Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 3631489, 9 pages
https://doi.org/10.1155/2018/3631489

Research Article

Evaluation and Application of Urban Traffic Signal Optimizing Control Strategy Based on Reinforcement Learning

Yizhe Wang[1,2] **Xiaoguang Yang**[1,2] **Yangdong Liu**[1,3] **and Hailun Liang**[1,2]

[1]*Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, China*
[2]*Intelligent Transportation System Research Center of Tongji University, 4801 Cao'an Road, Shanghai 201804, China*
[3]*Hangzhou Hikvision Digital Technology Co., Ltd., No. 555 Qianmo Road, Binjiang District, Hangzhou 310052, China*

Correspondence should be addressed to Xiaoguang Yang; yangxg@tongji.edu.cn

Received 6 August 2018; Revised 3 November 2018; Accepted 9 December 2018; Published 26 December 2018

Guest Editor: Hamzeh Khazaei

Reinforcement learning method has a self-learning ability in complex multidimensional space because it does not need accurate mathematical model and due to the low requirement for prior knowledge of the environment. The single intersection, arterial lines, and regional road network of a group of multiple intersections are taken as the research object on the paper. Based on the three key parameters of cycle, arterial coordination offset, and green split, a set of hierarchical control algorithms based on reinforcement learning is constructed to optimize and improve the current signal timing scheme. However, the traffic signal optimization strategy based on reinforcement learning is suitable for complex traffic environments (high flows and multiple intersections), and the effects of which are better than the current optimization methods in the conditions of high flows in single intersections, arteries, and regional multi-intersection. In a word, the problem of insufficient traffic signal control capability is studied, and the hierarchical control algorithm based on reinforcement learning is applied to traffic signal control, so as to provide new ideas and methods for traffic signal control theory.

1. Introduction

Traffic congestion has become a world-concerned problem all over the world. With the increasing number of vehicles, traffic congestion has deeply affected people's daily life and the development of social economy. Traffic control is one of the most important technological means of regulating traffic flow, improving obstruction, and improving its safety and even energy conservation and emission reduction. At present, traffic signal control problem not only has a long-time congestion phenomenon at peak time, but also has obvious ability of grooming in peak time. In order to ease the traffic pressure, rational analysis and control are considered as an important tool. Its progress and development are always keeping pace with the times, accompanied by information technology, computer technology, and system science.

According to the system's ability to adapt to the environment and the level of intelligent decision-making, Gartner proposed the evolution of urban transport control system development level in 1996 [1]. The first-generation self-adaptive control system adopts the multi-time timing control of fine division of period, or completely isolated self-adaptive control, to realize the simple regulation of traffic flow. The second-generation traffic signal control system dynamically adjusts the parameters of the signal timing scheme (cycle length, split, offset). Typical second-generation control systems include SCATS [2] and SCOOT [3]. The UK Transport Research Laboratory had a worldwide reputation for contributions to the field of traffic signal control, especially as originators of the TRANSYT and SCOOT signal coordination methods [4]. The third-generation control system uses similar idea to the second-generation to dynamically adjust the signal timing parameter in response to the fluctuation of the time-varying traffic flow at the intersection. HK Lo and HF Chow investigated the relationship of finer resolutions and larger errors in adaptive traffic control system through an extensive simulation of scenarios in Hong Kong with a recently developed dynamic traffic control model, DISCO

[5]. Aboudolas K and Papageorgiou M tested a preliminary simulation-based investigation of the signal control problem for a large-scale urban road network using store-and-forward modeling demonstrating the comparative efficiency and real-time feasibility of the developed signal control methods [6]. The fourth-generation traffic signal control system is an integrated traffic management and control system. Meneguzzer C presented two alternative deterministic, discrete-time DP models of the interaction between signal control and route choice, which are proposed and compared with the conventional iterative optimization and assignment (IOA) method for network traffic signal setting [7]. The fifth-generation traffic signal control system is based on the abilities of artificial intelligence and self-learning.

2. Literature Review

2.1. Traffic Big Data Environment. Digitized and informational infrastructure of urban road traffic and constructions of related systems have developed rapidly in the past ten years, and urban traffic control is developing from the "data poverty" times to the "data rich" times. Meanwhile, the appearance of ICV (intelligent connected vehicle) and autonomous vehicles will construct the future traffic environment jointly, which significantly differs from conventional manual driving vehicles in terms of individual information acquisition, perception ability, reaction time, interactive behavior, etc. New requirements of traffic control have formed a high-level demand for the next generation of traffic control [8]. The research on the next generation of traffic signal control for regional transportation under the "data rich" environment is on the agenda. Ma D et al. proposed the lane-based saturation degree estimation for signalized intersections and maximum queue length estimation for traffic lane groups, which enriched the way to obtain traffic parameters and increased the precision of estimation method. For example, the results show that the new method of maximum queue length estimation has a higher precision compared to the existing method based on a similar concept, with maximum and average deviations of 39.36% and 12.25%, respectively, over twenty cycles [9, 10].

