A novel reinforcement learning based traffic signal control strategy based on traffic pressure and traffic queue length

Jiazhe Sun, Yuanchao Zhang, Zepu Wang
University of Pennsylvania

***Abstract*—Traffic signal control (TSC) is a prominent area of research within Intelligent Transportation Systems, as it can substantially enhance operational efficiency across society. Among various methods, reinforcement learning-based approaches have garnered significant attention from researchers due to their flexibility. Concurrently, numerous studies have confirmed that an appropriate representation of traffic state is crucial for algorithms to extract information from extensive road networks effectively. In this paper, we propose a novel reward representation based on traffic pressure and overall traffic queue length, which has demonstrated high performance on the real-world Hangzhou Dataset.**

***Index Terms*—Intelligent Transportation System, Traffic signal control, Reinforcement learning, Adv-MP-Co-Light, Algorithmic resilience, Traffic management efficiency**

I. Introduction

In today's world, traffic congestion is a significant problem that negatively affects economic growth, environmental sustainability, and general quality of life. In this regard, controlling traffic signal control (TSC) is essential to promoting effective transportation networks and reducing traffic [1]. Intelligent transportation systems (ITS) now include TSC as a crucial component that optimizes traffic flow and improves overall system performance. Traditional TSC methods usually depend on specific flow dynamics and presumptions of traffic conditions [2] or require significant professional assistance to handle traffic pattern complexities [3], [4]. However, incorporating Reinforcement Learning (RL) into TSC has become a viable option with the development of machine learning and deep learning methodologies. Without depending on static models, reinforcement learning (RL) provides a dynamic method that can adjust to the intricacies of traffic dynamics. As a result, it is possible to immediately learn the best control techniques from interactions with the traffic environment. Our project's general goal is to provide more adaptable and flexible solutions to the problems associated with contemporary traffic management.

Within a broad vehicle network for TSC, several controllers leverage vehicle flow data to devise optimal control strategies. This concept is facilitated through the sharing of information, where vehicle flow data is crucial for decision-making. Compared to traditional Traffic Signal Control (TSC) algorithms, which are in scenarios of relatively low rewards, Adv-Colight and Adv-MPLight offer significant advantages. They provide adaptability to dynamic traffic conditions by dynamically adjusting signal timings based on real-time data, optimizing traffic flow in varying scenarios. Unlike traditional methods, these algorithms can optimize for multiple performance metrics simultaneously, balancing objectives such as minimizing delays and reducing congestion. Additionally, Adv-Colight and Adv-MPLight demonstrate scalability and generalization capabilities, making them suitable for larger and more complex traffic. Therefore, such two algorithms improve the performance of RL-based Algorithms applied to TSC in urban environments [5]–[7].

Our proposed method innovatively merges the strengths of Adv-Colight and Adv-MPLight, to introduce a novel, more efficient reward mechanism called Adv-MP-Co-Light. By combining insights from both approaches, we aim to optimize traffic flow dynamics while minimizing congestion and travel delays. The key innovation lies in the development of a hybrid reward function that incentivizes efficient traffic operations and promotes fairness in resource allocation across intersections. Through our experiment, we demonstrate the superiority of our approach in improving traffic efficiency and mitigating congestion, thereby advancing the state-of-the-art in TSC optimization and offering valuable insights for real-world traffic management scenarios.

The paper is organized as follows. Section II discusses some related work; Sections III makes some definitions of TSC; Section IV explains the core models and algorithms applied. Section V introduces simulation datasets and analysis of the experimental results; Section VI concludes the paper and discusses future research.

II. Related Work

Related work in Traffic Signal Control (TSC) encompasses various approaches to optimize traffic flow and reduce congestion in urban areas. Traditional methods have primarily relied on fixed-time signal plans, where signal timings are predetermined and not adaptable to real-time traffic conditions. Evolutionary algorithms, such as genetic algorithms, have optimized signal timings based on historical traffic data. Reinforcement learning techniques, such as profound reinforcement learning, have gained traction due to their ability to learn optimal signal control policies through environmental interactions. In addition, traffic simulation models have been utilized to evaluate the effectiveness of different control strategies before

deployment in real-world scenarios. Recent research has also explored the integration of TSC with emerging technologies like connected and autonomous vehicles, aiming to enhance efficiency and safety on road networks further.

