

Resource Competition in Multi-Agent Reinforcement Learning: A Literature Review

Reinforcement Learning - DSDM

Anastasia Chernavskaia

Moritz Peist

Nicolas Rauth

Abstract

Multi-agent reinforcement learning in competitive resource environments represents a critical domain where autonomous agents must learn optimal strategies while competing for limited resources. This literature review examines recent developments in competitive multi-agent systems, with particular emphasis on resource allocation scenarios. We survey foundational work including OpenAI's MADDPG framework, sequential social dilemmas, emergent complexity through competition, and practical applications across industries. The review identifies key algorithmic approaches, addresses fundamental challenges, including non-stationarity and credit assignment in competitive settings, and highlights current research directions toward more robust and scalable competitive multi-agent systems.

June 23, 2025

Contents

1	Introduction	2
2	Foundational Frameworks and Algorithmic Approaches	2
2.1	MADDPG and Centralized Training with Decentralized Execution	2
2.2	Value Decomposition and Policy Gradient Methods	3
3	Sequential Social Dilemmas and Emergent Complexity	3
3.1	Temporal Resource Competition	3
3.2	Autocurricula and Behavioral Complexity	4
4	Industry Applications and Practical Implementations	4
4.1	Cloud Computing and Network Resource Management	5
4.2	Financial Markets and Resource Competition	5
5	Current Challenges and Future Research Directions	5
5.1	Scalability and Robustness	6
5.2	Theoretical Foundations and Evaluation	6
6	Conclusion	7
A	Appendix	8

1 Introduction

Multi-agent reinforcement learning (MARL) in competitive resource environments presents unique challenges where agents must simultaneously learn optimal policies while adapting to the evolving strategies of competing agents. Unlike purely cooperative or zero-sum scenarios, resource competition introduces complex dynamics where agents compete for scarce resources while potentially benefiting from limited cooperation (**lowe_multi-agent_2017**).

The significance of this domain extends beyond theoretical interest to practical applications in cloud computing, wireless networks, autonomous vehicles, and financial markets, where multiple autonomous entities must efficiently allocate limited resources while pursuing individual objectives (**wang_qplex_2021**; **allahham_multi-agent_2022**). These environments exhibit fundamental properties that distinguish them from single-agent learning: non-stationarity from each agent's perspective, scalability challenges with increasing agent populations, and emergent behaviors that arise from competitive interactions (**hernandez-leal_survey_2019**; **ning_survey_2024**).

2 Foundational Frameworks and Algorithmic Approaches

The development of practical algorithms for competitive resource environments has required fundamental innovations in multi-agent learning architectures. The core challenge lies in addressing non-stationarity while enabling agents to learn effective competitive strategies. This section examines the foundational algorithmic contributions that have shaped the field, beginning with centralized training approaches that maintain decentralized execution, followed by value decomposition methods and policy gradient techniques adapted explicitly for competitive scenarios.

2.1 MADDPG and Centralized Training with Decentralized Execution

The seminal work by **lowe_multi-agent_2017** introduced Multi-Agent Deep Deterministic Policy Gradient (MADDPG), establishing a foundational framework for learning in mixed cooperative-competitive environments. MADDPG addresses the fundamental challenge of non-stationarity by employing centralized training with decentralized execution (CTDE), where agents use centralized critics during training that have access to all agents' observations and actions, while maintaining decentralized policies for execution.

The algorithm extends single-agent DDPG to multi-agent settings through a key innovation: each agent i maintains its own actor network $\mu_i(o_i|\theta_i)$ but uses a centralized critic $Q_i^\mu(x, a_1, \dots, a_N)$ where x represents the global state and a_1, \dots, a_N are all agents' actions. The policy gradient for agent i becomes:

$$\nabla_{\theta_i} J(\mu_i) = \mathbb{E}[\nabla_{\theta_i} \mu_i(o_i) \nabla_{a_i} Q_i^\mu(x, a_1, \dots, a_N) |_{a_i=\mu_i(o_i)}] \quad (1)$$

This approach demonstrates superior performance compared to independent learning ap-

proaches in environments featuring resource competition, such as the "simple tag" scenario, where agents compete to reach specific locations while avoiding adversaries (cf. our replication here). The OpenAI experiments revealed critical insights about how resource availability affects agent behavior and learning dynamics, establishing that multi-agent environments possess inherent curriculum properties where increasing agent capabilities create correspondingly more challenging environments (**lowe_learning_2017**).

2.2 Value Decomposition and Policy Gradient Methods

For competitive resource allocation, value decomposition methods address the credit assignment problem while maintaining competitive dynamics. QMIX (**rashid_monotonic_2020**) and its variants enable agents to learn joint action-value functions through monotonic mixing networks, ensuring that individual Q-value maximization leads to optimal joint actions in cooperative sub-tasks while maintaining competitive behavior where appropriate. Recent advances include QPLEX (**wang_qplex_2021**), which introduces a duplex dueling architecture for Q-function decomposition, and Weighted QMIX, which breaks monotonicity limitations to handle more complex competitive interactions.

