A Method of Reinforcement Learning Based Automatic Traffic Signal Control

Wang Yaping[1,2], Zhang Zheng[2,*]

1. Shaanxi College of Communication technology, Xi'an 710018, China
2. School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

***Abstract*—To improve performance of traffic signal control system in urban area, a novel method is proposed in this paper. The roads, vehicles and the traffic control systems are all modeled as intelligent agents. Wireless communication network provides the possibility of the cooperation of vehicles and roads. Based on all the information from vehicles and roads, a traffic control policy can be planned online according to the updated situation on the roads. The optimum intersection signals can be learned automatically on line based on reinforcement learning. An intersection signal control system is studied as an example of the method with a Q-learning based algorithm. The simulation results show that the proposed intersection signal control can improve traffic efficiency by about 30% over the traditional method.**

Keywords- Intelligent Transportation System; Traffic Control Mechanism; Reinforcement Learning

I. INTRODUCTION

Intelligent Transportation Systems (ITS) are those utilizing synergistic technologies and systems engineering concepts to develop and improve transportation systems of all kinds [1].

The current traffic signal control system is designed based on historic traffic flow data, which cannot adapt itself to the rapidly varying situation at a crossroad. In some extreme situations, there are no vehicles during a green light and simultaneously many vehicles waiting before a red light.

Many researchers have proposed scheme to solve this problem. However, all these papers solve the problem according to the history flow data but not current information [2][3][4].

This paper is organized as follows: Section 2 proposes an agent-based traffic control system architecture and Section 3 shows a reinforcement learning based traffic signal system. An illustrative case study, especially for an intersection signal control, is studied in section 4, which demonstrates the effectiveness of the proposed architecture and the control mechanism. Finally, section 5 provides a few concluding remarks from this study.

II. MULTIAGENT BASED TRAFFIC CONTROL ARCHITECTURE

Intelligent Transportation Systems Architecture has its roots in transportation. Although it deals with subjects involving the application of advanced technologies such as information and communications technologies, it addresses these topics from a transportation point of view [5].

Multiagent systems have received considerable attention during the last decade despite their many challenging issues [6].

In this section, a novel architecture for agent based traffic control is proposed, as shown in Fig.1.

A. Traffic control system

The traffic control system is the core of the ITS architecture. All the traffic rules, criteria and real time instructions are produced by this system. It can communicate with a Dedicated Communication Network (DCN) by transferring information.

B. Dedicated Communication Network (DCN)

Wireless communication is the infrastructure for cooperation between vehicle and highway. A dedicated wireless communication network is designed to be installed on the roadside [3].

A dedicated communication network will likely be realized within a few years. This paper supposes that the DCN is built well in the proposed architecture.

C. Vehicle Agent

Intelligent vehicles are increasing in prominence and are at the forefront of technical innovation in modern day vehicle technologies [7].

In the proposed architecture, all the vehicles are modeled as agents. The vehicle agent includes several modules. It can cooperate not only with other vehicles but also the traffic control system in the proposed architecture.

1) Environment perception module

Onboard digital maps combined with GPS can be considered to be a cooperative system, as positioning data is received from outside the vehicle. The environment perception module can play a crucial role in supporting vehicle-highway cooperation as well as active safety systems and navigation.

2) Coordination module

A coordination module can provide warnings to drivers in high-risk areas for situations that cannot be detected by onboard vehicle sensors. In the proposed architecture, based on information from the environment, the coordination module produces a strategy for vehicle control.

3) Planning module

The planning module is the interim module between the coordination module and the control module. It receives data from the internal perception module and instructions from the coordination module, and then creates a command for the control module. It produces a control algorithm by adapting the suggestion coming from the coordination module.

4) Vehicle internal perception

Using internal sensors, the vehicle agent knows its real time status, including speed, acceleration, direction and other necessary parameters for dynamic control. Detecting potential internal failures is another useful function of the internal perception module.

*Corresponding author. zhangzh@mail.xjtu.edu.cn;

The Project Sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

978-0-7695-4296-6/11 $26.00 © 2011 IEEE
DOI 10.1109/ICMTMA.2011.35

5) Control module

Based on the command from the planning module, this module controls the vehicle actuators directly, including both the longitudinal and the lateral directions, for lane keeping, lane switching, stopping and going and turning right or left.

With the above five modules, the vehicle agent will be an intelligent autonomous agent. In the whole traffic system, each vehicle agent will be one of the members of the multiagent system.