Under the conditions of limited cross-section traffic flow data, many existing adaptive traffic control systems have adopted traffic models to actively predict the evolution of network traffic flows and then adopted the aggregative indicator method to optimize and solve timing parameters. However, the real-time detection of the spatiotemporal data based on urban road network traffic status can provide rich and high-quality basic data and fine-grained assessment of control effects for traffic control. In the face of the main defects encountered in the existing self-adaptive traffic control system, a closed-loop feedback self-adaptive control system with better uncertainty response capability and higher intelligent decision-making level is an inevitable result of the objective needs of the development and application technologies [11]. Ma D proposed a calculation method for the occupancy per cycle under different traffic conditions presented, based on the relationship between the three basic traffic flow parameters, speed, traffic flow, and density [12]. The results show that the precision of this method was affected by the detector location and bus ratio insignificantly [13].

2.2. Reinforcement Learning Traffic Control. According to real-time collection of states, rewards, and punishments, the single intersection's signal control of reinforcement learning can find an optimization strategy of traffic signal control suitable for traffic flow characteristics through the interaction. In recent years, more and more domestic scholars have studied principles of reinforcement learning and discussed the applications of reinforcement learning algorithms in traffic control. Reinforcement learning has developed rapidly in the optimizing control [14, 15].

Scholars have done a lot of research on reinforcement learning theory, algorithms, and applications and have obtained many famous research results. Ma D proposed a new control method that brings significant and positive effects to the bottleneck link itself and to the entire test area [16]. Yang W concluded that critical issues in developing agent-based traffic control systems for integrated network were addressed as interoperability, adaptability, and extendibility [17]. Zhang L drew a conclusion that extensive simulation results for the designed Shanghai simulation scenarios indicate that most of the observed counts match quite well with the traffic simulation volumes and demonstrate the potential of MATSIM for large-scale dynamic transport simulation [18]. Aslani M developed adaptive traffic signal controllers based on continuous residual reinforcement learning (CRL-TSC) that was more stable, and the best setup of the CRL-TSC leads to saving average travel time by 15% in comparison to an optimized fixed-time controller [19].

Reinforcement learning control has the advantages of real-time online and feedback control, which especially accords with the control thoughts of signal adaptive control in urban intersections. However, there is a question as to whether the traffic signal optimization strategy based on the reinforcement learning is applicable to all the traffic environments.

3. Traffic Signal Control Strategy Based on Reinforcement Learning

Reinforcement learning is a typical data-driven control method. In this paper, the method of signal control scheme improvement is proposed. According to the different traffic flow characteristics, the subregions are divided. Based on the three key parameters of cycle length, arterial coordination signal offset, and green split, a set of hierarchical control algorithms based on reinforcement learning is constructed to optimize and improve the current signal timing scheme.

3.1. Control Subregions Division and Cycle Optimization. As for the regional coordination control, the primary content is the division of the coordination subregions. In the signal control road network, each intersection has its influence range, and the intersection and section within this range

are greatly affected by it. To quantify the impact and define the scope of influence, literature defines direct relevance to describe the relationship between adjacent intersections, finding that when the upstream node traffic flows into the downstream node, it is close to or greater than the downstream node's import capacity. It is found that the path correlation is mainly affected by the traffic network topology and OD distribution between the two intersections. The more the OD paths through two nodes at the same time, the stronger the correlation between nodes. The higher the flow rate of OD path passing through both nodes at the same time, the stronger the correlation between nodes. The more the OD paths that pass through both nodes at the same time are unique, the stronger the correlation between nodes will be.

The optimization range is region-level road network optimization. The control subregions are divided by characteristic parameters such as average travel time; vehicle OD amount between intersections and traffic coordination control subregions are finally determined.

The signal cycle refers to the time required for the signal color to display one cycle in the set phase order, that is, the sum of the steps of each control step in one cycle. The signal cycle is the key control parameter that determines the effectiveness of traffic signal control. If the signal cycle is too short, it is difficult to ensure that the vehicles in all directions can pass through the intersection smoothly, resulting in frequent stops at the intersection and a decline in the utilization rate of the intersection. If the signal cycle is too long, it will cause the driver to wait for too long, greatly increasing the delay time of the vehicle. The cycle in the green wave control is taken as the common cycle by the maximum signal cycle of the key intersection of the arterial, and the signal cycle of the remaining intersections is reallocated to each phase according to the traffic flow ratio.

According to different evaluation indexes, the optimal cycle is obtained by using model-based algorithm. Regarding the evaluation indicators of traffic efficiency at intersections, traffic capacity, saturation, service level, travel time, number of stops, and queue length are commonly used at home and abroad. The delay is mainly due to the travel time loss caused by traffic friction and traffic control. It is closely related to the cycle duration, green split, and saturation. It is an important indicator for evaluating the traffic service level and operational efficiency of signalized intersections, including queue delay, parking delay, control delay, and lane approach delay.

3.2. Offset Optimization Based on Bayesian Optimization Algorithm. The phase offset is also called the time offset or the green time offset. The phase offset includes the absolute phase offset and the relative phase offset. Absolute phase offset refers to the offset between the starting or ending point of the signal green light (red light) in the coordinated direction of the arterial at each intersection and the starting or ending point of the signal green light (red light) in the coordinated direction of the arterial at a certain intersection (generally a key intersection). Relative phase refers to the time offset between the starting or ending points of the green light (red light) signal in the coordinated direction of the arterial at adjacent intersections. The relative phase offset is equal to the difference value between the absolute phase offset of two intersections, which is determined by the actual vehicle speed.

According to the coordination effect between the intersections, it is divided into several control subregions, and internal coordination control is implemented for its traffic characteristics. The basic principles of control subregions division are as follows:

(1) The distance between adjacent intersections is less than 600 meters and control subregion contains no more than 10 intersections.

