PRGLight: A novel traffic light control framework with Pressure-based-Reinforcement Learning and Graph Neural Network

Chenguang Zhao[1*], **Xiaorong Hu**[1] and **Gang Wang**[1]
[1]School of Eletronics and Information Engineering, Beihang University
{zchenguang,hxiaorong,gwang}@buaa.edu.cn

Abstract

Existing ineffective and inflexible traffic light control at urban intersections can often lead to congestion in traffic flows and cause numerous problems, such as long delay and waste of energy. How to find the optimal light decision strategy is a significant challenge in urban traffic management. In this paper, we propose PRGLight, a novel traffic light control framework with Reinforcement Learning (RL) and Graph Neural Network (GNN). In PRGLight, we firstly design a novel Pressure index, which considers the remaining capacity of the outgoing lane, as the reward function to control the traffic light phase. Furthermore, to avoid the myopic policy of RL, future traffic information is predicted by GNN to adjust the light duration. The proposed PRGLight integrates RL with GNN and can improve the efficiency of the traffic network in a long term. Experiments demonstrate that PRGLight can yield lower average travel time than state-of-the-art algorithms, such as CoLight, MaCar, on both synthetic and real-world data-sets.

1 Introduction

With the rapid increase in vehicle quantity, traffic congestion has become an urgent problem to be solved in many places around the world, especially in big cities. In order to reduce the waiting time of vehicles and to increase the carrying capacity of urban road network, Intelligent Transportation System (ITS) has become one of the hottest research issues in recent years, which aims to optimize the coordination and control of traffic flow. Fig.1 gives a general road network structure and intersection setting in our paper.

Traffic light control is an important part of ITS. Recently, with increasing availability of large volumes of sensors, RL-based traffic light control algorithms have been widely used and have given superb performance compared to traditional algorithms [Qin *et al.*, 2019]. In RL algorithms, there is an agent interacting with the environment. At each time-step, the agent observes a state from the environment and chooses an action to change the environment state. After this, the agent will get a reward evaluating the performance of the chosen action. During the process, the agent chooses action with the aim to maximize the accumulated discount reward. As can been seen, the reward function is a key element in RL algorithms and has a powerful influence on the performance of the algorithm. Recently, the concept of "max pressure" (MP) has been utilized as the reward for RL control model optimization [Wei *et al.*, 2019a; Chen *et al.*, 2020]. The pressure is defined as the difference between the number of vehicles on incoming lanes and the number of vehicles on outgoing lanes. This setting of reward ignores the remaining capacity of the outgoing lane and can result in unnecessary waste of green light resources. In this paper, We propose Pressure with Remaining Capacity of the Outgoing Lane (PRCOL), a more rigorous way of calculating pressure to capture the real-time feature of the incoming and outgoing lane condition. The proposed PRCOL is further adopted as the reward function in a RL traffic light control algorithm to adjust the traffic light phase.

Considering the characteristics of traffic control, the RL algorithm can be myopic and yield unsatisfactory strategy. For

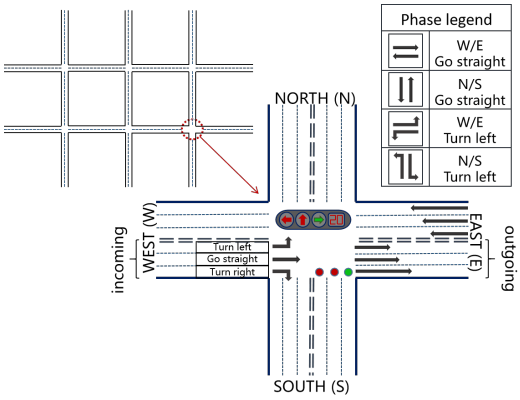


Figure 1: Road Network Structure and Intersection Setting. In this paper, we consider the widely seen multi-intersection grid road network, as shown in the upper left. For each intersection, there are four approaches with three incoming lanes on each, as shown in the downside. The traffic light has three states: red, green and yellow. We combine two non-conflicting lights as a phase and construct four phases as shown in the upper right box.

*Contact Author

instance, consider this example case: the traffic in the east-west is busy now, but few vehicles will come in following minutes; in contrast, few vehicles are in the north-south but there will be a busy traffic soon. Since the RL algorithm only considers the current or past traffic condition, it may set the east-west as green light and cause increase in the waiting time of the north-south traffic. This inspires us to consider the future traffic condition in order to increase the efficiency of traffic light control. Since we have no accurate information of the future traffic, traffic prediction is required to achieve the expected coordination.