Furthermore, multi-Agent Proximal Policy Optimization (MAPPO) (**yu_surprising_2022**) has demonstrated surprising effectiveness in competitive environments despite its conceptual simplicity, with success stemming from the proper implementation of centralized value functions rather than algorithmic complexity. Counterfactual Multi-Agent Policy Gradients (COMA) (**foerster_counterfactual_2018**) addresses credit assignment in competitive settings through counterfactual baselines that measure individual agent contributions while marginalizing out other agents' actions.

3 Sequential Social Dilemmas and Emergent Complexity

The evolution from simple matrix games to complex temporal environments has revealed fundamental insights about how resource competition shapes agent behavior over extended time horizons. This section explores how sequential social dilemmas capture the essence of real-world resource competition, where agents must balance immediate gains against long-term strategic positioning. We examine both the theoretical foundations of these environments and the remarkable complexity that emerges through competitive self-play, demonstrating how simple competitive pressure can give rise to sophisticated behavioral patterns.

3.1 Temporal Resource Competition

leibo_multi-agent_2017<empty citation> extended the traditional matrix game framework to sequential social dilemmas (SSDs), recognizing that real-world resource competition involves temporally extended decisions rather than single-shot actions. Their work introduced paradigmatic environments that capture essential aspects of resource competition, including the Fruit Gathering Game where agents navigate spatial envi-

ronments to collect finite fruit resources, and the Wolfpack Hunting Game that models predator coordination balancing individual reward maximization with collective hunting benefits.

The SSD framework reveals how environmental factors, particularly resource abundance, fundamentally shape learned behaviors. When resources are plentiful, agents pursue individual strategies with minimal interference. As scarcity increases, emergent phenomena occur including territorial establishment and defense, sophisticated temporal strategies for resource acquisition, and adaptive counter-strategies in response to opponents' evolving policies. These findings demonstrate that cooperativeness is a property of policies rather than elementary actions, requiring agents to learn complex behavioral patterns balancing immediate rewards with long-term strategic positioning.

3.2 Autocurricula and Behavioral Complexity

bansal_emergent_2018<empty citation> demonstrated that competitive multi-agent environments trained with self-play can produce behaviors far more complex than the environment itself. Their work in 3D physics-based competitive environments shows that agents develop sophisticated skills including running, blocking, tackling, and strategic deception through purely competitive pressure. The concept of autocurricula emerges as a critical feature, where agents' performance improvements effectively modify the learning environment for all participants, creating stacked layers of learning where each behavioral innovation provides the foundation for the next level of complexity.

This research identifies several mechanisms through which competitive pressure generates behavioral complexity, including arms race dynamics that create continuous pressure for counter-strategy development, natural curriculum adjustment that matches challenge levels to agent capabilities, skill transfer from competitive contexts to related tasks, and emergent tool use where agents discover and exploit environmental features for competitive advantage. The feedback loop ensures that, regardless of skill level, the environment contains opponents that are appropriately challenging.

4 Industry Applications and Practical Implementations

The transition from laboratory environments to real-world deployment has demonstrated the practical viability of competitive multi-agent learning across diverse industries. This section examines how theoretical advances in competitive multi-agent reinforcement learning (MARL) have been successfully applied to solve concrete resource allocation problems in cloud computing, wireless networks, and financial markets. These applications reveal both the potential and the practical challenges of deploying competitive agents in environments where resource constraints drive natural competition among autonomous systems.

4.1 Cloud Computing and Network Resource Management

belgacem_intelligent_2022<empty citation> developed the Intelligent Multi-Agent Reinforcement Learning Model (IMARM) for cloud resource allocation, where virtual machines compete for computational resources while optimizing energy consumption and fault tolerance. The multi-agent approach enables dynamic resource allocation responding to changing consumer demands, with Q-learning policies guiding virtual machines to optimal states based on current environmental conditions. Competitive elements arise from finite computational resources and the need to balance multiple objectives, including energy efficiency, load balancing, and quality of service.

Recent work in 6G wireless networks (**zhang_multi-agent_2025**) employs MARL for dynamic spectrum allocation and network selection, where edge nodes compete for radio access technologies and bandwidth resources. The multi-agent framework addresses natural network decentralization while enabling intelligent coordination. Key competitive elements include spectrum competition among multiple nodes for limited frequency bands, quality of service optimization that balances individual performance with network-wide efficiency, and energy management that optimizes battery lifetime while maintaining service quality.

