Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Application and development of reinforcement learning in traffic signal control

YuXiang Hong
College of Geoinformatics, Zhejiang University of Technology, Hangzhou, Zhejiang, China
Corresponding Author: YuXiang Hong, **E-mail**: 302025344013@zjut.edu.cn

| ABSTRACT

With the acceleration of urbanization, the traditional timing traffic signal control strategy is difficult to deal with the complex and variable nonlinear traffic flow. Reinforcement learning (RL) shows great potential in alleviating traffic congestion with its strong adaptive perception and optimization ability. However, reinforcement learning still faces many challenges in practical applications. This paper systematically reviews the latest research progress of reinforcement learning based traffic signal control (RL-TSC), focusing on control efficiency optimization, multi-objective constraints and overall architecture adjustment. Research shows that RL-TSC still has some problems, such as insufficient robustness to real-world noise, a lack of characterization of non-motor vehicle behavior, and a computing bottleneck of edge devices. The significance of this paper is to clearly point out the limitations and technical gaps of the current application in the rl-tsc field, and propose that the future research direction should be towards the interpretable RL, LLM hybrid architecture and global active scheduling, so as to provide theoretical guidance and forward-looking reference for the construction of a smart urban transportation system that takes into account safety, efficiency and sustainable development.

| KEYWORDS

Reinforcement learning, Traffic signal control, Intelligent transportation system

| ARTICLE INFORMATION

ACCEPTED: 01 May 2026 **PUBLISHED:** 17 June 2026 **DOI:** 10.32996/jcsts.2026.8.8.4

1. Introduction

With the growth of the global population and the expansion of urban scale, the density of urban traffic flow has increased significantly, and the nonlinearity and uncertainty of the system are increasing. The traditional timing traffic signal control scheme can not effectively deal with complex traffic flow because of its fixed phase configuration, which leads to difficulty in alleviating traffic congestion. The traffic signal control method based on reinforcement learning can perceive the traffic state, adaptively optimize the phase configuration, and effectively alleviate traffic congestion.

Most of the early studies in this field are about the simple situation of a single intersection, using tabular Q - learning and other basic algorithms, with the goal of maximizing throughput or minimizing waiting time. With the growth of data volume, the research gradually turns to deep reinforcement learning and multi-intersection coordination [1][2].

Most of the early models in this field were trained in the ideal simulation environment. However, in the real environment, there are not only many interference factors, but also obvious communication delays. Therefore, there is a significant sim2real gap. Some researchers have proposed a rigorous prediction mechanism to reduce the impact of network delay [3].

Early research on the single pursuit of traffic efficiency has also been unable to meet the increasingly stringent road safety standards and green low-carbon development goals. How to incorporate the safety of vulnerable road users and urban carbon emissions and other constraints into the optimization goals has become a new challenge. At the same time, in order to solve the above problems, the underlying level of the system is also facing innovation challenges. The pure reinforcement learning

architecture is not only difficult to train, but also has defects such as poor interpretability and difficulty dealing with special situations. Some researchers have proposed to integrate new architectures such as large language models, into the innovation architecture of agents [4].

To sum up, the actual implementation of reinforcement learning technology in the field of traffic signal control still faces challenges in terms of control efficiency, multi-objective constraints, and underlying architecture. Therefore, this paper reviews the latest breakthroughs in the field of traffic signal control through reinforcement learning, points out the potential research gaps, and predicts the future research directions.

2. Control efficiency optimization

The main purpose of traffic signal control using reinforcement learning model is to alleviate urban congestion and improve traffic efficiency, so the control efficiency of the optimization model is one of the main directions of current research. In order to further improve the control efficiency of the model, the current research is mainly divided into two directions: one is to improve the reward function and state representation to improve the theoretical efficiency; The second is to bridge the gap between the simulation environment and the display environment and improve the robustness of the model in the real environment.

In terms of reward mechanism and representation optimization. Ault et al. Used cumulative queue length, total waiting time or traffic pressure as punishment items to design rewards, and made a rigorous comparison between the performance of the Distributed Deep Q-Network (IDQN) and Proximal Policy Optimization (IPPO) in reducing average vehicle delay [5]. Boukerche et al. introduced the maximum pressure theory, defined the intersection pressure as $P_t = \sum_{l,m\in L_i} P_t(l,m)$, and then let the agent learn to minimize the total pressure, thus effectively solving the problem of frequent signal phase switching of local high-pressure lanes in the traditional RL [3].

