Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/14/20230787

Multi-Agent Reinforcement Learning in non-zero-sum games: Algorithms and applications

Xu Ma

University of Liverpool, Liverpool, UK

sgxma18@liverpool.ac.uk

Abstract. Multi-Agent Reinforcement Learning (MARL) algorithms have been utilized in resource allocation modelling, game theory analysis, formation of alliances and so on. The potential for future research and applications of MARL in non-zero-sum games is vast, including integrating MARL with game theory and mechanism design, developing online learning algorithms for dynamic environments, and exploring applications in various domains. Currently, there is a relative lack of review literature on this area of research. This paper, therefore, aims to fill this gap. This work will first introduce the role of MARL in non-zero-sum games. Then it will discuss in detail the practical applications of MARL in economics, social sciences, and political science, which are typical of non-zero-sum games while presenting possibilities and challenges for future research. Finally, a summary is given at the end of the paper. This article provides insights into the current and future possibilities of using MARL in non-zero-sum games.

Keywords: multi-agent, reinforcement learning, non-zero-sum games.

1. Introduction

Non-zero-sum games refer to situations in game theory where the outcomes for each player are not directly opposite or in conflict. Unlike zero-sum games, where the total gains and losses always balance out to zero, non-zero-sum games can result in mutual benefits or losses. These games often require cooperation and negotiation between players to achieve the best outcome for all involved parties.

Multi-Agent Reinforcement Learning (MARL) is a subfield of machine learning that deals with agents that learn to act and optimize their behaviors in a given environment. The study of MARL in non-zero-sum games is motivated by many real-world scenarios involving multi-agent systems with competing objectives. These scenarios often require agents to coordinate their actions, and it can be challenging to achieve a globally optimal solution without considering the activities of all agents [1]. In non-zero-sum games, the objectives of the agents may be aligned or conflicting, and the agents' actions can significantly impact each other's outcomes [2]. Studying MARL in non-zero-sum games can develop more efficient and effective decision-making strategies in various applications, such as traffic control, resource allocation, and robotics. Moreover, it can help us gain insights into how humans behave in complex social situations, leading to a better understanding of social behavior and decision-making. This article explores the role of MARL in non-zero-sum games and its practical applications in economics, social sciences, and games.

DOI: 10.54254/2755-2721/14/20230787

2. Applications of MARL in non-zero-sum games

2.1. Applications in economics

Applications of MARL to non-zero-sum games in economics include resource allocation modelling and game theory analysis. For example, modelling the resource allocation problem using the MARL algorithm allows multiple agents to learn and maximize their payoffs given limited resources. In addition, game-theoretic analysis using the MARL algorithm can reveal strategies and game equilibria among agents, helping economists to understand complex market environments better and predict the behavior of markets.

2.1.1. The potential for resource allocation optimization in economic scenarios. Figure 1 shows the resource allocation process in a non-zero-sum game.

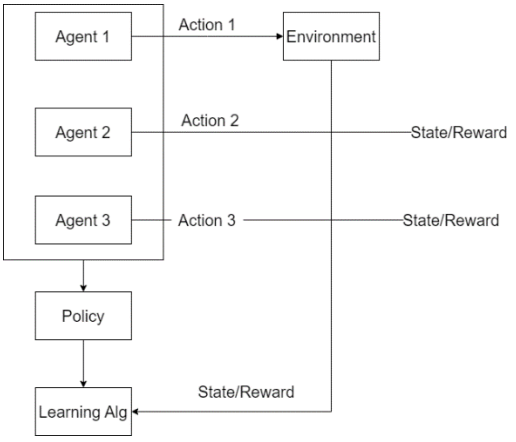

Figure 1. Resource allocation pro- cess in a non-zero-sum game.

Multi-agent reinforcement learning (MARL) can model complex resource allocation scenarios in non-zero-sum games within the economic domain. For instance, in a company, MARL can simulate the interplay of cooperation and competition among employees to maximize resource allocation efficiency [3]. When resources are scarce, MARL can assist agents, such as employees, to acquire rational resource allocation strategies to optimize their gains without adversely affecting other agents. This technique can be implemented in various domains, such as production scheduling, logistics management, energy allocation, etc. By leveraging the power of MARL, economists can enhance their ability to optimize resource allocation and utilization to achieve sustainable economic growth [4]. MARL is a promising approach that can help address resource allocation challenges in real-world scenarios.