The traffic prediction involves two aspects, to model the road network structure and to predict the traffic time series. A state-of-the-art approach to deal with these two tasks is GNN, which combines graph theory with the neural network algorithm. GNN-based solutions for traffic prediction have been adopted in previous literature and have achieved superior results than purely CNN- or RNN- based algorithms. [Cui *et al.*, 2020; Guo *et al.*, 2021; Diao *et al.*, 2019; Chen *et al.*, 2019].

In this paper, we integrate GNN with RL to dynamically adjust both the light phase and light duration. To be more specific, the proposed PRGLight algorithm is divided into two stages. Firstly, the RL agent chooses a green light phase based on the current traffic state. Then, to consider the future traffic condition, both the predicted traffic volume by GNN and the real-time traffic observation are used to help decide the light duration. To the best of our knowledge, this is the first work combining GNN prediction and RL decision in traffic light control to jointly consider the light phase and duration.

To understand the philosophy behind such approach, consider the case stated above. Since the traffic in east-west is busy now, the light phase should be set as east-west. Suppose the GNN predicts that only few vehicles will come in east-west, then there is no need to allocate long light duration to the east-west. To set the green light as north-south is inappropriate since there are only few vehicles in the current and the green light can increase the waiting time of east-west. To give the east-west a long light duration will also be inappropriate since there will only be few vehicles coming and therefore it can be helpful to set a short green light.

To summarize, the main contributions of this paper are as follows:

- We propose a novel pressure index PRCOL, which considers both the number of vehicles on the incoming lane and the remaining capacity of the outgoing lane.
- We design an RL algorithm with PRCOL as the reward function to decide the traffic light phase based on current traffic condition.
- We emphasize the importance of traffic prediction in traffic light control and adopt a GNN module to predict traffic condition. To improve coordination of intersections in the road network, we propose to decide the light duration by both the prediction from GNN and the light phase from RL.
- We conduct experiments on both synthetic and real-world data-sets. Extensive results demonstrate the effectiveness and rationality of the proposed algorithm. Detailed case studies are also presented to analyze the characteristic of the proposed algorithm.

2 Related Work

As a fundamental element in RL algorithms, the setting of reward will have a significant impact on the performance. Variables which are more easily to be observed, such as queue length or average delay, are often used as reward parameters, such as in [Wei *et al.*, 2019b; Chu *et al.*, 2019; Wei *et al.*, 2018; Genders and Razavi, 2020; Joo and Lim, 2020]. Nevertheless, such heuristic settings may cause high sensitivity and prolong learning process. [Wei *et al.*, 2019a] and [Chen *et al.*, 2020] propose a reward setting approach based on max pressure (MP) inspired by relevant research in the field of transportation [Varaiya, 2013]. The "pressure" is defined as the difference between the number of vehicles on incoming lanes and outgoing lanes. The MP does not take into account the carrying capacity of the lanes, so that excessive traffic flow may cause system inefficiency. In this paper, we design a novel pressure, PRCOL, and use it as the reward for RL algorithms.

Action is another basic component in RL algorithms.There are usually three action options for traffic light control problems. A simple way is to choose whether to switch to the next phase in a cycle-based light plan[Wei *et al.*, 2018], which is not flexible enough to cope with changing traffic condition. A most widely-used approach is to select the green phase for next state [Wei *et al.*, 2019a; Chen *et al.*, 2020; Wei *et al.*, 2019b; Chu *et al.*, 2019; Genders and Razavi, 2020]. The light duration in these work is set as a constant for all light phases. Such fixed duration may cause unnecessary delay when the vehicle's required pass time and the green phase duration do not match. In this paper, a more flexible approach is used, which can adjust the duration of the traffic light according to the traffic condition.

Despite the efforts towards intelligent traffic light control, most of exiting works consider only the current state. A light control approach that considers not only the current condition but also the future state may yield a better outcome. Regarding the spatial characteristics of road network, GNN-based algorithms have been taken as the state of the art traffic prediction algorithms. [Yu *et al.*, 2020] combines Communication Agent Network (CAN) and Traffic Forecasting Network (TFN) to adjust the light duration under a fixed light phase cycle. The reduction in the travel time proves the effectiveness of traffic prediction. In this paper, we design a more flexible traffic light control framework which can choose both the light phase and light duration.

3 Problem Formulation

In this section, we will introduce the road network and formulate the traffic prediction and traffic light control problem based on the road network.

