Neural Computing and Applications (2022) 34:1649–1651
https://doi.org/10.1007/s00521-021-06593-4

EDITORIAL

Special issue on adaptive and learning agents 2020

Felipe Leno da Silva[1] · Patrick MacAlpine[2] · Roxana Rădulescu[3] · Fernando P. Santos[4] · Patrick Mannion[5]

Received: 30 September 2021 / Accepted: 30 September 2021 / Published online: 12 January 2022

1 Introduction

The goal of the Adaptive and Learning Agents (ALA) workshop is to increase awareness of and interest in adaptive agent research, encourage collaboration, and give a representative overview of current research in the area of adaptive and learning agents. It aims at bringing together not only different areas of computer science (e.g. agent architectures, reinforcement learning and evolutionary algorithms) but also different fields studying similar concepts (e.g. game theory, bio-inspired control and mechanism design). The workshop serves as an inclusive forum for the discussion of ongoing or completed work in adaptive and learning agents and multi-agent systems and has been held annually since 2009 in conjunction with the International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), which attracts a wide audience working on those fields of interest. This special issue features selected papers from the 12th Adaptive and Learning Agents Workshop[1] (ALA 2020), which was held virtually on 9 and 10 May 2020, in conjunction with the 19th International Conference of Autonomous Agents and Multi-Agent Systems (AAMAS 2020).

✉ Patrick Mannion
patrickmannion@nuigalway.ie

Felipe Leno da Silva
leno@llnl.gov

Patrick MacAlpine
patmac@cs.utexas.edu

Roxana Rădulescu
roxana.radulescu@vub.be

Fernando P. Santos
f.p.santos@uva.nl

[1] Lawrence Livermore National Lab, San Francisco Bay Area, CA, USA

[2] Sony AI, Austin, USA

[3] AI Lab, Vrije Universiteit Brussel, Ixelles, Belgium

[4] University of Amsterdam, Amsterdam, The Netherlands

[5] School of Computer Science, National University of Ireland Galway, Galway, Ireland

2 Contents of the special issue

This special issue contains 8 articles, which were carefully selected from 45 initial submissions to ALA 2020. Preliminary versions of each article were initially presented at the workshop, before being extended and peer-reviewed again for this special issue. The articles in this special issue provide a comprehensive overview of current research trends within the ALA community.

In the first paper, *Dynamical systems as a level of cognitive analysis of multi-agent learning*, [1], W. Barfuss provides a multi-level conceptual framework to disentangle the different levels of analysis one can perform in the context of multi-agent learning. This framework comprises computational, algorithmic and implementation levels. The paper clarifies how dynamical systems can contribute to the computational and algorithmic levels of multi-agent learning analysis. To clarify the links between learning dynamics and multi-agent learning levels, the author studies dynamics in stochastic games under replicator-like equations, while also providing a new temporal-difference batch-learning algorithm that is shown to converge to deterministic replicator equations in the limit of large memory batches.

The second paper, *Useful policy invariant shaping from arbitrary advice* by Behboudian et al. [2], revisits the idea of dynamic potential based advice (DPBA), a method that should allow one to use arbitrary advice to shape the reward function and speed-up the learning process, without affecting the optimal policy. They demonstrate that DPBA does in fact alter the optimal policy and that, when adding a correction term, the method no longer provides effective shaping with good advice. The authors then propose policy invariant explicit shaping (PIES) as an alternative and

[1] https://ala2020.vub.ac.be/.

show that PIES can use arbitrary advice and speed-up learning, while leaving the optimal policy unchanged.

The third paper, *Lucid dreaming for experience replay: refreshing past states with the current policy*, [3], proposes a new framework that allows learning agents to use their current policy to refresh previous experiences stored in a replay buffer. Experience replay (ER) can improve off-policy reinforcement learning by providing agents the possibility to store and re-use past experiences. It is not clear, however, what experiences should be used in each time-step, given that previous experiences may become useless in the future. LiDER—the proposed framework—tackles this problem by moving agents to previous states, re-creating previous trajectories with the up-to-date policy, and if improved outcomes are observed with the refreshed trajectories, uses them in training.

The fourth paper, *Discrete-to-deep reinforcement learning methods* by Kurniawan et al. [4], combines the community knowledge from the vast literature on tabular reinforcement learning (RL) algorithms with the more recent desire of solving more challenging domains. Their paper contributes a 2-phase RL algorithm, that first learns a coarse policy using tabular RL, and then trains a classifier to map continuous states to actions. The speed with which the agent is able to learn the tabular policy is combined with the accuracy of considering continuous states, enabling agents to benefit from the best of the two worlds.

