Special issue on multi-objective reinforcement learning

Drugan, Madalina; Wiering, Marco; Vamplew, Peter; Chetty, Madhu

Published in:
Neurocomputing

DOI:
10.1016/j.neucom.2017.06.020

Publication date:
2017

Editorial

Special Issue on Multi-objective Reinforcement Learning

Madalina Drugan
Technical University of Eindhoven, Netherlands
E-mail address: madalina.drugan@gmail.com

Marco Wiering
Institute of Artificial Intelligence and Cognitive Engineering, University of Groningen, Netherlands
E-mail address: m.a.wiering@rug.nl

Peter Vamplew
School of Engineering and Information Technology, Federation University Australia
E-mail address: p.vamplew@federation.edu.au

Madhu Chetty
School of Engineering and Information Technology, Federation University Australia
E-mail address: madhu.chetty@federation.edu.au

Abstract. Many real-life problems involve dealing with multiple objectives. For example, in network routing the criteria may consist of energy consumption, latency, and channel capacity, which are in essence conflicting objectives. As in many problems there may be multiple (conflicting) objectives, there usually does not exist a single optimal solution. In those cases, it is desirable to obtain a set of trade-off solutions between the objectives. This problem has in the last decade also gained the attention of many researchers in the field of Reinforcement Learning (RL). RL addresses sequential decision problems in initially (possibly) unknown stochastic environments. The goal is the maximization of the agent's reward in an environment that is not always completely observable. The purpose of this special issue is to obtain a broader picture on the algorithmic techniques at the confluence between multi-objective optimization and reinforcement learning. The growing interest in multi-objective reinforcement learning (MORL) was reflected in the quantity and quality of submissions received for this special issue. After a rigorous review process, 7 papers were accepted for publication, and they reflect the diversity of research being carried out within this emerging field of research. The accepted papers consider many different aspects of algorithmic design and the evaluation and this editorial puts them in a unified framework.

Test environment. The practical motivation, such as novel approaches for challenging real-world applications or developing new algorithms with an improved computational efficiency for particular problems, is essential for any proposed technique. In this issue, all the papers use benchmark environments with two or three objectives. The Deep Sea Treasure task [2,3,6] is a bi-objective environment consisting of ten Pareto-optimal states, which has often been used for testing MORL algorithms. The Bonus World used in [7] is an original three objective environment. Another bi-objective environment that has been used to evaluate a novel multi-objective RL algorithm is the Linked Rings problem [3]. Some of the used environments consist of continuous state variables. The Cart Pole problem [5] has two objectives with continuous state values that reflect the position and velocity of the cart and the angle and the angular velocity of the pole. The Water Reservoir problem [1] models an agent controlling the water level from a reservoir with three conflicting objectives, which are flooding, water and electricity demand. The Dynamic Economic Emissions Dispatcher problem [6] involves scheduling electricity generators to meet the customers' demand while minimizing the fuel cost and emissions. Resource Gathering [2] and Predator Prey [5] are three objective environments with stochastic transitions which are related to strategic games. Two of the above multi-objective environments have stochastic transition functions [1,2]; the other environments are deterministic. In [4], an agent navigates through a maze with continuous states that contains obstacles and different kinds of areas. The problem has one primary objective, while other secondary objectives are found with an unsupervised learning method which are subsequently solved with off-policy RL techniques.

The methodological approach. Many of the proposed MORL algorithms use variants of the Q-learning algorithm [2-7]. In [5], multi-objectivization is used to create additional objectives next to solving the primary goal in order to improve the empirical efficiency. The objectives are assumed to be independent, and Q-values for each objective are learned in parallel. On top of the multi-objectivization mechanism, reward shaping is used to incorporate heuristical knowledge. The goal is to learn the Pareto front of optimal policies. The algorithm proposed in [6] uses scalarization functions and the hypervolume unary indicator to transform the reward vectors into scalar reward values. Similarly to [5], the goal is to identify the Pareto front of optimal policies when additional rewards are added to each objective through reward shaping functions. The hypervolume unary indicator is also used in [1] to measure the performance of a policy-search MORL algorithm. The empirical performance is improved using multiple importance sampling estimators. In [3], the authors use a variant of geometric steering for multi-objective stochastic games with scalarized reward vectors. The MORL algorithm in [4] is an interesting mixture of on-line learning for the first objective and off-line learning for two independently found secondary objectives. The secondary objectives are found using unsupervised learning and their corresponding learned policies are useful when the primary task changes in the environment. In [7], one objective is considered more important than the second objective with so-called lexicographic ordering, and to solve this problem the RL algorithm is integrated with new variants of the softmax exploration strategy. In another line of reasoning, in [2] the authors use Pareto dominance to partially order policies. Not one, but several policies with associated Q-value vectors are simultaneously optimized.

Theoretical analysis. There are only some papers in this special issue that give theoretical guarantees on the expected behavior of the algorithms. In this issue, in [5] and [6] proofs are provided for the convergence of MORL variants with reward shaping functions.

Short summary of papers in the current issue. The first three papers propose and evaluate the performance of reinforcement learning algorithms designed specifically for tasks involving multiple conflicting objectives.

- "Manifold-based Multi-objective Policy Search with Sample Reuse" by Simone Parisi, Matteo Pirotta, and Jan Peters: This paper extends prior approaches to policy-search learning of multiobjective policies by learning a manifold in policy-parameter space. Sampling points on this manifold can produce policies which accurately approximate the Pareto front of policies, which is more efficient than directly learning a set of these policies.

