Collaborative Strategy of Multi-Agent Deep Reinforcement Learning for Complex Tasks

Hang Yin

{hyin0027@student.monash.edu}

Monash University, Melbourne, Australia

Abstract: The rapid development of intelligent systems has made the research on collaborative optimization among multiple agents particularly important in multiple application fields such as multi-robot collaboration, intelligent transportation, and resource allocation. Traditionally, rule-based or model-based approaches have been difficult to achieve good scalability and adaptability in a constantly changing environment. In such situations, multi-agent reinforcement learning (MARL) can learn how to behave in collaboration by constantly interacting with their surroundings. This paper reviews basic frameworks for MARL like centralized learning, decentralized learning and centralized training with decentralized execution. It also looks into difficulties like environmental changes, working together well, and making good training. In summary, this paper gives directions for future work on coordination mechanism and communication strategy, and also makes MARL architecture can be used more widely in future application.

Keywords: Multi-agent Systems, Reinforcement Learning, Multi-agent Reinforcement Learning

1 Introduction

1.1 Background

Multi-agent system (MAS) has become an important research area in the field of intelligent control and collaboration decision making since it can coordinate many autonomous agents to achieve difficult jobs. It is like having many robots work together just like the swarm of flying drones, all the way down until each one makes an improvement by improving how its own agent speaks and communicate with others [1, 2]. However, when it is dynamic and ambiguous, each robot generally can only base its decisions on partial information from its own observations. At the same time, they also influence the entire situation. There is such a relationship linking single robot choice and global goal, which greatly increases how difficult it is and requires much better systems to learn things and work together.

ICIAAI 2026, May 22-24, Singapore, Singapore

DOI 10.4108/eai.22-5-2026.2365328

1.2 Research about multi-agent system

Although multi-agent systems have a lot of application areas in control cooperative systems and intelligent decision-making system, it also has many practical problems. For the traditional way to deal with many agents is basically relying on handmade rules, centralized coordination means. When the systems get more complicated, these tend to scale poorly and rely heavily upon good models of the environment [2]. So, the methods based on learning that the agent could have learnt about such a scenario instead of relying more upon pre-defined rules and correct systems are need. Reinforcement learning is an online approach which allows agents to learn the best way of acting in environment via interactions, and it has been increasingly applied for such problems.

In recent years, with the increasing introduction of reinforcement learning in the field of multi-agent, multi-agent reinforcement learning (MARL) has begun to take shape. MARL enables many agents to learn to interact both internally and with the environment. MARL is very useful in problems involving more agent working together on constantly changing things. Moreover, since the decisions of many agents are constantly changing the state of the world, they will learn simultaneously. Therefore, the non-stationary learning situation of each individual agent is obtained, which makes learning extremely difficult: previous work has also well recognized that the non-stationary nature often leads to unstable training and failure to converge. In addition, cooperative tasks can also lead to the issue of credit allocation to identify which one contributes the most to the common goal. This problem makes it even more difficult to establish a good cooperative learning approach [3-5].

1.3 Research motivation and planning

This paper reviews and analyzes the multi-agent collaborative optimization methods based on deep reinforcement learning, with a focus on sorting out the typical frameworks, key technical features of multi-agent reinforcement learning and their application ideas in collaborative decision-making. Then, the advantages and limitations of relevant methods in terms of coordination mechanisms and system stability is analysed. The motivation of this paper is giving directions for future work on coordination mechanism and communication strategy.

The structure of this paper is arranged as follows: The second part introduces the theoretical background and related research of this paper. The third part discusses the methodological framework and coordination mechanism of this paper. The fourth part analyzes the application and future development direction of this paper. Finally, a summary is made.

2 Theoretical foundation of multi-agent controlling

2.1 Multi-agent system

MAS are made up of many autonomous agents, each can make decisions. These agents interact, cooperate, and perform talks, work together and do stuff either for themselves or as a team. The general controlling and interaction loop of a MAS is shown in Fig. 1 [2]. The MAS is better than single agent system in dealing with large, structured and coupled problem. So, they are used in the area of robot cooperation, unmanned aerial vehicle formation flight, intelligent scheduling, and so on.

