Traffic Signal Control using Reinforcement Learning

Chitu Irina[ǀ], Iordache Bogdan[ǁ], Manghiuc Teodor[ǀ], Marchitan Teodor[ǁ], Sotir Anca[ǀ]
ǀ - group 507 Artificial Intelligence, ǁ - group 512 Natural Language Processing

Overview

- Introduction
- Survey of Related Work
- Experiments

Introduction

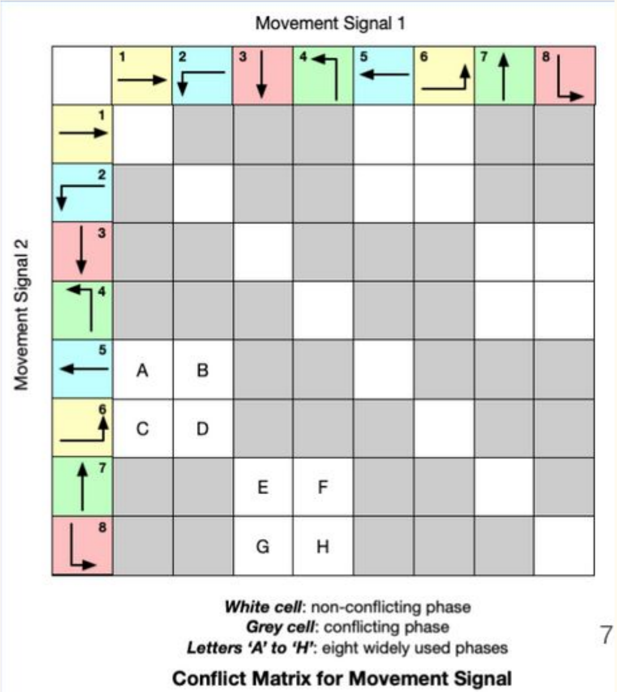
Introduction

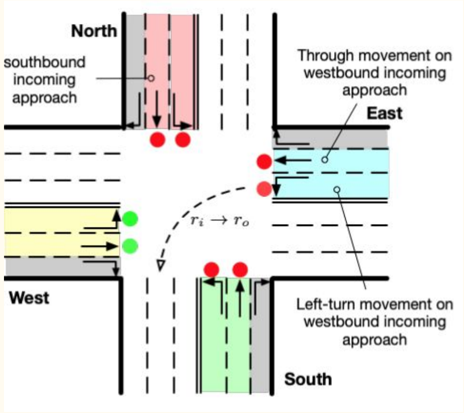
- Traffic congestion affects people’s daily lives
  - Wasted time on commute due to bad traffic conditions
  - Traffic congestion also contributes to fuel waste
  - It also increases harmful emissions (greenhouse gases and other particles)
- Provide designs for better traffic signal control systems
- RL is a promising solution

Terminology

- Movement signal
- Phase (combination of movement signals)
- Interval
- Phase sequence
- Signal plan (sequence of phases and their starting time)

Terminology





Various combinations of movement signals and commonly considered phases

Survey of Related Work

General Overview

Non-RL approaches based on heuristics:

- SOTL (Self-Organizing Traffic Lights)
- MaxPressure

General Overview

The **actions** of the agent correspond to changing the phase of one or more intersections.

Common features for representing the **state** of an intersection:

- queue length
- waiting time
- volume
- delay
- speed
- phase duration
- positions of vehicles
- phase

General Overview

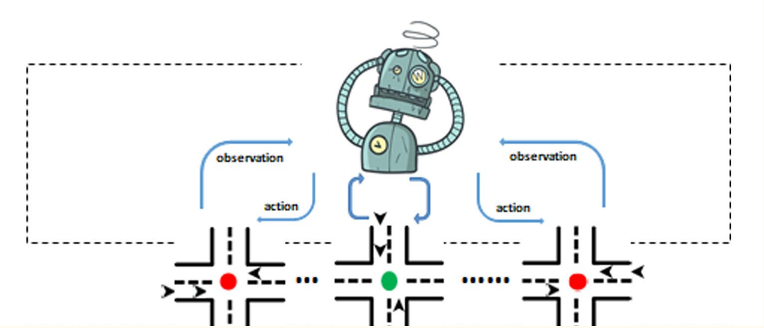
Commonly used parameters for the definition of the **reward** are:

- queue length
- waiting time
- change of delay
- speed
- number of stops
- throughput
- freq. of signal change
- pressure

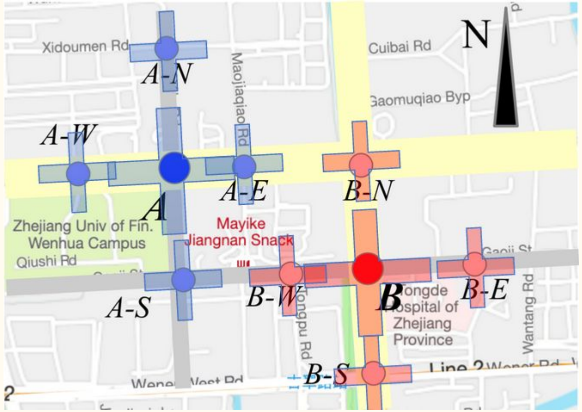
General Overview

Control in systems with multiple intersections:

