Optimizing Traffic Signal Control Using Deep Reinforcement Learning: A Case Study of Hangzhou

Anran Li*

Urban Transportation College, Beijing University of Technology, Beijing, 100124, China

*Lianran0511@emails.bjut.edu.cn

Abstract. Urban development faces the problem of traffic congestion as one of its major challenges. The conventional approaches to signal control work well in the context of optimization at a single point but do not scale well to large road networks or more complicated environments in which dynamics occur. Recently, Deep Reinforcement Learning (DRL) has proven useful in optimizing traffic signals where it can optimize both adaptively and globally due to the interactive learning process. The paper discusses the evolution of DRL in traffic signal control research, starting with single intersections and extending to multi-agent co-operation, and provides an insight into the Hangzhou intelligent traffic case study. It is found that this method can be highly efficient in terms of reducing the average delay and increasing the road network efficiency, however, there are still practical limitations, including data noise, multi-source uncertainty, and lack of robustness. The further research directions will focus on integrating multimodal data, improving robust modeling, and so on to make intelligent transportation be applied to real road instead of remaining in the theory.

Keywords: DRL, Traffic Control, Single Intersection Optimization, Traffic Jam, Multimodal Data Fusion.

1 Introduction

Traffic congestion is a popular issue of global urbanization and has negative effects on urban efficiency, energy use, environment quality, and living conditions. The congestion delay index at peak hours in the Chinese megacities such as Beijing, Shanghai and Hangzhou often surpass 2.0, limiting the prospects of China sustainable development. Thus, enhancing traffic efficiency has become one of the focal points in smart city development [1].

The optimization of the traffic signals is now the cheapest way to reduce the congestion. Nevertheless, conventional approaches are clearly limited. Fixed-time control is not flexible, whereas actuated and adaptive controls cannot be used to synchronize large-scale road networks [2]. These methods cannot successfully predict sudden accidents, non-linear traffic changes and unpredictable driver behavior because they use linear predictions and rule-based models. Their performance deteriorates in the face of the intense traffic flows of the megacities.

N. Malik et al. (eds.), *Proceedings of the 6th International Conference on Public Management and Intelligent Society (PMIS 2026)*, Advances in Engineering Research 301,
https://doi.org/10.2991/978-94-6239-721-7_2

The relatively recent Deep Reinforcement Learning (DRL) introduced a novel perspective, i.e., modeling signal control as a Markov Decision Process and dynamically learning policies by means of state-action-reward iterative interactions that can be used to adaptively control in more complex environments. In comparison to the traditional approaches, DRL has three main benefits: it is able to handle high-dimensional complex traffic conditions, multi-agent collaborative learning to coordinate globally across multi-intersections, and its ability to optimize itself continuously in simulation conditions. According to research, DQN and Actor-Critic solutions can reduce the time of vehicles and enhance the efficiency of the network [3].

Being a precursor of the smart transportation city, Hangzhou also has open traffic data platforms and successful experience in applying AI signals optimization, which can be considered as the perfect experimental setting to implement DRL applications with both local and international demonstration capability [4].

Consequently, this paper discusses the development path of DRL, its benefits, practical issues as well as opportunities in application to traffic signals control by reviewing literature and conducting a Hangzhou case study.

2 The Rise of Deep Reinforcement Learning and Its Application Value

The big data and machine learning have been used as the main drivers of intelligent transportation development in the early 21st century. Reinforcement Learning (RL) maximizes strategies by means of state-action-reward interactions, which is in accordance with the dynamic requirements of traffic signals control. Since 2013, DRL has made advancements with DQN and enhanced algorithms (Double DQN, Actor-Critic, Proximal Policy Optimization), which are capable of controlling high-dimensional complex environments and are at the centre of signal control research [2]. These techniques take into account queue size, waiting time and traffic speed rate, and are extended to optimization of networks on a regional scale with multi-agent models [5].

Experiments have shown that DRL-based models can greatly improve vehicle delay and intersection throughput compared to fixed control. In 2025, scientists proved that PPO-based DRL was highly effective in minimizing cumulative waiting time in small-scale simulations [6]. Another individual suggested a multi-agent system based on a graph neural network which outperforms the conventional system with regard to its multi-intersection cooperation [7]. Such developments make DRL a subject of current research interests in the field of signal control.

3 Model and Research Progress Review

The history of DRL research in traffic signal control can be described as the development of single intersection optimization into multi-intersection coordination, uncertainty treatment, and practical implementation. Many researchers have investigated various conditions of traffic, ranging over the development of algorithms and their application in practice. As a smart pilot city in transport, Hangzhou traffic data serves as a good research platform based on the real time data gathering and model creation by traffic management agencies and research centers.

