MO-Gym: A Library of Multi-Objective Reinforcement Learning Environments

Lucas N. Alegre[1,2], Florian Felten[3], El-Ghazali Talbi[4,5], Grégoire Danoy[3,5], Ann Nowé[1], Ana L. C. Bazzan[2], and Bruno C. da Silva[6]

[1] Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels, Belgium
[2] Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[3] SnT, University of Luxembourg, Luxembourg
[4] CNRS/CRIStAL, University of Lille, Lille, France
[5] FSTM/DCS, University of Luxembourg, Luxembourg
[6] CICS, University of Massachusetts, Amherst, USA
lnalegre@inf.ufrgs.br, florian.felten@uni.lu

Abstract. We introduce MO-Gym, an extensible library containing a diverse set of multi-objective reinforcement learning environments. It introduces a standardized API that facilitates conducting experiments and performance analyses of algorithms designed to interact with multi-objective Markov decision processes. Importantly, it extends the widely-used OpenAI Gym API, allowing the reuse of algorithms and features that are well-established in the reinforcement learning community. MO-Gym is available at: https://github.com/LucasAlegre/mo-gym.

Keywords: Reinforcement Learning · Multi-Objective Reinforcement Learning · Benchmarking.

1 Introduction

Recent successes of Reinforcement Learning (RL) brought a significant amount of attention to this research area. RL problems are typically modeled via Markov decision processes (MDPs), which allow to specify an objective that the agent should optimize. However, many real-world problems are composed of multiple—often conflicting—objectives, which might not be trivially expressed via a single scalar reward function [10]. As a solution, Multi-Objective RL (MORL) algorithms aim at optimizing sequential decision-making problems involving multiple objectives. Concretely, the difference between single-objective RL and MORL is that in the latter setting, the reward function is vector-valued. Depending on the particular preference (trade-off) of the agent for each objective, different decision-making policies can be optimal. Hence, a MORL algorithm's output may be a single policy, in case the preferences are known *a priori*, or a Pareto set of multiple policies, in cases where preferences are only known *a posteriori* [5].

As recently noticed [5,4], various benchmark problems have been proposed to evaluate MORL methods. These benchmarks have not, however, been made available via a standardized API or via a centralized repository. Arguably, this has made experimental reproducibility harder, time-consuming, and error-prone. MORL-Glue [11] is a first attempt to provide a centralized repository of MORL benchmark implementations. However, this library has not been widely adopted due to the fact that it is implemented in Java and targets tabular problems, whilst the community currently focuses on using Python and deep RL.

OpenAI Gym [3] is a standard API for designing and evaluating RL algorithms, providing various benchmark environments. It is one of the most widely-adopted implementation standards in the RL community and allows researchers to quickly test algorithms on various problems, thus speeding up the experimental phase of algorithm design. However, its API is currently limited to single-objective RL problems. Hence, Gym has been extended in various ways—e.g. PettingZoo [8] for multi-agent RL and Safety Gym [7] for RL with safety constraints. To the best of our knowledge, no extensions of Gym for MORL have been designed. We fill this gap by introducing MO-Gym: a standardized API for designing MORL algorithms and benchmark domains, as well as a centralized and extensible repository of multi-objective MDP implementations.

2 Multi-Objective Gym

MO-Gym is designed to be as close as possible to the original OpenAI Gym API. This allows it to benefit from most of the already-existing features (e.g., *wrappers*, a construction that allows for single aspects/properties of a domain to be modified) while extending the original API only where necessary so that it can be extensible and support a wide-range of MORL benchmark domains. The key difference between these two frameworks is that, in MO-Gym, the reward returned after the execution of each action (i.e., after each call of the *step* method) is a vector instead of a scalar. MO-Gym is available on PyPI and can be installed via `pip install mo-gym`.

Environments. Currently, MO-Gym supports 14 environments commonly used in the MORL literature—including environments with discrete and continuous state and action spaces—such as *deep-sea-treasure* [9,13], *four-room* [2], *mo-supermario* [13], *minecart* [1], and *mo-halfcheetah* [12].

Wrappers. MO-Gym introduces MORL-specific wrappers such as the *LinearReward* wrapper, which linearly scalarizes the reward function of a given environment and transforms a multi-objective problem into a standard MDP. This makes MO-Gym *directly compatible* with all widely-used RL libraries compatible with Gym such as Stable-Baselines [6].

By making MO-Gym open source and extensible—and thus open to contributions from other researchers—our hope is to provide a solid base for reproducible research in MORL. Currently, we are also developing *MORL Baselines*, a library of reliable implementations of MORL algorithms compatible with MO-Gym. We hope our work will allow researchers to seamlessly deploy existing algorithms on various MORL domains, speeding up experiments, facilitating algorithms comparisons, and being more conducive to reproducible experimental results.

