Two Intersections Traffic Signal Control Method Based on ADHDP

Lin Cao[1, 2], Bin Hu[1](Corresponding Author*), Xisong Dong [1,3], Gang Xiong [1,4], Fenghua Zhu[1, 4], Zhen Shen [1,5], Dong Shen[3], Yuliang Liu[5]

[1] The State Key Laboratory of Management and Control for Complex Systems,
Institute of Automation, Chinese Academy of Sciences,
Beijing, China

[2] College of Information Science and Technology, Beijing University of Chemical Technology, 100029, Beijing

[3] Institute of Smart Education Systems,
Qingdao Academy of Intelligent Industries,
Qingdao, China

[4] Dongguan Research Institute of CASIA,
Cloud Computing Center, Chinese Academy of Sciences,
Dongguan, China

[5] Beijing Engineering Research Center of Intelligent Systems and Technology,
Institute of Automation, Chinese Academy of Sciences,
Beijing, China

***Abstract*—With the rapid development of Chinese economy and automotive industry, urban traffic congestion has become increasingly serious. Therefore, how to effectively alleviate the traffic congestion and improve the efficiency of vehicles has become the main concern. Traffic signal control is one of the effective ways to solve urban traffic congestion. In this paper, a traffic signal control method based on Action-Dependent Heuristic Dynamic Programming (ADHDP) is investigated. The control algorithm is simulated on two intersections, both of which have two phases with four entrance approaches. The computer simulation results show that the control method has the better ability of on-line learning compared with traditional Fix-Time Control, and can effectively improve the average speed of vehicles, and reduce travel time and alleviate the traffic pressure.**

Keywords—ADHDP; Traffic Signal Control; Adaptive Control; Intelligent Transportation Systems

I. INTRODUCTION

With the rapid development of social economy and the increasing demand for traffic vehicles, the road congestion of the world's major cities are getting worse and worse. The speed of construction and improvement of traffic facilities is far behind the increasing traffic demand. Nowadays, urban inhabitants are aware of the importance of traffic time and efficiency, so how to effectively solve the problem of traffic congestion has become a common problem to solve in developed countries and developing countries. One way to solve the problem of traffic congestion is Intelligent Transportation System (ITS), which can help ensure the safety of the traffic. Furthermore, an important part of the intelligent transportation is urban transportation system [1].

In order to improve the traffic capacity of the urban transportation system, it is important to use traffic signal control. However, there are still many problems which need to be solved in the traffic signal control. Because of the characteristics of stochastic and non-linear of traffic conditions, consequently, many traditional control technologies cannot get the best results, and the signal control strategies must be highly responsive in real-time. Finding the optimal strategy becomes a rather complicated problem. With multiple information sources, it is more difficult to control them effectively with traditional theories and methods [2]. Because of the particularity of the urban traffic problems, it is difficult to establish accurate mathematical model. The simple traditional control of time-fixed method is difficult to adapt to the more and more complex traffic conditions. Thus, some advanced control theories and intelligent methods in control and optimization must apply to urban traffic control.

Traffic signal control system is a typical dynamic system. Furthermore, the dynamic system is universal. The research on dynamic system stability has been a hot research topic. In 1957, Bellman proposed the Dynamic programming (DP) method to solve the optimal control problem of nonlinear dynamic system [3].The core of the DP method is the optimal strategy for the state formed by the initial decision, regardless of the initial state and the initial decision. As a result, a basic recursive formula is summarized, in order to find the optimal action that transfers the system from the current state to a new state, the DP recursively calculates Bellman's equation backwards step-by-step. However, it is often computationally untenable to run true DP due to the backward numerical process required for its solutions, i.e., as a result of the well-known "curse of dimensionality" [4]. An approximation method can be used to overcome the shortcomings of the application of DP method, and maintain the basic function of dynamic control.

