ITM Web of Conferences **84**, 04004 (2026) https://doi.org/10.1051/itmconf/20268404004
ATCIDS 2026

The Cooperation and Competition Mechanisms In Multi-agent Reinforcement Learning

Puxuan Li[1*], *Shuo* Li[2], *Junxi* Wang[3]

[1] School of Mathematics and Physics, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China
[2] School of Advanced Technology, Xi' an Jiaotong-Liverpool University, Suzhou, Jiangsu, China
[3] School of Mathematics and Statistics, Central South University, Changsha, Hunan, China

Abstract. With the development of artificial intelligence, multi-agent systems have become an important research hotspot for achieving autonomous multi-agent collaboration and intelligent decision-making. Among them, multi-agent reinforcement learning (MARL) has become a key framework for solving uncertain and dynamic interaction problems. This article systematically reviews the cooperation and competition mechanisms in MARL. The article points out that centralized training and decentralized execution (CTDE) is the core paradigm for solving the problems of credit allocation and environmental instability in cooperation, and has derived key technologies such as value decomposition and actor-critic. For competition, the combination of game theory and deep reinforcement learning provides a theoretical basis for strategic interaction. Additionally, the article analyzes the complexity of mixed cooperative-competitive scenarios, summarizes a comprehensive framework integrating multiple technologies, and demonstrates its application potential through cases such as games and smart grids. Finally, in response to current bottlenecks, it looks forward to future directions such as combining with large models and improving generalization ability.

1 Introduction

With the rapid development of intelligent Artificial Intelligence (AI) in recent years, its applications in daily life have become increasingly widespread. This surge in demand has driven advancements in related technologies. Early artificial intelligence often operated in isolation, but people live in a world composed of complex systems where many problems cannot be solved by a single intelligent agent. As a result, multi-agent collaboration has emerged as a solution. Since optimal solutions are often difficult to identify and exhaustive search methods are highly inefficient, algorithms have continuously evolved. To clarify the current state of research in this field, the team has written this review to systematically organize and integrate the research findings in this area.

However, multi-agent reinforcement learning (MARL) collaboration and competition mechanisms hit key bottlenecks. In non-stationary environments—like sudden obstacles

* Corresponding author: Puxuan.li22@student.xjtlu.edu.cn

during robot collaboration—joint policy training loses stability. When agent numbers fluctuate (say, adding or removing robots), behavioral combinations get too complex to compute.

Offline learning misses collaboration temporal patterns (such as robot coordination sequences) and struggles to adapt to tasks, which limits cross-scenario use [1]. High computational needs hold back real-world deployment (e.g., autonomous driving), and no unified benchmarks make research harder to reproduce [2]. Unresolved credit assignment also slows collaborative optimization.

This study will present a systematic review of the subject matter, focusing on the core methods and technical evolution of cooperation mechanisms, the core methods and technical evolution of competition mechanisms, as well as mixed cooperation-competition mechanisms and their application cases.

To fix these, future research has several directions. It can pair large language models with cMALC-D to boost sample efficiency and ease nonstationary training instability [3], advance hierarchical skill learning for better dynamic adaptation and cross-scenario transfer [1], and use role-based frameworks (like CORD) to improve collaboration flexibility and credit assignment [4]. It can also set up standardized benchmarks.

Along with solving high-dimensional decision-making, these steps will speed up MARL’s use in smart grids and multi-robot collaboration [2].

2 Core Methods and Technical Evolution of Cooperation Mechanisms

In multi-agent reinforcement learning, the cooperative mechanism refers to the scenario where all agents share a common reward function, with the goal of maximizing the overall long-term return. To achieve this objective, the technology has undergone multiple generations of evolution.The first to be introduced is the independent learning paradigm, which was a commonly used optimization approach in the early days. It individualizes agents, that is, ignores the influence between agents, and thus can directly apply the single-agent reinforcement learning algorithm. However, this method is like closing the door to the outside world, not considering the external dynamic situation, resulting in difficult convergence of strategies and low collaboration efficiency. The first improvement is the explicit communication coordination paradigm, which enables agents to "communicate" with each other and be aware of each other's existence. Although it can adapt to complex tasks, it is also prone to information overload and has other shortcomings such as weak anti-interference ability. Then an extremely important event occurred - the paradigm revolution [1]. During this process, a large number of new paradigms were proposed, among which the most famous one was the Centralized Training and Distributed Execution (CTDE) paradigm, which changed the traditional training and execution mode of agents: The core of it is centralized training and decentralized execution. That is, during training, there is a "head coach" who acquires all the information and adjusts the strategy, while during execution, each agent operates independently, relying on the local information it receives to make individual decisions. This paradigm effectively resolves the core contradiction. Centralized training relies on global information and can accurately assess the value of joint actions, thereby solving the problems of credit distribution and environmental non-stationarity. Distributed execution ensures the feasibility and scalability of the algorithm in practical scenarios. Since then, almost all advanced collaborative algorithms have been developed based on the CTDE framework [2].

