IPDALight: Intensity- and phase duration-aware traffic signal control based on Reinforcement Learning

Journal of Systems Architecture, 2022

Wupan Zhao, Yutong Ye, Jiepin Ding, Ting Wang, Tongquan Wei, Mingsong Chen

발표자: 여지호

Introduction

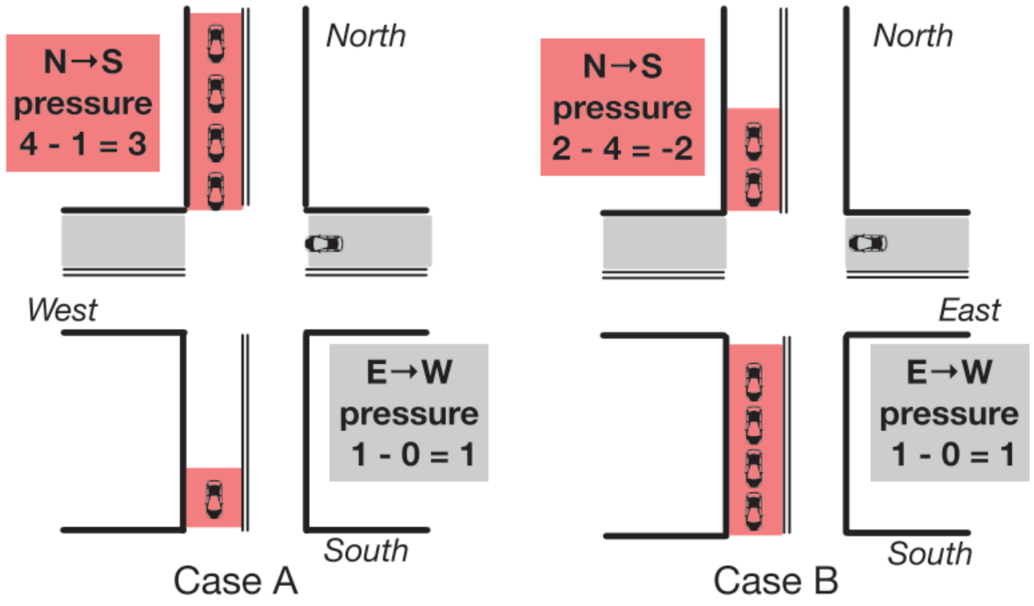
Previous RL-based traffic signal control

Table 1: Representative deep RL-based traffic signal control methods.

Citation	Method	Cooperation	Simulator	Road net (# signals)	Traffic flow*	
[2]	Value-based	With communication	Matlab	Synthetic (5)	2,4	
[5]	Policy-based	Without communication	Aimsun	Real (50)	5	
[10]	Policy-based	Without communication	Aimsun	Real (43)	5	
[11]	Value-based	Without communication	CityFlow	Real (2510)	5	MPLight
[12]	Policy-based	Joint action	SUMO	Real (30)	4	
[22]	Value-based	-	SUMO	Synthetic (1)	2	
[30]	Value-based	-	Paramics	Synthetic (1)	4	
[39]	Value-based	Without communication	SUMO	Synthetic (9)	2	
[42]	Both studied	-	SUMO	Synthetic (1)	1	
[44]	Value-based	With communication	SUMO	Synthetic (6)	2	
[48]	Value-based	Without communication	AIM	Synthetic (4)	1	
[49]	Both studied	Single global	GLD	Sythetic (5)	3	
[51]	Policy-based	-	SUMO	Real (1)	5	
[58]	Value-based	Joint action	SUMO	Synthetic (4)	2	
[60]	Value-based	With communication	SUMO	Real (4)	5	
[65]	Value-based	-	SUMO	Synthetic (1)	1,3,4,5	
[62]	Value-based	Without communication	CityFlow	Real (16)	2,5	PressLight
[63]	Value-based	With communication	CityFlow	Real (196)	2,5	
[74]	Value-based	Joint action	SUMO	Synthetic (36)	1,2,3,4	
[78]	Value-based	Without communication	CityFlow	Real (16)	3,5	
[77]	Value-based	Without communication	CityFlow	Real (5)	4,5	

Source: Zheng et al., 2021

The concept of pressure



- **Pressure**
  : # of vehicles on incoming lanes - # vehicles on outgoing lanes
- **Design an RL agent, PressLight**
  : using the pressure-based reward for long-term optimization.

