Incentive-driven Multi-Agent Reinforcement Learning Approach for Commons Dilemmas in Land-Use

Lukasz Pelcner[1 [0000−0002−6644−2296]], Matheus Aparecido do Carmo Alves[2 [0000−0003−4530−3331]], Leandro Soriano Marcolino[1 [0000−0002−3337−8611]], Paula Harrison[3 [0000−0002−9873−3338]], and Peter Atkinson[1 [0000−0002−5489−6880]]

[1] **Lancaster University, Bailrigg, Lancaster LA1 4YW, United Kingdom**
l.pelcner,l.marcolino,pma@lancaster.ac.uk
[2] **University of São Paulo, São Carlos, SP, 13566-590, Brazil**
matheus.aparecido.alves@usp.br
[3] **UK Centre for Ecology & Hydrology, Bailrigg, Lancaster LA1 4AP, United Kingdom**
paulaharrison@ceh.ac.uk

Abstract. We propose *ORAA*, a novel incentive-driven algorithm that guides agents in a property-based Multi-Agent Reinforcement Learning domain to act sustainably considering a common pool of resources in an online manner. *ORAA* implements 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 scenario and the incentivisation dynamics. We show significant improvement in the incentives' cost-efficiency, reducing the budget spent while increasing the collection of rewards by individual agents. Besides that, our application shows better results when using learned (approximated) models instead of using and simulating the true models of each agent for planning, saving up to 50% of the available budget for incentivisation.

Keywords: Multi-Agent · Reinforcement Learning · Commons Dilemmas · Incentivisation · Property-based Model.

1 Introduction

Historically, the management of common-pool resources (CPRs) has been crucial for maintaining social well-being. These resources include not only hard materials such as water and wood but also living components of ecosystems, among which pollinators like bees are particularly important. However, achieving a harmonious and equitable utilisation of CPRs remains a relevant challenge for the current state-of-the-art.

When a CPR is depleted or seriously wasted, we face a common dilemma named the "tragedy of the commons" [5]. This problem is often caused by the over-appropriation of CPRs by a group and the inherent self-interest of its constituent agents [3], which in the absence of effective supervision, tends to exploit the shared resource to attend to individual objectives and gains. To avoid it, the proposal of an accurate model to represent the spatial and temporal dynamics of such systems is necessary.

A possible solution for this problem is the design of an incentivisation system to handle such a complex Multi-Agent setting. Reward shaping and Multi-Agent RL

(MARL) techniques are commonly used to simulate the delivery of incentives in the environment [1, 4, 6]. For example, Yang et al. (2020) [7] propose shaping the rewards as incentives and embedding them into the knowledge of an ad-hoc agent to increase local cooperation. On the other hand, Perolat et al. (2017) [3] propose the implementation of an exclusion mechanism that enables individual agents (running a RL model) to prevent other agents from accumulating resources, if they are inside their exclusion zone, turning the problem into a competition game that guarantees a sustainable/equal distribution of the CPR in the environment.

In light of this context, we propose the following in this short paper:

- **Online Regulatory Agent Algorithm (ORAA):** a novel online learning and planning algorithm that optimises and establishes incentives for a target community to balance the longevity of CPRs and agents' income;
- **Property-based MADDPG (P-MADDPG):** a modified version of MADDPG that properly models and simulates the property-based environment, and;
- **Pollinators Game:** a novel problem that simulates the aforementioned context as a community of landowners who can decide how to utilise their land, balancing productivity and sustainability, while receiving incentives from a government.

These contributions explore and implement further mechanisms, in comparison to the literature, to enable the insertion of a regulatory agent (government) in an incentivisation system, capable of handling the predatory behaviour of decentralised agents. On this matter, our novel approach, ORAA, shows better cost-efficiency in terms of budget spent and incentive delivered. Our results present a significant improvement in terms of sustainability when the regulatory agent makes decisions based on estimated models for each agent, instead of directly accessing the decentralised agents' true model. We were also able to increase the agents' personal reward, i.e., the income per land while achieving the defined sustainability target while reducing the budget spending by up to 50% in our best case across different settings. All the appendices are available at GitHub[4], including an extended version of this work, with additional discussions about our solution, technical details of our contributions, pseudo-codes and further results.

