강화학습기반 신호 최적화 사례
(IntelliLight, PressLight, CoLight)

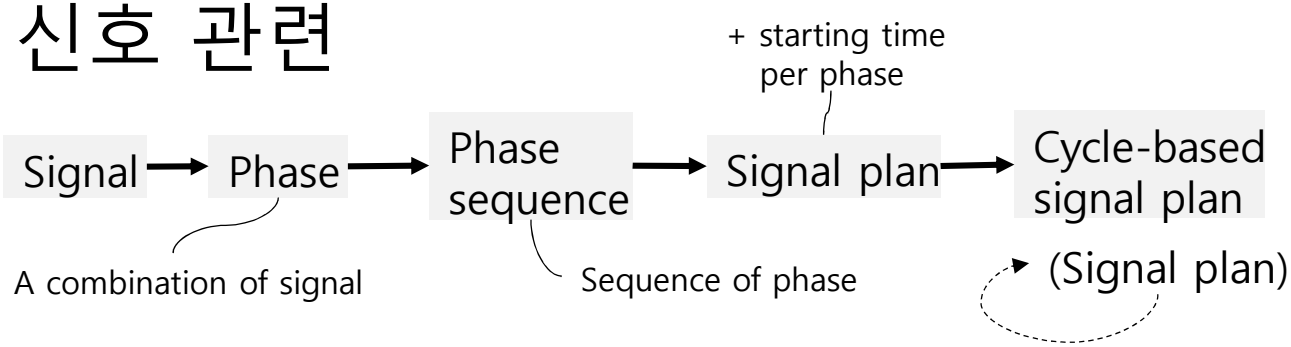
2020. 09. 17.

hunsoon@etri.re.kr

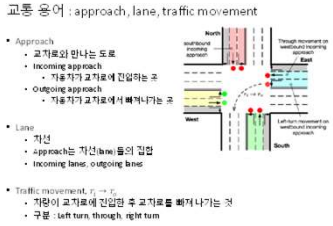
내용

- 교통 용어
- 교통 공학에서의 신호 최적화
- 강화학습기반 신호 최적화 개념
- 다중 교차로 환경에서의 강화학습
  - IntelliLight
  - PressLight
  - CoLight

교통 용어 : 교통 신호 관련



- (Movement) signal
- Phase
- Phase sequence
  - A sequence of phases
    - defines a set of phases and their order of change



- Signal plan
  - A sequence of phases and their corresponding starting time
  - (p_1, t_1) (p_2, t_2) .. (p_i, t_i) where p_i and t_i stand for a phase and its starting time
- Cycle-based signal plan
  - A kind of signal plan where the sequence of phases operated in a cyclic order

$$(p_1, t_1^1)(p_2, t_2^1)\ldots(p_N, t_N^1)(p_1, t_1^2)(p_2, t_2^2)\ldots(p_N, t_N^2)\ldots,$$

where $p_1, p_2, \ldots, p_N$ is the repeated phase sequence and t_i^j is the starting time of phase p_i in the j-th cycle.

$C^j = t_1^{j+1} - t_1^j$ is the cycle length of the j-th phase cycle

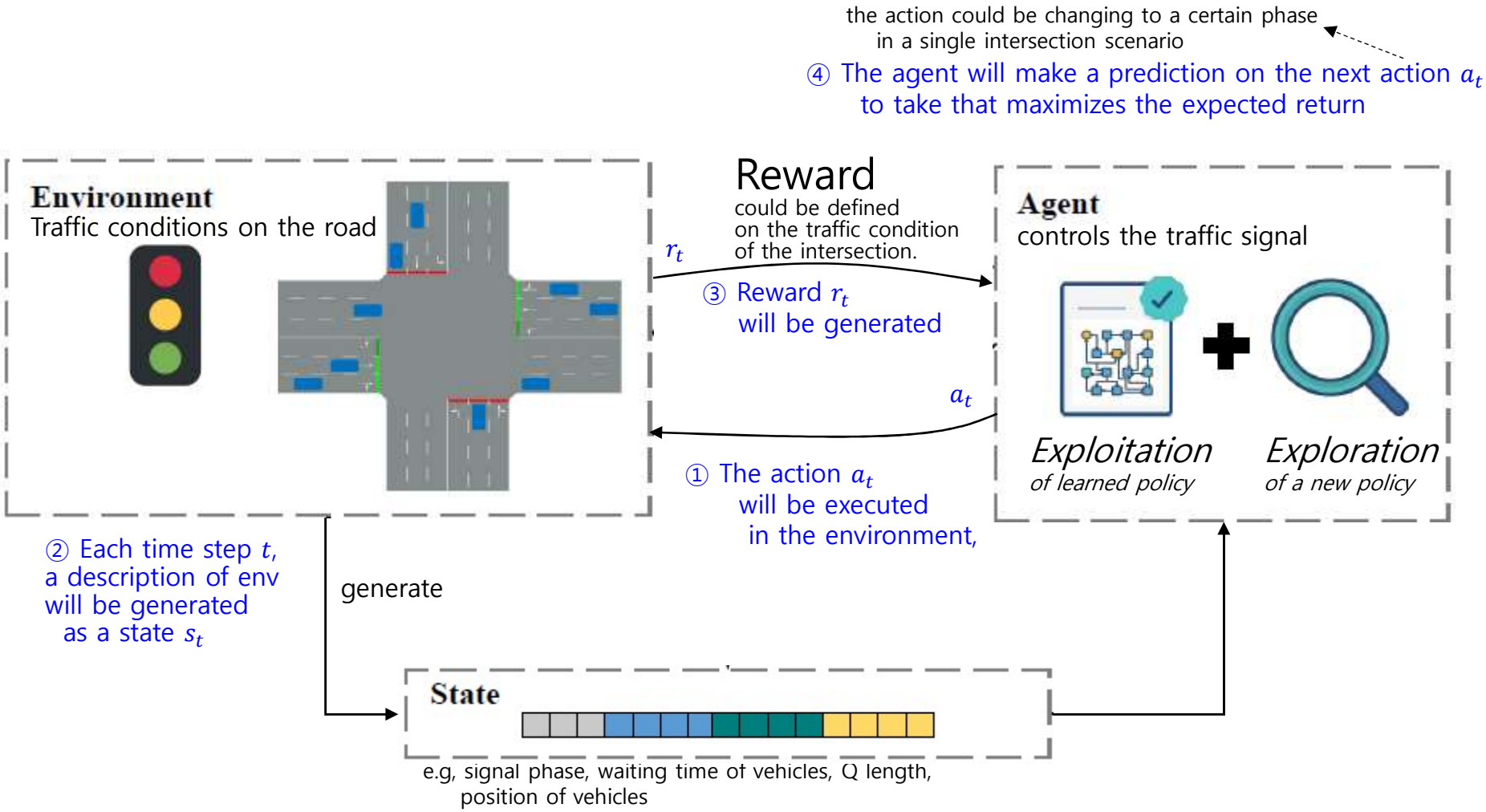
$\{\frac{t_2^j - t_1^j}{C^j}, \ldots, \frac{t_N^j - t_{N-1}^j}{C^j}\}$ is the phase split of the j-th phase cycle.