3.1 Road Network

We use a complex road network with multiple intersections as the scenario for traffic prediction and light control, as shown in Fig.1.

For each intersection, there are four approaches from the four corresponding directions: east, west, north and south, denoted by "E", "W", "N", "S". This is one of the most common situations in real-world traffic. The approach in each direction is further divided into two directions: incoming and outgoing. Vehicles approach the intersection from the incoming lane, pass the intersection and then leave from the outgoing lane. The incoming lane is also divided into three lanes: left, right and straight, which means the vehicles traveling on it will go to three different directions respectively. Clearly, there are totally 12 possible routes for a vehicle to go through a intersection, and we call each route a traffic movement.

Each intersection can use three traffic lights i.e., red, yellow and green, to control the traffic flow passing it. Suppose that turning right is allowed anytime, the other 8 traffic movement directions are combined into 4 pairs, which are referred as 4 light phases, as shown in Fig.1. Each phase contains two non-conflicting traffic movement directions. At the green time, vehicles in the two directions corresponding to this phase are allowed to pass, while the remaining directions are set as red lights. A yellow light between two different phases is added to clear the vehicles passing through the intersection.

To facilitate the formulation of traffic prediction and light control problem, we mathematically re-state the road network of Fig.1 as a weighted undirected graph $\mathcal{G} = (V, E, W)$ with V, E, W being the node sets, link matrix, and weighted adjacency matrix respectively. Each intersection is denoted as a node and suppose there are N nodes $V = \{v_1, v_2, \cdots, v_N\}$. E represents the links in this undirected graph $\mathcal{G}$, indicating the intersections' connectivity. For intersections v_x and v_y, $e_{x,y}$ has value 1 when these two intersections are directly connected and 0 otherwise. $W \in \mathbb{R}^{N \times N}$ is the weighted adjacency matrix of graph $\mathcal{G}$ with $w_{x,y}$ being the weight between nodes v_x and v_y.

3.2 Traffic Prediction

In the road network graph $\mathcal{G}$, each node is endowed with some attribute characteristics, which represent practical traffic condition. The node attribute characteristics can be modeled from any traffic information, such as traffic flow, vehicle speed, etc. Denote $s_t^i \in \mathbb{R}^D$ as the attribute of node i at time-step t, and $s_t = (s_t^1, \cdots, s_t^N) \in \mathbb{R}^{N \times D}$ as the information of the whole network at time-step t. Given the history information of past H time-steps, $S_t = (s_{t-H+1}, \cdots, s_t) \in \mathbb{R}^{H \times N \times D}$, the predicted traffic attribute of next F time-steps is given by

$$\begin{aligned} &(\hat{s}_{t+1}, \cdots, \hat{s}_{t+F}) \\ &= \arg\max P\{\hat{s}_{t+1}, \cdots, \hat{s}_{t+F} \mid s_{t-H+1}, \cdots, s_t\}, \end{aligned} \tag{1}$$

where P is the prediction function to optimize.

Intuitively, the prediction should be as close to the real traffic observation as possible, so the objective function for traffic prediction is:

$$\min \sum_{\tau=H}^{T} \|\hat{s}_\tau - s_\tau\|^2, \tag{2}$$

where T is the total time-steps.

3.3 Traffic Light Control

In this paper, we formulate the traffic light control in the road network as a Markov Decision Process $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, R, \pi, \gamma \rangle$ with state space $\mathcal{S}$, action space $\mathcal{A}$, transition probability $\mathcal{P}$, reward R, policy π and discount factor γ. We set one agent for each intersection, who interacts with the environment and controls the traffic light. At each time-step t, the agent will observe the environment state s_t and choose an action $a_t \in \mathcal{A}$ guided by the policy π. The environment will execute the chosen action and transfer to next state s_{t+1} following the transition probability p. After this, the agent can get a reward R_t to evaluate the performance of this action. The goal of the agent is to maximize the long term reward, i.e.,

$$\max \sum_{\tau=1}^{T} \gamma^{T-\tau} R_\tau. \tag{3}$$

4 Algorithm Design

4.1 Traffic Prediction

In this paper, we adopt the traffic prediction network structure proposed by the earlier work [Yu *et al.*, 2018]. Nevertheless, To make this paper self-contained, we will also introduce the basic work flow of the network. The structure of the traffic prediction module in PRGLight is shown in the left part of Fig.2. It consists of two spatial-temporal convolution blocks and a fully-connected layer, which are cascaded together. Each convolution block contains two gated temporal convolutional layers and a spatial graph convolutional layer in the middle of them. The spatio-temporal correlation information of traffic flow is extracted by convolution blocks, and the features obtained are integrated and processed by the fully-connected output layer to generate prediction.

