Incentive-based MARL Approach for Commons Dilemmas in Property-based Environments

Extended Abstract

Lukasz Pelcner
Lancaster University
Lancaster, United Kingdom
l.pelcner@lancaster.ac.uk

Matheus Aparecido do Carmo Alves
Lancaster University
Lancaster, United Kingdom
m.a.docarmoalves@lancs.ac.uk

Leandro Soriano Marcolino
Lancaster University
Lancaster, United Kingdom
l.marcolino@lancaster.ac.uk

Paula Harrison
UK Centre for Ecology & Hydrology
Lancaster, United Kingdom
paulaharrison@ceh.ac.uk

Peter Atkinson
Lancaster University
Lancaster, United Kingdom
pma@lancaster.ac.uk

ABSTRACT

We propose *ORAA*, a novel online incentive algorithm that guides agents in a property-based MARL domain to act sustainably with a common pool of resources. *ORAA* uses our proposed P-MADDPG model to learn and make decisions over the decentralised agents. We test our solutions in our novel domain, the "Pollinators' Game", which simulates a property-based MARL scenario and its incentivisation dynamics. We show significant improvement in the incentives' cost-efficiency when using learned models that approximate the behaviour of each agent instead of simulating their true models.

KEYWORDS

Reinforcement Learning; Multi-Agents System; Property-based Environment; Common-Pool Resources

ACM Reference Format:
Lukasz Pelcner, Matheus Aparecido do Carmo Alves, Leandro Soriano Marcolino, Paula Harrison, and Peter Atkinson. 2024. Incentive-based MARL Approach for Commons Dilemmas in Property-based Environments: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024,* IFAAMAS, 3 pages.

1 INTRODUCTION

In history, the effective management and distribution of sustainable resources have been pivotal in shaping societies. These resources include not only hard materials, such as water and wood extracted from groundwater basins and forested areas but also the living components within ecosystems, which are vital for regulating the ecosystem's trophic and pollination levels. These resources together are denominated as common-pool resources (CPRs) and are crucial for maintaining well-being, mainly in local communities. However, achieving a harmonious and equitable utilisation of CPRs remains a relevant challenge for state-of-the-art decision theory.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

MAS is commonly used to model the social dilemmas problem [1, 9–11]. In this context, stochastic models have been used widely to simulate the dynamics between agents [3–5]. Balaguer et al. (2022) [2] introduced a problem where learning agents have full access to information about other agents and the stochastic model for the domain, while a regulatory agent steers their behaviour into equilibrium (agents profit and CPR sustainability). However, in real domains, it may be a strong assumption to assume that learning agents have full access to information about other agents.

In contrast, Yi et al. (2021) [12] presents LToS, a hierarchically decentralised learning framework focused on incentivising cooperative behaviour among agents in networked Multi-Agent RL (MARL) settings. LToS operates on two levels of learning—regulatory and decentralised agent levels—similar to our approach. However, LToS employs a DDPG algorithm [7] for its high-level policy to learn reward-sharing strategies and objective decomposition while employing a DGN [6] for its low-level policy to optimise local objectives. Moreover, the system treats reward as an infinite resource which is undesirable for limited by budget institutions.

We present **Online Regulatory Agent Algorithm (ORAA)**, a novel online learning and planning algorithm that optimises and establishes global regulations to ensure the longevity of a target community using incentives dynamically defined by a regulatory agent. ORAA uses our novel version of MADDPG [8], the P-MADDPG, which properly models and simulates the property-based environment. ORAA is capable of learning models for agents and planning the best incentives to deliver to the community. For our experiments, we propose the ***"Pollinators Game"***, a novel problem that implements the interactions between decentralised agents and a regulatory agent in the context of a community of landowners who can decide how to utilise their land, balancing productivity and pollinators' sustainability. We performed various experiments which consider the heterogeneity or homogeneity between the decentralised agent and the regulatory agent. Our results present a significant improvement in the efficiency when the regulatory agent makes decisions based on estimated models, instead of accessing directly the decentralised agents' true model. We increase the agents' personal reward while still achieving the defined sustainability target and reducing the budget spending by up to 50%.

