Reinforcing the role of competition platforms

Jorge Montalvo Arvizu[1], Lachlan Mason[2] and John Moriarty*[2,3]

[1]Technical University of Denmark
[2]Alan Turing Institute, UK
[3]Queen Mary University of London, UK

Abstract

The RangL project encourages the wider uptake of Reinforcement Learning by supporting a cloud-native platform that hosts industrially focused sequential decision problems. In this opinion piece, members of the RangL team make the case for the potential of competition platforms to facilitate collaboration between academia and industry, and to drive progress in real-world applications of Reinforcement Learning, over the coming years.

Competition platforms have played a significant role in progressing machine learning (ML) to date. As an example, by making a large labelled image dataset available to participants, the annual *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) has fostered several breakthroughs in ML architectures and training techniques. For instance, the use of large-scale deep neural networks by the SuperVision team in ILSVRC2012 has been described as a turning point for large-scale object recognition [1]. Equally, however, competitions have accelerated progress in scientific applications. A recent example is the biennial *Critical Assessment of protein Structure Prediction* (CASP) competition, where participants are assessed on how well they predict the structure of a handful of benchmarked proteins. In CASP13, DeepMind's AlphaFold achieved significant advances in protein structure prediction using novel ML techniques, bringing closer the solution to this important problem in biology [2].

Competitions in Reinforcement Learning (RL), the area of ML addressing how intelligent agents learn to maximise rewards through interaction with the agent–environment framework [3], have also been regularly hosted, for example at the *Conference on Neural Information Processing Systems* and the *International Conference on Machine Learning*. By making available sets of benchmark problems with automated evaluation, even combining these into so-called poly-athlons, these competitions have helped stimulate the development of general RL algorithms. Despite this technical success, however, progress in applying RL to scientific and, more generally, real-world problems has been somewhat slower. Several reasons for this have recently been identified in [4], where the authors list nine technical challenge areas holding back RL from more widespread real-world use. Interestingly, the authors suggest that there are important additional challenges (for which we may use the shorthand 'non-technical'), including infrastructural, societal and problem-dependent barriers. Examples presented of these additional challenges are the modularisation of code; allocation of a fixed computing budget; designing simple interfaces so that real-world problems can be solved by people with limited RL knowledge; and identifying when a problem is suitable for RL. In this opinion piece we suggest that 'the facilitation of collaboration across disciplines and between academia and industry' could be added to this list of non-technical challenges, and make the case for competition platforms in driving progress in real-world applications of RL over the coming years. We also describe RangL [5], an RL competition platform conceived at the *2019 Mathematics of Energy Systems* research programme at the Isaac Newton Institute (INI) and subsequently developed at the Alan Turing Institute. During one of the INI workshops, participants from industry called for an updated set of standardised benchmark power systems problems on which a range of solution techniques could be demonstrated and compared. The aim was to contribute in both directions: to enable practitioners to apply specific recent advances (such as in RL) to the industrial problems in which they are expert, while providing industry with a platform on which to pose problems of interest.

Since unambiguous problem definition and solution evaluation are requirements for a competition, the resulting project became framed as the creation of a competition platform. This platform, RangL, aims to add value by facilitating the framing of sequential decision problems whose solution requires interdisciplinary and industrial collaboration. The RL agent–environment framework was chosen by virtue of being a minimal problem setup (that is, an action space, observations and rewards, and a means of simulating them), and due to the availability of general-purpose solvers such as Stable Baselines [6]. The aim is that collaboration on specific and clearly formulated real-world problems, between data scientists and domain experts with highly relevant knowledge and intuition, might overcome both technical and non-technical blockers to progress for that problem. In this way, competition platforms offer to help tackle a significantly wider range of real-world sequential decision problems than have as yet been attempted using RL.

As just one example of a class of sequential decision problems relevant both to industry and academic study, which could benefit from standardised expression in the agent–environment framework, we mention Real Options Analysis

*j.moriarty@qmul.ac.uk

[7]. The classical approach to valuing a capital investment is deterministic net present value (NPV) analysis, where cash flows in each future year are estimated, and rolled together (via discounting) to provide a net value in year zero. If this value is sufficiently positive then the project is seen favourably. However, a limitation of NPV analysis is that it avoids the explicit modelling of uncertainty, instead accounting for randomness through the crude tool of increased discount rates. This approach may not be adequate when future cash flows are highly uncertain. In contrast to NPV, Real Options Analysis applies financial option pricing models to value real assets.