The fifth paper, *Scalable multi-product inventory control with lead time constraints using reinforcement learning* by Meisheri et al. [5], applies deep RL to multi-product, multi-period inventory management. It approaches the inventory problem as a special class of dynamical system control and has minimal retraining requirements on the RL agent under system changes through the definition of an individual product meta-model while efficiently handling multi-period constraints that stem from different lead times of different products. Experiments show that the presented RL-based approaches scale to hundreds of products while performing better than baseline heuristics and close to the theoretical optimum.

The sixth paper, *Opponent learning awareness and modelling in multi-objective normal form games* by Rădulescu et al. [6], studies the effect of opponent modelling and learning with opponent learning awareness in a series of multi-objective normal form games, where agents have nonlinear utility functions and use the scalarised expected returns (SER) optimisation criterion. It contributes a set of learning approaches in the policy gradient family for multi-objective multi-agent settings, that incorporate opponent policy estimation, as well as modelling and anticipating the opponent's learning step using a Gaussian process. The authors demonstrate that opponent modelling can confer advantages to the agents implementing it, in settings where equilibria are present. For games with no Nash equilibria under SER, their proposed method allows agents to still converge to meaningful solutions that approximate equilibria.

In the seventh paper, *The impact of environmental stochasticity on value-based multiobjective reinforcement learning* [7], Vamplew et al. analyse the role of stochastic rewards and stochastic state transitions in multi-objective Q-Learning. Most of the previous empirical evaluations of multi-objective reinforcement learning and scalarisation methods assume that environments are deterministic. In [7], the authors find that stochasticity in rewards/transitions affects the optimal solution agents which can learn and, importantly, introduce important differences based on the choice of optimisation criterion (e.g. expected scalarised returns or scalarised expected returns). On top of pointing out limitations of multi-objective RL methods in stochastic environments, the authors explore a novel approach to learning optimal policies for environments with stochastic rewards and discuss potential alternative methods that may be more suitable for maximising returns under stochastic transitions.

Finally, in the paper *Value targets in off-policy AlphaZero: a new greedy backup*, Willemsen et al. [8] present a detailed exploration of approaches that combine planning with learning steps such as AlphaZero. Their paper contributes an exploration of how information can be backed-up when planning to construct training value targets when learning. This enables a deep understanding of the very successful Alpha family of algorithms and is very timely given that such algorithms have recently been successful when applied to many previously unsolved tasks.

Acknowledgements The ALA 2020 organisers would like to extend their thanks to all who supported the workshop and special issue by submitting and reviewing contributions. The organisers would also like to thank the Springer editorial staff, especially the Neural Computing and Applications Editor-in-Chief Prof. John MacIntyre, the Publishing Editor Rachel Moriarty and the Journals Editorial Office Assistant Rashmi Jenna for facilitating this special issue, and for sponsoring the Best Paper Award at the ALA 2020 workshop. F. L. Silva's part in this work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC. LLNL-ABS-827175.

References

1. Barfuss W (2021) Dynamical systems as a level of cognitive analysis of multi-agent learning. Neural Comput Appl. https://doi.org/10.1007/s00521-021-06117-0
2. Behboudian P, Satsangi Y, Taylor ME, Harutyunyan A, Bowling M (2021) Useful policy invariant shaping from arbitrary advice. Neural Comput Appl. https://doi.org/10.1007/s00521-021-06259-1
3. Du Y, Warnell G, Gebremedhin A, Stone P, Taylor ME (2021) Lucid dreaming for experience replay: refreshing past states with

the current policy. Neural Comput Appl. https://doi.org/10.1007/s00521-021-06104-5
4. Kurniawan B, Vamplew P, Papasimeon M, Dazeley R, Foale C (2021) Discrete-to-deep reinforcement learning methods. Neural Comput Appl. https://doi.org/10.1007/s00521-021-06270-6
5. Meisheri H, Sultana NN, Baranwal M, Baniwal V, Nath S, Verma S, Ravindran B, Khadilkar H (2021) Scalable multi-product inventory control with lead time constraints using reinforcement learning. Neural Comput Appl. https://doi.org/10.1007/s00521-021-06117-0
6. Rădulescu R, Verstraeten T, Zhang Y, Mannion P, Roijers DM, Nowé A (2021) Opponent learning awareness and modelling in multi-objective normal form games. Neural Comput Appl. https://doi.org/10.1007/s00521-021-06184-3
7. Vamplew P, Foale C, Dazeley R (2021) The impact of environmental stochasticity on value-based multiobjective reinforcement learning. Neural Comput Appl. https://doi.org/10.1007/s00521-021-05859-1
8. Willemsen D, Baier H, Kaisers M (2021) Value targets in off-policy alphazero: a new greedy backup. Neural Comput Appl. https://doi.org/10.1007/s00521-021-05928-5