A. Traditional Approaches

In 1958, the FixedTime strategy was introduced as an innovative method, initially applying static cycles to the management of traffic signals [1]. Subsequent advancements such as SCOOT [2], SCATS [8], and MP control [3] sought to refine the synchronization of traffic signals with actual traffic conditions, achieving the broad implementation. Initially designed for network packet scheduling [9], MP control was adapted to traffic signal control (TSC) by Varaiya [3], with additional improvements [10]–[12] proving its effectiveness in simulations.

Although MP has achieved some successes, its inflexibility and sensitivity to hyper-parameters, like phase duration and cycle length [11], [12], hinder its flexibility. This limitation becomes apparent when the approach does not consider ongoing traffic movements, but rather concentrates on vehicles in queues. This oversight is what SOTL [4], [13] aims to address by modifying signal timings based on current traffic conditions.

B. RL-Based Methods

The rise of RL in TSC can be attributed to advanced neural architectures and well-planned state and reward modeling. Developments such as FRAP [14], focusing on phase relationships for uneven flows, and Adv-Colight [7], which utilizes GAT for cooperation between intersections, highlight the effectiveness of this method. Both PressLight [15] and Adv-MPLight [5] emphasize the critical role of incorporating traffic pressure concepts into RL, markedly improving model performance. Efficient-CoLight [6] enhances this method by introducing 'efficient pressure' in state representation, establishing new standards in TSC performance. Additionally, other MBRL models that employ predictive system models for immediate rewards or transitions to generate synthetic samples for RL are also in use [16]. These models, which are more data-efficient and adaptable to partially observable environments compared to model-free RL, are mentioned [17], [18].

III. Preliminaries

A. Definitions for Traffic Signal Control

In this section, we first make some rigorous definitions of traffic signal control and then formulate our task under the condition of limited data. For a complete understanding of TSC, it is essential to define several foundational concepts, based on the established literature [5], [6].

Definition 1 (Signal Phase): Traffic movements were permitted into a coherent set by a Signal Phase group, which is represented as $s = \{set(l, m)\}$, where s belongs to the set of phases at intersection j, S_j. This group makes it easier to comprehend how traffic lights at crossings operate.

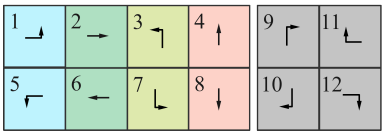

Fig. 1. Illustration of the 12 possible traffic movements at an intersection. In common cases, the last four cases, in which vehicles are turning right, are not under the traffic signal control [19]

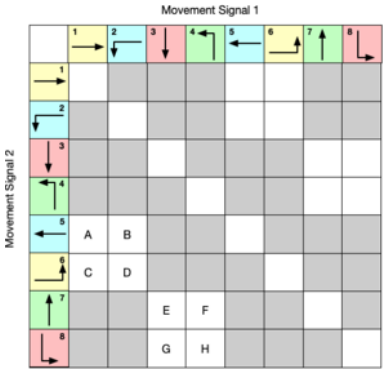

Fig. 2. The conflict matrix for the movement signal. The white cells indicate non-conflicting phases while the grey cells stand for conflicting phases. The letter from "A" to "H" stands for 8 frequently used phases as illustrated in Fig. 3 [19]

Definition 2 (Signal Phase Sequence): A set of phases and the order in which they change are defined by a phase sequence. To prevent misunderstanding among drivers, we usually adjust the signal phase sequence to $Red, Yellow, Green$ to correspond with their daily routine.