4.2 Traffic Light Control

As has been stated, we formulate the problem of road network light control as an MDP. In this subsection, we will explain the elements in this MDP in detail.

State. The state of each agent is a 12-dimension vector, representing the number of vehicles on the 12 incoming lanes.

Action. The action of each agent is to decide the traffic light of the intersection. The traffic light is decided by two parameters, the light phase and the light duration. Therefore, in this paper, we set the action a_t in each time-step t as the combination of two sub-actions: the light phase a_t^P and the light duration a_t^D. The light phase decides the traffic flow that can pass the intersection and the light duration decides how long the flow can pass. In most related work, only one of the two parameters is considered. In this paper, we jointly consider the light phase and light duration. We propose the light phase and the light duration should attach more importance to the current and future traffic condition respectively. The light phase should be set to alleviate the most busiest traffic but the light duration should further consider the future traffic.

In most related work, to deal with the huge state space, DQN is always utilized as the policy to get the action from the observed state. Inspired by the popular procedure in the DQN, the naive idea is to set the action space as both the light phase and the light duration. However, this can lead to a huge

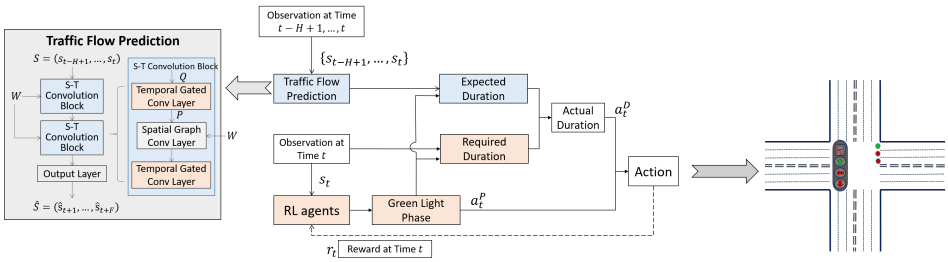


Figure 2: The framework of PRGLight. In PRGLight, the prediction of traffic volume is utilized to help decide the expected duration of green light, and the actual green light duration is decided by both the expected duration and the required duration derived from the DQN output.

action space and slow convergence rate. So in this paper, we propose to decide only the traffic light phase a_t^P by the DQN module and the light duration a_t^D is calculated by a simple equation that will be specified later in subsection 4.3. In such settings, the action space with respect to the DQN module is four dimension, corresponding to the 4 traffic phases shown in Fig.1.

Reward. In this paper, we improve the pressure algorithm [Wei *et al.*, 2019a] to consider the impact of the maximum carrying capacity of the outgoing lane. Our proposed pressure index is named as Pressure with Remaining Capacity of Outgoing Lane (PRCOL). The PRCOL is determined by the number of vehicles on the incoming and outgoing lanes and serves as an evaluation of the action. For a traffic movement i, the PRCOL is defined as:

$$P_i = N_{in} \times (1 - \frac{N_{out}}{N_{max}}) \,, \tag{4}$$

where N_{in} and N_{out} are the number of vehicles on the corresponding incoming and outgoing lane respectively, and N_{max} is the capacity of the outgoing lane. For an intersection, the reward is calculated from the sum of 12 traffic movements:

$$R = -\sum_i P_i. \tag{5}$$

The aforementioned settings of reward have achieved coordination between the incoming and outgoing lanes of an intersection. Note that in a road network, there should also be coordination between different intersections, since the traffic condition and phase setting of one intersection may affect another. For the communication between agents, we utilize the attention mechanism and adopt the similar Graph Attention Network used in CoLight [Wei *et al.*, 2019b].

4.3 Dynamic Light Duration

As mentioned earlier, to avoid the prohibitive computation introduced by merging the phase duration into the DQN action space, we aim to find a simple procedure to derive the light duration.

To consider the real-time traffic condition, suppose that there are N_{in} vehicles on the incoming lane and N_{out} vehicles on the outgoing lane and that the capacity of the outgoing lane is N_{max}, we then can get the estimated passing vehicles (EPV) as:

$$N_{pass} = \min\{N_{in}, N_{left}\}, \tag{6}$$

where $N_{left} = N_{max} - N_{out}$ is the number of remaining empty space of the outgoing lane. We assume there are N_{pass} vehicles waiting before the intersection statically and their gap is the minimum gap l_g. With the acceleration as a and the maximum speed as v, we can calculate the minimum required total time t_{req} for all the N_{pass} vehicles to pass the intersection.

