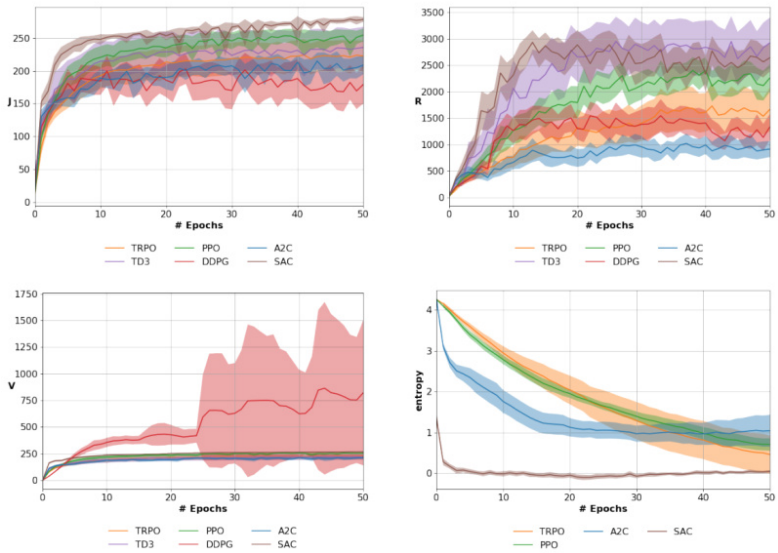
Hessisches Kompetenzzentrum
für Hochleistungsrechnen

Deep Reinforcement Learning: Benchmarking, Algorithms and Applications



Project Manager
Dr. Davide Tateo

Principal Investigator
Dr. Davide Tateo

Project Term
2020 - 2021

Clusters
Lichtenberg Cluster Darmstadt

Software
PyTorch

Additional Software
MuJoCo, PyBullet

Institute
Intelligent Autonomous Systems

University
Technische Universität Darmstadt

Introduction

Deep Learning is the major component of the success of most new Artificial Intelligence applications. A new promising field is the application of these techniques to Reinforcement Learning (RL), where they can be used to build AI that can beat humans in complex games, such as chess and Go, in the development of autonomous driving applications, or bring robotics application to the next level. However the field is relatively new, and the scientific method has not been adopted properly to assess the quality of the results in this field. Our objective is to found the Deep RL research starting from the basics, by first measuring with extensive benchmarks of the performances of the existing algorithms, and using the insights of this analysis, find out the best components of each approach to build novel and better deep reinforcement learning algorithms.

Methods

We considered many standard Reinforcement Learning methods such as DDPG [1], A2C [2], TRPO [3], PPO [4], TD3 [5] and SAC [6]. To ensure the statistical relevance of our results, web evaluated the algorithms over 25 runs and we reported mean and confidence intervals. Differently from standard RL research, not only we evaluate the performance w.r.t. cumulative return, but also we consider the —more appropriate in terms of theorydiscounted cumulative return, the policy entropy, and the value function in the initial state. Indeed, the discounted cumulative return is more appropriate than the cumulative one

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg

as is the actual measure the RL algorithms are maximizing. This metric is also more relevant in an infinite horizon scenario, where the undiscounted one diverges. The (expected) value function of the initial states is also a very important metric, as it is equivalent to the objective function, i.e. the cumulative discounted return. As deep actor-critic approaches are estimating the value function from data, by looking at the value function of the initial states we can understand if there is a problem of overestimation or underestimation. This metric, often neglected in current literature, gives us crucial insights to understand the properties of deep RL algorithms.

Results

We developed a Reinforcement Learning library called MushroomRL. This library provides a clean implementation of Deep Reinforcement Learning algorithms. To validate the implementation of the algorithms, we build an extensive benchmarking suite, MushroomRL Benchmark. We validated our implementation on the most common RL Benchmarks.

Discussion

The analysis of the results shows, as expected, that the cumulative reward and the cumulative discounted reward are not equivalent metrics: different algorithms may be regarded as better depending on the chosen metric in different tasks. A slight modification of the same environment (e.g. MuJoCo vs PyBullet environment) impacts massively the performance of the approaches. It is evident the importance of the value function of the initial state to understand the behavior of an algorithm, as overestimation issues can impact the overall performance as seen in DDPG. In general, the results show that is fundamental to improve the design of Deep RL experiments, in terms of the number of seeds, the variety of environments, and the performance metrics.

