Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3075

APPLICATION OF META-LEARNING IN MULTI-AGENT REINFORCEMENT LEARNING - A SURVEY

YanQiao Ji
Liaoning Equipment Manufacturing Vocational and Technical College, Shenyang 110161, Liaoning, China.
Corresponding Email: 491318458@qq.com

Abstract: This survey provides an comprehensive overview of the application of meta-learning in the field of multi-agent reinforcement learning (MARL). Meta-learning, also known as learning to learn, has emerged as a promising approach to enhance the learning efficiency and adaptability of reinforcement learning algorithms. This article explores the challenges and opportunities in applying meta-learning to MARL, highlighting the potential benefits such as faster convergence, improved generalization, and better coordination among agents.
Keywords: Meta-learning; Reinforcement learning; Artificial intelligence

1 INTRODUCTION

1.1 Meta-Learning

Machine learning and artificial intelligence models have been widely applied in various scenarios of life. For a specific artificial intelligence application scenario, the common approach is to train the model using specific data, so that the model can have a better performance in the specific scenario. With the continuous development of machine learning and artificial intelligence technology, its task scenarios have gradually expanded to various aspects of life. For the ever-changing application scenarios of machine learning tasks, how to make AI models adapt to complex and changeable task scenarios is a direction of artificial intelligence research.
In the process of learning new knowledge, different from the process of training a model, human usually reuse the known and effective methods in similar task scenarios based on existing experience and knowledge, simplifying the learning process and reducing the cost of trial and error, and finally realizing cross-task learning. By analogy with the human learning process, the researchers expect the model to "learn to learn" in a similar way. That is, on the basis of existing knowledge, the model can quickly learn new knowledge and use it to solve tasks in new application scenarios, which is called Meta Learning[1]. Different from machine learning, which is usually aimed at solving a specific task, in a meta-learning task, it is often necessary to prepare multiple tasks and their corresponding training and test data, and train the model on the train task so that the "prior knowledge" learned by the model can perform better on the test task.

1.2 Meta-Reinforcement Learning

Reinforcement Learning (RL) primarily focuses on how an agent can maximize the rewards it receives in a complex environment. The agent learns by sensing the state of the environment and the reaction to its actions, choosing better actions to achieve higher returns[2]. Despite the successful applications of RL in path planning, robot control, games and games, finance and education, there are problems with low data efficiency and lack of universality in the generation strategy, which limit the application of RL in many aspects1. Therefore, researchers consider introducing the idea of meta-learning, abstracting the process of obtaining better RL algorithms as another machine learning problem, and thus proposing meta-reinforcement learning (meta-RL). Meta-reinforcement learning is a meta-learning approach to reinforcement learning that tries to solve the common sampling inefficiency or ineffectiveness problems in RL by introducing meta-learning, thereby enabling RL to be applied to more task scenarios.

1.3 Multi-Agent Reinforcement Learning

Multi-Agent Reinforcement Learning (MARL) is a branch of reinforcement learning that aims to train a group of multiple agents and enable them to perform better in their interactions with the environment. Compared to the typical reinforcement learning tasks, the tasks in the MARL application environment usually involve multiple agents, which form a multi-agent system together. Each agent follows the same goal of reinforcement learning, trying to maximize the reward it can obtain. Due to the presence of multiple agents, the global state change of the environment will be affected by the joint action of all agents, which requires the agent to consider the impact of joint action on the multi-agent environment during its strategy learning process. It also brings unique challenges in modeling, algorithm design, and application. In MARL, a key challenge is the balance between cooperative learning and competitive learning. In some cases, agents need to cooperate to achieve common interests, while in other cases, they may need to compete to obtain limited resources. This makes the design of MARL complex, requiring consideration of communication, coordination, and competition strategies between agents.

2 RELATED RESERCH

2.1 Multi-Agent Reinforcement Learning

At present, MARL has made significant progress in a number of areas. One of the prominent applications is multi-agent collaborative control, such as traffic management, drone cooperative flight, etc. In addition, MARL has been widely used in fields such as game theory, social sciences, and economics. However, MARL still faces many challenges, such as learning in adversarial environments, scalability of large-scale multi-agent systems, etc.

For multi-agent reinforcement learning problems, the generalization of single-agent reinforcement learning is a more direct idea, that is, each agent regards other agents as factors in the environment, and still updates the strategy according to the way of single-agent learning and interaction with the environment. This type of approach, also known as value-based methods, includes Independent Q-learning[3] and Cooperative Q-learning[4], etc. This type of approach attempts to achieve collaborative decision-making by modeling the value function of each agent.