(2) The optimal period length of each intersection is an integer multiple relationship.

The following lines with inconsistent coordination effects should not be included in a subregional coordination:

(1) An excessively long connection, and the traffic flow along the connection is highly discrete.

(2) There are traffic production sources or attraction sources (such as large parking lots and shopping malls) and very frequent pedestrian activities along both sides of certain lines, which seriously interfere with traffic flow.

The Bayesian optimization algorithm belongs to the sequential model-based optimization (SMBO) algorithm. This algorithm determines the value of the next (optimal) sample set by analyzing historical observations of a loss function f. Since the Bayesian optimization algorithm was proposed around 2010, it has been used to optimize the hyperparameters of machine learning models in the field of machine learning in recent years. The so-called superparameter is the model parameter that needs to be set artificially. In this competition, due to the large number of timing parameters that need to be optimized, which includes the signal split and phase offset of multiple different intersections, the solution space dimension is relatively high and the optimization is quite difficult. The overall idea of the Bayesian optimization algorithm is as follows:

Calculate the posterior expectation of the loss function f using the observed sample set $\mathbf{X}_{1:n}$.

Generate a new set of samples $\mathbf{X}_{new}$ to sample the loss function f, which can maximize the expectation of f in the value range of independent variables.

Repeat the above steps until the preset convergence condition is reached. End the optimization process.

The algorithm will be described in detail below and the process will be summarized.

To calculate the posterior expectation of the loss function f, the likelihood model of the sample and the prior probability model of f should be obtained in advance. In the Bayesian optimization process, we can assume that the sample obeys the multivariate Gaussian distribution and obtain the Gaussian likelihood function:

$$y = f(\mathbf{X}) + \epsilon, \quad \epsilon \sim N\left(0, \sigma_\epsilon^2\right) \tag{1}$$

For the prior distribution, we assume that the loss function f can be described by a Gaussian process (GP). The essence of the Gaussian process is the generalization

of the multivariate Gaussian distribution to the function distribution. Therefore, just as the Gaussian distribution is determined by its expectation and variance, the Gaussian process is completely determined by its expectation function $m(\mathbf{X})$ and the covariance function $k(\mathbf{X}, \mathbf{X}')$. The Gaussian process is widely used in the application of all probabilistic models because its description of the posterior distribution of the loss function is easier for us to analyze and calculate.

One of the most widely used acquisition functions is the expected improvement (EI) function. The EI function is defined as

$$EI(\mathbf{X}) = \mathbf{E}\left[\max\left\{0, f(\mathbf{X}) - f\left(\widehat{\mathbf{X}}\right)\right\}\right] \tag{2}$$

where $\widehat{\mathbf{X}}$ is the current optimal sample set, and this function gives a new sample set that can best enhance the expectation of the loss function. Moreover, the expected lifting function can be calculated based on the Gaussian process model, namely,

$$EI(\mathbf{X}) = \begin{cases} \left(\mu(\mathbf{X}) - f\left(\widehat{\mathbf{X}}\right)\right)\mathbf{\Phi}(Z) + \sigma(\mathbf{X})\emptyset(Z) & \text{if } \sigma(\mathbf{X}) > 0 \\ 0 & \text{if } \sigma(\mathbf{X}) = 0 \end{cases} \tag{3}$$

$$Z = \frac{\mu(\mathbf{X}) - f\left(\widehat{\mathbf{X}}\right)}{\sigma(\mathbf{X})} \tag{4}$$

where $\mathbf{\Phi}(Z)$ and $\emptyset(Z)$ are the cumulative distribution function and probability density distribution function of the multivariate standard Gaussian distribution, respectively. When the posterior expectation μ(X) is higher than the current loss function optimal value $f(\widehat{\mathbf{X}})$, EI will get a larger value. When the uncertainty σ(X) of X is high, EI will get a larger value.

After the above analysis and introduction, the whole principle and process of Bayesian optimization can be summarized to form a Bayesian optimization algorithm:

Given the observed value $f(\mathbf{X})$ of the loss function, the posterior expectation of the loss function f is updated based on the Gaussian model.

Solve the expected lifting function (EI function) to find the new best sample set: $\mathbf{X}_{new} = \arg\max EI(\mathbf{X})$.

Calculate the value of the loss function at $\mathbf{X}_{new}$.

Repeat the above steps until the preset number of repetitions (i.e., the number of iterations) is reached or the convergence condition is met.

In (2) of the above steps, we can use the gradient-based solution method to optimize the EI function to get $\mathbf{X}_{new}$.

On the basis that the parameters such as the optimal cycle length are determined, the phase offset of the intersections after deduplication $\mathbf{X} = \{x_1, \ldots, x_n\}$ can be regarded as input loss function sample set. The sample $\mathbf{X}$ of the function, which is returned by the online feedback, can be iterated multiple times based on the Bayesian optimization algorithm.

3.3. Split Optimization Based on Q-Learning Algorithm. In the urban transportation system, the traffic flow, vehicle speed, and traffic density are the most intuitive reflections of traffic conditions. They are the three characteristic parameters of traffic flow and the research focus and foundation of traffic flow theory. Among them, the traffic flow refers to the number of vehicles passing through per unit time; the vehicle speed refers to the distance that the vehicle passes per unit time; and the traffic density refers to the number of vehicles on the section per unit length. The traffic flow theory is the basis for the establishment of urban traffic signal control system.

