Reinforcement Learning and Optimization in Stochastic Multi-objective Environments
Proposal for Special Session ADPRL 2014

Mădălina M. Drugan and Bernard Manderick

Computational Modeling Group, Artificial Intelligence Lab, Vrije Universiteit Brussel,
Pleinlaan 2, B-1050, Brussels, Belgium

1 Motivation

In real-life, we are very often confronted with selecting one of several options where the resulting payoff is stochastic and assumed to be unknown. However, in many applications, multiple objectives should be taken in account. We want to combine methods of machine learning (ML) and Multi-Objective Optimization (MOO) in order to learn and/or search difficult multi-objective environments that are possibly dynamic, uncertain and partially observable.

We are interested in exploring the potential synergies between reinforcement learning (RL), which is a well established sequential decision Machine Learning (ML) problem, and Multi-objective optimization, which is an sub-area of multi-criteria decision making. MOO considers the optimization of more than one objective simultaneously and a decision maker decides either which solutions are important for the user or when to present these solutions to the user for further consideration. We consider Multi-Objective Optimization (MOO) a sub field of multi-criteria decision (MCDM) concerned with optimization of more than one objective simultaneously and where a decision maker decides which solutions are important for the user and when to show these solutions to the user. Currently, MOO algorithms are seldom used for stochastic optimization. [5] gives an overview on the decision problems involving both multi-objective and stochastic optimization and concludes that this is a currently unexplored buy very promising research area.

We consider the extension of RL to multi-criteria stochastic rewards (also called utilities in decision theory). The resulting algorithms are hybrid between multi-criteria decision making and stochastic optimization. The RL algorithms are enriched with the intuition and computational efficiency of MOO in handing multi-objective problems. We call them multi-objective RL algorithms. Since several actions can be considered to be the best according to their reward vectors, in general, there are multiple Pareto optimal policies, each being optimal with respect to at least one of the criteria. Optimizing (stochastic) multi-objective problems can be expressed in different ways.

A general criterion to classify multi-objective approaches considers different order relationships of the goodness of solutions. Scalarization [4] transforms the multi-objective problem into a single objective problem, by combining the different values of the reward vectors into a scalar using linear or non-linear functions. The advantage of this approach a multi-objective problem is transformed into a single objective problem which can be solved by a standard RL. With this approach, the decision maker has little or no preference on the particular Pareto optimal policies returned. Pareto-based ranking methods perform the search directly in the multi-objective space using the Partial order relationships. They are called posterior methods and aim to produce all the Pareto optimal solutions. A decision maker can then afterwards select her/his preferred solution. The a-priori techniques assume that the decision maker has preferences for a particular region of the Pareto optimal set and uses and the interactive methods permanently interact with the user in order to select the preferred policy.

The most successful hybrid algorithms are also motivated and validated on real-world problems.

2 Aim and scope

The main goal of this special session is to start the process of unifying and streamlining research on learning and optimization in multi-objective stochastic environments which for time being seems to evolve independently and disconnected in Reinforcement learning and Multi-criteria decision making. We are considering machine-learning algorithms that are both theoretically and practically motivated. We want to bring together researchers from machine learning, optimization and artificial intelligence, interested learning and optimization in multi-objective stochastic environments. We also encourage submissions related to stochastic multi-objective optimization in other areas such as operation research, games and real-world applications.

Ideally, the special session will help researches with different background in Reinforcement Learning and Multi-objective Optimization to identify some common ground for their work.

3 Topics of interest

Topics of interests include but are not limited to

1. Multi-objective reinforcement learning
2. Multi-objective optimization algorithms such as metaheuristics, evolutionary algorithms, etc. for stochastic environments
3. Theoretical results on the learnability in multi-objective stochastic environments
4. Novel algorithmic frameworks for multi-objective stochastic environments
5. Multi-criteria aspects of robotics
6. Multi-objective self adapting systems
7. Multi-objective automatic configuration systems
8. Multi-objective games
9. Real-world applications in engineering, business, computer science, biological sciences, scientific computation, etc. in Stochastic Multi-objective Environments
10. Multi-criteria dynamic/reactive scheduling and planing

4 Organizers

Note that both organizers has a strong background in the field of Machine Learning and optimization and have experience in multi-objective optimization.