Under the mode of centralized training and distributed execution, to solve the key problem of credit allocation in collaboration, two main technical paths have been formed: the method based on value decomposition and the actor-critic method. First, let's talk about the technical route based on value decomposition. It needs to break down the overall value of the team into the individual value of each agent. When optimizing in the direction of "making the decomposed individual value consistent with the team's goals", it is necessary to ensure that the expression ability of each agent is strong enough. At the beginning, VDN used simple addition, adding up the Q values of each agent to represent the total Q value of the team. Although it was easy to understand, its expressive power was limited. Later, methods such as QMIX [3] and Qatten introduced monotonicity constraints and nonlinear hybrid networks, which significantly enhanced the model's expressive power while ensuring the consistency of individual and team goals. For instance, they performed exceptionally well in complex tasks like StarCraft II. Methods such as QTRAN and QPLEX, on the other hand, attempt to relax the monotonicity constraint, enabling the algorithm to more accurately depict the capabilities of teams and individuals in more complex scenarios. Looking at the technical route based on actors and critics, it mainly assigns a policy network to each agent, just like an "actor", allowing a centralized network of critics to guide its learning. MADDPG is a typical example of this approach [5, 6]. Its critics can obtain global information and thereby optimize individual strategies. Although MADDPG was designed for the CTDE mode, its idea of distributing "credit" through centralized critics has inspired a lot of subsequent work. COMA goes a step further by using the "counterfactual benchmark" method to compare the value of the actual actions of the agent with the default benchmark actions, accurately assessing the contribution of each individual, and providing a more reliable "credit basis" for the overall strategy update. These two technical routes have promoted the development of collaborative reinforcement learning from different perspectives and have also become the core methods in the CTDE framework. Under the CTDE framework, agents can also evolve from early broadcast communication to more efficient directional communication by learning "who to say" and "what to say", and coordinate in a targeted manner when needed, further enhancing their collaborative capabilities in complex environments [7, 8].

3 Core Methods and Technical Evolution of Competition Mechanisms

Game theory provides the primary theoretical framework for multi-agent competition. It abstracts the conflict of interest situations wherein the strategic choice of each agent determines the distribution of gains between agents. Continuous modification of the models keeps pace with the requirement of a steadily increasing complexity of environments. For example, a distributed coordination approach for multirobot systems, inspired by natural swarms, introduces the k-winner-take-all principle, thereby integrating a simple recurrent neural dynamics solver and a distributed solver—via consensus for global info conversion—within the discrete-time domain toward optimum stability and resource utilization [9]. A three-party dynamic multi-strategy evolutionary game theory method under asymmetry of information for USVs incorporates an incentive mechanism to enhance achievable payoffs, thus enabling it to build an interaction model that allows for higher decision-making accuracy [10]. A hierarchical game-based competitive incentive mechanism for demand response applies multiple agents and links pricing and responses using Stackelberg and evolutionary game theory methods. It also includes an energy system [11]. EGT also models swarm interest conflicts, utilizing the iterated prisoner's dilemma game to study cooperation evolution through four strategies[12].

The other main method is reinforcement learning. In it, agents learn from interacting with the environment and reward signals. However, non-stationarity and coordination problems have to be tackled. Some algorithms, such as MADDPG and MATRPO, are explicitly designed to improve stability and convergence, exist.

The CTDE framework embodies a nice trade-off between global optimization through central training-and local autonomy-in decentralized execution.

