Towards a Standardised Performance Evaluation Protocol for Cooperative MARL

Rihab Gorsane 1, Omayma Mahjoub 1, Siddarth Singh 1, Ruan de Kock 1, Roland Dubb 3, Wiem Khlifi 1,3, Abidine Vall 1,4

1InstaDeep Ltd, 2National School of Computer Science, Tunisia, 3University of Cape Town, South Africa, 4National School of Engineering of Tunis, Tunisia

Abstract

Multi-agent reinforcement learning (MARL) has emerged as a useful approach to solving decentralised decision-making problems at scale. Research in the field has been growing steadily with many breakthrough algorithms proposed in recent years. In this work, we take a closer look at this rapid development with a focus on evaluation methodologies employed across a large body of research in cooperative MARL. By conducting a detailed meta-analysis of prior work, spanning 104 papers accepted for publication from 2016 to 2022, we bring to light worrying trends that put into question the true rate of progress.

From RL to MARL: Lessons, trends and recommendations

#1 Know the true source of improvement and report everything

There is considerable variance in reported results in MARL. Performance clarity is poor.

Figure 1. Historical performance of QMIX on different SMAC maps across papers.

#2 Use standardised statistical tooling for estimating and reporting uncertainty

For each algorithm, a, there is an evaluation suite with shape elf (a), where each entry is the normalised absolute return for a specific training run on a specific task. 1000 Monte-Carlo samples are used to estimate the normalised return at each evaluation interval.

Figure 2. Statistical tools used across papers: a) Aggregate functions, b) Measure of spread c) Independent runs.

#3 Guard against environment misuse and overfitting

Algorithm A

Algorithm B

Algorithm C

Algorithm D

Figure 3. Reanalysis of original SMAC experiments [1]

Contributions

#1 We have open sourced our dataset and call for contributions from the MARL community as a whole

Figure 4. Environment adoption over time.

#2 Evaluation Tools Repo

We have open-sourced a repo to do all statistical aggregation and produce all plots given raw experiment data in the correct format.

#3 Standardised evaluation protocol

A Standardized Performance Evaluation Protocol for Cooperative MARL

Input: Environments with tasks f from a set F, Algorithms a ∈ A, including baselines and novel work.

1. Evaluation parameters – details
   - Number of training steps, T = 2 million.
   - Number of independent training runs, Rn = 10 from (Agarwal et al., 2022) [1].
   - Number of independent evaluation episodes per interval, Rn = 32.
   - Evaluation intervals, t ∈ T, every 10,000 timesteps.

2. Performance and uncertainty quantification
   - Performance metric: Always use return G (applicable to all environments), and the environment specific metric (e.g. Win rate).
   - Per task evaluation: Compute the mean G over E episodes at each evaluation interval, where G is the return of algorithm a on task t, with 95% CI, for all a.
   - Per environment evaluation:
     - Compute the normalised absolute return Colas et al. 2018 [2] for the mean return of the first 5M evaluation episodes using the best joint policy found during training and normalising the return to be in the range [0, 1] by taking (max returns) - (mean returns), where µt is the return for all algorithms on task t.
     - For each algorithm, a, from an evaluation suite with shape Elf(a), where each entry is the normalised absolute return for a specific training run on a specific task. 1000 Monte-Carlo samples are used to estimate the normalised return at each evaluation interval.

3. Reporting
   - Experiments: All hyperparameters, code-level optimisations, computational requirements and framework details.
   - Plots: All task and evaluation results as well as ablation study results.
   - Tables: Normalised absolute performance per task with 95% CI for all tasks, QMIX with stratified Bootstrap CIs per environment for all environments.
   - Public repository: Raw data and code implementations.

Additional Findings

We propose a standardised performance evaluation protocol, motivated in part by the literature on evaluation in RL, as well as by a meta-analysis of prior work in MARL. It is our hope that, if widely adopted, such a protocol could make comparisons across different works much faster, easier and more accurate leading to sustained progress in the field.

References