2 METHODOLOGY

Problem Description. *"A random group of farmers are invited by the government to live in a collective land. Each farmer receives a portion of this land with the only duty of planting and selling food to the government without harming the native pollinators. Therefore, each landowner can decide how to manage their land, but they need to manage the provision to pollinators and maintain the long-term sustainability of the community. Since they do not know each other and there is no previous organisation between them, each landowner decides independently which part of their land will be dedicated to the pollinators and which part to cultivate food. The greater the land designated for pollinators, the greater the sustainability, however, increasing the area of the land for pollinators means reducing the profit received by selling products in the short term."*

2.1 Online Regulatory Agent Algorithm

ORAA is a novel online planning algorithm capable of optimising the delivery of incentives in our property-based problem. It is a Monte Carlo-inspired approach for optimisation that samples and tests several incentives, simulates how agents react to each incentive within their property and, then, estimates the quality of each possibility to take the best action in the real world.

Agents' Simulation Approach. Considering the necessity to simulate the decentralised agents' actions to perform our incentivisation process, we propose two different approaches to model and simulate the interactions between ORAA and the agents:
(i) the Omniscient approach, which assumes the simulation of each agent's actions $\mathbf{a}_i$ using the true model of the agent ϕ_i.
(ii) the Model-Based approach, where we estimate the agents' behaviours using trained networks as a model (i.e., we assess and simulate the agents' actions $\mathbf{a}_i$ using self-trained networks for each decentralised agent ϕ_i in the environment). The regulatory agent trains neural networks to simulate each agent.

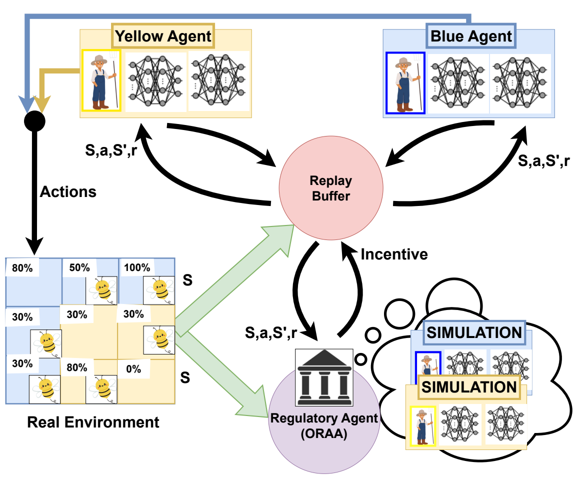


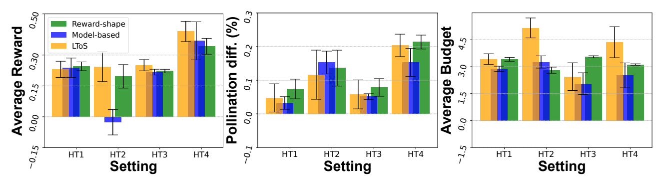
Figure 1: Simplified schematic representing our architecture and the interactions between entities to run ORAA.

2.2 P-MADDPG

Algorithm Outline. *Property-based MADDPG (P-MADDPG)* is a modified version of the traditional MADDPG [8] that we created for the optimisation and performance of the MARL decision-making process in property-based problems. Our algorithm models the decentralised agents' actions and rewards per cell, besides considering a multi-objective reward function (linearised by an α constant). Directly, P-MADDPG tries to address a relevant gap in the literature: *the shifting of dynamics from agents' movements to the dynamics of agents' property with applicability to the reality.* As presented earlier, in the property-based environment we have static agents (in terms of movement) that are capable of modifying their cells' parameters in an online manner by performing different activities in different properties and with different objectives. This characteristic makes the decision-making process of each agent unique and changes the application of the traditional MADDPG . Hence, we modified and extended it to our domain, presenting the P-MADDPG.