Frequency-domain cooperative jamming forms a hierarchical reinforcement learning model by decomposing the problem into subproblems that yield optimal solutions at each level of decomposition [13].

Distributed decision-making boosts real-time performance and scalability. A distance-based algorithm under the improved k-WTA framework generates dynamic signals for near-optimal multi-agent grouping [14].

For moving target defense, a mechanism switches agent interaction modes to optimize defense [15].

Technologically, the early research in multi-agent competition was quite simple in structure and low in agent number. The 1980s established the ideas of agents and multi-agent systems, but complex competition scenarios were absent. The 1990s applied game theory and simple reinforcement learning to things like robot soccer. The 21st century brought complex strategy learning based on deep learning, where AlphaX agents won out in the competitions. Recent trends involve combining it with blockchain for transparency and edge computing for low latency.

4 Mixed Cooperation-Competition Mechanisms and Application Cases

Due to the complexity of the environment, hybrid cooperation and competition opportunities in MARL encounter significant challenges, where agents must collaborate within teams and compete between teams. These complexities are quite evident in the application of different fields. In multiplayer online competitive games such as Dota 2 and League of Legends, agents in a team must coordinate their actions and consider the distribution of positive and negative rewards. They need to adapt to the strategies of competing teams and the rapidly changing environment [16]. This complexity is not only reflected in games but also in life. In autonomous vehicle systems, vehicles need to consider cooperative traffic management and choose route optimization, while also competing for limited roads [17]. In the smart grid system, the decentralized energy system must coordinate to optimize energy production and consumption, but it also faces competition in the energy market [18].

As shown in Figure 1, there are many algorithmic ideas to solve this hybrid mechanism. This framework is based on general techniques such as deep reinforcement learning and game theory. On this basis, specific methods for coordinating cooperation and competition were respectively formulated. Promote coordination within teams through cooperative algorithms, and handle competition among teams by using adversarial game strategies. Ultimately, by designing effective mechanisms and integrating strategies into adversarial applications, a operational hybrid cooperative - competitive system is formed to address challenges in complex environments. The main complexity lies in the rapidly changing environment, the reward distribution problem and the joint action space growing exponentially. However, due to the lack of unified information and the unpredictability of other subjects' behaviors, trust issues often lead to inefficiency [16]. The continuous changes in the environment further complicate the learning process, as the behavior of each agent affects the environment, making it difficult for agents to learn the optimal strategy in such a dynamic setting [19]. So MARL still has many application scenarios and development potential.

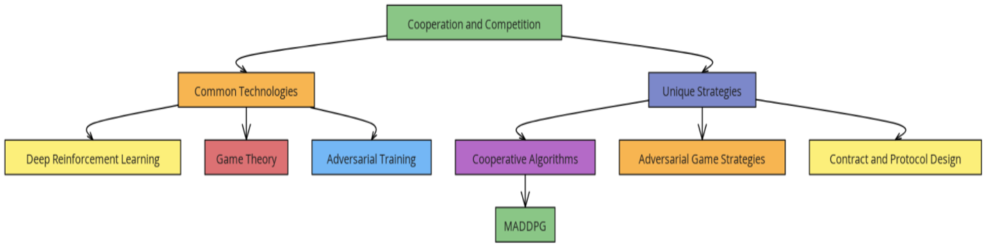

Fig. 1. Hybrid cooperation and competition. (Picture credit: Original)

5 Conclusion

MARL has made remarkable progress in addressing the issues of agent collaboration and competition through core paradigms such as CTDE. The collaboration mechanism optimizes credit allocation through methods such as value decomposition and actor-critic. The competition mechanism deeply integrates game theory and reinforcement learning. However, in the hybrid cooperation-competition scenario, issues such as non-stationary environment and reward distribution remain key challenges. In the future, integrating with large language models, developing hierarchical learning and role frameworks, and establishing standard benchmarks will be crucial directions for promoting the practical application of MARL in real and complex scenarios.