The traffic model uses a discrete-time difference equation or a continuous time subdivision tool to introduce a dynamic relationship between the concepts of traffic volume Q, vehicle speed V, and traffic density K, which summarizes the physical quantities of the traffic network and is used to describe the collective average behavior of a large number of vehicles. In the free flow, the interaction between vehicles can be neglected, and the traffic flow increases linearly with the vehicle density. The wide moving jam flow is usually characterized by stop-go-stop traffic, that is, a series of jams. The density of vehicles in the region is high and the average speed and flow of vehicles are small. The average velocity of the synchronized flow is significantly lower than that of the free flow.

At present, Q-learning algorithm is one of the most frequently used methods in the fields of reinforcement learning, proposed by Watkins in 1989 [20]. Q-learning algorithm is widely used in the fields of control, depending on the update mode of its special value function.

In Q-learning, the solution formula of the mainstream value function is as follows.

$$\begin{aligned} Q(s_t, a_t) &\longleftarrow Q(s_t, a_t) \\ &\quad + \alpha\left[\left(r_{t+1} + \gamma \max_{a \in A} Q(s_{t+1}, a) - Q(s_t, a_t)\right)\right] \end{aligned} \tag{5}$$

According to the formula, at the moment of t, the state of Q-learning is s_t. If the taken action is a_t, the corresponding value function will be Q(s_t, a_t). The update of the value function is determined by three factors. The first is the current value of the action state value function, Q(s_t, a_t), that needs to be updated. The second is to control the corresponding maximum value of all Q-values of actions in the postexecution state of s(t+1), and the third is the immediate return, r(t+1), after the action. Besides, there are also two model parameters, learning rate $\alpha \in [0,1]$ and discount factor $\gamma \in (0,1]$. The former is used to balance the relationship between the learning and utilization of the algorithm. When a$\longrightarrow$1, the controller tends to explore new knowledge; otherwise it will use the existing knowledge. The latter is used to coordinate the present relationship with the future. When $\gamma \longrightarrow 1$, the controller tends to consider the future return, and when $\gamma \longrightarrow 0$, the controller mainly considers immediate return [21].

Whether in theoretical research or in engineering practice analysis, road traffic density is an effective indicator for measuring the degree of traffic congestion. The operation of

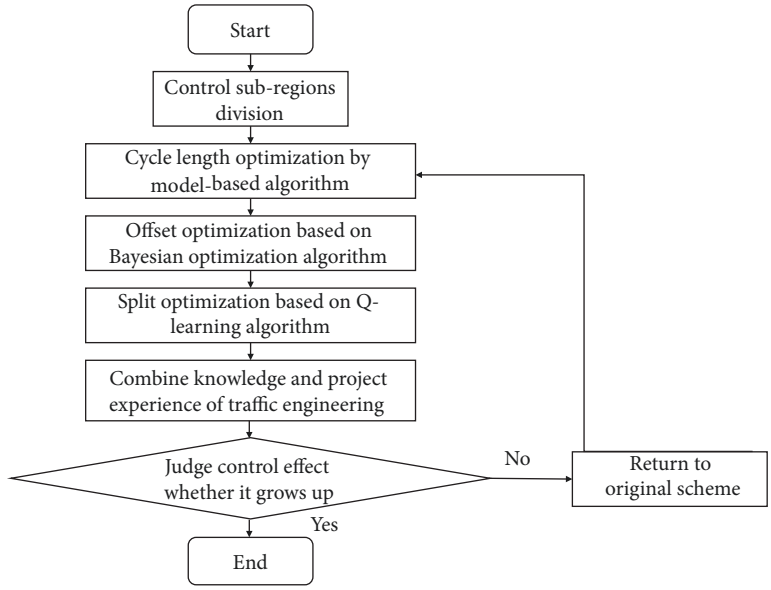

Figure 1: The flow chart of the hierarchical control algorithm based on reinforcement learning.

the traffic on the section is affected by the signal control of the upstream and downstream intersections. The release signal at the upstream intersection directly changes the density of the section, which indirectly affects the traffic capacity and saturation at the section of the stop line and indirectly affects the density of the queue section. The mutual influence of the two is especially noticeable in the supersaturated state. Since the penetration rate of the connected vehicles in different sections is unknown, it is impossible to visually reflect the actual flow of the road through the number of discrete connected vehicles. Even by expanding the sample, it is difficult to guarantee accuracy, but it can clearly reflect the speed of the overall traffic flow. Therefore, this paper uses traffic density as the core parameter to provide a basis for green split optimization.

3.4. The Flow of the Control Algorithm. Firstly, according to different evaluation indexes, the optimal cycle is obtained by using model-based algorithm. Using the combination of the average travel time of vehicles and the Bayesian optimization method based on the Gaussian process, which is commonly used in the optimization of machine learning algorithms, the arterial coordination control is set. The phase offset is optimized by the two-way flow ratio of the upstream and downstream roads and the reasonable setting of the pedestrian crossing phase. Then set different green wave bandwidths to match the upstream and downstream traffic of the morning rush hour and the tidal phenomenon with uneven travel speed. The intelligent algorithm such as Q-learning is used to optimize the green split of each intersection by using key traffic flow parameters at each intersection.

In conclusion, the flow of the traffic signal control strategy based on reinforcement learning is as shown in Figure 1.