For multiple agents, there may be a competitive, hybrid competitive or fully cooperative relationship between them, and in these relationship modes, the decision-making behavior of other agents will have different effects on the individual.

2.1.1 Competitive relationship

The Minimax Q-learning algorithm is commonly used in situations where two agents are preceded by a zero-sum random game in perfect competition. Based on the principle of least-maximum in game theory, the algorithm learns by modeling the adversarial relationship between agents to minimize the expected reward in the worst-case scenario. In Minimax Q-learning, each agent is treated as a minimized player and a maximized player. The minimized player works to minimize their expected reward, while the maximized player tries to maximize the expected reward of the minimized player. This adversarial learning allows agents to learn more robustly in uncertain and adversarial environments[5]. Specifically, Minimax Q-learning uses a Q-value function to represent the strategy for each agent. At each learning step, minimize the player's optimal strategy by updating its Q function to reflect its opponent's maximizing player. At the same time, the maximized player also adapts to the strategy of minimizing the player in an adversarial environment by updating its Q function.

Minimax Q-learning provides a powerful framework for dealing with cooperation and competition in multi-agent systems, especially when it comes to adversarial situations and incomplete information. Its algorithmic structure enables agents to effectively model uncertainties in the environment and adversary strategies to achieve a more robust and intelligent decision-making process.

2.1.2 Hybrid relationship

Nash Q-learning is a multi-agent reinforcement learning algorithm, which is commonly used in two-person zero-sum games between two agents and extends to more general multi-person general and game situations, and is committed to solving the Nash equilibrium problem in game theory. The algorithm coordinates the agent's strategy so that each agent cannot change its own strategy to increase its expected reward given the opponent's strategy. A Nash equilibrium is a combination of strategies in which each agent adopts a strategy that optimally responds to its opponent, forming a state of equilibrium that does not change from one another[6]. In Nash Q-learning, the learning goal of each agent is to find an equilibrium state, that is, a Nash equilibrium, through the evolution process of the game. By updating the Q function, each agent iteratively adjusts its strategy to approximate the Nash equilibrium of the game. This involves adversary modeling, which guides the adjustment of one's own strategy through the estimation of the adversary's strategy to achieve convergence to the equilibrium state.

Nash Q-learning provides an effective framework for coordinating cooperation and competition in multi-agent systems. The focus is on the application of game theory to enable the agent to gradually reach a stable equilibrium state in the game process. This method aims to deal with the policy interaction in the multi-agent system and ensure that each agent makes the optimal strategy choice under the Nash equilibrium, so as to form the cooperative behavior of the whole system.

2.1.3 Cooperation relationship

When multiple agents are fully cooperative in the environment, different agents need to cooperate with each other to achieve better overall performance. In dealing with the problem of mutual negotiation between agents to achieve optimal joint action, the mutual modeling between individuals can provide a potential coordination mechanism for the decision-making of agents. In the Joint Action Learner (JAL)[7] approach, agent i makes decisions based on the observed historical actions of other agents j and the modeling of their strategies. The Frequency Maximum Q-value (FMQ)[8] method introduces the frequency at which individual actions achieve the best return in joint actions to define individual Q values. In the learning process, the agent is guided to choose to maximize the probability of its own action in the joint action that obtains the best return.

Both JAL and FMQ methods are based on equilibrium solving, but these methods are often limited to dealing with small-scale (i.e., small number of agents) multi-agent problems. In practical problems, a large number of agents will be involved in the interaction and mutual influence between agents, and the traditional equilibrium solution method is limited by the computational efficiency and computational complexity, and it is difficult to cope with complex scenarios. In the problem of large-scale multi-agent learning, considering the effect of group joint actions, including the influence of the current agent and the role it plays in the group, is of great help to agent strategy learning.

2.2 Multi-Agent Deep Reinforcement Learning

The rise of deep reinforcement learning has also had a profound impact on MARL. The introduction of deep neural networks has enabled agents to learn more complex representations and strategies, improving the performance of MARL in real-world applications. However, deep MARL also introduces new challenges, such as training instability, sample efficiency, and generalization issues.

