Past to Present: Reassessing Evaluation in Multi-Agent Reinforcement Learning

Anonymous Author(s)
Affiliation
Address email

Abstract

Establishing sound experimental standards and rigour is important in any growing field of research. Deep Multi-Agent Reinforcement Learning (MARL) is one such nascent field. Although exciting progress has been made, MARL has recently come under scrutiny for replicability issues and a lack of standardised evaluation methodology, specifically in the cooperative setting. Although protocols have been proposed to help alleviate the issue, it remains important to actively monitor the health of the field. In this work, we extend the database of evaluation methodology previously published by Gorsane et al. [2022] containing meta-data on MARL publications from top-rated conferences and compare the findings extracted from this updated database to the trends identified in their work. Our analysis shows that many of the worrying trends in performance reporting remain. This includes the omission of uncertainty quantification, not reporting all relevant evaluation details and a narrowing of algorithmic development classes. Promisingly, we do observe a trend towards more difficult scenarios in SMAC-v1, which if continued into SMAC-v2 will encourage novel algorithmic development. Our data indicate that replicability needs to be approached more proactively by the MARL community to ensure trust in the field as we move towards exciting new frontiers.

1 Introduction

Multi-Agent Reinforcement Learning (MARL) is a rapidly growing field that has gained more attention in recent years. However, similar to other fields within machine learning, ensuring a high level of scientific rigour and sound experimental methodology has become difficult as algorithms become more complex and computational requirements to perform experiments more extreme[Colas et al., 2018, Henderson et al., 2018, Colas et al., 2019, Engstrom et al., 2020, Jordan et al., 2020, Agarwal et al., 2022].

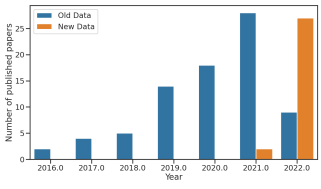


Figure 1: Recorded papers by year in the meta-analysis on evaluation methodologies in cooperative MARL.

Gorsane et al. [2022] compiled a dataset of evaluation methodologies used in deep cooperative MARL, from the first paper published in January 2016 up until April 2022. The authors discovered some worrying trends in research w.r.t the consistency of reported results. As the number of publications grows in the coming years, it is important to update this dataset with new publications and perform a continuous assessment of the field. Although the original analysis of [Gorsane et al., 2022] highlighted many issues, there was evidence of a more thorough evaluation beginning to gain traction in more recent years, with an increase in the number of ablation studies and an upward trend in the use of uncertainty quantification.

5th Deep Learning Indaba Conference (DLI 2023).

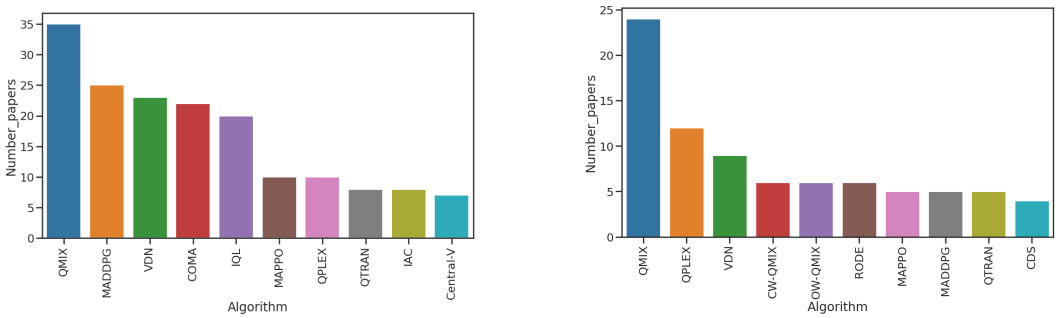
By supplementing this prior analysis with only 11 months of new data we show that these trends have shifted significantly from the original findings.

The updated dataset contains an additional 29 papers published at top-rated conferences (e.g. NeurIPS, ICLR, ICML etc.) during the period of April 2022 to December 2022 [1].

2 Historical trends vs new data

Inconsistencies in performance and a lack of ablation studies. The analysis of recent data indicates that the landscape of common algorithms has shifted greatly from historical analysis as illustrated in figure 2. Previous historic algorithms like COMA [Foerster et al., 2018] and MADDPG [Lowe et al., 2017] have lost popularity as baselines with none of the reviewed papers making use of COMA and MADDPG used in only 25.8% of new publications vs a historical use of 35.2%. As these methods are unable to achieve reasonable performance on current common benchmarks their decline reflects researchers moving towards more modern baselines [Papoudakis et al., 2021]. Qmix [Rashid et al., 2018] remains a strong baseline for newer cooperative MARL algorithms which take inspiration from it. Importantly Qmix is still competitive with newer methods when parameterised correctly and maintains relevance on modern benchmarks [Hu et al., 2021].