Acknowledgments. This work was funded by: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001; CNPq (Grants 140500/2021-9, 304932/2021-3); FAPESP/MCTI/CGI (Grant 2020/05165-1); the Fonds National de la Recherche Luxembourg (FNR), CORE program under the ADARS Project, ref. C20/IS/14762457; the Research Foundation Flanders (FWO) [G062819N]; the AI Research Program from the Flemish Government (Belgium); and the Francqui Foundation.

References

1. Abels, A., Roijers, D.M., Lenaerts, T., Nowé, A., Steckelmacher, D.: Dynamic weights in multi-objective deep reinforcement learning. In: Proceedings of the 36th International Conference on Machine Learning. vol. 97, pp. 11–20. International Machine Learning Society (IMLS) (2019)
2. Alegre, L.N., Bazzan, A.L.C., da Silva, B.C.: Optimistic linear support and successor features as a basis for optimal policy transfer. In: Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., Sabato, S. (eds.) Proceedings of the 39th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 162, pp. 394–413. PMLR (17–23 Jul 2022), https://proceedings.mlr.press/v162/alegre22a.html
3. Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., Zaremba, W.: OpenAI Gym (2016), arXiv preprint arXiv:1606.01540
4. Cassimon, T., Eyckerman, R., Mercelis, S., Latré, S., Hellinckx, P.: A survey on discrete multi-objective reinforcement learning benchmarks. In: Proceedings of the Adaptive and Learning Agents Workshop (ALA 2022) (2022)
5. Hayes, C.F., Rădulescu, R., Bargiacchi, E., Källström, J., Macfarlane, M., Reymond, M., Verstraeten, T., Zintgraf, L.M., Dazeley, R., Heintz, F., Howley, E., Irissappane, A.A., Mannion, P., Nowé, A., Ramos, G., Restelli, M., Vamplew, P., Roijers, D.M.: A practical guide to multi-objective reinforcement learning and planning. Autonomous Agents and Multi-Agent Systems **36**(1), 26 (Apr 2022). https://doi.org/10.1007/s10458-022-09552-y, https://doi.org/10.1007/s10458-022-09552-y
6. Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., Dormann, N.: Stable-baselines3: Reliable reinforcement learning implementations. Journal of Machine Learning Research **22**(268), 1–8 (2021), http://jmlr.org/papers/v22/20-1364.html
7. Ray, A., Achiam, J., Amodei, D.: Benchmarking safe exploration in deep reinforcement learning. https://cdn.openai.com/safexp-short.pdf (2019)
8. Terry, J.K., Black, B., Hari, A., Santos, L.S., Dieffendahl, C., Williams, N.L., Lokesh, Y., Horsch, C., Ravi, P.: Pettingzoo: Gym for multi-agent reinforcement learning. CoRR **abs/2009.14471** (2020), https://arxiv.org/abs/2009.14471
9. Vamplew, P., Dazeley, R., Berry, A., Issabekov, R., Dekker, E.: Empirical evaluation methods for multiobjective reinforcement learning algorithms. Mach. Learn. **84**(1–2), 51—80 (Jul 2011). https://doi.org/10.1007/s10994-010-5232-5
10. Vamplew, P., Smith, B.J., Källström, J., Ramos, G., Rădulescu, R., Roijers, D.M., Hayes, C.F., Heintz, F., Mannion, P., Libin, P.J.K., Dazeley, R., Foale, C.: Scalar reward is not enough: a response to silver, singh, precup and sutton (2021). Autonomous Agents and Multi-Agent Systems **36**(2), 41 (Jul 2022). https://doi.org/10.1007/s10458-022-09575-5
11. Vamplew, P., Webb, D., Zintgraf, L.M., Roijers, D.M., Dazeley, R., Issabekov, R., Dekker, E.: MORL-Glue: a benchmark suite for multi-objective reinforcement learning. In: Proceedings of the 29th Benelux Conference on Artificial Intelligence. pp. 389–390 (2017), https://pure.rug.nl/ws/portalfiles/portal/107483223/preproceedings_1_.pdf#page=405
12. Xu, J., Tian, Y., Ma, P., Rus, D., Sueda, S., Matusik, W.: Prediction-guided multi-objective reinforcement learning for continuous robot control. In: Proceedings of the 37th International Conference on Machine Learning (ICML) (2020)
13. Yang, R., Sun, X., Narasimhan, K.: A generalized algorithm for multi-objective reinforcement learning and policy adaptation. In: Wallach, H., Larochelle, H.,

Beygelzimer, A., d' Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems 32. pp. 14610–14621 (2019)