교통 공학에서의 신호 최적화(제어)

- 목적 : 교차로에서 차량의 안전하고 효율적인 이동 촉진
- 교차로 효율 측정 척도
  - Travel Time : 교통 시스템 진입과 진출 시간 차이
  - Queue Length : 교차로에서 대기 중인 차량 수
  - # of stops : 차량이 경험한 멈춤 회수
  - Throughput : 일정기간동안 도로 네트워크에서 여행(?)을 완료한 차량 수
- 방법 : 수식화하여 최적화 문제로 해결(가정 : 균등분포, 무한용량)

Method	Prior Knowledge	Data Input	Output
Webster (수식으로 cycle-length, phase split 계산)	하나의 Cycle을 구성하는 phase sequence	교통량	각 교차로에 대한 cycle-based 신호 계획(signal plan)
GreenWave(단방향), Maxband(양방향) (멈춤 최소화)	각 교차로의 cycle-based 신호 계획	교통량, 속도 제한, 차선 길이	cycle-based signal plan에서 오프셋(offset)
Actuated, SOTL (미리 정한 규칙에 따라)	phase sequence, phase 변경 규직	교통량	규칙과 데이터에 기반하여 다음 Phase 로의 변경
Max-pressure (Pressure 최소화)	None	대기열 길이	모든 교차로에 대한 signal plan
SCATS (포화도 최소화)	모든 교차로에 대한 signal plan	교통량	조정된 signal plan

RL framework for traffic light control



Ref. Hua Wei et al, "A Survey on Traffic Signal Control Methods,", ARXIV19

다중 교차로 환경에서 강화학습

- 협업(coordination)이 중요

- Global Single Agent
  - 하나의 전역 에이전트가 모든 교차로 제어
  - 모든 교차로의 state를 입력으로 모든 교차로의 action을 선택하는 법 학습

    $max_a\ Q(s, a)$

    where s is the global environment state,

    a is the joint action of all interactions

  - (-) Action space가 커짐에 따라 차원의 저주로 인해 학습이 안됨

  - 소규모 환경에서 실험됨 : 2개 연결 교차로, 5개 연결 교차로, 2X2 Grid

다중 교차로 환경에서 강화학습

- Joint Action Modeling
  - 에이전트들간의 협업 방법을 명시적으로 고려
    - Coordination graph, Max-Plus algorithm 이용
  - 전역 Q-함수를 지역 하위 문제(local-subproblem)의 선형 조합으로 분해

$$\hat{Q}(s,a) = \sum_{e} Q_e(s_e, a_e)$$

where e corresponds to a subset of neighboring agents

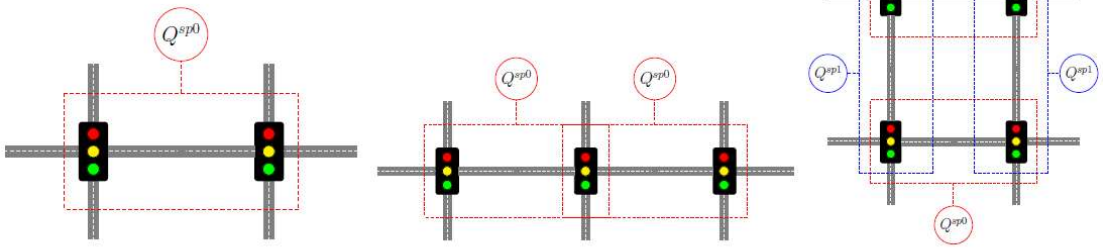
$$= \sum_{i,j} Q_{i,j}(o_i, o_j, a_i, a_j)$$

where i and j corresponds to the index of neighboring agents

Joint action

  - $Q_{i,j}$ 들의 예측 값들의 합을 최대화하는 a_i, a_j을 찾도록 학습

use the max-plus coordination algorithm
to optimize the joint global action over the entire coordination graph
(to know details, see ref. paper)



(a) Two agents (b) Three agents (c) Four agents

Ref. Elise van der Pol. 2016. Coordinated Deep Reinforcement Learners for Traffic Light Control. NeurIPS .... GRL

관련 논문

- Related concepts
  - coordination graphs; max-plus algorithm
- Joint Action Modeling
  - MA Wiering. "Multi-agent reinforcement learning for traffic light control," ICML2000
  - Lior Kuyer, Shimon Whiteson et al, "Multiagent reinforcement learning for urban traffic control using coordination graphs," In Joint European Conference on Machine Learning and Knowledge Discovery in Databases 2008. pp656–671.
  - Samah El-Tantawy and Baher Abdulhai. "Multi-agent reinforcement learning for integrated network of adaptive traffic signal controllers (MARLIN-ATSC)," ITSC2012, pp.319–326.
  - Samah El-Tantawy, Baher Abdulhai et al, "Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers (MARLIN-ATSC): methodology and large-scale application on downtown Toronto," ITS2013, vol.14, 3, pp.1140–1150.
  - Elise van der Pol. "Coordinated Deep Reinforcement Learners for Traffic Light Control," NeurIPS 2016. withSUMO

다중 교차로 환경에서 강화학습

- Independent RL w/o communication
  - 각각의 교차로가 하나의 에이전트에 의해 제어됨
  - 에이전트 간 communication 없음
  - 에이전트의 입력이 되는 상태는 담당하는 교차로의 교통 상황으로만 정의됨

    $max_{a_i} \sum_i Q_i(o_i, a_i)$

    where o_i is the local observation of intersection i
    and a_i is the action of intersection i

  - 단순한 환경(예, Arterial Network)에서는 GreenWave가 형성되어 좋을 수도 있음
  - 복잡한 환경에서는 다른 이웃하는 에이전트들로부터 일정하기 않은 영향을 받음 ➔ 학습이 수렴되지 않음