4. Verification of Traffic Signal Control Strategy Based on Reinforcement Learning

4.1. Verification of Single Intersection Signal Control Strategy Based on Reinforcement Learning. The intelligent algorithm such as Q-learning is used to optimize the green split of single intersection by using key traffic flow parameters at the intersection. We have compared the Q-learning control method adopted in this paper with the traditional timing signal control and the adaptive control method in second-generation traffic signal control system. The delay of intersection means the average delay for all vehicles passing through all of lane groups at the intersection in the same cycle. The results are shown in Figure 2.

Having compared the traditional timing control with the traffic signal control method based on Q-learning algorithm applied in this paper, the study has shown that the application of Q-learning control method has achieved good performance. In terms of effectiveness, compared with the traditional timing control, the optimization effects of traffic signal control based on Q-learning have, respectively, reached 31.68%, 30.10%, 37.59%, 38.07%, 40.69%, and 43.89%, which has shown that compared with traditional timing control, the traffic signal control based on Q-learning can achieve better optimization effects. However, compared with the existing traffic control strategy, the optimization effects of traffic signal control based on Q-learning have, respectively, reached -4.21%, -5.28%, 3.14%, 6.23%, 13.11%, and 9.72%. The optimization effects of traffic signal control based on reinforcement learning are inferior under low flow conditions, while they are better under medium and high flow conditions.

4.2. Verification of Arterial Intersection Signal Control Strategy Based on Reinforcement Learning. The green wave coordinated control has three important parameter conditions: the

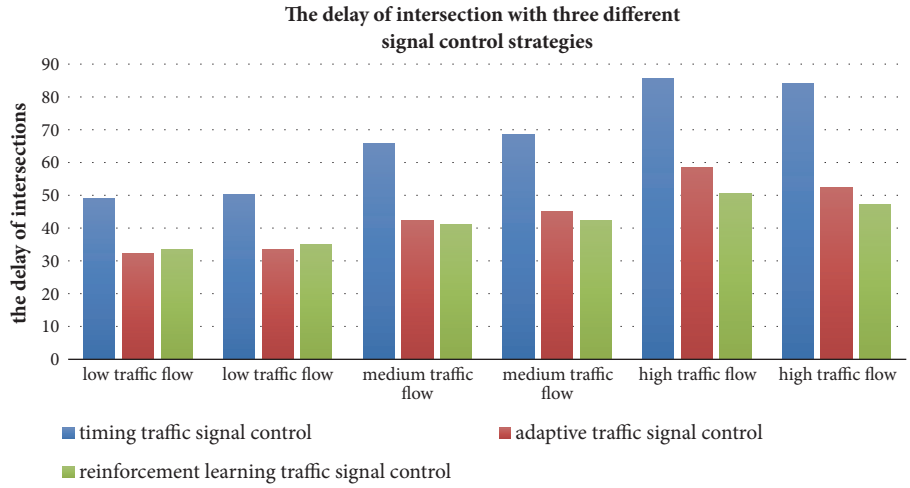

Figure 2: Comparisons of traffic control methods of single intersections.

signal clocks of each intersection should be synchronized; the signal cycle should be the same and have phase offset (the travel time calculated by the adjacent intersection based on the actual average speed). Only with these three conditions can the validity of green wave be guaranteed.

Discrete connected vehicle trajectory data cannot directly obtain the data required for signal distribution and intersection channelization scheme under traditional conditions but can obtain more detailed and complete trajectory-level data. The complete physical trajectory of the vehicle during driving can not only reflect the driving path of the vehicle on the road network, but also reflect the changing characteristics of the vehicle speed with time and space. It is the most comprehensive and complete expression form of the traffic flow operating state, containing a wealth of traffic flow information which is key parameter for offset optimization (for example, travel speed, queue length, delay, and stop times).

For the coordinated control problem of the supersaturation state during morning rush hour, the above research basically follows the idea based on the strong mathematical hypothesis model, but the control system deviates from the original trajectory due to the inaccuracy based on the strong mathematical hypothesis model and the interference from the outside of the control system. In response to these shortcomings, some scholars have further proposed the idea of predictive control, which enables the system to correct the trajectory deviation in real time and achieve the purpose of optimal control. However, the establishment of the optimal control model is still a centralized processing idea. In the application of intersection control problems, it focuses on the control problem of single-point intersections. From the perspective of the structure of the control algorithm, the hierarchical control structure can integrate more control personnel's design ideas, which is of great help to solve the problem of complex control state of the road network. With the development of intelligent control technology, Bayesian optimization methods based on Gaussian process, fuzzy control, reinforcement learning calculation, and neural network have been also widely used in traffic control. However, these applications present a similar feature that is loosely integrated with the actual traffic condition. The computational speed of the online control system is still a big obstacle, and it is more difficult to push it to practical applications. Therefore, traffic control at the network level for rush hours should be based on offline large-scale optimization calculation based on traffic model (based on travel time, then obtaining the relative phase offset between adjacent intersections) and intelligent algorithm (Bayesian optimization methods based on Gaussian process), seeking to achieve a system-optimized phase offset timing scheme.