In [8] one of the present authors applied Real Options Analysis in collaboration with a UK electricity distribution network operator to help reduce costs for a group of electricity customers. The study focused on uncertainty over the pace of decarbonisation on the UK power grid and its implications for capital planning. While investing in too much network capacity could lead to the 'stranding' or inefficient use of new network assets, too little investment could lead to network unreliability and additional costs resulting from unserved power demand. Real Options Analysis invites decision makers to model their own future actions and the associated costs and rewards over the coming years, as uncertainty progressively gives way to observation. It accounts for the fact that, in reality, decision makers will seek to take advantage of future better conditions when they occur and, conversely, will seek to minimise the impact of future poorer conditions should they arise. The value of the flexibility that is available in decision making is thus highlighted in a quantitative way. In particular, reasonable comparisons can be made between flexible and inflexible investment strategies by aiming to assign to both their fair value. Since Real Options Analysis involves modelling both a dynamic (uncertain) environment and the managerial actions taken in response to new information as it is observed, it can of course be interpreted in the agent–environment framework.

At the time of our study, the distribution network operator was considering the deployment of Demand-Side Response (DSR) – an emerging network solution which was much cheaper and more flexible than conventional asset-heavy network reinforcement. Its limitations are that it typically adds less capacity than asset-based network reinforcement, and also that its operation is more uncertain, since consumer demand (and hence the potential of demand to respond to a control signal) is uncertain. Because of these differences in the characteristics of DSR and asset-based solutions, there was no standard approach the company could use to compare the two. The Real Options Analysis methods available from the literature typically either lacked particular features, or took approaches whose validity in this context was unclear. Nevertheless, the company suspected that DSR could be the most appropriate solution for customers in a range of applications. Our joint research produced a software tool with a time-domain simulation of the growth in peak demand, and rule-based agents implementing the alternative technologies. This agent–environment loop was enough to reveal that in some but not all cases, DSR can delay or even avoid large irreversible capacity investments. Since the latter were an order of magnitude more costly, DSR had a potentially high value.

The software tool is now used as standard by the company and its use has expanded into other problems, saving an estimated £1 million per annum through more cost-effective capital investment decisions. The approach was also recognised as best practice in the industry in a report by National Grid, the UK power system operator [9]. Nevertheless, today our revised approach would be to formulate the problem on an RL competition platform. Firstly, the clarity of the agent–environment framework would accelerate the collaborative work and mutual understanding required to formulate the sequential decision problem. Then, general RL algorithms would allow attempts to be made at optimisation of the investment policy, in contrast to the more straightforward comparison of a fixed number of rule-based strategies undertaken in [8]. Further, if the problem was not commercially sensitive then the competition architecture would accelerate progress by enabling any interested party to collaborate in the solution. In any case, the result would be a substantially more flexible and extensible tool with a shortened software development cycle.

The cross-fertilisation from casting a wider range of real-world problems such as this into the RL framework is also capable of challenging and thus enriching the field of RL. Again we provide just one example, which comes from a recent study of sequential decision processes in [10] motivated by the use of weather forecasts in operational planning for power systems. Clearly, meteorological data can provide predictive information which is useful in power system planning. However, meteorological forecasts are complementary to the observations of the power system itself, and typically updated on a slower timescale. With this in mind, the authors of [10] confirm that different forecast update regimes lead to different sequential decision processes at a technical level: suggesting that the statistical role of exogenous forecasts in RL may be a fruitful area for further research.

