Volume 19 • Issue 3 • March 2015

Traffic Flow Optimization using Reinforcement Learning

Erwin Walraven

The MSc thesis 'Traffic Flow Optimization using Reinforcement Learning' [4] presents new algorithms to compute speed limits for highways, based on artificial intelligence. The resulting methods have been implemented in practice, and the thesis received the prestigious Ngi-NGN Informatie Scriptieprijs 2014, worth 5000 euros, from the Royal Holland Society of Sciences and Humanities.

Intelligent Solutions to Reduce Traffic Congestion

In modern society many people are faced with it every day: traffic congestion. When driving home after work, people often encounter traffic jams, causing not only an increased travel time, but also more fuel consumption and environmental pollution. A straightforward solution would be expanding the existing infrastructure to increase its traffic capacity, but due to space and budget limitations this is not always feasible in practice. In the last decade, Intelligent Transportation Systems have emerged as a potential solution to address this problem [1]. Designers of such intelligent systems aim to make traffic flow safer and more coordinated by using, for instance, in-car information systems and mobile applications. In this thesis project we contribute to the development of Smoover, a new Intelligent Transportation System in the Netherlands. Smoover can be used as a regular smartphone navigation app, but it also shows notifications with personalized advice regarding optimal driving speed, depending on the current position of the vehicle and predictions of congestion arising in the near future.

The problem we study is schematically shown in Figure 1, where the arrows indicate the direction of the vehicle flow. If congestion is either detected or predicted in the shaded area, then speed limits should be assigned to the preceding sections. However, it is initially unknown when and where speed limits should be assigned. Finding the right speed limits to reduce congestion is a problem that has been extensively studied in the past, but several existing methods are computationally expensive and limited attention has been paid to the application of artificial intelligence techniques to solve this problem. Therefore, we choose a rather different approach. In our work we observe that the assignment of speed limits to highway sections can be formulated as a sequential decision making problem, and we show that reinforcement learning can be applied to find a solution.

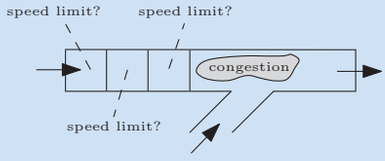


Figure 1: Highway with an on-ramp and three preceding sections.

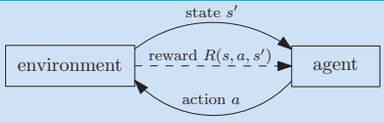


Figure 2: Agent executing action a in state s.

Reinforcement Learning

Reinforcement learning is a branch of machine learning that deals with agents interacting with their environment to optimize the cumulative reward they receive from the environment. Agents interact with their environment by executing actions, and the environment provides limited feedback in terms of a single reward. In reinforcement learning, agents are able to automatically learn how they should act, in such a way that the expected sum of future rewards is optimized.

Typically, the interaction between an agent and its environment is represented by a Markov Decision Process (MDP) [3]. In the MDP formalism, the state of the environment is represented by a state $s \in S$. An action $a \in A$ can be executed to change the state of the environment. The transition function $T(s, a, s')$ defines probability that the state transitions from state s to s' after executing action a. After every state transition, the agent receives a reward from the environment, defined by the reward function $R(s, a, s')$. An example of the interaction between an agent and the environment is shown in Figure 2, where the state transitions from s to s' after executing action a. A solution to a MDP is a policy $\pi : S \rightarrow A$ defining an optimal action for each state, such that the expected cumulative reward is optimized.

If the transition function T and reward function R are known, then an optimal policy can be found using dynamic programming (e.g., the value iteration algorithm). However, if the transition and reward function are unknown, then reinforcement learning can be applied to learn a policy from the sequence of rewards received by the agent. In the thesis we only consider the case without prior knowledge about the transition and reward function, and we apply the Q-learning algorithm [5] to learn policies.

Formulating Traffic Flow Optimization as MDP

Assigning speed limits to highway sections can be formulated as a Markov Decision Process. This section gives an example formulation of this problem for an highway consisting of eight sections and two on-ramps, as depicted in Figure 3. If the traffic demand of the origin and on-ramps is high, then congestion will arise near section 6 and 7, and propagates upstream (i.e., to the left). To reduce congestion in case of high demands, speed limits can be issued. For example, if congestion is detected or predicted near section 6 and 7, then the speed of section 2 to 6 can be temporarily decreased.