Limitations

- MP(Max-Pressure) only considers **the number of vehicles on the lanes**, while the **vehicle speed** and **position information** is neglected
- **No coordination** among intersections
- There is no selection of variable **phase duration(신호 길이)**
  : 신호가 빠르게 계속 바뀌는 문제
  : 현실 적용성이 떨어짐
  : 대부분의 교차로는 횡단보도도 존재하기 때문에 최소한의 보행자 신호가 확보가 되어야하지 않을까…

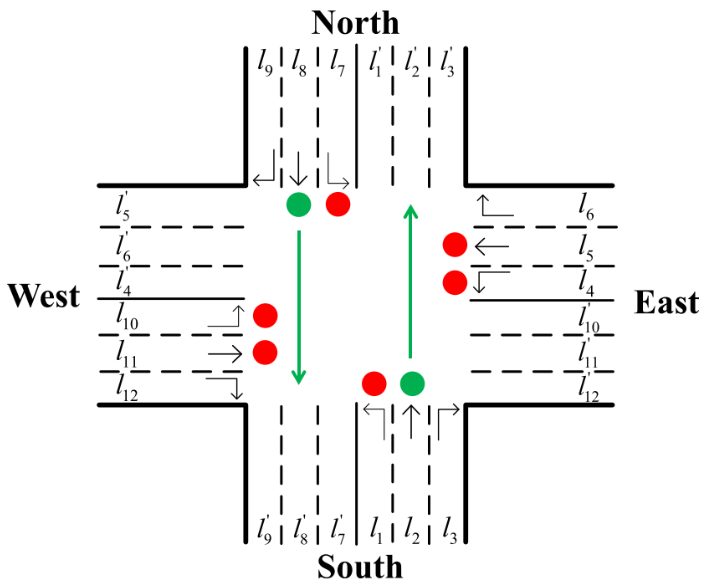
Propose IPDALight

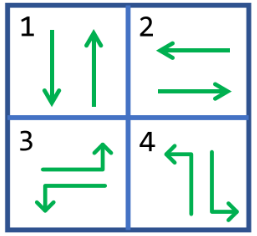
- **Propose concept of intensity**
  : considers detailed vehicle dynamics (i.e., speed and position)
  : coordination between neighboring intersections
- **Propose a heuristic algorithm**
  : support the fine-tuning of phase duration

IPDALight approach

Intersection modeling

- **Definition 1 (Incoming Lanes and Outgoing Lanes of an Intersection)**
  - An incoming lane of some intersection is a lane on which vehicles can enter the intersection.
  - An outgoing lane of some intersection is a lane on which vehicles can leave the intersection.





12 incoming and 12 outcoming lanes

Intersection modeling

- **Definition 2 (Movement, Signal and Phase)**
  - A movement refers to the behavior of vehicles crossing an intersection in a given direction from certain incoming lanes to their target outgoing lanes.
  - A signal is used to determine whether a movement is allowed at a certain time, where the green color implies a permitted movement, and the red color forbids any movement.
  - A phase is defined as a combination of signals that do not conflict, indicating the rights-of-way signaled to vehicles by traffic lights.

Intensity modeling

- **Unlike the pressure concept**, the definitions of intensity presented as follows consider more vehicle dynamics.
- **Definition 3 (Intensity of Vehicles)**

$$\mathcal{T}_{veh} = log\left(\frac{L-x}{L} \times \frac{\delta \times (v_{max} - v)}{v+1} + 1\right)$$

where x (unit: meters) indicates the distance between the vehicle and the intersection, L (unit: meters) denotes the lane length, v (unit: meters/second) represents the current vehicle speed, v_{max} (unit: meters/second) denotes the maximum allowed speed of the lane, and δ is a weight factor to adjust the influence of speed on intensity.