The data are road network structure descriptions, e.g., intersection geometry, lane configurations and signal phase settings, in addition to dynamic traffic data, i.e., the sequence of arrivals of vehicles and simulations of behaviors. Their properties show that they are highly compatible with DRL requirements, allowing interaction to perceive the state, e.g., queue length and waiting time, and accumulate experience to improve decision making. The important peak-hour properties can challenge models to be adaptable to changes in traffic, whereas diverse sources of uncertainty are good to train on how to work with real-world complexities. Moreover, this data enables extrapolation of single-point to multi-point networks, which confirms the validity of multi-agent methods.

Such attributes are highly similar to those that the city of Hangzhou requires in its management in reality, including synchronizing the traffic lights at nearby intersections when the traffic is at its heaviest to avoid congestion spread, which can be used as an experimental value in the development of evacuation strategy. The data are useful in model training to test the possible use of DRL in dynamic traffic pattern management and multi-intersections coordination situations [4].

Nevertheless, there are also constraints. The data scale is generally restricted to local situations, it does not incorporate multimodal information like effects of weather or events, and it is unable to represent a day-long pattern of traffic. Multi-source data fusion needs to be extended in future studies to increase the level of richness of models and their practical applicability so that finally there will be no gap between theoretical studies and actual application in intelligent transportation systems.

3.1 Single Intersection Optimization: The Starting Point of Deep Reinforcement Learning

The study of DRL in the sphere of traffic signal control started with single intersection cases. The primary purpose of this phase was to determine whether the approach could be extended to approaches that are better than fixed timing or actuated control when used in small-scale settings.

The first systematic review of how DQN has been implemented to optimize the traffic signal of a single intersection was performed by Rasheed et al. in 2020. The findings indicated that the average wait time per vehicle was much lower than with conventional fixed control lights. Thereafter, other scientists came up with another signal timing plan using Deep Reinforcement Learning with a delay reduction rate of more than 20 percent better than the traditional ones on the SUMO simulation system [8]. The Hangzhou dataset offered a realistic environment in which these experiments were made possible. The model of DQN that learned on this data was able to acquire the features of the morning rush hour vehicle flow and learn how to switch between the stages when the queue length hit some threshold, so as to prevent the build-up of excess waiting times. This finding suggests that DRL models that are trained on real data have the same level of flexibility and usefulness when applied to single intersections.

The single intersection model has a number of drawbacks such as lack of consideration of reciprocal effects of neighboring intersections. The cars travel through a particular intersection point but queue at the next, and there would be no overall delay.

3.2 Evolution of Multi-Intersection Cooperative Control Algorithms

With the growing tendency of urban traffic congestion to become networked and regionalized, Deep Reinforcement Learning models that are based on optimizing a single intersection have difficulty alleviating the overall deterioration of traffic due to upstream spillback. Therefore, studies on traffic signal control have been changed to consider multi-agent systems instead of single-agent systems using Multi-Agent Reinforcement Learning (MARL). Research conducted in the last five years indicates that effective communication and coordination between various intersections within high-dimensional traffic networks are the main reasons behind the technical iteration in the field.

Paradigm Shift from Independent Learning to Multi-Agent Cooperation. The initial multi-intersection signal control schemes tried to utilize the Independent Q-Learning (IQL) approach, whereby every intersection was considered as an independent entity and would maximize its own reward depending on its local state. Nevertheless, as is shown by numerous reviews and comparative experiments, there is severe non-stationarity in IQL in multi-agent environments. As the policies of the neighboring intersections keep changing, the environment of a particular intersection becomes dynamic, which makes it very hard to converge the model.

In order to address this problem, a later research proposed MARL models, with Multi-Agent Deep Deterministic Policy Gradient as an example of algorithms. Centralized Training with Decentralized Execution is commonly used in this paradigm. In training, the agents exchange global states and actions, which helps them coordinate, and in execution, they simply use local observations to make their decisions. The method is very useful in alleviating non-stationarity and optimizing regional signal policies throughout the region [9]. However, with the expansion of road networks, the size of the state space increases exponentially, resulting in enormous communication loads in traditional MARL models and drastic reductions in computational feasibility.

Application of Graph Neural Networks for Enhancing Topological Information. In order to address the challenge of the dimensionality curse and state representation bottlenecks in large-scale MARL, researchers resorted to Graph Neural Networks. Because the road network in a city is necessarily non-Euclidean, conventional Convolutional Neural Networks cannot be used to extract the irregular topological relationships between intersections.