On the other hand, in the traditional arterial coordinated control scheme, the green wave velocity, the forward green wave bandwidth, and the reverse green wave bandwidth between the starting point and the ending point are always the same or almost the same, and no or less consideration is given to the individualization of the velocity distribution between the sections and the tidal of the traffic flow. Therefore, we adopted different green wave speed optimization methods for different road sections, combined with the unique tidal phenomenon. In the direction of large traffic flow, based on the calculated green wave bandwidth, the bandwidth of the reverse green wave is appropriately increased to match the traffic demand of rush hour. At the same time, the vehicle traffic phase is, respectively, covered to the forward and reverse two-way green wave at the pedestrian crossing, which minimizes the probability that the vehicle stops at the signal control pedestrian crossing.

As is shown in Figure 3, the comparison of original signal control scheme and hierarchical traffic signal control method in different traffic flows has been given. Compared with the original signal control scheme, the arterial traffic signal optimization control method based on hierarchical traffic signal control performs better in medium and high traffic flow. The average delay per vehicle of the original signal control scheme is, respectively, 28.47 seconds, 40.34 seconds, and 61.38 seconds. While with regard to the arterial traffic signal optimization control method based on hierarchical traffic

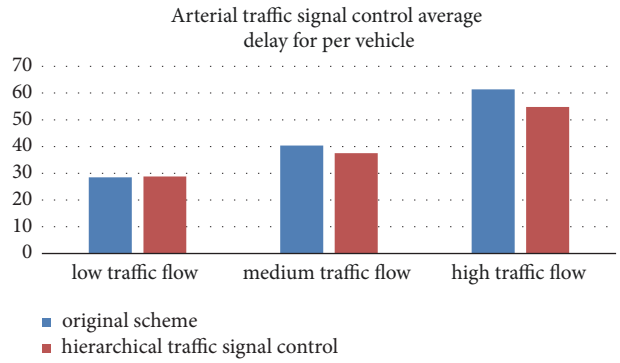

Figure 3: Evaluation of the traffic control method of arterial multiple intersections.

signal control, the average delay per vehicle is, respectively, 28.77 seconds, 37.56 seconds, and 54.79 seconds. Then, the optimization percentage is, respectively, -1.05%, 6.89%, and 10.74%. However, due to the randomness of traffic flows, although there is some increasing average delay in low traffic flow, the general trend has similar conclusions.

4.3. Verification of Regional Traffic Signal Control Strategy Based on Reinforcement Learning. The concept of regional signal control can be divided into a narrow sense and a broad sense. In a narrow sense, regional signal control is a signal control method that unifies several intersections with strong correlation and carries out mutual coordination, namely, the so-called regional signal coordination control. In a broad sense, regional signal control refers to the monitoring of all intersections within the region under the management of a command and control center. It is a comprehensive signal control for single isolated intersection, multiple intersections of the arterial and the highly connected intersection group. It can be classified according to control strategy (timed offline control system, adaptive online control system), control mode (scheme selection, solution generation), and control structure (centralized, distributed).

The purpose of vehicle path feature extraction is to obtain the information of the nodes (i.e., intersections and OD points) that each vehicle passes through, so as to be able to calculate other dynamic features of the sections and road networks (such as traffic, average speed, and road network OD matrix). However, the trajectory data does not contain information such as when the vehicle passed through which node, and we only know the coordinate points of the vehicle trajectory. Then, can we extract the vehicle trajectory by using the trajectory coordinate points and the node information?

In the beginning, we tried clustering based on clustering methods, trying to cluster the coordinate points according to the separated road segments. However, after experimenting with various mainstream clustering methods, it is found that clustering cannot solve the tagging problem of coordinate points. How do we mark the vehicle coordinate points with the tag of the node? After further experimentation, we thought that the vehicle path extraction can be carried out by using the function inpolygon that comes with MATLAB. The core idea is as follows:

(1) Taking each node as the center, the regular polygons that can cover a certain section range are constructed, respectively. These regular polygons represent the in-out range of the node, which is referred to as the node regular polygon. If a coordinate point is within a node regular polygon, it is considered to belong to the node; otherwise, it does not belong to the node.

(2) Traverse the trajectory data and find out which node the coordinate points belong to. If a coordinate point does not belong to all nodes, it means that the coordinate point is on the road segment and special labeling is carried out.

(3) Compress the trajectory information of each vehicle, and record the starting line number of each vehicle passing through the node and the number of coordinate points at each node.

(4) Continue to compress the trajectory information of each vehicle, record all the nodes that each vehicle passes through (in the form of a string and a vector), and derive the OD characteristics of the vehicle.

In this paper, the method of signal control scheme improvement is proposed. According to the different traffic flow characteristics, the subregions are divided. Based on the three key parameters of cycle, arterial coordination signal offset, and green split, a set of hierarchical control algorithms based on reinforcement learning is constructed to optimize and improve the current signal timing scheme. Firstly, according to different evaluation indexes, the optimal period is obtained by using model-based algorithm. Using the combination of the average travel time of vehicles and the Bayesian optimization method based on the Gaussian process, which is commonly used in the optimization of machine learning algorithms, the arterial coordination control is set. The phase offset is optimized by the two-way flow ratio of the upstream and downstream roads and the reasonable setting of the pedestrian crossing phase. Then set different green wave bandwidths to match the upstream and downstream traffic of the rush hour and the tidal phenomenon with uneven travel speed. The intelligent algorithm such as Q-learning is used to optimize the green split of each intersection by using key traffic flow parameters such as traffic flow, density, and speed at each intersection. In the end, this paper uses the hierarchical traffic signal control algorithm based on reinforcement learning and combines the relevant knowledge of traffic engineering and engineering project experience to fine-tune the phase offset and green split to solve the problem of green wave bottleneck point of the arterial and signal interference caused by the right-turning vehicle and then obtain the optimal solution.