1. Centralized control
2. Individual RL without coordination
3. Individual RL with coordination



centralized control

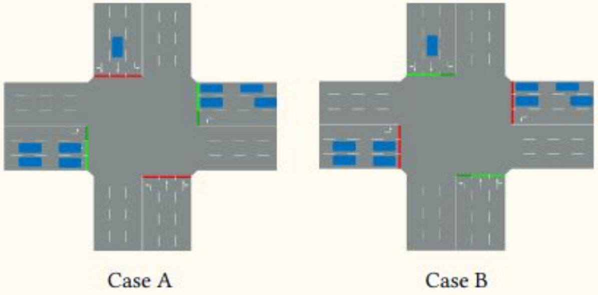


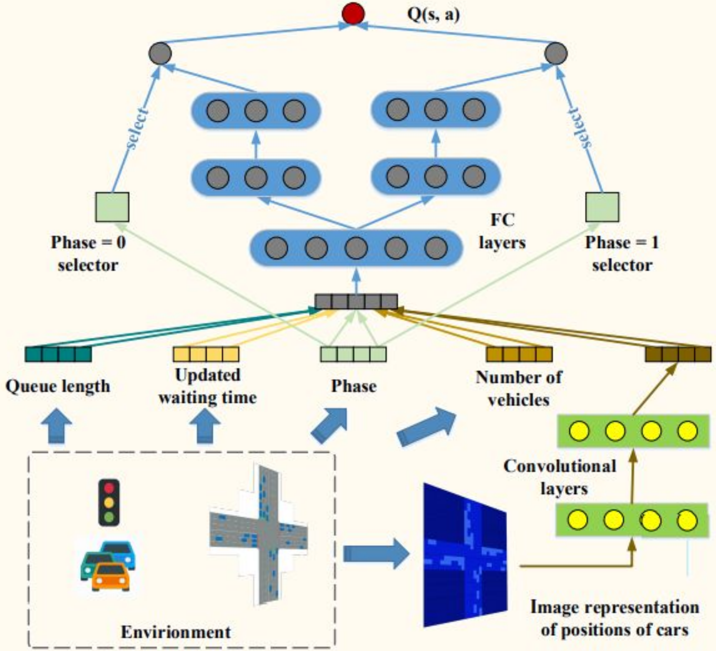
individual RL w/ coordination

IntelliLight: A Reinforcement Learning Approach for Intelligent Traffic Light Control

Key contributions:

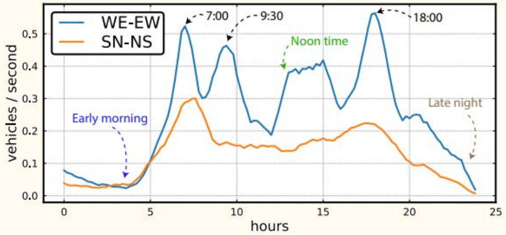
- Real-world traffic data (24 intersections, China - Jinan, 31 days)
- Phase-Gated Deep Q-Network with memory palace



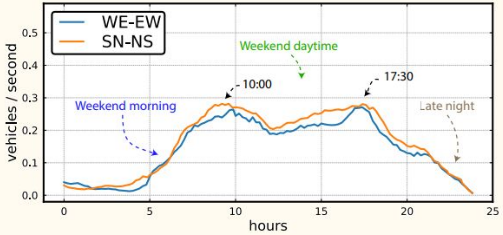


Interpret policies learned from real data:

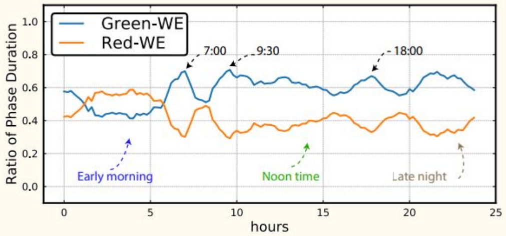
- peak hour vs. non–peak hour
- weekday vs. weekend
- major arterial vs. minor arterial



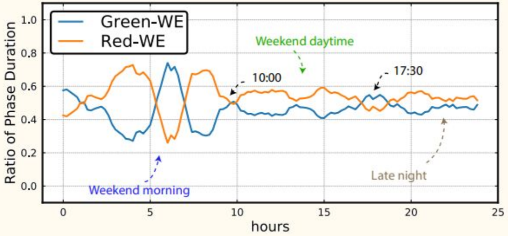
(a) Average arrival rate of August 1st (Monday)



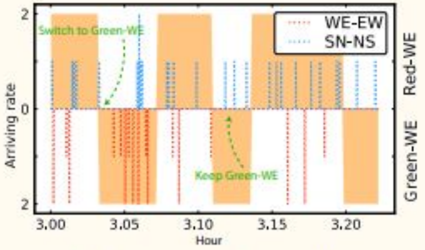
(b) Average arrival rate of August 7th (Sunday)



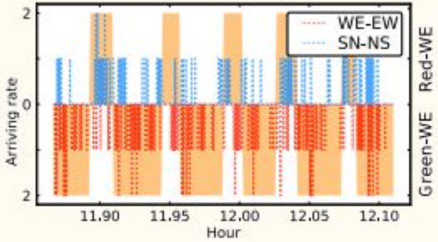
(c) Phase time ratio from learned policy on August 1st (Monday)



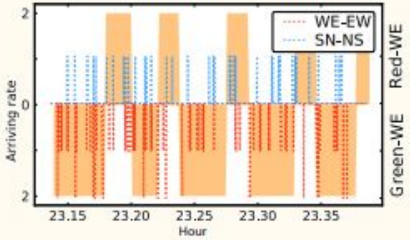
(d) Phase time ratio from learned policy on August 7th (Sunday)



(a) Early morning when traffic on *WE-EW* is less than *SN-NS*



(b) Noon when traffic on *WE-EW* is more than *SN-NS*

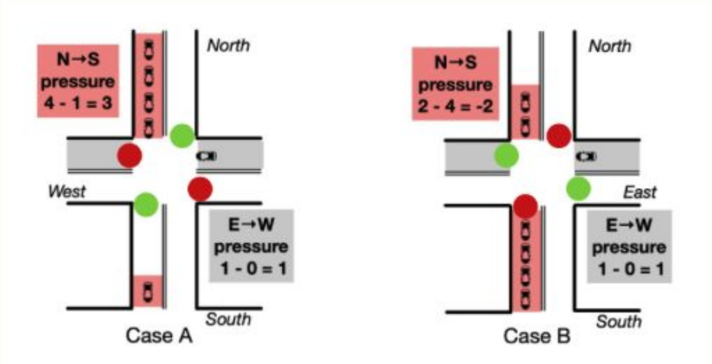


(c) Late night when traffic on *WE-EW* is more than *SN-NS*

PressLight: Learning Max Pressure Control to Coordinate Traffic Signal in Arterial Network

- MaxPressure (MP) is proven to **maximize the system throughput**
- **State** includes the current phase, the number of vehicles on each outgoing lane and the number of vehicles on each segment of every incoming lane
- **Reward** is defined as minus the pressure of the intersection

Deep Q-Network is used as function approximator to estimate the Q-value function.



