2nd collaboration workshop on Reinforcement Learning for Autonomous Accelerators (RL4AA'24)

Contribution ID: **43**

Type: **Contributed Talk**

Explainability in Reinforcement Learning: An Application for Powertrain Control

Wednesday, February 7, 2024 2:00 PM (30 minutes)

Reinforcement learning (RL), a subgroup of machine learning, has gained recognition for its astonishing success in complex games, however it has yet to show similar success in more real-world scenarios. In principle, the ability for RL to generalise past experience, act in real time, and its resilience to new states makes it particularly attractive as a robust decision-making support for real-world scenarios. However, such scenarios bring unique challenges that aren't present in the game-like domains, such as complex and contradictory reward functions and a necessity for explainability. In this presentation we will discuss some of these challenges in the context of using RL for automotive powertrain control. We will discuss the problem setup, including reward definition, as well as one approach to explainability. This approach is to first learn a neural network based policy (which can learn effectively and efficiently) and then extrace a rule-based policy (which is easier to interpret and can be directly implemented in current control software). The results are benchmarked with an optimised MATLAB policy, using a simulink simulation.

Possible contributed talk

Yes

Are you a student?

No

Author: LAFLAMME, Catherine (Fraunhofer Austria Research GmbH)

Co-authors: DOPPLER, Jörg (Bosch Engineering GmbH); DOMNIKA, Sven (Bosch Engineering GmbH); CZARNETZKI, Leonhard (Fraunhofer Austria Research GmbH); PALVOLGYI, Bence (SZTAKI: Institute for Computer Science and Control); VIHAROS, Zsolt (SZTAKI: Institute for Computer Science and Control)

Presenter: LAFLAMME, Catherine (Fraunhofer Austria Research GmbH)

Session Classification: Contributed Talks