In the MDP formulation, the highway represents the environment in which the agent assigns speed limits to sections. Therefore, the action space A consists of speed limits values that can be assigned to sections 2 to 6: $A = \{60, 70, 80, 90, 100, 120\}$. A state is a vector representing the current highway

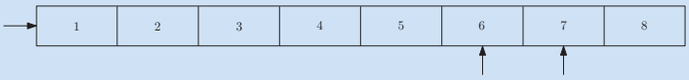


Figure 3: Example highway stretch with an origin and two on-ramps.

state. More formally, the state $s_t \in S$ of the highway shown in Figure 3 at time t is defined as follows:

$$s_t = \left(\frac{v_1(t)}{v_f}, \frac{v_2(t)}{v_f}, \frac{v_3(t)}{v_f}, \frac{v_4(t)}{v_f}, \frac{v_5(t)}{v_f}, \frac{v_6(t)}{v_f}, \frac{v_7(t)}{v_f}, \frac{v_8(t)}{v_f} \right)$$

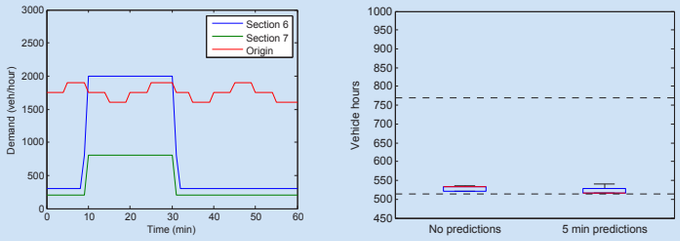
where $v_j(t)$ denotes the speed of section j at time t. The free flow speed v_f, which is the speed without traffic congestion, is used as a normalizing constant. The agent receives rewards equal to the number of vehicle hours after the last speed limit assignment, which is a value proportional to the delay incurred by car drivers. This means that the agent will be able to reduce congestion by assigning speed limits that minimize the cumulative reward, and thus delay. A policy $\pi : S \rightarrow A$ is a mapping from highway states to speed limits, which we call a speed limit policy.

Learning Speed Limit Policies

Learning algorithms cannot be used to interact directly with real highways, because this would lead to dangerous situations. Therefore, we use the traffic flow simulation model METANET [2] as a representation of the highway, and the Q-learning algorithm interacts with this model to obtain speed limit policies. To evaluate our method, we can define a traffic scenario, for which we show that the quality of the generated policies is close to the optimal speed limit assignment. In general it is intractable to compute optimal solutions, but for small scenarios optimal speed limits can be obtained by enumerating and evaluating all solutions.

Figure 5a shows the demand profile of an example scenario, defining the demand flows of the origin and both on-ramps during 60 minutes. The distribution of policy quality for this scenario is shown in Figure 5b. The horizontal lines represent the number of vehicle hours without control (i.e., the baseline) and the number of vehicle hours realized by the optimal speed limit assignment, which is considered as a lower bound. It shows that the generated policies are close to the optimal solution, and they reduce congestion significantly in comparison to the number of vehicle hours without speed control. In the thesis we also study the influence of predictions on policy quality. Figure 6 illustrates how congestion is reduced in the scenario defined by Figure 5a. It shows how speed changes over time, and the colors represent the speed measures. In the leftmost picture there is no speed control, and the picture in the middle shows the same scenario with speed control based on speed limit

Figure 4: Speed advice shown as a notification in the Smoover app.



(a) Scenario demand profile. (b) Policy quality distribution.

Figure 5: Scenario demand profile and policy quality.

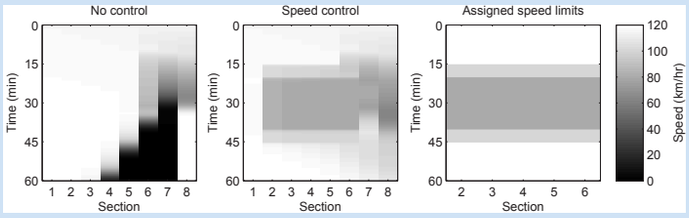


Figure 6: Scenario simulation without control (left) and with control (middle) using speed limits assigned to section 2 to 6 (right).

policies. Without control section 4 to 7 become congested after 60 minutes, whereas the policies ensure that congestion resolves within an hour.

Smoover: Smart Sharing Smooth Driving

Speed limit policies have been used to build Smoover[1], a new Intelligent Transportation System. Smoover is a regular smartphone application to navigate to a destination, but also shows notifications regarding advised driving speed if congestion is either detected or predicted. A photo of the running system is shown in Figure 4, where an advice is presented as a notification. The speed limits shown are based on a set of precomputed speed limit policies. The app is currently being tested on the A67 highway, and until February 2015 participants traveled more than 190000 km while using Smoover. Although the pilot study focuses on the A67, Smoover can be used on any highway in the Netherlands.

Acknowledgements

The thesis project was carried out within the Algorithmics group, under supervision of Matthijs Spaan, in collaboration with Bram Bakker from Cygnify.

Currently we are working on planning algorithms for smart energy grids. If you are interested in an MSc project in this field, or if you have any other question, feel free to contact me: *e.m.p.walraven@tudelft.nl*.

References

[1] L. Figueiredo, I. Jesus, J. Machado, J.R. Ferreira, and J.L. Martins de Carvalho. Towards the Development of Intelligent Transportation Systems. In *Intelligent Transportation Systems*, volume 88, pages 1206–1211, 2001.

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

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

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

[5] C.J.C.H. Watkins. *Learning from Delayed Rewards*. PhD thesis, King's College, Cambridge, UK, 1989.

[1] Smoover can be downloaded from Google Play. More info: smoover.nl.