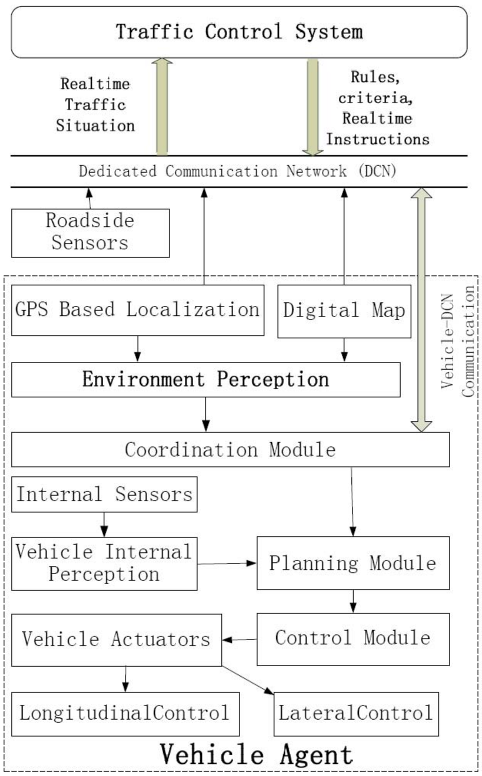

Figure 1. Agent based traffic control architecture

D. Roadside Sensors

Roadside sensors detect phenomena that vehicles cannot, such as the obstacles on the road and road hazards. The real-time situation of the road will be accessible in the near future.

III. REINFORCEMENT LEARNING BASED TRAFFIC SIGNAL CONTROL MECHANISM

An intersection control unit that is isolated from the traffic control system can be considered as an intelligent agent. It takes charge of the traffic flow management for a separate intersection. A reinforcement learning based intersection control mechanism is proposed in this section, as shown in Fig.2.

In Fig.2, we assume that the intersection control unit knows all the situations around the crossroad by communicating with vehicle agents, traffic control systems and roadside sensors.

The input of the control unit is the current state of the intersection, and the output is instruction for each vehicle agent real time. The received current status will be the initial environment state for the reinforcement learning module. The inputs for reinforcement learning are composed of an initial environment state, an expected environment state and basic actions of vehicle agents. The basic actions of vehicle agents are preloaded into the intersection control agent. For this application, the expected environment state must consist of all the vehicles passing into the crossroad, in other words, the number of vehicles needing to cross the intersection should be zero. The best solution from the initial state to the expected state will be produced by reinforcement learning and then be transferred to the vehicle agents by the dedicated communication network.

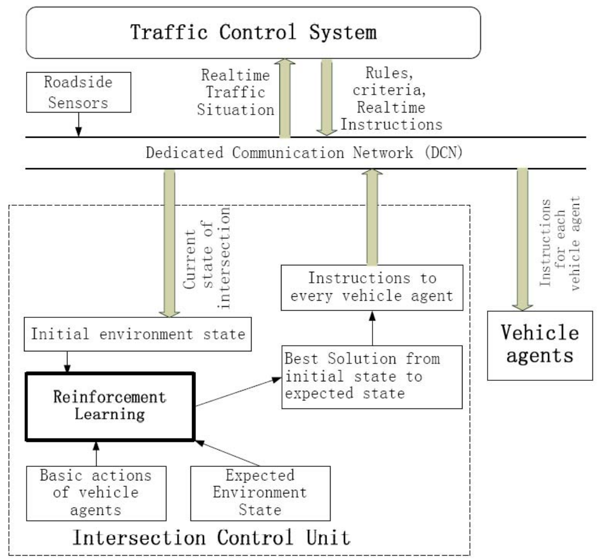

Figure 2. RL based single intersection signal control mechanism

In the standard reinforcement learning model, an agent is connected to its environment via perception and action [8][9].

IV. STUDY OF INTERSECTION SIGNAL CONTROL

In this section, a Q learning algorithm will be used to create a real time cooperation policy for an isolated intersection control under the proposed Traffic Control Mechanism.

A. Q-learning algorithm

Q learning, a type of reinforcement learning, can develop optimal control strategies from delayed rewards, even when an agent has no prior knowledge of the effects of its actions on the environment [8].

B. Model of the intersection signal system

A traffic system consists of various components, among which the traffic intersection is one of the most important. Our method is applied to a traffic intersection that consists of two intersecting roads, each with several lanes and a set of synchronized traffic lights that manage the flow of vehicles, as shown in Fig.3.