There are four key evaluation indexes included the number of vehicles leaving the road network at the end of the simulation, and the total delay time, the total travel time, and the total stopping number have been reduced to varying degrees shown in Figure 4.

For the different traffic flow states, the optimization effect caused by the Q-learning signal control method proposed in this paper in the high flow conditions (32.73%) is better than that in the medium flow and low flow conditions (22.32%,

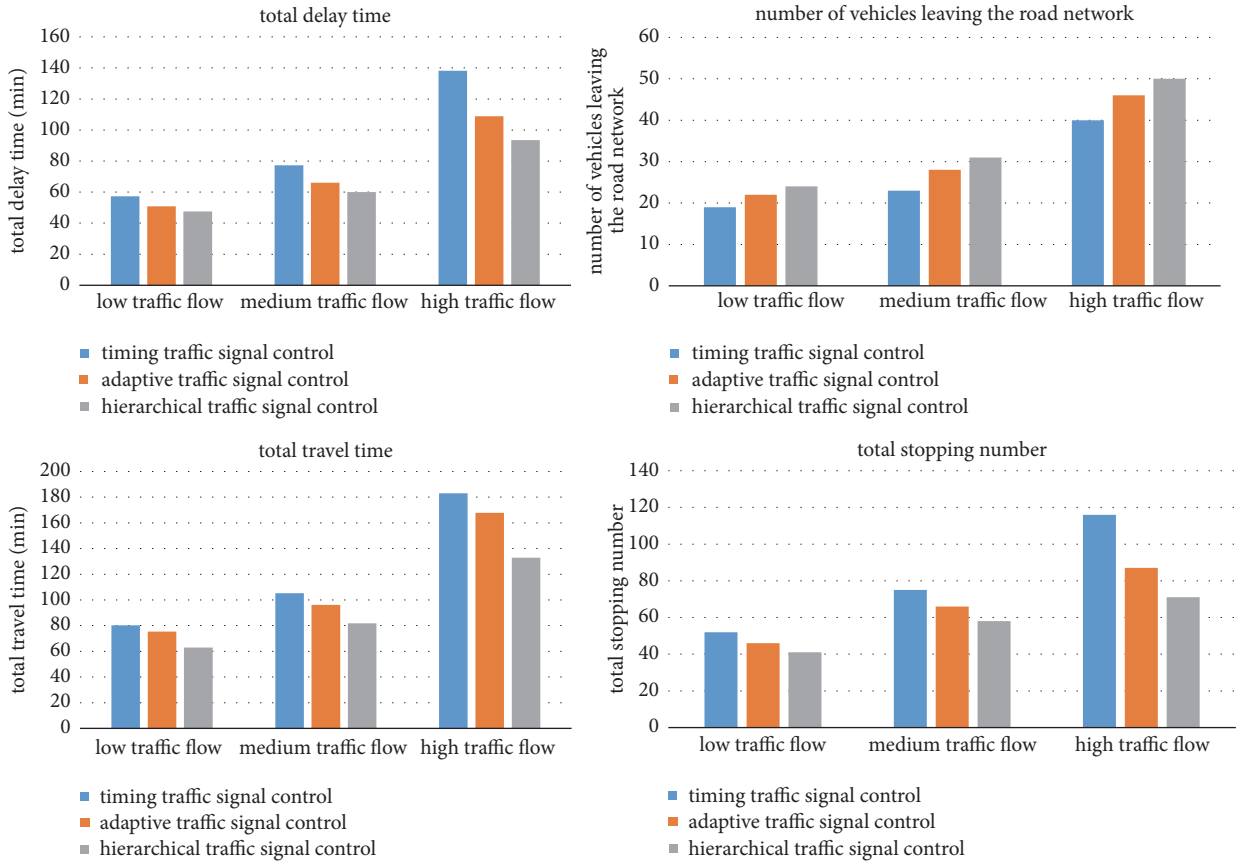

Figure 4: Evaluation of the traffic control method of regional road network.

17.11%). This also proves that the traffic control strategy based on reinforcement learning is more suitable for complex traffic environment (medium and high traffic flow, multi-intersection).

4.4. Analysis of Simulation Results. Through the above comparisons and analysis, it can be concluded that the traffic signal optimization strategy based on reinforcement learning is not applicable for all traffic environments. As to single intersections and arteries, their control effects are inferior to the current adaptive traffic signal control strategy in the low flow conditions. However, the traffic signal optimization strategy based on reinforcement learning is suitable for complex traffic environments (high flows and multiple intersections), and the effects of which are better than the current optimization methods in the conditions of high flows, as to both single intersections and arteries. In the future, we will focus on the network, continue to study the network traffic signal optimization method based on reinforcement learning, and then compare the effects with traditional optimization algorithms.

5. Conclusion and Discussion

This paper uses the hierarchical traffic signal control algorithm based on reinforcement learning and combines the relevant knowledge of traffic engineering and engineering project experience to fine-tune the phase offset and green split to solve the problem of green wave bottleneck point of the arterial and signal interference caused by the right-turning vehicle and then obtain the optimal solution.