The RangL project [5] encourages growth in RL uptake within wider audiences by supporting a cloud-native platform that hosts industrially focused sequential decision problems. For each RangL competition we invite control strategy submissions from a wide audience, and foster the fast transfer of results towards industrial use. We provide a code repository, containing a modular reference environment (typically implemented using OpenAI Gym) and helper scripts to enable rapid testing of modified environments and agent training, evaluation and debugging. RangL evaluates RL-based agents in a problem-specific environment, supporting both the definition (codification) and solution of infrastructure problems as sketched in Figure 1. To encourage collaboration we provide a channel on a communications platform, while competition aspects such as agent submission and leaderboard rankings are handled by a web-based frontend. We have captured our learning from early experimentation in the form of environment templates which modularise the problem definition into common steps including, where relevant, the handling of exogenous forecasts. The platform's initial competition modelled the scheduling of power generation under uncertainty and attracted 10 teams from academia, industry and the third sector and, at the time of writing, we are collaborating with the UK Net Zero Technology Centre and Offshore Renewable Energy Catapult on a competition to analyse the optimal pathway to net zero carbon emissions for the UK energy industry. In this spirit, we endeavour to reinforce the role of competition platforms in the wider uptake of machine learning approaches to real-world sequential decision problems.

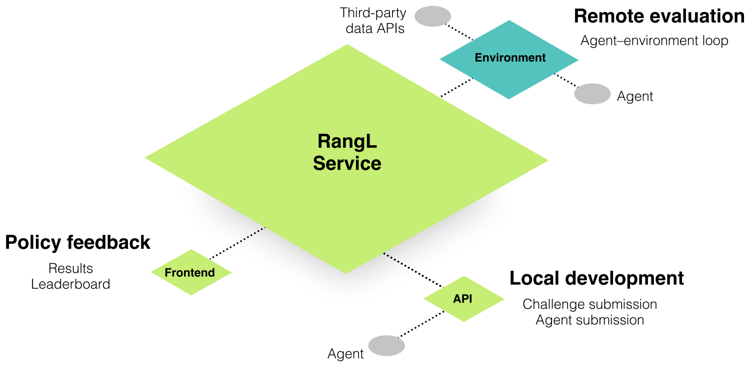

Figure 1: RangL provides a competition platform for real-world Reinforcement Learning (RL) problems based on evaluation of an agent–environment loop.

Group biography. Jorge Montalvo Arvizu is an MSc candidate in Sustainable Energy from the Technical University of Denmark. His work focuses on machine learning applications for power systems with a focus on reinforcement learning and causality.

Lachlan Mason is a Data-centric Engineering Programme Group Leader at the Alan Turing Institute, UK. His research interests include industrial simulation and machine learning engineering.

John Moriarty is Professor of Mathematics in the School of Mathematical Sciences at Queen Mary University of London, UK and a Data-Centric Engineering Programme Research Fellow at the Alan Turing Institute.

References

[1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vision*, vol. 115, pp. 211–252, 2015.

[2] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, and D. Hassabis, "Improved protein structure prediction using potentials from deep learning," *Nature*, vol. 577, no. 7792, pp. 706–710, 2020.

[3] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.* MIT press, 2018.

[4] G. Dulac-Arnold, N. Levine, D. J. Mankowitz, J. Li, C. Paduraru, S. Gowal, and T. Hester, "An empirical investigation of the challenges of real-world reinforcement learning," *CoRR*, vol. arXiv:2003.11881 [cs], 2020.

[5] "RangL." `https://github.com/rangl-labs`, 2021.

[6] A. Hill, A. Raffin, M. Ernestus, A. Gleave, A. Kanervisto, R. Traore, P. Dhariwal, C. Hesse, O. Klimov, A. Nichol, M. Plappert, A. Radford, J. Schulman, S. Sidor, and Y. Wu, "Stable Baselines." `https://github.com/hill-a/stable-baselines`, 2018.

[7] R. K. Dixit and R. S. Pindyck, *Investment under uncertainty.* Princeton University Press, 2012.

[8] J. A. Schachter, P. Mancarella, J. Moriarty, and R. Shaw, "Flexible investment under uncertainty in smart distribution networks with demand side response: Assessment framework and practical implementation," *Energy Policy*, vol. 97, pp. 439–449, 2016.

[9] "Network options assessment methodology review," National Grid plc, 2017.

[10] J. A. de Chalendar and P. W. Glynn, "On incorporating forecasts into linear state space model Markov decision processes," *Philos. Trans. R. Soc. London, Ser. A*, vol. 379, no. 2202, p. 20190430, 2021.