2 Methodology

Problem Introduction We propose and study the *Pollinators' Game*, which implements a problem that considers a property-based environment where agents control different parts of the environment and make local decisions to modify it. The following story describes the context that inspires our game:

"A random group of farmers are invited by the government to live in a wide 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 on their property. 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

[4] GitHub webpage: `https://github.com/lsmcolab/oraa`

between them, each landowner decides independently which portion of the land will be dedicated to the pollinators and which portion to cultivate food. The higher the portion of the land designated for pollinators, the higher the sustainability, however, increasing the size of the land for pollinators means reducing the profit received by selling products in the short term. How should the agents organise themselves towards sustainability?"

In this context, we propose the *Pollinators Game* to test and evaluate incentive systems. Its model model extends a *n*-player Partially Observable Stochastic Games (POSGs) considering MARL applications. We refer to the reader our Technical Appendix and our GitHub webpage for more details about it.

2.1 Property-based MAADPG

Overview In the property-based environment we have static agents (in terms of movement) which 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 turns the application of the traditional MADDPG unfeasible without modification. Hence, we present P-MADDPG. *P-MADDPG* is a modified version of the traditional MADDPG algorithm [2] focused on 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' movement to the dynamics of agents' property*. To do so, it implements an actor-critic approach that focus on evaluating and optimising the regulatory and the decentralised agents in real-time.

2.2 Online Regulatory Agent Algorithm

Overview *ORAA* is a novel online planning algorithm that optimises 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 estimates the quality of each possibility to take the best action. To do so, we propose two different approaches to model and simulate the interactions between ORAA and the agents: the Omniscient approach, which assumes the simulation of each agent's actions using their true model, and; the Model-Based approach, where we estimate the agents' behaviours using trained networks as a model.

Straightforwardly, the regulatory agent trains neural networks to simulate each agent and perform one of this approaches. Both **P-MADDPG** and **ORAA** represent sequential decision-making processes, which focus on supporting the estimation of the best incentive by enabling the performance of successive simulations of the environment. The application of both solutions together can improve performance in property-based scenarios. We present their pseudo-codes in Technical Appendix in **Section A** and Figure 1 presents the general schematic to understand our training process.

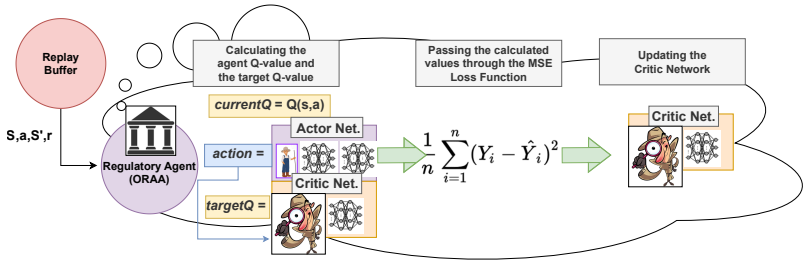


Fig. 1: General training schematic. The difference between ORAA and P-MADDPG's training is the output neural network (a critic or an actor network, respectively).

3 Experiments

Agents Types We define two different types for the landowners and the government:
(i) Homogeneous community (HM): All landowners share the same parameters, meaning all agents use the same weight to balance local and global rewards.
(ii) Heterogeneous community (HT): Each landowner has a different parameters, i.e., they weigh local and global rewards differently according to their parameters.
(iii) Homogeneity-based regulatory agent (HMC): The government considers a single pollination policy to incentivise the community, i.e., it defines a common permitted (target) percentage of pollinators application across all lands.
(iv) Heterogeneity-based regulatory agent (HTC): The government defines a specific pollination policy for each landowner, i.e., it defines the best target percentage of pollinators application for each landowner and its respective crop fields situation.

Experimental settings We combine each landowner type with each possible government policy to define the groups of analysis using the Pollinators' Game and different settings to test each group. We refer the reader to our Technical Appendix for more details about the weights across settings and agents.
(i) HM group: each "HM#" setting presents a specific α parameter common to every agent in the community, following $HM = \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$.
(ii) HT group: in the "HT#" settings, each landowner acts according to specific (predefined) α parameters, which can change across scenarios for each agent.
(iii) HMC group: the HMC government defines, across all possible experiments, the pollination target percentage equals to 0.35, and is common for every agent.
(iv) HTC group: an HT# scenario with a specific pollination target for each agent in the environment.