To take into account the further traffic demand, suppose that there will be N_P vehicles according to the prediction of the GNN, the expected time t_{exp} for these N_P vehicles can also be derived with the following parameters l_g, a, and v. The final duration of the light phase is set as the minimum of t_{exp} and t_{req}:

$$a_t^D = \min\{t_{exp}, t_{req}\}. \tag{7}$$

4.4 Framework of PRGLight

The PRGLight framework is divided into two parts: traffic flow prediction and traffic light control, as shown in Fig.2. In PRGLight, the DQN agent first chooses a light phase according to the real-time traffic condition following ϵ-greedy policy. The GNN module records the history traffic state $(s_{t-H+1}, \ldots, s_t)$ and predicts the future traffic state $(\hat{s}_{t+1}, \ldots, \hat{s}_{t+F})$. The light duration is then decided by both the real-time traffic state s_t and the predicted traffic state $(\hat{s}_{t+1}, \ldots, \hat{s}_{t+F})$.

The pseudo-code of PRGLight algorithm is shown in Algorithm 1.

5 Experiment and Analysis

In this section, to demonstrate the effectiveness of the proposed PRGLight algorithm, we perform experiments and analyze the results. The experiments are performed based on the CityFlow[Zhang *et al.*, 2019] simulator .

5.1 Data-sets and Experiment Setting

In the experiment, we use one synthetic and three real-world data-sets. For the synthetic data-set, there is only one intersection and the traffic flow is generated manually. For the three real-world data-sets, Hangzhou, Jinan, and New-York, there are 16, 12, and 196 intersections respectively. Experiments are conducted in the same setup and simulation environment as CoLight[Wei *et al.*, 2019b] and MaCAR[Yu *et al.*, 2020], and the results are fairly compared.

Algorithm 1 PRGLight: Traffic Light Control based on GNN Prediction

Input: Graph $\mathcal{G} = (V, E, W)$, episode length T, greedy ϵ, update step-size η, target network replacement frequency C

Initialize Q with parameters θ, $\hat{Q}$ with parameters $\hat{\theta}$
for each $episode$ **do**
  Initialize step number t and total time t_{sum} to be 0
  while $t_{sum} < T$ **do**
    Predict traffic state: Get the predicted traffic state $(\hat{s}_{t+1}, \hat{s}_{t+2}, \cdots, \hat{s}_{t+F}) \leftarrow GNN(s_{t-H+1}, s_{t-T+2}, \cdots, s_t)$
    Choose Light Phase: $a_t^P \leftarrow DQN(s_t)$
    Decide Light Duration: $a_t^D \leftarrow (a_t^P; s_t, \hat{s}_{t+1}, \hat{s}_{t+2}, \cdots, \hat{s}_{t+F})$
    Execute $a_t \leftarrow \{a_t^P, a_t^D\}$, observe new state s_{t+1}, get reward R_t
    $t_{sum} \leftarrow t_{sum} + a_t^D$, $t \leftarrow t + 1$
    Update Q by gradient descend with step-size η
    Update $\hat{Q}$: $\hat{Q} \leftarrow Q$ every C steps
  end while
end for

The GNN module for traffic prediction is pre-trained on Beijing data-set[Zhang *et al.*, 2017]. Note that the GNN requires the same topology between training and prediction, we cut out a portion of the origin Beijing data-set according to the topology of traffic light control scene. The distance between each node is calculated according to the respective road network. The data of past 10 minutes is utilized to predict the traffic volume of next 5 minutes.

Regarding the calculation of the proposed PRCOL in Eq. (4), the number of vehicles on the incoming lane N_{in} and the number of vehicles on the outgoing lane N_{out} can be measured by some sensors or camera in practice. There are also APIs in CityFlow that can give these measurements. For N_{max}, the maximum number of vehicles that can fit in the outgoing lane, we assume that the length of the outgoing lane is l_l, the average length of the vehicle is l_v and the minimum gap between two vehicles is l_g. The N_{max} then can be simply calculate as $\lfloor l_l/(l_v + l_g) \rfloor$. ($\lfloor x \rfloor$ means the maximum integer that does not exceed x.) In the experiment, for all the four data-sets, we take $l_v = 5\ m$ and $l_g = 2.5\ m$. Regarding the calculation of t_{exp} and t_{req}, we take $a = 2\ m/s^2$ and $v = 40\ km/h$. For all traffic light control algorithms, a yellow light of 5 seconds is added to clear the traffic and avoid collision whenever the green light swift to another phase.

