Reinforcement Learning for Automated Scientific Discovery

Mattia Cerrato[1], Jannis Brugger[2], Nicolas Schmitt[3], and Stefan Kramer[1]

[1]Johannes Gutenberg-Universität Mainz
[2]TU Darmstadt
[3]Universität Tübingen

March 23, 2023

In the talk, we will discuss reinforcement learning (RL) as one element of *automated scientific discovery*. This is much in line with the notion of "agents of exploration and discovery" recently proposed by Pat Langley [2], although the use of RL has not been mentioned or elaborated there explicitly. One obvious problem with the idea is that "nature does not provide rewards" - as one might put it - to a learning agent. However, clearly, any agent of discovery follows some kind of policy to come up with interesting results and perhaps a multitude of sub-policies to design and run experiments to validate its theories. So, for each of the tasks involved in automated discovery, RL might be a suitable approach to obtain a policy whenever it is more effective to learn it from rewards than to implement it directly or learn it from text or being advised by another agent. Our focus is on the discovery of human-comprehensible knowledge in physics, not on the optimization of some property (e.g., in material science or drug development) without being able to communicate the results. In this way, it also differs from earlier attempts to include the measurement costs into deep reinforcement learning [1] without any option for explanation.

For our purposes, we view discovery as a sequence of tasks, where (i) interesting states are discovered (e.g., equilibria or steady states) by trial-and-error or reward-based positive reinforcement, then (ii) the agents learn to reach those states consistently by following a policy, and (iii), once this is achieved, they aim to characterize the conditions under which this can be achieved with a symbolic equation, to be able to communicate the discoveries. The process can be repeated in multiple settings (i.e. environments), to learn more and more policies to discover unusual states, be able to reach them consistently and finally, create layers of symbolic knowledge to "explain" phenomena. As a testbed, we have developed 4 scientific RL environments as an extension of OpenAI Gym. These environments let agents experiment with, and rediscover i) the law of the lever; ii) the motion of objects subjected to gravity; iii) projectile motion; iv) Lagrange points in the orbit of two bodies.

From a broader perspective, we propose to cast the scientific discovery process as a RL problem in which an agent concurrently *acts* by performing experiments in an external scientific environment and *reasons* about their outcomes in an internal theoretical environment so as to form theories. The acting and reasoning processes are formalized as two separate partially-observable Markov decision processes (PO-MDPs). Solving the theoretical PO-MDP entails approximating the available experimental data as well as possible by means of a free-form symbolic equation – a theory for a phenomenon of interest. On the other hand, an experiment-level policy can be driven to create empirical evidence which disproves the current theory. The goal of this interaction between the theoretical and experimental policies is to achieve autonomous experimental design as an emergent behavior.

References

[1] Colin Bellinger et al. “Active Measure Reinforcement Learning for Observation Cost Minimization”. In: *Proc. of the 34th Canadian Conference on Artificial Intelligence, Canadian AI 2021, https://doi.org/10.21428/594757db.72846d04* (2021).

[2] Pat Langley. “Agents of Exploration and Discovery”. In: *AI Magazine* 42.4 (Jan. 2022), pp. 72–82. DOI: 10.1609/aaai.12021.