2.1.2. Game-theoretic analysis of bargaining. Game-theoretic analysis of bargaining is an essential aspect of multi-agent reinforcement learning (MARL) in non-zero-sum games. MARL has shown tremendous potential in analyzing economic negotiation and bidding game scenarios. Nash-Q algorithm, proposed by Hu and Wellman in their seminal work [5], has been widely used in perfect information games to find the Nash equilibrium and optimize the strategy of agents involved in the game. In contrast, the Minimax-Q algorithm, introduced by Bowling and Veloso [6], is more suitable for imperfect information games, where the optimal strategy is found by minimizing the maximum loss. These algorithms have been applied to economic scenarios such as auctions and bidding games [7]. Applying these game-theoretic algorithms in MARL provides a new tool for economists to understand better and predict market trends, thus contributing to the development of the economic field.

DOI: 10.54254/2755-2721/14/20230787

2.2. Applications in social science

2.2.1. Formation of alliances and bargaining. In non-zero-sum games in social sciences, MARL can be used for the analysis of problems such as coalition formation and negotiation. Commonly used MARL algorithms include Q-learning and deep reinforcement learning, which are able to model the interactions between agents and update strategies by learning the reward function in the game. These algorithms can be used to analyze and predict behavior and outcomes in areas such as coalition formation and negotiation, thus helping sociologists to better understand and predict trends in various situations in human societies.

As mentioned above, Q-learning is a model-free reinforcement learning algorithm used to find the optimal action-selection policy for any given Markov decision process (MDP) [8]. It works by learning an action-value function, also known as Q-function, that maps a state-action pair to an expected reward. The Q-function is learned through an iterative update process, where the agent interacts with the environment, observes the current state, takes action, and receives a reward. The update rule for Q-learning is given by:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma\, max_{a'} Q(s',a') - Q(s,a)] \tag{1}$$

s and a are the current state and action, respectively
s′ is the next state after taking action an in state s
r is the reward received after taking action a instate s
α is the learning rate, controlling the weight given to new information
γ is the discount factor, controlling the importance of future rewards

The update rule essentially updates the Q-value for the current state-action pair based on the difference between the observed reward and the estimated value of the next state-action pair[9]. Over time, the Q-function converges to the optimal Q-values, which can then be used to select the best action in each state.

2.2.2. Analysis of provision of public goods. In a non-zero-sum game involving public goods provision, the Nash equilibrium may need to be more efficient due to the free-riding problem. Cooperative mechanisms can be designed to encourage participants to contribute to the public good. MARL algorithms can be used to study the effectiveness of these mechanisms and identify optimal resource allocation solutions. For instance, research has shown that the Nash-Q algorithm can effectively incentivise participants to contribute to the public good while ensuring a fair distribution of the benefits [10]. By utilising MARL algorithms in analysing public goods provision, researchers can gain insights into how to design effective cooperative mechanisms to promote the provision of public goods.

The Nash-Q algorithm is a model-free reinforcement learning algorithm used to find a Nash equilibrium in a non-zero-sum game. In the algorithm, each agent learns a Q-value function, which represents the expected reward for taking a certain action in a certain state. The agents update their Q-values based on their own experiences and the experiences of other agents, and they select actions according to a probability distribution based on their Q-values.

The key idea of the Nash-Q algorithm mentioned above is to incorporate the Nash equilibrium into the learning process. Specifically, the agents learn to play a best response to the average strategy of the other agents, which converges to a Nash equilibrium in the long run [10]. The update rule of the Q-value function in Nash-Q algorithm is as follows:

$$Q_i(s,a) = Q_i(s,a) + \alpha\left[r_i(s,a) + \gamma \sum_{j=1}^{n} p_j(s') \max 1_{a'}\, Q_i(s',a') - Q_i(s,a)\right] \tag{2}$$

where $Q_i(s,a)$ is the Q-value function of agent i in state s taking action a, $r_i(s,a)$ is the immediate reward obtained by agent i in state s taking action a, $p_j(s')$ is the probability of reaching state s' after taking actions according to the current joint strategy, and α and γ are learning rate and discount factor, respectively.

DOI: 10.54254/2755-2721/14/20230787

In non-zero-sum games, the Nash-q algorithm has been optimised several times to deal with unknown deterministic continuous-time linear systems [11]. Through these results, MARL in the field of social sciences can significantly expand the number of people and situations analysed in games and obtain optimal solutions for cooperation.

2.3. Applications in games

Multi-agent reinforcement learning (MARL) has emerged as a vital technique for enabling agents to make intelligent decisions in complex, competitive environments. This section explores two notable applications of MARL in non-zero-sum games: poker and the online multiplayer game Dota 2.

In poker, specifically Texas Hold'em, multiple agents interact in a competitive environment where strategies are formed through imperfect information. DeepStack, a cutting-edge algorithm developed by Moravčík et al. [12], employs MARL to create artificial intelligence capable of defeating professional poker players. By leveraging a combination of deep neural networks and counterfactual regret minimization (CFR), DeepStack evaluates its strategy iteratively, updating its policy based on past experiences. This results in the agent being able to adapt to various opponents and situations, demonstrating the efficacy of MARL in non-zero-sum games.