  - 사례 : IntelliLight, PressLight

FT : Fixed Time
SOTL : 대기 차량 수가 일정 기준 넘는 경우만 변경
(Self-Organized Traffic Light Control)
DRL : DQN + image only

IntelliLight

- DQN 이용해서 real-world traffic data로 실험
- Agent
  - State : Q길이, 대기 차량 수,차량들의 대기시간, Phase, 차량 위치(이미지)
  - Action : 다음 신호 Phase로 변경 여부(keep, change)
  - Reward : 아래 항목의 Weighted SUM
    - Q길이, ∑지연시간, ∑대기시간, 신호변경여부, 변경 이후 통과차량 수, 변경 이후 통과차량의 여행시간 합
  - 네트워크 구조 : DQN
  - Phase와 action에 따라 분리된 replay memory에 저장
- Experiments
  - Env : SUMO, syntactic data, Real-world data(Jinan@중국), 2 phase (동-서, 남-북)
  - FT, SOTL, DRL 대비 모든 측면(Reward, Q길이, 지연시간, 여행시간)에서 우수

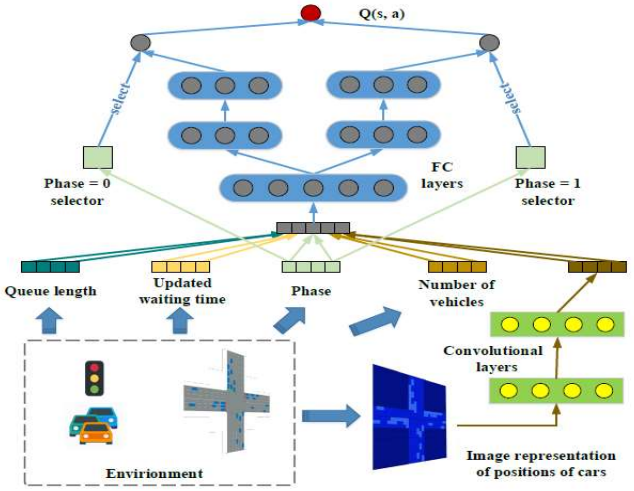


Figure 5: Q-network

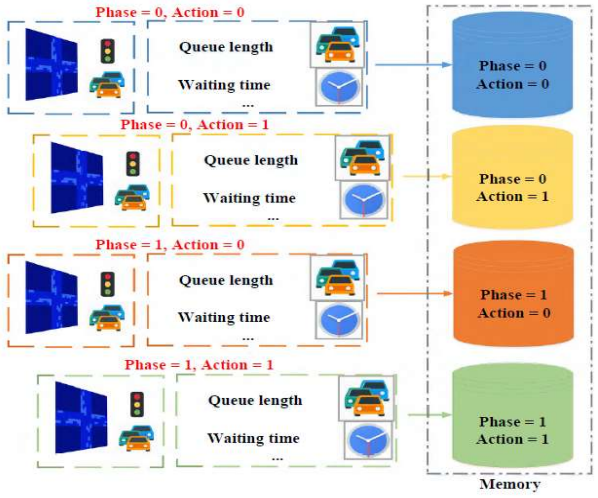
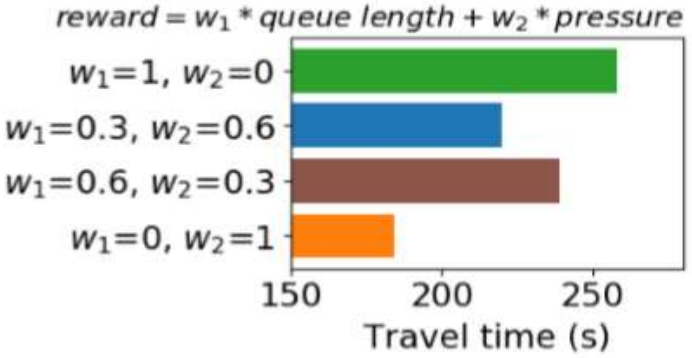


Figure 6: Memory palace structure

Hua Wei et al, "IntelliLight : A Reinforcement Learning Approach for Intelligent Traffic Light Control," KDD18

PressLight

- 강화학습 기반 교통 최적화에 교통 공학 이론 도입
  - 기존 : 휴리스틱에 기반한 상태(State) 보상(Reward) 설계
    - 성능이 일정하지 않고(Sensitive performance), 학습 시간이 길어짐
  - 제안 : 교통 공학 분야의 state-of-the-art인 Max-Pressure 기법을 상태와 보상에 접목
    - 단순한 상태로 학습시간 단축
    - 교통 공학 이론에 근거한 보상 설정으로 성능 향상



(a) Performance w.r.t. reward

상태: 진입차량 정보
보상 : - Q길이

(b) Convergence w.r.t. state

성능이 가중치에 따라 달라짐

상태 정보가 복잡해지면서
학습 시간 길어지지만 성능 향상은...

Hua Wei et al, "PressLight : Learning Max Pressure Control to Coordinate Traffic Signal in Arterial Network," KDD19

PressLight

차선 l 에서 차선 m으로 이동의 Pressure, $w(l,m)$

$$w(l,m) = \frac{x(l)}{x_{max}(l)} - \frac{x(m)}{x_{max}(m)}$$

$x(l)$: 해당 차선의 현재 차량 수

$x_{max}(l)$: 해당 차선의 최대 차량 수

교차로 i Pressure, P_i

$$P_i = \left| \sum_{(l,m)\in i} w(l,m) \right|$$

- Agent
  - State : current phase, 진입/**진출** 차선의 **segment별** 차량 수
  - Action : 다음 phase 임의 선택(4개 phase 중 하나)
  - Reward : -1 * 교차로의 **pressure**
  - 네트워크 구조 : DQN
- Experiments
  - Environment : CityFlow Simulator, syntactic data, Real-world data(NY@미국, Jinan@중국 등)
  - average travel time 측정

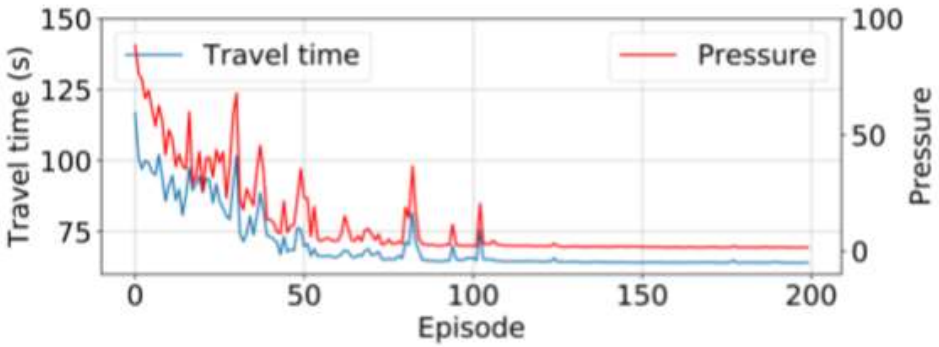
GRL : coordination graph + joint local Q-func on 2 adj intersection