PressLight

	Synthetic traffic				Real-world traffic					
	LightFlat	LightPeak	HeavyFlat	HeavyPeak	Qingdao Rd., Jinan	Beaver Ave., State College	8th Ave., NYC	9th Ave., NYC	10th Ave., NYC	11th Ave., NYC
FixedTime	93.29	109.50	325.48	246.25	317.40	336.29	432.60	469.54	347.05	368.84
GreenWave	98.39	124.09	263.36	286.85	370.30	332.06	451.98	502.30	317.02	314.08
MaxPressure	74.30	82.37	262.26	225.60	567.06	222.90	412.58	370.61	392.77	224.54
GRL	123.02	115.85	525.64	757.73	238.19	455.42	704.98	669.69	676.19	548.34
LIT	65.07	66.77	233.17	258.33	58.18	338.52	471.30	726.04	309.95	340.40
PressLight	**59.96**	**61.34**	**160.48**	**184.51**	**54.87**	**92.00**	**223.36**	**149.01**	**161.21**	**140.82**

DemoLight: Learning Traffic Signal Control from Demonstrations

- the first work that tries to integrate demonstrations into RL for traffic signal control
- exploits demonstrations collected from SOTL to accelerate an actor-critic RL algorithm
- actor and critic are trained with demonstrations in order to provide expert-like initialization
- ablation studies for assessing the role of demonstrations in overall performance

DemoLight

Training using demonstrations:

- **Actor:** the sampled actions are drawn from the categorical distribution provided by the policy

$a_{\text{soft}} = \text{softmax}((g+\pi)/\tau)$, where g corresponds to random re-parametrization.

$L_{pre}(\theta_\pi) = \text{Cross-Entropy}(a_{\text{soft}}, a_D)$, where a_D is the action of the demo.

DemoLight

Training using demonstrations:

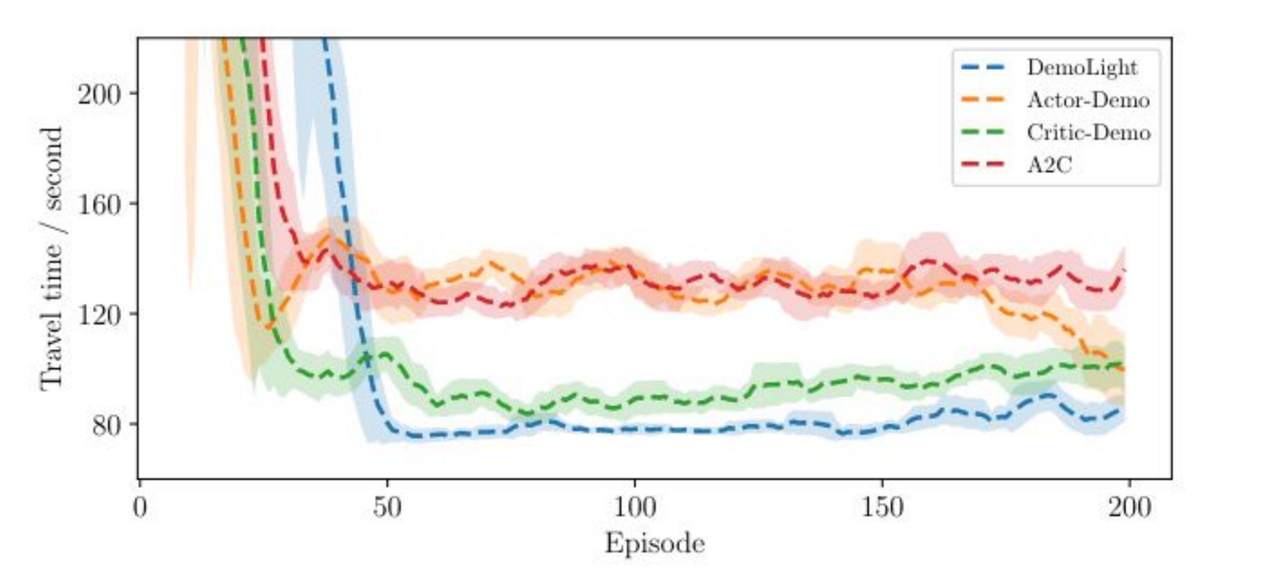
- **Critic:** cloning through four losses
  - 1-step TD
  - n-step TD
  - large margin classification loss
  - L2 regularization

$$L_{TD}(\theta_Q) = \frac{1}{2}(R(s,a) + \gamma Q(s',a') - Q(s,a|\theta_Q))^2$$

$$L_{\text{margin}}(\theta_Q) = \max_a [Q(s,a) + l(a_D,a)] - Q(s,a_D)$$

where $l(a_D,a) = 0.8$ if $a_D \neq a$ else 0.

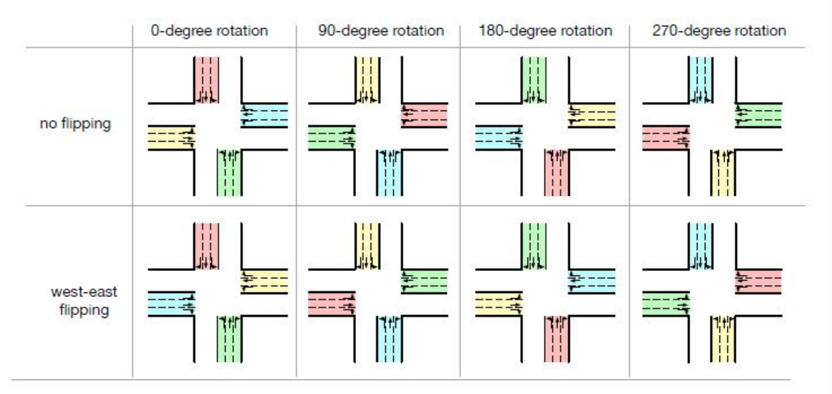
DemoLight



Ablation study for learning traffic signal control through demonstrations

FRAP: Learning Phase Competition for Traffic Signal Control

- The authors introduced FRAP in 2019
  - Especially designed for traffic signal control
  - The aim is to capture **the competition between phases**
- The most important advantage: **invariance to symmetrical traffic conditions**
- **Major contributions**
  - Propose a **novel design**, FRAP
  - Demonstrate its **faster convergence**
  - Demonstrate its **generalizability**
    - Different intersection structure
    - Different traffic flows
    - Multi-intersection environments



FRAP

- A single RL agent manages an individual intersection (standard 4-way)
- **States** = the **number of vehicles** for each traffic movement & the **active phase**
- **Reward** = based on the **average queue length** for each traffic movement

Design

- The FRAP network is made up of **embedding** and **convolutional** layers
  - Represent **the demand for each phase** and then **the competition between pairs of phases**
  - Intuitively: when t**wo phases are in conflict**, the **higher demanding one should be chosen**
- DQN method
  - Predict the Q-value for each phase and the agent will choose the highest one

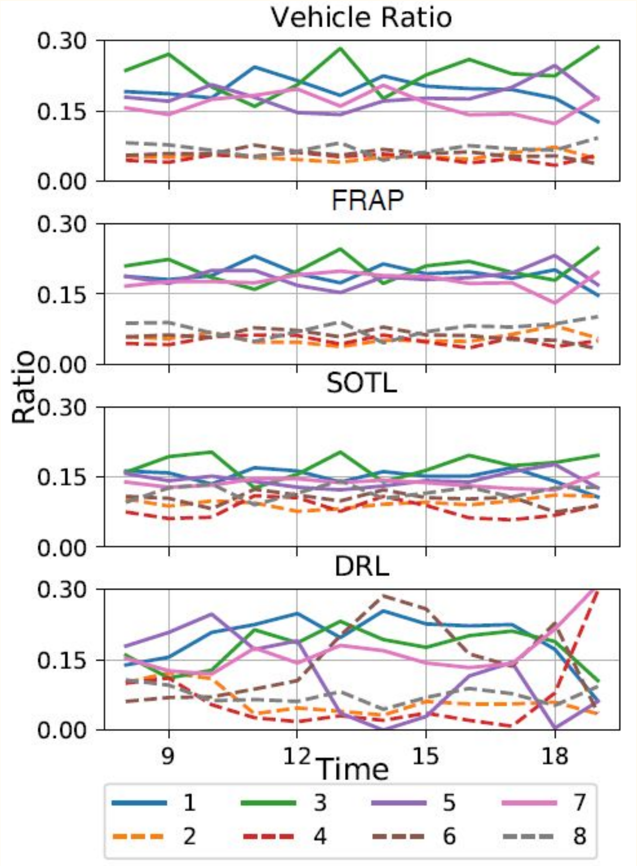
FRAP

Experiments on real world datasets, using CityFlow.

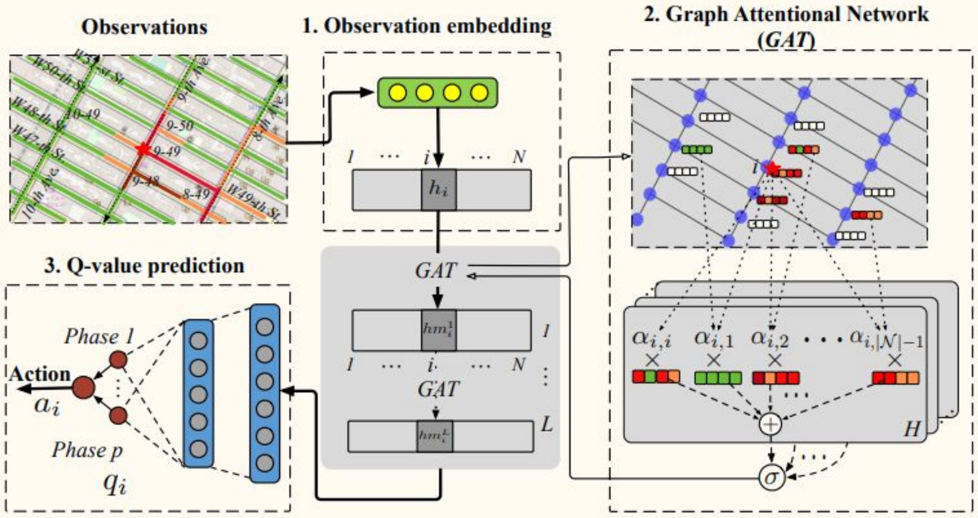
- 2 private sets (Jinan, Hangzhou) and 1 public set (Atlanta)
- All show **superior results compared with previous models**

The authors' experiments have led to the following conclusions regarding FRAP's performance:

- Good results for **multi-intersection envs** (even **without explicit coordination**)
- Saves time - it does **not need retraining when traffic patterns drastically change**
- Easily adapts to **different intersection structures** (from 4-way to 3 and 5-way)



CoLight: Learning Network-level Cooperation for Traffic Signal Control

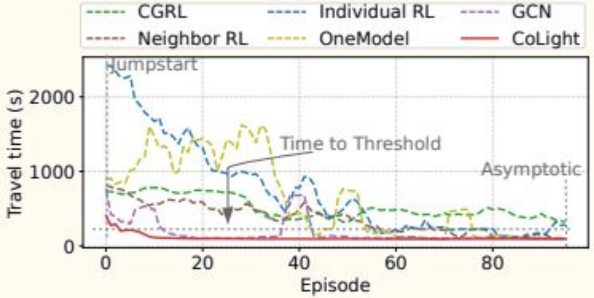


Model framework

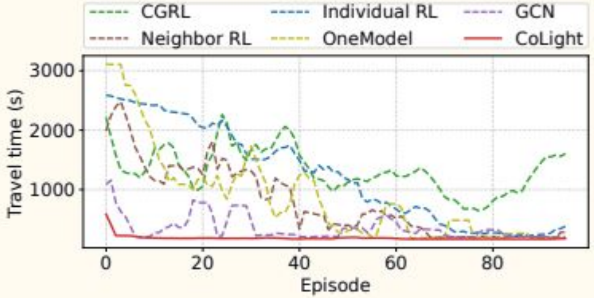
Key contributions:

- Use Graph Attentional Network to learn the dynamics of the traffic trends
- Index-free model learning with parameter sharing
- Scalable to hundreds of intersections

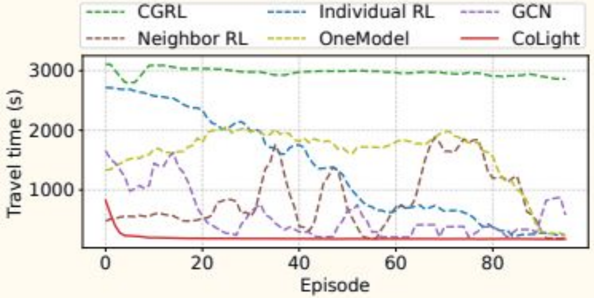
Performance Comparison



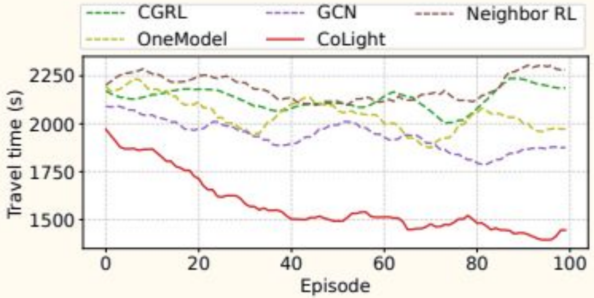
(a) $Grid_{3\times3}$



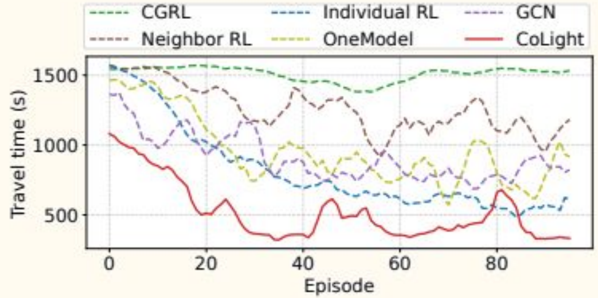
(b) $Grid_{6\times6}$-Uni



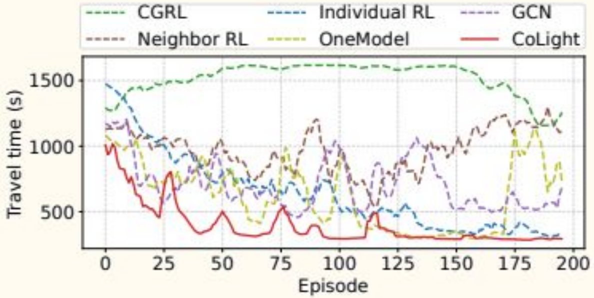
(c) $Grid_{6\times6}$-Bi



(d) $D_{NewYork}$



(e) $D_{Hangzhou}$



(f) D_{Jinan}

MetaLight: Value-based Meta-reinforcement Learning for Traffic Signal Control

Main advantages:

- Easy to adapt in new situations
- No need for a large number of samples

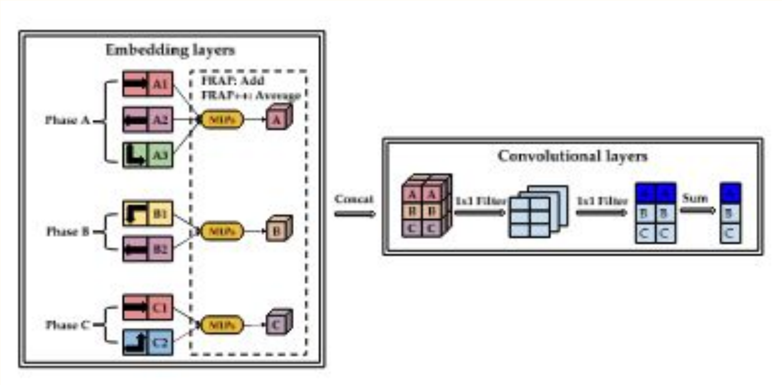
MetaLight

Contributions:

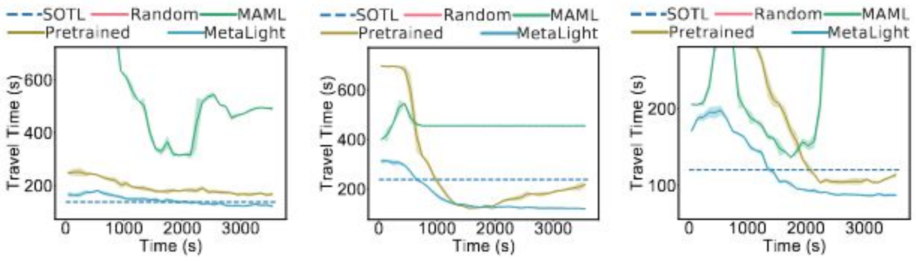
- **FRAP++:** Removed the influence of difference in the lane number under each phase
- **Improved MAML:** A generalization of the parameters is learnt in order to be quickly adapted

How to transfer knowledge to new intersections?

- Given a new intersection, the learnt parameters are used and then quickly optimized



MetaLight



(a) Jinan-6a (b) Atlanta-4d (c) Los Angeles-4c

City	Homogeneous JN	Homogeneous AT	Homogeneous LA	Heterogeneous JN	Heterogeneous AT	Heterogeneous LA
Random	451.88	379.16	262.23	363.59	602.60	684.15
Pretrained	128.20	186.86	104.59	156.04	351.39	331.75
MAML	173.13	301.29	135.11	335.81	618.84	393.58
MetaLight	**95.01**	**161.37**	**77.23**	**137.02**	**310.39**	**308.71**
Improvement	25.89%	13.64%	26.16%	10.17%	11.67%	6.94%

MPLight: Toward A Thousand Lights: Decentralized Deep Reinforcement Learning for Large-Scale Traffic Signal Control

- **No other model** was tested on networks containing **more than 1K traffic lights.**

- The authors propose **3 key issues** that must be addressed by an **effective model**
  - **Scalability** (cannot be satisfied by centralized methods)
  - **Coordination** (is not easily satisfied by decentralized methods)
  - **Data feasibility** (some models use data which cannot realistically be obtained in real scenarios)
- They propose MPLight, a model which **combines FRAP with PressLight**

MPLight

- Decentralized approach → satisfy **scalability**
- Parameter sharing → satisfy **coordination**
    - FRAP as base model (also using deep Q-learning)
        - Faster convergence
        - Agents essentially follow the same logic even when managing different intersection types
- State and reward based on PressLight → satisfy **coordination**
    - Balances the distribution of vehicles and maximizes the throughput
- The pressure of an intersection → satisfy **data feasibility**
    - It is derived from simple features as queue length

MPLight

Experiments made using CityFlow, on both synthetic and real-world data.