4.1 Madalina M. Drugan

She is senior researcher at the Artificial Intelligence Lab, Vrije Universiteit Brussels, Belgium. She received a PhD (2006) from the Computer Science Department, University of Utrecht, The Netherlands. Her PhD thesis "Conditional log-likelihood MDL and Evolutionary MCMC" is researching (designing, analyzing, experimenting) various Machine Learning and optimization algorithms in fields like Bayesian Network classifiers, Feature Selection, Evolutionary Computation, and Markov chain Monte Carlo. She did research in Evolutionary Computation related algorithmic design for Bioinformatics, Multi-objective optimization, Meta-heuristics, Operational Research, and Evolutionary Computation for more than 10 years. Recently, she is involved in developing a theoretical and algorithmic framework of the new branch of Reinforcement Learning using multi-objective rewards. She has experience with research grants, reviewing services and a strong publication record in international peer-reviewed journals and conferences, various academic prices. She initiated and is currently organizing a special session at WCCI 2014.

Key publications for this tutorial include [1,2,9,10,3,8,7,6].

4.2 Bernard Manderick

Since 1994, he is professor at the Artificial Intelligence Lab, Vrije Universiteit Brussels, Belgium, where he also obtained in 1991 his PhD (with Greatest Distinction) entitled "Selectionism as a Basis for Categorization and Adaptive Behavior" supervized by Prof. Dr. L. Steels. And, for which he received the IBM-prize for Informatics from the Fund for Scientific Research. From 1992 till 1994, he was assistent professor at the Department of Computer Science (Faculty of Economics) at the Erasmus University Rotterdam. In 1993, he was post-doctoral researcher at Electrotechnical Lab in Tsukuba, Japan.

So far, he is (co-)author of over 130 papers covering several machine learning techniques and several of its applications. In the domain of evolutionary computation, he worked on fine-grained parallel GAs, the relation between fitness landscapes and GA-performance, and evolvable hardware. He also supervized 11 PhD theses.

4.3 The organizer's relevant experience

The Computational Modeling group (COMO) was experience in organizing special sessions, workshops, and tutorials at major conferences:

- COMO organized a special session at the 2012 and 2014 IEEE World Congress of Computational Intelligence (IEEE WCCI 2012 and IEEE WCCI 2014).
- COMO co-organized in 2012 and 2013 the Adaptive Learning Agents workshop (ALA) that is part of the eleventh international conference on Autonomous Agents and Multiagent Systems (AAMAS 2012 and AAMAS 2013)

- COMO organized Multi-agent learning tutorials at major conferences including ECML 2013, AAMAS 2013 and EVOLVE 2014

The organizers have already published a number of important publications in the area of multi-objective reinforcement learning:

- Yahyaa, S. Q., Drugan, M. M., & Bernard, M.. (2014). Exploration vs Exploitation in the Multi-Objective Multi-Armed Bandit Problem. In International Joint Conference on Neural Networks (IJCNN). presented at the 07/2014, Beijng: IEEE.
- Drugan, M. M., & Nowe, A.. (2014). Scalarization based Pareto optimal set of arms identification algorithms. In International Joint Conference on Neural Networks (IJCNN). presented at the 07/2014, Bejing, China: IEEE.
- Van Moffaert, K., Drugan, M. M., & Now, A.. (2014). Learning Sets of Pareto Optimal Policies. In Thirteenth International Conference on Autonomous Agents and Multiagent Systems - Adaptive Learning Agents Workshop (ALA).
- Drugan, M. M., & Nowe, A.. (2013). Designing multi-objective multi-armed bandits algorithms: a study. International Joint Conference on Neural Networks (IJCNN'13).
- Van Moffaert, K., Drugan, M. M., & Nowe, A.. (2013). Hypervolume-based Multi-Objective Reinforcement Learning. Lecture Notes in Computer Science, Evolutionary Multi-Criterion Optimization (EMO 2013).
- Van Moffaert, K., Drugan, M. M., & Nowe, A.. (2013). Scalarized Multi-Objective Reinforcement Learning: Novel Design Techniques. In IEEE SSCI. Singapore.
- Van Moffaert, K., Drugan, M. M., & Nowe, A.. (2013). Multi-Objective Reinforcement Learning using Sets of Pareto Dominating Policies. In 22nd International Conference on Multiple Criteria Decision Making.
- Brys, T., Van Moffaert, K., Van Vaerenbergh, K., & Nowe, A.. (2013). On the Behaviour of Scalarization Methods for the Engagement of a Wet Clutch. In Proceedings of the 12th International Conference on Machine Learning and Applications (ICMLA 2013). Miami, Florida, USA: IEEE.