With the introduction of Graph Convolutional Networks (GCNs), there was a paradigm shift. GCNs represent the physical topology of the network by regarding intersections as nodes and street segments as edges and aggregate upstream state information (queue lengths and speed of vehicles) to downstream intersections. There are some works that have successfully implemented the GCNs in MARL structures. With the local aggregation capabilities of graph convolutions, these models reduce the dimensionality of the state space significantly but preserve the structure of the network, which is why large scale networks can be computed with relative ease [10].

When evaluated horizontally, although GCN-based approaches reduce the communication redundancy of conventional MARL, they have a significant drawback in that they are fixed-weight. The adjacency matrices in GCNs tend to be fixed depending on the actual distance between nodes or the fixed links. In contrast, real-world traffic propagation is very dynamic (e.g., tidal traffic during peak hours or localized traffic jams due to an unexpected accident). The fixed spatial aggregation approach will not respond to these dynamic traffic conditions.

Dynamic Graph Coordination and the Rise of Spatio-Temporal Attention. The weaknesses of the static GCN weights are solved by applying Graph Attention Networks (GATs) and Spatio-Temporal Attention Mechanisms to the current state-of-the-art research. This development is a shift in the mode of coordination that was based on physical distance to data-driven dynamic association on signal control.

With such architecture, graph attention mechanisms permit any intersection agent to dynamically assign various attention weights to their neighbors depending on the traffic conditions of the current time (e.g., lane flow variability or congestion wave velocity) [11]. E.g. in case of large queuing at a preceding node, the target agent will automatically raise the level of its attention to the state of that neighbor, changing its own phase in advance to control the incoming traffic. In contrast, in late night hours when traffic is light, the attention mechanism filters away remote interference by default.

Many empirical researches confirm this mechanism. The experiment with a multi-agent DRL framework on a real-life Hangzhou network dataset revealed that the addition of spatio-temporal graph modeling and attention mechanisms enabled the model to learn long-term time-series characteristics of morning rush hour traffic and spatial interactions at complex intersections. The comparison of the spatio-temporal attention model to independent control or native GCN approaches also showed lower global average delays of 15-25 per cent in complex real-world road networks [11].

Current Status and Application Bottlenecks. To summarize the technical evolution history, the conceptual model of multi-intersection cooperative control has mostly been unified into the concept of using graph structures to represent spatial correlations, attention parameters to filter redundant information and MARL to make decisions that are globally oriented.

Nevertheless, in the context of modern project practice, these state-of-the-art algorithms are also subject to some serious engineering challenges. Multi-agent frequent interaction requires a lot of communication bandwidth (e.g., V2X communication infrastructure). Also, very powerful graph attention networks have large quantities of parameters and frequently lead to prohibitive inference latency on resource-constrained traffic signal controllers (edge computing devices). The subsequent important checkpoint of massive engineering deployment in this area is balancing cooperative performance with light-weight requirements of edge computing.

3.3 Uncertainty in Traffic Flow and Robust Modeling

Classic DRL techniques usually presuppose deterministic traffic conditions and ideal Markovian environment changes. However, actual traffic networks in cities are far too complicated and unpredictable to be simulated in software. Harsh weather conditions may drastically reduce the speed of vehicles, an accident may suddenly decrease the capacity of roads, and traffic flow between vehicles and pedestrians is highly unpredictable on the micro-level. In the absence of adaptive mechanisms to manage these uncertainties, an idealized simulation trained model is very susceptible to catastrophic failure in the real world. Consequently, there has been a swift shift in scholarship both in academia and engineering to focus on the modeling of robustness and the management of uncertainty. The recent advances can be sorted into three technical strategies:

The initial approach is using recurrent neural networks to capture temporal dependencies. The randomness of traffic arrivals makes it impossible to depend only on the current static observations. Many researchers have applied time-series structures, including the Long Short-Term Memory (LSTM) networks, with the aim to thoroughly model traffic dynamics at any time. Correctly identifying long- and short-term correlations in the patterns of arrival of vehicles enables LSTMs to endow RL agents with a dynamic capacity to predict the future evolution of flows. It greatly enhances the resilience and anti-interference properties of the model to real-world traffic changes [12].