3 RESULTS

Heterogeneous Community (HT+HTC). Figure 2 shows the results for experiments where landowners are heterogeneous, i.e., each agent has a different α parameter which means they weigh local and global rewards differently according to their own α value.



(a) Average reward **(b) Pollination difference** **(c) Average budget delivery**

Figure 2: HT+HTC settings result.

Overall Result. ORAA increased the individual achievement of the agent's personal goals by 5% compared to the **Reward-shaping** baseline. **P-MADDPG** surpassed the original pollinator count's target by 6.7% and reduced the budget spending by 50% compared to the **Reward-Shaping** baseline. A smaller reduction is also observed when comparing our results against **LToS**, up to 23%.

4 CONCLUSIONS

We propose ORAA, a novel online planning and incentivisation algorithm for property-based MARL domains. We consider rewards in reinforcement learning as a precious resource. Our objective was to minimise the reliance on reward incentives, yet achieve agent success comparable to LToS, using over 20% less reward (in the form of incentive). Introduced a versatile algorithm capable of Online Planning and Learning, we demonstrated its effectiveness in a novel and relevant environment.

REFERENCES

[1] Hamoud Alhazmi, Ahmed Imran, and Mohammad Abu Alsheikh. 2021. Digital Divide and Social Dilemma of Privacy Preservation. https://doi.org/10.48550/ARXIV.2110.02669

[2] Jan Balaguer, Raphael Koster, Christopher Summerfield, and Andrea Tacchetti. 2022. The Good Shepherd: An Oracle Agent for Mechanism Design. https://doi.org/10.48550/ARXIV.2202.10135

[3] C.Daniel Batson. 2012. *A history of prosocial research.* Psychology Press, In Handbook of the History of Social Psychology. 243–264 pages.

[4] Zhenbo Cheng, Xingguang Liu, Leilei Zhang, Hangcheng Meng, Qin Li, and Xiao Gang. 2021. Improved cooperation by balancing exploration and exploitation in intertemporal social dilemma tasks. https://doi.org/10.48550/ARXIV.2111.09152

[5] Daniel B. Cooney, Dylan H. Morris, Simon A. Levin, Daniel I. Rubenstein, and Pawel Romanczuk. 2022. Social dilemmas of sociality due to beneficial and costly contagion. https://doi.org/10.48550/ARXIV.2202.09905

[6] Jiechuan Jiang, Chen Dun, Tiejun Huang, and Zongqing Lu. 2018. Graph convolutional reinforcement learning. *arXiv preprint arXiv:1810.09202* (2018).

[7] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. 2015. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971* (2015).

[8] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. 2017. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. https://doi.org/10.48550/ARXIV.1706.02275

[9] David O'Callaghan and Patrick Mannion. 2021. Exploring the Impact of Tunable Agents in Sequential Social Dilemmas. https://doi.org/10.48550/ARXIV.2101.11967

[10] Julien Perolat, Joel Z. Leibo, Vinicius Zambaldi, Charles Beattie, Karl Tuyls, and Thore Graepel. 2017. A multi-agent reinforcement learning model of common-pool resource appropriation. https://doi.org/10.48550/ARXIV.1707.06600

[11] Arnu Pretorius, Scott Cameron, Elan van Biljon, Tom Makkink, Shahil Mawjee, Jeremy du Plessis, Jonathan Shock, Alexandre Laterre, and Karim Beguir. 2020. A game-theoretic analysis of networked system control for common-pool resource management using multi-agent reinforcement learning. https://doi.org/10.48550/ARXIV.2010.07777

[12] Yuxuan Yi, Ge Li, Yaowei Wang, and Zongqing Lu. 2021. Learning to Share in Multi-Agent Reinforcement Learning. *arXiv preprint arXiv:2112.08702* , (2021), 16.