Research on Multi-Intersection Signal Control Optimization Based on Deep Reinforcement Learning (Postprint)

Authors: Zhao Chun, Dong Xiaoming, Ren Yiying

Date: 2022-04-07T15:01:57+00:00

Abstract

Emerging intelligent transportation systems hold significant promise for improving traffic flow, optimizing fuel efficiency, reducing delays, and enhancing overall driving experience. Currently, traffic congestion constitutes an extremely serious challenge for humanity, particularly severe at urban intersections with dense traffic. The reward mechanism of the signal control system is improved by transitioning from a shared reward scheme across all intersections to a unique reward per intersection, and through the combination of a dense sampling strategy with multi-intersection signal control, leveraging the currently prevalent deep reinforcement learning to solve traffic signal timing problems. Simulation experiments are conducted based on the internationally mainstream traffic simulation software (SUMO), and results demonstrate that the improved deep reinforcement learning multi-intersection signal control method achieves superior control performance compared to traditional reinforcement learning methods.

Full Text

Multi-Junction Signal Control Optimization Based on Deep Reinforcement Learning

Zhao Chun[1]**, Dong Xiaoming**[1]**†, Ren Yiying**[2]
(1. School of Computer & Information, Anqing Normal University, Anqing, Anhui 246000, China;
2. School of Electronic Engineering & Intelligent Manufacturing, Anqing Normal University, Anqing, Anhui 246000, China)

Abstract

Emerging intelligent transportation systems hold significant promise for improving traffic flow, optimizing fuel efficiency, reducing delays, and enhancing the overall driving experience. Today, traffic congestion represents an extremely serious challenge for humanity, particularly severe at urban intersections with dense traffic. This paper improves the reward mechanism of signal control systems by transitioning from a shared reward across all intersections to a unique reward for each individual intersection. Combined with a dense sampling strategy for multi-intersection signal control, the proposed approach leverages the popular deep reinforcement learning technique to solve traffic signal timing problems. All simulation experiments are conducted using the internationally mainstream traffic simulation software SUMO. Experimental results demonstrate that the improved deep reinforcement learning method for multi-junction signal control achieves superior performance compared to traditional reinforcement learning approaches.

Keywords: intelligent transportation system; deep reinforcement learning; traffic flow; multi-junction signal control

0 Introduction

With the continuous growth in the number of motor vehicles, traffic congestion has become an extremely complex and troubling problem facing humanity, particularly acute in large metropolitan areas with complicated traffic conditions [1]. Traditional traffic signals operate with fixed timing, leading to unnecessary waiting during green phases and causing substantial resource waste. Multi-junction traffic signal control based on deep reinforcement learning can effectively alleviate traffic congestion, reduce accidents, and improve system efficiency and rationality.

Conventional Markov decision processes and reinforcement learning suffer from poor scalability, resulting in state space explosion. Reinforcement learning is an adaptive control strategy where one or more agents autonomously learn to solve tasks in an environment through experience generated by interacting with the environment itself [2]. Early traffic signal control relied heavily on manual feature extraction, requiring significant human resources while being prone to state fluctuations and loss of critical state information. Traditional Q-learning, proposed by Watkins in 1989, is a model-free online reinforcement learning algorithm [3]. In Q-learning, the green light duration for each time step should increase with rising traffic intensity. However, configuring excessively high or low phase green times for a given traffic state is highly unreasonable. EL-Tantawy et al. [4] summarized reinforcement learning approaches for traffic signal control from 1997 to 2010, which were limited to tabular Q-learning and typically used linear functions to estimate Q-values. Due to technological constraints at the time, state space definitions often employed simple data types such as queue lengths and traffic flow volumes, which could not fully capture the complexity of traffic systems, preventing reinforcement learning from achieving optimal results in traffic signal control. Balaji et al. [5] combined traditional Q-learning with traffic signal control, validating its effectiveness. However, traditional Q-learning may lead to an excessively large action space, ultimately causing dimensionality explosion.