Definition 3 (Signal Plan & Circle-based Signal Plan): A signal design for a single junction consists of a set of phases and their matching intervals. A cycle-based signal plan is one type of signal plan in which the phases function in cyclic order. The cyclic-based signal scheme is often observed at crossings where the $FixedTimePolicy$ is implemented [8].

Definition 4 (Efficient Pressure): Efficient Pressure is defined as the average discrepancy in queue lengths between upstream and downstream lanes for each traffic movement, calculated as,

$$ep(l, m) = \frac{1}{M}\sum_{i=1}^{M} ql(l'_i) - \frac{1}{N}\sum_{j=1}^{N} ql(m'_j), \tag{1}$$

where $ql(l')$ and $ql(m')$ quantify the queue lengths of lanes l' and m', respectively. This metric serves as a crucial indicator of congestion levels and traffic flow efficiency.

Definition 5 (Phase Pressure): The Phase Pressure represents the combined effective pressures exerted by all individual traffic movements within a phase, which is indicated as,

$$p(s) = \sum_{(l,m)\in s} ep(l, m), \tag{2}$$

This allows for a thorough assessment of the amount of traffic being managed by each signal phase.

Definition 6 (Intersection Pressure): Intersection Pressure measures intersection congestion by the net difference in queue lengths across all incoming and departing lanes, codified as

$$P_j = \sum_{l' \in L_{in}^i} ql(l') - \sum_{m' \in L_{out}^j} ql(m'). \tag{3}$$

This description emphasizes key traffic concentration sites and potential bottlenecks.

Definition 7 (Action Duration): The phrase "Action Duration" refers to the duration of time that a traffic signal phase remains active, represented by the symbol t_{dr}. The length of each phase can be extended by multiples of t_{dr}, allowing for flexibility in making adjustments to signal timing.

IV. Methods

Although traditional Traffic Signal Control (TSC) methods with Reinforcement Learning (RL) algorithms often fail to ensure satisfactory rewards, the Adv-Colight and Adv-MPLight Algorithms offer comparatively higher rewards.

Several steps are crucial to take advantage of each algorithm's strengths and achieve synergistic benefits: (1) understanding such two algorithms' objectives, identifying complementary features, (2) integrating such two algorithms strategically, merging data, combining rewards through weighted sums, (3) implementing adaptive control for dynamic adjustments, (4) evaluating and validating performance, and establishing a feedback loop for continuous enhancement.

By integrating the cooperative decision-making prowess of Adv-Colight with the optimized traffic signal timing of Adv-MPLight, our hybrid Adv-MP-Co-Light Algorithm maximizes traffic flow efficiency and incentives across diverse traffic scenarios, ultimately enhancing the transportation system's overall performance.

Three Algorithms' Action: At time t, each agent chooses a phase $\hat{s}$ as its action a_t, and the traffic signal will be set to phase $\hat{s}$.

Adv-Colight Algorithm's Rewards: CoLight generates rewards through cooperative decision-making among vehicles, aiming to improve overall traffic flow by optimizing interactions at intersections. These rewards are typically based on reduced congestion, decreased travel times, and smoother traffic flow achieved through vehicle coordination.

$$r_i = -\sum_{l' \in \text{Lin}} q(l'), \tag{4}$$

Adv-MPLight Algorithm's Rewards: On the other hand, Adv-MPLight generates rewards by optimizing traffic signal timing based on real-time traffic conditions, aiming to minimize delays and maximize throughput at intersections.

$$r_i = -|Pi| \tag{5}$$

Adv-MP-Co-Light Algorithm's Hybrid Rewards: The hybrid reward is achieved by integrating the rewards generated by both algorithms, often through a weighted sum approach in which the importance of the reward for each algorithm is adjusted based on factors such as traffic conditions, road network topology, and user preferences. This hybrid reward reflects the overall improvement in traffic flow and efficiency achieved by integrating Adv-Colight and Adv-MPLight, providing a comprehensive system performance measure.