1. Synthetic experiments
   - 4x4 network, 4 configurations (different traffic patterns and arrival rates)
2. Real-world experiments
   - The road network of Manhattan (obtained from OpenStreetMap)
   - Contains circa **2.5K street lights**

MPLight outperformed all previous models in both synthetic and real scenarios.

Other interesting experiments

- The positive impact of using **pressure** for three different model architectures
- The positive impact of **parameter sharing** (less episodes needed to converge)

Comparing Methods

Not a trivial task, due to some limitations:

- differences in synthetic data generation
- single intersection performance vs. multi-intersection networks
- using various subsets of a roadnet (just a main road, or the whole network)
- different simulation environments (SUMO vs. CityFlow)

Comparing Methods

Model	Jinan							Hangzhou					
	1	2	3	4	5	6	7	1	2	3	4	5	6
Fixedtime	118.82	250.00	233.83	297.23	101.06	104.00	146.66	271.16	192.32	258.93	207.73	259.88	237.77
Formula	107.92	195.89	245.94	159.11	76.16	100.56	130.72	218.68	203.17	227.85	155.09	218.66	230.49
SOTL	97.80	149.29	172.99	64.67	76.53	92.14	109.35	179.90	134.92	172.33	119.70	188.40	171.77
DRL	98.90	235.78	182.31	73.79	66.40	76.88	119.22	146.50	118.90	218.41	80.13	120.88	147.80
IntelliLight	88.74	195.71	100.39	73.24	61.26	76.96	112.36	97.87	129.02	186.04	81.48	177.30	130.40
A2C	135.81	166.97	226.82	43.28	67.05	148.69	236.17	110.91	98.56	187.41	86.56	116.70	128.88
FRAP	**66.40**	**88.40**	**84.32**	**33.83**	**54.43**	**61.72**	**72.31**	**80.24**	**79.43**	**110.33**	**67.87**	**92.90**	**88.28**
Improvement	25.17%	40.79%	16.01%	47.69%	11.15%	19.72%	33.87%	18.01%	33.20%	35.98%	15.30%	23.15%	32.30%

Model	Jinan	Hangzhou	Atlanta
Fixedtime	880.18	823.13	493.49
Formula	385.46	629.77	831.34
SOTL	1422.35	1315.98	721.15
DRL	1047.52	1683.05	769.46
IntelliLight	358.83	634.73	306.07
A2C	316.61	591.14	244.10
FRAP	**293.35**	**528.44**	**124.42**

FRAP performs better than **IntelliLight** on both single-intersection and multi-intersection environments

Comparing Methods

Model	$Grid_{6\times6}$-Uni	$Grid_{6\times6}$-Bi	$D_{NewYork}$	$D_{Hangzhou}$	D_{Jinan}
Fixedtime [15]	209.68	209.68	1950.27	728.79	869.85
MaxPressure [24]	186.07	194.96	1633.41	422.15	361.33
CGRL [23]	1532.75	2884.23	2187.12	1582.26	1210.70
Individual RL [30]	314.82	261.60	-*	345.00	325.56
OneModel [5]	181.81	242.63	1973.11	394.56	728.63
Neighbor RL [1]	240.68	248.11	2280.92	1053.45	1168.32
GCN [18]	205.40	272.14	1876.37	768.43	625.66
CoLight-node	178.42	176.71	1493.37	331.50	340.70
CoLight	**173.79**	**170.11**	**1459.28**	**297.26**	**291.14**

*No result as *Individual RL* can not scale up to 196 intersections in New York's road network.

CoLight was also compared with **IntelliLight**, and it achieves better average travel time on multi-intersection settings

Comparing Methods

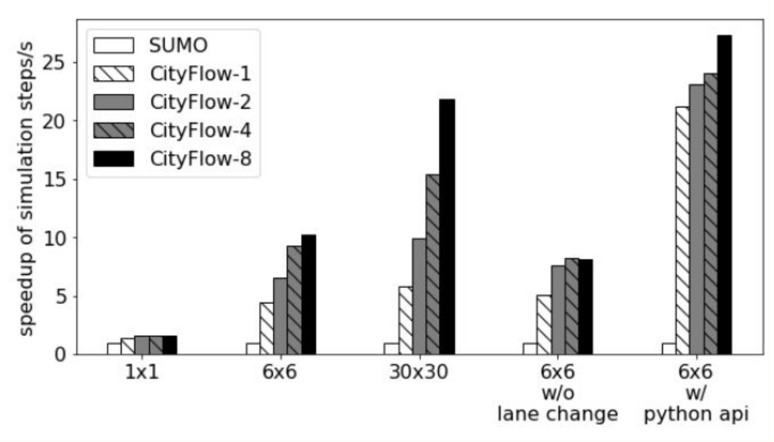
Model	Travel Time				Throughput			
	Config 1	Config 2	Config 3	Config 4	Config 1	Config 2	Config 3	Config 4
FixedTime	573.13	564.02	536.04	563.06	3555	3477	3898	3556
MaxPressure	361.17	402.72	360.05	406.45	4702	4324	4814	4386
GRL	735.38	758.58	771.05	721.37	3122	2792	2962	2991
GCN	516.65	523.79	646.24	585.91	4275	4151	3660	3695
NeighborRL	690.87	687.27	781.24	791.44	3504	3255	2863	2537
PressLight	354.94	353.46	348.21	398.85	4887	4742	5129	5009
FRAP	340.44	298.55	361.36	598.52	5097	5113	5483	4475
MPLight	**309.33**	**262.50**	**281.34**	**353.13**	**5219**	**5213**	**5652**	**5060**

MPLight achieves better performance than **PressLight** and **FRAP** on multi-intersection environments using synthetic data.