  - 전통적 방법, 다른 강화학습 기반 방법(GRL, LIT) 보다 우수
  - State 와 reward의 평균 여행 시간에의 영향(아래 표)

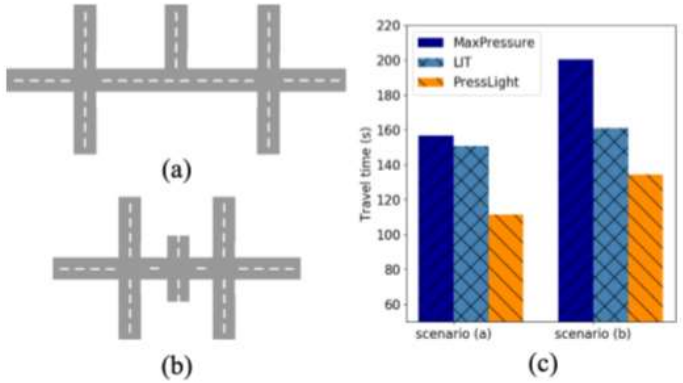
	Heavy Flat	Heavy Peak	state	reward	비고
LIT	233.17	258.33	Phase, 진입 차선별 차량 수	Q-length	
LIT + out	201.56	281.21	+ 진출 차선별 차량 수	Q-length	이웃 교차로 고려
LIT + out + seg	200.28	196.34	차량 수를 segment별로 세분화	Q-length	세분화→ Offset 학습
PressLight	160.48	184.51	Phase, 진입/진출 차선의 segment별 차량 수	Pressure	Reward 효과

Hua Wei et al, "PressLight : Learning Max Pressure Control to Coordinate Traffic Signal in Arterial Network," KDD19

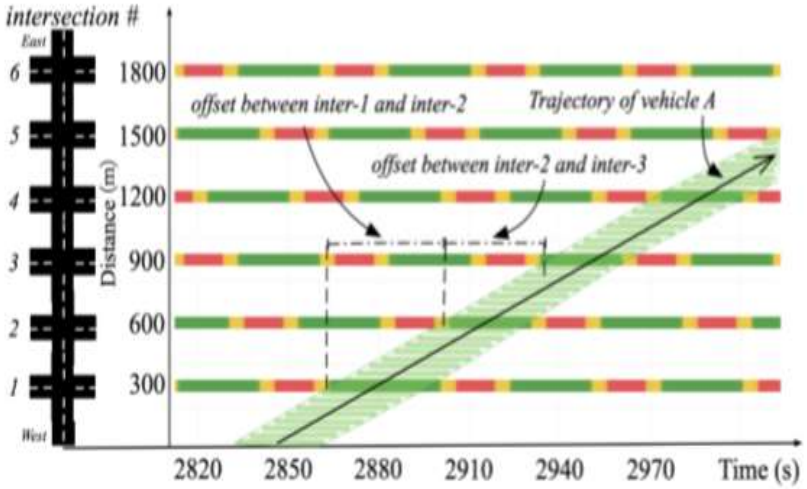
PressLight



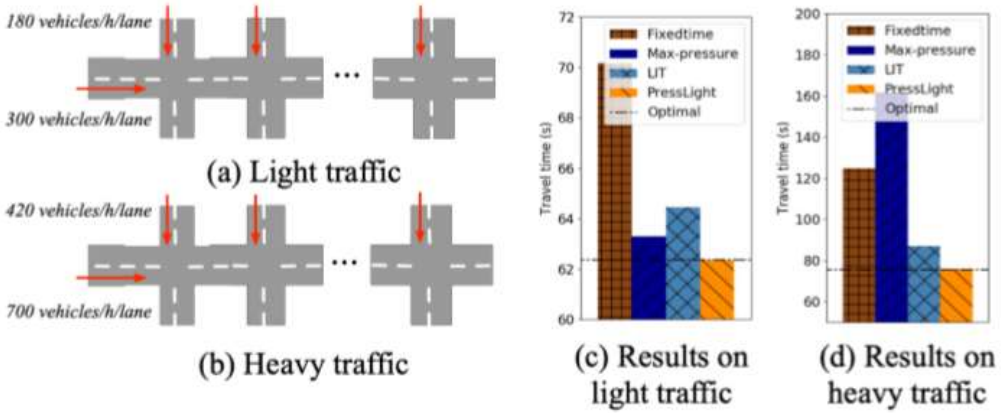
▲ Travel Time과 Pressure가 동일한 Trend로 수렴



▲ 이질적((a)leg, (b)length) 교차로에서도 평균 여행시간 우수



▲ Uni-directional uniform traffic 환경에서 Offset이 보임



▲ GreenWave가 optimal solution이라고 알려진 uniform uni-directional traffic 환경에서도 우수 (PressLight만 optimal 달성)

Hua Wei et al, "PressLight : Learning Max Pressure Control to Coordinate Traffic Signal in Arterial Network," KDD19

관련 논문

- Independent RL w/o communication
  - Mohammad Aslani, Stefan Seipel, Mohammad Saadi Mesgari, and Marco Wiering. "Traffic signal optimization through discrete and continuous reinforcement learning with robustness analysis in downtown Tehran," Advanced Engineering Informatics 38 (2018), 639–655.
  - Tianshu Chu, Jie Wang, Lara Codeca, and Zhaojian Li. "Multi-Agent Deep Reinforcement Learning for Large-scale Traffic Signal Control," arXiv 2019 (A2C… ITS2020) withSUMO
  - Yuanhao Xiong, Guanjie Zheng, Kai Xu, and Zhenhui Li. "Learning Traffic Signal Control from Demonstrations," CIKM2019 withCityFlow
  - Hua Wei, Chacha Chen, Guanjie Zheng, Kan Wu, Vikash Gayah, Kai Xu, and Zhenhui Li. "PressLight: Learning Max Pressure Control to Coordinate Traffic Signals in Arterial Network," KDD2019 withCityFlow
  - Guanjie Zheng, Yuanhao Xiong, Xinshi Zang, Jie Feng, HuaWei, et al. "Learning Phase Competition for Traffic Signal Control," CIKM2019 withCityFlow
  - Guanjie Zheng, Xinshi Zang, Nan Xu, Hua Wei, Zhengyao Yu, Vikash Gayah, Kai Xu, and Zhenhui Li. "Diagnosing Reinforcement Learning for Traffic Signal Control," arXiv 2019 withCityFlow
  - Xinshi Zang, Huaxiu Yao, Guanjie Zheng, Nan Xu, Kai Xu, and Zhenhui Li. "MetaLight: Value-based Meta-reinforcement Learning for Online Universal Traffic Signal Control," AAAI 2020 withCityFlow
  - Chacha Chen, Hua Wei, Nan Xu, Guanjie Zheng, et al. "Toward A Thousand Lights: Decentralized Deep Reinforcement Learning for Large-Scale Traffic Signal Control," AAAI2020 withCityFlow