Intensity modeling

$$\mathcal{T}_{veh} = log\left(\frac{L-x}{L} \times \frac{\delta \times (v_{max} - v)}{v+1} + 1\right)$$

- assume that the intensity of vehicle increases when it approaches some intersection or the speed of vehicle decreases due to traffic congestion
- the vehicles crossing an intersection with higher speed will pose lower intensity on the intersection

Intensity modeling

- **Definition 4 (Intensity of Lanes)**
  : Intensity of a lane is the sum of intensity of all vehicles on this lane

$$\mathcal{T}_{lane} = \sum_{veh_i \in lane} \mathcal{T}_{veh_i}.$$

- **The intensity of a movement indicates the difference between the mean intensity of all the incoming lanes and the mean intensity of all the outgoing lanes.**

Intensity modeling

- **Definition 5 (Intensity of Movements)**

  : The intensity of a movement indicates the difference between **the mean intensity of all the incoming lanes and the mean intensity of all the outgoing lanes**.

$$\mathcal{T}_{movement} = \frac{\sum_{lane_i \in lane_{in}} \mathcal{T}_{lane_i}}{|lane_{in}|} - \frac{\sum_{lane_j \in lane_{out}} \mathcal{T}_{lane_j}}{|lane_{out}|}, \quad (3)$$

where $|lane_{in}|$ and $|lane_{out}|$ denotes the cardinality of incoming lanes and the cardinality of outgoing lanes, respectively.

Intensity modeling

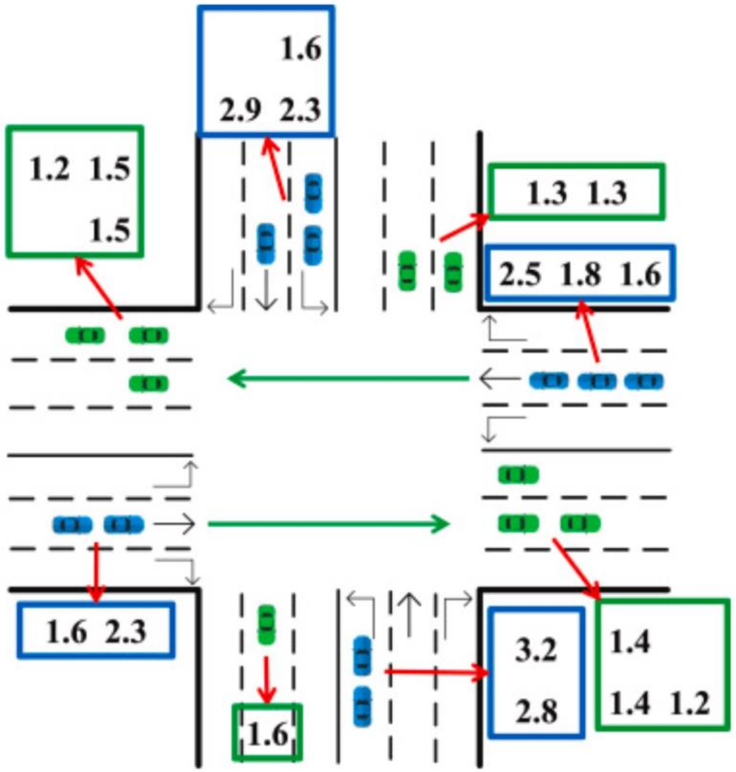
- **Definition 6 (Intensity of Phases)**

  : The intensity of a phase is the sum of the intensity of all the movements involved in this phase

$$\mathcal{I}_{phase} = \sum_{movement_i \in phase} \mathcal{I}_{movement_i}\,.$$

Example

$(1.6 + 2.3)/1 - (1.4 + 1.4 + 1.2)/3 + (2.5 + 1.8 + 1.6)/1 - (1.2 + 1.5 + 1.5)/3 \approx 7.07$



