Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-5 (2016) 291–296

A Learning-based Adaptive Group-based Signal Control System under Oversaturated Conditions ⋆

Junchen Jin * **Xiaoliang Ma** *

* *System Simulation & Control, Department of Transport Science, KTH Royal Institute of Technology, Sweden (e-mail: junchen@kth.se & liang@kth.se).*

Abstract: The operation of traffic signal control is of significant importance in traffic management and operation practice, especially under oversaturated condition during the morning and afternoon peak hours. However, the conventional signal control systems showed the limitations in signal timing and phasing under oversaturated situations. This paper proposes a multi-agent adaptive signal control system in the context of group-based phasing techniques. The adaptive signal control system is able to acquire knowledge on-line based on the perceived traffic states and the feedback from the traffic environment. Reinforcement learning with eligibility trace is applied as the learning algorithm in the multi-agent system. As a result, the signal controller makes an intelligent timing decision. Feature-based function approximation method is incorporated into reinforcement learning framework to improve the learning efficiency as well as the quality of signal timing decisions. The learning process of the learning-based signal control is carried out with the aid of a microscopic traffic simulation model. A benchmarking system, an optimized group-based vehicle actuated signal control system, is compared with the proposed adaptive signal control systems. The simulation results show that the proposed adaptive group-based signal control system has the potential to improve the mobility efficiency under different congested situations.

Keywords: Intelligent transport system; adaptive signal control; group-based phasing; multi-agent system; reinforcement learning; oversaturated signal control.

1. INTRODUCTION

Adaptive signal control is one of the most important active traffic managements. In the adaptive signal control system, the timing scheme changes according to the live traffic conditions. Currently, several adaptive signal control systems have been developed and even deployed in urban areas around the world, such as ACS-Lite (Luyanda et al., 2003), FITS (Jin et al., 2016), CRONOS (Boillot et al., 2006) and so on. The recent developments of adaptive signal control system focus more on the learning-based method to capture the uncertainties in the transportation system. For example, El-Tantawy et al. (2013) implemented a learning-based signal controller (MARLIN-ATSC) on a large-scale network. Their study shows that MARLIN-ATSC system can significantly improve traffic mobility and environmental efficiency. However, the utmost challenge of all the learning-based signal control systems is to learn efficiently from traffic environment and further properly behave under the dynamic traffic situations.

Besides, most of the currently deployed signal controllers are using group based phasing techniques in the European countries. Group-based signal control owns its ability to allocate the signal times to individual traffic movements rather than a collection of compatible traffic movements. Due to the flexibility of signal timings, group-based signal control can be well-performed when traffic demands are not balanced in different directions at an isolated intersection (Jin and Ma, 2014). However, most of the prevalent group-based signal controllers apply conventional signal timing strategies that do not consider the current traffic conditions for the whole intersection. A sudden change in traffic demand may also downgrade the performances, on the efficiency of traffic mobility, of group-based control using the conventional signal timing strategies.

The conventional signal control system does not work well in during oversaturated or unusual load conditions due to the limitations of signal timing and phasing. For instance, a commonly used signal control system, vehicle actuated control, extends green time depending the detected traffic demand. The green extension is limited to a pre-defined maximum green time. However, the green extension is probably authorized under oversaturated conditions due to the high demand detected. Therefore, vehicle actuated cannot adjust its timing under oversaturated conditions. This study aims to propose an adaptive group-based signal control system which can address the limitations of conventional signal control systems under oversaturated conditions. In this paper, the group-based signal control scheme is formulated as a multi-agent system capable of learning from the traffic environment. Reinforcement learning with eligibility trace is implemented as the learn-

⋆ This work is partially supported by J. Gust. Richert Foundation.

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.07.128

ing algorithm in this multi-agent framework. Simulation-based test-bed experiments are carried out to assess the performance of the proposed learning-based signal control system under different congested traffic situations.

2. MULTI-AGENT GROUP-BASED SIGNAL CONTROL SYSTEM

In the concept of group-based signal control, signal group and phase are two basic components. Signal group denotes a traffic movement or a collection of a few traffic movements. Timings are directly assigned to each signal group. All of the compatible signal groups have the possibility to form a phase. During the operation of group-based phasing, if a signal group is ordered to terminate, the system will automatically search for another signal group as the candidate. The set of substitute signal groups is defined as the "candidate signal groups". If the candidate signal groups are not existing, the ordered-to-terminate signal group has to wait until all signal groups in the current phase are ready to terminate. During the waiting time, the ordered-to-terminate signal group shows green indication but detection information is not reported to determine signal timings. Such a green period is named as a passive green time.