With the development of deep learning, the powerful expressive power of neural networks is used to build value approximation models and policy models (commonly found in policy-based DRL methods). Deep reinforcement learning methods can be divided into value-based and policy-based, when considering the multi-agent problem, the main way is to introduce multi-agent related factors into the definition of value function or strategy, and design the corresponding network structure as a value function model and strategy model, and finally the trained model can adapt (directly or potentially learn the complex relationship between agents). Get good results on specific tasks.

2.2.1 Value-based method

The value function-based approach can be said to be the first attempt of multi-agent reinforcement learning algorithms (e.g., IQL algorithm). However, for more complex environments, IQL is not able to handle the problems caused by non-stationary environments. The centralized method, that is, the method of merging the state space and action space of all agents as an agent, can better deal with the problem of environmental non-stationarity, but there are also problems such as poor algorithm scalability in a large-scale multi-agent environment, and the strategies made by agents with slower progress will hinder agents who have learned some strategies, thus reducing the global debriefing. Therefore, the researchers proposed a VDN (Value-Decomposition Networks) method[4]. The basic idea is that a federated Q network is centrally trained, but this federated network is obtained by the sum of the local Q networks of all agents, so that not only can the problems caused by the non-stationary environment be dealt with through centralized training, but also the complex interrelationships between agents are decoupled because they are actually learning the local model of each agent. Finally, since each agent has a Q network based only on its own local observations after training, decentralized execution can be realized.

VDNs make strong assumptions about the relationship between agents; however, such assumptions are not applicable to all cooperative multi-agent problems. In order to overcome this limitation, the researchers proposed an improved method, QMIX (Q-Mixing). In the QMIX framework, each agent is given its own Q-value function, and these individual Q-values are combined to form a global Q-value. QMIX uses a neural network known as a "hybrid network" whose task is to learn how to effectively integrate the individual Q values of individual agents into a global Q value, thereby motivating the agents to cooperate efficiently when performing actions. QMIX achieves two key improvements over VDN: first, global information is introduced during training to provide assistance; Secondly, the hybrid network is used to fuse the local value function of a single agent instead of simple linear addition to further improve the performance of the model. QMIX is designed so that agents can make collaborative decisions by maximizing global rewards without direct communication.

2.2.2 Policy-based method

When scaling reinforcement learning algorithms from a single agent to a multi-agent environment, the most straightforward approach is to adopt an independent Q learning (IQL) class approach. However, in complex environments, there are some difficulties in dealing with such methods due to the non-stationary nature of the environment. Again, while a centralized approach is able to cope with these issues, it sacrifices scalability to a certain extent relative to IQL. Therefore, the researchers propose an Actor-critic based approach[9]. The basic idea of this algorithm is to decompose the learning problem into two main components: the actor and the critic. The actor is responsible for performing the action, and the critic is responsible for evaluating how good or bad the action chosen by the actor is. There is a synergistic relationship between the actor and the critic, with the actor rewarding the improvement strategy by maximizing expectations, while the critic directs the actor's improvement by providing an evaluation of the actor's movements. This decoupled architecture enables actor-critic algorithms to deal more efficiently with complex environments and high-dimensional state spaces.

Multi-Agent Deep Deterministic Policy Gradient (MADDPG) is a reinforcement learning method for solving cooperative and competitive multi-agent problems. MADDPG is a multi-agent extension of Deep Deterministic Policy Gradient (DDPG) designed to address the challenges of inter-agent interaction and collaborative decision-making between agents. The goal of MADDPG is to enable agents to maximize cumulative rewards in a multi-agent environment by learning appropriate strategies. This approach excels in multi-agent scenarios involving collaborative decision-making or competing tasks, enabling agents to learn effective collaboration strategies. MADDPG establishes a centralized critic for each agent, who is able to obtain global information, including the global state and the actions of all agents, to provide the corresponding value function. To a certain extent, this design helps to alleviate the problem of environmental instability of multi-agent systems. At the same time, the actors of each agent only need to make decisions based on their local observations, so as to achieve distributed control of the multi-agent system.

3 APPLICATION OF META-REINFORCEMENT LEARNING MULTI-AGENT REINFORCEMENT LEARNING

As an important branch of reinforcement learning, multi-agent reinforcement learning(Multi-Agent Reinforcement Learning，MARL) is in the stage of continuous evolution and improvement. With the growing demand for modeling and decision-making in complex systems, researchers are faced with the challenge of generalizing the unknown environment. To address this challenge, they began to consider introducing a mindset of meta-learning. Meta-learning,

as a learning mode of learning how to learn, provides a new idea for solving the generalization problem in multi-agent reinforcement learning. Researchers try to combine meta-reinforcement learning with multi-agent reinforcement learning to solve the generalization difficulties and nonstationary problems that arise in multi-agent environments. This integration effort aims to make the system more adaptable, enabling efficient decision-making in unknown environments.

