complex as the scale of the problem increases. Heuristic search methods, despite their efficiency in handling large-scale issues, still struggle with the problem of local optima. Swarm intelligence algorithms, by simulating natural group behaviors, provide new solutions for multi-robot task allocation but also need to overcome issues like premature convergence.

Reinforcement learning methods, due to their self-learning capabilities and advantages in interacting with the environment, have gradually become a research hotspot in the field of multi-robot task allocation. Through continuous exploration and feedback adjustment, reinforcement learning can progressively approximate the optimal solution in dynamic and complex environments. However, the existing reinforcement learning algorithms still face challenges in computational complexity and training efficiency when applied in multi-agent systems, necessitating further research and optimization. The next direction for research in multi-robot task allocation using reinforcement learning should focus on the application of lifelong reinforcement learning, which can enhance algorithmic efficiency and reduce costs in this domain.

Looking forward, as technological advancements continue and computing power increases, multi-robot task allocation algorithms will become more efficient and intelligent. Developing hybrid task allocation strategies that combine the advantages of various algorithms could be an effective way to solve large-scale complex problems. Additionally, with the widespread adoption of emerging technologies such as the Internet of Things (IoT) and 5G, the real-time communication and collaborative capabilities of multi-robot systems will be enhanced, broadening the prospects for the application of multi-robot task allocation algorithms.

References

[1] 顾大强, 郑文钢. 多移动机器人协同搬运技术综述[J]. 智能系统学报, 2019, 14(1): 20-27. (*references*)

[2] Guerrero, J.; Miñana, J.-J.; Valero, O.; Oliver, G. Indistinguishability Operators Applied to TaskAllocation Problems in Multi-Agent Systems. Appl. Sci. 2017, 7, 963. https://doi.org/10.3390/app7100963

[3] Chakraa H, Guérin F, Leclercq E, et al. Optimization techniques for Multi-Robot Task Allocation problems: Review on the state-of-the-art[J]. Robotics and Autonomous Systems, 2023: 104492.

[4] P. R. Wurman, R. D’Andrea, M. Mountz, Coordinating hundreds of cooperative, autonomous vehicles in warehouses, AI Magazine 29 (1)(2008) 9. doi:10.1609/aimag.v29i1.2082.

[5] K. Yakovlev, A. Andreychuk, T. Rybecký, M. Kulich, On the Application of Safe-Interval Path Planning to a Variant of the Pickup and Delivery Problem:, in: Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics, 2020, pp.521–528. doi:10.5220/0009888905210528.

[6] B. Mishra, D. Garg, P. Narang, V. Mishra, Drone-surveillance for search and rescue in natural disaster, Computer Communications 156 (2020) 1–10.

[7] C. Ju, J. Kim, J. Seol, H. I. Son, A review on multirobot systems in agriculture, Computers and Electronics in Agriculture 202 (2022) 107336. doi:https://doi.org/10.1016/j.compag.2022.107336.

[8] Z. -Y. Zhang, Y. -F. Chen, D. Xu, et al. Research on Multi-robot Task Allocation Algorithm Based on HADTQL// 2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI). IEEE, 2020: 164-169

[9] 付光远, 李源, 付文宇,等. 改进合同网在多机器人围捕任务分配中的应用[J]. 兵器装备工程学报, 2019, 40(03):98-102.

[10] 马焱, 赵捍东, 张玮,等. 基于自适应烟花算法的多无人机任务分配[J]. 电光与控制, 2018, 025(001):37-43.

[11] Cui R , Guo J , Gao B . Game theory-based negotiation for multiple robots task allocation[J]. Robotica, 2013, 31(6): 923-934.

[12] Elango M , Nachiappan S , Tiwari M K . Balancing task allocation in multi-robot systems using K-means clustering and auction based mechanisms[J]. Expert Systems with Applications, 2011, 38(6): 6486-6491.

[13] Janakiraman N , Kumar P N . Multi-objective hardware/software partitioning technique for dynamic and partial reconfigurable system-on-chip using genetic algorithm[J]. International Journal of Engineering & Technology, 2014, 6(2).

[14] Arulkumaran K, Deisenroth M P, Brundage M, et al. Deep reinforcement learning: A brief survey[J]. IEEE Signal Processing Magazine, 2017, 34(6): 26-38.

[15] 张子迎,宫思远,徐东,等. 多机器人任务分配与路径规划算法[J].哈尔滨工程大学学报,2019,40(10):1753-1759.

