Off-the-Grid MARL: Datasets and Baselines for Offline Multi-Agent Reinforcement Learning

Extended Abstract

Claude Formanek
InstaDeep & University of Cape Town
South Africa
c.formanek@instadeep.com

Asad Jeewa
University of KwaZulu-Natal[0]
South Africa
jeewaa1@ukzn.ac.za

Jonathan Shock
University of Cape Town, NiTheCS & INRS Montreal
South Africa
jonathan.shock@uct.ac.za

Arnu Pretorius
InstaDeep
South Africa
a.pretorius@instadeep.com

ABSTRACT

Being able to harness the power of large, static datasets for developing autonomous multi-agent systems could unlock enormous value for real-world applications. Many important industrial systems are multi-agent in nature and are difficult to model using bespoke simulators. However, in industry, distributed system processes can often be recorded during operation, and large quantities of demonstrative data can be stored. Offline multi-agent reinforcement learning (MARL) provides a promising paradigm for building effective online controllers from static datasets. However, offline MARL is still in its infancy, and, therefore, lacks standardised benchmarks, baselines and evaluation protocols typically found in more mature subfields of RL. This deficiency makes it difficult for the community to sensibly measure progress. In this work, we aim to fill this gap by releasing *off-the-grid MARL (OG-MARL)*: a framework for generating offline MARL datasets and algorithms.

KEYWORDS

Multi-Agent Reinforcement Learning; Offline Reinforcement Learning; Reinforcement Learning

ACM Reference Format:
Claude Formanek, Asad Jeewa, Jonathan Shock, and Arnu Pretorius. 2023. Off-the-Grid MARL: Datasets and Baselines for Offline Multi-Agent Reinforcement Learning: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

Reinforcement learning (RL) has proven to be a powerful computational framework for sequential decision-making, both in single-agent [2, 13, 19], and multi-agent autonomous systems [11, 17, 20]. However, training RL algorithms typically requires extensive online interactions with an environment, making RL impractical for real-world applications. More recently, the field of offline RL has offered a solution to this challenge by bridging the gap between RL and supervised learning, developing algorithms that can leverage large existing datasets of sequential decision-making tasks, to learn optimal control strategies that can be deployed online [10]. Although the field of offline RL has experienced a flurry of research interest in recent years, the focus on offline approaches specific to the multi-agent setting has remained relatively neglected, despite the fact that many real-world problems are naturally formulated as multi-agent systems (e.g. traffic management [23], a fleet of ride-sharing vehicles [20], a network of trains [14] or electricity grid management [8]). The importance of open-access datasets to the progress we have seen in machine learning cannot be understated. Offline RL research in the single agent setting has benefited greatly from the now widely-adopted public datasets and benchmarks available such as D4RL [3] and RL Unplugged [6]. It is essential that multi-agent datasets follow suit since it is currently very challenging to gauge the state of the field and reproduce results from previous work without a common benchmark. Ultimately, to develop new ideas that drive the field forward, a standardised repository of tasks and baselines is required. To fill this gap we present OG-MARL, a framework for dataset generation with baselines for cooperative offline MARL. It is our hope that OG-MARL becomes an ever-growing, evolving repository of offline MARL datasets, that helps foster the development of new offline MARL algorithms, whilst also making it easier for new researchers to enter the field.

2 A FRAMEWORK FOR OFFLINE MARL

In this section, we present our first contribution, OG-MARL as a framework for generating offline MARL datasets. In order to make the generation of datasets easier, we have developed a simple Python package[1] that can be used to wrap any MARL environment with minimal effort to record experiences for new datasets. In addition, we provide a website to host and distribute the OG-MARL datasets[2].

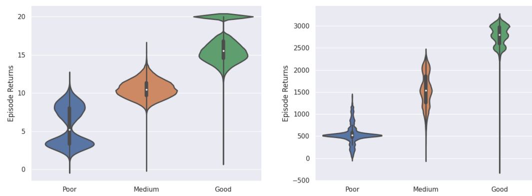
3 DATASETS

In this section, we describe our second contribution: a diverse suite of datasets for cooperative offline MARL. We provide datasets for several popular MARL benchmark environments including the

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

[0]Work done while at InstaDeep.
[1]https://github.com/instadeepai/og-marl
[2]https://sites.google.com/view/og-marl

Starcraft Multi-Agent Challenge (SMAC) [18], Multi-Agent MuJoCo [16], Flatland [14] and environments from PettingZoo [21]. Together these environments cover a broad range of task characteristics including: *i)* discrete and continuous action spaces, *ii)* vector and pixel-based observations, *iii)* dense and sparse rewards, *iv)* a varying number of agents (from 2 to 27 agents), and finally *v)* heterogeneous and homogeneous agents.



(a) SMAC 27m_vs_30m **(b) MAMuJoCo 4-Ant**

Figure 1: Violin plots of the probability distribution of episode returns for selected datasets in OG-MARL. In blue the `Poor` datasets, in orange the `Medium` datasets and in green the `Good` datasets.