Despite the discovery of the effectiveness of Proximal Policy Optimisation (PPO) [Schulman et al., 2017] to match the more complex value-decomposition methods in performance and sample efficiency [Yu et al., 2021], Policy Gradient (PG) approaches still seem unpopular. With both the centralised training and decentralised execution (CTDE) and decentralised training and decentralised execution (DTDE) variants of the Advantage Actor-Critic (A2C) [Mnih et al., 2016], Independent Actor-Critic (IAC) and Central-V [Foerster et al., 2018] not being present in newer publications. However multi-agent PPO (MAPPO) was still used in 16.1% of the new papers and will hopefully gain traction as a common baseline. Another notable development is the decline of Independent Learners (IL). Although normally used as baselines, the findings in [Papoudakis et al., 2021] show that they are important to assess the trade-offs that newer methods may have in certain settings.



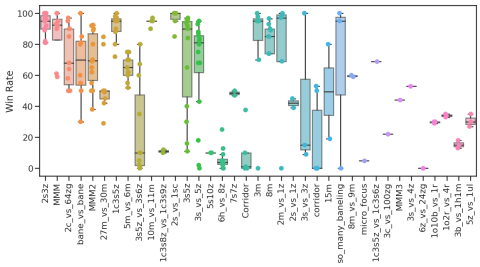
(a) Algorithm usage from pre-2022 database

(b) Algorithm usage from recent database

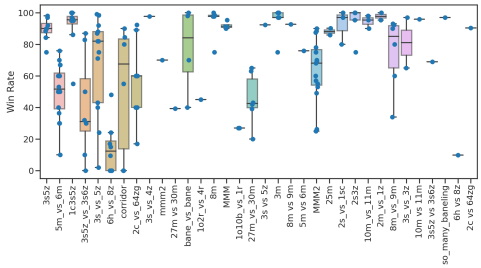
Figure 2: Number of occurrences of algorithms from historic and recent data

In the early years of MARL, there was naturally a high variance in results even when using the same algorithm. Initially, standardised frameworks were uncommon and MARL algorithms suffered from high implementation variance. Fortunately, MARL experienced some standardisation with the release of the Starcraft Multi-Agent Challenge (SMAC) [Samvelyan et al., 2019] along with the EPyMARL framework [Papoudakis et al., 2021]. Since then it has become and remains the most common benchmark in cooperative MARL by a sizable margin as shown in Figure 5. However, even with this framework and common testing environment, there was still a high variance in reported performance as shown in 3a. From the most recent data, we can see from figure 3b that these trends persist. Seemingly, variance has only reduced significantly in the 'easy' category of scenarios like 8m which are trivial for newer algorithms. In practice, this makes it very difficult to determine the true performance gains of recently developed and future works against the historical baselines they aim to improve on.

[1]The updated dataset is available at:
https://drive.google.com/file/d/1psSUrS8ywPAJY1Qm4ySfTYcYYJs0TlLF/view?usp=drive$_{l}ink$



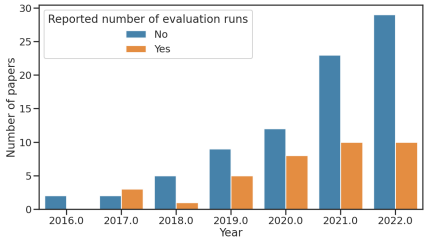
(a) Performance spread for Qmix in SMAC from historical data



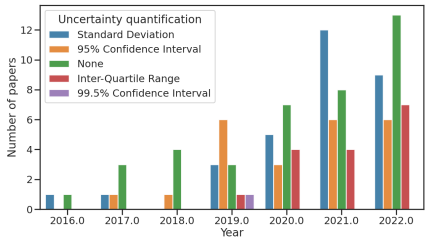
(b) Performance spread for Qmix in SMAC from recent publications

Figure 3: Performance spread for Qmix in SMAC

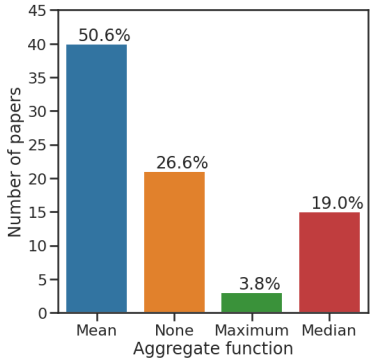
Tends in performance reporting. Metric reporting reached a promising peak of 75% in 2021 and looked to be a rising trend over previous years. However, we see this drop to 63% in 2022 in figure 4a. This is alarming given the goal of using MARL in real-world settings where reliability is often more important than absolute performance. Additionally, aggregation metrics are not present in 20% of recent publications and with the computational complexity of modern MARL, it is very difficult to evaluate over enough seeds to account for the variance in performance over different runs [Agarwal et al., 2022].