Adaptive dynamic programming (ADP) combines DP, reinforcement learning (RL) and function approximation method, using function approximation structure on the basis of online data and off-line data to estimate the performance index function of the system and then obtain approximate optimal control measure based on the principle of optimality [5]. Usually ADP consists of action module, critic module, and model module. Because each module adopts neural network to achieve, it is called the three modules for the Critic network, the model network and Action network. The basic idea of the theory comes from the action-critic structure of the RL, in which approximate the optimal control strategy with the action structure network and approximate the optimal performance index function by the critic structure. Combined the action of network and the critic structure is equivalent to an Agent. The adaptive dynamic programming method can effectively solve the limitations of traditional dynamic programming method, and can realize the on-line learning without requiring the known system model [6].

The action dependent heuristic dynamic programming (ADHDP) is a typical adaptive dynamic programming method and model network does not contain. Urban traffic signal has strong randomness, so by classical control methods, such as timing control, traffic signal control is difficult to achieve more optimal results, especially when the traffic flow is large random fluctuations. It is difficult to establish accurate mathematical model for urban traffic problems. Because ADHDP is model-free, in the actual control process, it is not necessary to establish accurate control mathematical model of the controlled object. In order to improve the system performance, online learning control with controlled object information interactively is used.

In this paper, a traffic signal control method based on ADHDP is investigated. The organization of this paper is as follows: Section II gives a comprehensive introduction to ADP algorithm and introduces the comparative algorithms of ADHDP, at the same time, the traffic signal controller designed based on ADHDP is presented. Section III presents the results of the experiments and their analysis. Then the conclusions are drawn out, and future research directions are given in Section IV.

Fig. 1. The traffic jams and congestion on the road

II. ADHDP Principle and Algorithm

The action dependent heuristic dynamic programming (ADHDP) is a typical adaptive dynamic programming method, whose main idea is that under the effects of the critic neural network, the performance index function of the system in dynamic programming can be estimated, in order to avoid the problem of "curse of dimensionality". What's more, because ADHDP has more practical characteristics than the traditional methods, for example, the system model is not necessary to be known, and this approach has adaptive capacity, it can satisfy the requirement include the system parameters changing frequently, highly responsing in real-time, establishing the accurate model of urban traffic control system with difficulty. ADHDP is also composed of Critic module, Action module and Model module.

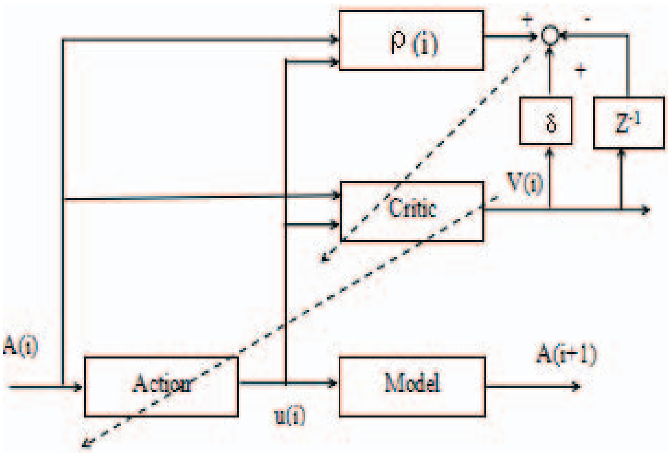

Fig. 2. Schematic diagram of structure of ADHDP

Fig. 2 shows a schematic diagram of our proposed ADHDP control scheme. What follows is that the specific structure of ADHDP which is model-free and the specific control process methods should be introduced. Based on the data which have been already collected, the system traffic state $A(i)$ is obtained, then the state $A(i)$ can be as the input of the action network, the output of the action network is a control variable $u(i)$. $u(i)$ can be as the input to the model module, then the next system status $A(i+1)$ can be obtained. The inputs to the critic network are $A(i)$ and $u(i)$. The critic network's output $V(i)$ is output performance index of critic module. According to the output result of the critic network, the error is calculated and then training the critic network. After that, the action network should be trained. Control process is executed in this way until the desired goal is reached. ADP consists of action module, critic module, and model module and the training error of the Critic module is defined as:

$$e_c(t)=\delta V(t)-V(t-1)+\rho(t)$$

$\rho(t)$ is the reinforcement signal and δ ($0 < \delta < 1$) is a discount factor. If the training error gets closer to zero, the action network's optimal traffic control performance can be evaluated and the following equation can be recursively obtained:

$$V(t)=\sum_{i=t+1}^{\infty}\delta^{i-t-1}\rho(t)$$

The function of the Model module is to simulate the controlled object. If the accurate mathematical model of the controlled object is known, then the model can be directly used as the Model module. From Fig.2 it can be seen that the error back propagation is not through the Model module, which means action network and critic network do not need to know the internal information of the controlled object, only needs to detect the external information. So ADHDP algorithm is a kind of model free and adaptive method. If the offline learning is put into use, Model module should be the artificial system or system simulation model, on the other hand, if the online learning is adopted, then Model can be the actual controlled object.

The structure of the critic neural work is chosen as a three-layer feedforward network. The Critic module can be made up of artificial neural network, the input is the system control variable $u(i)$ and the system state $A(i)$, the training error is shown in the following formula:

$$E_c(i)=\frac{1}{2}\left[\delta V(i)-V(i-1)+\rho(i)\right]^2$$

The weights for the critic network are updated using the following methods:

$$w_c(i+1)=w_c(i)+\Delta w_c(i)$$

$$\Delta w_c(i)=-\frac{\partial E_c(i)}{\partial w_c(i)}=-\frac{\partial E_c(i)}{\partial V(i)}\frac{\partial V(i)}{\partial w_c(i)}$$

The action module can be made up of artificial neural network, the input is the system state $A(i)$, control variable $u(i)$ is the output of the action network. The training error is shown in the following formula:

$$E_a(i)=\frac{1}{2}\left[V(i)-Q(i)\right]^2$$

where $Q(i)$ is the control target.

The weights for the action network are updated using the following methods:

$$w_a(i+1)=w_a(i)+\Delta w_a(i)$$

$$\Delta w_a(i)=-\frac{\partial E_a(i)}{\partial w_a(i)}=-\frac{\partial E_a(i)}{\partial V(i)}\frac{\partial V(i)}{\partial u(i)}\frac{\partial u(i)}{\partial w_a(i)}$$

Fig. 3 is flow chart of traffic signal control method of ADHDP. ADHDP is a complex nonlinear system control algorithm. The action network of ADHDP is the controller traditionally. It constantly collects the state of the controlled object and sends out the approximate optimal control variable according to the control algorithm to meet the desired value of the control. The critic network can monitor and evaluate the performance of the system to meet the function of optimization principle. The ADHDP controller continuously improves the control performance of the system through training the critic network and the action network. ADHDP is a kind of model-free method, and it can be online adaptive learning, but in order to improve the stability of the system, the controller need to go through off-line design. ADHDP adopts offline learning mode and online learning mode, so it can adapt to the changes of system parameters in real time, and enhance the robustness of the system.

