Traffic signal control environment at intersections

Yifan Wang, Yu Li

LUH|AI

Leibniz Universität Hannover

1 TL;DR

- We propose an RL environment used to optimize traffic by letting and agent learn a policy for setting traffic lights.
- Goal: reduction of CO2-emissions and waiting times/congestion

2 Motivation & Problem Setting

Motivation

- IntelliLight: A Reinforcement Learning Approach for Intelligent Traffic Light Control [Wei et al. 2018]
- Adaptive traffic signal control system using composite reward architecture based deep reinforcement learning [Nower et al. 2020]

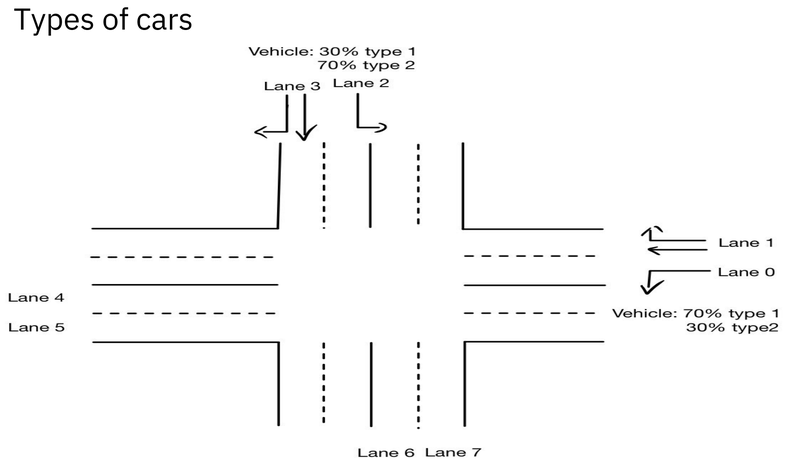
Problem Setting

Firstly we want to simulate an environment that is relevant to the intersection. And add some random variables to the environment. We also added some car attributes as observed attributes for reinforcement learning. In order to expect the reinforcement learning algorithm to also take into **account the CO2 emissions** and the **waiting time** of the cars during the optimization process. as a way to **ease traffic congestion**.

3 Approach

MDP Modell

- state space:
  - current traffic light phase
  - queue length on each lane
  - Waiting time for all cars
  - CO2 emissions from all cars
- action space:
  - 4 traffic light phases
- Transition Probabilities:
  - traffic flow
  - Types of cars



Reward Function

- Queue reward:

$$R_Q = \sum_{j \in Lin}^{n} (N_{j,t-1} - N_{j,t})$$

- Waiting time reward:

$$R_T = 0{,}01 * \sum_{j \in Lin}^{n} (1/T_{wait})$$

- CO2 reward:

$$R_c = \sum_{j \in Lin}^{n} (C_{j,t-1} - C_{j,t})$$

Test

- Tests were conducted in a test environment using different random parameters and compared with the results of FIXED CHANGE.
- Testing using **PPO** with **A2C** as a reinforcement learning algorithm.
- Design **different reward functions** for the same target. Compare their optimization differences.

4 Key Insights

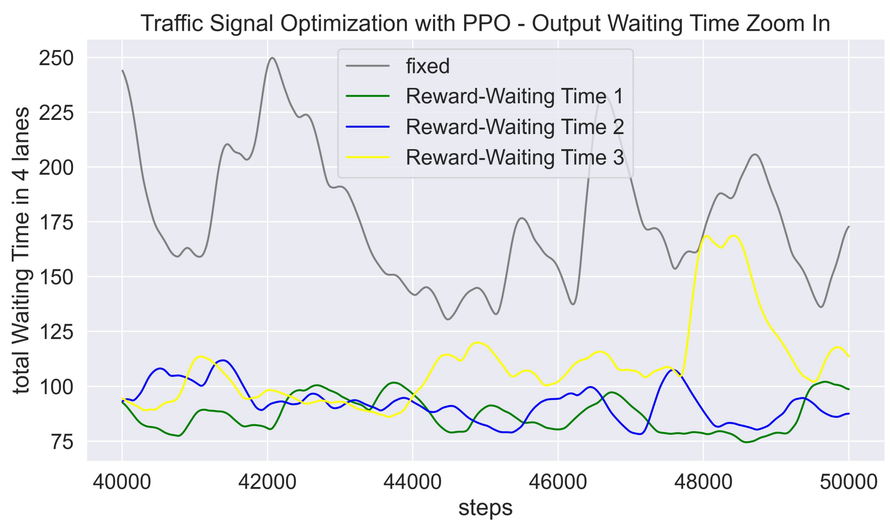
- Testing in our designed environment, the PPO performed better.

  [Mean value after convergence (10000 steps).]

	Queue	Waiting time	CO2 Emission
Fixed	59,38	181,60	1497,56
PPO	41,50	87,59	743,51
A2C	42,60	94,14	792.02

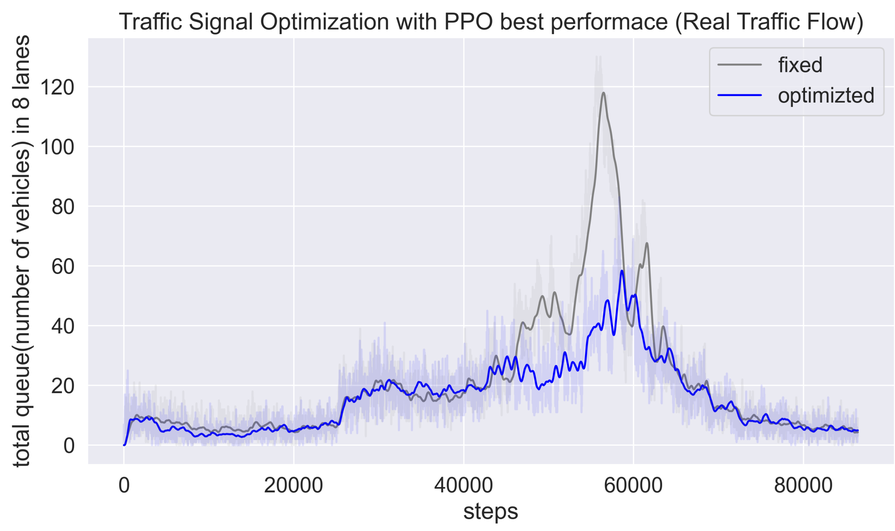
- A number of different Reward functions were tested and the best Reward function was selected in terms of mean as well as variance. Calculating the state difference and the inverse as a reward function usually yields better stability.

  [Waiting time curves using different reward functions.]



- Expanding on the simple environment, the test was conducted using actual traffic data collected as a basis, with eight lanes having independent traffic flows. The results are judged using the column of cars as the basis of the test.

  [Testing with real data.]



- Different steps in the test represent different time periods. During periods of low traffic volume, the difference is not significant. However, during the hours with more traffic, better optimization results can be obtained using the model.

5 Future Works

- The discrete model is transformed into a continuous model so that the behavior of the traffic lights can be controlled in more detail.
- Finding **more relevant** reward functions and car parameters to enrich the design environment.