다중 교차로 환경에서 강화학습

- Independent RL w/ communication
  - 각각의 교차로가 하나의 에이전트에 의해 제어됨
  - 에이전트 간 Communication을 함

    $max_{a_i} \sum_i Q_i(\Omega(o_i, N_i), a_i)$

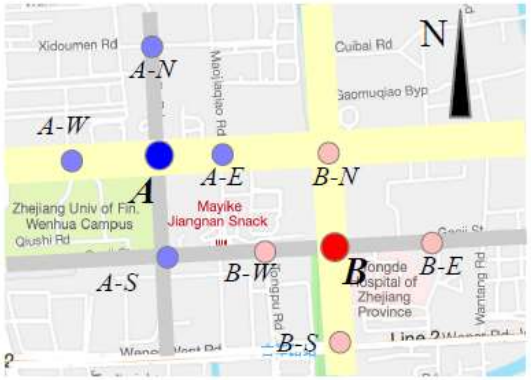
    where N_i is the neighborhood representation of intersection i,
    $\Omega(o_i, N_i)$ is **the func that models local observations and the obs.s of neighborhoods**
  - 에이전트 간 communication 을 통해 개별적이 아닌 그룹으로 동작 가능
  - 교통량이 심하게 변하고 교차로가 근접한 환경에서 협업이 중요

  - 이웃 교차로의 교통 상황을 o_i에 추가하기도 함
  - Graph Convolution Network을 이용하여 에이전트들간에 Communication을 학습
    - 이웃 에이전트의 은닉 상태간 상호 작용 학습
    - 교차로 사이의 Multi-hop의 영향도 학습
  - 사례 : CoLight

GCN(Graph Convolution Network) : 그래프 구조를 가진 데이터에서 예측을 하는 딥러닝 모델

CoLight

- 교차로의 교통 신호등들 사이에 어떻게 협업을 할수 있을까?
- 전통적 : 미리 계산된 두 교차로 신호간 Offset 이용
  - Uniform arrival rate, unlimited lane capacity 가정 → 실세계의 교통 환경(dynamic)에 부적합
- 기존 강화학습 : 중요도 고려없는 Observation(state)의 단순 취합
  - (예1) 진출 차량 정보, (예2)모든 이웃 교차로 State, (예3) 이웃의 은닉 상태(hidden state)
- 제안 : Graph Attention Network을 활용하여 동적 주변 교차로의 영향을 학습하여 협업
  - 이웃 교차로들의 시공간적 영향에 대해 학습 → 중요도
    - Up-Stream .vs. Down-Stream, Major-Road .vs. Side-Road, Morning .vs. Night
  - 이웃 교차로에 대한 Index-free modeling 이용
- 협업을 통해 대규모 교통 네트워크(196 교차로@Manhattan)에서도 좋은 성능
  - 기존 70개 교차로 이내가 최고



(a) Two target intersections in a real-world road network

(b) Model input alignment

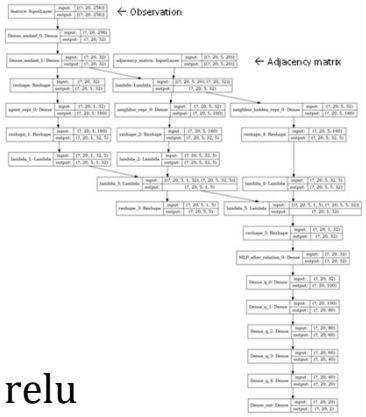
(c) Actual influence to target intersection

Hua Wei et al, "CoLight : Learning Network-level Cooperation for Traffic Signal Control," CIKM19

CoLight

$$\mathcal{L}(\theta_n) = \mathbb{E}[(r_i^t + \gamma \max_{a'} Q(o_i^{t\prime}, a_i^{t\prime}; \theta_{n-1}) - Q(o_i^t, a_i^t; \theta_n))^2] \quad (1)$$

- Agent : 하나의 교차로 담당
  - State : 자신&이웃의 o_i^t (current phase, 진입/진출 차선의 차량 수), adjacency matrix
  - Action a_i^t: Δt 시간 동안 유지될 다음 phase 선택
  - Reward r_i^t : $-\sum_l u_{i,l}^t$ (진입 차선의 대기 차량 수 합)
  - Loss : $E[$ 타겟 $-$ 예측$]^2$ …. (1)
  - 네트워크 구조
    : Observation embedding + **Graph Attentional Network** + Q-value prediction
- Observation embedding
  - K-차원 데이터를 m-차원의 잠재 공간으로 임베딩 $\in R^k$
  - Multi-Layer Perceptron 층 이용



  - *현재 교차로 i의 교통 상황* h_i

$$h_i = Embed(o_i^t) = \sigma(o_i W_e + b_e) \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

$where\ o_i^t \in R^k$ 시간 t에서 교차로i의 $observation$
k는 o_i^t의 특징의 차원, W_e 는 $weight\ matrix$, b_e는 $bias\ vector$, σ는 relu

Hua Wei et al, "CoLight : Learning Network-level Cooperation for Traffic Signal Control," CIKM19

CoLight

- Graph Attentional Network
  - Observation interaction
    - 교차로 i(target intersection)의 policy를 정하는데 교차로 j(source intersection)의 **중요도 계산**(attention score)

      $e_{i\,j} = (h_i\ W_t) \cdot (h_j\ W_s)^T$ .....(3)