In this intersection, the rule of traffic management is right-hand based, which is used in China and South Korea. The vehicles in lanes ①, ③, ⑤ and ⑦, are approaching the intersection. Vehicles in ②, ④, ⑥ and ⑧, are leaving the intersection. For each of the approaching lanes, there are three directions for vehicles to choose: turn left, turn right and go straight, as shown in Fig.3. We will not consider the turn right direction because it does not impact other directions. In order to make this problem easy to model, we will not consider the pedestrian crossing the road. It will be very easy to add an additional rule for a pedestrian under our proposed mechanism.

Therefore, this problem can be modeled as 8 queues for different paths, as shown in Table 1.

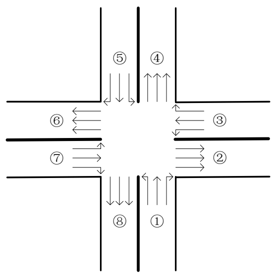

Figure 3. Isolated intersection

Table 1. Basic action definition of different queues

Queue	Basic action symbol	Path
Que1	A1	① → ④
Que2	A2	① → ⑥
Que3	B1	⑤ → ⑧
Que4	B2	⑤ → ②
Que5	C1	③ → ⑥
Que6	C2	③ → ⑧
Que7	D1	⑦ → ②
Que8	D2	⑦ → ④

We assume that there are a random number of vehicles spreading on different queues at the beginning of a signal period. This is the initial state of the environment. The final state must be that all the vehicles in the initial state have crossed the intersection. The intersection signal control system is modeled as a leader agent to manage the actions of all vehicle agents around the intersection.

Table 2. Action combination symbol

Phase	Action combination symbol	Component
Ph1	Ac1	A1+A2
Ph2	Ac2	A1+B1
Ph3	Ac3	B1+B2
Ph4	Ac4	C1+C2
Ph5	Ac5	C1+D1
Ph6	Ac6	D1+D2

If two of the actions from A1 to A8 are nonintervention, they are possible action combinations. We call these different combinations a signal phase. All possible combinations are shown in Table.2.

Therefore, the problem can be described as how to find the optimum sequence of action combinations to reach the final state. This is the main function of the intersection signal control agent.

C. *Parameters of learning process*

How shall we set the leaning parameters for this deterministic Markov decision process?

1) *Cost function*

We suppose that the vehicle number is $\boldsymbol{n}$ at state $\boldsymbol{s}$. After the selected action $\boldsymbol{a}$ completed, the current vehicle number will be $\boldsymbol{n_1}$. The cost of this action depends on the waiting time $\boldsymbol{t}$, and the remainder of vehicles $\boldsymbol{n_1}$.

$$g(s,a) = n_1 \times (t + t_{transition}). \tag{1}$$

where $\boldsymbol{t_{transition}}$ equals one of the three numbers {0, 1.5, 3} shown in Table 3. The average time for each vehicle passing the crossroad is supposed to be 3 seconds.

Table 3. $\boldsymbol{t_{transition}}$ of different phase transition

Phase transition type	Comment	$t_{transition}$ (s)
No transition	Current phase is the same as the previous one	0
Half transition	$Ac1 \Leftrightarrow Ac2; Ac2 \Leftrightarrow Ac3;$ $Ac4 \Leftrightarrow Ac5; Ac5 \Leftrightarrow Ac6;$	1.5
Full transition	Phase transfer except half transition	3

Table 4. Rules of state transition

Phase	Current state	Successor state
Ph1	{Que1,Que2, Que3,Que4, Que5,Que6, Que7,Que8}	{(Que1-1),(Que2-1), Que3,Que4, Que5,Que6, Que7,Que8}
Ph2	{Que1,Que2, Que3,Que4, Que5,Que6, Que7,Que8}	{(Que1-1),Que2, (Que3-1),Que4, Que5,Que6, Que7,Que8}
Ph3	{Que1,Que2, Que3,Que4, Que5,Que6, Que7,Que8}	{Que1,Que2, (Que3-1),(Que4-1), Que5,Que6, Que7,Que8}
Ph4	{Que1,Que2, Que3,Que4, Que5,Que6, Que7,Que8}	{Que1,Que2, Que3,Que4, (Que5-1),(Que6-1), Que7,Que8}
Ph5	{Que1,Que2, Que3,Que4, Que5,Que6, Que7,Que8}	{Que1,Que2, Que3,Que4, (Que5-1),Que6, (Que7-1),Que8}
Ph6	{Que1,Que2, Que3,Que4, Que5,Que6, Que7,Que8}	{Que1,Que2, Que3,Que4, Que5,Que6, (Que7-1),(Que8-1)}

2) *State transition function*

Table 4 explains how to determine the successor state when performing different phase actions. In this table, we suppose that each queued vehicle number is greater than zero. If one of the queue numbers is zero, the corresponding action will not change the current state.