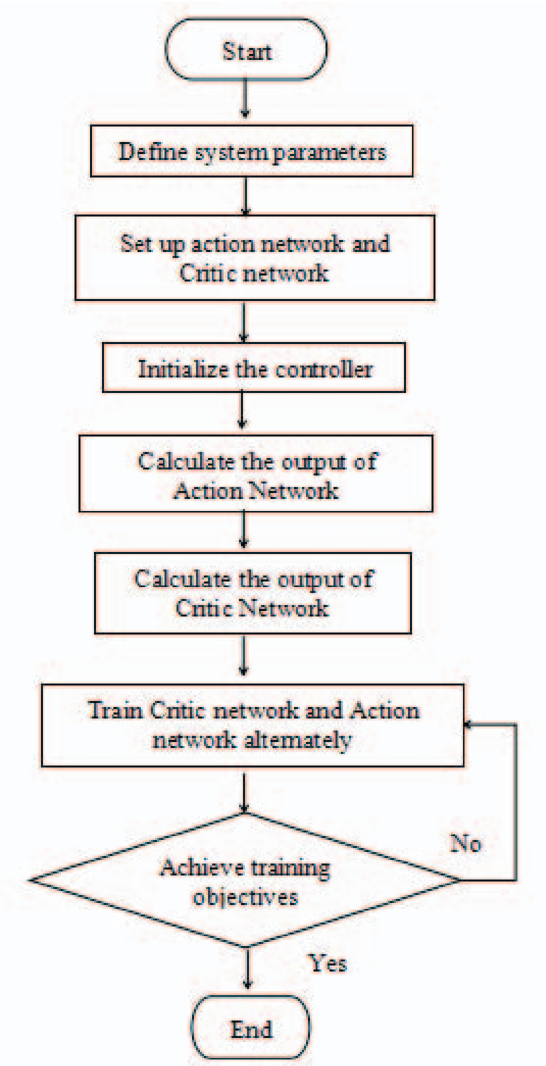

Fig. 3. Flow chart of traffic signal control method of ADHDP

III. Experiments and Results

Due to the rapid development of society, the single intersection control cannot meet the needs of social development, so in this section, a case of two traffic intersections is simulated. Microscopic traffic simulation software can simulate various road traffic conditions, and on this basis it can simulate the possible effects of various traffic signal control scheme. In contrast to other software, such as TSIS, VISSIM, PARAMICS can be more perfect to simulate network simulation. From the relevant research experiments, whether it is in the intersection simulation or larger scale network simulation, PARAMICS has higher efficiency and more perfect function. In the simulation of large and medium-sized urban traffic complex network, it has a strong ability to simulate and analyze. PARAMICS is a excellent micro kit for traffic simulation, PARAMICS is composed of monitoring module, programming module, analysis module, processing module and the modeler and so on. This software can theoretically support about one million intersection points, about four million road and more than 32000 regional road network. In terms of use, it has a good portability and scalability. The application program interface (API) on the C++ platform is introduce. With the help of API, a large number of external functions can be called to modify or use the traffic model. With the help of programming module, users can reduce workload in the model of operational changes and other aspects. The experiments presented in this paper were conducted using the platform based on PARAMICS traffic simulator in order to prove the effectiveness of the proposed algorithm. Fig. 4 shows the intersection case to be studied where the model has two intersections, both of which have two phases with four entrance approaches, each of which has two lanes and every lane has a detector. In the paper, the structure of Action neural network and the Critic neural network are shown in Fig. 5.