Second is the lightweight structural compression and fault tolerance on abnormal data. In addition to natural traffic variation, physical sensors (such as cameras and radars) can create missing or incorrect data in adverse weather, which feeds bad data into deep models. To address this, investigators came up with an advanced structural compression model. Both structural and network compression methods are used to reduce bloated DRL models. Not only does it remove redundant parameters in order to improve generalization, but it also minimizes the sensitivity of models to extreme outliers. The mechanism employed by the framework is called Small Model Ensembles, which filters out corrupt knowledge representations due to traffic anomalies or distribution changes. It helps achieve stability in decision making in situations of uncertainty and reduces computational and memory costs, so that more sophisticated algorithms can be run on edge devices [13].

To sum up, the multimodal data fusion can be adapted to the extreme situation, as seen in the Hangzhou case. The use of a single source of video or radar data alone cannot assure the awareness of the environment. This is an example of how current research uses Sensor Fusion Technology to map probe vehicle GPS trajectories, real-time meteorological warnings and even social media reports about incidents into the reinforcement learning state space. This ensures that the control strategies are better aligned with actual problems.

Changes in traffic patterns in Hangzhou confirm the relevance of this multimodal configuration. In case of typhoons in the summer or unexpected rainstorms, the dynamics of the city roads change structurally: the speed of motor vehicles decreases but due to the high demand in food deliveries, large numbers of e-bikes and their riders take over the roads, resulting in the phenomenon of wrong-way driving and lane sharing. This geometrically blows up the matrix of spatial conflicts within the intersection.

Strong incorporation of weather variables, and multi-source event information into RL training allows the model to identify environmental shocks (such as flipping a "rainstorm" switch). The system then automatically alters its approach, including more yellow clearance time, more leeway on slow non-motorised vehicles, or tighter green ratio limits on large arteries.

The empirical data indicate that the multimodal uncertainty modeling increases the resilience of traffic optimization strategies in extreme conditions and is an effective way to fill the final gap between the theory of DRL and its practice implementation.

4 Conclusion

The paper provides a systematic overview of the theoretical development and frontier uses of Deep Reinforcement Learning (DRL) in urban traffic signal control. The research path indicates that control methods have now gone beyond initial single-agent independent optimization. Academia and industry alike have led a paradigm shift towards Multi-Agent Reinforcement Learning (MARL) in order to overcome the ubiquitous congestion spillback as well as spatio-temporal coupling effects in city-level road networks. Incorporating spatio-temporal graph neural networks with dynamic attention mechanisms, the most recent cooperative control algorithms are able to successfully eliminate local disturbance. Furthermore, they are able to correctly represent the extremely non-linear spatial relationships among intersections, significantly eliminating the bottleneck of computation and dimensionality in the case of global coordination in simulated architectures.

Nevertheless, implementing these sophisticated models into physical world settings exposes a fundamental obstacle: the inability to manage the high level of real-world traffic system uncertainty and constraints of the underlying hardware. Hence, this paper examines the developmental trajectory of robust modeling of highly uncertain environments. To obtain long- and short-term memory in high-frequency micro-fluctuations in traffic flow, researchers have proposed time-series dependency networks. They also use structural model compression tools at the deployment end to minimize fat cloud-level deep models significantly. It makes it possible to deploy pipelines to devices with limited computing capabilities and also helps to filter out exceptional noise due to a sensor failure or a severe weather condition. Moreover, the examination of the real-life development of traffic in Hangzhou in particular its reaction to summer typhoons and food delivery riders spikes completely confirms the need to leave behind one-dimensional thinking. The control models prove to be highly resilient and adaptive in extreme traffic conflict conditions through deep fusion of multimodal data such as meteorological warnings, probe vehicles trajectories, and multi-source emergencies. Finally, next-generation intelligent signal control is no longer solely an iterative process of algorithms but has evolved into a total systems engineering project that includes centralized cloud coordination, a wide range of multi-source perception, and light weight edge execution.

As we look into the future, the control algorithms based on DRL have enormous theoretical potential in expanding as the new generation of intelligent transportation systems are being developed. The initial significant change concerns the very essence of the physical traffic participants. Due to the increasing accessibility of Vehicle-to-Everything (V2X) communication infrastructure and the long-term state of mixed traffic flows, the mass adoption of Connected and Automated Vehicles (CAVs) will inevitably change the traditional perception and decision-making principles. Future reinforcement learning models should overcome the existing one-sided control logic of optimizing the timing of roadside signals. They will have to transform into a three-dimensional control platform comprising the vehicle pathway coordination, platoon coordination and active roadside signal playing. Such transformation will allow the full realization of the natural benefits of CAVs as high precision mobile perception points and pacifiers of traffic flow.

