Traffic Flow Optimization using Reinforcement Learning

Erwin Walraven

Algorithmics group, Delft University of Technology, The Netherlands

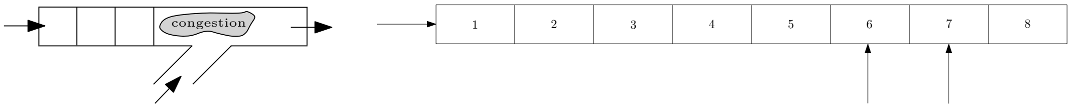
Abstract

Traffic congestion is an important problem that causes unnecessary delay, environmental pollution and more fuel consumption. In this thesis project we address this problem by proposing a new method to assign speed limits to highways. Our approach combines a macroscopic traffic flow model with Q-learning to generate speed limit policies, and we show that traffic predictions can be included in our method to regulate traffic flow more efficiently.

1 Problem Description

We study the problem of finding speed limits that can be assigned to a unidirectional highway to reduce congestion. An example highway is shown in Figure 1a, where the arrows indicate the direction of the vehicle flow. The rectangles represent a partitioning of the highway into sections. If the demand flow of the origin and the on-ramp is high, congestion may arise in the shaded region.

We aim to find a method that defines the speed limits that should be assigned to the sections, such that the global delay car drivers incur is minimized. A few additional constraints can be imposed. For example, speed limits should be increased and decreased smoothly and alternating sequences of speed limits should be prevented. Note that the schematic representation shown in Figure 1a can easily be generalized to highways with more sections and on-ramps, such as the highway shown in Figure 1b. The problem we defined occurs naturally in practice near on-ramps and road interchanges.



(a) Unidirectional highway with one on-ramp.

(b) Unidirectional highway consisting of eight sections.

Figure 1: Schematic representation of highways.

2 Learning Speed Limit Policies

We define a traffic flow optimization problem as a Markov Decision Process (MDP) [2]. Since we are unable to let our algorithm interact with a real highway, we use a macroscopic traffic flow model as a representation of the environment in which the agent assigns speed limits. The model we use is called METANET, and provides a set of formulas to compute speed, density and flow values for highway sections in discrete time steps [1]. It can be adapted such that speed limits are taken into account.

In the MDP formulation we consider the unidirectional highway shown in Figure 1b. The action space A consists of the speed limits 60, 80, 100 and 120 that can be assigned to section 3 to 6. We

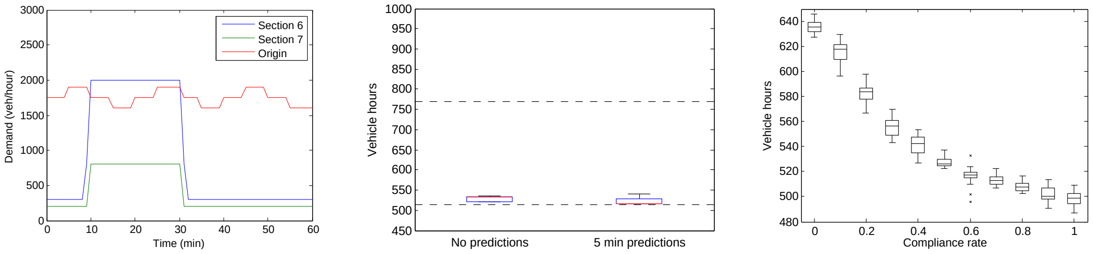
The full version of the thesis can be found in [3].

define a six dimensional state space S, where each state contains the last two speed limits that have been selected, as well as the speed values of section 4 to 7. The reward function represents a punishment received by the agent if speed limits are decreased too much, or if they are alternating. If there is no congestion, no punishment is given. In all other cases, the agent receives a punishment proportional to the number of vehicle hours that vehicles have made after the last speed limit assignment. The reward function ensures that speed limits are only activated if needed, and implicitly defines that the number of vehicle hours should be minimized.

We apply Q-learning to find a policy $\pi : S \rightarrow A$ for a given traffic scenario [4]. We enhanced this algorithm with artificial neural networks as a value function approximation technique, to handle a larger state space and it generalizes learning experiences. In addition to our work on speed limit policy learning, we made a first step to generalize this idea to multi-agent coordination of variable speed limits. In this approach, there is one agent associated with each highway section. We have also shown that knowledge regarding traffic flow regulation, represented by a policy, can be reused to learn policies for new traffic scenarios more efficiently. To investigate whether traffic predictions can be integrated in our method, we included model-based traffic predictions in the state description, and we have shown that policy quality improves.

3 Simulation Results

To evaluate our method, we defined several traffic scenarios, for which we have shown that the quality of the generated policies is close to lower bounds we were able to compute. Figure 2a shows the demand profile of one of the scenarios. The distribution of policy quality for this scenario is shown in Figure 2b. The horizontal lines represent the baseline and optimal number of vehicle hours, which is considered as a lower bound. It shows that the generated policies are close to the lower bound we found, and the policy quality is slightly better if model-based predictions are included in the state description. Additionally, we performed a case study involving a more realistic simulation of the A67 highway in The Netherlands. We derived a real traffic demand pattern for the A67 from NDW, the Dutch national database containing historical traffic data, and we performed a simulation in a microscopic simulation environment. We applied a speed limit policy during the simulations, for different compliance rates. The results are presented in Figure 2c, which shows that congestion can be reduced, and that small compliance rates already yield a significant improvement.



(a) Scenario demand profile.

(b) Policy quality distribution.

(c) A67 scenario simulation.

Figure 2: Simulation results.

References

[1] A. Messner and M. Papageorgiou. METANET: A macroscopic simulation program for motorway networks. *Traffic Engineering & Control*, 31(8-9):466–470, 1990.

[2] M.L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994.

[3] E. Walraven. Traffic Flow Optimization using Reinforcement Learning. Master's thesis, Delft University of Technology, 2014.

[4] C.J.C.H. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8(3-4):279–292, 1992.