In the experiment, the greedy ϵ in Algorithm 1 is decreasing from 0.8 to 0.2. The discount factor γ for calculating the accumulated reward is set as 0.8. The maximal sample size is $10,000$. The target Q network $\hat{\theta}$ is updated every 5 steps. All source code and data-sets are available online at https://github.com/wangf622.

5.2 Baseline

For the synthetic data-set, PRGLight is compared with FixedTime, MaxPressure [Varaiya, 2013], and an advanced reinforcement learning control approach PressLight [Wei *et al.*, 2019a]. For the three real-world data-sets, PRGLight is compared with several state-of-the-art approaches, including CGRL [Van der Pol and Oliehoek, 2016], NeighborRL [Arel *et al.*, 2010], GCN [Nishi *et al.*, 2018], OneModel [Chu *et al.*, 2019], Individual RL [Wei *et al.*, 2018], CoLight [Wei *et al.*, 2019b], and MaCAR [Yu *et al.*, 2020]. Among them, MaCAR takes into account the role of traffic prediction as our approach, and performs best among these baseline approaches.

Methods	Average Travel Time	Throughput
FixedTime	572.73	1206
MaxPressure	280.14	2062
PressLight	236.31	1995
PR-Light	221.75	2248
G-Light	198.41	2217
PRGLight	**191.31**	**2378**

Table 1: Average travel time and Throughput on Synthetic Data-set

Methods	Hangzhou	Jinan	New-York
CGRL	1582.25	1210.70	2187.12
NeighborRL	1053.45	1168.32	2280.92
GCN	768.43	625.66	1876.37
OneModel	394.56	728.63	1973.11
Individual RL	345.00	325.56	-
CoLight	297.26	291.14	1459.28
MaCAR	291.18	279.49	1425.00
PRGLight	**283.06**	**276.00**	**1406.12**
	± 9.26	**± 4.98**	**± 167.90**

Table 2: Average travel time on real-world data-sets. In the last two lines, we list the average and standard variation respectively.

5.3 Result and Analysis

In Table 1, we list the average travel time and throughput of the synthetic data-set. PR-Light only keeps the PRCOL module and adopts fixed light duration. G-Light only utilizes the GNN module and abandons the PRCOL. In Table 2, we list the average travel time of the three real-world data-sets. Among all data-sets, the proposed PRGLight yield lowest average time. In table 3, we test the result with variations of the proposed PRGLight algorithm. For FIXED, the light duration is set 10 seconds. For DYNAMIC, the light duration is set based only on the current traffic condition without prediction. For HA, the GNN module is replaced by the history averaging prediction.

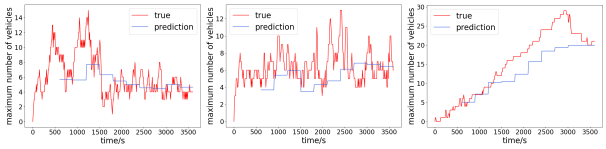
5.4 Case Study

To gain a more clear and intuitive understanding of how the PRGLight makes choices and controls traffic flow, we present some case studies in this subsection.

Prediction of traffic volume. In the proposed PRGLight algorithm, an accurate prediction of the future traffic volume is important, because the maximum green light duration will adjust according to the predicted number of vehi-

	Hangzhou	Jinan	New-York
FIXED	291.14	291.20	1616.98
DYNAMIC	289.73	287.49	1559.39
HA	286.56	281.51	1541.75
PRGLight	**283.06**	**276.00**	**1406.12**

