Q-learning for intersection traffic flow Control based on agents

Q-learning for intersection traffic flow Control based on agents

주선 , 정길도
ZhouXuan and Kil To Chong

Abstract - In this paper, we present the Q-learning method for adaptive traffic signal control on the basis of multi-agent technology. The structure is composed of sixphase agents and one intersection agent. Wireless communication network provides the possibility of the cooperation of agents. As one kind of reinforcement learning, Q-learning is adopted as the algorithm of the control mechanism, which can acquire optical control strategies from delayed reward; furthermore, we adopt dynamic learning method instead of static method, which is more practical. Simulation result indicates that it is more effective than traditional signal system.

Key Words : agent; cooperation; Q-learning; reinforcement learning; dynamic; practical;

1. Introduction

As the number of vehicles on the road increases, optimal intersection traffic flow control method becomes more and more important. The traditional method which distributes the same fixed time into every phase and static learning method can't effectively meet the requirement. Hence, searching for new methods becomes the hotspot.

In recent years, reinforcement learning algorithm which doesn't need to depend on exterior model has attracted rapidly increasing interest in machine learning and artificial intelligence communication. By the interaction between state and action, we can finally get the most optimal control policy. Meanwhile, A multi-agent system can divide big and complex system into small and simple systems which can communicate with each other. So, A new method is proposed in this paper using reinforcement learning combined with multi-agent system technology

There are two kinds of agents in the method we propose: the phase agent and the intersection agent, and they interact with each other. In the Q-learning process, the phase agents base their decision according to the policy coming from the "most reward value". The phase agents can make intersection control become modulated. The exactly structure is given in section 2. For the intersection control, there are several algorithms on how to get the reward value. For example, reward value can be inversely proportional to the waiting time of the vehicles. In this paper we apply the queue length's variety to calculate the reward. Suppose we have a fixed phase time and multiply it with the departure or arriving rate, we will then get the reward value. This approach is presented in section3. A simulation about Q-learning process is given in section 4.

2. Multi-agent system

Agents are computer systems with two important capabilities. First, they are at least to some extent capable of autonomous action- deciding what they need to do in order to satisfy designer's acquirements. Second, they have the ability of interacting with other agents. In this paper, there are two kinds of agents-intersection agent and phase agent. The structure of multi-agent system is showed as figure1.. Meanwhile, the relationship between them is showed in figure2

저자 소개
* 周 璇 : 全北大學 電子情報工學科 碩士課程
* 丁吉道 : 全北大學 電子情報工學科 敎授・工博

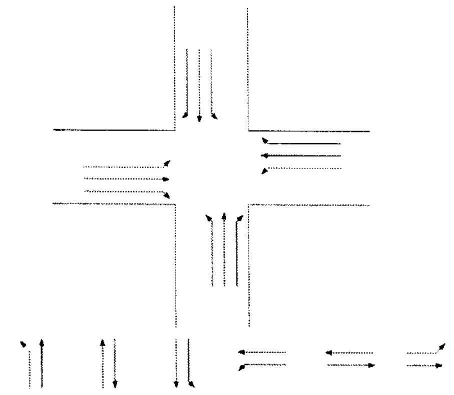

Fig.1. Intersection and phases

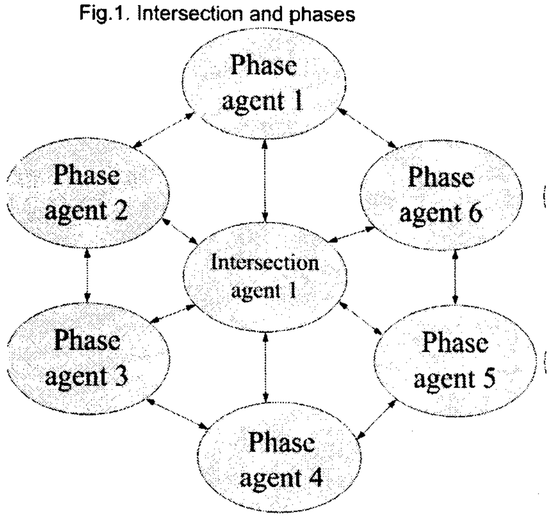

Fig.2.The relationship between phase agent and intersection agent

2.1. Intersection agent

Intersection agent stores the whole information about the lanes and control results. Furthermore, it has some kinds of rules on how to modify phase order and select the control method. After the learning process, it can make decision for which rule to choose, set the final policy for the phase agents.

2.2. phase agent

Phase agent mainly contains environment apperceiving module, information disposing module, communicating module and executing module. model structure of agent is showed in Fig.3.

3. Q-learning algorithm

Reinforcement learning is learning how to map

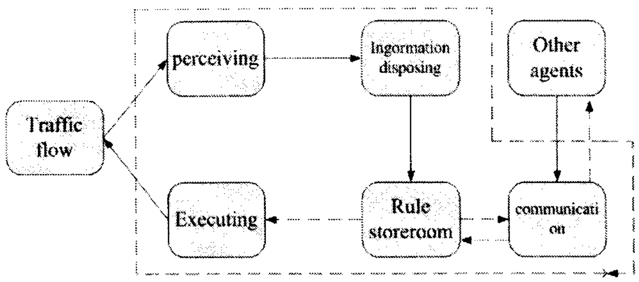

Fig.3. model structure of the agent

situations to actions, so as to maximize a numerical reward or cost signal. In most forms of the machine learning, the learner is not told which actions to take, but instead can discover which action yields the most reward by trying them. As one kind of reinforcement learning, Q-learning can acquire optimal control strategies from rewards, even when the agent has no prior knowledge of the effects of its actions on the environment.