The blue vehicles are incoming vehicles and the green ones are outgoing vehicles

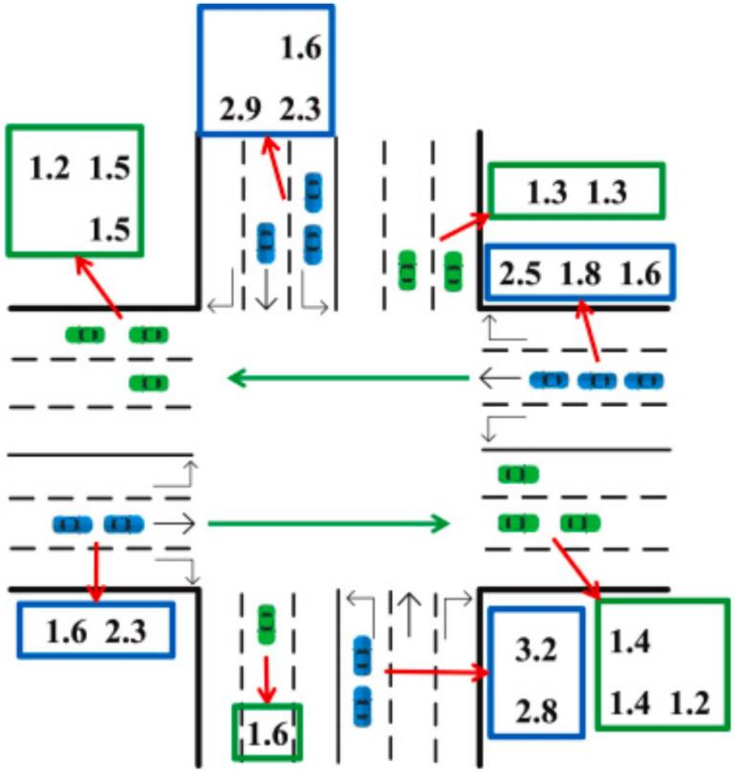
Intensity modeling

- **Definition 7 (Intensity of Intersections)**

  : intensity difference between the incoming lanes and the outgoing lanes

$$\mathcal{T}_{\mathcal{I}} = \sum_{lane_i \in lane_{in}} \mathcal{T}_{lane_i} - \sum_{lane_j \in lane_{out}} \mathcal{T}_{lane_j}$$

$$\mathcal{T} = (1.6 + 2.9 + 2.3 + 1.6 + 2.3 + 3.2 + 2.8 + 2.5 + 1.8 + 1.6) - (1.2 + 1.5 + 1.5 + 1.6 + 1.4 + 1.4 + 1.2 + 1.3 + 1.3) = 10.2$$



Intensity modeling

- **Definition 8 (Impacts of Neighboring Intersections)**

Assume that the current intersection is I, and I' is a neighboring intersection of I. The impact of I' on I is formulated as

$$\mathscr{P}_{I,I'} = \omega \times \sum_{lane_i \in nlane_{in}} \left(\mathscr{T}_{nlane_i} \times min\left(\frac{n \times t}{N_i}, 1 \right) \right),$$

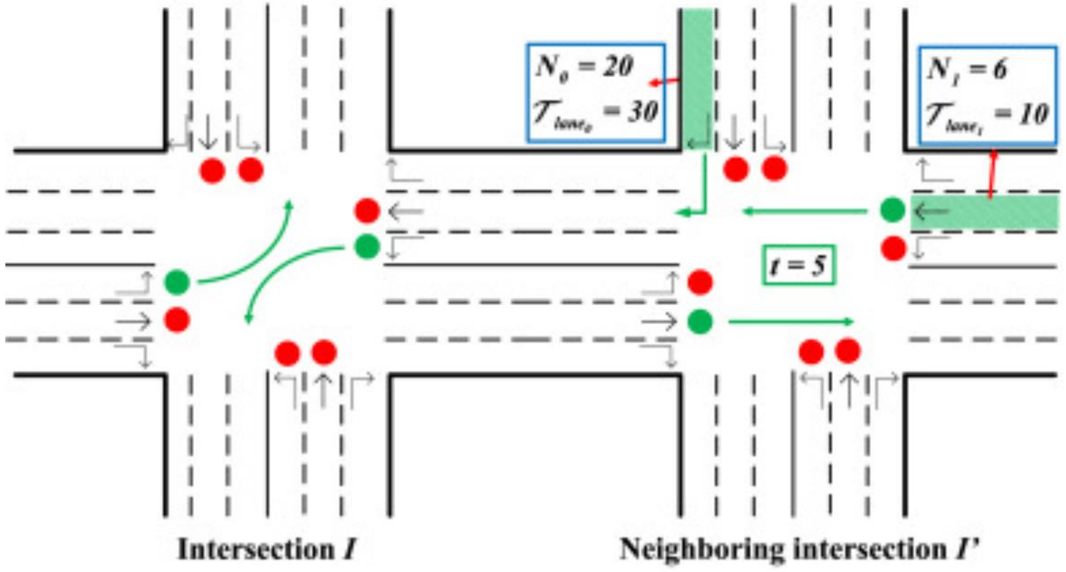
ω is the shrinkage factor to limit the impacts value
n is the number of vehicles traveling across the neighboring intersection within a unit time,
t denotes the remaining time of the current phase for the neighboring intersection.

