



Model Based Multi-agent Reinforcement Learning

Ignacio Carlucho[†], Giuseppe Vecchio[‡], Simone Palazzo[‡]

[†]School of Engineering and Physical Sciences, Heriot-Watt University
Campus The Avenue, Edinburgh, UK

[‡]Department of Electrical Electronic and Computer Engineering, University of Catania
Via Santa Sofia 64, Catania 95123, Italy

Email: ignacio.carlucho@ed.ac.uk, giuseppe.vecchio@phd.unict.it, simone.palazzo@unict.it

Multi-agent Reinforcement Learning (MARL) has proven to be a powerful approach for training agents to perform a wide range of tasks [1]. However, traditional MARL methods rely on trial-and-error learning, which can be slow, costly, and can cause damage to the robots when applied to real-world systems, especially during learning stages. On the other hand, Model-Based Multi-agent Reinforcement Learning (MBMARL) uses a model of the environment to generate predictions and plan actions, which can significantly speed up the learning process [2]. This model-generated data can reduce the cost of training considerably, which is particularly helpful in robotic applications.

While the model free MARL domain has been widely studied, MBMARL remains largely unexplored. However, we argue that MBMARL could provide higher data efficiency, and would be safer to train when compared to model free MARL. At the same time, using data-driven modelling techniques does not require any domain knowledge. All these points make these types of methods extremely promising in robotics for the development of control and planning systems.

MBMARL has not been studied in detail with only a limited number of works in the literature [3]. These works can be divided into three broad categories: i) Dyna-style, ii) Dreamer style, and iii) Planning over a learned model. In the first class of methods, both a model and an optimal solution are learnt using transitions from real data, such is the case of MAMBPO [4]. In dreamer types models [5], a model is first learnt and then a solution is trained using only these imaginary transitions, such is the case of M³-UCRL [6]. In the second class of methods, a model is learned, and then it is used to plan for optimal actions. This is similar to what is done in Model Predictive Control (MPC) formulations [7].




One of the main issues in MARL is non-stationary [1]. As agents learn, their policies change, but these changes generate differences between the expectations

individual agents have about the behaviour of other agents in the team. These issues will translate to MBMARL and the generated models. As agents learn their behaviour changes, which changes the models. We believe that utilising shared ensemble models can help reduce these non-stationarities [8].

The other main issue in MBMARL is the need for accurate models. In robotic applications, issues such as partial observability, uncertainty, and compounding errors can degrade the model, affecting the ability of the agent to select the optimal action. New methodologies in data-driving modelling could potentially improve the performance of MBMARL for robotics applications. Particularly, neural network architectures that utilise attention mechanisms, such as transformers, could improve model accuracy. These types of models can better handle sequential data which can enhance coordination [9].

References

- [1] G. Papoudakis, F. Christianos, L. Schäfer, and S. V. Albrecht, “Benchmarking multi-agent deep reinforcement learning algorithms in cooperative tasks,” 2021.
- [2] T. M. Moerland, J. Broekens, A. Plaat, and C. M. Jonker, “Model-based reinforcement learning: A survey.”
- [3] X. Wang, Z. Zhang, and W. Zhang, “Model-based multi-agent reinforcement learning: Recent progress and prospects,” 2022.
- [4] D. Willemsen, M. Coppola, and G. C. H. E. de Croon, “Mambpo: Sample-efficient multi-robot reinforcement learning using learned world models,” 2021.
- [5] D. Hafner, T. Lillicrap, J. Ba, and M. Norouzi, “Dream to control: Learning behaviors by latent imagination,” 2020.
- [6] B. Pasztor, I. Bogunovic, and A. Krause, “Efficient model-based multi-agent mean-field reinforcement learning,” 2021.
- [7] M. Morari and J. H. Lee, “Model predictive control: past, present and future,” vol. 23, no. 4, pp. 667–682.
- [8] R. Mendonca, O. Rybkin, K. Daniilidis, D. Hafner, and D. Pathak, “Discovering and achieving goals via world models,” 2021.
- [9] Y. Tay, M. Dehghani, D. Bahri, and D. Metzler, “Efficient transformers: A survey,” *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–28, 2022.

ORCID iDs Ignacio Carlucho:  0000-0002-6262-480X,
Giuseppe Vecchio:  0000-0001-5009-4365, Simone Palazzo:
 0000-0002-2441-0982



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivatives 4.0 International.