In traditional multi-agent reinforcement learning, the generalization problem has become particularly prominent due to the interaction of multiple agents and complex game strategies. The introduction of meta-learning provides a way for researchers to improve the performance of systems in new domains by learning the ability to adapt quickly in different environments. However, generalization difficulties in multi-agent reinforcement learning are not the only challenges. The issue of non-stationarity is also an important aspect to be addressed. The dynamics and interaction complexity of the multi-agent environment make traditional learning algorithms unable to cope with nonstationary. The introduction of meta-reinforcement learning provides a new way to deal with this problem, so that the system can better adapt to the changes in the environment through learning adaptability and flexibility. The integration of meta-reinforcement learning and multi-agent reinforcement learning aims to overcome the generalization and non-stationarity challenges brought about by the multi-agent environment. This research direction provides new theories and methods for building more intelligent and adaptive systems.

Meta-learning has been widely used in the field of multi-agent reinforcement learning to deal with diverse and complex problems. On the one hand, meta-learning solves the problem of information transfer and cooperation in multi-agent systems by learning which agents to communicate with1[10]. This learning mechanism enables the agent to adaptively select the agent it communicates with in the case of environmental changes or uncertainties, so as to maximize the performance of the whole system. On the other hand, meta-learning provides a mechanism for multi-agent systems to adjust and improve autonomously by learning agent-specific reward functions to automatically design mechanisms[11]. This approach allows the agent to better adapt to changing external conditions by optimizing its behavior through dynamic responses to the environment. A similar example is the ability to effectively deal with leader-follower dynamics in multi-agent systems[12] by performing agent learning to compute the Stackelberg equilibrium. This approach enables the system to quickly provide an adaptive, optimal response strategy when training a fixed leader strategy, resulting in an overall performance improvement in the system. This series of applications demonstrates the broad and far-reaching impact that meta-learning offers in multi-agent systems. The application of meta-learning in multi-agent reinforcement learning is not only limited to solving specific problems, but also provides a more flexible and intelligent learning paradigm for the system. This research direction not only promotes the development of multi-agent system theory, but also provides strong support for the design of more intelligent and adaptive systems in practical applications.

3.1 Problems Solved by Meta-Reinforcement Learning Methods in Multi-Agent Reinforcement Learning

This section focuses on two main problems to be solved by meta-reinforcement learning in a multi-agent environment. Firstly, the generalization problem of unseen agents and unstable agents is introduced, and how meta-reinforcement learning can solve these problems in general is discussed. Secondly, the types of meta-reinforcement learning methods used to solve each problem are discussed, and the PPG method that proposes additional mechanisms for each problem is elaborated.

We first consider the generalization of introducing new agents in a multi-agent environment. In multi-agent reinforcement learning, many agents act in a shared environment. In general, there may be large differences between the strategies of different agents, which creates the problem of generalization of unknown agents. This generalization can occur between opponents [13], or between teammates[14], or for specific tasks[15].The complexity of the generalization problem lies in the fact that the identities of other agents may be highly variable, either through learned strategies or actual human decision-makers.

In the context of meta-reinforcement learning, researchers consider the existence of other agents as part of the task and assume that the distribution of these agents is known and available for training. This assumption and perspective make meta-reinforcement learning methods effective in dealing with generalization problems in multi-agent systems. By considering the variability of other agents and incorporating them into the learning framework, meta-reinforcement learning provides a potential solution to improve the agent's ability to generalize to other agents in an unknown environment.

In multi-agent reinforcement learning, the problem of unsteady state is also an important and complex challenge. From the point of view of any one agent, all other agents are constantly learning, causing the environment of the problem to change. This makes it difficult for traditional learning algorithms to effectively cope with this unsteady state. The meta-reinforcement learning method considers the unsteady state as a part of the task by including other learning agents in the definition of the task, which enables the meta-learning algorithm to better adapt to the changes introduced by other agents in the learning process, so as to improve the adaptability of the system to environmental unsteady. By repeatedly resetting other learning agents during meta-training, we can meta-learn how to handle changes introduced by other agents. From the perspective of meta-learning agents, the distribution to other agents remains the same. This effectively solves the problem of unsteady state of multi-agent reinforcement learning.