Intensity modeling

- **Definition 8 (Impacts of Neighboring Intersections)**

Fig. 3 shows an example of two neighboring intersections, i.e., I and I', where the remaining time of green light of I' is 5 s. Assuming that $n = 2$, $\omega = 0.5$, the impact of I' on I can be calculated as

$$\mathscr{P}_{I,I'} = 0.5 \times \left(30 \times min\left(\frac{2\times5}{20},\ 1\right) + 10 \times min\left(\frac{2\times5}{6},\ 1\right)\right) = 0.5 \times (30 \times 0.5 + 10 \times 1) = 12.5$$



Agent design - Observation (State)

- **Information around the intersection and its immediate neighbors**
  : the intensity of each phase
  : the impacts derived from each immediate neighbor in four directions
  : the current phase

Under the four-phase setting, a state of the intersection can be encoded as $(\mathscr{T}_{phase_1}, \mathscr{T}_{phase_2}, \mathscr{T}_{phase_3}, \mathscr{T}_{phase_4}, \mathscr{P}_{I,I'_1}, \mathscr{P}_{I,I'_2}, \mathscr{P}_{I,I'_3}, \mathscr{P}_{I,I'_4}, phase)$, assuming that I'_1, I'_2, I'_3 and I'_4 are the four neighboring intersections of the current intersection I.

Agent design - Action

- The main task of an RL agent is to select the best phase to minimize **the intensity of the intersection**.
- When the time of the current phase is exhausted, the agent needs to select a new feasible phase

Agent design - Reward

- The **intensity of intersections** can reflect the **average travel time of vehicles** crossing the intersection more accurately

$r = -\mathscr{T}_I$, where $\mathscr{T}_I$ is the intensity of I

Phase duration design

- our approach needs to figure out **a set of phase durations** based on the given t_{min} and t_{max} , which represent the minimum and maximum allowable duration, respectively.
- When M=1, the phase duration is fixed and equals to t_{max}. When M>1, the phase duration set *D* can be constructed as follows

$$D = \left\{ t_{min} + i \times \Delta t : i \in \mathbb{N} \quad \& \quad i < M \quad \& \quad \Delta t = \frac{t_{max} - t_{min}}{M-1} \right\}$$
$$where\ M \in \mathbb{N}\ and\ M > 1.$$

Phase duration design

- **After selecting a phase**, the agent will **determine a proper duration for the phase** from the set *D* based on the observed number of vehicles on the incoming lanes under the restriction of the chosen phase.

$$duration = \left[\underset{t}{argmin} \left(t - \left\lceil \frac{N}{n} \right\rceil \right) \right]_{t_{min}}^{t_{max}}$$
$$where \; t \in D, \; and \; t \geq \left\lceil \frac{N}{n} \right\rceil.$$

Here, n is the number of vehicles passing through the intersection per unit time, and N represents the sum of total number of vehicles on all the incoming lanes. The operator $y = [x]_a^b$ means that: (i) if $x \geqslant a$ and $x \leqslant b$ then $y = x$; (ii) if $x \leqslant a$ then $y = a$; or (iii) if $x \geqslant b$ then $y = b$.

Learning process

- In our approach, we define one episode as the whole process of a certain period (e.g., one hour)
- Within an episode, an RL agent continuously interacts with the environment, where each interaction can be divided into five parts
  : (i) observing the environment to get the state;
  : (ii) selecting the phase action;
  : (iii) figuring out the duration of the selected phase action based on the current state;
  : (iv) simulating the traffic network based on the phase action;
  : (v) updating the weights of the action-value function

Experiment

Datasets

- **Synthetic datasets**: We considered four synthetic traffic datasets with different scales (i.e., 1×3, 2×2, 3×3, 4×4). The simulator generated 500 vehicles/h/lane on average following a Gaussian distribution. All the vehicles entered and left the network from rim edges (see details in [36]). The ratios of vehicles turning left, going straight, and turning right are 10%, 60% and 30%, respectively.
- **Real-world datasets**: We used two datasets collected from the real-world traffic of two cities (i.e., Hangzhou and Jinan) in China via roadside surveillance cameras. To enable the simulation of these two datasets on Cityflow, we adopted the traffic networks exported from GoogleMap as shown in Fig. 4. Fig. 4(a) shows the traffic network of Dongfeng sub-district used for the Jinan dataset, which contains 12 intersections in the form of a 3×4 grid. Fig. 4(b) shows the traffic network of Gudang sub-district for the Hangzhou dataset, which contains 16 intersections in the form of a 4×4 grid.

