The Knowledge Engineering Review, Vol. 36, e7, 1 of 3.
doi:10.1017/S0269888921000047

Special issue on adaptive and learning agents 2018

PATRICK MANNION[1], ANNA HARUTYUNYAN[2], BEI PENG[3] and KAUSHIK SUBRAMANIAN[4]

[1]*School of Computer Science, National University of Ireland Galway, University Road, Galway H91 TK33, Ireland*
e-mail: patrick.mannion@nuigalway.ie
[2]*DeepMind, 6 Pancras Square, London N1C 4AG, UK*
e-mail: harutyunyan@google.com
[3]*Department of Computer Science, University of Oxford, Parks Road, Oxford OX1 3QD, UK*
e-mail: bei.peng@cs.ox.ac.uk
[4]*College of Computing, Georgia Institute of Technology, 801 Atlantic Drive, Atlanta, GA 30332, USA*
e-mail: ksubrama@cc.gatech.edu

1 Introduction

The Adaptive and Learning Agents (ALA) community designs and develops autonomous agent-based systems. How best to design such systems is a difficult question, and researchers in ALA take inspiration from many related fields, such as multi-agent systems (MASs), machine learning, reinforcement learning (RL), evolutionary computation, game theory, planning and multi-objective optimization, among others.

The ALA community have made great advances in both theoretical and applied research in recent years, exploring lines of research such as transfer learning, integrating expert human knowledge, safe decision making, multi-task learning, multi-objective decision making and distributional RL, to name a few. Examples of complex application areas that have been successfully tackled by ALA researchers in recent years include energy systems and smart grid, game-playing agents, wind farm control, unmanned aerial vehicles, inventory management and transportation systems.

This special issue features selected papers from the 10th Adaptive and Learning Agents Workshop workshop (ALA 2018), which was held on 14 & 15 July 2018 at the Federated AI Meeting (FAIM) in Stockholm, Sweden. Major international AI conferences that were co-located at the FAIM include AAMAS, ICML, IJCAI and ECAI. ALA 2018 had over 100 attendees and was the largest workshop at the FAIM in terms of submissions received. ALA workshops have been held annually since 2009, and the workshop series aims to promote awareness of and interest in adaptive agent research, to encourage collaboration between ALA researchers and to provide a representative overview of current research in the field. The workshop serves as an interdisciplinary forum for the discussion of ongoing or completed work in ALA and MASs.

2 Contents of the special issue

This special issue contains 10 papers, which were selected out of 55 initial submissions to the ALA 2018 workshop. All papers were initially presented at the workshop, before being extended and reviewed again for this special issue. These articles provide a comprehensive overview of current research directions that are being explored within the ALA community.

In the first paper, *Team learning from human demonstration with coordination confidence*, Banerjee *et al*. (2019) deal with the problem of how best to use expert human demonstrations to speed up RL in collaborative multi-agent problems. The authors evaluate the effectiveness of a newly proposed technique called Coordination Confidence (CC) and compare it with relevant baselines such as the previously proposed Human-Agent Transfer (HAT) technique. Empirical results in the Furniture Movers,

Guided Navigation and Block Dudes domains demonstrate that CC performs well relative to the baseline algorithms.

The second paper *Pre-training with non-expert human demonstration for deep reinforcement learning* by de la Cruz *et al.* (2019) aims to improve the data efficiency and training time of deep RL agents, presenting a method that leverages supervized learning techniques to pre-train agents using small sets of non-expert human demonstrations. The proposed approach is empirically evaluated using the asynchronous advantage actor-critic (A3C) algorithm in six different Atari games, and the results demonstrate that the proposed technique leads to significant speedups in learning when compared to baseline algorithms. The authors also provide Grad-CAM visualizations of feature maps in the final convolutional layers of trained agents' neural networks, demonstrating that RL agents can learn meaningful high-level information about the game through pre-training using the proposed technique.

In the third paper, *Safe Option-Critic: Learning Safety in the Option-Critic Architecture*, Jain *et al.* (2021) deal with the problem of designing safe RL algorithms, that is, algorithms that learn to avoid regions of the state space where there is a high degree of uncertainty about the outcomes of actions. The authors tackle this problem in the options framework, where temporally abstract actions allow an agent to use sub-policies with start and end conditions. Experimental results collected in a variety of grid world environments and Atari games demonstrate that the proposed approach achieves a reduction in the variance of returns and boosts performance in environments with intrinsic variability in the reward structure when compared to baseline algorithms.

The paper *Introspective Q-learning and learning from demonstration* by Li *et al.* (2019a) proposes a method that extends the basic Q-learning algorithm with the concept of introspection. Taking inspiration from prior works that allow RL agents to learn from external demonstrations, the authors' new method allows an RL agent to use its own experiences that have produced high rewards in the past to guide future learning, in place of external expert demonstrations. The authors present experiments in two different domains: the 4-dimensional CartPole domain and the Super Mario AI domain. The experimental results demonstrate that the proposed method improves learning speed significantly when compared to a baseline Q-learning agent.

The paper *Two-level Q-learning: learning from conflict demonstrations* by Li *et al.* (2019b) considers reinforcement learning from demonstrations (LfDs) by multiple experts, where the experts provide conflicting advice. The authors propose a two-level variant of Q-learning, which allows an agent to learn the correct action while simultaneously learning to select the expert that provides the best advice for the current state, thus removing a common assumption in traditional LfD methods, that is, that demonstrations are largely consistent with one another and come from one expert only. Experimental results in a maze navigation domain, a coloured flags visiting domain and the Atari game Pong demonstrate that the proposed two-level Q-learning algorithm outperforms the basic Q-learning algorithm and the previously proposed confidence-based HAT algorithm in settings with multiple expert demonstrations and conflicting advice.

