Transportation Research Part C 68 (2016) 300–310

Contents lists available at ScienceDirect

Transportation Research Part C

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

Network sensor health problem

Zhe Sun [b], Wen-Long Jin [a,*], ManWo Ng [c]

[a] Department of Civil and Environmental Engineering, California Institute for Telecommunications and Information Technology, Institute of Transportation Studies, 4038 Anteater Instruction and Research Bldg, University of California, Irvine, CA 92697-3600, United States

[b] Department of Civil and Environmental Engineering, Institute of Transportation Studies, 4191 Anteater Instruction and Research Bldg, University of California, Irvine, CA 92697-3600, United States

[c] Department of Information Technology and Decision Sciences, Strome College of Business, 2165 Constant Hall, Old Dominion University, Norfolk, VA 23529, United States

ARTICLE INFO

Article history:
Received 27 August 2015
Received in revised form 1 April 2016
Accepted 19 April 2016

Keywords:
Network flow conservation
Least inconsistent base set
Health index
Greedy search algorithm
Caltrans Performance Measurement System (PeMS) database

ABSTRACT

Many existing studies on the sensor health problem determine an individual sensor's health status based on the statistical characteristics of collected data by the sensor. In this research, we study the sensor health problem at the network level, which is referred to as the network sensor health problem. First, based on the conservation principle of daily flows in a network, we separate all links into base links and non-base links, such that the flows on the latter can be calculated from those on the former. In reality, the network flow conservation principle can be violated due to the existence of unhealthy sensors. Then we define the least inconsistent base set of links as those that minimize the sum of squares of the differences between observed and calculated flows on non-base links. But such least inconsistent base sets may not be unique in a general road network. Finally we define the health index of an individual sensor as the frequency that it appears in all of the least inconsistent base sets. Intuitively, a lower health index suggests that the corresponding sensor is more likely to be unhealthy. We present the brute force method to find all least inconsistent base sets and calculate the health indices. We also propose a greedy search algorithm to calculate the approximate health indices more efficiently. We solve the network sensor health problem for a real-world example with 16 nodes and 30 links, among which 18 links are monitored with loop detectors. Using daily traffic count data from the Caltrans Performance Measurement System (PeMS) database, we use both the brute-force and greedy search methods to calculate the health indices for all the sensors. We find that all the four sensors flagged as unhealthy (high value) by PeMS have the lowest health indices. This confirms that a sensor with a lower health index is more likely to be unhealthy. Therefore, we can use such health indices to determine the relative reliability of different sensors' data and prioritize the maintenance of sensors.

1. Introduction

Since the introduction of the first known vehicle sensor[1] in 1928 at a signalized intersection, researchers have devoted significant efforts to create and improve sensor systems that monitor vehicle presence and passage at critical locations on

* Corresponding author.
E-mail address: wjin@uci.edu (W.-L. Jin).

[1] We use sensor and detector interchangeably in this research.

http://dx.doi.org/10.1016/j.trc.2016.04.015

streets and freeways. The collected data are then used to monitor traffic congestion and incidents on freeways, estimate travel times, and support decision making for transportation agencies. Many different types of traffic sensors, including inductive loop detectors, magnetic sensors, video image detectors, and microwave detectors, have been developed. Because of their high reliability under different weather conditions, loop detectors have been widely deployed to provide uninterrupted traffic measurements, including occupancy and flow counts.

In California, loop detectors are embedded in many freeway pavements, and 30-s and 5-min occupancies and traffic counts are available through the Caltrans Performance Measurement System (PeMS). However, loop detector data can be corrupted by noises and errors, due to pavement/saw-cut failures, intermittent communications, double counting of lane-changing vehicles, and so on (Coifman, 2006). According to PeMS, which is a widely used data source for the freeway sensor system in California, only 67% of the detectors were working properly in May, 2014 (Chen and Petty, 2001). Some districts (e.g. district 6 in Los Angeles County) have even lower proportions of working detectors. Thus, it is important to determine the health status of a detector, before one can use its data to estimate congestion levels and other traffic characteristics. Such sensor health information is also critical for prioritizing the maintenance of detectors.