Experiments

CityFlow

- simulation environment designed for large-scale traffic signal control benchmarking
- leverages multithreaded computations in order to efficiently simulate a large amount of traffic
- faster than previous approaches (SUMO)



CityFlow

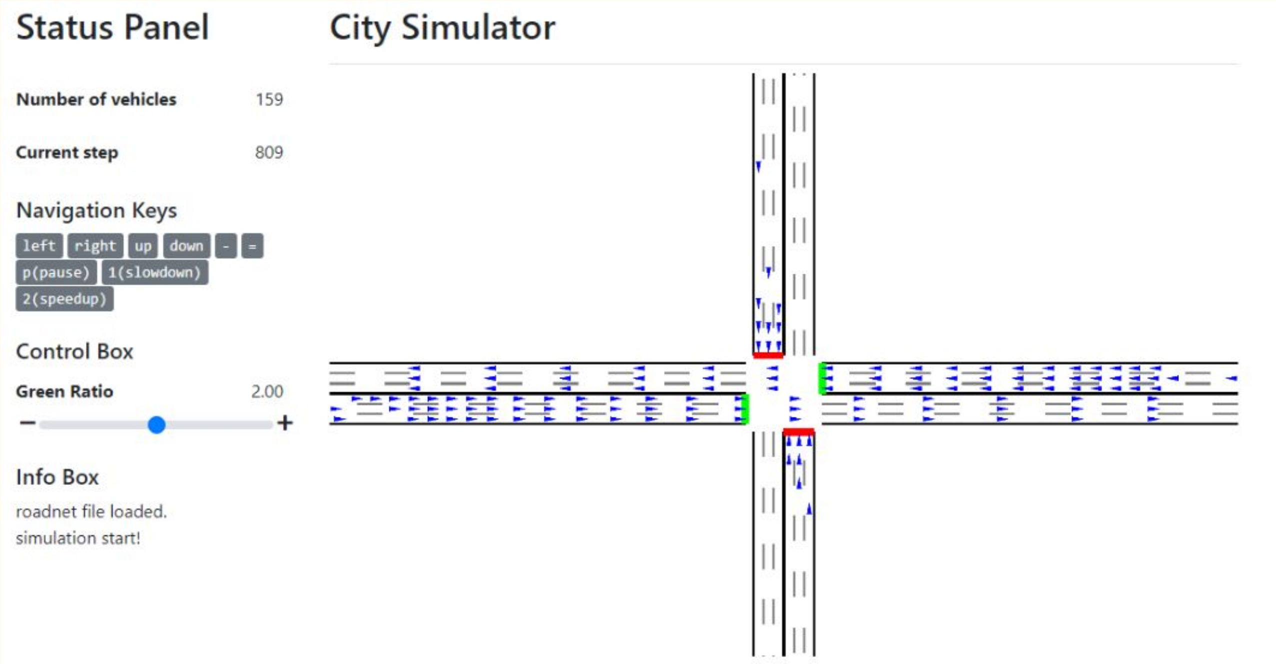

```
{
    "interval": 1.0,
    "seed": 0,
    "dir": "data/",
    "roadnetFile": "roadnet/testcase_roadnet_3x3.json",
    "flowFile": "flow/testcase_flow_3x3.json",
    "rlTrafficLight": false,
    "saveReplay": true,
    "roadnetLogFile": "frontend/web/testcase_roadnet_3x3.json",
    "replayLogFile": "frontend/web/testcase_replay_3x3.txt"
}
```


- **roadnet description**
  - roads, and their corresponding lanes
  - intersections, with corresponding road-links and lane-links
  - traffic light phases description for each intersection
- **flow configuration** (vehicle-related information)
  - dimensions
  - maximum speed
  - maximum acceleration
  - intersection related speed

We have wrapped the simulation environment inside an **OpenAI Gym** environment.

1x1 two-lane intersection with **one hour** of traffic recorded in **Hangzhou, China**

CityFlow



CityFlow WebGL Frontend

Problem Representation

For the **state** we consider:

- the number of waiting vehicles for each incoming lane
- the total number of vehicles on each lane, both incoming and outgoing (inspired from the definition of pressure)

The **actions** performed by the agent can change the current phase of the intersection, at each time-step.

The **reward** is computed as the total number of waiting vehicles on incoming lanes.

*We also tried to penalize phase changes that are performed **too often**, but with no sensible improvements.*

Stable Baselines 3

Method	Policy	Observations
Q-learning	off-policy	discrete observation space
SARSA	on-policy	discrete observation space
Deep Q-network (DQN)	off-policy	
Deep Deterministic Policy Gradient (DDPG)	off-policy	only continuous action space
Advantage Actor Critic (A2C)	on-policy	
Proximal Policy Optimization (PPO)	on-policy	