Table 3: Ablation Study. Average travel time on real-world data-sets



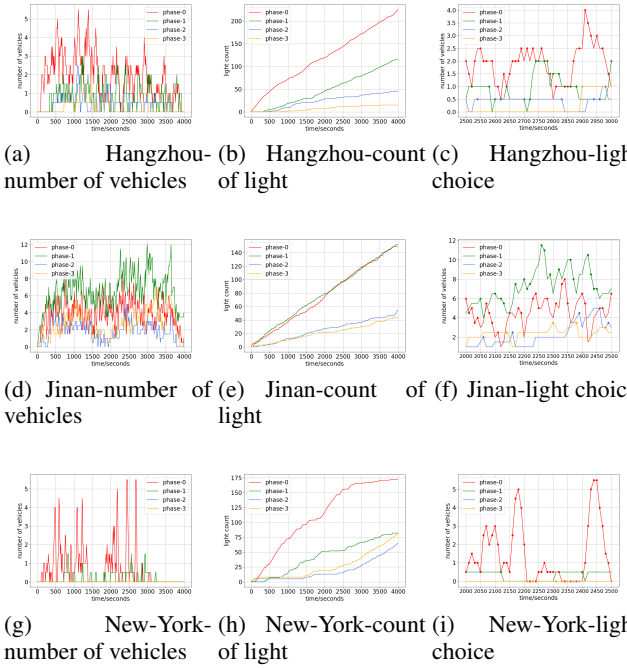
(a) Hangzhou (b) Jinan (c) New-York

Figure 3: Predicted and real traffic volume. We have numbered each intersection in the road network. For example, Hangzhou-inter-4 represents the forth intersection in the Hangzhou data-set.

cles. Fig.3 draws the *predicted* and *real* number of vehicles on three data-sets. The *prediction* curve begins at around 600-th seconds since there is not enough history data to be input to the GNN during the first 10 minutes. The *prediction* and *real* curves show approximate trend with the *real* curve being more variant.

One can observe a delay from the *real* and *prediction*. Take the Hangzhou data-set as an example, the *real* surges at around 1000-th seconds as the manual flow drives to the network, while the *prediction* increases at around 1200-th seconds. The main reason can be summarized as follows. The prediction is based on the history data, and if the history shows no clear tendency, it will be unreasonable to predict a heavy traffic. Before the arrive of heavy vehicle flow, the existing traffic condition presents little sign of heavy traffic. The heavy flow arrives at 900-th second and such traffic feature is captured and input to the GNN after 5 minutes (300 seconds) according to the experiment setting. The prediction thus increases at 1200-th second. In practice, however, data of last day or last week can help to predict the traffic peak and avoid such delay. We list this as a future research direction.

Choice and duration of green light. PRGLight controls the traffic flow by deciding the phase and duration of the green light. A detailed survey into the choice of PRGLight can be therefore helpful and instructive. Take Hangzhou as an example, Fig.4(a) draws the number of vehicles corresponding to the four phases; Fig.4(b) draws the accumulated time a light phase has been chosen; Fig.4(c) gives the detail of traffic condition and light control. In Fig.4(c), the curve represents the number of vehicles of each phase, and the point on each curve means one choice of the corresponding phase. For the Hangzhou data-set, it can be seen that for the phase-1 or phase-2 or phase-3, the interval between two choices is relatively long, compared to phase-0. One rational explanation for this can be stated as follows. The aim of traffic light control is to minimize the *average* travel time of all vehicles. For a same interval, green light on the phase with more number of vehicles means less average results. Therefore, the phase with maximum number of vehicles will be chosen priorly. However, if the vehicles of one phase have been waiting long enough, it will be reasonable to set the green light for this phase since some extremely huge numbers may have a deep impact on the result. The analysis on the Jinan and New-York presents similar result.



(a) Hangzhou-number of vehicles (b) Hangzhou-count of light (c) Hangzhou-light choice

(d) Jinan-number of vehicles (e) Jinan-count of light (f) Jinan-light choice

(g) New-York-number of vehicles (h) New-York-count of light (i) New-York-light choice

Figure 4: Number of vehicles and the detail of light choice on the three real-world data-sets

6 Conclusion

In this paper, we have proposed PRGLight, a reinforcement learning algorithm that combines traffic prediction and light control for intelligent traffic control problem. The proposed algorithm first uses GNN, which combines graph theory with convolutional neural network, to predict the traffic flow in the short-term. A RL module with the novel PRCOL as the reward function has been applied to choose the light phase. After that, the predicted information and the real-time observation have been used in the traffic light control to decide the light duration. Experiments on both synthetic and real-world data-sets have shown that the proposed PRGLight algorithm decreases delay of the traffic network compared to the state of the art algorithms.

Acknowledgments

This work was supported by the Funds of the National Natural Science Foundation of China (Grant No. U2033215), and the National Key R&D Program of China (Grant No. 2018YFB1601200).

References

[Arel *et al.*, 2010] Itamar Arel, C. Liu, T. Urbanik, and Airton Kohls. Reinforcement learning-based multi-agent system for network traffic signal control. *Intelligent Transport Systems, IET*, 4(2):128–135, July 2010.