How to realize the Q-learning algorithm? The key point is to find a feasible way to estimate training values for Q, given a sequence of immediate reward r spread out over time. Using following equation, we can realize it.

$$Q^{*}(s,a) \leftarrow r+\gamma \max Q^{*}(s',a') \tag{1}$$

Here, r denotes the reward value, coefficient γ is randomly chosen from 0 to 1.

There are four work steps for getting the value of r.

Stepl: randomly choose one as the original phase. And it is supposed to be the green phase.

step2: compare the length of other phases, choose the longest one as the next green phase. In order to realize it, two sensors are put on each lane for the velocity test, one is 50m far away from the stop line, the other is 5m far away. parameters are set as follows:

V_a stands for the arriving rate of the vehicles during the green phase.

V_l stands for the leaving rate of vehicles during the green phase.

V_{ra} stands for the leaving rate of vehicles of the other red phases.

we suppose the phase time is t.

The length of the green phase is denoted by L_g,

$$L_g = L_g' + (V_a - V_l)t \tag{2}$$

The length of the other red phase is denote by L_{ro},

$$L_{ro} = L_{ro}' + V_{ra}t \tag{3}$$

Here, L_{ro} is a vector, which contains five elements. now, we can do the comparison of the each element, choose the largest one as the next green phase.

Step3: calculate the value of the reward.

$$r = (L_g' + L_{ro}') - (L_g + L_{ro}) \tag{4}$$

Step4: do the circulation. After one circle-finishing the six phase's passing, we can get a cumulative reward. then, go back to the first step, do the circulation again. fix times later, we can attain six cumulative value. then choose the largest one as the policy of the action. Its process is called learning process.

4. simulation

A simulation test with mat-lab is done on the basis of Q-learning algorithm. In table1, each row denotes the length of every phase and its arriving rate, which also represents the current state. After the learning process, we can finally get the order of the phases as shown in table2, which is the most optimal policy.

Table.1.

50	0.1	50	0.2	40	0.2	40	0	50	0.1	50	0.2
46	0.2	52	0.2	42	0	30	0.1	46	0	52	0.1
48	0.1	54	0.1	32	0	26	0	36	0	48	0.2
44	0.1	50	0.2	22	0	16	0.2	26	0.2	50	0.1
40	0.1	52	0	12	0.2	18	0	28	0	46	0.2
36	0.2	42	0.2	14	0.1	8	0.1	18	0.1	48	0.2
38	0	44	0.2	10	0.2	4	0.2	14	0.1	50	0.1
28	0.2	46	0.1	12	0	6	0.2	10	0.1	46	0.1
30	0.2	42	0	2	0.1	8	0.1	6	0.1	42	0
32	0	32	0.1	0	0	4	0.2	2	0.1	32	0.2
22	0.2	28	0.1	0	0	6	0.2	1	0	34	0.2
24	0	24	0.1	0	0	8	0.1	0	0	36	0
14	0.2	20	0.1	0	0.1	4	0.1	0	0.1	26	0
16	0.2	16	0.2	0	0.2	2	0.2	1	0.1	16	0.1
18	0.1	18	0.1	2	0	4	0	0	0.1	12	0
14	0.2	14	0.2	0	0	0	0.1	2	0	2	0.2
16	0.2	16	0	0	0.2	2	0	0	0.2	6	0.2
18	0.1	6	0.2	2	0.2	0	0.1	4	0	8	0
14	0	8	0	4	0	1	0.1	0	0.2	0	0.1
4	0.1	0	0.2	0	0.2	2	0.2	6	0.2	1	0.2

Table.2

6	2	5	1	3	4
2	6	1	5	3	4
2	6	1	5	3	4
2	6	1	5	4	3
2	6	1	5	3	4
6	2	1	5	3	4
6	2	1	5	3	4
6	2	1	5	4	3
6	2	1	4	5	3
6	2	1	4	5	3
6	2	1	4	5	3
6	2	1	4	5	3
6	2	1	4	5	3
2	1	6	4	3	5
2	1	6	5	4	3
2	1	6	4	5	3
1	2	6	5	3	4
1	2	6	3	4	5
1	2	5	4	6	3
1	5	4	6	3	2

5 conclusion

On the basis of the Q-learning algorithm, intersection agent carries out the learning progress, sets down the most efficient policy and conveys it to the phase agents. Finally, phase agents execute the policy, which all base on the performance of the multi-agent-apperceiving communicating, making decision, executing capability. From the simulation, we can find that the method in this paper will shorten the length of the longest phase, which solve the problem of stacking in one phase.

참 고 문 헌

[1] Liu Guang-ping,Zhai Run-ping and Pei Yu-long. "A Calculating Method of Intersection Delay under Signal Control", IEEE Proceedings of the Intelligent Transportation Systems Conference, Seattle, WA, USA,pp 1114-1119. Sept. 30 - Oct. 3, 2007..

[2] Liu Xiao-ming,Gong Xiao-yan," Study of intersection traffic flow control on the basis of agents" IEEE Proceedings of Intelligent Transportation Systems, Vol. 2, pp. 1755-1758, Oct. 2003

[3] D. Srinivasan and M.C. Choy." Cooperative multi-agent system for coordinated traffic signal control". IEE Proc. Intell. Transp. Syst, Vol. 153, No. 1,pp 41-50 March 2006.

[4] Richard S. Sutton and Andrew G.Barto,"Reinforcement Learning: An introduction"MIT Press, Cambridge, MA, A Bradford Book, pp127-134,1998.

[5[A. D'Ambrogio et al., "Simulation model building of traffic intersections, Simulat. Modell. Pract. Theory", doi:10.1016/j.simpat. November2008.

[6] Wei wu. zhsng Qwn, "A Traffic Signal Control Method Based on Furry Logic Controller at Urban Road Intersection,.Journal of Traffic and transportation Engineering, pp99-102.June,2001.