With the development of reinforcement learning and deep learning, researchers have proposed combining them as deep reinforcement learning to estimate Q-values. Li et al. [6] applied deep reinforcement learning to single-intersection control problems with improvements. LEE et al. [7] combined Convolutional Neural Networks (CNN) with Q-learning to propose the DQN algorithm, which utilizes experience replay to break sample sequence correlations and improve learning efficiency. Advances in vehicle communication technology now provide more detailed information about vehicle positions and speeds. This enables comprehensive real-time information combined with edge cloud computing to implement more flexible traffic light control policies for effective flow improvement, and in the long term, could directly drive fully autonomous driving scenarios. While the potential benefits are enormous, the technical challenges are equally significant. Moreover, such control systems involve unprecedented scale in terms of intrinsic complexity, geographic scope, and number of objects. Real-world traffic signal timing is often distributed, hybrid, and difficult to predict. Overcoming these challenges requires introducing deep reinforcement learning concepts—DQN is an algorithm with strong perception capabilities and rapid decision-making ability.

The main advantages of the proposed method are: (a) improving the reward mechanism of traffic signal control systems by changing from a shared reward across all intersections to a unique reward for each intersection; (b) combining dense sampling strategy with multi-junction signal control to enhance control performance; (c) conducting all simulation experiments using the internationally mainstream traffic simulation software (Simulation of Urban Mobility, SUMO) to significantly improve experimental reliability and stability; and (d) employing reasonable parameter settings and multiple experiments to reduce randomness and enhance control system stability.

---

1 Intersection Model Establishment

This paper establishes two types of road intersection models and presents optimization solutions, described separately below.

1.1 Single Intersection Model

The single intersection model established in this paper is shown in Figure 1, where $Q_i(t)$ represents the number of vehicles waiting to pass through traffic flow i, and the intersection state is represented by $P(t) \in \{0, 1, 2, 3\}$. The traffic light configuration is defined as: "0" : direction 1 green, direction 2 red; "1" : direction 1 yellow, direction 2 red; "2" : direction 2 green, direction 1 red; "3" : direction 2 yellow, direction 1 red.

As shown in equation (1), the action decision $A(t)$ at time t is selected, where $A(t) \in \{0, 1\}$ is represented by a binary variable: "0" means continue, "1" means switch.

These rules generate a strict cyclic control sequence. As shown in Figure 2, the queue state evolves over time under recursive control. Next, we examine the vehicle calculation function for a single intersection.

$Q_i(t)$ represents the number of vehicles waiting to pass through traffic flow i at time t. $S_i(t)$ represents the number of vehicles of traffic flow i appearing at the intersection at time t, and $W_i(t)$ represents the number of vehicles of traffic flow i leaving the intersection.

1.2 Multi-Junction Intersection Model

To investigate the performance and scalability of the DQN algorithm in large-scale scenarios on more complex roads, this paper considers a linear network topology [8], as shown in Figure 3, examining a multi-junction intersection model structure with N intersections and bidirectional traffic flow.

At this point, the dimensionality changes, requiring an upgrade to the single-intersection function. The system state $P(t)$ at time t must be described by a 5-tuple $(Q_{n1}(t), Q_{n2}(t), Q_{n3}(t), Q_{n4}(t), P_n(t))$ where $n = 1 \dots N$.

The following is the queue state transition function for the multi-junction intersection model structure:

Next, we examine the vehicle calculation function for the multi-junction intersection model structure: $S_{ni}(t)$ represents the number of vehicles appearing in direction i at intersection n at time t, $W_{ni}(t)$ represents the number of vehicles leaving in direction i at intersection n at time t, and $S_{n1}(t), S_{n2}(t), S_{n3}(t), S_{n4}(t)$ (for $n = 1 \dots N$) correspond to all vehicles approaching the intersection from the external environment:

Equations (5) and (6) indicate that vehicles passing through direction 1 of intersection n during time period t appear as vehicles in direction 1 of intersection $(n + 1)$ during time period u moving eastward. Similarly, vehicles passing through direction 3 of intersection $(n+1)$ during time period t appear as vehicles in direction 3 of intersection n during time period u moving westward. This creates highly complex interactions between vehicles across intersections along the main road, presenting additional challenges for optimizing control strategies.

---

2 Deep Reinforcement Learning Framework

2.1 State Representation

On each arm of multi-junction intersections, incoming vehicles are discretized into cells that can identify whether vehicles are present. The system state S

is fed into the DQN as input to both the target and evaluation networks. The algorithm' s environmental state is represented as road surface discretization to inform the agent of vehicle positions at specific times. The input for a single intersection is $S = (Q_1, Q_2; P)$, while for multi-junction scenarios it becomes $S = (Q_{n1}, Q_{n2}, Q_{n3}, Q_{n4}; P_n)$, resulting in a dimensional change.

2.2 Action Behavior

The action set represents the available interaction methods for the agent, defined as the configuration in Section 1.1. Executing an action means turning some traffic lights green on a set of lanes for a fixed duration.

2.3 Reward Mechanism

In Sun et al.' s experiment [9], the delay time for vehicles entering each lane was set as d, the sum of queue lengths of all waiting vehicles in entering lanes as q, the waiting time of all vehicles in entering lanes as w, phase state switching as p, emergency braking stops as e, and the number of vehicles leaving after action execution as n. The comprehensive reward formula is:

This paper improves the reward mechanism for multi-junction signal control systems by transforming the R_t function into a two-dimensional function $R_t[x][y]$, changing from shared rewards across all intersections to unique rewards for each intersection. The formula is:

This means subtracting the cumulative reward value of all previous intersections from the reward value of all previously passed vehicles. After i iterations, the current intersection' s reward value is obtained—the so-called unique reward. This way, each intersection has its own reward, significantly improving the precision of experimental results after implementing this improved mechanism.

2.4 Q-Learning Update Formula

This paper uses the following update formula:

The reward r_{t+1} is obtained after taking action a_t in state s_t, and s_{t+1} is the next state after taking the relevant action. The discount factor γ indicates that future rewards are increasingly penalized compared to immediate rewards as time step t progresses. This formula updates the current action' s Q-value in state S_t using immediate rewards and discounted future Q-values. The term representing the implicit value of future actions holds s_{t+1} and also possesses the maximum discounted return after the next state, i.e., the maximum discounted return of the state. This demonstrates that regardless of how the agent chooses the next action, decisions are based not only on immediate rewards but also on expected future discounted rewards. During simulation, the agent continuously iterates to acquire knowledge about action sequence values, ultimately selecting action sequences that achieve higher cumulative returns for optimal performance.

2.5 Deep Neural Network

This paper employs the Deep Q-Learning algorithm, mapping observed environmental states s_t to action-related Q-values through a deep neural network. Its input is the IDR (environmental state vector) at time step t, and the network' s output is the Q-value of actions from state s_t. Generally, the neural network input $in^n_{k,t}$ is defined as the k-th element of vector IDR at time step t, representing the n-th input to the neural network at time t. In this paper, the input is the system state $S = (Q_{n1}, Q_{n2}, Q_{n3}, Q_{n4}; P_n)$. The neural network output is defined as $out^n_{v,t} = Q(s_t, a_{v,t})$, representing the Q-value of taking the v-th action at time step t.

This paper first presents the DQN algorithm for single-intersection scenarios, then demonstrates its effectiveness for linear topology structures with N intersections. Even in the latter case, a "single-agent" DQN algorithm with global access is adopted. This approach differs from "multi-agent" methods by using only one agent to reduce intersection complexity and redundancy. Although the single-agent method involves a larger state space, it achieves more intelligent control and coordination. Figure 4 clearly illustrates the connections between layers in the deep neural network:

As shown in the figure, n IDR vectors are input to the deep neural network and transmitted to neural network layers for training. After training, Q-values related to time step t are output.

---

3 Simulation Experiments

All experiments are conducted using the internationally recognized traffic simulation software SUMO [10] (Simulation of Urban Mobility), an open-source, microscopic, multi-modal traffic simulation platform that allows individual route planning for each vehicle on the road. It enables simulation of given traffic demands composed of individual vehicles and their movement in a specified road network, as illustrated in Figure 5.

3.1 System Input

Before training begins, the system first generates simulations of vehicles and intersections. As shown in Figure 5, the system randomly generates vehicles and traffic signal states. The specific state transition process is reflected in Figure 2, which only magnifies the generation process for one intersection in the multi-junction network. The complete multi-junction road network generation process is shown in Figure 6, forming an entire multi-junction road network simulation.

After simulation completion, the intersection system state S is fed as input to both target and evaluation networks [11], where $S = (Q_{n1}, Q_{n2}, Q_{n3}, Q_{n4}; P_n)$.

