

Research Article

Optimizing the Junction-Tree-Based Reinforcement Learning Algorithm for Network-Wide Signal Coordination

Yi Zhao , Jianxiao Ma, Linghong Shen, and Yong Qian

College of Automobile and Traffic Engineering, Nanjing Forestry University, Nanjing 210037, China

Correspondence should be addressed to Yi Zhao; zhaoyi207@126.com

Received 31 October 2019; Revised 1 January 2020; Accepted 24 January 2020; Published 21 February 2020

Guest Editor: Zeyang Cheng

Copyright © 2020 Yi Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study develops three measures to optimize the junction-tree-based reinforcement learning (RL) algorithm, which will be used for network-wide signal coordination. The first measure is to optimize the frequency of running the junction-tree algorithm (JTA) and the intersection status division. The second one is to optimize the JTA information transmission mode. The third one is to optimize the operation of a single intersection. A test network and three test groups are built to analyze the optimization effect. Group 1 is the control group, group 2 adopts the optimizations for the basic parameters and the information transmission mode, and group 3 adopts optimizations for the operation of a single intersection. Environments with different congestion levels are also tested. Results show that optimizations of the basic parameters and the information transmission mode can improve the system efficiency and the flexibility of the green light, and optimizing the operation of a single intersection can improve the efficiency of both the system and the individual intersection. By applying the proposed optimizations to the existing JTA-based RL algorithm, network-wide signal coordination can perform better.

1. Introduction

Signal control system is an important method of improving the operation of urban traffic. With the development of people's understanding on traffic and technology, urban traffic signal control systems have undergone three stages: single-point, linear coordinated, and regional coordinated. Traffic signal coordination is considered to be more effective in alleviating traffic congestion than single-point and linear coordinated.

1.1. Review of the Literature on Signal Coordination.

Signal coordination has been studied extensively over the past 30 years. The first developed signal coordination control systems include SCOOT [1], SCATS [2], PROLYN [3], OPAC [4], RHODES [5], UTOPIA [6], CRONOS [7], and TUC [8]. Although the signal coordination control can achieve better effects than the single-point signal control and the inductive signal control, there are also many restrictions on the signal coordination control, such as difficulty in

parameter calibration, computational complexity, and poor adaptability and stability.

Considering these restrictions and the fact that the dynamic characteristics of the traffic environment also provide the need for interactive environment-based learning from the environment, machine learning algorithms are proposed to be used in signal coordination control research. Among the machine learning algorithms, the reinforcement learning (RL) algorithm is the most widely used in the field of traffic signal control.

Liang et al. [9] proposed a deep reinforcement learning model to control the traffic light cycle. Aslani et al. [10] introduced the actor-critic method to solve the problem of the trade-off between exploration of the traffic environment and exploitation of the knowledge already obtained. Aslani et al. [11] developed adaptive traffic signal controllers based on continuous residual reinforcement learning to improve their stability. Jeon et al. [12] suggested a novel artificial intelligence that only uses video images of an intersection; the image-based RL model outperformed both the actual operation of fixed signals and a fully actuated operation. Aziz

et al. [13] applied R-Markov Average Reward Technique-based reinforcement learning algorithm for vehicular signal control problem leveraging information sharing among signal controllers in the connected vehicle environment. Darmoul et al. [14] suggested a Immune Network Algorithm-based Multiagent System to control a network of signalized intersections, which is able to handle different traffic scenarios.

Graph theory models can reduce the computational complexity of RL, especially when joint action of multiagents needs to be calculated. But not much research has been done in this area. Some work has included developments in the max-plus algorithm and junction-tree algorithm (JTA); these have been applied to signal coordination control research at the road network level.

Medina and Beenekohal [15] applied the max-plus algorithm as a coordinating strategy in the network-wide signal control problem. However, the max-plus algorithm has two key limitations. Firstly, it is only applicable to tree-structured networks and cannot guarantee the convergence to an optimal solution for general cyclic networks. Secondly, this algorithm only provides a brief loopy propagation that refers to inexact messages received at a node. Thus, it only provides an approximate inference of the exact message being passed. Zhu et al. [16] first proposed the JTA instead of the max-plus algorithm to obtain the best joint action for traffic signals and to realize network-wide signal coordination. JTA was first proposed by Jensen et al. [17]. The advantage of JTA is that it is computationally efficient and can handle looped or acyclic road networks and accurately infer the best joint scheme.

1.2. Motivations and Contributions of this Study. Zhu et al. [16] demonstrated that the test network can perform better under the JTA compared to an adaptive or single-agent RL-based control. Although the network system improved, some intersections still experienced poor operations. Zhu et al. [16] also noted that it is necessary to assess the variance of performance metrics at the intersection level, and modified schemes should be developed to optimize the system to ensure desired level of performance at local intersections.

To summarize, the research goals are as follows:

- (1) To optimize the basic parameters of the JTA algorithm so that the signal coordination control scheme is consistent with actual requirements
- (2) To evaluate the impact of existing algorithms on local intersection operations
- (3) To propose optimization measures for local intersections to improve the practical application value of the algorithm

2. Introducing the Junction-Tree-Based RL Algorithm

2.1. Reinforcement Learning (RL) and Its Application in Signal Control. The basic RL model is shown in Figure 1. It

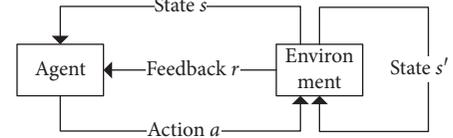


FIGURE 1: The basic reinforcement learning model.

contains an environment, agents, learners, and strategies. The agent obtains the state “s” from the environment and selects action “a” according to the state. The action “a” interacts with the environment, which then returns to a new state “s’” and sends a certain feedback “r” to the agent. After repeated interactions, the agent can learn an optimal strategy for the situations presented.

In the application of RL to traffic signal control, the road network is the environment and the signal control machine is the agent. During the decision period, the signal control machine takes an action to activate a signal phase, and the state of the environment changes accordingly. The goal of the algorithm is to obtain the optimal strategy that can achieve the maximum return. The optimal strategy is to map the activation phase and state of the traffic. The feedback can include average delay and the number of stops. Its value can be extracted directly from the environment.