      $where\ W_s, W_t \in R^{mXn}$ source와 target 교차로에 대한 임베딩 파라미터
  - Attention Distribution within Neighborhood scope
    - Target 교차로i와 이웃 교차로 사이의 Attention score를 **normalize**
    - $a_{i\,j} = softmax(e_{i\,j}) = \frac{\exp(e_{ij}/\tau)}{\sum_{j \in N_i} \exp(e_{ij}/\tau)}$ .....(4)
    - 이웃교차로 : road distance, node distance
  - Index-free neighborhood cooperation
    - 이웃하는 교차로의 표현을 그들의 **중요도에 따라 결합**
    - $hs_i = \sigma\left(W_q \cdot \sum_{j \in N_i} \alpha_{ij}\,(h_j W_c) + b_q\right)$ .....(5)
    - 이웃의 표현을 더함으로써, 모든 에이전트가 이웃 교차로를 인덱스에 따라 정렬할 필요가 없으므로 모델이 index-free이다.
  - Multi-head attention

Hua Wei et al, "CoLight : Learning Network-level Cooperation for Traffic Signal Control," CIKM19

CoLight

- Graph Attentional Network
  - Observation interaction
  - Attention Distribution within Neighborhood scope
  - Index-free neighborhood cooperation
  - Multi-head Attention
    - Attention을 여러 개 만들어서 **다양한 특징에 대한 Attention**을 보기 위함
    - 입력 데이터를 **head 수** 만큼으로 **나누어서** 병렬로 **계산**
      - 앞에서 살펴본 Single-head Attention을 여러 개 만듦

예, Shape가 (2, 3, 5)일 때
head 5이면
➔ shape가 (2, 3, 1)인 5개로 나누어서 attention 계산

$$e_{ij}^h = (h_i W_t^h) \cdot (h_j W_s^h)^T \tag{6}$$

$$\alpha_{ij}^h = \mathrm{softmax}(e_{ij}^h) = \frac{\exp(e_{ij}^h/\tau)}{\sum_{j\in\mathcal{N}_i} \exp(e_{ij}^h/\tau)} \tag{7}$$

$$hm_i = \sigma\left(W_q \cdot \left(\frac{1}{H}\sum_{h=1}^{h=H}\sum_{j\in\mathcal{N}_i} \alpha_{ij}^h (h_j W_c^h)\right) + b_q\right) \tag{8}$$

H : # of attention heads

Hua Wei et al, "CoLight : Learning Network-level Cooperation for Traffic Signal Control," CIKM19

CoLight

- Q-Value prediction : 각 action(신호)별 Q값을 계산하여 최적의 action 선택

$$
\begin{aligned}
hi &= Embed(o_i^t), \\
hm_i^1 &= GAT^1(h_i), \\
&\cdots \\
hm_i^L &= GAT^L(hm_i^{L-1}), \\
\widetilde{q}(o_i^t) &= hm_i^L W_p + b_p,
\end{aligned}
\tag{9}
$$

where $W_p \in \mathbb{R}^{c \times p}$ and $b_p \in \mathbb{R}^p$ are parameters to learn,
p is the number of phases (action space),
L is the number of GAT layers,
$\widetilde{q}$ is the predicted q-value

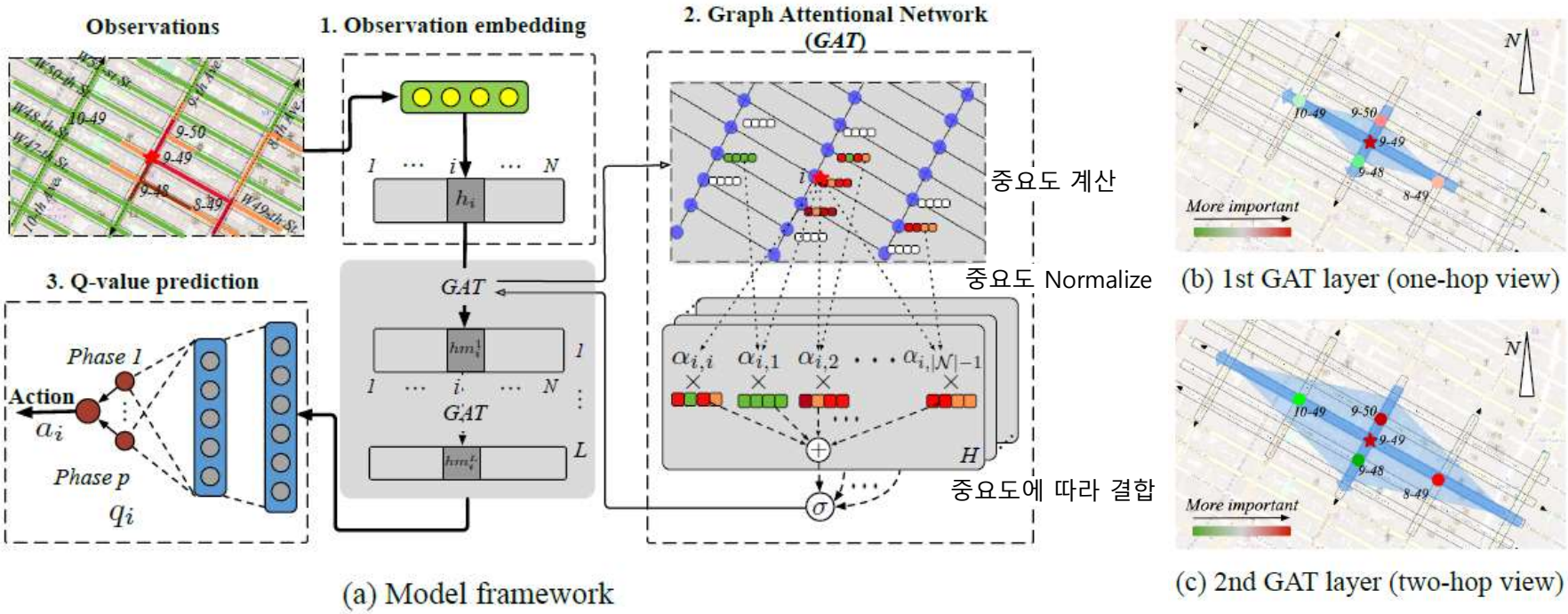
$$
L(\theta) = \frac{1}{T} \sum_{t=1}^{t=T} \sum_{i=1}^{i=N} \left(q(o_i^t, a_i^t) - \widetilde{q}(o_i^t; a_i^t, \theta) \right)^2, \tag{10}
$$

where T is the total number of time steps that contribute to the network update,
N is the number of intersections in the whole road network,
θ represents all the trainable variables in *CoLight*.