Here, $Q_{n1}, Q_{n2}, Q_{n3}, Q_{n4}$ represent vehicles approaching from four directions at each intersection, while P_n represents the vehicle state transition probability. The vector $S = (Q_{n1}, Q_{n2}, Q_{n3}, Q_{n4}; P_n)$ is ultimately input into the DQN algorithm for training.

3.2 Dense Sampling Strategy

The dense sampling strategy enhances model implementation and testing, thereby improving agent performance during training when the γ value is high. The agent' s training phase involves finding the most valuable action given an environmental state. However, during early training stages, the most valuable actions are unknown. To overcome this, the agent should initially explore action consequences without concern for performance. The hyperparameters for agent model training are set as follows:

a) Neural network: 5 layers, each containing 400 neurons.

b) γ value: increased from 0.25 to 0.75.

c) Reward function: unique reward, as described in Section 2.3.

The sampling method in Figure 8 collected approximately 2.5 million samples over 4,000 training episodes. To achieve a qualitative improvement in training episodes, the γ value was increased to 0.75. Figures 9 and 10 show that 5,000 training episodes collected over 60 million samples, demonstrating that the proposed dense sampling method provides a qualitative improvement. This combination of the new reward function and sampling strategy helps solve Q-value instability issues and significantly reduces the likelihood of misguidance by future optimal actions.

3.3 System Training Process

Q_1 to Q_{14} represent only a portion of the multi-junction traffic network; the actual experimental scenario is much more complex. Target Q-values provide the foundation for updating the neural network approximator through Q-Learning, while the evaluation network is updated via gradient descent and greedy strategies.

As established in Sections 1.1 and 1.2, two models are created: single intersection and linear topology. Experimental comparisons between these models clearly demonstrate the advantages of the proposed method. By combining dense sampling strategy, the agent' s training dataset is substantially increased, making $Q(s, a)$ more stable and convergent. Specific experimental results are presented in Section 4.

The interaction method for vehicles at intersections is shown in Figure 7, implemented through formulas (4), (5), and (6) from Section 1.2. The numbers on the right side of the figure represent the number of vehicles waiting on each road, while black rectangles indicate vehicles entering from surrounding roads. This creates highly complex interactions between vehicles for coordinated and stable training.

4 Experimental Results Analysis

This paper compares experimental results between single and multi-junction scenarios. Figure 8 shows the cumulative negative reward values obtained from single-intersection training [12]. The results are unsatisfactory, with reward values showing excessive fluctuations and large value ranges, indicating high instability.

Next, we examine experimental results comparing the original shared reward and improved unique reward for multi-junction scenarios, shown in Figures 9 (shared) and 10 (unique). The left figure' s stability is significantly weaker than the right figure' s, and the multi-junction reward value range is much smaller than that of single-junction [13], demonstrating greater stability. The dense sampling strategy [14] yields an order of magnitude larger sample size than the single-junction case, further proving the superiority and stability of the proposed algorithm.

Three trained network models are tested, with results shown in Figure 11. Here, x_1 represents the vehicle queue length at single intersections, while x_2 and x_3 represent queue lengths for multi-junction shared and unique reward scenarios, respectively. The figure clearly shows x_1 has the longest queue length, averaging nearly 10m. x_2 shows some improvement, while x_3 performs best, reducing average queue length to approximately 2.5m—a substantial performance improvement. Testing clearly demonstrates the advantages of the proposed method, which significantly reduces average vehicle queue length and improves agent performance and system stability.

5 Conclusion

Traffic intelligence and informatization represent a prevailing trend in modern society. Due to the complexity and dynamic nature of traffic systems [15], coupled with continuously expanding control scope and exponentially increasing traffic state information data, control complexity grows exponentially, yet traffic network signal control problems remain unsolved effectively.

This paper explores both single-junction and more complex linear network topology scenarios [16], applying deep reinforcement learning algorithms to both cases. Comparative results clearly demonstrate that the proposed method effectively reduces intersection congestion and significantly saves energy consumption, providing substantial improvements in efficiency and performance. The agent maximizes global vehicle passage speed within limited time, continuously adjusting its internal parameters through reinforcement learning according to different policies. Ultimately, deep reinforcement learning discovers more complex cross-road network features, enabling direct learning of effective control strategies from high-dimensional data. This allows the agent to significantly improve average vehicle speed, minimize average travel time, reduce average waiting queue length, and select optimal traffic control strategies by observing current traffic states. Experimental results show that the improved multi-junction control method substantially enhances system control performance.