The second is that the optimization objectives should be moved away to a multidimensional value orientation. Current RL reward systems can be reduced to modeling limitations, concentrating too much on reducing the delays in local passages of motor vehicles. Future reward function engineering is bound to shift towards a very complex multi objective joint optimization paradigm to address the macro-level needs of sustainable urban development. It needs to include the green ecological indicators explicitly, like energy use per vehicle or carbon emission, at the micro-computing level. In addition, more rights of way should be given to vulnerable transportation users (such as scheduled buses, pedestrians, and non-motorized cars) and emergency vehicles when synthesizing policies. The algorithms will ensure that the principles of travel fairness and safety are not compromised and that the throughput of traffic is increased.

Last but not least, current models are currently very unreliable in engineering terms in the context of rare disaster situations where there is no previous sample of how the event occurred before. Regardless of whether it is a case of sudden contiguous gridlock or large scale traffic blockage due to force majeure, the conventional RL approach of relying on online random trial and error has structural delays and can lead to unpredictable secondary safety hazards. Hence, a very promising avenue is finding ways of combining the commonsense reasoning logic and human-like deduction abilities of emerging Traffic Foundation Models with the very efficient performance of lightweight edge DRL algorithms. With this new architecture, which will be based on the combination of knowledge-driven logic and data fusion, urban traffic control centers will have the ability to generalize anything about their environment and make safe and rational decisions even in the face of an entirely new distribution shift. It is a major milestone towards the accomplishment of artificial general intelligence in urban traffic control, which can actually respond to complex and dynamic environments.

References

1. M. Kurth, W. Kozlowski, A. Ganin, A. Mersky, B. Leung, J. Dykes, M. Kitsak, I. Linkov: Absence of Resilience in Transportation Networks: Economic Consequences. Transportation Research Part D, Transport and Environment 86, 102419-102419 (2020).
2. J Tan, Q Yuan, W Guo, N Xie, F Liu, J Wei and X Zhang: Deep Reinforcement Learning Application in Traffic Signal Control Model and Adaptation Research. Sensors 22(22) (2022).

3. A survey on reinforcement and deep reinforcement learning in coordination of intelligent traffic light control. Saadi, A.; Abghour, N.; Chiba, Z.; Moussaid, K.; Ali, S. Journal of Big Data 12(1), 1-20 (2025).
4. Pan, T.: The Traffic Light Control with the Reinforcement Learning. Applied and Computational Engineering 43(1), 26-43 (2024).
5. Rasheed, F.; Alvin Yau, K.; R. Md. Noor; Wu, C.; Low, Y.: Deep Reinforcement Learning of Traffic Signal Control: A Literature Review. IEEE Access 8, 208016-208044 (2020).
6. Michailidis, P., Michailidis, I., Rafail Lazaridis, C., Kosmatopoulos, E.: A Review on Applications and Innovations of Traffic Signal Control Using Reinforcement Learning. Infrastructures 10(5) (2025).
7. Bie, Y., Ji, Y., Ma, D.: Multi-agent Deep Reinforcement Learning Collaborative Traffic Signal Control Method that takes into account intersection heterogeneity. Transportation Research Part C-Emerging Technologies 164, 104663-104663 (2024).
8. B. Wang, Z. He, J. Sheng and Y. Chen: Deep Reinforcement Learning Applied to Traffic Light Timing Optimization. Processes 10(11) (2022).
9. T. Hu, Z. Li: Multi-agent Deep Reinforcement Learning-Based Traffic Signal Coordination Method. IET Intelligent Transport Systems, vol. 18, no. 8, pp. 1428-1444 (2024).
10. G. Yang, X. Wen and F. Chen: Multi-Agent Deep Reinforcement Learning with Graph Attention Network on Traffic Signal Control of Multiple-Intersection Urban Areas. Transportation Research Record (2025).
11. Jia, W., Ji, M.: Deep Reinforcement Learning with Multiple Agents in Traffic Signal Control of Large Scale Traffic Flows using Spatio-Temporal Attention Mechanism. Applied Sciences-Basel (2025).
12. Abdoos, M., Bazzan, A.L.C.: Hierarchical Optimization of Traffic Signals Based on Reinforcement Learning and Traffic Prediction by Long-Short Term Memory. Expert Systems with Applications 171, 114580 (2021).
13. Xu, D., Liao, X., Yu, Z., Gu, T., Guo, H.: Through Structure Compression, Robustness Enhancement of Deep Reinforcement Learning-Based Traffic Signal Control Model. Knowledge-Based Systems 310 (2025).