ACP based self-learning control for urban intersection

Feng Chen, Lingyun Zhu, Chuyue Han, and Gang Xiong

Abstract—**The intersection models, such as delay models and queuing length models, are the foundations of optimal signal timing for urban intersection. Lack of the field data of intersection, it is highly difficult to calibrate parameters of the intersection models. Due to the effects of intersection topology, channelization and traffic conditions on these models, obviously it is impossible for single model to be suitable for optimal control of various intersections.**

ACP is emerging technique for Intelligent Traffic Systems (ITS). It provides an effective means for problem-solving of complex traffic issues by constructing artificial system which is consistent with corresponding real counterpart.

Based on ACP approach, we propose a self-learning optimal control strategy for typical intersection. In this approach, optimal control policy is found by systematic interaction with the traffic environment, so as to adapt dynamic traffic conditions and different intersection topologies. Furthermore, according to traffic characteristics analysis of intersection, joint control of all approaches can be reduced to optimal control of only one approach. Thus, the computational and storage complexities are decreased significantly.

The experiment results demonstrate that our approach has considerably lower average delay and higher traffic capacity of intersection than these optimal control methods which are respectively based on HCM2000 and Webster models.

I. Introduction

Intersection is the bottleneck of urban traffic, and it also basic unit of traffic management and control. The optimal control of intersection has still been attracted considerable attentions in ITS community. Till now, various research achievements have been obtained in intersection optimal control fields. These control methods primarily rely on intersection models, which include average vehicle delay models [1],[2], queue length models [3], and their modified versions [4-6]. The achievements above have greatly improved current traffic situation, but it is extremely difficult to gain field data for modeling delay and queuing length of intersections. In addition, multiple uncertain factors impact on these models in different degrees. Thus the model parameters can only be calibrated roughly. Furthermore, it is impossible for an intersection model to be suitable for all intersections with different topology and road channelization.

Reinforcement learning is a class of unsupervised machine learning approach. It can find optimal behavior policy by systematic trial-and-error interaction with environment [7], and thus offers an effective way for the control of complex traffic system. In the past decade, some researchers have studied optimal control of intersection traffic signal by the use of reinforcement learning [8]-[11]. In their works, agent learns how to control the traffic light through its trial-and-error interaction with the environment and reward received. However, it is extremely difficult to perceive the states of traffic environment. Most RL approaches are only for the theoretical purpose. Meanwhile, these methods suffer from the exponential growth in the number of states and actions, thus they cannot meet the demands of practical application.

The ACP approach was originally proposed in [12]–[14] for the purpose of modeling, analysis, and control of complex systems. And it represents another new milestone in solving the management difficulty of real-world complex systems. Through parallel interaction between an actual transportation system and its corresponding artificial counterparts (one or often more), ACP approach offers a verifiable, repeatable, and manipulated way for ITS research in the condition of lack of available large scale data.

Based on the novel ACP approach, we develop the Artificial Traffic System USTCATS2.0. In this artificial traffic system platform, virtual traffic scenarios are built, which are consistent with their corresponding real traffic counterparts. We propose a self-learning strategy for optimal control of typical intersection. The traffic flow characteristics are analyzed and defined. On this basis, the joint actions for all approaches of intersection are reduced to the action for single approach, so as to decrease both computational complexity and storage complexity. The optimal control of intersection is achieved by the interaction between artificial traffic system and corresponding real counterpart.

II. Self-learning control principle for intersection based on ACP approach

A. The Principle Description

ACP (Artificial systems, Computational experiments, and Parallel execution) approach is novel means for modeling, analysis, and control of complex systems [15]. This approach consists of three steps: 1) modeling and representation using artificial societies; 2) analysis and evaluation by Computational experiments; and 3) control and management through Parallel execution of real and artificial systems.

Based on ACP approach, we develop the Artificial Traffic System USTCATS2.0. For the purpose of studying self-learning control of traffic flow in intersection, virtual traffic scenarios and the corresponding real traffic counterparts are built respectively. In this work, we focus on optimal control of traffic flow for typical intersection. Figure 1 shows our

*Research is supported by NSFC (71232006 and 61233001).

Feng Chen is with the University of Science and Technology of China, Hefei, 230027 China (phone: 86-0551-63600917; e-mail: chenfeng@ ustc.edu.cn).

Lingyun, Zhu, is with the University of Science and Technology of China, Hefei, 230027 China (e-mail: zzzlyun@163.com).

Chuyue Han, is with the University of Science and Technology of China, Hefei, 230027 China (e-mail: hcy0551@qq.com).

Gang Xiong is with The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China (e-mail: gang.xiong@ia.ac.cn).

978-1-4799-6058-3/14/$31.00 ©2014 IEEE

constructed virtual traffic scenarios in a typical intersection, which include vehicles, traffic signal controller, inductive loop sensors, and road channelization. The virtual inductive loop sensors are used to calculate average vehicle delay and traffic capacity of intersection.

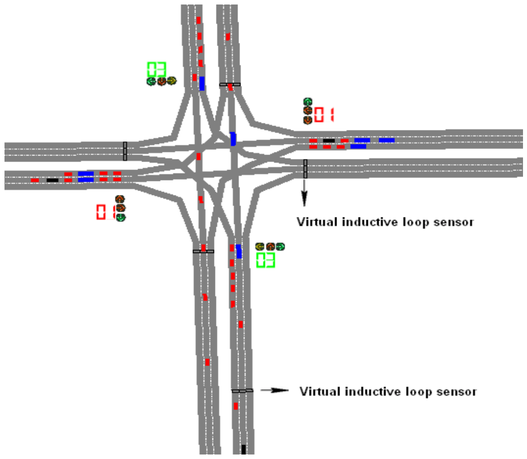

Figure 1. Traffic scenarios of intersection in artificial traffic system

In order to find optimal control scheme for real intersection, we propose a self-learning control approach. Q-learning, a model-free machine learning algorithm, is introduced to search for optimal control policy of intersection based on the artificial traffic system platform. In this artificial system, respect to dynamic traffic status, agent seeks for optimal control policy through trail-and-error interaction with its traffic environment. When agent finds the optimal control policy, artificial traffic system will transfer this policy to its corresponding real traffic system to execute. Then, these two systems will respectively run in parallel manner and interact individual information each other. Therefore, the two systems are improved and perfected continuously. Their parallel evolution is achieved accordingly.

This self-learning control system of intersection is defined in the following:

Definition: The self-learning control system of intersection can expressed as a quadruple $< S, A, P, R >$. Where $S =< S_1, S_2, S_3, S_4 >$ is discrete and finite set of joint states, and $S_1, S_2, S_3,$ and S_4 are sub-states of joint state which respectively correspond to the queuing lengths of different approaches in intersection. A=$< A_1, A_2, A_3, A_4 >$ is discrete set of joint actions which denote signal timing scheme. A_1, A_2, A_3 and A_4 are possible sub-actions respect to sub-state $S_1, S_2, S_3,$ and S_4 respectively. P is state transition probability, and R is reward-reciprocal of vehicle delay. The system aims at searching for optimal traffic signal control policy to maximize the discount cumulative rewards.

According to the above definition, the computational and storage complexity of this self-learning control strategy are analyzed as follows:

Let maximal size of sub-state sets S_i $(i = 1,2,3,4)$ be m and maximal size of sub-action sets A_i $(i = 1,2,3,4)$ be n. The computational complexity and storage complexity of this approach are respectively $O(n^4m^4)$ and $S(n^4m^4)$, which exponentially increase with size of state and action sets. This limits its application to optimizing control scheme for intersection, especially for large-scale urban road network.