(a) Dongfeng, Jinan, China (b) Gudang, Hangzhou, China

Comparison

- **Fixed Time**
- **SOTL(Self-Organized Traffic Lights)**
- **3 RL algorithms**
  - **GRL**
  - **CoLight**
    : Use graph attentional networks to facilitate the communication among intersections.
  - **PressLight**
    : Select control phases for intersection pressure minimization

Table 2
Comparison of different traffic signal control methods.

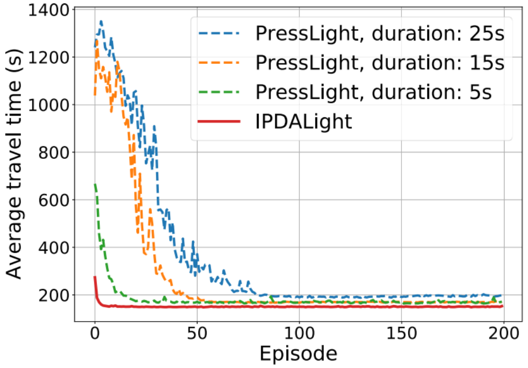
	RL-based	Number of vehicles	Vehicle dynamics	Pressure/Intensity	Intersection coordination	Variable phase duration
Fixed-Time						
SOTL		✓				
GRL	✓		✓		✓	
CoLight	✓	✓			✓	
PressLight	✓	✓		✓		
IPDALight	✓	✓	✓	✓	✓	✓

Comparison of average travel time

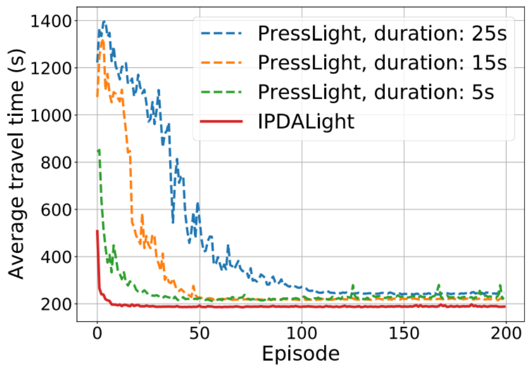
Table 3
Comparison results of average vehicle travel time.

Type	Method	Average vehicle travel time (s)					
		Syn_1 × 3	Syn_2 × 2	Syn_3 × 3	Syn_4 × 4	Jinan	Hangzhou
Non-RL	Fixed-Time [37]	384.47	454.01	508.87	565.99	405.91	488.51
	SOTL [38]	247.07	331.64	424.66	474.32	410.65	505.53
RL	GRL [39]	208.21	239.13	431.43	523.01	562.91	598.17
	CoLight [9]	210.01	312.29	328.70	397.07	327.65	337.45
	PressLight [8]	98.74	123.90	166.28	215.32	285.65	341.99
Ours	PressLight+Duration	90.85	113.27	153.52	196.68	281.31	331.12
	Intensity-Only	97.58	123.26	163.40	209.44	275.07	313.09
	IPDALight	**88.01**	**109.66**	**146.92**	**184.54**	**255.35**	**298.99**
Improvement over PressLight		10.87%	11.49%	11.64%	14.30%	10.61%	12.57%