Another prominent application of MARL is in Dota 2, a complex multiplayer online battle arena (MOBA) game where two teams of five players compete to destroy the opposing team's base. OpenAI's Five, a team of artificial intelligence agents, has showcased the power of MARL by defeating top human players in Dota 2 [13]. OpenAI's Five employs a variant of the Proximal Policy Optimization (PPO) algorithm, enabling the agents to learn from individual and team experiences (The in-game timely calculations screen is shown in Figure 2). By employing MARL, OpenAI's Five has achieved remarkable coordination, strategy formation, and decision-making abilities, surpassing human-level performance.

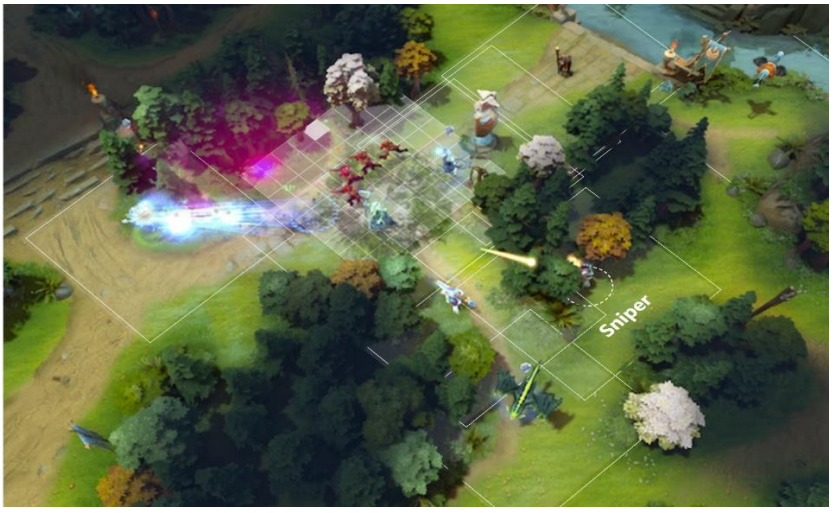

Figure 2. Applications of MARL in games: OpenAI Five in Dota 2 [13].

These two examples underscore the potential of MARL in non-zero-sum games. The success of DeepStack in poker and OpenAI's Five in Dota 2 exemplifies the capability of MARL algorithms to handle complex, competitive environments. These applications highlight the importance of continued research and development of MARL techniques to improve the performance of artificial intelligence agents in non-zero-sum games.

2.4. Potential for future research and applications

The potential for future research and applications of Multi-Agent Reinforcement Learning (MARL) in non-zero-sum games is vast and promising. In order to advance the field, it is essential to explore novel directions and methods to address the challenges in such games.

DOI: 10.54254/2755-2721/14/20230787

One potential direction is the integration of MARL with game theory and mechanism design, which can lead to more effective and efficient solutions for complex real-world problems [14]. Game theory provides a mathematical framework for understanding strategic interactions among multiple agents, while mechanism design focuses on designing the rules of a game to achieve specific objectives. Combining these approaches with MARL makes it possible to develop algorithms that consider the strategic nature of agents and the desired system-wide outcomes.

Another area of future research is the development of MARL algorithms for dynamic and changing environments. Traditional reinforcement learning algorithms often assume a stationary environment, which is only sometimes accurate in real-world scenarios. Developing online learning algorithms that can adapt to changes in the environment and the behaviours of other agents can significantly enhance the applicability of MARL in practical settings [15].

Additionally, the potential applications of MARL in non-zero-sum games span numerous domains, including cybersecurity, transportation, and healthcare. In cybersecurity, MARL can be employed to model the interactions between attackers and defenders, leading to more robust and adaptive defence strategies. In transportation, MARL can be utilized to optimize traffic signal timings or route planning, leading to improved traffic flow and reduced congestion. In healthcare, MARL can be applied to optimize patient treatment plans or allocate resources more efficiently, ultimately leading to better patient outcomes.

The future research and applications of MARL in non-zero-sum games offer numerous opportunities for advancement. The combination of MARL with game theory and mechanism design, the development of online learning algorithms for dynamic environments, and the exploration of applications in various domains can significantly improve the performance and applicability of MARL algorithms.

2.5. Challenges and limitations of using MARL in non-zero-sum games

Multi-Agent Reinforcement Learning (MARL) has demonstrated significant potential in addressing non-zero-sum games. However, several challenges and limitations must be considered when utilizing MARL in such games.

One challenge lies in the complexity of the models and algorithms, which can hinder convergence and scalability [16]. As the number of agents increases, the state and action spaces grow exponentially, making it computationally expensive to find optimal policies.

Another area for improvement is the need for more transparency in agents' decision-making processes. MARL algorithms often produce black-box models that make interpreting and understanding the results challenging. This opaqueness can hinder the adoption of MARL in sensitive applications where explainability is crucial.