Figures

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg

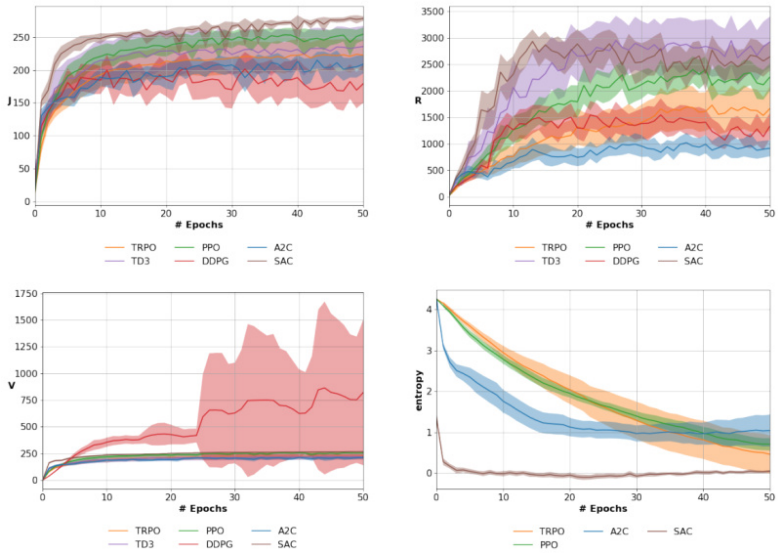


Figure 1: Discounted return (J), Cumulative return (R), Value function on the initial state (V), and policy entropy on the MuJoCo Hopper-v3 Task

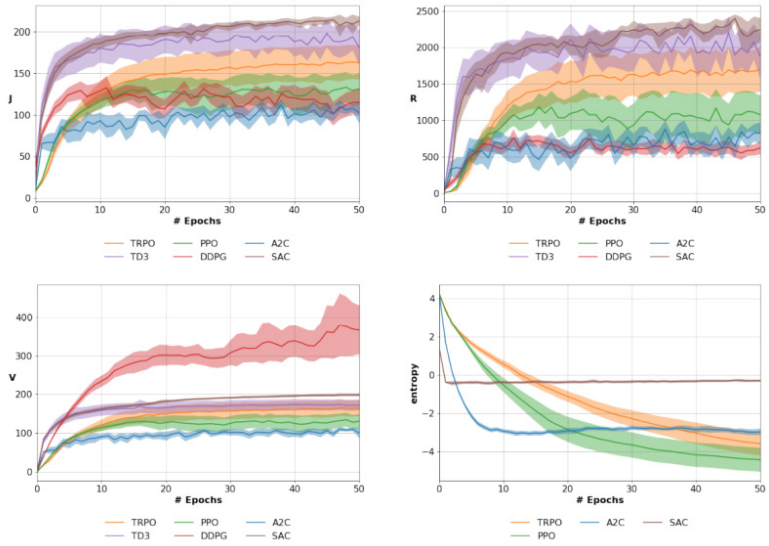


Figure 2: Discounted return (J), Cumulative return (R), Value function on the initial state (V), and policy entropy on the PyBullet HopeprBulletEnv-v0 Task

Publications

D'Eramo, C.; Tateo, D.; Bonarini, A.; Restelli, M.; Peters, J. Mushroomrl: Simplifying reinforcement learning research. Journal of Machine Learning Research, 22(131):1, 2021 https://www.jmlr.org/papers/v22/18-056.html

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg

Reference

[1] Lillicrap, T. P.; Hunt, J. J.; Pritzel, A.; Heess, N.; Erez, T; Tassa, Y.; Silver, D.; Wierstra, D.: “Continuous Control with Deep Reinforcement Learning,” in International Conference on Learning Representations (ICLR), 2016 https://doi.org/10.48550/arXiv.1509.02971

[2] Mnih, V.; Badia, A. P.; Mirza, M.; Graves, A.; Lillicrap, T.; Harley, T.; Silver, D.; Kavukcuoglu, K.: “Asynchronous methods for deep reinforcement learning,” in International conference on machine learning, pp. 1928–1937, PMLR, 2016 https://doi.org/10.48550/arXiv.1602.01783

[3] Schulman, J.; Levine, S.; Abbeel, P.; Jordan, M.; Moritz, P.: “Trust region policy optimization,” in International conference on machine learning, pp. 1889–1897, PMLR, 2015 https://doi.org/10.48550/arXiv.1502.05477 Focus to learn more

[5] Fujimoto, S.; Hoof, H.; Meger, D.: “Addressing function approximation error in actorcritic methods,” in International conference on machine learning, pp. 1587–1596, PMLR, 2018 https://proceedings.mlr.press/v80/fujimoto18a.html

[6] Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine,S.: “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” in International conference on machine learning (J. Dy and A. Krause, eds.), vol. 80, pp. 1861–1870, PMLR, 2018 https://proceedings.mlr.press/v80/haarnoja18b.html

Last Update: 2022-07-08 15:19