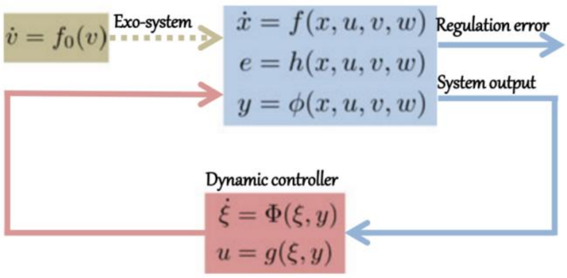

Fig. 1. The General control and interaction model for a multi-agent system [2].

2.2 Deep reinforcement learning

Reinforcement learning is a type of learning which learns by interacting with an environment and then makes decision based on rewards, it is applicable for sequential decisions and controls problems. Reinforcement learning is where the goal of agent is to get the most rewards from all time [3]:

$$J = E_{x_0 \sim p_0, x_{t+1} \sim P, u_t \sim \Pi}\left[\sum_{t=0}^{\infty} \gamma^t R\left(x_t, u_t, x_{t+1}\right)\right] \tag{1}$$

Where γ stands for the discount factor and $R(x_t, u_t, x_{t+1})$ stands for reward function. Deep Reinforcement Learning (DRL) can be widely seen in single-agent problems. The training of DRL still requires a great deal of interaction data for training and is rather sensitive towards certain hyperparameters as well. As it goes on to use in many agents, the interaction of strategies between different groups will make an even less stable learning atmosphere, which requires greater stability when training as well as better cooperation. And these are the problems that will become the problems to focus on as deep reinforcement learning is being used for multi-agent systems.

From the point of view of DRL, learning is done by constantly interacting with environment and other agent by receiving reward signal so as to maximize return over time during learning. But in reality, agent rely on local information when it comes to systems being based global environment. In multi-agent, DRL is different from single-agent where local decisions maybe completely different to the overall goal, which can add up drastically for models and learning.

2.3 Multi-agent reinforcement learning

MARL applies reinforcement learning to more than one agent that is simultaneously learning and acting. Compared to simple-agents reinforcement learning, DRL is a bigger problem because agents can interact and use various different tactics while MARL is happening. So there arise some problem like non-stationary environment and also cooperative credit assignment, these become main problem in MAMR research.

Problem formulation. Due to the high complexity of multi-agent systems, collaborative optimization problems are usually difficult to solve by traditional analytical modeling or precise model-based methods. Policy-based methods directly learn parameterized policies that map observed states to actions, making them applicable to high-dimensional decision-making and continuous action spaces [3]. In practice, they tend to be used along with the actor-critic formulation. Actor based on the current policy produces action, and critic evaluate if it is a good policy and give feedback for updating the policy. This method can help reduce the variance of policy gradient method and also make training more stable [4]. But policy-based multi-agent still have some problems. When many agent’s policies are updated at once, that makes it harder for the training because things aren’t stationary. Also, agents mostly rely on their own observations while being carried out therefore do not have much grasp about the whole picture. Therefore, how to make use of global information without centralization is still an important topic on multiagent collaborative optimization.

Cooperative learning framework. In mult-agent reinforcement learning the choice of what is learned has a strong influence on whether one can get a stable, scalable and cooperatively efficient system. The collaborative learning structure can be divided in to three categories: Centralized Learning, Totally Decentralize Learnining, Centralized Training and Decentralized Execution (CTDE).

The centralized learning method uses global state info in training by having one unified model make decision for all the agents. It could be helpful to mitigate some degree of the non-stationary problem. But as the quantity of the agent grows, it brings about serious problems of computation and scale. Thus, the centralized learning method is hard to apply in a huge system.

Compared to fully decentralized, where every agent learns its own policy without sharing it with anyone and decides what to do without global information. From this standpoint, it will do well in the handing out of work. But without global information support, decentralization will have poor collaboration and instability in converging on tasks that need a lot of cooperation.

To maintain stability when the system scale expands, CTDE is also popular in multi-agent reinforcement learning. Under this framework, agents can use the global knowledge known during training, but can only utilize their own local observations at runtime. It alleviates the instability problem during training while retaining the advantages of distributed decision-making.

