Model-based Reinforcement Learning in Shadow Mode

Background

With its successes in recent years, reinforcement learning (RL) has become very popular for training agents to complete a wide variety of tasks. However, it is not yet competitive for many cyber-physical systems, such as robotics, process automation, and power systems. One of the main reasons is that training on a system with physical components cannot be accelerated because simulation models do not exist or the simulation-to-reality gap is too big. During the long training time required, expensive equipment cannot be used and might even be damaged due to inappropriate actions of the reinforcement learning agent.

We want to address this problem by training a reinforcement agent in a so-called shadow mode while the system is operated by a conventional controller, which instantaneously performs well as it does not have to be trained. While learning the actual control task in shadow mode, the agent simultaneously learns in which situations it performs better than the conventional controller. and takes over, once confident enough and can keep learning, thus increasing the number of situations where it outperforms the baseline. We therefore always ensure that the performance is superior compared to only using conventional controllers or reinforcement learning and minimizes regret during learning.

One relevant domain for applying this principle is Autonomous Driving.

Description

The focus of this work lies on using model-based learning to not only improve training but also allow for a better decision algorithm for handing control to the RL agent. The idea of training in shadow mode is useful for systems which cannot be fully and accurately simulated and can be applied to all systems for which some (sub-optimal) controller already exists. Initially, the system is controlled by the existing baseline controller. The RL agent is trained by simulating a subsystem, for which the dynamics are known and under the control of the agent. The observations of the RL agent are taken from the actual environment by adapting them according to the simulation. For example, a human might steer a vehicle, while the RL agent trains in the background. It simulates the ego-vehicle (the vehicle it is steering) and uses the real observations (such as distances to other vehicles) to infer the observations for its next state. A decision algorithm decides which action is executed on the real vehicle for each time step, i.e. it decides when the RL agent takes over.

In model-based RL methods we want to explicitly learn the dynamics of our system. While it can be difficult to train such a model, it enables us to use the predictions of the model not only for standard training and to obtain better sample complexity, but also to improve our shadow training: It provides predictions that can be used to adapt observations and rewards during shadowing and uncertainty estimates that can be used to make more comprehensive decisions on whether to use the baseline agent or the RL agent on the actual system. In this work, you will implement model-based methods for RL in shadow mode and evaluate the performance of the training algorithms compared to existing model-free approaches.

All programming will be done in our python-based CommonRoad-RL library.

Technical University of Munich

Department of Informatics

Chair of Robotics, Artificial Intelligence and Real-time Systems

Supervisor:
Prof. Dr.-Ing. Matthias Althoff

Advisor:
Philipp Gassert

Research project:
Reinforcement Learning in Shadow Mode

Type:
BA/MA

Research area:
Reinforcement Learning (model-based)

Programming language:
Python

Required skills:
some familiarity with RL

Language:
English, German

Date of submission:
6. August 2023

For more information please contact us:

Phone: +49 (89) 289 18100

E-Mail: philipp.gassert@tum.de

Website:
www.ce.cit.tum.de/air/home/

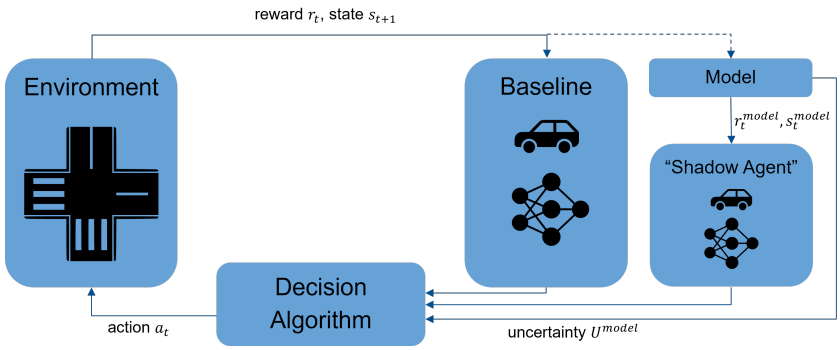

Figure 1: Framework for RL in shadow mode using a model-based approach

Tasks

- Literature review on model-based RL and uncertainty estimation
- Setup of model-based method in CommonRoad-RL's shadow mode feature
- Training and tuning of models in shadow mode
- Evaluation of performance and identification of exemplary scenarios

References

[1] Owen Lockwood and Mei Si. A review of uncertainty for deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 18, pages 155–162, 2022.

[2] Thomas M. Moerland, Joost Broekens, and Catholijn M. Jonker. Model-based reinforcement learning: A survey. *CoRR*, abs/2006.16712, 2020.

[3] Tianhe Yu, Garrett Thomas, Lantao Yu, Stefano Ermon, James Y Zou, Sergey Levine, Chelsea Finn, and Tengyu Ma. Mopo: Model-based offline policy optimization. *Advances in Neural Information Processing Systems*, 33:14129–14142, 2020.

Technical University of Munich

Department of Informatics

Chair of Robotics, Artificial Intelligence and Real-time Systems