[Chen *et al.*, 2019] Cen Chen, Kenli Li, Sin G Teo, Xiaofeng Zou, Kang Wang, Jie Wang, and Zeng Zeng. Gated residual recurrent graph neural networks for traffic prediction. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*, Hawaii, USA, February 2019. AAAI.

[Chen *et al.*, 2020] Chacha Chen, Hua Wei, Nan Xu, Guanjie Zheng, and Zhenhui Li. Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control. In *Proceeding of the Thirty-fourth AAAI Conference on Artificial Intelligence (AAAI-20)*, New York, USA, February 2020. AAAI.

[Chu *et al.*, 2019] Tianshu Chu, Jie Wang, Lara Codecà, and Zhaojian Li. Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE Transactions on Intelligent Transportation Systems*, 21(3):1086–1095, 2019.

[Cui *et al.*, 2020] Zhiyong Cui, Kristian Henrickson, Ruimin Ke, and Y. H. Wang. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 21(11):4883–4894, 2020.

[Diao *et al.*, 2019] Zulong Diao, Xin Wang, Dafang Zhang, Yingru Liu, Kun Xie, and Shaoyao He. Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*, Hawaii, USA, February 2019. AAAI.

[Genders and Razavi, 2020] Wade Genders and Saiedeh Razavi. Policy analysis of adaptive traffic signal control using reinforcement learning. *Journal of Computing in Civil Engineering*, 34(1):04019046, 2020.

[Guo *et al.*, 2021] Kan Guo, Yongli Hu, Zhen Qian, Hao Liu, and Baocai Yin. Optimized graph convolution recurrent neural network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 22(2):1138–1149, 2021.

[Joo and Lim, 2020] Hyunjin Joo and Yujin Lim. Reinforcement learning for traffic signal timing optimization. In *2020 International Conference on Information Networking (ICOIN)*, Barcelona, Spain, January 2020.

[Nishi *et al.*, 2018] Tomoki Nishi, Keisuke Otaki, Keiichiro Hayakawa, and Takayoshi Yoshimura. Traffic signal control based on reinforcement learning with graph convolutional neural nets. pages 877–883. IEEE, 11 2018.

[Qin *et al.*, 2019] Zhiwei (Tony) Qin, Jian Tang, and Jieping Ye. Deep reinforcement learning with applications in transportation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, page 3201–3202, New York, NY, USA, 2019. ACM.

[Van der Pol and Oliehoek, 2016] Elise Van der Pol and Frans A. Oliehoek. Coordinated deep reinforcement learners for traffic light control. In *NIPS'16 Workshop on Learning, Inference and Control of Multi-Agent Systems*, December 2016.

[Varaiya, 2013] Pravin Varaiya. The max-pressure controller for arbitrary networks of signalized intersections. *Advances in Dynamic Network Modeling in Complex Transportation Systems*, pages 27–66, February 2013.

[Wei *et al.*, 2018] Hua Wei, Guanjie Zheng, Huaxiu Yao, and Zhenhui Li. Intellilight: A reinforcement learning approach for intelligent traffic light control. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, (KDD 2018)*, pages 2496–2505, London, UK, August 2018. ACM.

[Wei *et al.*, 2019a] 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. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, Alaska USA, August 2019. ACM.

[Wei *et al.*, 2019b] 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. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing, China, November 2019. ACM.

[Yu *et al.*, 2018] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, Stockholm, Sweden, July 2018. IJCAI.

[Yu *et al.*, 2020] Zhengxu Yu, Shuxian Liang, Long Wei, Zhongming Jin, Jianqiang Huang, Deng Cai, Xiaofei He, and Xian-Sheng Hua. Macar: Urban traffic light control via active multi-agent communication and action rectification. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI 2020)*, pages 2491–2497, Yokohama, Japan, July 2020. ijcai.org.

[Zhang *et al.*, 2017] Junbo Zhang, Yu Zheng, and Dekang Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proceedings of the Thirty-first AAAI Conference on Artificial Intelligence (AAAI-17)*, California, USA, February 2017. AAAI.

[Zhang *et al.*, 2019] Huichu Zhang, Siyuan Feng, Chang Liu, Yaoyao Ding, Yichen Zhu, Zihan Zhou, Weinan Zhang, Yong Yu, Haiming Jin, and Zhenhui Li. Cityflow: A multi-agent reinforcement learning environment for large scale city traffic scenario. In *The World Wide Web Conference*, California, USA, May 2019. ACM.