MORL-Glue: A Benchmark Suite for Multi-Objective Reinforcement Learning

Peter Vamplew, Dean Webb, Luisa M. Zintgraf, Diederik M. Roijers, Richard Dazeley, Rustam Issabekov, and Evan Dekker

FLAG, Federation University, Ballarat, Australia
AI laboratory, Vrije Universiteit Brussel, Brussels, Belgium
p.vamplew,d.webb,r.dazeley,r.issabekov,e.dekker@federation.edu.au &
lzintgraf,droijers@ai.vub.ac.be

Many — if not most — real-world decision problems involve multiple (possibly conflicting) objectives. If a model of the environment is not readily available, agents must interact with this environment, in order to learn what the best possible policies are. Unlike in single-objective problems, in such *multi-objective reinforcement learning (MORL)* [2] problems it is not immediately obvious what the best policy is, as different users may have different preferences with respect to the different objectives. Furthermore, even with only one user, it is typically very hard, if not impossible, for users to specify a *utility function* that reflects their preferences. Therefore, in MORL, we typically learn a set of possibly optimal policies, that reflect different trade-offs between the objectives. Specifically, agents aim to learn a *coverage set (CS)* with at least one optimal policy for every possible utility function that a user might have.

It has been argued [3] that MORL is underrepresented in the reinforcement learning (RL) literature. Perhaps one reason for this is that while many good benchmarks and benchmarking software exist for single-objective RL [4, 6], MORL research papers typically employ only one or two benchmarks that do not necessarily overlap, and those are often not freely available. This situation leads to only partial comparability and reproducibility. In this work we aim to mitigate this by proposing a benchmark suite that not only contains the most popular benchmarks for MORL used in the literature, but also proposes and includes new benchmarks that are generalised [6, 7]. Not only do these generalised benchmarks help to prevent method overfitting, they also provide researchers with an easy-to-use tool to investigate how their algorithms perform under different characteristics of their MORL problems (such as the size of state and action spaces, the amount of stochasticity in the transition function, etc.).

To create a good benchmark suite, we previously outlined the following criteria [5, 7]: 1) two or more objectives, 2) stochasticity in transition dynamics and/or rewards, 3) different Pareto front features such as concavities and discontinuities, 4) state dimensionality high enough to require the use of function approximation, 5) continuous state or action spaces, 6) partially-observable states, 7) a mixture of episodic and continuing tasks. In the proposed MORL-Glue benchmark suite, criteria 1, 2, and 3 are implemented via standard benchmarks, as well as the new generalised Deep Sea Treasure (G-DST) and random multiobjective Markov Decision Process (MOMDP) problems, and the Collecting Traveller problem (CTP)

[7]. The latter three problems are generalised, and can thus scale up with respect to the number of objectives, and the size of the state and action spaces (criterion 4). Furthermore, the CTP can be made continuous as well as partially observable (criteria 5 and 6). Furthermore, CTP and G-DST are episodic problems, while random MOMDPs are continuing tasks (criterion 7).

Our implementation builds on RL-Glue [4]. It runs a server to which separate agent, environment, and experiment processes connect via sockets. Therefore, these can be implemented in any (and different) languages, facilitating the reuse of standard benchmark implementations with new learning algorithms.

1 The Demo

The demo consists of training an agent on one of our generalised benchmarks, the generalised Deep Sea Treasure (G-DST) problem. We show how an algorithm, in this case Q-learning in combination with thresholding and lexicographic ordering over objectives [1], can be evaluated across different parameters of the problem. Specifically, we designed an experiment in which the size of the state-space is increased: at every instance, we determine the size of the state-space, and generate a random solution set (i.e., coverage set). We then show how many samples it takes before the agent can accurately estimate this coverage set, as a function of the size of this state-space. The demo contains a visualisation component, intended to show the behaviour of the agent at different times during learning. A video accompanying this demo can be found at https://eportfolios.federation.edu.au/view/view.php?id=74593, and the MORL-Glue software is available from https://github.com/FedUni/MORL.

Acknowledgements

Diederik M. Roijers is a postdoctoral fellow of the Research Foundation – Flanders (FWO). This research was in part supported by Innoviris – Brussels Institute for Research and Innovation.

References

1. Z. Gabor, Z. Kalmar, and C. Szepesvari. Multi-criteria reinforcement learning. In *ICML*, pages 197–205, July 24-27 1998.
2. D.M. Roijers, P. Vamplew, S. Whiteson, and R. Dazeley. A survey of multi-objective sequential decision-making. *JAIR*, 47:67–113, 2013.
3. D.M. Roijers, S. Whiteson, P. Vamplew, and R. Dazeley. Why multi-objective reinforcement learning? In *European Workshop on Reinforcement Learning*, 2015.
4. B. Tanner and A. White. RL-Glue: Language-independent software for reinforcement-learning experiments. *Journal of Machine Learning Research*, 10(Sep):2133–2136, 2009.
5. P. Vamplew, R. Dazeley, A. Berry, E. Dekker, and R. Issabekov. Empirical evaluation methods for multiobjective reinforcement learning algorithms. *Machine Learning*, 84(1-2):51–80, 2011.
6. S. Whiteson, B. Tanner, M.E. Taylor, and P. Stone. Generalized domains for empirical evaluations in reinforcement learning. In *ICML*, 2009.
7. L.M. Zintgraf, T.V. Kanters, D.M. Roijers, F.A. Oliehoek, and P. Beau. Quality assessment of MORL algorithms: A utility-based approach. In *Benelearn*, 2015.