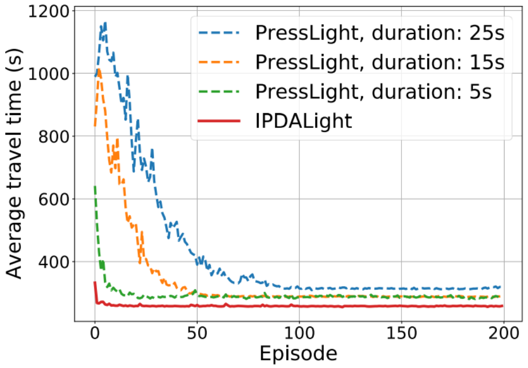
Comparison of learning convergence rate



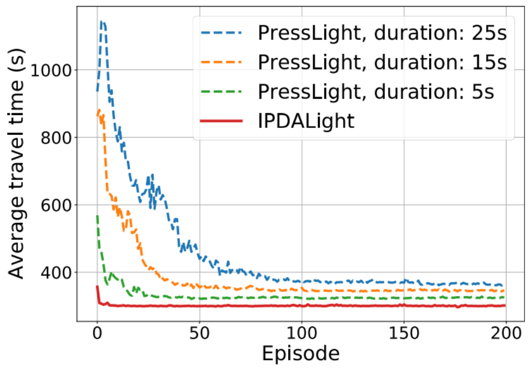
(a) Syn_3×3



(b) Syn_4×4



(c) Dongfeng, Jinan



(d) Gudang, Hangzhou

- **Investigat PressLight with different fixed phase durations (i.e., 5 s, 15 s and 25 s)**

Table 4
Convergence Information of PressLight and IPDALight.

Method/Duration	Start episode # of converge			
	Syn_3 × 3	Syn_4 × 4	Jinan	Hangzhou
PressLight/5 s	20	42	25	38
PressLight/15 s	55	55	60	73
PressLight/25 s	90	110	102	111
IPDALight	**5**	**11**	**5**	**7**

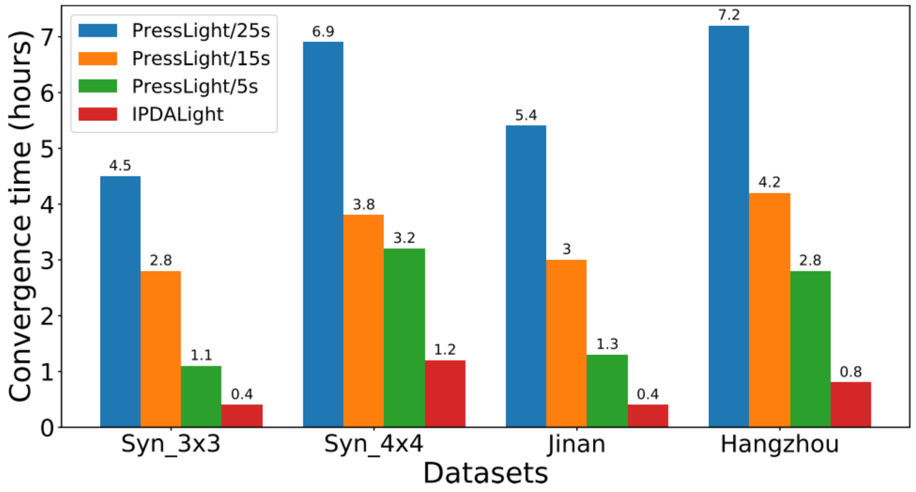


Fig. 7. RL training convergence time comparison.