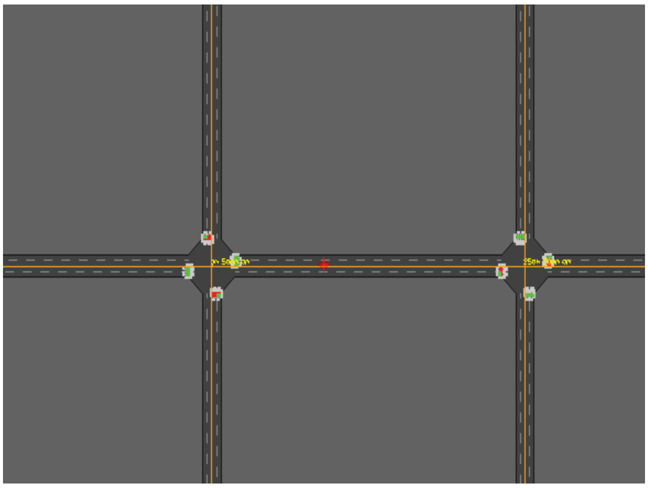

Fig. 4. Road network model

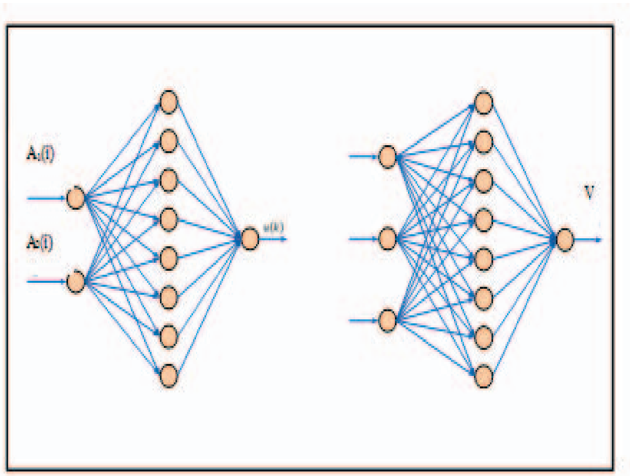

Fig. 5. Network structure of Action and Critic

By using the microscopic traffic simulation platform PARAMICS, the algorithm is trained off-line until the effect is close to that of the Fix-Time Control (FTC). And then, PARAMICS is used to simulate the real traffic scene and the method of ADHDP is applied in order to control the intersections online. The simulation time is 60 minutes.

The experimental results are shown in the Fig. 6 & Fig. 7, the control effect of ADHDP control method is compared with that of the traditional control method (Fixed-Time Control, FTC). At the beginning period of the simulation (0~16 minutes), the two curves are relatively close (shown on Fig. 6 and Fig. 7). However, when the training and learning progresses more and more in-depth, the control effect and advantages of ADHDP algorithm gradually becomes more and more obvious. In the vicinity of 21 minutes, the average speed of ADHDP control has been significantly higher than FTC control as shown in Fig. 6. In addition, the average travel time of ADHDP control is apparently less than the average travel time of FTC. By the end of the simulation, the average speed of FTC is only 32.9km/h, and the global average speed of ADHDP control is 44.8km/h, that is to say, the average speed of ADHDP control method is 11.9km/h faster than that under the control of the FTC method. The average speed increased by 36.1%. Moreover, the average travel time of FTC is only 116.3s, and the average travel time of ADHDP control is 84.5s, by comparison, the average speed increased by 27.3%. In other words, the traffic control method based on ADHDP can effectively decrease the average vehicle travel time and improve the efficiency of the vehicle.