In order to overcome the drawbacks of this self-learning control strategy based on Q-learning, the traffic flow characteristics of intersection are further studied. When there is conflict relation among different approaches in intersection, different sub-action sets are correlated and they must be handled as a whole. Thus the computational and storage complexity of this approach cannot be decreased. In this work, all approaches of intersection are divided groups according to running directions of vehicles. For instance, east-west through, east-west left turning, north-south through, and north-south left turning lanes or directions. Obviously, there are not any conflicted relations among the traffic flow of different groups in intersection. Accordingly, different sub-action sets are mutual independent as well as different sub-state sets. Therefore, optimal control of all approaches can be reduced to that of only one approach in intersection. Search space of this self-learning control is simplified significantly. The computational and storage complexity of this approach are reduced to $O(nm)$ and $S(nm)$, respectively.

B. Self-learning control for intersection

According to the above definition and analysis, the self-learning control algorithm for intersection needs a training stage before its practical application. The training stage is addressed as follows:

Suppose state S=$\{0,1,2, \ldots, L_{max}\}$, where$L_{max}$ is possible maximum of queuing vehicles for all approaches. Goal state $s_g = \{0\}$, which means all vehicles of an approach are released. Action set A is possible timing schemes for a phase, and A= $\{0, a_{min}, \cdots, a_{max}\}$. a_{min} is minimal green time and a_{max} is maximal green time. These two parameters can be determined empirically. The state-action pair function $Q(s_t, a_t)$ is defined in Equation 1. It is a discounted cumulative reward given that an agent starts in state s_t, takes an action a_t once, and follows a control policy thereafter.

$$Q(s_t, a_t) = r_t + \gamma\{\max_{a\in A} Q(s_{t+1}, a)\} \tag{1}$$

In Eq. 1, γ is discount factor.

For searching for optimal control policy, n vehicles are generated for entrance lanes, such as east-west through lanes, by the Artificial Traffic System platform, here $n < L_{max}$. Assume there is not any new arrival vehicle. In this Artificial Traffic System platform, at time t, agent senses states of the environment-the number of queuing vehicles, selects an action, that is green time of the corresponding phase, to perform. Then the environment makes a transition from state s_t to s_{t+1}. The reward r_t is obtained. Considering traffic capacity of intersection is much easier than the other traffic parameters, we define reward function in term of traffic capacity per signal cycle. Function $Q(s, a)$ is updated by the following equation:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha(r_t + \gamma \max_{b\in A} Q(s_{t+1}, b) - Q(s_t, a_t)) \tag{2}$$

where α is learning rate.

One loop ends if goal state is reached. Repeat the above process until all $Q(s, a)$ values converge. As a result, optimal control policy is found and applied to real traffic control system.

In searching for optimal control policy process, an agent faces with the tradeoff problem between exploration and exploitation. That is, an agent needs to make the balance between exploiting previous knowledge and exploring new actions. Exploring new actions can improve the long-term performances of agent's policy. In contrast, exploiting agents' previously acquired knowledge can obtain good short-term performance of the policy to minimize the learning costs. However, it may converge to a sub-optimal control policy. In our approach, the exploration-exploitation dilemma is handled effectively based on the Boltzmann exploration method.

Let $p(a_j|s)$ denote choice probability of action a_j under state s. We define probability function $p(a_j|s)$ as follows:

$$p(a_j|s) = \frac{e^{Q(s,a_j)/T}}{\sum_{b \in A} e^{Q(s,b)/T}} \quad (3)$$

where T is a positive parameter called the temperature. High temperatures cause the actions to be all (nearly) equiprobable. Low temperatures cause a greater difference in selection probability for actions that differ in their value estimates of Q-function. The costs of search process can be decreased considerably through selecting parameter T carefully.