[16] Noureddine D B, Gharbi A, Ahmed S B. Multi-agent Deep Reinforcement Learning for Task Allocation in Dynamic Environment[C]//ICSOFT. 2017: 17-26.

[17] 秦智慧,李宁,刘晓彤,刘秀磊,佟强,刘旭红.无模型强化学习研究综述[J].计算机科学, 2021,48(03):180-187

[18] Lippi M, Marino A. A mixed-integer linear programming formulation for human multi-robot task allocation[C]//2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN). IEEE, 2021: 1017-1023.

[19] Chopra S, Notarstefano G, Rice M, et al. A distributed version of the hungarian method for multirobot assignment[J]. IEEE Transactions on Robotics, 2017, 33(4): 932-947.

[20] 石岭. 基于改进的模拟退火 PSO 无人机资源分配[D].南京航空航天大学,2015.

[21] 李珍萍,杨光,韩倩倩.考虑工作量均衡的成品油二次配送车辆路径问题 [J/OL]. 系统仿真学报 :1-14[2021-03-08].http://kns.cnki.net/kcms/detail/11.3092.V.20210306.1302. 005.html.

[22] Souza S S F , Romero R , Pereira J , et al. Artificial immune algorithm applied to distribution system reconfiguration with variable demand[J]. International Journal of Electrical Power & Energy Systems, 2016, 82(nov.): 561-568.

[23] Alhaqbani A, Kurdi H, Youcef-Toumi K. Fish-inspired task allocation algorithm for multiple unmanned aerial vehicles in search and rescue missions[J]. Remote Sensing, 2020, 13(1): 27.

[24] 张辉,赵晨曦,王杨,等.基于改进粒子群算法的空间众包任务分配模型[J].计算机应用研究,2020,37(09):2698-2700+2705.

[25] Gronauer S, Diepold K. Multi-agent deep reinforcement learning: a survey[J]. Artificial Intelligence Review, 2022: 1-49.

[26] Noureddine D B, Gharbi A, Ahmed S B. Multi-agent Deep Reinforcement Learning for Task Allocation in Dynamic Environment[C]//ICSOFT. 2017: 17-26.

[27] Sukhbaatar, S., Szlam, A., and Fergus, R. (2016). Learning multiagent communication with backpropagation.In 29th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

[28] Agrawal A, Bedi A S, Manocha D. Rtaw: An attention inspired reinforcement learning method for multi-robot task allocation in warehouse environments[C]//2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023: 1393-1399.

[29] [34]Agrawal A, Hariharan S, Bedi A S, et al. DC-MRTA: Decentralized Multi-Robot Task Allocation and Navigation in Complex Environments[C]//2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2022: 11711-11718.

[30] Christianos F, Schäfer L, Albrecht S. Shared experience actor-critic for multi-agent reinforcement learning[J]. Advances in neural information processing systems, 2020, 33: 10707-10717.

[31] Ming Tan. “Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents”. In: Machine Learning Proceedings. Elsevier, 1993, pp. 330–337.

[32] 胡华,张强,胡海洋,等.基于 Q-learning 的移动群智感知任务分配算法[J].计算机集成制造系统,2018,24(07):1774-1783.

[33] 郑延斌,樊文鑫,韩梦云,等.基于博弈论及 Q 学习的多 Agent 协作追捕算法[J].计算机应用,2020,40(06):1613-1620.

[34] Ghazanfari B , Mozayani N . Enhancing Nash Q-learning and Team Q-learning mechanisms by using bottlenecks[J]. Journal of Intelligent & Fuzzy Systems, 2014, 26(6): 2771-2783.

[35] Huang J, Yang B, Liu D. A distributed q-learning algorithm for multi-agent team coordination[C]//2005 International Conference on Machine Learning and Cybernetics. IEEE, 2005, 1: 108-113.

[36] Richard S. Sutton, Doina Precup, Satinder Singh. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. 1999, 112(1- 2): 181-211.

[37] Parr R E . Hierarchical control and learning for markov decision processes[M]. University of California, Berkeley, 1998.

[38] Dietterich T G . Hierarchical Reinforcement Learning with the MAXQ Value Function Decomposition[J]. Journal of Artificial Intelligence Research, 2000, 13: 227-303.

[39] Menache I , Mannor S , Shimkin N . Q-Cut - Dynamic Discovery of Sub-Goals in Reinforcement Learning[J]. 2002: 295-306.

[40] Jin Y, Wei S, Yuan J, et al. Hierarchical and stable multiagent reinforcement learning for cooperative navigation control[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021.