Considered methods for RL experiments

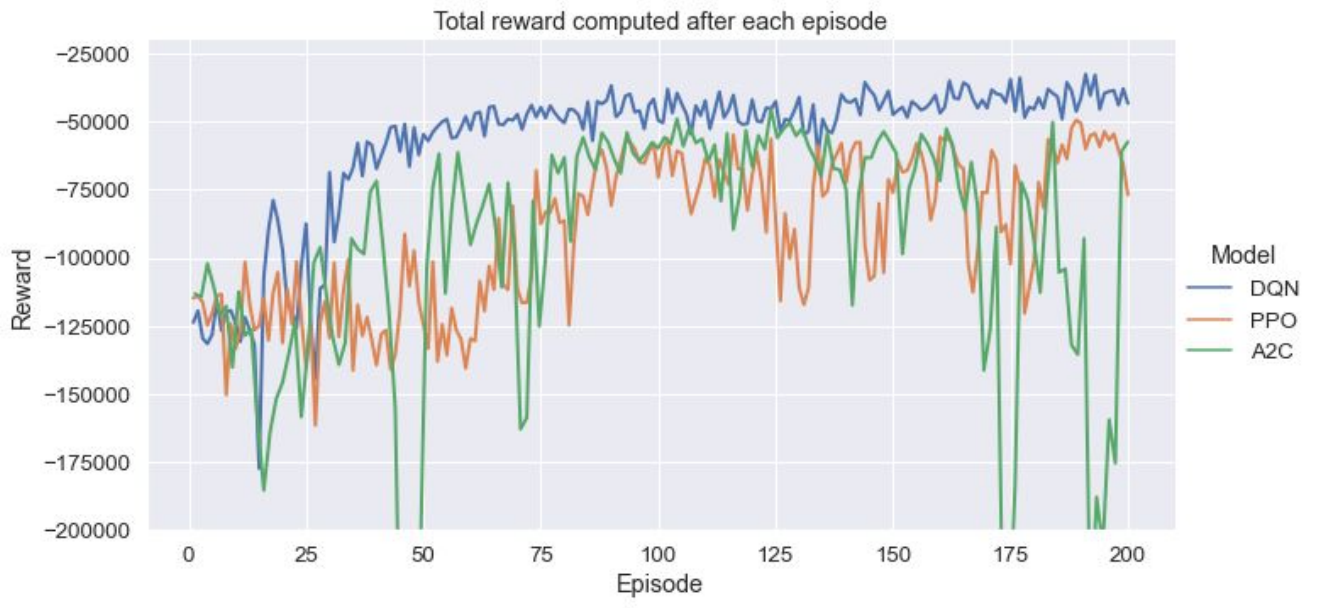
Results

Through hyperparameter tuning, we have decided on the following settings:

- **DQN**: batch size = 128, learning rate of 0.0005
- **A2C**: num. steps per update = 15, learning rate of 0.0005
- **PPO**: batch size = 256, learning rate of 0.0007

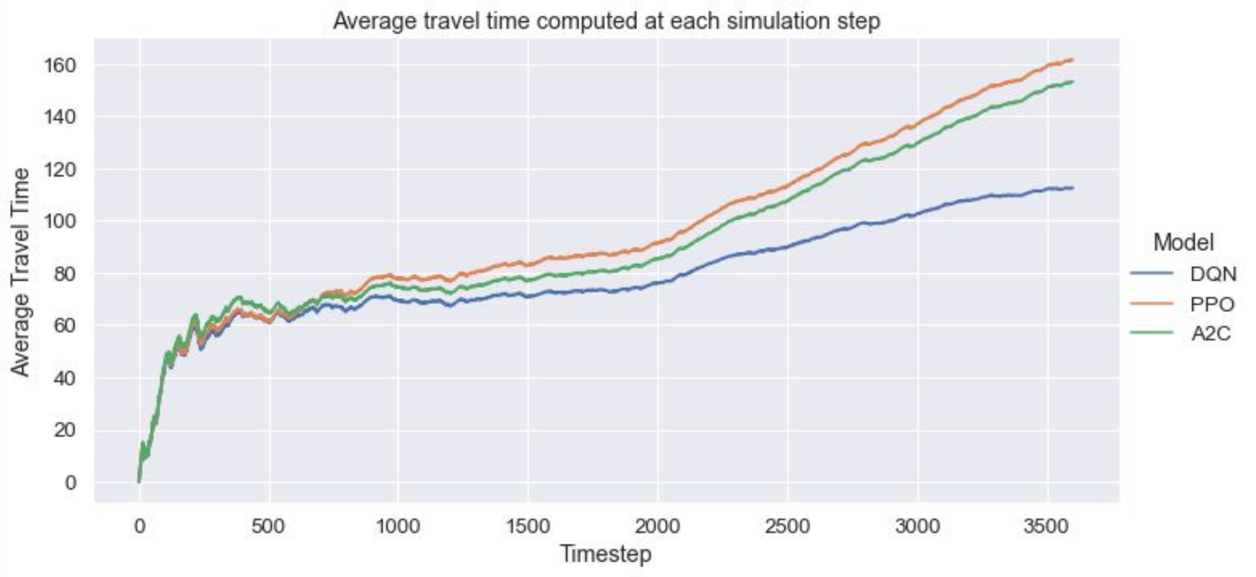
The training is done for **200** episodes, each episode consisting of **3.600** steps.

Results



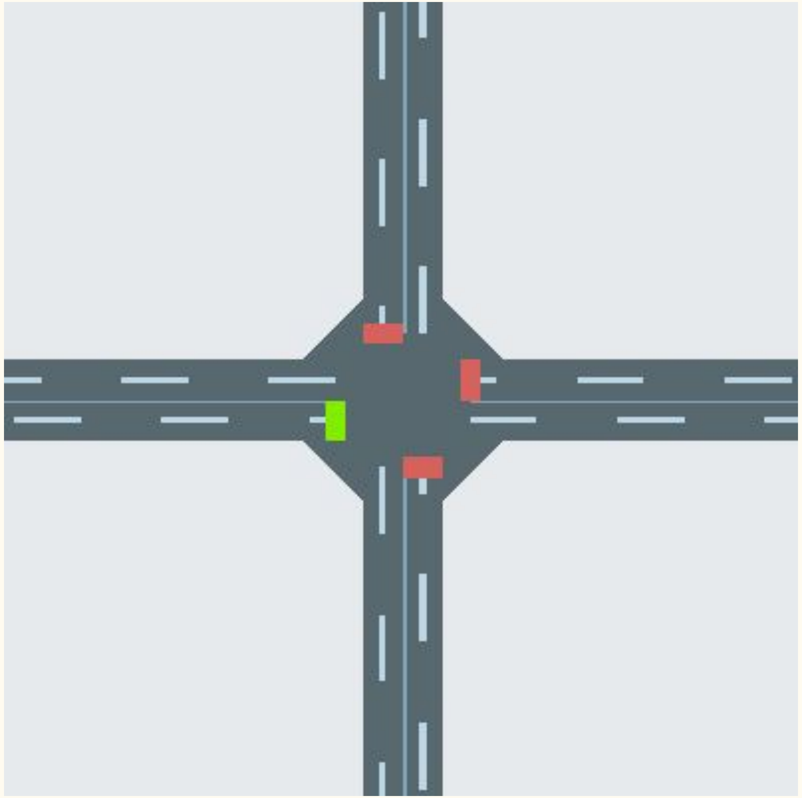
Evolution of total reward computed for each training episode

Results



Average vehicle waiting time at each step, using best performing models

Results



Conclusions

- a brief overview of most recent RL methods for traffic signal control
- how various formulations of state and reward can improve traffic
- a comparison of popular RL algorithms on a minimal representation of the state

Thank you!