In the paper *Improving trust and reputation assessment with dynamic behaviour*, Player and Griffiths (2020) propose Reacting and Predicting in Trust and Reputation (RaPTaR), a method that extends existing trust and reputation models to give agents the ability to monitor the output of interactions with a group of agents over time to identify any likely changes in behaviour and adapt accordingly. Experiments were conducted where agents selected partners for tasks using small-world networks, scale-free networks and fully connected networks, and a number of previously published trust and reputation models were evaluated alongside the proposed RaPTaR method. The results demonstrate that RaPTaR effectively predicts which agents make good partners in settings where agent behaviour is predictable and repetitive, as well as in situations where agent behaviour is dynamic and changes at random.

The paper *Toll-based reinforcement learning for efficient equilibria in route choice* by Ramos *et al.* (2020) explores how RL methods can be leveraged to achieve efficient outcomes in route choice problems in transportation networks. In route choice problems, each driver usually aims to selfishly reduce their own travel time; however, such an approach can lead to outcomes that are inefficient from a system perspective. To mitigate the impact of selfishness, the authors present Toll-based Q-learning (TQ-learning), which employs the idea of marginal-cost tolling where each driver is charged according to the cost it

imposes on others. The authors present a theoretical analysis which proves that TQ-learning converges to a system-efficient equilibrium in the limit, along with empirical results demonstrating that TQ-learning converges to the optimal outcomes in simulated route choice problems.

The paper *Action learning and grounding in simulated human–robot interactions* by Roesler and Nowé (2019) deals with the problem of developing robots that can understand and act on commands given in human language. Their framework is set up to learn the meaning of object, action, colour and preposition words and phrases. The proposed framework is evaluated using a simulated interaction experiment between a human tutor and a robot, and the results demonstrate that the proposed framework performs well in this task.

In the paper, *Effects of parity, sympathy and reciprocity in increasing social welfare*, Sen *et al*. (2020) aim to develop an understanding of how socially desirable traits like sympathy, reciprocity and fairness can survive in environments that include aggressive and exploitative agents. The authors investigate how parity, sympathy and reciprocity can be used by agents to seek out cooperation possibilities while avoiding exploitation traps in dynamic societies. Experimental investigations in small-world social networks are presented, where agents repeatedly interact with neighbouring agents by playing normal form games such as the prisoner's dilemma and the battle of the sexes. The analysis by the authors highlights some interesting findings, including that purely selfish behaviour is often self-defeating, and that sympathy is useful in increasing social and individual welfare in many situations, while parity is rarely useful in increasing welfare.

Finally, in the paper *Fully Distributed Actor-Critic Architecture for Multitask Deep Reinforcement Learning*, Valcarcel Macua *et al*. (2021) propose the Diff-DAC algorithm, which they apply to multitask reinforcement learning. Diff-DAC uses a distributed learning process, where multiple agents work together to learn a policy, communicating their value and policy parameters to their neighbours and diffusing information across a network of agents with no need for a central learning coordination mechanism. Experiments in a variety of control tasks including inverted pendulum, cart-pole, Acrobot and MuJoCo Hopper demonstrate that Diff-DAC achieves higher performance than the baseline Dist-MTPS algorithm.

Acknowledgements

The ALA 2018 organisers would like to extend their thanks to all who served as reviewers for the workshop and special issue and to the Cambridge University Press staff and the KER Co-Editors-in-Chief Prof. Peter McBurney and Prof. Simon Parsons for facilitating this special issue.

References

Banerjee, B., Vittanala, S. & Taylor, M. E. 2019. Team learning from human demonstration with coordination confidence. *The Knowledge Engineering Review* **34**, e12.

de la Cruz, G. V., Du, Y. & Taylor, M. E. 2019. Pre-training with non-expert human demonstration for deep reinforcement learning. *The Knowledge Engineering Review* **34**, e10.

Jain, A., Khetarpal, K. & Precup, D. 2021. Safe option-critic: learning safety in the option-critic architecture. *The Knowledge Engineering Review* **36**, e4.

Li, M., Brys, T. & Kudenko, D. 2019a. Introspective q-learning and learning from demonstration. *The Knowledge Engineering Review* **34**, e8.

Li, M., Wei, Y. & Kudenko, D. 2019b. Two-level q-learning: learning from conflict demonstrations. *The Knowledge Engineering Review* **34**, e14.

Player, C. & Griffiths, N. 2020. Improving trust and reputation assessment with dynamic behaviour. *The Knowledge Engineering Review* **35**, e29.

Ramos, G. d. O., Da Silva, B. C., Râdulescu, R., Bazzan, A. L. C. & Nowé, A. 2020. Toll-based reinforcement learning for efficient equilibria in route choice. *The Knowledge Engineering Review* **35**, e8.

Roesler, O. & Nowé, A. 2019. Action learning and grounding in simulated human–robot interactions. *The Knowledge Engineering Review* **34**, e13.

Sen, S., Crawford, C., Dees, A., Nanda Kumar, R. & Hale, J. 2020. Effects of parity, sympathy and reciprocity in increasing social welfare. *The Knowledge Engineering Review* **35**, e31.

Valcarcel Macua, S., Davies, I., Tukiainen, A. & Munoz de Cote, E. in press. Diff-dac: Fully distributed actor-critic for average multitask deep reinforcement learning. *The Knowledge Engineering Review* **36**.