3.2 Meta-Reinforcement Learning Methods Applied in Multi-Agent Reinforcement Learning

Various types of meta-reinforcement learning methods can be applied in the process of multiagent reinforcement learning, among which the most typical ones are PPG (Proximal Policy Gradients) method[16], black-box method[17] and task inference methods[14], etc. If another agent is also learning, that agent can even use a non-adaptive strategy such as Markov[18]. Most of these methods can be directly applied to the underlying meta-reinforcement learning problem to solve the generalization and unsteady problem of multiple agents.

The PPG method is a strategy gradient optimization method in reinforcement learning, which aims to solve the decision-making problem in the continuous action space. The core idea of the PPG method is to stabilize and accelerate the policy optimization process by controlling the magnitude of policy updates and maintaining the "proximal invariance" of the policy distribution. PPG's goal is to maximize cumulative returns while ensuring that the difference between the new strategy and the old strategy is bounded. This mechanism for controlling for differences helps to prevent excessive policy updates from being introduced in training, thereby improving the stability of training. This approximate invariance makes the PPG method an excellent performer in the face of highly variable environments. It can be applied to collaboration, adversarial scenarios, and other complex multi-agent systems. Furthermore, some scholars have studied additional mechanisms based on the idea of PPG to further deal with the generalization and unsteady problems in multi-agent reinforcement learning.

3.3 For Generalization Problems in Multi-Agent Reinforcement Learning

Some of the existing work is dedicated to improving the distribution of other agents through meta-gradients to improve the generalization ability of new agents. In 2021, Abhinav et al.[19] proposed a meta-learning method based on dynamic populations for multi-agent communication with natural language. In this way, populations are built dynamically in an iterative manner and diversity is introduced over time. In this way, an agent can be trained using population-based meta-learning algorithms that allow it to generalize among known and unknown partners and humans, and ultimately to obtain an agent that can cooperate with known, unknown, and humans in a multi-agent communication environment. This work iterates between PPG meta-learning and fixed adversary distributions, and adds the optimal response of meta-learning agents back to the population, leveraging the distribution of agents to create a robust agent capable of generalizing among many other agents.

3.4 For Unsteady State Problems in Multi-Agent Reinforcement Learning

Other works introduce additional mechanisms to deal with unsteady-state. In order to account for astability, all adaptive or non-adaptive agents must eventually be repeatedly reset to their initial policy. However, PPG-based approaches tend to allow other agents to continue learning [17] or even meta-learning cite{A policy gradient algorithm for learning to learn in multiagent reinforcement learning} without resetting.In 2021, Kim et al.[16] proposed a meta-multiagent policy gradient theorem Meta-MAPG (meta-multiagent policy gradient theorem).

Due to the presence of multiple agents in the environment, the overall environment is unsteady due to the change of strategies of other agents when each agent perceives the environment, and the distribution of experiences encountered by each agent itself is also astable. In order to solve this steady-state problem, the algorithm proposed in this work directly considers the inherent unsteady policy dynamics in the multi-agent learning environment, and models the gradient update to consider both the agent's own unsteady policy dynamics and the unsteady policy dynamics of other agents in the environment, so it can quickly adapt to the unsteady dynamics of other agent strategies in the shared environment. At the same time, in the process of interacting with other learning agents, the future strategies of other agents will also be affected to a certain extent, and this influence will also help to improve the performance of the meta-agents themselves to a certain extent.

4 CONCLUSION

Meta-reinforcement learning, a transformative approach that equips agents with the ability to learn how to learn, holds immense promise for addressing the challenges of generalization and non-stationarity in multi-agent reinforcement learning. By enabling agents to adapt their strategies to new agents and changing environments, meta-reinforcement learning empowers multi-agent systems to navigate complex and dynamic scenarios effectively. The Proximal Policy Gradients (PPG) algorithm, along with its extensions, has demonstrated its effectiveness in tackling both generalization and non-stationarity challenges. However, further research is needed to explore alternative meta-reinforcement learning algorithms, address scalability issues, and integrate this approach with other machine learning techniques. By pushing the boundaries of meta-reinforcement learning and its applications, we can unlock the full potential of multi-agent systems and pave the way for intelligent and adaptable solutions in various domains.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

The project was supported by Research on the Construction of Virtual Simulation Training Base for Vocational Education (LZJG2023041).

