Enhanced Multi-Agent Multi-Objective Reinforcement Learning for Urban Traffic Light Control

Mohamed A. Khamis and Walid Gomaa

Department of Computer Science and Engineering

Egypt-Japan University of Science and Technology (E-JUST)

Alexandria, Egypt

Email: {mohamed.khamis, walid.gomaa}@ejust.edu.eg

Traffic Hazards

- Social Stress & Accidents,
- Congestion & Delays,
- CO_2 Emissions.



US Statistics in 2010:

- Congestion (based on wasted time and fuel) cost about $115 billion in 439 urban areas.
- 32,885 people died in motor vehicle traffic accidents.

Our Focus: Urban Traffic Light Control

Our Contributions in:

- Traffic modeling
  - Traffic demand and acceleration/deceleration
- Traffic control
  - Traffic lights configurations
- Traffic simulation
  - For experimentation of model and control

Motivation:

- Safe life,
- Save time,
- High flow,
- Clean environment,
- Adaptive to traffic dynamics,
- Applied on large scale networks.

Traffic Management Challenges

Methods include:

- Traffic Lights in Urban areas
- Ramp metering in High ways
- Traffic-dependent route guidance



Challenges:

- 70% of world population will live in cities by 2050


- Non-linear traffic dynamics
- Construction of new infrastructure is expensive!

Extending the GLD Traffic Simulator

Features need enhancements:

- Discrete time/discrete space simulator
- Oversimplifications in modeling the driving behavior
- Some simplifications in computing the statistics

Our extensions to GLD:

- Varying distributions of traffic demand (modeling non-stationarity)
- Applying the Intelligent Driver Model (M. Treiber *et al.*, 2000)
  - Acceleration/deceleration model
  - Continuous time/continuous space model

$$\frac{\mathrm{d}v}{\mathrm{d}t} = a\left[1 - \left(\frac{v}{v_0}\right)^{\delta} - \left(\frac{s^*}{s}\right)^2\right],$$

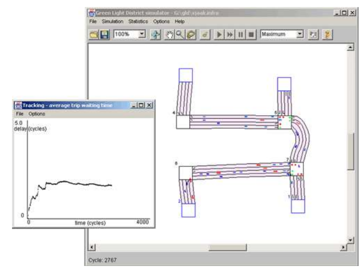
$$s^* = s_0 + min\left[0, \left(vT + \frac{v\Delta v}{2\sqrt{ab}}\right)\right]$$

- Synchronization between **three** timers:
  - Model actual time,
  - Controller time,
  - Simulation time

$$speed_{new} = speed_{old} + acceleration_{IDM} * \delta t,$$

$$position_{new} = position_{old} - speed_{new} * \delta t.$$

- Open source, developed by Wiering *et al.* in early 2000's

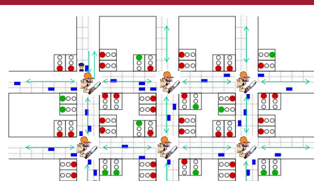


- Develop and experiment traffic light controllers
- Various Performance Indices

Reinforcement Learning in Traffic Light Control

Traffic Light Control:

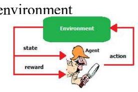
- Finding the optimal traffic light configuration
  - Red/Green consistent configurations
- **Multi-Agent System (MAS)** modeling:
  - Vehicle: passive agent; Junction: active agent
- Online learning using **Reinforcement Learning**



Reinforcement Learning:

- Markov Decision Process (MDP): Suitable for **Sequential decision making tasks**;
- Learning from **trial-and-error interaction** between the agent & surrounding environment
- Based on: Control theory - Dynamic Programming - Bellman Equation

$$Q(s,a) = \sum_{s'} \Pr(s,a,s')(R(s,a,s') + \gamma V(s'))$$

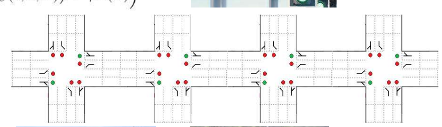


Multi-Objective Q-Function

$$Q(s,a) = \sum_{s'} \Pr(s,a,s')\Big(\big(W_{FR}(s,a,s') * R_{FR}(s,a,s') + R_{ATWT}(s,a,s') + R_{ATT}(s,a,s') + R_{AJWT}(s,a,s') + R_S(s,a,s') + R_{GW}(s,a,s') + R_{FC}(s,a,s')\big) + \gamma V(s')\Big)$$

Objectives:

- Flow Rate (FR),
- Average Trip Waiting Time (ATWT),
- Average Trip Time (ATT),
- Average Junction Waiting Time, (AJWT),
- Safety (S) ,
- Green Wave (GW),
- Fuel Consumption (FC),



Rewards :

- Function in road types,
- Residential area /Main street.


Handling Non-Stationarity

- Using Bayesian probability interpretation rather than frequentist approach
- Current estimation becomes prior in the next time step
- More stable & adaptable to the changing conditions

Starting with Bay's rule: $\Pr(P_t|x_i) = \dfrac{\Pr(x_i|P_t)\Pr(P_t)}{\Pr(x_i)}$; i = [1, …, t+1],

Ending with all weighted experiences:

$$P_t = \frac{2}{t(t+1)} \sum_{i=1}^{t} \sum_{j=1}^{i} x_j.$$

Exploration and Cooperation

Decayed Boltzmann Exploration:

$$\epsilon = e^{-t/k_t}$$

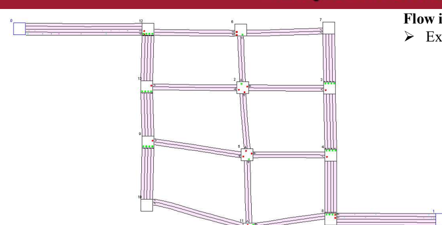
t: the current simulation time step,
k_t Boltzmann temperature parameter ➔ used to **increase the exploration effect initially**
k_t decreases gradually ➔ where all traffic light configurations selected according to their **cumulative gain**

Traffic Lights Cooperation:

$$Q_{new} = Q_{own} + \alpha_t[Q_{transferred} - Q_{own}]$$

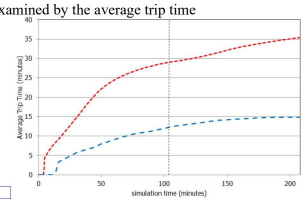
Agents **transfer knowledge** from the **external layer** of the traffic network to the **internal layer**.
$\alpha_t \in [0,1]$ is the **agent's learning rate** ➔ decrease as the **temperature parameter** falls down.

Experimentations



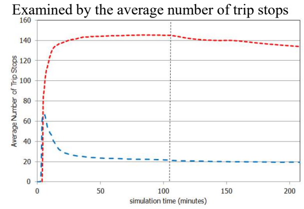
Flow in arteries:

- Examined by the average trip time



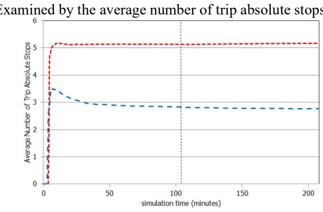
Fuel Consumption:

- Examined by the average number of trip stops



Green Wave:

- Examined by the average number of trip absolute stops



- TC-1 is the **single-objective controller** (M. Wiering 2000) that is based on the **frequentist probability interpretation**.

The 11th International Conference on Machine Learning and Applications (ICMLA 2012)
December 12-15, 2012, Marriott Hotel, Boca Raton, Florida, USA