Most of the methods discussed in the existing literature use idealized assumptions and take data from the benchmark environment. In common benchmarks such as the reso test platform, it is generally assumed that the intelligent physical ability can obtain the queue length, speed and waiting time of all vehicles within 200 meters without delay [3]. Although the assumption that the agent is omniscient is conducive to the fair and objective comparison of different algorithms, there is no doubt that this assumption deviates from reality.

At present, the problem of communication delay has been discussed in several literatures. Boukerche et al. Clearly pointed out that v2x communication in reality has transmission delay that can not be ignored, so they redesigned the state representation (link partition matrix), introduced the forward-looking traffic state prediction module based on physical kinematics, and estimated the current real position of the vehicle with delayed data to eliminate the interference of delay on the RL strategy [3].

3. Multiobjective constraint

The traditional signal control method and even the early reinforcement learning control method usually take maximizing intersection traffic efficiency and minimizing vehicle delay as the single optimization objective. However, this single goal has been unable to meet the requirements of modern cities for traffic safety and sustainable development. Therefore, the current research has gradually turned to multi-objective optimization, including the safety and ecological benefits of vulnerable road users [4].

For traffic safety issues. Because the pure pursuit of vehicle traffic efficiency is often at the expense of pedestrian safety, Ren et al. developed a multi-objective reward design method, introduced Post Encroachment Time (PET) into the reward function, and gave a maximum traffic conflict penalty (r_{CTC}) when the pet is less than 3 seconds [1]. In addition, it gives appropriate weight (ω_{CTD}) to vehicle delay time and pedestrian waiting time, so it can reduce pedestrian vehicle conflict (the number of conflicts decreases by 41%) without reducing the overall throughput of the intersection, and truly achieve the balance between safety and efficiency.

On the other hand, driven by the concept of sustainable development, the existing literature has begun to systematically discuss the ecological benefits. Literature [5] and literature [6] demonstrate that it is very effective to reduce urban carbon emissions with the reinforcement learning signal control enabled by big data. By reducing the frequent start and stop and idle time of vehicles, the RL agent can smooth the traffic flow and directly reduce the fuel consumption level [7]. Through the joint optimization of signal control and speed guidance, it can not only reduce the idle time, but also achieve energy conservation and emission reduction at the vehicle level, further strengthening the application value of RL in ecological transportation [8].

4. Overall structure adjustment

The traditional pure reinforcement learning architecture is not only difficult to train and poor in interpretability, but also difficult to deal with special situations. In addition, it can be seen from the experiments on the road network of real cities (Cologne and Ingolstadt) in literature [3]: because the sharing of complex collaborative parameters in irregular and asymmetric real road network easily causes training divergence at heterogeneous intersections, the completely decentralized independent DQN (IDQN) model in such road network is better than various well-designed collaborative architectures (FMA2C or MPLight). It can be seen that the adjustment of the underlying model is the only way to improve the performance of the model.

At present, several studies have reasonably integrated the connected vehicle (CV) data and the semantic representation of large models. The vehicle-level trajectory data provided by CV is used to construct an ultra-high resolution gridded spatial matrix (12 × 20 characteristic matrix), so the position of vehicles and pedestrians can be estimated very accurately [1]. Some researchers have added the classic traffic flow theory (Max-Pressure) to the RL architecture in the form of hard-coding rules. Their method simultaneously uses the local pressure and the "input pressure" of adjacent intersections to shorten the unnecessary yellow light time [3]. Ren et al. Made a detailed analysis of the problem of mixed traffic of people and vehicles. They designed a two-step deep reinforcement learning (TSDRL - TSC) framework: first, the dynamic action set was screened, and then the d3qn (Dueling Double Deep Q - Network) architecture combining a convolutional neural network and a short - and long-term memory network was used to find the optimal signal phase configuration [1]. In order to solve the problem of information sharing in collaboration, some researchers use an attention mechanism (such as CoLight) to capture the spatial correlation between intersections, so as to improve the overall traffic efficiency [9].

The LLMLight framework proposed by Lai turns the traffic state into a prompt word, giving agents the ability of zero sample traffic control and common-sense reasoning of large language models [4]. In addition, the latest architecture adjustment is not limited to transforming traffic status into cue words. More research has proposed a deep hybrid architecture of large model and reinforcement learning, such as using LLM's common sense reasoning to adjust RL's multi-objective reward weight in real time in a dynamic environment to make up for the decision-making weakness of pure reinforcement learning in a small sample environment [10].