3) *Discount factor*

In the simulation we set the discount factor, $\gamma = 0.8$.

D. *Simulation and results*

We wrote some MATLAB code to perform the simulation.

In the traditional mechanism, the signal phase transition is in a fixed sequence as shown by *Ph1, Ph2, Ph3, Ph4, Ph5* and *Ph6*. However, our proposed method can determine the optimum phase sequence automatically based on the updated situation.

In the following, we will show the comparative result for three different periods $\boldsymbol{T}$ and different phase time interval $\boldsymbol{t_{phase}}$.

In the following tables, P_s is the simulation period series, $\boldsymbol{N_{IV}}$ is the total number of vehicles at the initial state, ***Random Queues*** is the number of vehicle queues that are randomly created, $\boldsymbol{T_{IQ}}$ is the time interval from the initial state to the final state for a Q learning method, $\boldsymbol{T_{WQ}}$ is the total waiting time of all vehicles for the Q learning method, $\boldsymbol{T_{IT}}$ is the time interval from the initial state to the final state for the traditional method and $T_{IT} = 6 \times t_{phase}$, $\boldsymbol{T_{WT}}$ is the total waiting time for the traditional method, $P_{EI} = \frac{T_{IT} - T_{IQ}}{T_{IT}} \times 100\%$ is the percent improvement in the traffic efficiency, $\boldsymbol{O_A}$ is the optimum phase sequence from Q learning, $P_{WD} = \frac{T_{WT} - T_{WQ}}{T_{WT}} \times 100\%$ is the percent decrease in total waiting time, $\boldsymbol{T_L}$ is the running time of the Q learning program on the above mentioned computer.

E. Analysis of the results

From Table 5, we find that all the running times of the Q learning program $\boldsymbol{T_L}$ in every period are less than one second. This is short enough for the application of the intersection signal control system.

At the same time, the percent traffic efficiency improvement $\boldsymbol{P_{EI}}$, is located in [4.17% 56.67%]. The percent total waiting time decrease $\boldsymbol{P_{WD}}$ is located in [1.07% 61.36%]. The average percents of $\boldsymbol{P_{EI}}$ is 29.6%. The average percents of $\boldsymbol{P_{WD}}$ is 30.6%.