In terms of the temporal dynamics of traffic control, reinforcement learning does not have complex optimizing modules and instant decisions can be made to respond to the uncertainty of time-varying traffic flow according to the characteristics of traffic flow observed in real time, which also corresponds with the actual conditions. Therefore, this paper focuses on the application of reinforcement learning in the field of traffic control and concludes that the traffic control method based on reinforcement learning has a better applicability in the complex traffic environment (high flows and multiple intersections), but it is not applicable to all traffic conditions. Furthermore, different from single intersection signal control, facing the integrated control for mainline and networked level traffic, it is still necessary to make further analysis in the aspects of data models and samples, coordination optimization techniques and multi-agent strategies, and mechanism analysis of the interaction between heuristic guidance and higher-level optimization mechanisms such as the pure stochastic optimization and hierarchical algorithm.

Data Availability

According to the data support, the authors have obtained data in the field and simulated them by VISSIM—C# to implement secondary development with kernel algorithm.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to acknowledge the Intelligent Transportation System Research Center of Tongji University for data support. The research is supported by Project of National Natural Science Foundation of China (Project No. 61773293 and No. 61773288) and Key Project of National Natural Science Foundation of China (Project No. 51238008).

References

[1] N. H. Gartner, C. Stamatiadis, and P. J. Tarnoff, "Development of advanced traffic signal control strategies for intelligent transportation systems: multilevel design," *Transportation Research Record*, no. 1494, pp. 98–105, 1995.

[2] A. G. Sims, "The Sydney coordinated adaptive traffic system," in *Proceedings of the Engineering Foundation Conference on Research Directions in Computer Control of Urban Traffic Systems*, Calif, USA, 1979.

[3] Y.-T. Wu and C.-H. Ho, "The development of Taiwan arterial traffic-adaptive signal control system and its field test: A Taiwan experience," *Journal of Advanced Transportation*, vol. 43, no. 4, pp. 455–480, 2009.

[4] R. Vincent, "Safer signalized junction design and self-optimizing control," *Journal of Advanced Transportation*, vol. 28, no. 3, pp. 217–226, 2010.

[5] H. K. Lo and H. F. Chow, "Adaptive traffic control system: Control strategy, prediction, resolution, and accuracy," *Journal of Advanced Transportation*, vol. 36, no. 3, pp. 323–347, 2010.

[6] C. Meneguzzer, "Dynamic process models of combined traffic assignment and control with different signal updating strategies," *Journal of Advanced Transportation*, vol. 46, no. 4, pp. 351–365, 2012.

[7] K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "Store-and-forward based methods for the signal control problem in large-scale congested urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 2, pp. 163–174, 2009.

[8] J. Hao, *Studies of the application of data-driven control method in traffic control*, Beijing Jiaotong University, Beijing, China, 2013.

[9] D. Ma, X. Luo, S. Jin, W. Guo, and D. Wang, "Estimating Maximum Queue Length for Traffic Lane Groups Using Travel Times from Video-Imaging Data," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 3, pp. 123–134, 2018.

[10] D. Ma, X. Luo, S. Jin, D. Wang, W. Guo, and F. Wang, "Lane-based saturation degree estimation for signalized intersections using travel time data," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 3, pp. 136–148, 2017.

[11] Y. Wang, X. Yang, H. Liang, and Y. Liu, "A Review of the Self-Adaptive Traffic Signal Control System Based on Future Traffic Environment," *Journal of Advanced Transportation*, vol. 2018, Article ID 1096123, 12 pages, 2018.

[12] D.-F. Ma, D.-H. Wang, F. Sun, Y.-M. Bie, and S. Jin, "Method of spillover identification in urban street networks using loop detector outputs," *Journal of Central South University*, vol. 20, no. 2, pp. 572–578, 2013.

[13] D. Ma, D. Wang, Y. Bie, S. Jin, and Z. Mei, "Identification of spillovers in urban street networks based on upstream fixed traffic data," *KSCE Journal of Civil Engineering*, vol. 18, no. 5, pp. 1539–1547, 2014.

[14] Littman M. L., "Markov games as a framework for multi-agent reinforcement learning," in *Proceedings of The 11th International Conference on Machine Learning*, pp. 157–163, 1994.

[15] S. J. Bradtke and Michael O. D., "Reinforcement learning methods for continuous-time Markov decision problems," *Advances in Neural Information Processing Systems*, pp. 393–400, 1995.

[16] D. Ma, F. Fu, S. Jin et al., "Gating control for a single bottleneck link based on traffic load equilibrium," *International Journal of Civil Engineering*, vol. 14, no. 5, pp. 281–293, 2016.

[17] W. Yang, L. Zhang, Y. Shi, and M. Zhang, "Applications of Agent Technology in Urban Traffic Signal Control Systems: A Survey," *Journal of Wuhai University of Technology (Transportation Science Engineering)*, vol. 38, no. 4, pp. 709–718, 2014.

[18] L. Zhang, W. Yang, J. Wang, and Q. Rao, "Large-scale agent-based transport simulation in shanghai, china," *Transportation Research Record*, vol. 2399, no. 1, pp. 34–43, 2013.

[19] M. Aslani, S. Seipel, and M. Wiering, "Continuous residual reinforcement learning for traffic signal control optimization," *Canadian Journal of Civil Engineering*, vol. 45, no. 8, pp. 690–702, 2018.

[20] C. J. Watkins, *Learning from delayed rewards*, University of Cambridge, UK, 1989.

[21] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279–292, 1992.