Furthermore, the heterogeneity of agents and the environment dynamics create difficulties in developing accurate models. Designing MARL algorithms capable of handling diverse agents with different objectives and adapting to changing environments remains complex.

Lastly, the potential for adversarial attacks and strategic manipulation can compromise the integrity and reliability of the results [17]. Ensuring robustness against such adversarial behaviour is essential to apply MARL in non-zero-sum games successfully.

3. Conclusion

In conclusion, Multi-Agent Reinforcement Learning (MARL) has shown significant promise in addressing non-zero-sum games, with applications spanning various domains such as economics, social sciences, games, and other real-world scenarios. Through its game theory and mechanism design integration, MARL can help develop more effective and efficient solutions for complex problems. Exploring online learning algorithms for dynamic environments can further enhance their applicability.

However, challenges and limitations remain in utilising MARL in non-zero-sum games. These include the complexity of models and algorithms, the need for increased transparency and explainability, heterogeneity of agents and environmental dynamics, and robustness against adversarial attacks and

DOI: 10.54254/2755-2721/14/20230787

strategic manipulation. Addressing these challenges will be crucial for the continued advancement of MARL in non-zero-sum games.

Future research directions should focus on integrating MARL with game theory and mechanism design, developing online learning algorithms for dynamic environments, and exploring applications in various domains. By tackling these challenges and limitations, researchers can unlock the full potential of MARL in non-zero-sum games, enabling better decision-making and improved performance across a wide range of applications. Ultimately, the continued advancement of MARL in non-zero-sum games can lead to a deeper understanding of complex social, economic, and political interactions and contribute to developing more efficient and effective multi-agent systems.

References

[1] J. Subramanian, A. Sinha, and A. Mahajan, Robustness and sample complexity of model-based marl for general-sum markov games, *Dynamic Gam. Appl*., 2020 2: 1–33.

[2] C. Aldaibis, Investigating relaxed probability updating games, Master's thesis, 2022.

[3] J. Cui, Y. Liu, and A. Nallanathan, Multiagent reinforcement learning-based resource allocation for uav networks, *IEEE Trans. Wirel. Comm.*, 2019, 19(2), 729–743.

[4] X. Fang, Q. Zhao, J. Wang, Y. Han, and Y. Li, Multi-agent deep reinforcement learning for distributed energy management and strategy optimization of microgrid market, *Sustain. Citi. Soc*., 2021 17(4) 103163.

[5] J. Hu, M. P. Wellman, et al., Multiagent reinforcement learning: theoretical framework and an algorithm., *Inter. Conf. Mach. Learn*., 1998 98, 242–250.

[6] M. Bowling and M. Veloso, Multiagent learning using a variable learning rate, *Arti. Intel.*, 2002 36(2) 215–250.

[7] D. C. Parkes and M. P. Wellman, Economic reasoning and artificial intelligence, 2015 *Science*, 349 (6245), 267–272.

[8] C. J. Watkins and P. Dayan, Q-learning, Machine learning, 1992 8(1) 279–292.

[9] J. Clifton and E. Laber, Q-learning: Theory and applications, Annual Review of Statistics and Its Application, 2020 (7) 279–301.

[10] J. Hu and M. P. Wellman, Nash q-learning for general-sum stochastic games, *J. mach. Learn. Res*., 2003 11(1) 1039–1069.

[11] K. G. Vamvoudakis, Non-zero sum nash q-learning for unknown deterministic continuous-time linear systems, *Automatica*, 2015 61(1). 274–281.

[12] M. Moravčík, M. Schmid, N. Burch, V. Lisý, D. Morrill, N. Bard, T. Davis, K. Waugh, M. Johanson, and M. Bowling, Deepstack: Expert-level artificial intelligence in heads-up no-limit poker, *Science*, 2017 356(6337) 508–513.

[13] OpenAI, Openai five. https://blog.openai.com/openai-five/.

[14] M. L. Littman, Markov games as a framework for multi-agent reinforcement learning, *Mach. learn. proceed.* 1994 157–163.

[15] S. Omidshafiei, J. Pazis, C. Amato, J. P. How,and J. Vian, Deep decentralized multi-task multi-agent reinforcement learning under partial observability, *Inter. Conf. Mach. Learn.*, 2017 2681–2690.

[16] L. Busoniu, R. Babuska, and B. De Schutter, A comprehensive survey of multiagent reinforcement learning, *IEEE Trans. Sys., Man, Cyber.*, 2008, 38(2), 156–172.

[17] S. Li, Y. Wu, X. Cui, H. Dong, F. Fang, and S. Russell, Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient, *AAAI Conf Arti. Intel.*, 2019 33, 4213–4220.