2.2. Junction-Tree Algorithm and Application in Signal Control. The key idea of the JTA is to find a way to decompose the global computation of joint probability into a set of related local computations. The JTA is introduced to reveal the important connections between global and local probabilistic reasoning using graph theory.

The essence of the JTA is information transmission. The forward transmission is the transfer from the root node to the leaf node, while the reverse transmission is from the leaf node to the root node. The process of information transfer can be expressed by equations (1)–(4).

Forward transmission from v to s :

$$(\psi'_v, \phi'_s) = \arg \max_{v/s} \psi_v. \quad (1)$$

Forward transmission from s to w :

$$\psi'_w = \frac{\phi'_s}{\phi_s} \psi_w. \quad (2)$$

Reverse transmission from w to s :

$$(\psi''_w, \phi''_s) = \arg \max_{w/s} \psi'_w. \quad (3)$$

Reverse transmission from s to v :

$$\psi''_v = \frac{\phi''_s}{\phi'_s} \psi'_v. \quad (4)$$

In the equations above, v is the root node; w is the leaf node; s is the separation node; ψ_v , ψ_w , and ϕ_s denote potential functions of v , w , and s ; ψ'_v , ϕ'_s , and ψ'_w denote

potential functions after forward transmission; and ψ_v'' , ϕ_s'' , and ψ_w'' denote potential functions after reverse transmission.

JTA and RL have the same objective function in terms of calculating the maximum posteriori probability. They both decompose the whole network optimization problem into local subproblems, and both use their Markov attributes to do so. In the probability model, the probability of a node depends on the adjacent nodes. In the coordinated traffic signal control, the phase selection of the intersection depends on the phase of the adjacent intersection. Therefore, JTA is selected to solve a coordinated traffic signal control problem. JTA has great advantages in dealing with coordinated traffic signal control problems because it is the fastest and most accurate inference algorithm.

2.3. The Junction-Tree-Based RL Algorithm. The control flow of the JTA-based RL algorithm method is shown in Figure 2. In the applied method, the RL is the core algorithm of signal control, and the JTA is used to find the signal control scheme with the highest rate of return. Existing research verifies that the applied method is better than the timing signal control, the independent Q learning signal control, and the maximum queue length priority signal control under different traffic intensities.

It should be noted that the RL algorithm can learn the Q value under specific traffic demand and signal control scheme for one or two adjacent intersections. But, the RL algorithm cannot learn the Q value for the whole network with too many intersections because of the large scale of knowledge to be learned. JTA is adopted to achieve the best signal control scheme so that the Q value for the whole network is the best one. In the proposed algorithm, there is no cycle time and split. If the frequency of running JTA is 1 s, then the algorithm can only decide which phase is green light for each intersection in the next 1 s.

3. Optimizing the Junction-Tree-Based RL Algorithm

3.1. Optimizing Basic Parameters

3.1.1. Frequency of Running the JTA. As the JTA determines the phase-switch at intersections, the lower frequency running it, the longer a given phase duration will be. To adjust the signal control scheme according to feedback in time, the frequency to run the JTA should not be lower than the headway of queueing vehicles passing the parking line.

Both Shao et al. [18] and Zhao et al. [19] have verified that the headway is less than 2 s when the queue length is longer than 10 vehicles. However, in existing research on JTA, the frequency is 5 s which cannot meet actual control requirements. In order to improve the sensitivity of the signal control scheme, and considering the minimum step size of the signal control scheme, 1 s is employed in this study.

3.1.2. Intersection Status Division. The JTA-based RL algorithm selects the phase scheme with the highest return according to the state of the road network. Phase schemes

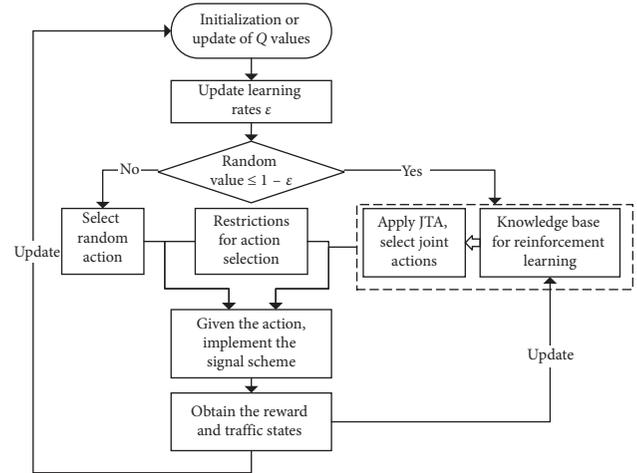


FIGURE 2: The control flow of the JTA-based RL algorithm method.

are determined by the number of intersections and the phases of a single intersection, which are relatively fixed. Therefore, the accuracy of the applied method for signal control is determined by the state of the road network. However, the large number of intersections available when signal coordination control is performed provides a status division that is too detailed and may lead to a long learning time. Existing studies treat the saturation as the evaluation index of intersection entrance, and saturations of all phases are summed and divided into three levels. That is, each intersection contains three states, and the state of two adjacent intersections is divided into nine. In general, this state division is rough and makes the signal control scheme less sensitive to the traffic state of the road network.

Considering that the state will be defined as an eight-dimensional vector in the program of the applied method, the saturation of each intersection entrance is divided into three levels, and then each intersection is divided into 81 states. In future applications, the status of the intersection can be divided in more detail based on specific requirements.

3.2. Analysis of the JTA Information Transmission Mode. The JTA uses the continuity function while calculating the maximum posteriori probability, which should not be directly applied to the information transmission in traffic signal coordination control. Therefore, a new information transmission mode that will be applied in signal coordination control is defined. The new transmission mode, taking four intersections as the example, is shown as follows.