교차로별 $[E$타겟 $- E$예측$]^2$
모든 교차로에 대한 합
모든 step의 평균

Hua Wei et al, "CoLight : Learning Network-level Cooperation for Traffic Signal Control," CIKM19

CoLight



Left: Framework of the proposed CoLight model.
Right: variation of cooperation scope (light blue shadow, from one-hop to two-hop) and attention distribution(colored points, the redder, the more important) of the target intersection.

Hua Wei et al, "CoLight : Learning Network-level Cooperation for Traffic Signal Control," CIKM19

CoLight

Table 1: Data statistics of real-world traffic dataset

Dataset	# intersections	Arrival rate (vehicles/300s) Mean	Std	Max	Min
$D_{NewYork}$	196	240.79	10.08	274	216
$D_{Hangzhou}$	16	526.63	86.70	676	256
D_{Jinan}	12	250.70	38.21	335	208

- 실험
  - Environments : CityFlow Simulator
    - 단/양방향 합성 데이터 : $Arterial_{1X3}$ (spatial attention), $Grid_{3X3}$, $Grid_{6X6}$
    - 실 데이터 : NY@미국, Hangzhou/Jinan@중국 등
  - 평균 여행시간(Average travel time) 측정
  - 전통적 방법, 기존 RL 기반 방법과 비교 : CoLight이 가장 좋음
    - 평균 여행시간이 짧음
    - 빨리 수렴, 확장성 있음

(Individual RL, IRL)
이웃 교차로 정보 X,
파라미터 공유 X

(OneModel)
이웃 교차로 정보 X,
+ 에이전트간 파라미터 공유

(Neighbor RL)
+ 이웃 교차로 정보 단순 취합,
파라미터 공유

(GCN)
GCN 이용 이웃 교차로 정보 추출
이웃교통정보 동일하게 취급

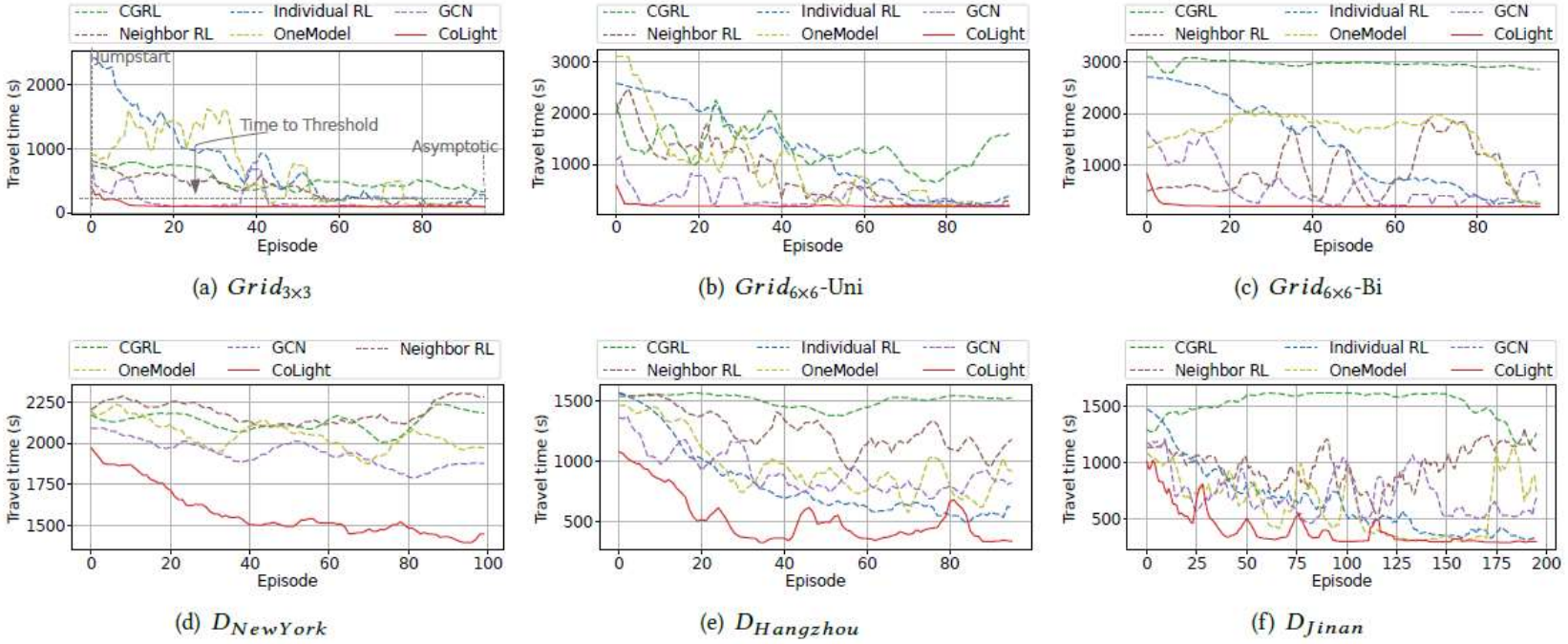
Model	$Grid_{6\times6}$-Uni	$Grid_{6\times6}$-Bi	$D_{NewYork}$	$D_{Hangzhou}$	D_{Jinan}
Fixedtime [15]	209.68	209.68	1950.27	728.79	869.85
MaxPressure [24]	186.07	194.96	1633.41	422.15	361.33
CGRL [23]	1532.75	2884.23	2187.12	1582.26	1210.70
Individual RL [30]	314.82	261.60	-*	345.00	325.56
OneModel [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 (geo-distance)	**173.79**	**170.11**	**1459.28**	**297.26**	**291.14**

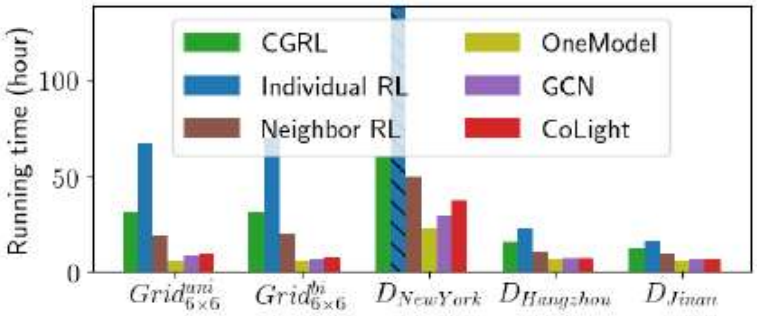
*No result as *Individual RL* can not scale up to 196 intersections in New York's road network.