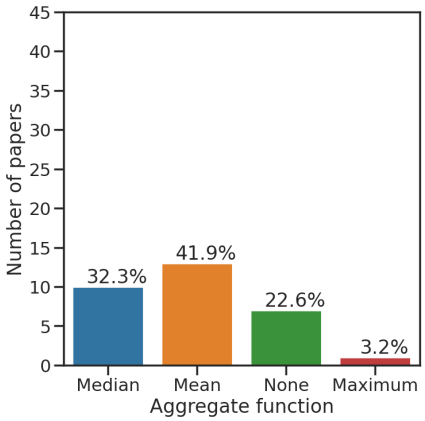
(a) Historical tendencies for reporting evaluation runs



(b) Historical tendencies for reporting uncertainty metrics



(c) Aggregation function historical usage



(d) Aggregation function recent publications

Environment usage trends. As mentioned previously, SMAC is the most popular environment in MARL by a large margin followed by MPE which can be seen in figure 5. Evidence suggests that these settings have both reached a point of being over-fit to[Papoudakis et al., 2021, Gorsane et al., 2022] which makes determining algorithmic ranking difficult. Newer settings have generally not gained much popularity and despite the insight RWARE and LBF were shown to provide on MARL algorithms by [Papoudakis et al., 2021] recent data shows they are underutilised. However, environment overfitting on SMAC may be naturally resolved as users move over to SMAC-v2 [Ellis et al., 2022] which provides a new host of challenges through a well-understood interface.

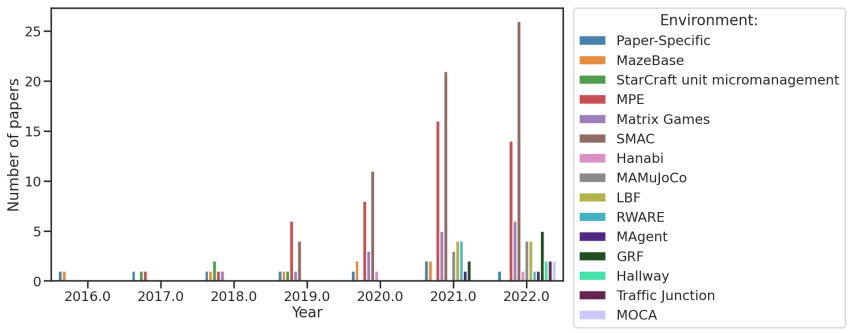
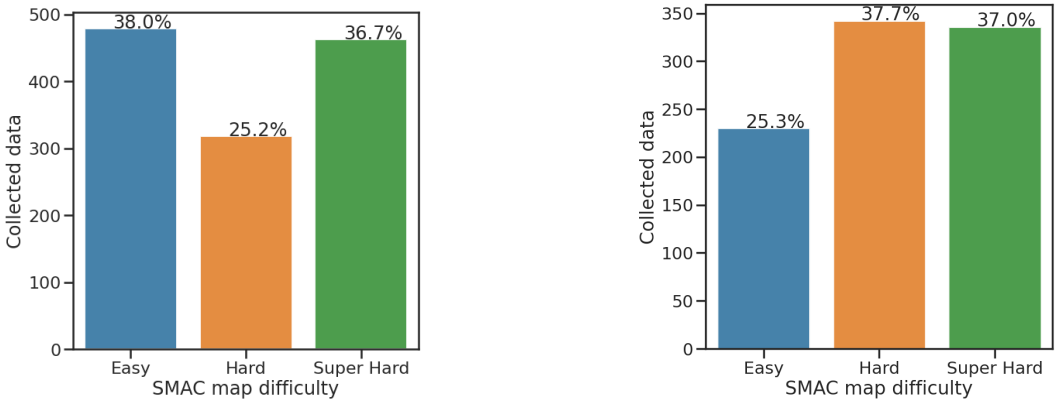


Figure 5: Historic environment usage trends from 2016 to 2022

SMAC categorises scenarios into ‘easy‘, ‘hard‘ and ‘super hard‘ difficulties. Over time, publications have been able to render no in the ‘super hard‘ category as trivial to solve. Due to the large number of scenarios in SMAC, authors will often use a subset of the scenarios to reduce compute time which produces misleading aggregate performance[Gorsane et al., 2022]. A helpful development seen in figure 6 is the reduction in the use of the ‘easy‘ scenarios as a focus on difficult challenges streamlines the total set of evaluation scenarios required for comparative evaluation.



