Benchmarking Provably Safe Reinforcement Learning Approaches

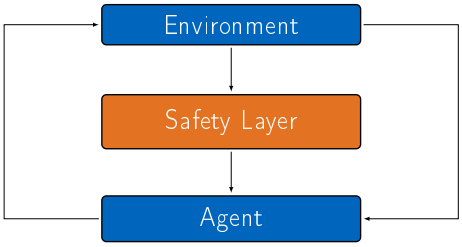
Technische Universität München

Fakultät für Informatik

Lehrstuhl für Echtzeitsysteme und Robotik

Background

Machine learning approaches emerged in recent year as computational resources and enough data is now available to produce good results with these techniques. But data-driven approaches like reinforcement learning are based on random exploration which can lead to unsafe and non-verifiable behaviors. This is problematic for real world tasks where unsafe behavior can lead to material damage or even human injuries. To overcome this challenge concepts are developed to adapt reinforcement learning in order to achieve safe reinforcement learning where the exploration and optimization is not entirely random anymore.



Safe reinforcement learning with safety layer.

However, the different safe reinforcement learning approaches are usually tested on different tasks and with different algorithms which makes it hard to compare their computation efficiency and transferability. Therefore, existing implementations have to be improved in a transferable way and tested on the same task to identify the benefits and challenges of the different approaches in detail.

Description

The goal of this thesis is to implement provably safe reinforcement learning approaches in order to benchmark them on the same learning algorithm and task. In particular, provably safe reinforcement learning implementations in the literature, mainly masking [1, 2], shielding [3] and using control barrier functions [4], have to be reviewed and adapted to make them transferable between reinforcement learning algorithms. Further, a common safety metric for the inverted pendulum task[1] has to be identified. The provably safe reinforcement learning implementation will be compared at the inverted pendulum task with a discrete action space. Optionally, the masking concept could be extended to continuous action spaces and thus a comparison for continuous actions spaces can be performed as well.

Tasks

- Perform a literature review on provably safe reinforcement learning approaches
- Familiarize the existing safe reinforcement learning implementation of the literature
- Improve and adapt the approaches in order to benchmark different reinforcement algorithms [5]
- Train and test the approaches on the inverted pendulum task
- *Optional:* Extend the masking concept to continuous action spaces

References

[1] Cheng-Yen Tang, Chien-Hung Liu, Woei-Kae Chen, and Shingchern D. You. Implementing action mask in proximal policy optimization (PPO) algorithm. *ICT Express*, 6(3):200 –

[1] https://gym.openai.com/envs/Pendulum-v0/

Supervisor:
Prof. Dr.-Ing. Matthias Althoff

Advisor:
Hanna Krasowski, M.Sc.
Xiao Wang, M.Sc.

Research project:
-

Type:
Bachelor's Thesis

Research area:
Safe reinforcement learning, machine learning

Programming language:
Python

Required skills:
Good programming skills, interest in reinforcement learning

Language:
English

Date of submission:
2. Februar 2021

For more information please contact us:

Phone: -

E-Mail: hanna.krasowski@tum.de

Internet: www.in.tum.de/i06

203, 2020.

[2] Hanna Krasowski, Xiao Wang, and Matthias Althoff. Safe Reinforcement Learning for Autonomous Lane Changing Using Set-Based Prediction. In *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2020.

[3] Mohammed Alshiekh, Roderick Bloem, Ruediger Ehlers, Bettina Koenighofer, Scott Niekum, and Ufuk Topcu. Safe Reinforcement Learning via Shielding. In *Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, 2018.

[4] Richard Cheng, Gabor Orosz, Richard M. Murray, and Joel W. Burdick. End-to-End Safe Reinforcement Learning through Barrier Functions for Safety-Critical Continuous Control Tasks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3387–3395, 2019.

[5] Ashley Hill, Antonin Raffin, Maximilian Ernestus, Adam Gleave, Anssi Kanervisto, Rene Traore, Prafulla Dhariwal, Christopher Hesse, Oleg Klimov, Alex Nichol, Matthias Plappert, Alec Radford, John Schulman, Szymon Sidor, and Yuhuai Wu. Stable baselines. https://github.com/hill-a/stable-baselines, 2018.

Technische Universität München

Fakultät für Informatik

Lehrstuhl für Echtzeitsysteme und Robotik