REFERENCES

[1] Vanschoren, J. Meta-Learning: A Survey. arXivOctober8, 2018. DOI: https://doi.org/10.48550/arXiv.1810.03548.
[2] François-Lavet, V, Henderson, P, Islam, R, et al. An Introduction to Deep Reinforcement Learning. Found. Trends® Mach. Learn, 2018, 11(3-4): 219–354. DOI: https://doi.org/10.1561/2200000071.
[3] Tampuu, A, Matiisen, T, Kodelja, D, et al. Multiagent Cooperation and Competition with Deep Reinforcement Learning. Plos One, 2017, 12(4): e0172395. DOI: https://doi.org/10.1371/journal.pone.0172395.
[4] Sunehag, P, Lever, G, Gruslys, A, et al. Value-Decomposition Networks For Cooperative Multi-Agent Learning. arXiv June 16, 2017. DOI: https://doi.org/10.48550/arXiv.1706.05296.
[5] Littman, ML. Markov Games as a Framework for Multi-Agent Reinforcement Learning. In Proceedings of the Eleventh International Conference on International Conference on Machine Learning; ICML ' 94; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1994, 157-163.
[6] Hu, J, Wellman, MP. Nash Q-Learning for General-Sum Stochastic Games. J. Mach. Learn. Res, 2003, 4 (null), 1039-1069.
[7] Claus, C, Boutilier, C. The Dynamics of Reinforcement Learning in Cooperative Multiagent Systems. In Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence; AAAI '98/IAAI '98; American Association for Artificial Intelligence: USA, 1998, 746-752.
[8] Kapetanakis, S, Kudenko, D. Reinforcement Learning of Coordination in Cooperative Multi-Agent Systems. In Eighteenth national conference on Artificial intelligence; American Association for Artificial Intelligence: USA, 2002, 326-331.
[9] Konda, V, Tsitsiklis, J. Actor-Critic Algorithms. In Advances in Neural Information Processing Systems. MIT Press, 1999, 12.
[10] Zhang, Q, Chen, D. A Meta-Gradient Approach to Learning Cooperative Multi-Agent Communication Topology. 5th Workshop on Meta-Learning at NeurIPS 2021. 2021.
[11] Yang, J, Wang, E, Trivedi, R, et al. Adaptive Incentive Design with Multi-Agent Meta-Gradient Reinforcement Learning. In Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems; AAMAS' 22; International Foundation for Autonomous Agents and Multiagent Systems: Richland, SC, 2022, 1436-1445.
[12] Gerstgrasser, M, Parkes, D. C. Meta-RL for Multi-Agent RL: Learning to Adapt to Evolving Agents. Workshop: NeurIPS 2022 Workshop on Meta-Learning. 2022.
[13] Papoudakis, G, Albrecht, SV. Variational Autoencoders for Opponent Modeling in Multi-Agent Systems. 2020. DOI: https://doi.org/10.48550/arXiv.2001.10829.
[14] He, J ZY, Erickson, Z, Brown, DS, et al. Learning Representations That Enable Generalization in Assistive Tasks. Proceedings of The 6th Conference on Robot Learning, PMLR, 2023, 205: 2105-2114.
[15] Stone, P, Kaminka, GA, Kraus, S, et al. Ad Hoc Autonomous Agent Teams: Collaboration without Pre-Coordination. In Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence; AAAI'10; AAAI Press: Atlanta, Georgia, 2010, 1504-1509.
[16] Kim, DK, Liu, M, Riemer, MD, et al. A Policy Gradient Algorithm for Learning to Learn in Multiagent Reinforcement Learning. In Proceedings of the 38th International Conference on Machine Learning, PMLR, 2021, 5541-5550.
[17] Foerster, J- Chen, RY, Al-Shedivat, M, et al. Learning with Opponent-Learning Awareness. In Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems; AAMAS' 18; International Foundation for Autonomous Agents and Multiagent Systems: Richland, SC, 2018, 122-130.
[18] Lu, C, Willi, T, Letcher, A, et al. Adversarial Cheap Talk. In Proceedings of the 40th International Conference on Machine Learning; ICML' 23; JMLR.org: Honolulu, Hawaii, USA, 2023, 202, 22917-22941.
[19] Gupta, A, Lanctot, M, Lazaridou, A. Dynamic Population-Based Meta-Learning for Multi-Agent Communication with Natural Language. In Advances in Neural Information Processing Systems; Curran Associates, Inc., 2021, 34, 16899-16912.