Suppose that all four intersections have only two phases, A and B; phase A is for north-south traffic, and phase B is for east-west traffic. The virtual road network can be transferred into a junction tree using moralization and triangulation, see Figure 3. Intersections 1–3 form a root node; intersections 2–4 form a leaf node, and intersections 2 and 3 form a separation node. The key parameter Q is the value of two adjacent intersections and is shown in Table 1.

The target function of JTA is $\arg \max(Q_{12} + Q_{13} + Q_{24} + Q_{34})$.

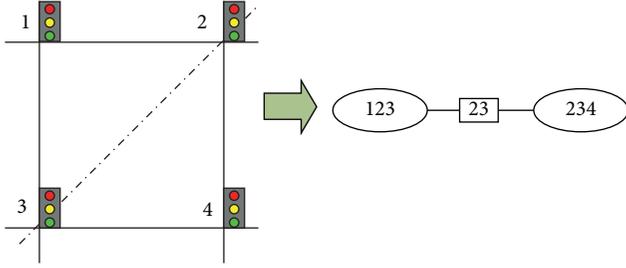


FIGURE 3: Virtual road networks and the corresponding junction tree.

3.2.1. Initialization: Define the Potential Function of all Nodes. The potential functions of the root and leaf nodes are the sum of the Q values of three intersections that form the node. The potential function of the separation node is the phase combination of two intersections that form the node; the initial value is null.

The potential function of the root node is $\psi_{123} = Q_{123} = Q_{12} + Q_{13}$

The potential function of the separation node is $\phi_{23} = \text{Null}$

The potential function of the leaf node is $\psi_{234} = Q_{234} = Q_{24} + Q_{34}$

3.2.2. Forward Transmission from the Root Node to the Separation Node. The transmission function is $(\psi'_{123}, \phi'_{23}) = \arg \max \psi_{123} = \arg \max (Q_{12} + Q_{13})$.

After transmission, ψ_{123} should achieve the max value ψ'_{123} under all possible potential functions ϕ_{23} and also achieve the best phase combination ϕ'_{23} . The transmission result is shown in Table 2.

3.2.3. Forward Transmission from the Separation Node to the Leaf Node. The transmission function is $\psi'_{234} = \phi'_{23} \psi_{234} = \phi'_{23} (Q_{24} + Q_{34})$.

After transmission, the potential function of leaf node ψ_{234} changes to ψ'_{234} .

3.2.4. Reverse Transmission from the Leaf Node to the Separation Node. The transmission function is $\arg \max (Q_{24} + Q_{34})$, $(\psi''_{234}, \phi''_{23}) = \arg \max \psi'_{234} = \arg \max \phi'_{23} (Q_{24} + Q_{34})$.

After transmission, ψ'_{234} should achieve the max value ψ''_{234} under all possible potential functions ϕ'_{23} and the best phase combination ϕ''_{23} . The transmission result is shown in Table 3.

By combining ϕ''_{23} and $\arg \max (Q_{24} + Q_{34})$, it is easy to understand that $\phi''_{23} (Q_{24} + Q_{34})$ achieves the maximum value only when ϕ''_{23} selects combination 4. In other words, $\phi''_{23} (Q_{24} + Q_{34})$ can achieve the maximum value only when intersections 2, 3, and 4 are all in phase B; at the same time, ψ''_{234} must be 13.

3.2.5. Reverse Transmission from the Separation Node to the Root Node. The transmission function is $\psi''_{123} = \phi''_{23} \psi'_{123}$.

After transmission, ψ'_{123} changes to ψ''_{123} based on ϕ''_{23} . At this time, ψ''_{123} is 16, and intersection 1 is in phase B. The

result of applying JTA is obtained after the above information transmission occurs, that is, after the joint action of the four intersections becomes (B, B, B, B), which will result in the joint tree achieving its highest potential function.

3.3. Optimizations for Single Intersection's Operation. Network-wide signal coordination control both pursues the system optimization and the requirements of the individual intersection. For example, the queue length of a single intersection entrance should not be too long when the network has a low average queue length. The JTA-based RL algorithm considers system optimization to be the goal; however, this tends to cause the queue lengths of some entrance lanes to be too long.

To improve the performance of single intersections, optimization should be studied.

3.3.1. Information Transmission Rule-Based Optimization.

In the JTA-based RL algorithm, the root and leaf nodes determine the direction of information transmission along the junction tree. Existing study, Zhu et al. [16], simply assigns the endpoints of the junction tree as the root and leaf nodes, without considering the signal control requirements. Analyses of the JTA information transmission modes show that the intersection's phase is determined in the reverse transmission process. For these reasons, it is proposed that the phase of the intersection with poor operation should be determined first. Therefore, the worst running node should be taken as the leaf node while all endpoints of the junction tree are taken as root nodes. The information transmission rule before and after optimizations is shown in Figure 4.

3.3.2. Differentiated Return-Based Optimization.

System Q value of the JTA-based RL algorithm is determined by the Q values of every two adjacent intersections. For example, A and B are adjacent intersections, and entrances of two connecting sections between A and B are saturated with a and b , and then the Q value of A and B can be expressed as $Q(A, B) = a + b$. When $a = 0.1$, $b = 0.8$, then $Q(A, B) = 0.9$; when $a = b = 0.45$, then $Q(A, B) = 0.9$. Saturations of 0.1, 0.45, and 0.8 indicate different service levels, but there is no difference in calculating $Q(A, B)$; thus, the differences cannot be learned in signal timing. Therefore, the differentiated return-based optimization method is proposed to optimize the definition of Q values.

If the saturation q is taken as the evaluation index and varies from 0 to 1, q should be divided into n levels, and the return of the k^{th} level should be 2^k ($k \in [1, n]$). When the saturations of the adjacent intersections A and B are q_1 and q_2 , q_1 belongs to level k_1 , and q_2 belongs to level k_2 . Therefore, the Q value of the adjacent intersections is expressed as follows:

$$Q(A, B) = q_1 * 2^{k_1} + q_2 * 2^{k_2}, \quad (5)$$

TABLE 1: The given Q -value matrix of the virtual road network.

Phase combination	Intersection and phase		Q_{12}	Intersection and phase		Q_{13}	Intersection and phase		Q_{34}	Intersection and phase		Q_{24}
	1	2		1	3		3	4		2	4	
Combination 1	A	A	4	A	A	8	A	A	7	A	A	8
Combination 2	A	B	5	A	B	7	A	B	6	A	B	9
Combination 3	B	A	3	B	A	5	B	A	4	B	A	5
Combination 4	B	B	7	B	B	9	B	B	6	B	B	7

TABLE 2: ψ'_{123} under possible combinations and corresponding ϕ'_{23} .

Phase combination	Intersection and phase		$\max(Q_{12} + Q_{13})$	Intersection and phase
	2	3		
Combination 1	A	A	12	A
Combination 2	A	B	12	B
Combination 3	B	A	13	A
Combination 4	B	B	16	B

where $Q(A, B)$ is the Q value of adjacent intersections A and B , q_1 and q_2 are the saturations of adjacent intersections A and B , and k_1 and k_2 are the levels of q_1 and q_2 .

4. Test Case Study

4.1. Network Description. This study used VISSIM5.4 to build a virtual road network and test the validity of optimizations on the JTA-based RL algorithm. Details about the modules in VISSIM (e.g., car-following, lane-changing, traffic light control) can be found in the VISSIM manual. The JTA-based RL algorithm is coded in VB.net and interacts with VISSIM through the component object model (COM) interface.

A virtual road network same to the one in Zhu et al.'s study [16] was built. Under the same test environment, the results of this study should be more convincing. The network uses a structure with six horizontal and three vertical roads. The number of lanes is randomly set. There are 18 intersections in the network, and each entrance has an independent left turn lane, as shown in Figure 5. Also, the given network is transformed into a junction tree, as shown in Figure 6.

The length of the road section in the test network is set randomly, and channelization schemes of 18 intersections are also not uniform. All 18 intersections in the test network are coordinated intersections. Four phases are considered: (a) E-W + W-E bound through and right turn, (b) N-S + S-N bound through and right turns, (c) dual left from E-S + W-N bound, and (d) dual left from S-W + N-E bound.

The performance of the JTA-based RL algorithm is tested at three levels of congestion: low, medium, and high. The traffic demand is input into the network through the 18 link origins in Figure 5. The congestion levels are reflected in the ranges of the demand inputs, which are 500 vph to 600 vph, 600 vph to 800 vph, and 900 vph to 1200 vph, respectively.

4.2. Test Group Settings. In the test case, queue length w_{ij} is adopted to build the return and objective functions. The

TABLE 3: Combination forms for $\arg \max(Q_{24} + Q_{34})$.

Phase combination	Intersection and phase		$\max(Q_{24} + Q_{34})$	Intersection and phase
	2	3		
Combination 1	A	A	15	A or B
Combination 2	A	B	15	B
Combination 3	B	A	13	B
Combination 4	B	B	13	B

objective function is created to achieve the shortest queue length for the system. The return function is as follows:

$$f_r^{ij}(t) = \frac{q_{ij}^t}{J_{ij}} \times \frac{1}{l_{ij}}, \quad (6)$$

where $f_r^{ij}(t)$ is the return of intersection i in phase j and time t , q_{ij}^t is the traffic volume of the key entrance of intersection i in phase j and time t , J_{ij} is the density of the key entrance when it is congested, and l_{ij} is the lane length available for queueing of intersection i in phase j .

Three test groups are set to test the effectiveness of optimization methods. The details of the settings are as follows:

Group 1: existing research of Zhu et al. [16] applying JTA in signal coordination

- (1) Frequency of running JTA: 5 s
- (2) Intersection status division: each intersection contains three states, and the state of the two adjacent intersections is divided into nine parts
- (3) JTA information transmission mode: the mode introduced in Section 2.2
- (4) Root and leaf node: $V(1, 2, 4)$ is the root node, and $V(14, 16, 17)$ and $V(15, 17, 18)$ are the leaf nodes
- (5) Q value: calculated without regard to the differentiated returns

Group 2: optimizations of basic parameters and information transmission modes

- (1) Frequency of running JTA: 1 s
- (2) Intersection status division: the saturation of each intersection entrance is divided into three levels, and each intersection is divided into 81 states
- (3) JTA information transmission mode: the mode introduced in Section 3.2
- (4) Root and leaf node: same as group 1
- (5) Q value calculated same as group 1

Group 3: optimizations on the information transmission rule and the return

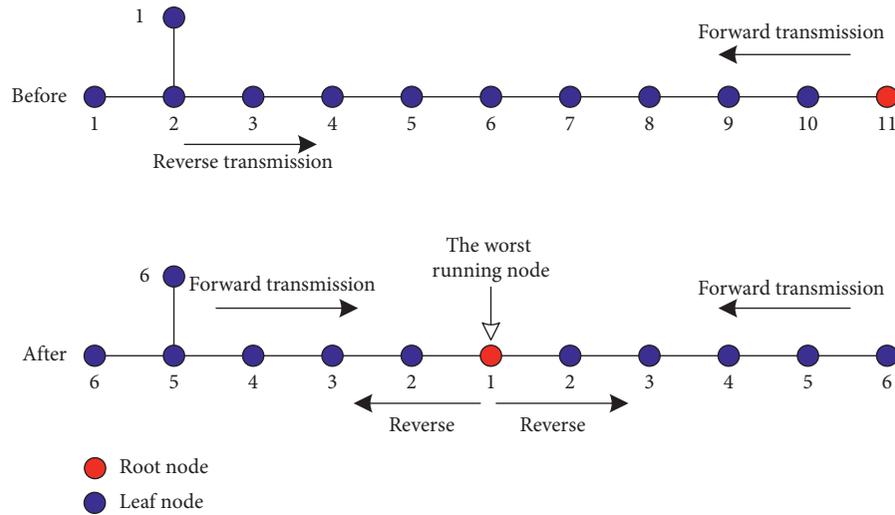


FIGURE 4: Optimization of the JTA information transmission rule.

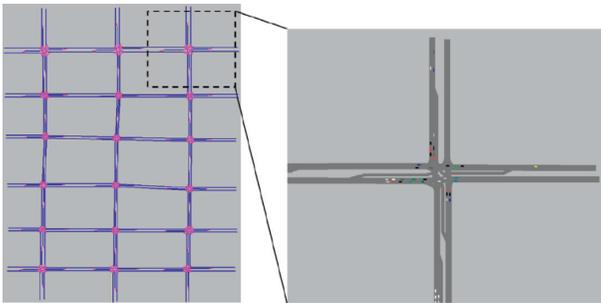


FIGURE 5: Center line representation of the network and a zoomed-in view of one intersection.

- (1) Frequency of running JTA: same as group 2
- (2) Intersection status division: same as group 2
- (3) JTA information transmission mode: same as group 2
- (4) Root and leaf node: the worst running node is taken as the leaf node while all endpoints of the junction tree are taken as root nodes
- (5) Q value-differentiated returns are calculated and applied

In addition to the above settings, the training time of group 1 is 5 h, while that of groups 2 and 3 is 10 h. After training, the three groups are applied in signal coordination; each group contains 10 simulation runs (each with a different random seed), and each simulation lasts 1 h.

The differentiated-return-based optimization method adopted in group 3 is necessary to classify the queue length w_{ij} . This is divided into three levels in this study: the first level is $[0, 0.4)$, the second is $[0.4, 0.7)$, and the third is $[0.7, 1]$. The return of each level is 2, 4, and 8, respectively.

4.3. Test Result Analysis. By comparing the test results of three groups, several conclusions can be drawn as follows.

4.3.1. The Green Light of Each Phase Is More Flexible. Taking intersection 8 as an example, 50 randomly selected continuous phases under medium congestion levels are extracted, and the corresponding green light durations are shown in Figure 7. As the frequency of calling the JTA in group 1 is 5 s, the green time of all phases is a multiple of 5, while the green time in group 2 is not subject to this constraint. The green time in group 2 can be adjusted according to the length of the queue. It can be concluded that optimization of the basic parameters can increase the flexibility of the green light duration, which in turn makes the green light more reasonable.

4.3.2. The Efficiency of Signal Coordination Is Improved. The queue length of the system and the intersection at different congestion levels are shown in Table 4. The queue length of the intersection is the longest queue length of all the entrance lanes while the phase is being switched. The average queue length of the system is the average queue length of all 18 intersections. As the traffic demand is input into the network via link origins, the outermost intersections of the network are directly affected by the traffic input, which may then also affect the evaluation result. Considering the above reasons, only intersections 5, 8, 11, and 14 are selected and analyzed.

In terms of the queue length of the system, the table shows that the length of group 2 is shorter than group 1 by over 10%. It can be concluded that optimizations of basic parameters and the JTA information transmission mode can improve the efficiency of signal coordination. The lengths of group 2 and group 3 are not significantly different, which means that optimizing the operation of a single intersection has little effect on the system operation.

4.3.3. Problems after Parameter Optimization and the Information Transmission Mode Are Still Significant. Optimization methods improve system operation, but the operations of some intersections are still poor. Table 4 shows

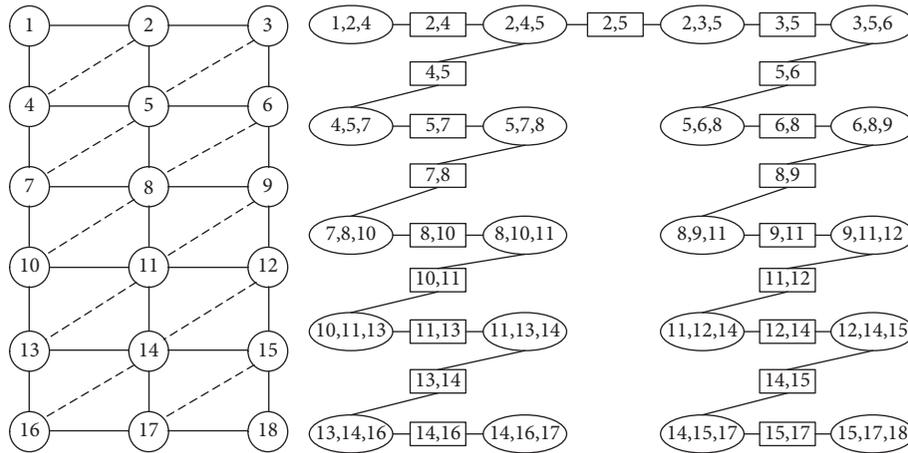


FIGURE 6: Triangulation and junction-tree construction of the test network.

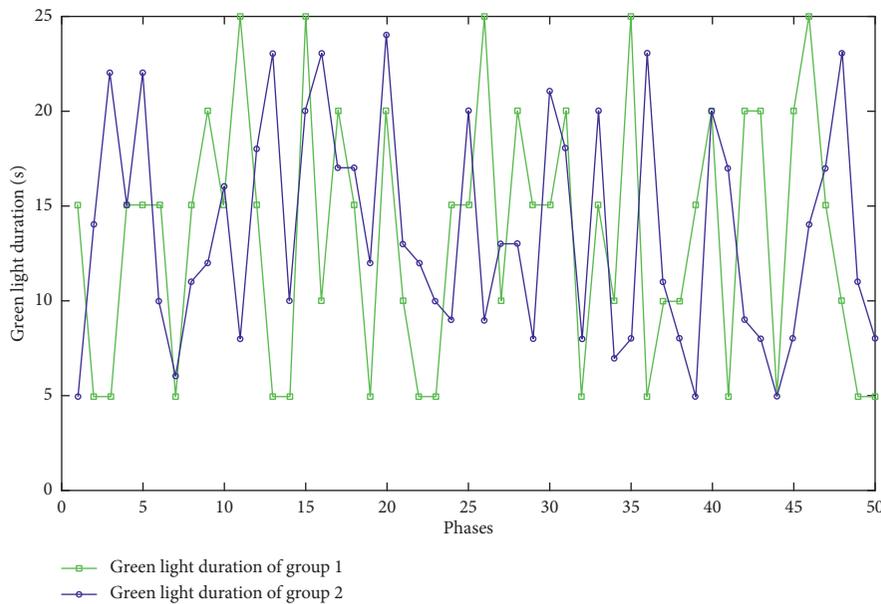


FIGURE 7: The green light duration of intersection 8 in groups 1 and 2.

that the average queue length of some intersections in group 2 is longer than that in group 1; for example, intersection 5 under a low congestion level and intersection 8 under a high congestion level. Queue lengths of 50 randomly selected continuous phases of these two intersections are also shown in Figures 8 and 9. The two figures show intersections with large fluctuations in queue length, such as intersection 5 under low congestion level with a maximum queue length of 0.55 and a minimum queue length of 0.16.

In other words, after optimizing basic parameters and the information transmission mode, the operation of a single intersection still needed to be improved.

4.3.4. Optimizations for Operating Single Intersections Can Reduce the Maximum Queue Length of the System. The maximum queue length of the system under low and

high congestion levels is counted at intervals of 10 s and shown in Figures 10 and 11. It is obvious that the queue length of group 3 is the lowest. In other words, the maximum queue length of the system is reduced after the optimizations for the operation of a single intersection were adopted.

4.3.5. Optimizations for the Operation of a Single Intersection Can Reduce the Fluctuation of the Queue Length at the Intersection. After applying a differentiated return-based optimization, group 3 should be more sensitive towards returns than groups 1 and 2. The queue length of intersection 5 under low congestion levels in different groups can be taken as an example. The variations in the queue lengths are shown in Figure 12. The queue length in group 1 varies from 0.18 to 0.53, group 2 varies from 0.17 to 0.55, and group 3 varies from 0.32 to 0.44. The fluctuation of queue length shows that the

TABLE 4: Queue length-based improvement analysis of optimization methods.

Congestion level	Test group	Average queue length of the system	Average queue length of the intersection			
			Intersection 5	Intersection 8	Intersection 11	Intersection 14
Low	Group 1	0.36	0.35	0.32	0.43	0.37
	Group 2	0.32	0.38	0.30	0.32	0.31
	Group 3	0.33	0.36	0.31	0.35	0.29
Medium	Group 1	0.51	0.50	0.47	0.45	0.53
	Group 2	0.43	0.40	0.42	0.44	0.48
	Group 3	0.44	0.41	0.45	0.42	0.46
High	Group 1	0.83	0.84	0.82	0.76	0.81
	Group 2	0.71	0.73	0.87	0.70	0.69
	Group 3	0.71	0.75	0.81	0.72	0.71

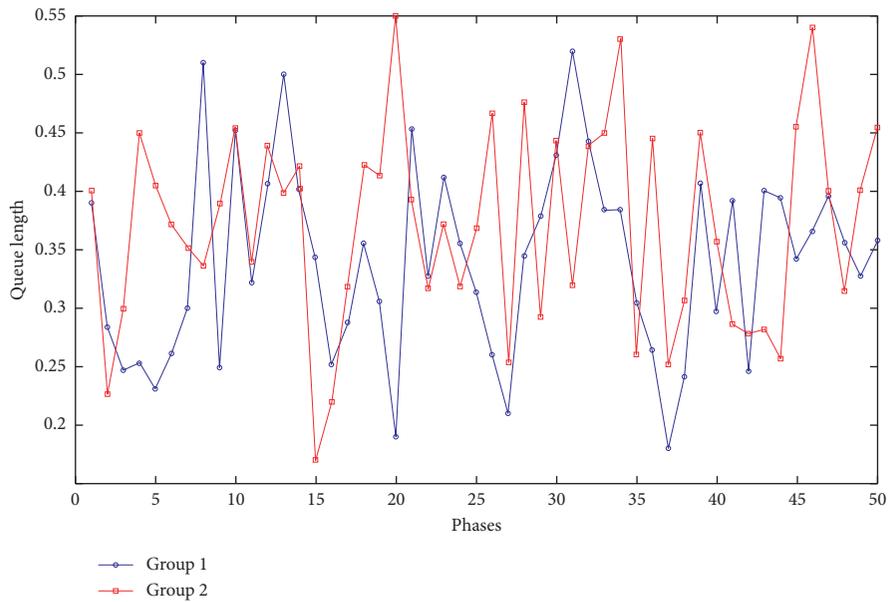


FIGURE 8: Variations in queue length at intersection 5 under low congestion levels.

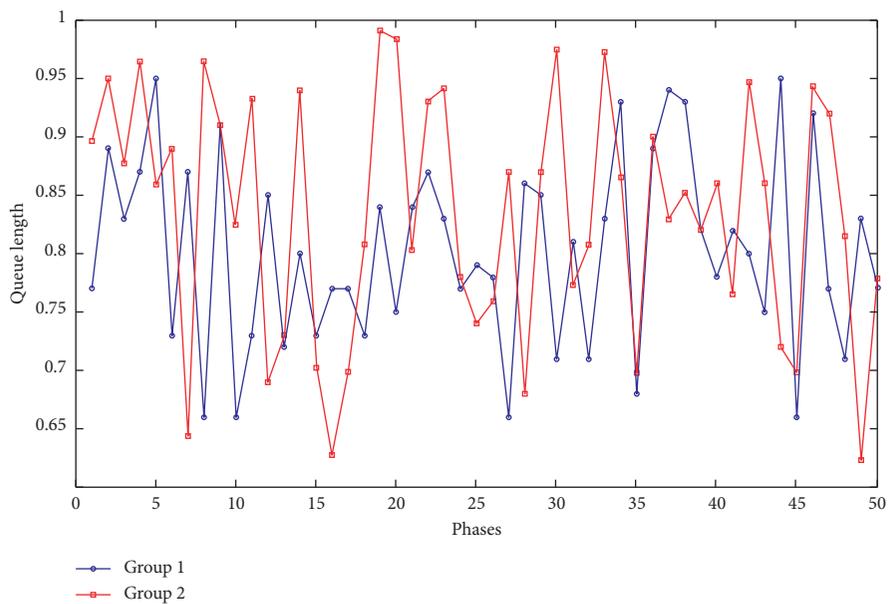


FIGURE 9: Variations in queue length at intersection 8 under high congestion levels.

