A Model-based Deep Q-Learning Algorithm for Traffic Signal Timing Control at Isolated Intersections

Yun Yuan, Xianfeng Terry Yang* (x.yang@utah.edu), Hao Wang, Tian Zhao, Yang Liu
Department of Civil and Environmental Engineering, University of Utah

INTRODUCTION
To contend the considerable externalities of traffic congestion, researchers and engineers have made great efforts to improve traffic-responsive algorithms for Adaptive Traffic Signal Control (ATSC). Deep reinforcement learning (DRL) methods have been tested on simplified traffic light timing problem and show promising potentials in addressing the curse of dimensionality. However, previous studies ignore the limitation of sensors and heavily rely on simulators. The following figure shows the framework of the DRL methods.

METHODS

- **Advantages**
  - the effectiveness of RL and DRL methods (deep policy gradient, value estimation, Q-learning) in traffic signal timing in simulated environment
  - the effectiveness of traffic simulators (such as VISSIM, SUMO) in developing RL methods

- **Disadvantages**
  - computational cost of the simulation sampling
  - heavily rely on simulators, which cannot provide plausible results despite costly highly-informative data and fine-tuned parameters
  - not based on real data or experiments in real-world scenarios and are hard to implement due to simplistic assumptions

This paper customizes a Deep Q-learning Learning (DQL) method to optimize traffic signal timings at single intersections, where a Deep Q Network (DQN) is proposed to estimate the value function (i.e., delays) and another Deep Neural Network is used to synthesize the future state. Then the machine learning-based methods are evaluated on a real-world case with Automatic Number-Plate Recognition (ANPR) data.

- **a)** A deep reinforcement learning algorithm is proposed with vehicle delay distribution state presentation and partially observed flawed inputs.
- **b)** A deep model is proposed to speed up the computation.
- **c)** ANPR field data is used for training the deep neural networks.
- **d)** The field SCOOT data is used as the baseline for benchmarking.

RESULTS/DISCUSSION

- **The compatible, realistic simulator and deep model**
  - Note it takes about 40 hours on average to test on the simulator SUMO while 1.2 hours on the deep model on a workstation with Intel (R) Xeon CPU.

- **Performance of the proposed deep model**
  - The figure shows the deep model can accurately estimate the delay distribution of 6 lane groups (out of 10 lane groups) at a 4-leg intersection in a morning peak cycle in comparison to the real data, where NB, SB, L, T, and R represent northbound, southbound, left-turn, through, right-turn, respectively. The y-axis is for the probability of that the corresponding vehicle appears in this cycle. The average R2 and relative mean squared error between the observed and synthesized sample are 0.69 and 0.0072, respectively.

- **Performance benchmarking of the proposed DRL method**
  - All the proposed DRL operations

CONCLUSIONS

- **This paper customizes a Deep Q-learning Learning (DQL) method to optimize traffic signal timings at single intersections and contributes in**
  - a) Designing better state representation (vehicle delay distribution) and reasonable assumptions (compatible across simulators, deep neural networks and real sensors);
  - b) Developing a deep model to speed up the sampling
  - c) Testing the proposed methods with data from real-world scenarios (ANPR data);
  - d) Comparing with commercial systems (SCOOT, SYNCHRO)

Experiments show the machine learning-based model can predict the traffic state in limited computational time and the deep Q-learning algorithm is 3.9% better than the field experiment performance from the adaptive control system, SCOOT, and 22% better than the time-of-day plan by SYNCHRO

ACKNOWLEDGEMENTS

- Data source from Suzhou Industrial Park, China
- NVIDIA Corp for GPU donation (4th author)
- NAI-IT Fund RCQD16-01 (5th author)