For each environment scenario, we provide three types of datasets: `Poor`, `Medium` and `Good`. The dataset types are characterised by the quality of the joint policy that generated the trajectories in the dataset, which is the same approach taken by previous works such as [6, 12, 15, 22]. In Figure 1) we provide violin plots to visualise the distribution of expected episode returns for a sample of the datasets in OG-MARL.

4 BASELINES

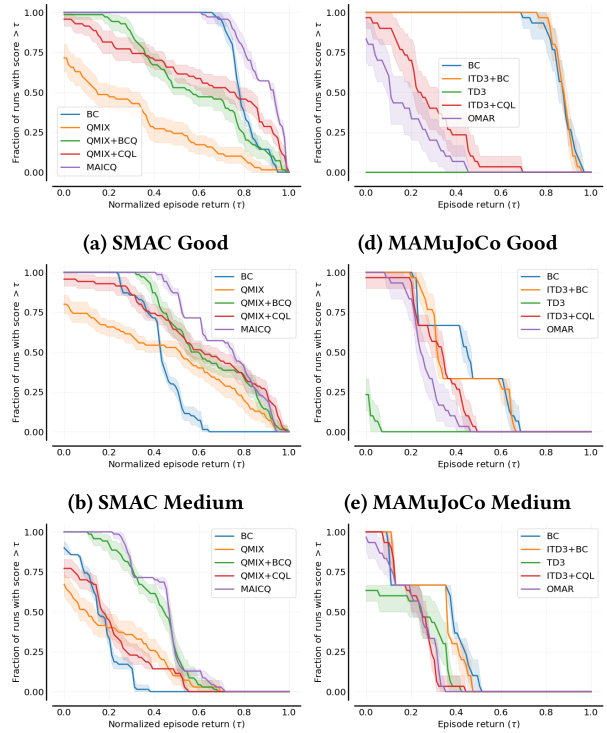
In this section, we describe the third contribution from this work: a stable suite of offline cooperative MARL algorithm implementations. To date there have been a limited number of algorithmic contributions in cooperative offline MARL, but we have included implementations for most of them [15, 22]. In addition, we also propose two new baseline algorithms for cooperative offline MARL: QMIX [17] with Conservative Q-Learning [9] and QMIX with Batch Constrained Q-Learning [4].

Table 1: An overview of cooperative offline MARL algorithms from the literature grouped by the work that proposed them as a novel algorithm or baseline.

Algo Name	Open-Sourced	OG-MARL
MABCQ [7]	✗	✗
MAICQ [22]	✓	✓
DOP+CQL	✗	✗
DOP+BCQ	✗	✗
OMAR [15]	✓	✓
ITD3+CQL	✓	✓
ITD3+BC	✗	✓
MATD3+CQL	✗	✓
MATD3+BC	✗	✓
QMIX+CQL	n/a	✓
QMIX+BCQ	n/a	✓

5 BENCHMARKING

In this section, we present a sample of the results from the benchmarking we performed using our baselines and the datasets in OG-MARL. For each of the benchmarks, SMAC and MAMuJoCo, we aggregate the results across all of the respective tasks using the *MARL-eval* [5] tools. In Figure 2 we give the performance profiles for the `Good`, `Medium` and `Poor` datasets.



(a) SMAC Good **(d) MAMuJoCo Good**

(b) SMAC Medium **(e) MAMuJoCo Medium**

(c) SMAC Poor **(f) MAMuJoCo Poor**

Figure 2: Performance profiles [1] for SMAC and MAMuJoCo. Shaded regions show pointwise 95% confidence bands based on percentile bootstrap with stratified sampling. BC (in blue) is simple behaviour cloning.

6 CONCLUSION

In this extended abstract, we provide a sample of the contributions we made to the field of cooperative offline MARL in our full-paper, which is available on ArXiv [3]. The goal of this work was to highlight the importance of cooperative offline MARL as a research direction in order to make progress towards applying RL to real-world problems. We specifically focused on the lack of standardisation in the field to date, where the absence of a common set of benchmark datasets and baselines is a significant obstacle to progress. To address this issue, we presented a set of relevant and diverse datasets and baselines for offline MARL. It is our hope that the research community will adopt OG-MARL as a framework for offline MARL research and that it helps to drive progress in the field.

[3]https://arxiv.org/abs/2302.00521

ACKNOWLEDGMENTS

We would like to kindly thank the following people for useful discussions and feedback on this work: Ruan de Kock, Omayma Mahjoub and Andries Petrus Smit. As well as Chaima Wichka and Remi Debette for their assistance with running experiments on the compute cluster. We would also like to thank InstaDeep for providing the necessary funding and compute for this research.

REFERENCES

[1] Rishabh Agarwal, Max Schwarzer, Pablo Samuel Castro, Aaron C Courville, and Marc Bellemare. 2021. Deep reinforcement learning at the edge of the statistical precipice. *Advances in Neural Information Processing Systems* 34 (2021).

[2] Adrià Puigdomènech Badia, Bilal Piot, Steven Kapturowski, Pablo Sprechmann, Alex Vitvitskyi, Zhaohan Daniel Guo, and Charles Blundell. 2020. Agent57: Outperforming the atari human benchmark. In *International conference on machine learning*.