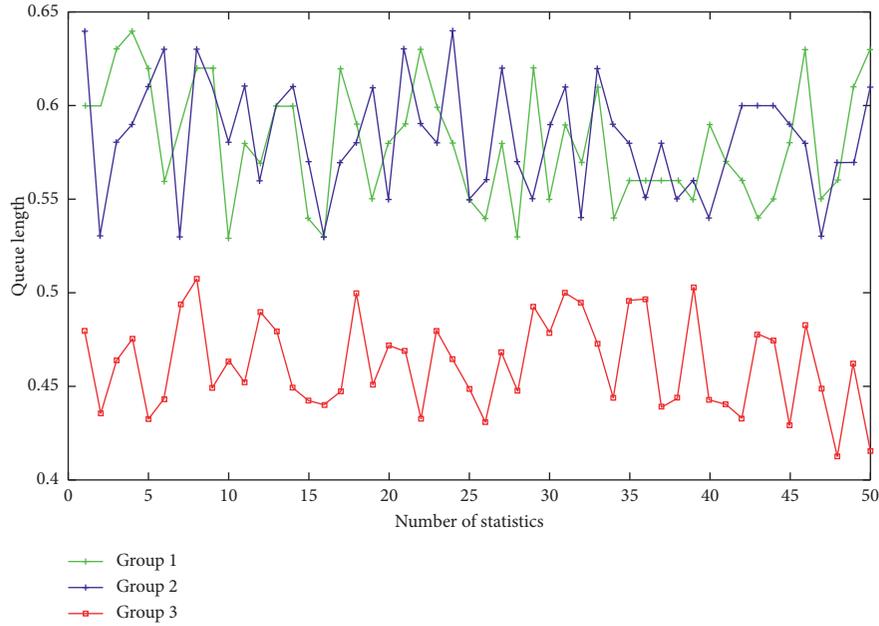


FIGURE 10: The maximum queue length of the system under a low congestion level.

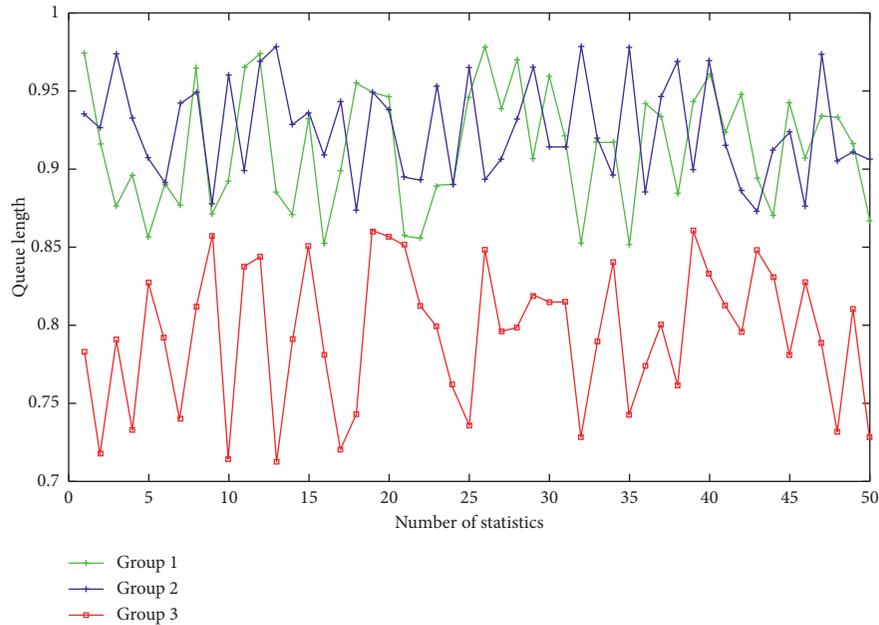


FIGURE 11: The maximum queue length of the system under a high congestion level.

intersections in group 3 have better operations, which benefit from a differentiated return-based optimization.

5. Discussion and Conclusion

The study proposed three optimization methods for the JTA-based RL algorithm which can be used for network-wide signal coordination. Three test groups were built to analyze the optimization effect.

Group 1 used the existing algorithm applying JTA in signal coordination; this group was taken as the control group

Group 2 applied optimizations on basic parameters and the information transmission mode relative to group 1

Group 3 applied optimizations on the transmission rule and the return relative to group 2

Detailed grouping and improvement effects are shown in Table 5.

Table 5 shows that the optimizations proposed in this paper play a good role in improving the operation of the JTA-based RL algorithm used for network-wide signal coordination. Optimizations of basic parameters and

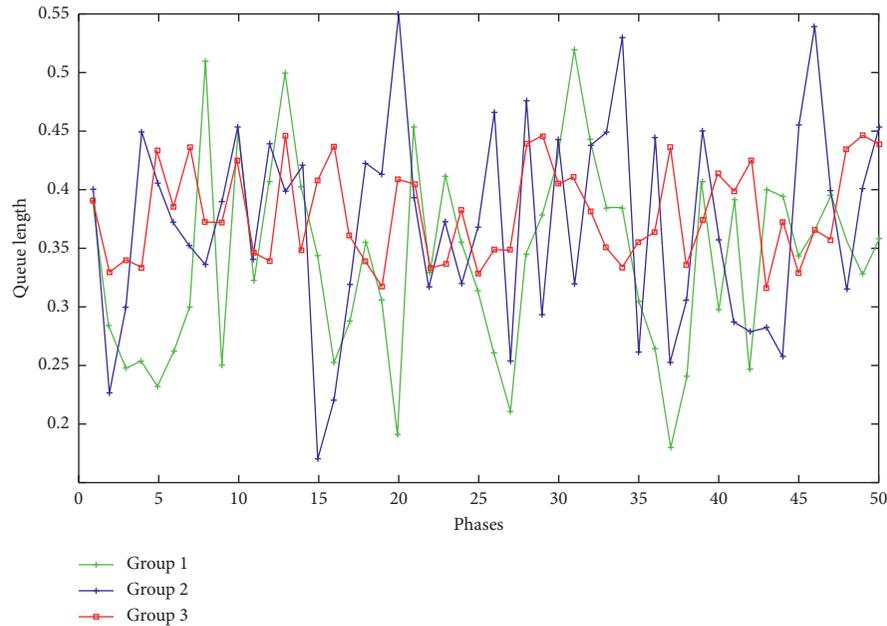


FIGURE 12: Fluctuations in queue length at intersection 5 under differentiated returns and a low congestion level.