In the transportation literature, there are only a few studies on the sensor health problem. The study by Turochy and Smith (2000) assesses a detector's health based on the time series of flow and occupancy measurements. The proposed method places thresholds on the maxima of occupancy and volume, the numbers of samples with non-zero volume but zero speed, and the average effective vehicle lengths. A sensor's health status is then determined by the total number of its faulty records. The study by Chen et al. (2003) developed a similar method to determine the health status of a sensor by using four statistics: the number of samples with zero occupancy, the number of samples with zero flow and non-zero occupancy, the number of samples with extremely high occupancy, and the variance of flow and occupancy. The four statistics are calculated every day per sensor. The algorithm makes decisions by comparing the statistics with the predefined thresholds. Furthermore based on a classification algorithm, PeMS categorizes a sensor's health status into ten different diagnostic states, such as "line down", "controller down", and "high value" (Rajagopal and Varaiya, 2007).

The aforementioned methods are all designed to solve the sensor health problem for individual sensors, by examining whether the data produced by an individual sensor looks statistically correct. Thus we refer to these methods as local methods. Such local methods have two major limitations. First, the thresholds in the algorithms could be challenging to determine. In practice, the thresholds may vary by locations and are subject to exogenous factors such as traffic incidents, constructions, and weather conditions. Second, all existing local methods focus on each sensor's performance individually and fail to take into account the correlations between neighboring sensors in a road network.

In contrast, in this study we attempt to study the sensor health problem at the network level. Thus we refer to this problem as the Network Sensor Health Problem (NSHP). The basic idea is to cross-check the consistency among traffic flow sensors based on flow conservation. As the prerequisites, we assume a certain level of redundancy in terms of observed link flows and the majority of the flow sensors are working properly.

First, based on the conservation principle of daily flows in a network, we separate all links into base links and non-base links, such that the flows on the latter can be calculated from those on the former. In reality, the network flow conservation principle can be violated due to the existence of unhealthy sensors. Then we define the least inconsistent base set of links as those that minimize the sum of squares of the differences between observed and calculated flows. But such least inconsistent base sets may not be unique in a general road network. Finally we define the health index of an individual sensor as the frequency that it appears in all the least inconsistent base sets. Intuitively, a lower health index suggests that the corresponding sensor is more likely to be unhealthy. In the study by Waller et al. (2008), an index of network consistency was introduced to describe the agreement in the conservation of cumulative vehicle counts between neighboring sensors at a node. However, this problem is still local for individual nodes. In this study, the proposed method evaluates the flow consistency among observed traffic flows in the whole network. Thus the NSHP is more versatile, as it applies even not all links at a node are monitored (equipped with flow sensors), and more powerful as it combines information from multiple sensors.

Unlike the binary "health label" (healthy vs. unhealthy) used in most existing local methods, the health index is a number in the range of zero to one. The health indices can provide guidelines for researchers to filter unreliable traffic data and for transportation agencies to prioritize tasks for repairing and replacing sensors.

The rest of the paper is organized as follows. In Section 2, we briefly review the node-based flow conservation formulation, which was proposed by Ng (2012) to solve the network sensor location problem (NSLP). Based on this formulation, define the health index of a sensor in three steps. In Section 3, we first apply the brute-force method to solve the problem. Then we propose a greedy search algorithm to calculate the health index more effectively. In Section 4, we solve the NSHP for a freeway network on SR-91 and compare the results with the health statuses provided by PeMS. In Section 5 we summarize the results of this research.

2. Conceptual framework

The NSHP formulation is developed based on a node-based formulation of network flow conservation, which was also used by Ng (2012) to solve the Network Sensor Location Problem (NSLP). We try to identify potentially unhealthy sensors by comparing the estimated link flows with the observations. The intuition behind this is, the flow estimated from the healthy sensors measurements are expected to be more consistent with observations. The NSLP is an observability problem