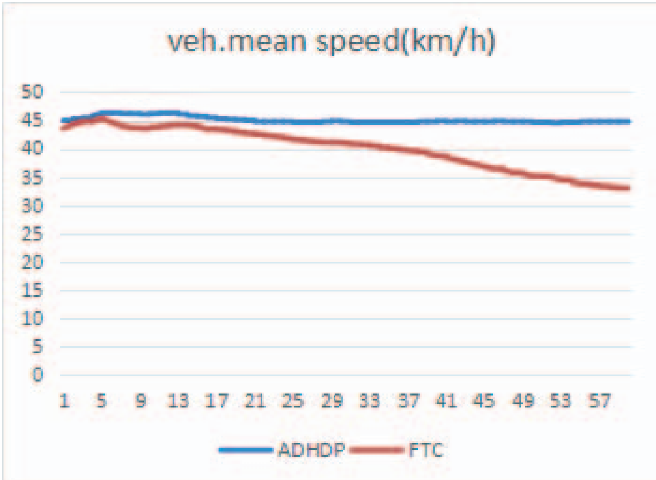

Fig. 6. Average speed using different method

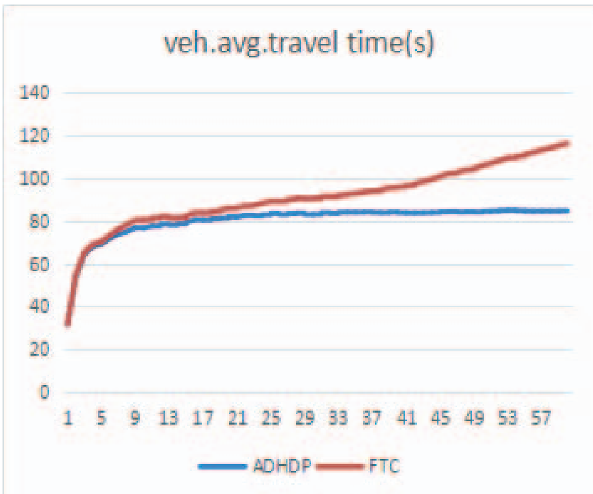

Fig. 7. Travel time using different method

IV. CONCLUSION

With the rapid development of social economy and the continuous improvement of people's living standard, the problem of traffic congestion has become a serious problem in many cities. Traffic signal control is one of the effective ways to solve urban traffic congestion, and it is of great practical value and significance to establish an effective traffic signal control system. The traffic signal control research has already become one of the hot spots in the research of intelligent transportation system.

In this paper, a traffic signal control method based on Action-Dependent Heuristic Dynamic Programming (ADHDP) is proposed. The ADHDP method does not need to establish the mathematical model of the controlled object and has the ability of on-line learning, furthermore the control effect can be continuously improved with the deepening of learning. On the basis of the ADHDP method, simulation experiment has been carried on, and the experimental results were analyzed. The computer simulation results show that the ADHDP method has the ability of on-line learning compared with traditional Fix-Time Control, can effectively improve the average speed of vehicles, reduce travel time and alleviate the traffic pressure. In this paper, the ADHDP algorithm is studied, but this paper's simulation is based on smaller traffic region, there still exists a certain gap with the practical application of intelligent transportation system and this paper's simulation is based on smaller traffic regional, so future work is to study more effective control algorithm applied to a larger area.

ACKNOWLEDGMENTS

This work is supported in part by Natural Science Foundation of China (61233001, 71232006, 91520301, and 61304201); "Intelligent Transportation & Internet of Vehicle" project of 2015 College Students' innovation and practice training program in Chinese Academy of Sciences; Guangdong Province S&T Department projects (2014A010103004, 2014B090902001, 2014B010118001, 2015B010103001, 2016B090918004)

REFERENCES

[1] F.Y. Wang. Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications. IEEE Transactions on Intelligent Transportation Systems, 2010, 11(3): 630-638.

[2] Z.Y. Liu. Intelligent Transportation Control Theory and Application. Beijing: Science Press,2003

[3] R.E. Bellman. Dynamic Programming. Princeton: Princeton University Press, 1957.

[4] F.Y. Wang, H.G. Zhang, D.R. Liu. Adaptive dynamic programming: an introduction. IEEE Computational Intelligence Magazine, 2009, 4(2): 39-47

[5] J.J. Murray, C.J. Cox, G.G. Lendaris, R. Saeks. Adaptive dynamic programming. IEEE Transactions on Systems, Man, and Cybernetics, Part C, 2002, 32(2): 140-153

[6] D.R. Liu, H.L. Li, D. Wang. Data-based self-learning optimal control: research progress and prospects. Acta Automatica Sinica, 2013, 39(11): 1858-1870

[7] H.G. Zhang, X. Zhang, Y.H. Luo, J. Yang. An overview of research on adaptive dynamic programming. Acta Automatica Sinica, 2013, 39(4): 303-311

[8] T. Li, D.B. Zhao, J.Q. Yi, Adaptive Dynamic Programming for Multi-intersections Traffic Signal Intelligent Control,11th International IEEE Conference on Intelligent Transportation Systems, Beijing, 2008, pages: 286-291.

[9] D.B. Zhao, D.R. Liu, J.Q. Yi. An overview on the adaptive dynamic programming based urban city traffic signal optimal control. Acta Automatica Sinica, 2009, 35(6): 676-681