Comparison of greenwave control effects

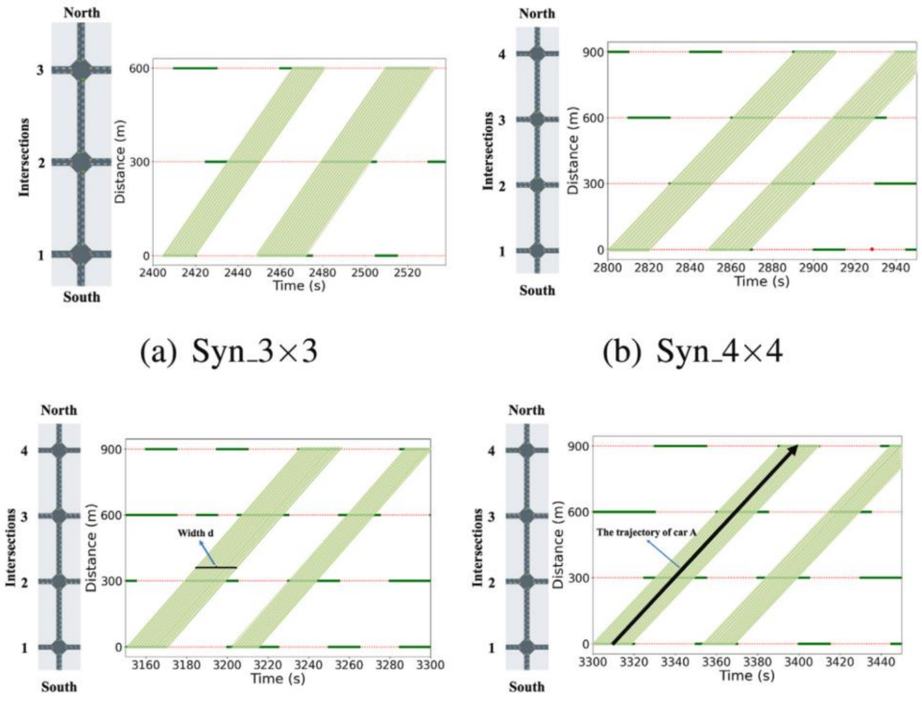
- A greenwave refers to a scenario that occurs when a series of traffic lights coordinate with each other to enable continuous traffic flow along one given direction.
- a key indicator is the **offset** between two adjacent intersections, which equals to **the distance between the two intersections divided by the free-flow speed of vehicles**.

Comparison of greenwave control effects

- we conducted experiments on the four collected datasets, i.e., Syn_3 × 3, Syn_4 × 4, Jinan and Hangzhou. In the experiments, the distance between two adjacent intersections is set to **300 m,** and the average speed of vehicles is set to **10 m/s**. Therefore, the offset here is 30s.
- For each dataset, we recorded the phase and duration change information of the traffic signals during the simulation under the control of models generated by IPDALight
- We randomly selected one arterial road with the north–south direction from each dataset and collected all the related simulation results for greenwave generation.

Comparison of greenwave control effects

- Vehicles traveling in the green oblique shaded area will fortunately experience greenwaves
- Table 6 presents the comparison results in terms of overall lasting time of greenwaves



(a) Syn_3×3 (b) Syn_4×4 (c) Dongfeng, Jinan (d) Gudang, Hangzhou

Table 6
Comparison of overall greenwave lasting time.

Method/Duration	Lasting time of greenwave (in s).			
	Syn_3 × 3	Syn_4 × 4	Jinan	Hangzhou
PressLight/5 s	257	142	110	35
PressLight/15 s	306	156	186	120
PressLight/25 s	345	175	198	101
IPDALight	**382**	**200**	**206**	**365**

두 논문을 리뷰해 보며...

- **DQN, Actor-critic과 같은 유명하지만 오래된 모델을 사용한 논문이 대부분**
  : 스터디에서 리뷰한 논문은 모두 DQN을 사용
  : DDQN, PPO와 같은 알고리즘 사용해보았으나 학습이 전혀 되지 않음... (이유를 찾는중)

- **현실에 적용하는 것에 대한 어려움**
  : 보행자에 대한 고려를 전혀 하지 않고 있음
  : 보행신호를 Signal Phase에 포함시키거나 Minimum phase time을 길게 줄 필요가 있음
  : 하지만 이 경우 신호제어의 flexibility를 크게 떨어뜨릴 수 있음

- **진짜 Performance가 좋은지에 대한 직접 검증이 필요**
  : 대전시를 대상으로 테스트 환경 구축해서 실험해볼 예정
  : 다들 자기들 논문이 가장 성능이 좋다고 하기 때문에, 검증이 필요해보임
  : Source Code, Dataset을 공개했기 때문에 검증이 가능할 것으로 기대됨
  : 환경구축 하는데 수작업이 많이 필요하고, 어려움이 있음

Additional information

- **Source code**
  : https://github.com/Dokyyy/IPDALight
- **Reinforcement learning for traffic signal control**
  : https://traffic-signal-control.github.io/
- **Cityflow**
  : https://arxiv.org/abs/1905.05217
  : Twenty times faster than SUMO
- **Tutorial**
  :https://docs.google.com/presentation/d/12cqabQ_V5Q9Y2DpQOdpsHyrR6MIxy1CJlPmUE3Ojr8o/edit

감사합니다

Q & A

Comparison of fairness