Karlsruher Institut für Technologie

BOSCH
Invented for life
Bosch Center for Artificial Intelligence

Autonome Lernende Roboter (ALR)
Prof. Gerhard Neumann

Project Type

- ☑ Master Thesis
- ☐ Bachelor Thesis
- ☐ Research Project

Supervisors

- Fabian Otto
- Fabian.Otto@de.bosch.com
- Maximilian Hüttenrauch
- m.huettenrauch@kit.edu

Difficulty

Algorithmic

Math

Application

Advanced Entropy Control for Efficient Exploration in Deep Reinforcement Learning

Description

The exploration-exploitation trade-off is one of the main problems in reinforcement learning (RL). While the agent needs to explore the problem space adequately, it eventually has to decide on exploiting the most promising direction. Vanilla policy gradient (PG) methods work well at exploiting a given direction, but often at the cost of insufficient exploration. Trust-region methods, such as PPO [3], have shown to improve upon this by providing more stability, yet they still suffer from this issue.
Recently, we introduced a novel approach to such trust regions [2], based on differentiable projections. Compared with PPO, those match or improve performance on standard RL tasks while being more robust to implementation details and allowing for more fine-grained control of updates. On another note, past work demonstrated that trust-region methods benefit from an additional entropy control, especially for complex exploration problems. This behavior is also already observable for the trust-region projections while only utilizing a simple form of entropy control.
The goal of this thesis is to explore the effects of complementing the trust-region projections with an advanced entropy control strategy. This could include, for example, adapting entropy control methods from stochastic search to the deep RL setting. Specifically, CMA-ES [1] uses a heuristic that observes the correlation of past updates to decide whether entropy should be increased or decreased. In the stochastic search setting, this measure was, so far, only leveraged for the parameter space. We are interested in extending it to the action space where exploration typically happens in reinforcement learning.

Tasks

- Literature Review: Getting familiar with the basics and reviewing existing approaches for advanced entropy control.
- Implementation: Reimplementing the existing entropy control approaches and fusing them with the trust-region layers.
- Evaluation: Choosing (and potentially implementing) suitable benchmark tasks with complex exploration behavior. Evaluating the performance in contrast to existing methods.

Qualifications

- Background in computer science, mathematics, physics, or similar.
- Experience with basics in machine learning (e. g. deep learning, (Lagrange) optimization, reinforcement learning)
- Experience with programming in Python (PyTorch is a plus).

References

[1] Nikolaus Hansen. The CMA evolution strategy: A tutorial. In *arXiv preprint*, 2016.

[2] Fabian Otto, Philipp Becker, Ngo Anh Vien, Hanna Carolin Ziesche, and Gerhard Neumann. Differentiable Trust Region Layers for Deep Reinforcement Learning. In *International Conference on Learning Representations*, 2021.

[3] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms. In *arXiv preprint*, 2017.