Hua Wei et al, "CoLight : Learning Network-level Cooperation for Traffic Signal Control," CIKM19

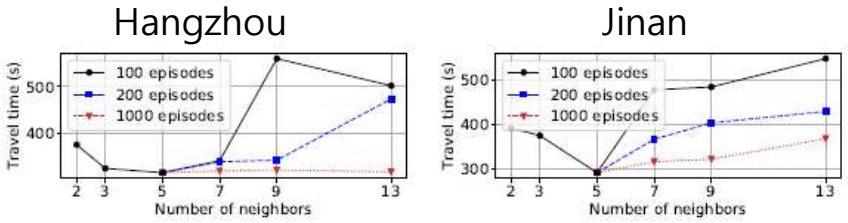
CoLight

▼ 수렴 시간: 짧음





▲ 100 episode에 대한 학습 시간 : 짧음



▲ 이웃이 많으면 성능이 좋으나
이웃 수가 5를 넘어가면 학습 시간이 많이 걸림

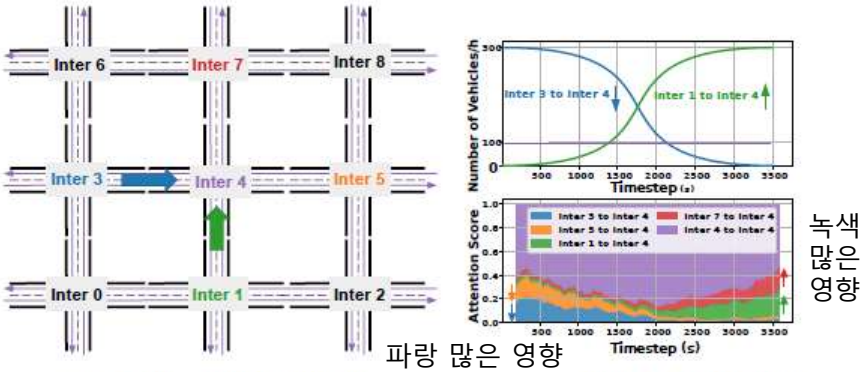
Hua Wei et al, "CoLight : Learning Network-level Cooperation for Traffic Signal Control," CIKM19

CoLight

Table 3: Performance of *CoLight* with respect to different numbers of attention heads (H) on dataset $Grid_{6\times6}$. More types of attention ($H \leq 5$) enhance model efficiency, while too many ($H > 5$) could distract the learning and deteriorate the overall performance.

#Heads	1	3	5	7	9
Travel Time (s)	176.32	172.47	170.11	174.54	174.51

▲ 헤드 수가 증가하면서 성능이 좋아지나 5를 넘어가면 오히려 나빠짐.($Grid_{6X6}$)



(a) $Grid_{3\times3}$ road network

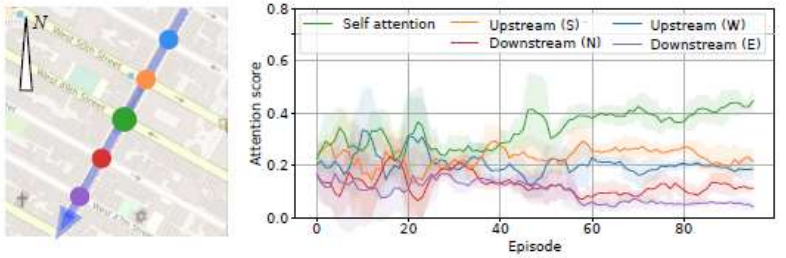
(b) Traffic flows and attention scores of neighbors to Inter#4

파랑 많은 영향

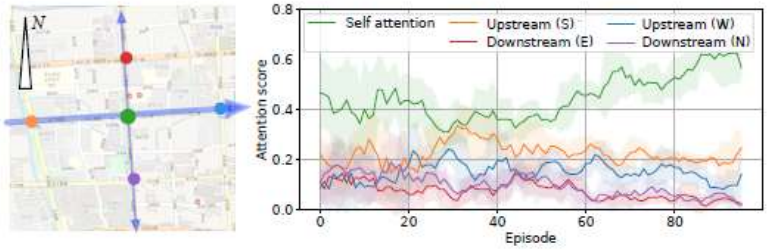
녹색 많은 영향

Figure 8: Temporal distribution of attention score learned by *CoLight* corresponds with temporally changing traffic.

▲ Attention Score의 변화가 시간에 따른 교통량의 변화와 일치



(a) Intersection A in New York



(b) Intersection B in Hangzhou

Figure 7: Spatial difference of attention distribution learned by *CoLight* during training process in real-world traffic. Different colored lines in the right figures correspond with the colored dots in the left figures. Up: For intersection A in $D_{NewYork}$, major concentration is allocated on upstream intersections and A itself. Down: For intersection B in $D_{Hangzhou}$, major concentration is allocated on arterial intersections and B itself.

▲ 공간에 따른 영향
(a) 자신과 Upstream의 영향을 많이 받음
(b) 자신과 주요 도로의 영향을 많이 받음

Hua Wei et al, "CoLight : Learning Network-level Cooperation for Traffic Signal Control," CIKM19

관련 논문

- Independent RL w/ communication
  - Itamar Arel, Cong Liu, T Urbanik, and AG Kohls. "Reinforcement learning-based multi-agent system for network traffic signal control," IET ITS 4, 2 (2010), pp.128–135
  - Samah El-Tantawy and Baher Abdulhai. "An agent-based learning towards decentralized and coordinated traffic signal control," ITSC2010, pp.665–670.
  - Sainbayar Sukhbaatar, Rob Fergus, et al. 2016. Learning multiagent communication with backpropagation. In NeurIPS. 2244–2252.
  - Tomoki Nishi, Keisuke Otaki, Keiichiro Hayakawa, and Takayoshi Yoshimura. "Traffic Signal Control Based on Reinforcement Learning with Graph Convolutional Neural Nets," ITSC 2018, pp.877–883. withSUMO
  - Zhi Zhang, Jiachen Yang, and Hongyuan Zha. "Integrating independent and centralized multi-agent reinforcement learning for traffic signal network optimization," arXiv 2019
  - Hua Wei, Nan Xu, Huichu Zhang, Guanjie Zheng, Xinshi Zang, Chacha Chen, Weinan Zhang, Yanmin Zhu, Kai Xu, and Zhenhui Li. "CoLight: Learning Network-level Cooperation for Traffic Signal Control," CIKM2019 withCityFlow