5 Potential contributors

The organizers consider the following possible contributors with background in Machine Learning and optimization from labs of all continents

1. Marc G. Bellemare, University of Alberta
2. Peter Bosman, CWI
3. Michael Bowling, University of Alberta
4. Juergen Branke, University of Coventry
5. Dimo Brockhoff, INRIA Lille
6. Robert Busa-Fekete, University Marburg
7. Tim Byrs, Vrije Universiteit Brussels
8. Darwin G Caldwell, Italian Institute of Technology
9. Andrea Castelletti, Politecnico di Milano
10. Richard Dazeley, University of Ballarat
11. Kalyanmoy Deb, Michigan State University
12. Damien Ernst, University of Liege
13. Christian Fischer, Ulm University
14. Johannes Fürnkranz, TU Darmstadt
15. Raphael Fonteneau, INRIA Lille
16. Xu Haichi, Chiba University
17. Hisashi Handa, Okayama University
18. Kazuyuki Hiraoka, Tokyo Polytechnic University
19. Matthew W. Hoffman, University of Cambridge
20. Eyke Hüllermeier, Universität Marburg
21. Hisao Ishibuchi, Osaka Prefecture University
22. Rustam Issabekov, University of Ballarat
23. Yaochu Jin, University of Surrey
24. Joshua Knowles, University of Manchester
25. Petar Kormushev, Italian Institute of Technology
26. Arnaud Liefooghe, INRIA Lille
27. Daniel J. Lizotte, University of Waterloo
28. Francis Maes, K.U. Leuven
29. Bernard Manderick, Vrije Universiteit Brussel

30. Shie Mannor, Israel Institute of Technology
31. Susan A. Murphy, University of Michigan
32. Sriraam Natarajan, Indiana University
33. Gabriela Ochoa, University of Stirling
34. Timo Oess, Ulm University
35. Yew-Soon Ong, Nanyang Technological University
36. Mohamed Oubbati, Ulm University
37. Günther Palm, Ulm University
38. Francesca Pianosi, Politecnico di Milano
39. Jan Peters, TU Darmstadt
40. Marcello Restelli, Politecnico di Milano
41. Diederik Roijers, Vrije Universiteit Amsterdam
42. Michele Sebag, INRIA Saclay
43. Nahum Shimkin, Israel Institute of Technology
44. Thomas Stutzle, Free University of Brussels
45. Csaba Szepesvári, University of Alberta
46. Prasad Tadepalli, Oregon State University
47. El-Ghazali Talbi, University of Lille
48. Alexandru Tantar, University of Luxembourg
49. Emilia Tantar, University of Luxembourg
50. Dirk Thierens, Utrecht University
51. Kristof Van Moffaert, Vrije Universiteit Brussels
52. Martijn van Otterlo, Radboud University Nijmegen
53. Peter Vamplew, University of Ballarat
54. Sebastien Verel, INRIA Lille
55. Weijia Wang, University of Toronto
56. Shimon Whiteson, Vrije Universiteit Amsterdam
57. Marco Wiering, University of Groningen
58. Xin Xu, National University of Defense Technology
59. Shengxiang Yang, University of Leicester
60. Xin Yao, University of Birmingham

The majority of these possible contributors are part of the academic network of the organizer or of the COMO group and will be considered in the reviewing process.

References

1. M. M. Drugan. Multi-objective multi-armed bandits algorithms: an alternative optimizer for stochastic environments. Set-Oriented and Indicator-Based Multi-Criteria Optimization (SIMCO 2013), 2013. Lorentz Center, Leiden.
2. M. M. Drugan and D. Thierens. Generalized adaptive pursuit algorithm for genetic pareto local search algorithms. In *GECCO*, pages 1963–1970, 2011.
3. M.M. Drugan and A. Nowe. Designing multi-objective multi-armed bandits: an analysis. In *Proc of International Joint Conference of Neural Networks (IJCNN)*, 2013.
4. G. Eichfelder. *Adaptive Scalarization Methods in Multiobjective Optimization*. Springer, 2008.
5. W. J. Gutjahr and A. Pichler. Stochastic multi-objective optimization: a survey on non-scalarizing methods. *Annals of Operations Research: Special Volume*, 2013.
6. K. Van Moffaert, M. M. Drugan, and A. Nowe. Multi-objective reinforcement learning using sets of pareto dominating policies. In *International Conference on Multiple Criteria Decision Making*, 2013.
7. K. Van Moffaert, M. M. Drugan, and A. Nowe. Scalarized multi-objective reinforcement learning: Novel design techniques. In *IEEE Symposium Series on Computational Intelligence*, 2013.
8. A. Nowe, K. Van Moffaert, and M. M. Drugan. Multi-objective reinforcement learning. In *European Workshop on Reinforcement Learning*, 2013.
9. F. Puglierin, Madalina M. Drugan, and M. Wiering. Bandit-inspired memetic algorithms for solving quadratic assignment problems. In *IEEE Congress on Evolutionary Computation*, pages 2078–2085, 2013.
10. K. van Moffaert, M.M. Drugan, and A. Nowe. Hypervolume-based multi-objective reinforcement learning. In *Proc of Evolutionary Multi-objective Optimization (EMO)*. Springer, 2013.