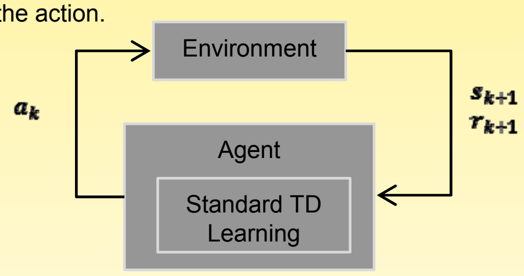
The Netherlands Research School for Transport, Infrastructure and Logistics
TRAIL

Freeway Traffic Control Using Q-learning

- A standard reinforcement learning (RL) algorithm is applied to control the density of a freeway via ramp metering in a macroscopic level.
- RL algorithms are effective tools for letting an agent learn from its experiences generated by its interaction with an environment.
- The performance of the algorithm as well as its robustness against communication failure is studied. The results of the simulations demonstrated the effectiveness the technique.

- In RL, the learner perceives environment state, takes an action and receives a scalar signal providing evaluative information on the quality of the action.



- The signal does not provide any instructive information on the best behavior in that state.
- At each time step, the scalar signal can be positive, negative or zero.
- The goal for the agent is to maximize the expected cumulative discounted rewards by finding an optimal action selection.
- The learner starts with almost random actions, but by seeking a balance between exploration and exploitation, gradually finds actions that lead to high values of reward function.
- Knowing the value of each state, which is the expected long-term reward that can be earned when starting from that state, the agent can choose the best action to take.

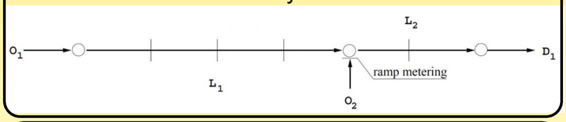
- Temporal difference (TD) methods are a class of incremental learning procedures that are designed to learn the value function.
- Among them, table lookup representation of value function is widely used. For example, the Q-learning algorithm stores the return obtained by taking action in state and choosing the greedy action w.r.t. the current Q-values.
- Q-learning recursively updates the estimate of values of state-action pairs.

- States: The state of the network is represented by two variables namely the density on the downstream of the diversion point and current metering rate.
- The density and the ramp metering rate are discretized to 11 equispaced grid points.

- Actions: The agent has three actions {-0.1, 0, 0.1}. The ramp metering rate is a finite number of distinct values in [0,1].

- Rewards function: The reward can be either positive or negative, in accordance with the outcome, based on whether a benefit or penalty is accrued.
- Since the usual control goal is outflow maximization, here at each time step, the reward is defined as a function of outflow of the network.

Case study network:

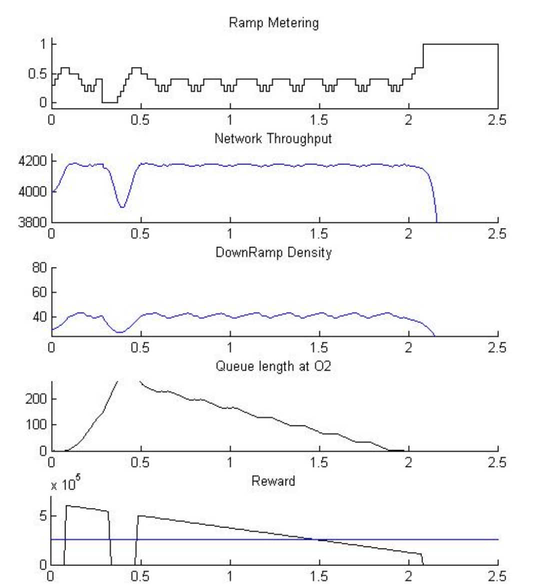


Simulation Results

•1000 iterations

•The traffic flow throughput volume is maintained in capacity during the high demands.

•To see if the agent can cope with a communication failure (that prevents the agent in communicating with the measures for some time) or any other possible failures, the ramp metering rate is set to 0 for about 5 minuets.



(a) Applied ramp metering control, (b) Resulted flow throughput, (c) Density at fist segment of L2, (d) Queue length at O2, (e) Obtained reward values

Future work: An interesting setting of ramp metering control problem is to pose maximum permissible queue length. We will work to extend the Q-learning based density control to the scenario where the maximum on ramp queue length is bounded.

Mohsen Davarynejad, Andreas Hegyi,
Jos Vrancken, Yubin Wang

TU Delft
Delft University of Technology

UNIVERSITY OF TWENTE.

Erasmus Universiteit Rotterdam

TU/e Technische Universiteit Eindhoven University of Technology

Radboud University Nijmegen