C. Algorithm description

According to the aforementioned principle, the self-learning control approach is described as follows:

```
Step1:  $Q(s,a)$ ← a set of initial values for
    $s \in \{0,1,2,\cdots,L_{\max}\}$ and $a \in \{0, a_{\min}, \cdots, a_{\max}\}$
Step2:  parameter setting for temperature parameter T, learning
    rate $\alpha$, and discount factor $\gamma$
Step3:  repeat
Step4: generate an initial state x randomly:
        $x = L_{\max} \times random()$
Step5:   if x is the goal state then goto step 9 endif
Step6:   Seek the action $a \in A$ in lookup table in term of
condition $s = x$
        Calculate $p(a_j|s) = \frac{e^{Q(s,a_j)/T}}{\sum_{b \in A_x} e^{Q(s,b)/T}}$
        Select action $a_i$ with $\varepsilon$ −greedy strategy
Step7:   Execute action $a_i$, state x is transformed to y, reward
$r_i$ is obtained
      Update Q(x, a) according to the following formula:
  $Q(x,a_i) = Q(x,a_i) + \alpha(r_i + \gamma \max Q(y,b) - Q(x,a_i))$
        The number of Iteration k=k+1
        $T = \beta^k T, 0 < \beta < 1$
Step8:   $x \leftarrow y$,  goto step 5
Step9:  until function $Q(s,a)$ are convergent for all state s and
action a
```

III. EXPERIMENTS AND DISCUSSION

USTCsim1.0 simulator, a project which is financially supported by National 863 plan of China, is employed to validate our proposed approach. In the experiments, a typical intersection respectively with nonsaturation and oversaturation traffic conditions is selected for tests. These saturations include 0.8, 1.2, and 1.5, which can reflect current urban traffic conditions.

For the purpose of performance evaluation, we compare our approach to the other typical methods including optimal cycle equations respectively based on Webster delay model and HCM2000 delay model. These two methods are used for comparisons because they are classic optimal control methods that are reported to perform well on their studied problems.

For a fair comparison among all three approaches, in all cases, we test them using the same intersection saturation, intersection topology, and road channelization. In the experiments, we chose the following evaluation metrics to examine these three approaches.

- The average vehicle delay of intersection
- Traffic capacity of intersection per cycle

In the experiments, temperature $T = 10^5$, $\beta = 0.95$, learning rate $\alpha = \frac{1}{n+1}$, where n is learning times of a state, $a_{\min} = 10$, $a_{\max} = 60$, and $L_{\max} = 50$. After the training process, optimal control policy is obtained and applied to the signal control of typical intersection. The simulation results are shown in Figure 2-7.

All the three methods can be utilized to implement optimal control of traffic flow in the case of non-saturated traffic conditions of intersection (e.g. 0.8). Figure 2 and 3 show the experiment results by using these three methods. From the two figures, we can see that our approach has much lower delay and high traffic capacity than the methods respectively based on Webster and HCM2000 models. Compared to Webster and HCM2000 methods [16], our approach reduce intersection delay by 13.5% and 13%, and traffic capacity increases by 4% and 3.9% respectively.

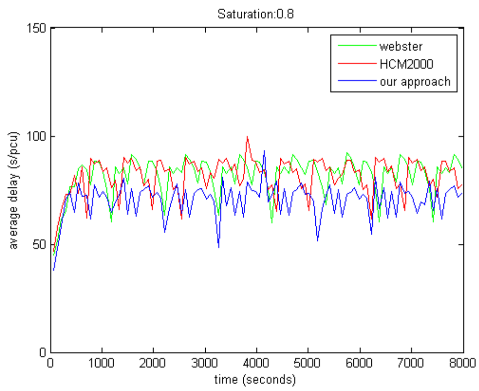

Figure 2. Average vehicle delay of intersection with the three approaches for saturation 0.8

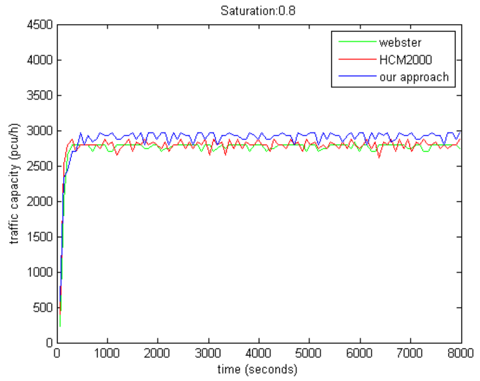

Figure 3. Traffic capacity of intersection with the three approaches for saturation 0.8