- "A Temporal Difference Method for Multi-Objective Reinforcement Learning" by Manuela Ruiz-Montiel, Lawrence Mandow and José-Luis Pérez-de-la-Cruz: Like Parisi et al, this work addresses the task of learning multiple policies which represent different Pareto-optimal tradeoffs between objectives. However, rather than policy-search, this paper extends the temporal-difference Q-learning algorithm to the task of learning multiple Pareto-optimal policies.

- "Steering Approaches to Pareto-Optimal Multiobjective Reinforcement Learning" by Peter Vamplew, Rustam Issabekov, Richard Dazeley, Cameron Foale, Adam Berry, Tim Moore, and Douglas Creighton: This paper adapts the geometric steering algorithm, originally designed for stochastic multi-criteria games, to learning Pareto-optimal non-stationary policies for multiobjective Markov Decision Processes. It also provides an example of the application of the steering approach to the problem of controlling local battery storage for a household's solar power system.

The next two papers both address the incorporation of additional objectives into an existing reinforcement learning task.

- "Identification and Off-Policy Learning of Multiple Objectives Using Adaptive Clustering" by Thommen Karimpanal George and Erik Wilhelm: In this paper additional objectives are dis covered by the agent itself during its exploration of the environment, using online unsuper- vised clustering. It is shown that Q-learning can be used to learn, at least partially, the values associated with these additional objectives in parallel with learning to solve the primary goal, thereby minimizing the need for additional exploration in case the goal would change.

- "Multi-objectivization and Ensembles of Shapings in Reinforcement Learning" by Tim Brys, Anna Harutyunyan, Peter Vrancx, Matthew Taylor and Ann Nowé: This paper examines the use of multi-objectivization to improve the performance of a reinforcement learning agent on a single-objective task. Additional objectives are introduced either by decomposition of the original objective or based on external heuristic knowledge. This introduces an additional source of diversity, which supports the use of ensemble methods which significantly improve the learning performance.

The final two papers in the issue examine how methods which are widely used in single-objective reinforcement learning can be applied in the context of multiobjective reinforcement learning.

- "Policy Invariance under Reward Transformations for Multi-Objective Reinforcement Learning" by Patrick Mannion , Sam Devlin, Karl Mason, Jim Duggan and Enda Howley: Potential-Based Reward Shaping (PBRS) has been shown to be an effective means of accelerating learning in single-objective problems, with proven guarantees that it does not interfere with the final optimal policy. This paper extends these theoretical guarantees to the case of multiple objectives, for both single-agent and multi-agent systems. It also provides the first empirical results for the use of PBRS within multiobjective reinforcement learning.

- “Softmax Exploration Strategies for Multiobjective Reinforcement Learning” by Peter Vamplew, Richard Dazeley, and Cameron Foale: The effectiveness of exploration strategies has been widely studied in single-objective reinforcement learning, but this paper provides one of the first intensive studies of these techniques in the context of multiple objectives, showing that unexpected complications may arise due to the introduction of additional objectives. It also proposes and evaluates two multiobjective adaptations of the widely used softmax approach to exploration.

Acknowledgements

We would like to thank all of the authors who submitted their work for this issue, as well as the reviewers who generously gave their time and expertise during the review process. We also wish to thank the editors of Neurocomputing who supervised an independent review process for those papers for which we had a conflict of interest.

References

[1] Simone Parisi, Matteo Pirotta , and Jan Peters (2017) “Manifold-based Multi-objective Policy Search with Sample Reuse”. In Neurocomputing, special issue on Multi-Objective Reinforcement Learning.

[2] Manuela Ruiz-Montiel, Lawrence Mandow and José-Luis Pérez-de-la-Cruz (2017) “A Temporal Difference Method for Multi-Objective Reinforcement Learning”. In Neurocomputing, special issue on Multi-Objective Reinforcement Learning.

[3] Peter Vamplew, Rustam Issabekov, Richard Dazeley, Cameron Foale, Adam Berry, Tim Moore, and Douglas Creighton (2017) “Steering Approaches to Pareto-Optimal Multiobjective Reinforcement Learning”. In Neurocomputing, special issue on Multi-Objective Reinforcement Learning.

[4] Thommen Karimpanal George and Erik Wilhelm (2017) “Identification and Off-Policy Learning of Multiple Objectives Using Adaptive Clustering”. In Neurocomputing, special issue on Multi-Objective Reinforcement Learning.

[5] Tim Brys, Anna Harutyunyan, Peter Vrancx, Matthew Taylor and Ann Nowe (2017) “Multi-objectivization and Ensembles of Shapings in Reinforcement Learning”. In Neurocomputing, special issue on Multi-Objective Reinforcement Learning.

[6] Patrick Mannion , Sam Devlin, Karl Mason, Jim Duggan and Enda Howley (2017) “Policy Invariance under Reward Transformations for Multi-Objective Reinforcement Learning”. In Neurocomputing, special issue on Multi-Objective Reinforcement Learning.

[7] Peter Vamplew, Richard Dazeley, and Cameron Foale (2017) “Softmax Exploration Strategies for Multiobjective Reinforcement Learning”. In Neurocomputing, special issue on Multi-Objective Reinforcement Learning.