Transportation Research Part B 113 (2018) 91–120

Contents lists available at ScienceDirect

Transportation Research Part B

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

Data-driven distributionally robust optimization approach for reliable travel-time-information-gain-oriented traffic sensor location model

Ning Zhu, Chenyi Fu*, Shoufeng Ma

Institute of Systems Engineering, College of Management and Economics, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 14 December 2017
Revised 13 May 2018
Accepted 14 May 2018

Keywords:
Traffic network sensor location
Travel time information gain
Distributionally robust optimization
ϕ-divergence

ABSTRACT

Travel time is one of the most intuitive pieces of traffic information to help decision makers to control real-time traffic conditions and to guide travelers to choose a reasonable route. An optimal sensor location scheme can obtain reliable route travel time information. Most current travel-time-oriented sensor location models are deterministic and assume a given and correct travel time probability density function. Nevertheless, due to widespread observational and systematic errors, prior travel time information is not accurate or reliable. In our study, a novel data-driven link-based network sensor location method is proposed to maximize travel time information gain. The effect of route differentiation is considered, and the sensors are located at links rather than at nodes. In addition, to account for the uncertainty in the prior travel time distribution, the distributionally robust travel time information gain sensor location (DRTTIGSL) model is presented. The prior distribution information is taken into account based on a statistical measure called ϕ-divergence. The ϕ-divergence is used to construct the uncertainty set. The reformulation of DRTTIGSL is dependent on the choice of ϕ-divergence and is tractable. Extensive numerical experiments are conducted to verify the effectiveness of the DRTTIGSL model. Compared with the optimal solutions for the deterministic model, the optimal solutions for the DRTTIGSL model can reduce the worst-case situation with a small price of the average objective value, especially when the total budget is not large.

1. Introduction

Traffic sensors play an important role in collecting real-time information for transportation systems. Well-designed transportation surveillance systems help to build effective traffic control and guidance systems. Generally, the traffic sensor location problem in the literature is divided into several problems, including the OD flow estimation problem (Yang and Zhou, 1998; Castillo et al., 2008a; Hu and Liou, 2014), link flow observation or estimation problem (Hu et al., 2009; Ng, 2012; He, 2013; Ng, 2013; Zhu et al., 2014), route reconstruction problem (Castillo et al., 2008b; Zangui et al., 2015; Fu et al., 2016; 2017), and travel time estimation problem (Li and Ouyang, 2011; Xing et al., 2013; Zhu et al., 2016). There are two main types of sensor: passive and active. Passive sensors are relatively cheap and can obtain flow volume, speed and occupancy

* Corresponding author.
E-mail addresses: zhuning@tju.edu.cn (N. Zhu), chenyi_fu_201509@163.com (C. Fu), sfma@tju.edu.cn (S. Ma).

https://doi.org/10.1016/j.trb.2018.05.009
0191-2615/

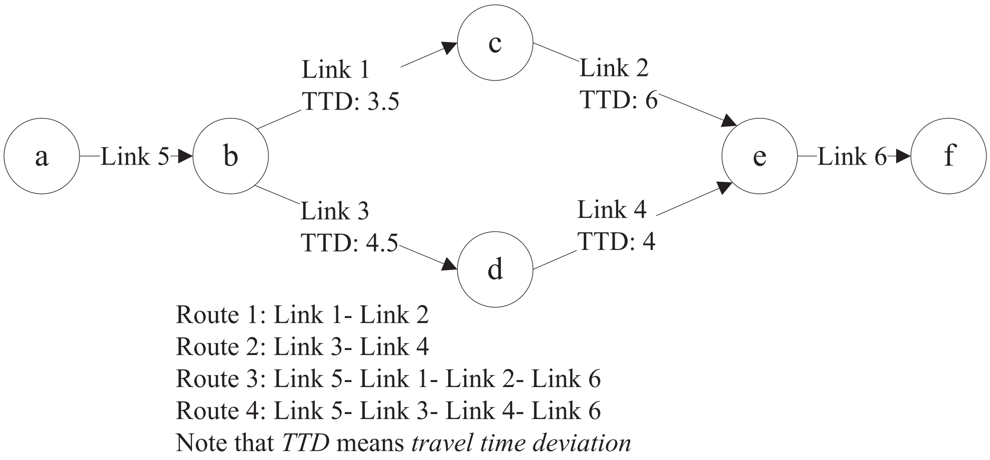

Fig. 1. Simple network with six nodes and six links.

Table 1
Travel time deviations based on different scenarios in the simple network.

Scenario	Travel time deviation.					
	Link 1	Link 2	Link3	Link 4	Route 1	Route 2
1	3	4	4	3	7	7
2	1	3	6	5	4	11
3	6	8	3	4	14	7

information. Active sensors have the capabilities of passive sensors and can also recognize vehicle plates, but they are more expensive than passive sensors.

Prior data are widely used in traffic sensor location problems. From the view of mathematical programming, prior data are normally adopted as a measurement in the objective function (Yang and Zhou, 1998; Ehlert et al., 2006; Zhou and List, 2010; Xing et al., 2013; Castillo et al., 2008b). The purpose of these models is to obtain an estimation or observation that is as close to the prior data as possible. For the travel-time-oriented traffic sensor location models, the commonly adopted objective function is to make the travel time coverage or variation as large as possible (Mirchandani et al., 2009; Li and Ouyang, 2011) or to make the gap between the estimation and the prior data as small as possible (Danczyk and Liu, 2011; Zhu et al., 2016). Therefore, prior data play a critical role in determining the location of traffic sensors.

Most studies have assumed that prior information is accurate and reliable. However, in practice, collected traffic data are far from accurate. There are various error sources, including measurement error, sensor failure, severe weather, accidents, and unpredictable travel behavior. These unavoidable errors make it very difficult to obtain accurate information. From another angle, we can state that it is impossible to determine the true travel time probability density function. For example, the travel time information derived from GPS in urban area is often biased due to GPS signal issues. Fu et al. (2017) show that the travel time data between two toll stations of a freeway network may be multimodal because of the existence of multiple routes. For these cases, traditional statistical approaches may fail to fit a reliable distribution.

Traditional traffic sensor location models aim to maximize the coverage of travel time information, which is based on the fact that the travel time probability density function is known and correct (Danczyk and Liu, 2011; Xing et al., 2013; Park and Haghani, 2015). Statistical measurements of the true probability density function are used to calculate the travel time coverage. However, it is highly likely that the prior data are low quality because of the many error sources. The true probability density function is unknown. Fig. 1 shows a simple example with six nodes, six links, and two routes. Route 1 contains links 1 and 2 while route 2 consists of links 3 and 4. The travel time deviation is shown in Fig. 1. For the traditional coverage model, which is deterministic, the traffic sensors are located to cover route 1 since the total travel time deviation of link 1 and 2 is high. Nevertheless, the travel time probability density function is unknown. There are a number of scenarios for the travel time deviation for each link (see Table. 1). Each scenario is associated with a probability. The true travel time probability density function is hidden in these scenarios. Furthermore, assume that there are two additional routes in the network, i.e., route 3 contains links 5, 1, 2 and 6 while route 4 consists of links 5, 3, 4 and 6. Thus, routes 1 and 2 can be regarded as subroutes of routes 3 and 4, respectively. If two sensors are placed at the end of link 5 and the beginning of link 6, the travel time deviations of routes 1 and 2 are covered but cannot be differentiated because they are obtained by the same sensor pair. In practice, the mean travel time deviation of routes 1 and 2 will be used to estimate their true travel time deviation, which will result in overestimation or underestimation. Hence, to obtain reliable travel time deviation information