Coordination mechanism. Multi-agent cooperation is about how to optimize the collaborative efforts of different agents inside it. The main areas covered are things like rewards, contributions, and sharing information with other agents. Reward design is very important to make agents work together. They usually get a total bonus on the whole system in virtually all circumstances. But if only global rewards is used, it is hard for one agent to know how effective there actions are. Good systems must be built for evaluating individual contributions, it should help improve agent's cooperation. The evaluating individual contributions also help agent in making them better coordinate as they exchange useful observation and features [6]. However, too much information exchaging will make things more complicated and costier. Therefore, good coordination methods must weigh the benefits of cooperation against system scalability and training stability.

3. Research status of multi-agent deep reinforcement learning

In the past few years, people have begun to use more real-world multi-agent learning for optimization. Because it can learn to collaborate by interacting with dynamic surroundings, MARL is very promising in a number of fields [7]. And the most typical direction is multi robot cooperation, intelligent transportation, and resource allocation and schedule optimization.

3.1 MultiRobot cooperation

Multi-robot systems are an important application scenario of multi-agent reinforcement learning. As shown in Fig. 2, the general real-world scenarios of multi-robots include cooperative navigation, formation control and so on [8]. They require at least two or more such machines to work together. When compared with using rules，multi-agent reinforcement learning lets robots keep on acting and discovering by engaging constantly not only with the things in its surrounding environment but also with all other robots [9].

At present, many studies apply MARL methods to different aspects of multi-robot systems, such as cooperative path planning (CPP), collision avoidance (CA), and distributed task allocation (DTA) [8]. Robot will adopt mutual teach and cooperative learn to gather some usefully division-of-labor tactic plans, can adapt their actions and cooperate with the rest if there's incomplete information. And it could let the learning way improve being flexile and adapt to multi robots, then face a scenario where it is difficult to make accurate predictions.

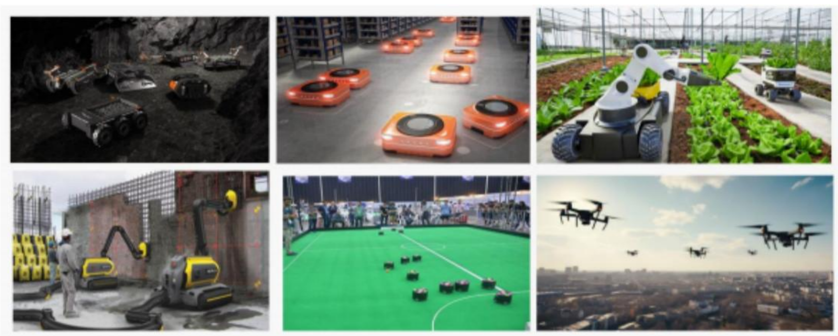

Fig. 2. Real world application of multi-robot system [8].

In addition, MARL has also been applied in industrial production and smart factory environments [10].

3.2 Intelligence transportation interact decision

Another important application field of MARL is the intelligence of transportation and automation drive [11]. In traffic system, traffic lights etc interacting in complex ways creating lots of decision-making scenarios that are also very coupled and uncertain. Traffic control or rule-based decision-making in traditional ways have difficulty coping with such things. Multi-agent reinforcement learning provides a promising approach to model and optimize these interactions. In scenarios such as vehicle following, intersection management, fleet coordination, and traffic signal control, each vehicle or control unit can be modeled as an autonomous agent that learns decision strategies through interaction with the environment and other agents. By considering system-level objectives such as traffic efficiency, safety, and congestion reduction, multi-agent learning frameworks can improve overall transportation performance [12].
Recent research has applied MARL methods to cooperative driving strategies, adaptive traffic signal control, and autonomous vehicle coordination at intersections. These approaches allow traffic participants to learn cooperative behaviors and optimize global traffic efficiency while maintaining safety constraints. However, due to the strong interaction among agents and the dynamic nature of traffic environments, ensuring training stability and reliable decision policies remains a key challenge in this research area [13]. Fig. 3 shows a representative MARL-based traffic signal control framework [12].