5. Existing limitations

Although the application of the reinforcement learning method has achieved good results in the field of TSC, there are still the following research gaps to be solved.

At present, most RL algorithms (including IDQN, IPPO, etc.) are trained and tested in Sumo and other simulation environments. Therefore, it is necessary to propose a method to ensure the robustness of the model when the model is faced with data loss caused by sensor noise and bad weather in the real world.

Some studies have shown that the introduction of pedestrian protection can improve driving safety, but the current model is not able to comprehensively describe the changeable interpenetrating behavior of non-motor vehicles, and does not have a collaborative reward mechanism that can be applied to all vulnerable groups (such as non-motor vehicles) [1].

Models such as MP-CTSC (high-dimensional state prediction) and LLMLight(using large language model as controller) require strong computational power support and long reasoning time, which cannot meet the limited computational power of roadside edge computing equipment. Therefore, a model lightweight method without reducing accuracy is needed [1][4].

6. Conclusion

Starting from the gap of the existing literature and the problems discussed, this paper makes an appropriate outlook on the future development direction of RL - TSC.

First, because traffic control is directly related to public safety, the deep reinforcement learning of pure black box should turn to "interpretable RL", so that traffic engineers can understand the green light logic determined by AI.

Secondly, because the traditional RL needs to learn tens of thousands of rounds to get good results, and the large model has excellent decision-making ability with few samples because it has learned a lot of traffic related knowledge in advance, the mixture of LLM and traditional RL is expected to become the mainstream, that is, using LLM to do macro high-level intention reasoning and crisis handling, and using lightweight RL to do micro second level signal switching.

Finally, due to the full development of big data and Internet of vehicles technology, it is expected to change TSC from the existing "passive response" to "active scheduling", so as to realize the leap from "single point control" to "global coordination". RL agent can accurately calculate the arrival time of each vehicle and dynamically adjust the signal light, so that the vehicles can form a group to pass the intersection at perfect speed, completely eliminate the green light loss of 'stop and go', and promote the overall traffic efficiency and capacity of urban roads to the extreme, which is more conducive to taking into account the goals of safety, efficiency and urban carbon neutralization.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Ren, A. D., Zhang, B. G., Chang, C. F., & Huang, D. H. (2025). Two-step deep reinforcement learning for traffic signal control to improve pedestrian safety using connected vehicle data. Accident Analysis and Prevention, 222, 108161.

[2] Machine Learning. (2023). Algorithmic advancements in reinforcement learning for traffic systems. s10994-023-06412-y, 1-20.

[3] Boukerche, A., Zhong, D., & Sun, P. (2022). A Novel Reinforcement Learning-Based Cooperative Traffic Signal System Through Max-Pressure Control. IEEE Transactions on Vehicular Technology, 71(2), 1187-1198.

[4] Lai S, Xu Z, Zhang W, et al. LLMLight: Large language models as traffic signal control agents[C]//Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1. 2025: 2335-2346.

[5] Ault J, Sharon G. Reinforcement learning benchmarks for traffic signal control[C]//Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1). 2021.

[6] Wu K, Ding J, Lin J, et al. Big-data empowered traffic signal control could reduce urban carbon emission[J]. Nature Communications, 2025, 16(1): 2013.

[7] Zhang S, Li S. The Impact of Traffic Signal Control on Emissions [J]. Sustainability, 2023, 15: 3479.

[8] Y. Bentaleb, H. Asaidi, M. Bellouki and N. E. Allali, "Eco-Intelligent Traffic Signal Control via Deep Reinforcement Learning," 2025 International Conference on Circuit, Systems and Communication (ICCSC), Fez, Morocco, 2025, pp. 1-6.

[9] Wei H, Xu N, Zhang H, et al. Colight: Learning network-level cooperation for traffic signal control[C]//Proceedings of the 28th ACM international conference on information and knowledge management. 2019: 1913-1922.

[10] S. Choi and Y. Lim, "Optimizing Traffic Signal Control Using LLM-Driven Reward Weight Adjustment in Reinforcement Learning," Journal of Information Processing Systems, vol. 21, no. 1, pp. 43–51, Mar. 2025, doi: 10.3745/JIPS.04.0334.