Authors Contribution

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

References

1. G. Zhu, R. Zhou, W. Ji, H. Zhang, D. Wang, S. Zhao, Multi-Task Multi-Agent Reinforcement Learning via Skill Graphs, IEEE Robot. Autom. Lett. (2025)
2. B. Luo, T. M. Hu, Y. H. Zhou, T. W. Huang, C. H. Yang, W. H. Gui, Survey on Multi-Agent Reinforcement Learning for Control and Decision-Making, Acta Automat. Sin. 51(3), 510–539 (2025)
3. A. Satheesh, K. Powell, H. Wei, cMALC-D: Contextual Multi-Agent LLM-Guided Curriculum Learning with Diversity-Based Context Blending, arXiv preprint arXiv:2508.20818 (2025)
4. K. Matsuyama, K. Su, J. Wang, et al., CORD: Generalizable Cooperation via Role Diversity, arXiv preprint arXiv:2501.02221 (2025)
5. P. Sunehag, G. Lever, A. Gruslys, W. M. Czarnecki, V. Zambaldi, M. Jaderberg, M. Lanctot, ... , Value-Decomposition Networks for Cooperative Multi-Agent Learning, arXiv preprint arXiv:1706.05296 (2017)
6. X. Liu, B. Jin, Information-Theoretic Multi-Agent Algorithm Based on the CTDE Framework, in 2024 9th International Conference on Electronic Technology and Information Science (ICETIS), IEEE, pp. 511–516 (2024)
7. W. Chen, S. Huang, J. Schneider, Soft-QMIX: Integrating Maximum Entropy for Monotonic Value Function Factorization, arXiv preprint arXiv:2406.13930 (2024), 1
8. L. Zhang, J. Li, Y. Zhu, et al., Multi-Agent Reinforcement Learning by the Actor-Critic Model with an Attention Interface, Neurocomputing 471, 275–284 (2022)

9. M. Liu, Y. Li, Y. Chen, Y. Qi, L. Jin, A Distributed Competitive and Collaborative Coordination for Multirobot Systems, IEEE Trans. Mob. Comput. 23(12), 11436–11448 (2024)
10. S. Jin, Y. Wang, P. Duan, H. Zhang, G. Li, Z. Cai, Multiagent Confrontation Method Based on Three-Party Dynamic Multistrategy Evolutionary Game, IEEE Trans. Comput. Soc. Syst. (2025)
11. T. Jiang, P. Ju, Z. Lin, Q. Wu, Z. Lin, C. Y. Chung, Competitive Incentive Mechanism for Multi-Agents in Demand Response via a Hierarchical Game Considering Joint Uncertainties, IEEE Trans. Smart Grid (2025)
12. Y. Zhu, Q. Li, J. Zhang, Multiple Strategy Competition Among Structured Populations, in 2017 Chinese Automation Congress (CAC), pp. 3197–3202, IEEE (2017)
13. W. Zhang, T. Zhao, Z. Zhao, Y. Wang, F. Liu, An Intelligent Strategy Decision Method for Collaborative Jamming Based on Hierarchical Multi-Agent Reinforcement Learning, IEEE Trans. Cogn. Commun. Netw. 10(4), 1467–1480 (2024)
14. L. Wang, Q. Guo, Distance-Based Formation Stabilization and Flocking Control for Distributed Multi-Agent Systems, in 2018 IEEE Int. Conf. Mechatronics and Automation (ICMA), pp. 1580–1585, IEEE (2018)
15. H. Zhao, R. Cui, W. Yan, L. Chen, Competition and Cooperation of Multiagent System for Moving Target Defense With Dynamic Task-Switching, IEEE Trans. Syst., Man, Cybern.: Syst. (2025)
16. Z. Ning, L. Xie, Deep Reinforcement Learning in Multi-Agent Systems for Cooperation and Competition, J. Artif. Intell. 34(2), 105–115 (2024)
17. I. Perera, F. de Nijs, J. García, Learning to Cooperate Against Ensembles of Diverse Opponents in Multi-Agent Systems, Neural Comput. Appl. 37(4), 18835–18849 (2025)
18. A. Haupt, P. Christoffersen, M. Damani, D. Hadfield-Menell, Formal Contracts Mitigate Social Dilemmas in Multi-Agent Reinforcement Learning, Auton. Agents Multi-Agent Syst. 38(1), 51–67 (2024)
19. M. Yasir, A. Howes, V. Mavroudis, C. Hicks, Environment Complexity and Nash Equilibria in Multi-Agent Reinforcement Learning, Int. J. Comput. Games 22(1), 50–62 (2024)