Over the past few years, reinforcement learning techniques for traffic signal control have matured significantly with the popularization of deep learning. Future work will investigate algorithms in more complex road scenarios, integrating the proposed method with vehicle communication technology to provide more detailed vehicle status information. Combining comprehensive real-time information with edge cloud computing will ultimately achieve effective traffic flow improvement and flexible intelligent traffic control.

References

[1] Ge Z. Reinforcement learning based signal control strategies to improve travel efficiency at Urban intersection [C]// International Conference on Urban Engineering and Management Science (ICU-EMS). Zhuhai, China: IEEE, 2020: 347-351.

[2] Zhang Hongsen, Liu Tian, Zhao Yuhong, et al. Research and design of intelligent traffic signal lamp [J]. Industrial Control Computer, 2020, 33 (10): 132-133.

[3] Hatri C E, Boumhidi J. Q-learning based intelligent multiobjective particle swarm optimization of light control for traffic urban congestion management [C]// The 4th IEEE International Colloquium on Information Science and Technology. Tangier, Morocco: IEEE, 2016: 794-799.

[4] Eltantawy S, Abdulhai B, Abdelgawad H. Design of Reinforcement learning parameters for seamless application of adaptive traffic signal control [J]. Journal of Intelligent Trans Systems, 2014, 18 (3): 227-245.

[5] Balaji P, German X, Sxinivasan D. Urban traffic signal control using reinforcement learning agents [J]. IET Intelligent Transport Systems, 2010, 4 (3): 177-188.

[6] Li L, Yisheng L, Feiyue W. Traffic signal timing via deep reinforcement learning [J]. IEEE/CAA Journal of Automatica Sinica, 2016, 3 (03): 247+254+248-253.

[7] Jeehyong L, Hyunglee K. Distributed and cooperative fuzzy controllers for traffic intersections group [J]. IEEE Trans on Systems Man and Cybernetics,

1999, 29 (2): 263-271.

[8] Cheng Yuyang, Zhou Bintao, Shi Chengxi. Traffic signal timing optimization based on long and short term memory artificial neural network and SUMO simulation [J]. Scientific and Technological Innovation, 2021 (26): 67-70.

[9] Sun Hao, Chen Chunlin, LIU Qiong, et al. Traffic signal control method based on deep reinforcement learning [J]. Computer science, 2020, 47 (02): 169-174.

[10] Ma Dongfang, Xiao Jiawang, Ma Xiaolong. A decentralized model predictive traffic signal control method with fixed phase sequence for urban networks [J]. Journal of Intelligent Transportation Systems, 2021, 25 (5): 62-78.

[11] Wang Juanjuan, Wang Yanan, Zhou Hongfang. Design of online simulation system for signal control of Urban intersections based on visual sensing technology [J]. Journal of Physics: Conference Series, 2021, 1982 (1): 136-149.

[12] Guo Mengjie, Ren Anhu. Single intersection signal control algorithm based on deep reinforcement learning [J]. Electronic Measurement Technique, 2019, 42 (24): 49-52.

[13] Hao Huang, Zhi Qun hu, Zhao Minglu, et al. Network-scale traffic signal control via multiagent reinforcement learning with deep spatiotemporal attentive network [J]. IEEE Trans on System Man and Cybernetics, 2021, 10 (1109): 1-13.

[14] Gao J, Shen Y, Liu J, et al. Adaptive traffic signal control: deep reinforcement learning algorithm with experience replay and target network [J]. Ithaca: Cornell University Library, arXiv.org, 2017, 21 (5).

[15] Curran N, Sun J, Joowha H. Anthropomorphizing alphaGo: A content analysis of the framing of google deepmind' s alphaGo in the chinese and american press [J]. AI & Society, 2020, 35 (3): 727-735.

[16] Wu N, Li D, Xi Y. Distributed weighted balanced control of traffic signals for urban traffic congestion [J]. IEEE Trans on Intelligent Transportation Systems, 2019, 20 (10): 3710-3720.