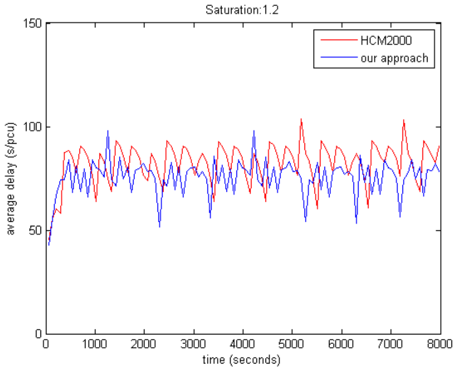

Figure 4. Average vehicle delay of intersection with the three approaches for saturation 1.2

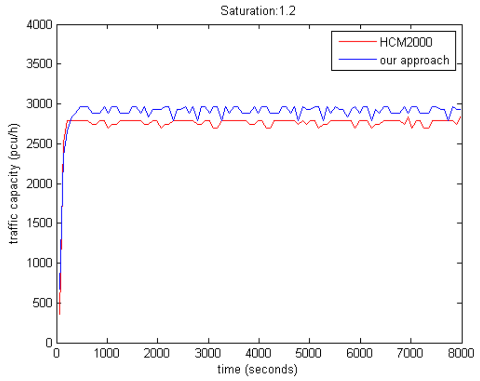

Figure 5. Traffic capacity of intersection with the three approaches for saturation 1.2

Webster model cannot deals with traffic signal control when traffic condition of intersection is oversaturated. For oversaturated traffic conditions in intersection, HCM2000 method and our proposed approach are used to carry out optimal control for intersection. Figure 4 and 5 show the simulation results when intersection saturation is about 1.2. Accordingly, figure 6 and 7 are the experimental results corresponding to saturation 1.5.

For intersection saturation 1.2, in comparison to the HCM2000 based method, our approach has 8.2% reduction in intersection delay and 4.8% increase in traffic capacity. When intersection saturation increases to 1.5, intersection delay using our approach is reduced by 10.1% than that using the HCM2000 based method. Traffic capacity using our approach increases by 7.9% than the HCM2000 based method.

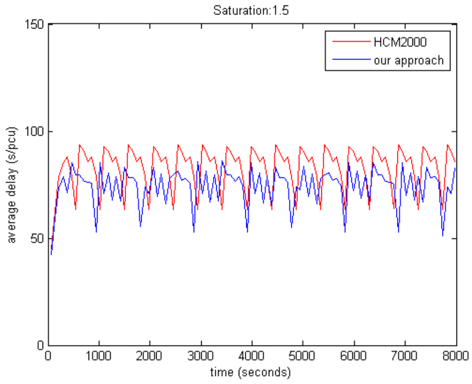

Figure 6. Average vehicle delay of intersection with the three approaches for saturation 1.5

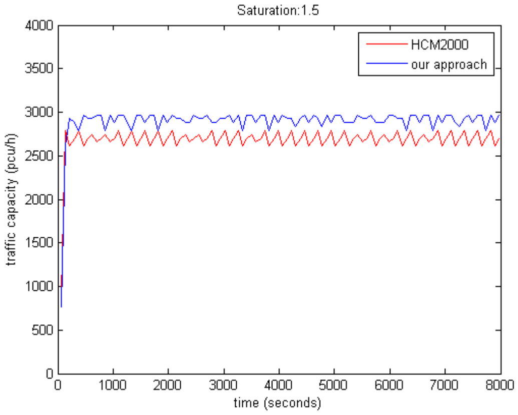

Figure 7. Traffic capacity of intersection with the three approaches for saturation 1.5

In summary, for different traffic conditions of intersection, our proposed approach achieves the best overall performances in comparison to these methods respectively based on Webster and HCM2000 models. Furthermore, this proposed approach has better performances with intersection saturation increases. It is suitable for the control of oversaturated traffic conditions.

