Knowledge-Based Systems 110 (2016) 275–292

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Learning traffic signal phase and timing information from low-sampling rate taxi GPS trajectories

Juan Yu, Peizhong Lu*

School of Computer Science and Technology, Fudan University, Shanghai 200433, China

ARTICLE INFO

Article history:
Received 29 February 2016
Revised 14 July 2016
Accepted 27 July 2016
Available online 28 July 2016

Keywords:
Traffic signal phase and timing information (TSPaT)
Phasing scheme
Timing plan
Cycle length
Green length
Taxi GPS trajectory
Approximate greatest common divisor (AGCD)

ABSTRACT

Traffic signal phase and timing (TSPaT) information is valuable for various applications, such as velocity advisory systems, navigation systems, collision warning systems, and so forth. However, the acquisition of the TSPaT information in the city-scale is very challenging. In this paper, we propose a framework to learn the TSPaT information from low-sampling rate taxi GPS trajectories. Specifically, our framework could learn: the phasing scheme, i.e., the number of phases and the assignment of traffic movements to phases; timing plans, including the cycle length and green lengths of phases within a cycle, for each given fixed-time signalized intersection. In our framework, the cycle length is the first important parameters to be learned. We formalize the cycle length estimation problem as a general approximate greatest common divisor (AGCD) problem, and propose the most frequent AGCD (MFAGCD) algorithm to solve the problem. The MFAGCD algorithm is robust to noises and outliers, and could estimate the cycle length with a high accuracy using a small number of green-start times extracted from taxi GPS trajectories. Based the correlation between phases, we propose an all-direction joint determination method to jointly estimate green lengths using green-start times and cross-over times from all phases. The effectiveness of our framework is experimentally evaluated on three selected fixed-time signalized intersections in Shanghai, China.

1. Introduction

City-scale traffic signal phase and timing (TSPaT) information, including the phasing scheme, the cycle lengths and green lengths of timing plans, is an indispensable input for various in-vehicle applications. It enables velocity advisory systems to calculate optimal velocity trajectories which help vehicles to reduce idling time at red signals, thereby improving the fuel efficiency and lowering green house gas emissions [1–3]. Koukoumidis et al. [4] showed that their velocity advisory system can reduce fuel consumptions by 20.3% on average. Audi company has also conducted a nationwide field trial on its driving assistance system after integrating traffic light information, and demonstrated that CO_2 emissions could be reduced by up to 15% [5]. The traffic signal timing information could also help navigation systems to find arterial routes with less idling delay, and to provide more accurate trip time estimation [6]. In addition, collision warning systems could benefit from the signal timing information by warning against potential signal violations.

However, the acquisition of the TSPaT information in the city-scale is very challenging. The diversity of traffic signal control strategies and uncertainties caused by clock drift render the direct access to the information from local authorities prohibitively difficult [7]. The US and European transportation agencies have promoted to equip traffic signals with short range antennas for broadcasting the traffic signal state information. However, this proposal has not been widely implemented because of the significant cost of antennas' deployment. Koukoumidis et al. [4] designed an infrastructure-free framework to collaboratively collect the state information of traffic signals using mobile phones. However, their framework could bring about heavy burdens to users' mobile phones, and requires a large amount of users to participate in.

In this paper, we aim to learn the TSPaT information of fixed-time signalized intersections from low-sampling rate taxi GPS trajectories. Taxis acting as moving sensors are able to sense the state information of traffic signals during their operation. For a specific intersection, each passing taxi could sample a small piece of traffic signal state information. Through aggregating all pieces of the state information sensed by all passing taxis, a relative larger set of observations about the state indication of traffic signals could be collected, and therefore makes it possible to learn the TSPaT information via the crowdsourcing manner.

Taxi GPS trajectories are economic resources for learning the TSPaT information in a large city scale. As taxis in many big cities of China have installed GPS devices for real-time delivery of empty

* Corresponding author.
E-mail addresses: juanyu10@fudan.edu.cn (J. Yu), pzlu@fudan.edu.cn (P. Lu).

http://dx.doi.org/10.1016/j.knosys.2016.07.036
0950-7051/

taxis and fleet management, large volumes of taxi GPS trajectories can be collected from tens of thousands of taxis during their operation every day. Therefore, unlike the method proposed in [4], there is no requirement on the participation of other system users, and no extra cost involved for data collection. In addition, the coverage of taxis over the road network is relatively wide compared with other types of probe vehicles, such as buses, private vehicles, and so forth, because driving routes of taxis are diverse and cover more road segments than buses and private vehicles.

Accurately learning the TSPaT information from taxi GPS trajectories is not easy. Firstly, sampling rates of taxi GPS devices are often relatively low, e.g., one point every 30 s –5 min, and we refer to low-sampling rate as one point every 30 s or above in this paper. Although the sampling rates are highly enough for most existing applications, they pose a great challenge for the TSPaT information learning problem. For one thing, the low sampling rate would make it difficult to accurately infer the state information of traffic lights from taxi GPS trajectories, as the travel speed of a taxi may vary a lot during the sampling interval. For another, the low sampling rate would result in the incomplete observation about the state changes of traffic lights, because the duration of green/red lights within each cycle might be shorter than 30 s at some intersections. Secondly, taxi locations recorded in GPS trajectories are not accurate, and the GPS positioning error is often about 10–20 m in city environment. Thirdly, the taxi penetration rate at each intersection is often relatively low, seeing from Table 2 that the least average taxi arrival interval is about 2.5 min. Finally, fixed-time intersections might use different timing plans at different times of day. For example, an intersection might partitions a day into peak hours and off-peak hours, and uses one timing plan at peak hours and the other timing plan at off-peak hours. By different timing plans we mean that the cycle lengths are different.