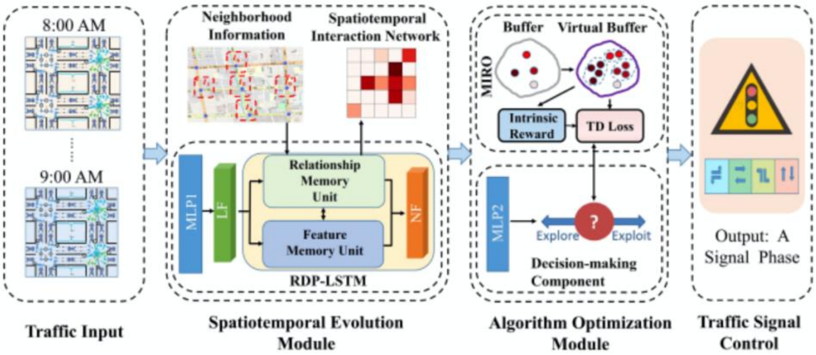

Fig. 3. A MARL-based traffic signal control framework [12].

3.3 Optimize resources distribute and planning

Multi-agent reinforcement learning has also been widely studied in resource allocation and scheduling optimization problems. In many real-world systems, such as communication networks, energy management systems, and cloud computing platforms. Multiple decision-making entities must coordinate their actions to achieve efficient utilization of limited resources. In these scenarios, each agent may represent a server, device or subsystem, and they make local decisions while affecting the overall system performance. As shown in Fig. 4, recent studies have explored MARL-based approaches for dynamic resource allocation, distributed scheduling, and load balancing in large-scale systems [14]. Compared with traditional optimization methods, learning-based optimization methods are more adaptable to dynamic workloads, uncertain environments and constantly changing system conditions. MARL enables these agents to learn adaptive allocation strategies through repeated interactions with the environment. Through the optimization of long-term cumulative rewards, the system can gradually formulate efficient scheduling strategies to balance resource utilization, system throughput and operational stability. Therefore, MARL demonstrates strong potential in enhancing the flexibility and scalability of complex resource management systems [15].

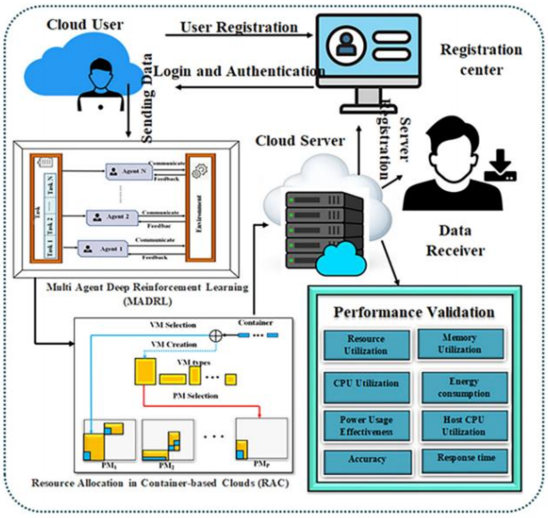

Fig. 4. A MARL-based resource allocation framework [14].

4. Challenges and future directions

4.1 Environment non-stationary

Environmental non-stationarity is one of the fundamental challenges in MARL. As multiple agents learn and update their policies simultaneously, the environment from the perspective of each agent is constantly changing. This dynamic nature leads to unstable training processes and makes policy convergence significantly more difficult. Addressing non-stationarity remains a key issue for improving learning stability in MARL systems.

4.2 Collaboration efficiency

Collaboration efficiency is another key issue, especially in large-scale multi-agent systems. In many real-world scenarios, agents can only access a limited number of local observations, which restricts their ability to coordinate effectively. Insufficient information sharing will reduce cooperation efficiency, while excessive communication will increase system complexity and computational overhead. Therefore, designing an effective communication and coordination mechanism is crucial for enhancing the overall performance of the system.

4.3 Training stability and scalability

As the number of agents increases, training stability and scalability remain the main issues. If there is a more complex system with more interactions among agents, the learning process will be more challenging. Developing scalable and stable algorithms in multiple environments is also an important research topic in the future.

In the days to come, research on more effective coordination mechanisms, adaptive communication strategies, and scalable learning architectures is expected to further enhance the applicability of multi-agent reinforcement learning in complex real-world systems.

