ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Master Thesis

Empirical study of Value approximation in Reinforcement Learning

Reinforcement Learning (RL) has gained a lot of attention in recent years. In RL an agent interacts with an environment through actions and gets rewards when certain conditions are met. The goal of the agent is to learn the sequences of actions that maximize the reward it gets. This approach has been successfully used to learn all kinds of different tasks, from learning to play Atari games [2] to complex robotic control [5].

Current techniques very often rely on vast amounts of compute to achieve very impressive results. However, this has led to many scientist building new algorithms based on intuitions and guesswork, rather than proper experimental analysis. What's more, due to the large costs of single experiments, many times it's unfeasible to run enough repetitions to make statistically meaningful claims. In fact, very recent work has shown that many state-of-the-art published algorithms owe a big part of their performance to the random seed that was used to generate the published experiments [1].

Recently, there has been a push in the community to create RL environments that allow for fast, and consequently more rigorous, experimentation. The two most famous examples are MinAtar[6] and BRAX[3]. The first one, implements a simplified version of several Atari games which run on Numpy. This makes the RL task faster to learn and also the environment slightly faster to simulate while still preserving most of the structure. The second example, BRAX, goes a step further and implements a physics simulator for robotic tasks using the JAX library. This allows running the simulator in the GPU and optimizing the computation graph of the environment together with the training procedure. The end results are RL agents that learn several orders of magnitude faster than any other available library. Most importantly, it has been shown that the insights from these environments translate nicely to more complex and computationally expensive ones[4, 3].

Building on insights from a previous thesis in our lab, we will conduct a proper empirical study of current Reinforcement Learning approaches. While we experimented on MinAtar (reimplemented in JAX) before, you will be working with Brax. Brax is very fast, but still significantly slower than MinAtar. This means that the experimental design will be more involved due to computational limitations. But, as the environments are more complex, we will be able to get a better understanding of RL. As a first step, we will focus on the value network. Here we can also build on the previous thesis. If time allows, afterwards we will try to leverage the insights to build better value learning algorithms.

More information and grading scheme can be found on:
`https://www.cadmo.ethz.ch/education/thesis/guidelines.html`

Co-Supervisor: Robert Meier, CAB J 21.4, romeier@inf.ethz.ch
Co-Supervisor: Asier Mujika, CAB J 21.2, asierm@inf.ethz.ch

Supervising Professor: Prof. Dr. Angelika Steger, CAB G 37.2, steger@inf.ethz.ch

References

[1] R. Agarwal, M. Schwarzer, P. S. Castro, A. Courville, and M. G. Bellemare. Deep reinforcement learning at the edge of the statistical precipice. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021.

[2] A. P. Badia, B. Piot, S. Kapturowski, P. Sprechmann, A. Vitvitskyi, Z. D. Guo, and C. Blundell. Agent57: Outperforming the atari human benchmark. In *International Conference on Machine Learning*, pages 507–517. PMLR, 2020.

[3] C. D. Freeman, E. Frey, A. Raichuk, S. Girgin, I. Mordatch, and O. Bachem. Brax - a differentiable physics engine for large scale rigid body simulation, 2021.

[4] J. S. Obando-Ceron and P. S. Castro. Revisiting rainbow: Promoting more insightful and inclusive deep reinforcement learning research. *arXiv preprint arXiv:2011.14826*, 2020.

[5] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[6] K. Young and T. Tian. Minatar: An atari-inspired testbed for thorough and reproducible reinforcement learning experiments. *arXiv preprint arXiv:1903.03176*, 2019.