[3] Justin Fu, Aviral Kumar, Ofir Nachum, George Tucker, and Sergey Levine. 2020. D4rl: Datasets for deep data-driven reinforcement learning. *arXiv preprint arXiv:2004.07219* (2020).

[4] Scott Fujimoto, David Meger, and Doina Precup. 2019. Off-policy deep reinforcement learning without exploration. In *International conference on machine learning*.

[5] Rihab Gorsane, Omayma Mahjoub, Ruan John de Kock, Roland Dubb, Siddarth Singh, and Arnu Pretorius. 2022. Towards a Standardised Performance Evaluation Protocol for Cooperative MARL. In *Advances in Neural Information Processing Systems 36 (2022).*

[6] Caglar Gulcehre, Ziyu Wang, Alexander Novikov, Thomas Paine, Sergio Gómez, Konrad Zolna, Rishabh Agarwal, Josh S Merel, Daniel J Mankowitz, Cosmin Paduraru, et al. 2020. Rl unplugged: A suite of benchmarks for offline reinforcement learning. *Advances in Neural Information Processing Systems* 33 (2020).

[7] Jiechuan Jiang and Zongqing Lu. 2021. Offline decentralized multi-agent reinforcement learning. *arXiv preprint arXiv:2108.01832* (2021).

[8] Vanshaj Khattar and Ming Jin. 2022. Winning the CityLearn Challenge: Adaptive Optimization with Evolutionary Search under Trajectory-based Guidance.

[9] Aviral Kumar, Aurick Zhou, George Tucker, and Sergey Levine. 2020. Conservative Q-learning for offline reinforcement learning. *Advances in Neural Information Processing Systems* 33 (2020).

[10] Sergey Levine, Aviral Kumar, George Tucker, and Justin Fu. 2020. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643* (2020).

[11] Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems* 30 (2017).

[12] Linghui Meng, Muning Wen, Yaodong Yang, Chenyang Le, Xiyun Li, Weinan Zhang, Ying Wen, Haifeng Zhang, Jun Wang, and Bo Xu. 2021. Offline Pre-trained Multi-Agent Decision Transformer: One Big Sequence Model Conquers All StarCraftII Tasks. *arXiv preprint arXiv:2112.02845* (2021).

[13] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. 2015. Human-level control through deep reinforcement learning. *Nature* 518 (2015).

[14] Sharada Mohanty, Erik Nygren, Florian Laurent, Manuel Schneider, Christian Scheller, Nilabha Bhattacharya, Jeremy Watson, Adrian Egli, Christian Eichenberger, Christian Baumberger, Gereon Vienken, Irene Sturm, Guillaume Sartoretti, and Giacomo Spigler. 2020. Flatland-RL : Multi-Agent Reinforcement Learning on Trains. https://doi.org/10.48550/ARXIV.2012.05893

[15] Ling Pan, Longbo Huang, Tengyu Ma, and Huazhe Xu. 2022. Plan better amid conservatism: Offline multi-agent reinforcement learning with actor rectification. In *International Conference on Machine Learning*.

[16] Bei Peng, Tabish Rashid, Christian Schroeder de Witt, Pierre-Alexandre Kamienny, Philip Torr, Wendelin Böhmer, and Shimon Whiteson. 2021. Facmac: Factored multi-agent centralised policy gradients. *Advances in Neural Information Processing Systems* 34 (2021).

[17] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2018. Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning. In *International Conference on Machine Learning*.

[18] Mikayel Samvelyan, Tabish Rashid, Christian Schroeder de Witt, Gregory Farquhar, Nantas Nardelli, Tim GJ Rudner, Chia-Man Hung, Philip HS Torr, Jakob Foerster, and Shimon Whiteson. 2019. The StarCraft Multi-Agent Challenge. In *International Conference on Autonomous Agents and MultiAgent Systems.*

[19] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. 2016. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* 529 (2016).

[20] Quinlan Sykora, Mengye Ren, and Raquel Urtasun. 2020. Multi-Agent Routing Value Iteration Network. In *International Conference on Machine Learning.*

[21] J Terry, Benjamin Black, Nathaniel Grammel, Mario Jayakumar, Ananth Hari, Ryan Sullivan, Luis S Santos, Clemens Dieffendahl, Caroline Horsch, Rodrigo Perez-Vicente, et al. 2021. Pettingzoo: Gym for multi-agent reinforcement learning. *Advances in Neural Information Processing Systems* 34 (2021).

[22] Yiqin Yang, Xiaoteng Ma, Li Chenghao, Zewu Zheng, Qiyuan Zhang, Gao Huang, Jun Yang, and Qianchuan Zhao. 2021. Believe what you see: Implicit constraint approach for offline multi-agent reinforcement learning. *Advances in Neural Information Processing Systems* 34 (2021).

[23] Huichu Zhang, Siyuan Feng, Chang Liu, Yaoyao Ding, Yichen Zhu, Zihan Zhou, Weinan Zhang, Yong Yu, Haiming Jin, and Zhenhui Li. 2019. CityFlow: A Multi-Agent Reinforcement Learning Environment for Large Scale City Traffic Scenario. In *The World Wide Web Conference.*