$$r_i = -|Pi| - \sum_{l' \in \text{Lin}} q(l') \tag{6}$$

Our Contributive Adv-MP-Co-Light Algorithm is updated by the Bellman Equation:

$$Q(s_t, a_t) = R(s_{t+1}, a_t) + \gamma \max Q(s_{t+1}, a_{t+1}) \tag{7}$$

V. Experimental Results

A. Experiment Platform

Our experiments are conducted using the CityFlow simulator [20], which is specifically designed for large-scale traffic signal control (TSC) simulations. CityFlow provides a realistic simulation environment that includes signal transition features such as a three-second yellow light and a two-second all-red period to ensure safe signal changes. Experiments are executed within a Docker Desktop Image.

B. Simulation Datasets and Parameters

For our TSC experiments, we use parameters from the Efficient-MP strategy [6], setting four phases with a minimum of 15 seconds per phase. These parameters are chosen to reflect typical urban traffic systems, enhancing the simulation's real-world applicability.

Road Network Data: Provides the structural layout of the traffic system, detailing road links, traffic movements, and signal phases.

Traffic Flow Data: Consists of vehicle trajectories, described by tuples (t, u), where t is the time and u is the vehicle's route through the network.

HangZhou Datasets: There are 16 (4 × 4) junctions in the road network. Every intersection has two 800-meter-long (East-West) and two 600-meter-long (South-North) road segments, making it a four-way intersection. This traffic road network dataset contains two traffic flow datasets at different time[1].

Fig. 3. Illustration of Hangzhou Dataset

[1]https://traffic-signal-control.github.io/

C. Evaluation Benchmarks

In our research, we evaluate the selected strategies against established benchmarks that span both conventional traffic control techniques and modern RL-based approaches. Uniform hyperparameters ensure consistency across RL models during training. We simulate 60-minute episodes, with results averaged over the final ten tests, and report means from three distinct experiments.

Traditional Methods:

- **FixedTime** [1]: a policy gives a fixed cycle length with a predefined phase split among all the phases.
- **MaxPressure** [3]: the max pressure control selects the phase that has the maximum pressure.
- **Efficient-MP** [21]: it selects the phase with the maximum efficient pressure. It is a SOTA method that has superior performance than MPLight [5].
- **Adv-MP** [22]: it is based on MaxPressure and SOTL algorithm to reach the optimal MP strategy.

Advanced RL Methods:

- **Adv-MPLight** [5]: using FRAP [14] as the base model, and introduces pressure into the state and reward design. By experiment, Adv-MPLight can realize city-level traffic signal control.
- **Adv-CoLight** [7]: using graph attention network to realize intersection cooperation. By experiment, Adv-CoLight can realize large-scale traffic signal control.
- **Efficient-MPLight** [21]: FRAP [14] based model, using current phase and efficient traffic movement pressure as observation, intersection pressure as reward.
- **AttendLight** [23]: using attention mechanism to construct phase feature and predict phase transition probability.
- **PRGLight** [24]: using graph neural network to predict traffic state and adjust the phase duration according to the currently observed traffic state and predicted state.

Our Contributive Method:

- **Adv-MP-Co-Light**: We self-design a niche algorithm that combines Adv-MPLight and Adv-Colight algorithms to improve Rewards significantly.

D. Results and Analysis

Table I reports the experimental results in the HangZhou real-world data sets with respect to the average travel time. Due to the time limit, we cited some results from other authors [19].

Our Adv-MP-Co-Light method outperforms all previous methods. we can know that generally, RL-based models have a better performance compared to traditional methods, indicating that RL methods have become a dominate mainstream for TSC research.

For RL methods, we can tell that We can see from the results that Adv-based model consistantly have a better performance than other RL methods, indicating that introducing noval advanced traffic state(ATS) can really better represent the traffic information in the general road network, enhancing the model performance .Integrating AST with the RL-based approaches brings excellent improvements. The performance of Adv-CoLight is improved by 13.54% and 6.01% from the previous SOTA methods, Efficient-MP and Efficient-CoLight, respectively [19].