4.2 Financial Markets and Resource Competition

Multi-agent systems in financial markets represent natural competitive resource environments where trading agents compete for profitable opportunities in markets with limited liquidity (**shavandi_multi-agent_2022**). Recent developments employ MARL for algorithmic trading, where agents learn to adapt to market conditions and competitor strategies. This includes portfolio optimization, which involves dynamic capital allocation across multiple assets, and risk management, balancing individual returns with systemic risk considerations. The competitive nature arises from zero-sum trading aspects where one agent's profit often corresponds to another's loss, combined with finite liquidity available at any price level.

In competitive resource environments, communication presents challenges where agents may benefit from coordination to avoid wasteful conflicts while maintaining competitive advantages. Recent work explores learned communication protocols that balance information sharing with competitive considerations (**foerster_learning_2016**), strategic information sharing that maximizes individual utility while enabling beneficial coordination (**eccles_biases_2019**), and deceptive communication, where agents learn to provide misleading information for competitive advantages (**cao_emergent_2018**).

5 Current Challenges and Future Research Directions

Despite significant progress in competitive multi-agent learning, several fundamental challenges remain that limit the scalability and practical deployment of these systems. This section addresses the primary obstacles facing the field, including computational scalability as agent populations grow, robustness against adversarial exploitation, and the need

for stronger theoretical foundations. We examine current research approaches to these challenges and identify promising directions for future investigation that could enable more robust and scalable competitive multi-agent systems.

5.1 Scalability and Robustness

Current MARL approaches face significant scalability challenges as the number of competing agents increases, with joint action spaces growing exponentially and centralized training becoming computationally prohibitive. Research directions include mean-field approaches that approximate large-scale interactions through average effects, graph neural networks that leverage structural relationships for efficient computation, and hierarchical methods that decompose large-scale problems into manageable sub-problems (liu_scaling_2024; ma_efficient_2024).

Competitive environments pose unique robustness challenges as adversarial interactions can exploit learned vulnerabilities. Current research focuses on population-based training, maintaining diverse agent populations to improve robustness, adversarial training that explicitly targets worst-case opponents, and domain randomization, which enhances generalization across varied competitive scenarios (zhou_malib_2023).

5.2 Theoretical Foundations and Evaluation

The theoretical analysis of competitive resource allocation builds on game-theoretic foundations, particularly the existence and computation of Nash equilibria in multi-agent settings. However, traditional equilibrium concepts often fail to capture dynamic learning in resource-constrained environments. Recent theoretical work explores correlated equilibria as more robust solution concepts for multi-agent learning, as well as no-regret learning algorithms with theoretical guarantees in competitive settings, and theoretical frameworks for opponent modeling that facilitate learning about and adapting to competitor strategies (fuente_game_2024).

Independent learning in competitive environments typically lacks convergence guarantees due to non-stationarity. Research addresses this issue through multi-timescale learning, which utilizes different learning rates to stabilize competitive dynamics, best response dynamics, and theoretical analysis of agent response patterns, as well as regret minimization algorithms that minimize worst-case performance against adaptive opponents.

The competitive MARL community lacks standardized benchmarks that capture the full complexity of resource competition. Recent efforts include PettingZoo environments, which provide standardized multi-agent environments with competitive scenarios, Melling Pot, DeepMind’s evaluation suite for social interaction, including resource competition, and JaxMARL, offering high-performance implementations that enable large-scale experimentation (liu_scaling_2024).

6 Conclusion

Multi-agent reinforcement learning in competitive resource environments has evolved from foundational frameworks, such as MADDPG, to sophisticated applications across various industries, including cloud computing, wireless networks, and financial markets. The field has demonstrated that resource scarcity fundamentally shapes emergent behaviors, leading to complex strategic interactions extending far beyond simple competition.

Key contributions include establishing sequential social dilemmas as a paradigm for understanding temporal resource competition, demonstrating emergent complexity through competitive pressure, and developing scalable algorithms for real-world resource allocation problems. Current challenges focus on scalability, robustness, and theoretical understanding of learning dynamics in competitive settings.

Future research directions emphasize the development of more sophisticated evaluation frameworks, improving theoretical guarantees for learning in competitive environments, and exploring hybrid cooperative-competitive scenarios that better reflect real-world resource allocation challenges. The integration of large language models and foundation model approaches presents new opportunities for developing more capable and generalizable competitive agents. For instance, **fish_algorithmic_2025**<empty citation> demonstrated collusion in multi-agent environments under specific prompt settings, where large language models competing in a fictional market created new coordination challenges rather than genuine competition.

The field's evolution toward practical deployment demonstrates the maturity of competitive multi-agent reinforcement learning (MARL) as a technology for addressing real-world resource allocation challenges, while continuing to provide insights into the fundamental nature of multi-agent learning and strategic interaction.

A Appendix