(a) Historic difficulty of SMAC scenarios used in publications

(b) Recent difficulty of SMAC scenarios used in publications

Figure 6: Percentage usage of different SMAC scenario difficulties

3 Conclusion

We compiled 29 papers from the latest MARL publications to extend the MARL publication database to take into account recent development trends. With only 11 months of data, we can see marked changes from historical data.

Despite a push for increased rigour, certain worrying historical trends are still prevalent in MARL w.r.t replicability. Qmix, despite its longstanding use as a baseline, still shows large variance across different publications, the details required to replicate evaluation methodologies are often unreported and IL baselines are absent. Together, these issues make it difficult to get a true measurement of algorithmic development against historic algorithms. Additionally, environmental usage also seems to have not yet diversified and research is focused on SMAC and MPE. However, authors do seem to be moving towards a more streamlined set of scenarios which will reduce computational overhead and aid better evaluation. Given these trends, it is important that replicability is approached proactively by the MARL community to ensure trust in the field is maintained as it grows.

References

R. Agarwal, M. Schwarzer, P. S. Castro, A. Courville, and M. G. Bellemare. Deep Reinforcement Learning at the Edge of the Statistical Precipice, 2022.

C. Colas, O. Sigaud, and P.-Y. Oudeyer. How Many Random Seeds? Statistical Power Analysis in Deep Reinforcement Learning Experiments, 2018.

C. Colas, O. Sigaud, and P.-Y. Oudeyer. A Hitchhiker's Guide to Statistical Comparisons of Reinforcement Learning Algorithms, 2019.

B. Ellis, S. Moalla, M. Samvelyan, M. Sun, A. Mahajan, J. N. Foerster, and S. Whiteson. Smacv2: An improved benchmark for cooperative multi-agent reinforcement learning, 2022. URL `https://arxiv.org/abs/2212.07489`.

L. Engstrom, A. Ilyas, S. Santurkar, D. Tsipras, F. Janoos, L. Rudolph, and A. Madry. Implementation matters in deep policy gradients: A case study on ppo and trpo, 2020.

J. Foerster, G. Farquhar, T. Afouras, N. Nardelli, and S. Whiteson. Counterfactual multi-agent policy gradients. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

R. Gorsane, O. Mahjoub, R. de Kock, R. Dubb, S. Singh, and A. Pretorius. Towards a standardised performance evaluation protocol for cooperative marl, 2022.

P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger. Deep Reinforcement Learning that Matters, 2018.

J. Hu, S. Jiang, S. A. Harding, H. Wu, and S.-w. Liao. Rethinking the implementation tricks and monotonicity constraint in cooperative multi-agent reinforcement learning, 2021. URL `https://arxiv.org/abs/2102.03479`.

S. M. Jordan, Y. Chandak, D. Cohen, M. Zhang, and P. S. Thomas. Evaluating the Performance of Reinforcement Learning Algorithms, 2020.

R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. *Neural Information Processing Systems (NIPS)*, 2017.

V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu. Asynchronous methods for deep reinforcement learning, 2016.

G. Papoudakis, F. Christianos, L. Schäfer, and S. V. Albrecht. Benchmarking multi-agent deep reinforcement learning algorithms in cooperative tasks, 2021.

T. Rashid, M. Samvelyan, C. S. de Witt, G. Farquhar, J. Foerster, and S. Whiteson. Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning, 2018.

M. Samvelyan, T. Rashid, C. S. de Witt, G. Farquhar, N. Nardelli, T. G. J. Rudner, C.-M. Hung, P. H. S. Torr, J. Foerster, and S. Whiteson. The starcraft multi-agent challenge, 2019. URL `https://arxiv.org/abs/1902.04043`.

J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

C. Yu, A. Velu, E. Vinitsky, Y. Wang, A. Bayen, and Y. Wu. The surprising effectiveness of ppo in cooperative, multi-agent games, 2021. URL `https://arxiv.org/abs/2103.01955`.