The main idea of our proposed framework for learning the TSPaT information is that we firstly propose a procedure to extract critical times for each intersection individually, including green-start times and cross-over times, from taxi GPS trajectories; then we use the extracted critical times to learn the TSPaT information for corresponding intersection, i.e., the phasing scheme and timing plans used at the intersection. The critical times extraction procedure is motivated by the observations:

(1) taxis crossing an intersection with stop-and-go movement patterns could sample green-start times;
(2) the times at which taxis crossed over corresponding stop lines, i.e., cross-over times, are snapshot green times of corresponding traffic lights.

The TSPaT information learning procedure using the extracted critical times for each intersection consists of three stages. The first stage is to learn the daily pattern of timing plans from a sequence of cycle lengths estimated hourly using green-start times. The second stage is to learn the phasing scheme using the green-start times and the learned daily pattern. Finally, green lengths of phases for each timing plan are estimated in the third stage using the extracted green-start times and cross-over times incorporated with the learned daily pattern and phasing scheme.

Our main contributions are as follows. Firstly, we formalize the cycle length estimation problem as the general AGCD problem, and propose the Most-Frequent Approximated Great Common Divisor (MFAGCD) algorithm to solve it. The MFAGCD algorithm is robust to noises and outliers, and could estimate the cycle length with high probability from a small number of green-start times. Secondly, we introduce a method to learn the phasing scheme using the estimated cycle length, green-start times, and the prior knowledge about the minimal length of phase green time. The phasing scheme is beneficial for green lengths estimation. Because we can bring sets of green-start times (cross-over times) belonging to the same phase together according to the phasing scheme. Finally, we propose an all-direction joint determination method to estimate green lengths for each timing plan. The method transforms the green lengths estimation problem into an optimization problem, which takes advantage of the correlation between phases to jointly determine green lengths from all extracted green-start times and cross-over times, and thereby improving the estimation accuracy. Then, we propose an exhaustive search algorithm to solve the optimization problem.

The remainder of this paper is organized as follows. In Section 2, we survey related works about traffic signal information collection. Section 3 introduces correlated definitions and gives the formal statement about the TSPaT information learning problem. Section 4 presents the procedure of extracting green-start times and cross-over times from taxi GPS trajectories. In Section 5, we introduce how to learn the phasing scheme and timing plans of given intersections according to these extracted green-start times and cross-over times. Experimental results based on the real taxi trajectory data are demonstrated in Section 6. Finally, a conclusion is given in Section 7.

2. Related work

To seek the convenient and economic way of collecting the TSPaT information, several researchers have recently started to leverage GPS traces of probe vehicles, including taxis, buses, and so forth. Kerper et al. [8] designed a traffic light state estimation method using shared velocity profiles from participated drivers. However, they only evaluated their methods on simulation data, and did not consider the impact of non-uniform sampling rate of velocity profiles on the performance of their method. Fayazi et al. [9] demonstrated the feasibility of estimating cycle lengths, red durations, and starts of greens using bus traces collected in San Francisco, CA, USA. Their methods were special for fixed-time intersections which use constant cycle length and green(red) lengths every day. In addition, the accuracy of their methods could be significantly affected by the accumulated error, as they required long-term observations for their estimation. Chuang et al. [10] introduced shockwave models for discovering the phase timing information of target intersections from GPS traces. The DBSCAN [11] clustering algorithm was used to estimate the cycle length from a set of stop events in the situation that the cycle length was unknown. However, their methods focused on high-sampling rate GPS traces with the sampling rate 1 Hz. Zhu et al. [12] proposed a supervised method to learn the state sequences of target traffic signals from taxi GPS trajectories. However, the supervised method depends on labeled datasets for learning the joint probability distributions of speed and distance to traffic light conditioned on the state of traffic lights, and the acquisition of the labeled training data is costly and intersection dependent.

In comparison with these closely related works [8], [9] and [10], advantages of our work are as follows. Firstly, our work could learn both the phasing scheme and timing plans of traffic signals, whereas the existing works focused on learning timing plans, including cycle length and green lengths. Secondly, our work could detect cycle length changes, whereas the related works did not consider this situation. For example, the authors in [9] considered offset changes of green-start times, but they did not consider the problem of cycle length changes which is considered in this paper. Thirdly, our method could support a relatively larger scale of the TSPaT information collection, due to the wide coverage of the taxi GPS trajectory data. Finally, the proposed cycle length estimation algorithm has better noise and sparsity tolerant capabilities than the method in [9].

Trajectory data management and mining has become one of the most popular research topics in data mining area in recent