TABLE 5: Summary of improvement effect on the JTA-based RL algorithm.

Grouping	Features	Improvement
Group 1	Existing algorithm	The control group
Group 2	Optimizations of basic parameters and the information transmission mode	(i) The green light of each phase was more flexible (ii) The system efficiency of signal coordination improved (iii) The operations of some intersections were still poor which need to be improved
Group 3	Optimizations of the information transmission rule and the return	(i) The maximum queue length of the system was reduced (ii) The fluctuation of the queue length at the intersection was optimized

information transmission modes can improve the system efficiency and the flexibility of green lights. Optimizations of the information transmission rule and the return can improve the efficiency of both the system and of the single intersection. It can be concluded that better operational results can be achieved in network-wide signal coordination by applying the proposed optimizations to existing JTA-based RL algorithms.

However, the results reported here are based on a hypothetical network. Results from real-world implementation should be studied in future research. This would make our conclusions stronger. What is more, each intersection is divided into only 81 states; the possibility of more detailed states division should be studied.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was sponsored by the Natural Science Research Project of Colleges and Universities in Jiangsu Province (19KJB580012).

References

- [1] P. B. Hunt, D. I. Robertson, R. D. Rrtherton, and M. C. Royle, "The SCOOT on-line traffic signal optimization technique," *Traffic Engineering & Control*, vol. 4, no. 23, pp. 190–192, 1982.
- [2] P. R. Lowrie, "The Sydney coordinated adaptive traffic system—principles, methodology, algorithms," in *Proceedings of the International Conference on Road Traffic Signaling*, pp. 67–70, London, UK, April 1982.
- [3] J. J. Henry, J. L. Farges, and J. Tuffal, "The PROLYN real time traffic algorithm," in *Control in Transportation Systems*, pp. 305–310, Pergamon Press, Oxford, UK, 1984.
- [4] N. H. Gartner, "OPAC: a demand-responsive strategy for traffic signal control," *Transportation Research Record*, vol. 906, pp. 75–81, 1983.
- [5] P. Mirchandani and L. Head, "A real-time traffic signal control system: architecture, algorithms, and analysis," *Transportation Research Part C: Emerging Technologies*, vol. 9, no. 6, pp. 415–432, 2001.

- [6] V. Mauro and D. Taranto, "Utopia," in *Proceedings of the Sixth IFAC/IFIP/IFORS Symposium on Control, Computers, Communications on Transportation*, pp. 245–252, Oxford, UK, August 1990.
- [7] F. Boillot, S. Midenet, and J.-C. Pierrelée, "The real-time urban traffic control system CRONOS: algorithm and experiments," *Transportation Research Part C: Emerging Technologies*, vol. 14, no. 1, pp. 18–38, 2006.
- [8] C. Diakaki, M. Papageorgiou, and K. Aboudolas, "A multi-variable regulator approach to traffic-responsive network-wide signal control," *Control Engineering Practice*, vol. 10, no. 2, pp. 183–195, 2002.
- [9] X. Liang, X. Du, G. Wang, and Z. Han, "A deep reinforcement learning network for traffic light cycle control," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1243–1253, 2019.
- [10] M. Aslani, M. S. Mesgari, S. Seipel, and M. Wiering, "Developing adaptive traffic signal control by actor-critic and direct exploration methods," *Proceedings of the Institution of Civil Engineers—Transport*, vol. 172, no. 5, pp. 289–298, 2019.
- [11] 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.
- [12] H. Jeon, J. Lee, and K. Sohn, "Artificial intelligence for traffic signal control based solely on video images," *Journal of Intelligent Transportation Systems*, vol. 22, no. 5, pp. 433–445, 2017.
- [13] H. M. A. Aziz, F. Zhu, and S. V. Ukkusuri, "Learning-based traffic signal control algorithms with neighborhood information sharing: an application for sustainable mobility," *Journal of Intelligent Transportation Systems*, vol. 22, no. 1, pp. 40–52, 2017.
- [14] S. Darmoul, S. Elkosantini, A. Louati, and L. Ben Said, "Multi-agent immune networks to control interrupted flow at signalized intersections," *Transportation Research Part C: Emerging Technologies*, vol. 82, pp. 290–313, 2017.
- [15] J. C. Medina and R. F. Beenekohal, "Traffic signal control using reinforcement learning and the max-plus algorithm as a coordinating strategy," in *Proceedings of the 2012 15th International IEEE Conference on Intelligent Transportation Systems*, pp. 596–601, Alaska, USA, September 2012.
- [16] F. Zhu, H. M. A. Aziz, X. Qian, and S. V. Ukkusuri, "A junction-tree based learning algorithm to optimize network wide traffic control: a coordinated multi-agent framework," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 487–501, 2015.
- [17] F. V. Jensen, S. L. Lauritzen, and K. G. Olesen, "Bayesian updating in causal probabilistic networks by local computations," *Computational Statistics Quarterly*, vol. 916, no. 4, pp. 269–282, 1990.
- [18] C. Q. Shao, X. H. Du, and G. Q. Li, "Queuing vehicle headway statistic analysis at signalized intersection," *Journal of Highway and Transportation Research and Development*, vol. 4, no. 20, pp. 76–79, 2003.
- [19] Y. Zhao, Y. Ma, J. Lu, Y. Zong, and Q. Wan, "Variability of green time to discharge a specified number of queued vehicles at a signalized intersection," *Mathematical Problems in Engineering*, vol. 2017, Article ID 3723575, 8 pages, 2017.