IV. Conclusion

Urban congestion and oversaturation have become serious issues in social and economic concerns around the world. Lack of field data which is necessary for traffic modeling and model parameter calibration, it is highly difficult for typical intersection models such as Webster and HCM2000 to achieve optimal control for complex and oversaturation traffic flow of intersection. Based on ACP approach, we develop an artificial traffic system and proposed a self-learning control strategy for typical intersection. In order to avoid difficulty of obtaining intersection field data, this proposed approach can find optimal control policy through agent trial-and-error interaction with its traffic environment. The state space is reduced significantly by introducing the definitions of non-conflicted traffic flow for all approaches in the intersection. Our approach can be apply to optimal control of traffic flow under different saturation of intersection, and has considerably higher traffic capacity and lower delay than the optimal control methods which respectively based on Webster and HCM2000 models

This approach offers an interesting paradigm for problem-solving of complex traffic system. The future work is to extend our proposed approach to coordination control for large-scale urban road network.

REFERENCES

[1] Webster F V. Traffic signal settings [R]. Road Research Technical Paper No. 39. London: HMSO, 1958.

[2] Transportation Research Board. Highway capacity manual 2000 [R]. Washington D C: National Research Council, 2000.

[3] Q. Wang, X. Tan, and S. Zhang, Signal timing optimization of urban single-point intersections, Journal of Traffic and Transportation Engineering, vol. 6, pp. 60-64, June 2006.

[4] Roger P R, Prassas E S, Mcshane W R., Traffic Engineering [M]. 3rd ed. New J ersey: Pearson Prentice Hall, Pearson Education Inc., 2004: 494 - 495.

[5] Akcelik R, Rouphail N M., Estimation of delays at traffic signals for variable demand conditions [J]. Transportation Research, 27B(1):109-131, 1993.

[6] Akcelik R, Rouphail N M., Overflow queues and delays with random and platooned arrivals at signalized intersections [J]. Journal of Advance Transportation, 28 (2): 227-251, 1994.

[7] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning). Cambridge, MA: MIT Press, Mar. 1998.

[8] B. Abdulhai, R. Pringle, and G. Karakoulas, “Reinforcement learning for true adaptive traffic signal control,” J. Transp. Eng., vol. 129, no. 3, pp. 278–285, May/Jun. 2003.

[9] M. Shoufeng, L. Ying, and L. Bao, “Agent-based learning control method for urban traffic signal of single intersection,” Journal of Systems Eng., vol. 17, no. 6, pp. 526-530, 2002.

[10] B. Abdulhai, R. Pringle, and G. Karakoulas, “Reinforcement learning for true adaptive traffic signal control,” Journal of Transportation Engineering, vol. 129, pp. 278 – 285, 2003.

[11] Prashanth L. A. and Shalabh Bhatnagar, Reinforcement Learning With Function Approximation for Traffic Signal Control, IEEE transaction on intelligent transportation systems, vol.12, no.2,pp.412-421, June 2011

[12] F.-Y. Wang, “Computational theory and methods for complex systems,” China Basic Sci., vol. 6, no. 41, pp. 3-10, 2004.

[13] F.-Y. Wang, “Artificial societies, computational experiments, and parallel systems: An investigation on computational theory of complex social economic systems,” Complex Syst. Complexity Sci., vol. 1, no. 4, pp. 25–35, 2004.

[14] F.-Y. Wang, “Parallel system methods for management and control of complex systems,” Control Decision, vol. 19, no. 5, pp. 485-489, 2004.

[15] Fei-Yue Wang, Parallel control and management for ITS: concepts. Architecture, and applications, IEEE transaction on intelligent transportation systems, vol.11, no.3, pp.630-638, September 2010.

[16] JIANG Jinyong, YUN Meiping , YANG Peikun, Optimal Cycle Length Estimation Equations Based on Delay Models of HCM 2000, JOURNAL OF TONGJI UNIVERSITY (NATURAL SCIENCE), vol.37, no. 8, pp. 1024-1028, August 2009.