Reward Definitions in Reinforcement Learning for Traffic Light Control

Cian Jansen

Delft University of Technology

Reinforcement Learning

Reinforcement Learning (RL) is a kind of machine learning where an agent operates in an environment. An agent takes actions that alter the state of the environment, and gets feedback on how desireable each state of the environment is by receiving rewards. The agent tries to optimize a policy so that reward over time is maximized.

Reward Functions

A reward function used in RL determines what is seen as “desirable”. By rewarding/punishing certain actions or states, the reward function controls what kind of behavior gets reinforced/avoided.

Reward Shaping

Sometimes a reward function can not be created that directly rewards desirable behavior. The average speed of vehicles can only be calculated after a simulation is finished, whereas a reward needs to be calculated during every step. In this case, reward shaping can be used to reward behavior that is seen as indicative of a higher average speed at the end. A good shaping reward can be calculated at any time, even during a simulation, and reinforces actions that ultimately improve desirability.

“What effect can different reward functions have on the performance of a Reinforcement Learning system for traffic light control?”

[1] Establish the goal of an RL agent controlling traffic:

The goal is to minimize average travel time by maximizing the Average Speed of all vehicles

[3] Test and compare agents trained with different rewards in SUMO traffic simulator:

Compare which reward function leads an agent to the highest average speed of all cars consistently. Best performance + stability comes from a combination of Waiting Time and Speed Change

[2] Draft reward functions that estimate average speed:

Functions used:

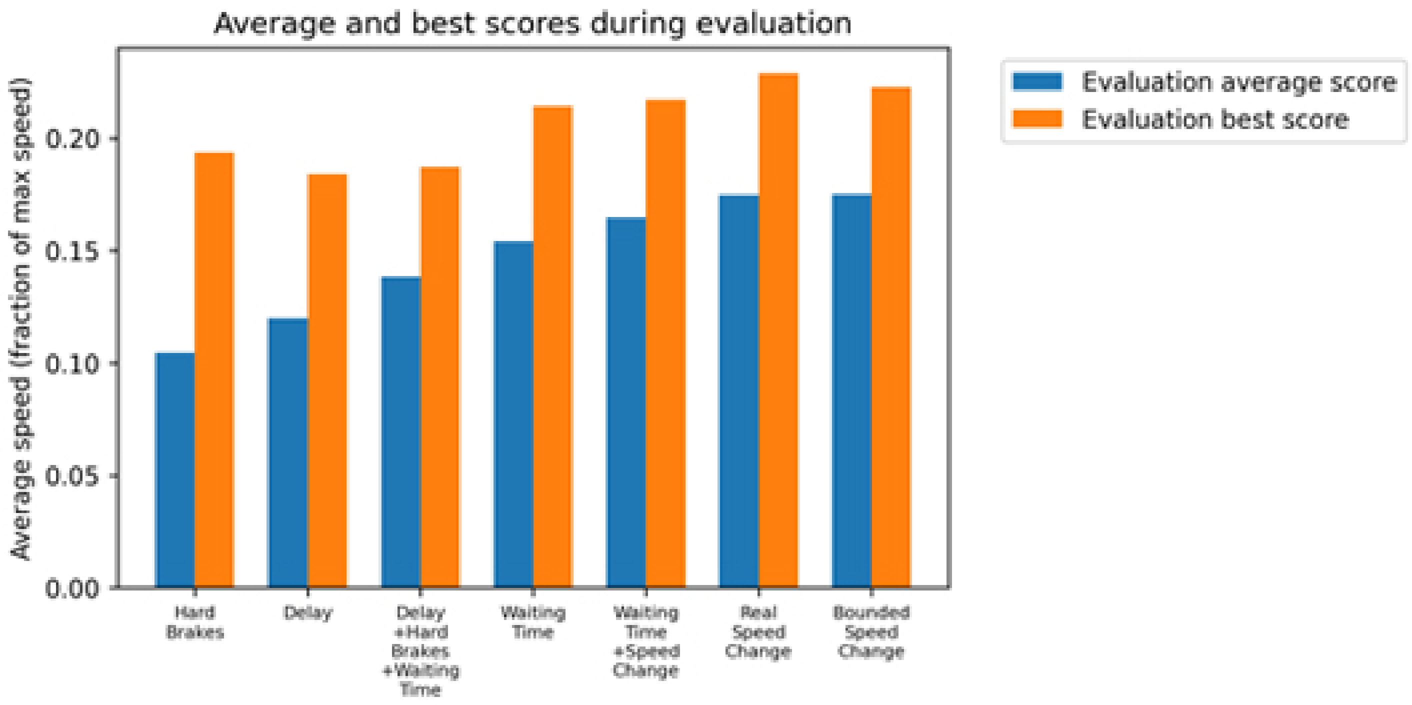
Minimize Waiting Time

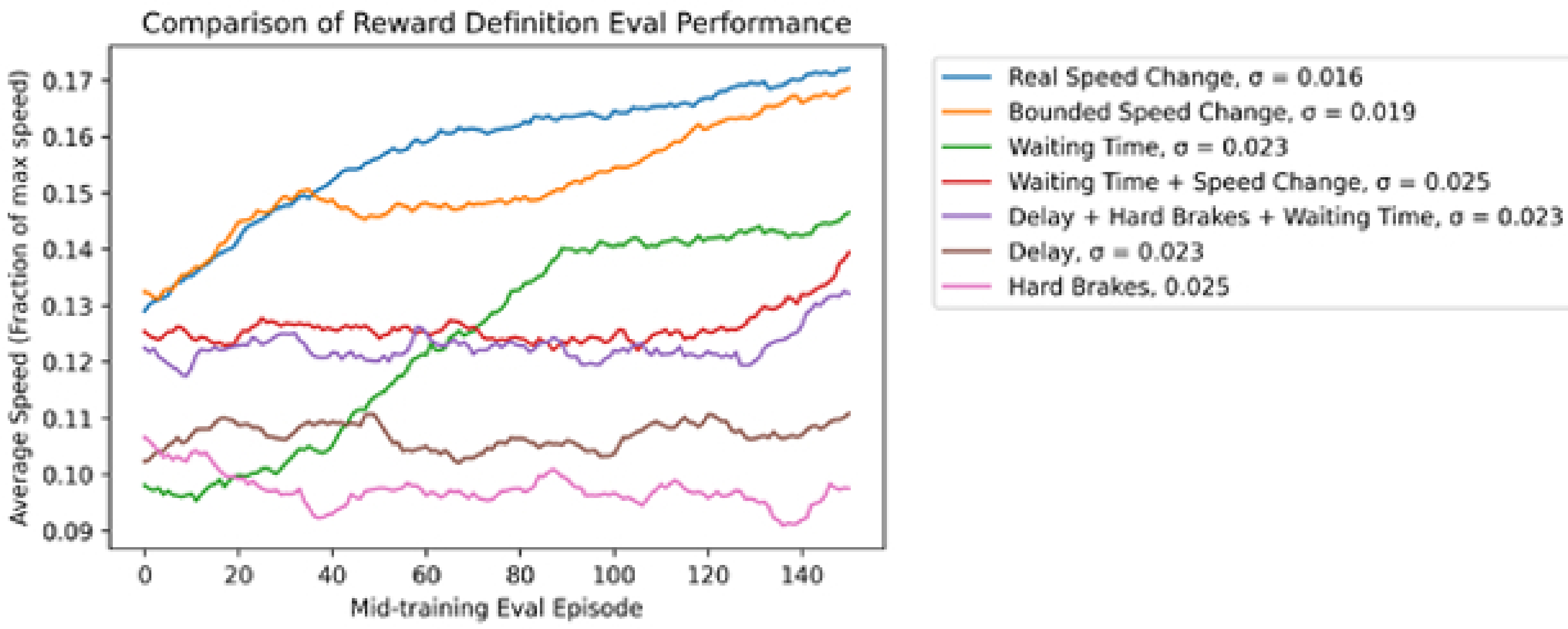
Minimize Delay (vehicle speed ÷ allowed speed)

Maximize positive changes in vehicle speed

Minimize Hard brakes

Punish changing of traffic lights (from green to red or vice versa)





[4] Recommendation for further research:

Machine learning techniques like Linear Regression could be used to find a reward function that is perfectly aligned with optimizing maximum speed

23-06-2020
CSE3000
Supervisors:
F. Oliehoek
M. Suau de Castro