5 Conclusion

In conclusion, this paper reviews the multi-agent collaborative optimization methods based on deep reinforcement learning. The theoretical basis, learning framework and coordination mechanism of MARL are reviewed, and its advantages in complex dynamic environments are emphasized. The representative applications such as multi-robot systems, intelligent transportation and resource allocation show that MARL has great potential in the real world. However, in large-scale systems, there exist problems such as unstable environment, poor cooperation efficiency, and scale issues. Future work will require more attention to perfect coordination, better communication, and the creation of a broader learning environment, so that MARL can be used in harsher real-life settings. This paper gives directions for future work on coordination mechanism and communication strategy, and also makes MARL architecture can be used more widely in future application.

References

[1] Ma, L.: Consensus control of stochastic multi-agent systems: a survey. Science China Information Sciences, Vol. 60, pp. 120201 (2017)

[2] Qin, J., Ma, Q., Shi, Y.: Recent advances in consensus of multi-agent systems: a brief survey. IEEE Transactions on Industrial Electronics, Vol. 64, pp. 7776972 (2017)

[3] Gronauer, S., Diepold, K.: Multi-agent deep reinforcement learning: a survey. Artificial Intelligence Review, Vol. 55, pp. 895–943 (2022)

[4] Oroojlooy, A., Hajinezhad, D.: A review of cooperative multi-agent deep reinforcement learning. Applied Intelligence, Vol. 53, pp. 13677–13722 (2023)

[5] Du, W., Ding, S.: A survey on multi-agent deep reinforcement learning: from the perspective of challenges and applications. Artificial Intelligence Review, Vol. 54, pp. 3215–3238 (2021)

[6] Zhu, C., Dastani, M., Wang, S.: A survey of multi-agent deep reinforcement learning with communication. Autonomous Agents and Multi-Agent Systems, Vol. 38, pp. 4 (2024)

[7] Canese, L., Cardarilli, G. C., Di Nunzio, L., Fazzolari, R., Giardino, D., Re, M., Spanò, S.: Multi-agent reinforcement learning: a review of challenges and applications. Applied Sciences, Vol. 11, pp. 4948 (2021)

[8] Arjun, K., Parlevliet, D., Wang, H., Yazdani, A.: Optimizing coalition formation strategies for scalable multi-robot task allocation: a comprehensive survey of methods and mechanisms. Robotics, Vol. 14, pp. 93 (2025)

[9] Orr, J., Dutta, A.: Multi-agent deep reinforcement learning for multi-robot applications: a survey. Sensors, Vol. 23, pp. 3625 (2023)
[10] Bahrpeyma, F., Reichelt, D.: A review of the applications of multi-agent reinforcement learning in smart factories. Frontiers in Robotics and AI, Vol. 9, pp. 1027340 (2022)
[11] Dinneweth, J., Boubezoul, A., Mandiau, R., Espié, S.: Multi-agent reinforcement learning for autonomous vehicles: a survey. Autonomous Intelligent Systems, Vol. 2, pp. 27 (2022)
[12] Chen, W., Yang, S., Li, W., Hu, Y., Liu, X., Gao, Y.: Learning multi-intersection traffic signal control via coevolutionary multi-agent reinforcement learning. IEEE Transactions on Intelligent Transportation Systems, Vol. 25, pp. 15947–15963 (2024)
[13] Chu, T., Wang, J., Codeca, L., Li, Z.: Multi-agent deep reinforcement learning for large-scale traffic signal control. IEEE Transactions on Intelligent Transportation Systems, Vol. 21, pp. 8667868 (2020)
[14] Cui, J., Liu, Y., Nallanathan, A.: Multi-agent reinforcement learning-based resource allocation for UAV networks. IEEE Transactions on Wireless Communications, Vol. 19, pp. 8807386 (2020)
[15] Nagarajan, S., Rani, P. S., Vinmathi, M. S., Subba Reddy, V., Saleth, A. L. M., Abdus Subhahan, D.: Multi agent deep reinforcement learning for resource allocation in container-based clouds environments. Expert Systems, Vol. 42, pp. 13362 (2025)