Specially, our Adv-MP-Co-Light has a better performance than other Adv-based methods. Although the two rewards representation, to some extent, both represents the status of ATS and are related to each other, making them a hybrid reward can improve the performance of final model.

TABLE I
PERFORMANCE (THE AVERAGE TRAVEL TIME IN SECONDS) COMPARISON OF DIFFERENT METHODS EVALUATED ON HANGZHOU REAL-WORLD DATASET (THE SMALLER THE BETTER).

Method	Hangzhou1	Hangzhou2
FixedTime	495.57	406.65
MaxPressure	288.54	348.98
Efficient-MP	284.44	327.62
MPLight	314.60	357.61
CoLight	297.02	347.27
AttendLight	293.89	345.72
PRGLight	301.06	369.98
Efficient-MPLight	284.49	321.08
Efficient-CoLight	282.07	324.27
Adv-MP	280.62	327.89
Adv-MPLight	273.26	312.68
Adv-CoLight	270.45	310.74
Adv-MP-Co-Light	267.79	308.79

Algorithm 1 Adv-MP-Co-Light

Parameter: Current phase time t, minimum action duration t_{duration}

```
for each time-step do
  t = t + 1
  if t = t_duration then
    Get Advanced Traffic State (ATS) by
        ATS(l, m) = {(e(l, m); r(l, m))}
    for each intersection
    Set the phase by combining Adv-CoLight and Adv-MPLight using an RL model
    t = 0
  end if
end for
```

However, a potential limitation of Adv-MP-Co-Light is that, compared to Adv-Adv-MPLight and Adv-CoLight, it is much harder to train. It requires more epochs to get converge, a more careful selection of hyperparameters, and it is highly sensitive to the dataset. Therefore, we do not have enough time to train Adv-MP-Co-Light on other dataset.

VI. Conclusion and Future Work

Our study investigates the efficacy of conventional traffic signal control algorithms without information. The initial report examines essential aspects of intelligent traffic signal control and establishes the research aims and inquiries. Following this, the report provides an overview and assessment of current intelligent traffic signal control techniques, with an emphasis on investigating and appraising methods based on reinforcement learning. This document introduces a technique for forecasting traffic flow at unfamiliar intersections by utilizing data from familiar intersections to aid agents in managing traffic signals. Overall, the Adv-MP-Co-Light Algorithm continues to outperform conventional approaches.

To broaden our methodology, additional datasets that cover an extensive transportation network could be utilized. The outcomes from a small dataset might not reflect the true efficacy of the algorithm when applied to a broader traffic system. Moreover, we could incorporate considerations of more intricate city dynamics, like patterns of human movement. It is possible that such crossroads will require consideration of additional phases.

References

[1] P. Koonce and L. Rodegerdts, "Traffic signal timing manual." United States. Federal Highway Administration, Tech. Rep., 2008.

[2] P. B. Hunt, D. I. Robertson, R. D. Bretherton, and M. Royle, "The scoot on-line traffic signal optimisation technique," *Traffic engineering and control*, vol. 23, 1982.

[3] P. Varaiya, "Max pressure control of a network of signalized intersections," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 177–195, 2013.

[4] S.-B. Cools, C. Gershenson, and B. D'Hooghe, *Self-Organizing Traffic Lights: A Realistic Simulation*. Springer London, Nov. 2007, p. 41–50.

[5] C. Chen, H. Wei, N. Xu, G. Zheng, M. Yang, Y. Xiong, K. Xu, and Z. Li, "Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control," in *Proc. AAAI*, vol. 34, no. 04, 2020, pp. 3414–3421.

[6] J. Wu, Y. Wang, B. Du, Q. Wu, Y. Zhai, J. Shen, L. Zhou, C. Cai, W. Wei, and Q. Zhou, "The bounds of improvements toward real-time forecast of multi-scenario train delays," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 2445–2456, 2021.