Table 5 Simulation result when $\boldsymbol{t_{phase}}$=60s

P_s	N_{IV}	Random Queues	T_{IQ}	T_{WQ}	T_{WT}	P_{EI}	P_{WD}	O_A	T_L
1	180	{20 20 40 20 20 20 20 20}	312s	18900s	24000s	13.33%	21.25%	{4 5 6 1 2 3}	0.8438s
2	175	{20 20 19 19 38 19 20 20}	303s	17916s	23280s	15.83%	23.04%	{1 2 3 6 5 4}	0.9375s
3	176	{20 20 18 18 40 20 20 20}	306s	18048s	23640s	15.00%	23.65%	{1 2 3 6 5 4}	0.4375s
4	202	{17 17 36 18 36 18 40 20}	345s	28479s	30600s	4.17%	6.93%	{1 2 3 6 5 4}	0.8438s
5	183	{38 19 16 16 17 17 40 20}	345s	21939s	26880s	4.17%	18.38%	{3 2 1 4 5 6}	0.8906s
6	189	{19 19 36 18 20 20 38 19}	351s	23418s	28380s	2.50%	17.48%	{1 2 3 4 5 6}	0.9063s
7	157	14 14 16 16 38 19 20 20}	276s	14331s	22740s	23.33%	36.98%	{3 2 1 6 5 4}	0.8750s
8	174	{38 19 26 13 20 20 19 19}	282s	18852s	22020s	21.67%	14.39%	{4 5 6 1 2 3}	0.9063s
9	123	{30 15 14 14 13 13 12 12}	219s	8946s	14280s	39.17%	37.35%	{4 5 6 3 2 1}	0.8906s
10	133	{15 15 17 17 30 15 12 12}	234s	10413s	16440s	35.00%	36.66%	{3 2 1 6 5 4}	0.9219s
11	130	{11 11 34 17 22 11 12 12}	249s	11007s	16140s	30.83%	31.80%	{6 5 4 1 2 3}	0.8594s
12	152	{22 11 15 15 16 16 38 19}	285s	14904s	24480s	20.83%	39.12%	{3 2 1 4 5 6}	0.8750s
13	171	{36 18 32 16 24 12 22 11}	273s	19292s	19500s	24.17%	1.07%	{4 5 6 1 2 3}	0.9531s
14	183	{26 13 36 18 26 13 34 17}	300s	22265s	25560s	16.67%	12.89%	{6 5 4 3 2 1}	0.9375s
15	120	{32 16 20 10 20 10 6 6}	216s	10221s	11760s	40.00%	13.09%	{6 5 4 1 2 3}	0.8750s
16	128	{19 19 20 10 15 15 15 15}	219s	9768s	15900s	39.17%	38.57%	{4 5 6 1 2 3}	0.8906s
17	112	{14 14 9 9 28 14 16 8}	189s	7905s	14220s	47.50%	44.41%	{1 2 3 4 5 6}	0.8906s
18	100	{16 8 8 8 34 17 6 3}	195s	6480s	11760s	45.83%	44.90%	{3 2 1 6 5 4}	0.9375s
19	128	{16 16 32 16 16 8 16 8}	228s	9960s	15120s	36.67%	34.13%	{4 5 6 1 2 3}	0.8906s
20	128	{15 15 30 15 14 7 16 16}	237s	10431s	16620s	34.17%	37.24%	{6 5 4 1 2 3}	0.8906s
21	108	{16 8 36 18 14 14 1 1}	189s	6906s	10860s	47.50%	36.41%	{4 5 6 3 2 1}	0.8750s
22	81	{5 5 24 12 18 9 4 4}	165s	4365s	10140s	54.17%	56.95%	{6 5 4 1 2 3}	0.4844s
23	109	{14 14 18 9 30 15 6 3}	207s	7740s	11400s	42.50%	32.11%	{1 2 3 6 5 4}	0.4375s
24	144	{9 9 16 16 40 20 17 17}	258s	11772s	21660s	28.33%	45.65%	{3 2 1 6 5 4}	0.5156s
25	77	{5 5 30 15 8 4 5 5}	156s	3501s	9060s	56.67%	61.36%	{6 5 4 1 2 3}	0.4219s

V. Conclusions

A new traffic control mechanism based on a combination of machine learning and multiagent modeling methods is proposed for future intelligent transportation systems.

The control method for an isolated intersection was studied specifically and was introduced in detail. The simulation results show that the proposed intersection control mechanism can improve traffic efficiency by about 30% over the traditional method. It proved that the proposed traffic control mechanism will be applicable in the near future.

References

[1] http://www.ewh.ieee.org/tc/its/.

[2] Min chee Choy, Dipti Srinivasan, Ruey Long Cheu. Cooperative, Hybrid Agent Architecture for Real-Time traffic Signal Control. IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 33, NO. 5, pp.597-607, Sep. 2003.

[3] Bishop, Richard. Intelligent Vehicle Technology and Trends. Boston, MA: Artech House, 2005. 344 pages. ISBN: 1580539114.

[4] Weihua Bao, Quanlin Chen and Xiaoxia Xu. An Adaptive Traffic Signal Timing Scheme for Bus Priority at Isolated Intersection. Proceedings of the 6th World Congress on Intelligent Control and Automation, June 21 - 23, 2006, Dalian, China, pp 8712-8716.

[5] Bob McQueen, Judy McQueen. Intelligent Transportation Systems Architectures. © 1999 ARTECH HOUSE, INC. 685 Canton Street, Norwood, MA 02062. ISBN 0-89006-525-X.

[6] Y. Shoham, R. Powers, and T. Grenager. Multi-agent reinforcement learning: a critical survey. Technical report, Stanford University, 2003.

[7] Ljubo Vlacic, Toshio Fukuda. Cooperative Intelligent Vehicles. IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 7, NO. 4, DECEMBER, pages 398-400,2006.

[8] Tom M. Mitchell. Machine Learning. McGraw-Hill , March 1, 1997. ISBN: 0070428077.

[9] M. Tan. Multi-agent reinforcement learning: Independent vs. cooperative agents. In Pro-ceedings of the Tenth International Conference on Machine Learning, pages 330-337, 1993.