KI WISSEN
Automotive AI Powered by Knowledge

2.12 Behavior-Conditioned Driving Policies

Tim Joseph | FZI Forschungszentrum Informatik

Dream to Drive

Current reinforcement learning (RL)-based driving policies learn a fixed behavior resulting from a fixed reward function. To alter the behavior after deployment, the policy needs to be fine-tuned or trained from scratch with a modified reward function, which is cumbersome. On the other hand, human drivers can adjust their preference for safety, comfort, and efficiency according to the current driving situation. In this work, we explore the concept of conditional RL policies in the context of autonomous driving.

The key idea to facilitate such a driving policy is to explicitly condition the policy on the parameters of the reward function.

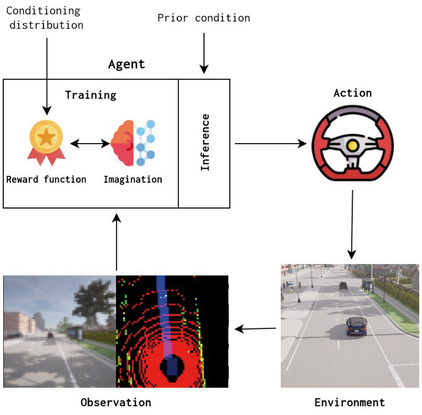

Figure 1: Our reinforcement learning agent interacts with the environment through direct steering and acceleration commands. The agent acts based on sensor inputs. In training, we randomly condition the reward function of the agent so that at inference time, we can adapt the agent's behavior by changing the conditioning prior.

This enables us to modify the preferences of different parts of the reward function, e.g., lane distance, target speed, or other factors during inference time, instead of learning a new policy for each configuration. Furthermore, since our approach is model-based, we can gather data in the actual environment using a fixed reward function that might be focused on safety but use the world model to learn different behaviors in imagination, which we can subsequently use during inference.

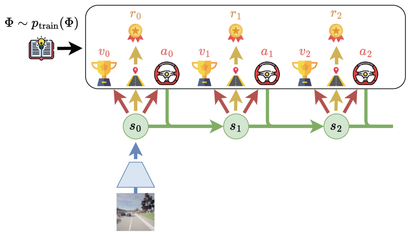

Figure 3: We sample different reward-shaping parameters in the imagination of DreamerV2 and train a universal actor-critic policy to learn many different behaviors.

Compared to a model predictive control based approach, our method is not constrained to some fixed planning horizon. Furthermore, we can change the policy's behavior without retraining, as opposed to other RL-based approaches.

Our method is based on DreamerV2[1] a state-of-the-art model-based RL agent that simultaneously learns a world model and policy. We extend this approach such that it can be conditioned explicitly on speed and maximum route deviation. Without loss of generality, the agent can be conditioned on other factors of the reward function. We evaluate our approach on CARLA[2], and show that we are able to alter the behavior without retraining the agent by changing the aforementioned parameters. We show that we can effectively train a single policy that displays diverse behaviors dependent on the conditioning.

References:

[1] D. Hafner, T. P. Lillicrap, M. Norouzi, and J. Ba, “Mastering atari with discrete world models,” ICLR, 2021

[2] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “Carla: An open urban driving simulator,” CoRL. PMLR, 2017, pp. 1–16

Figure 2: The agent in the top row is conditioned with a maximum route deviation of 0.5m. It stays in its lane as it can not overtake the vehicle in front without deviating too far from the route. In contrast, the agent in the bottom row is allowed to deviate up to 4.5m. Instead of waiting in traffic, it overtakes to reach its speed target.

Partners

External partners

For more information contact:
joseph@fzi.de

KI Wissen is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

Funded by the European Union
NextGenerationEU