Metrics We define three different metrics for analysis:
(i) Average reward obtained by each agent in the domain while performing its individual decision-making;
(ii) Percent difference of pollination between the pollination target (defined by the regulatory agent) and actual landowner application (percentage of land used), and;

(iii) Average budget spending of the government to incentivise landowners and benefit environment sustainability.

We present our Baselines in Technical Appendix in **Section F**.

4 Results

ORAA: Omniscient vs Model-based (Figure 2) – We compare our two proposed approaches for ORAA, highlighting the advantages and limitations of each one. Although the Omniscient ORAA achieves slightly better results in terms of Pollination difference, the average budget spent by it increases over time; a surprising result since the Omniscient agent has access to more knowledge.

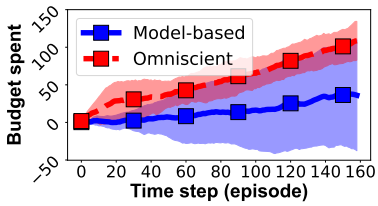
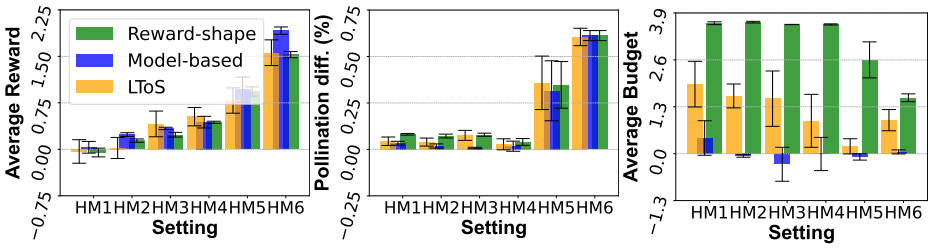


Fig. 2: Average cumulative budget spent in HT+HMC.

Overall result (Figure 3) – ORAA could increase the individual agent's personal goals by 5% compared to the **Reward-shaping** baseline. We surpassed the original pollinator count's target by 6.7% and reduced the budget spending by 50% when comparing our results to the **Reward-Shaping** baseline. A smaller reduction is also observed when comparing our results against **LToS** (literature baseline), up to 23%.



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

Fig. 3: HM+HMC settings result.

5 Conclusions

We propose ORAA, a novel online planning and incentivisation algorithm for property-based MARL domains. Our objective was to minimise the reliance on reward incentives, yet achieve agent success comparable to LToS, using over 20% less incentives. By introducing this versatile online planning and learning algorithm, we demonstrated its effectiveness in a novel and realistic environment, the Pollinators' Game.

Acknowledgements

We want to acknowledge the São Paulo Research Foundation (FAPESP), 2018/15472-9, and Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP)/Petrobras for partially funding this research and resulting paper. We also thank the staff working on the Lancaster University's High End Computing (HEC) Cluster for providing the necessary computational resource and support to this project.

Bibliography

[1] E. Hughes, J. Z. Leibo, M. Phillips, K. Tuyls, E. DueñezGuzman, A. G. Castañeda, I. Dunning, T. Zhu, K. R. McKee, and R. Koster. Inequity aversion improves cooperation in intertemporal social dilemmas. In *NIPS*. 2018.

[2] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments, 2017.

[3] J. Perolat, J. Z. Leibo, V. Zambaldi, C. Beattie, K. Tuyls, and T. Graepel. A multi-agent reinforcement learning model of commonpool resource appropriation. In *NIPS*, page 3643–3652. 2017.

[4] A. Peysakhovich and A. Lerer. Prosocial learning agents solve generalized stag hunts better than selfish ones. In *Proceedings of the 17th AAMAS*, page 2043–2044, , 2018. .

[5] D. J. Rankin, K. Bargum, and H. Kokko. The tragedy of the commons in evolutionary biology. *Trends in ecology & evolution*, 22(12):643–651, 2007.

[6] J. X. Wang, E. Hughes, C. Fernando, W. M. Czarnecki, E. A. Duéñez-Guzmán, and J. Z. Leibo. Evolving intrinsic motivations for altruistic behavior. In *Proceedings of AAMAS'19*, page 683–692, , 2019. .

[7] J. Yang, A. Li, M. Farajtabar, P. Sunehag, E. Hughes, and H. Zha. Learning to incentivize other learning agents. page 20, 2020.