[7] H. Wei, N. Xu, H. Zhang, G. Zheng, X. Zang, C. Chen, W. Zhang, Y. Zhu, K. Xu, and Z. Li, "Colight: Learning network-level cooperation for traffic signal control," in *Proc. ACM CIKM*, 2019, pp. 1913–1922.

[8] L. PR, "Scats: A traffic responsive method of controlling urban traffic control/pr lowrie," *Roads and Traffic Authority*, 1992.

[9] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," in *Proc. IEEE CDC*, 1990, pp. 2130–2132.

[10] J. Gregoire, E. Frazzoli, A. de La Fortelle, and T. Wongpiromsarn, "Back-pressure traffic signal control with unknown routing rates," 2014.

[11] J. Gregoire, X. Qian, E. Frazzoli, A. de La Fortelle, and T. Wongpiromsarn, "Capacity-aware back-pressure traffic signal control," 2014.

[12] T. Wongpiromsarn, T. Uthaicharoenpong, Y. Wang, E. Frazzoli, and D. Wang, "Distributed traffic signal control for maximum network throughput," in *Proc. IEEE ITSC*, 2012, pp. 588–595.

[13] C. Gershenson, "Self-organizing traffic lights," 2005.

[14] G. Zheng, Y. Xiong, X. Zang, J. Feng, H. Wei, H. Zhang, Y. Li, K. Xu, and Z. Li, "Learning phase competition for traffic signal control," in *Proc. ACM CIKM*, 2019, pp. 1963–1972.

[15] H. Wei, C. Chen, G. Zheng, K. Wu, V. Gayah, K. Xu, and Z. Li, "Presslight: Learning max pressure control to coordinate traffic signals in arterial network," in *Proc. ACM SIGKDD*, 2019, pp. 1290–1298.

[16] F.-M. Luo, T. Xu, H. Lai, X.-H. Chen, W. Zhang, and Y. Yu, "A survey on model-based reinforcement learning," 2022.

[17] Y. J. Ma, A. Shen, O. Bastani, and J. Dinesh, "Conservative and adaptive penalty for model-based safe reinforcement learning," in *Proc. AAAI*, vol. 36, no. 5, 2022, pp. 5404–5412.

[18] E. Vértes and M. Sahani, "A neurally plausible model learns successor representations in partially observable environments," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[19] L. Zhang, Q. Wu, J. Shen, L. Lü, B. Du, and J. Wu, "Expression might be enough: representing pressure and demand for reinforcement learning based traffic signal control," in *Proc. ICML*. PMLR, 2022, pp. 26 645–26 654.

[20] H. Zhang, S. Feng, C. Liu, Y. Ding, Y. Zhu, Z. Zhou, W. Zhang, Y. Yu, H. Jin, and Z. Li, "Cityflow: A multi-agent reinforcement learning environment for large scale city traffic scenario," in *The world wide web conference*, 2019, pp. 3620–3624.

[21] Q. Wu, L. Zhang, J. Shen, L. Lü, B. Du, and J. Wu, "Efficient pressure: Improving efficiency for signalized intersections," *arXiv preprint arXiv:2112.02336*, 2021.

[22] H. Zhang, H. Chen, D. Boning, and C.-J. Hsieh, "Robust reinforcement learning on state observations with learned optimal adversary," *arXiv preprint arXiv:2101.08452*, 2021.

[23] A. Oroojlooy, M. Nazari, D. Hajinezhad, and J. Silva, "Attendlight: Universal attention-based reinforcement learning model for traffic signal control," *Advances in Neural Information Processing Systems*, vol. 33, pp. 4079–4090, 2020.

[24] C. Zhao, X. Hu, and G. Wang, "PRGLight: A novel traffic light control framework with pressure-based-reinforcement learning and graph neural network," in *IJCAI 2021 Reinforcement Learning for Intelligent Transportation Systems (RL4ITS) Workshop*, Virtual, Aug. 2021.