The signal group cannot be nominated as a "candidate signal group" concerning if it has already been activated in the current cycle, or it has conflicts with the rest of signal groups in the current phase. Conflict matrix is used to represent the conflicts among signal groups. The left diagram in Fig. 1 shows a typical example for conflict matrix. The gray square indicates that signal groups can be served simultaneously. Inter-green times between signal groups are assigned to gray squares. In addition, the right diagram of 1 gives an example of group-based phasing operations. Assume signal group $SG1$ has been activated. Signal group $SG1$ is able to combine with either $SG2$ or $SG3$ or $SG4$. Phase $PH1$, phase $PH2$ and phase $PH3$ respectively represent three possible combinations. In practice, the decision, regarding which signal group to combined with, depends on the received detection information. Consequently, multi-phase pictures can be generated by group-based phasing techniques in the real-world operations.

In principle, group-based phasing suits the principle of the multi-agent framework. Every signal group is considered as an individual agent. Timing scheme is a consequence of actions made by agents. In this multi-agent framework, central level of manipulation is not required such that each agent pursues its goal based on own knowledge. From a practical point of view, the interactions between traffic environment and signal group agents occur at the discrete time. At each learning step, all of the signal group agents perceive states and feedback from the traffic environment. Signal group agents are also able to receive information from other agents and incorporate the information into their decision-making process. The cooperation between agents is achieved by sharing partial information of the states with their neighborhoods. Therefore, the final signal timing decision is made by considering a trade-off between the agent's preferences against those of the other agents. The active agent learns knowledge based on the received states as well as the immediate feedback caused by the previous action. The learning algorithm is implemented in this multi-agent signal control system. Accordingly, actions are selected by the agents concerning a certain selection strategy and also the newly acquired knowledge.

3. INTELLIGENT TIMING BY REINFORCEMENT LEARNING

3.1 Temporal Difference(λ)

In the multi-agent signal control system, signal group agent $i, i \in n$ possesses a finite state set $\mathcal{X}_i$ and a finite action set $\mathcal{U}_i$, where n is the number of signal groups associated with the intersection. The state of signal group agent is interpreted by the detection information. Therefore, the generalization of this signal control system at an isolated intersection can be represented by a tuple $< \mathcal{U}_1, ..., \mathcal{U}_n, \mathcal{X}_1, ..., \mathcal{X}_n >$. The operation process of group-based signal control is considered as a finite and discrete-time stochastic decision process. That is, signal group agent i perceives a sequence of states $\{\boldsymbol{x}_{i,t}\}$ and is governed by a control sequence $\{\boldsymbol{u}_{i,t}\}$, where $\boldsymbol{x}_{i,t}$ and $\boldsymbol{u}_{i,t}$ respectively denote the state variable and control variable at time t. The notation $\boldsymbol{u}_{i,t_1:t_2}$ and $\boldsymbol{x}_{i,t_1:t_2}$ for $t_1 \leq t_2$ where $\boldsymbol{u}_{i,t_1:t_2} = \boldsymbol{u}_{i,t_1}, \boldsymbol{u}_{i,t_1+1}, \boldsymbol{u}_{i,t_1+2}, ..., \boldsymbol{u}_{i,t_2}$ and $\boldsymbol{x}_{i,t_1:t_2} = \boldsymbol{x}_{i,t_1}, \boldsymbol{x}_{i,t_1+1}, \boldsymbol{x}_{i,t_1+2}, ..., \boldsymbol{x}_{i,t_2}$ are respectively denoted as a sequences of actions and states from t_1 to t_2. Suppose that signal group agent i starts from state $\boldsymbol{x}_{i,0}$. The initial action made by this agent in the initial state is defined as $\boldsymbol{u}_{i,0}$.

Reinforcement learning (RL) is able of finding an optimal solution without completely knowing the environment dynamics. RL runs with a stochastic iterative algorithm using the observations obtained from online samples of state-action trails. Temporal Difference(TD) algorithm is one classical type of reinforcement learning algorithms capable of recursively estimating the maximum expected cumulative reward (Barto, 1998). TD algorithms aim at finding an optimal solution without completely knowing the environment dynamics. TD algorithms work by an on-line updating procedure by which Q-factor is immediately updated after the state being ever visited. In practice, decisions made by signal controllers usually have effects on the several following states. Temporal difference algorithm with multiple-step backups enables a signal group agent to look backward all the way to the beginning of the defined learning horizon. The traces decay gradually over time. TD(λ) is the multiple-step backups version of TD algorithm. TD(λ) utilizes eligibility trace to achieve the average effects of multiple-step backups. SARSA(λ) is an on-policy TD(λ) algorithm that estimates Q-factor concerning a specific behavior policy.

Consider the following general scenario, the state of an active signal group agent i is $\boldsymbol{x}_{i,t}$ and the agent takes action $\boldsymbol{u}_{i,t}$ at time point t. Then the agent receives reward value $r_{i,t+1}$ and its state vector becomes $\boldsymbol{x}_{i,t+1}$. The estimated optimal cumulative reward corresponding to state-action pair $(\boldsymbol{x}_{i,t}, \boldsymbol{u}_{i,t})$ is denoted as $Q_{i,t}(\boldsymbol{x}_{i,t}, \boldsymbol{u}_{i,t})$ at time t. The following equation presents the update process for signal group agent i in regards to all the state-action pairs at time step $t+1$ by applying TD(λ) algorithm: