DRISI
Caltrans Division of Research, Innovation and System Information

Research Results

Transportation Safety and Mobility

JUNE 2020

Project Title:
Improving Transportation Information Resilience: Error Estimation for Networked Sensor Data

Task Number: 3393

Start Date: August 23, 2018

Completion Date: December 31, 2019

Task Manager:
John Slonaker
Sr. Transportation Electrical Engineer (specialist)
john.slonaker@dot.ca.gov

Improving Transportation Information Resilience: Error Estimation for Networked Sensor Data

Identifying systematic errors in networked traffic detector data from heterogeneous sources using spatial interdependence for correlation

WHAT WAS THE NEED?

Reliable traffic data serves as a critical foundation for almost all mobility technologies in a modern transportation system, playing a significant role in improving efficiency and safety. According to Caltrans' Performance Measurement System (PeMS), about one third of the vehicle count stations that measure traffic volume and speed are not reporting data. PeMS incorporates vast amounts of networked data from various detectors and heterogeneous sources. There was a need to estimate and correct the systematic error caused by inaccurate calibration or malfunction of traffic detection instruments so that true parameter values can be recovered from poorly measured ones.

WHAT WAS OUR GOAL?

This task aimed to develop a mathematical model to identify systematic errors caused by inaccurately calibrated or malfunctioning vehicle count stations and recover true parameter values from the poorly measured ones. The proposed model sought to combine data science and network modeling techniques, which could provide greater modeling flexibility to incorporate spatial correlation of networked data, which was thought to lead to a better estimation quality.

WHAT DID WE DO?

Most sensor health monitoring studies have focused on identifying 'bad' sensors whose data should be discarded. However, the problem of estimating and correcting the systematic bias of sensor data had not been addressed, especially in the context of large-scale networked data pieces from heterogeneous sources. Systematic sensor data bias can be caused by unbalanced detector sensitivity and can result in counting neighboring lane traffic, missing motorcycles, and reporting more than one count for long vehicles. However, this data may still be valuable in revealing important information about traffic flow. In this project, the researchers sought to estimate and correct the systematic bias of the sensor data they collected by combining data science and network modeling techniques. Their model proved to be more flexible in incorporating spatial correlation of networked data, which lead to better estimation quality.

The first main research task was the creation of their mathematical model. The researchers combined a sensor measurement error model with a traffic network model. This new model uses a Generalized Method of Moments (GMM) based estimation approach to determine the parameters of both systematic and random errors of traffic sensors in a road network. The model detected both completely and partially malfunctioning sensors, denoised the data, corrected for systematically erroneous data and, finally, imputed missing data. This is illustrated in Image 1, with steps 1 and 2 representing the detection of completely and partially malfunctioning sensors, respectively. Step 3 represents standard denoising procedure. Step 4 is to correct systematically erroneous data, and step 5 is to impute missing data.

The second main task was to test the estimation quality of the model. The researchers used computer simulated and real-world traffic data for modal testing and validation. The real-world traffic data came from Caltrans' Performance Measurement System (PeMS) and included road segments and networks in urban congested areas such as Orange County and the San Francisco Bay Area.

WHAT WAS THE OUTCOME?

This research created an open source computer program that takes data from PeMS as input and identifies erroneous and missing data caused by inaccurately calibrated or malfunctioning vehicle count stations and corrects and imputes the erroneous and missing data respectively. The researchers tested the program on data from The North Orange County freeway network shown below is to demonstrate the scalability of the developed method.

WHAT IS THE BENEFIT?

The major contribution of this research is two-fold. First, the developed method can evaluate level-of-data issues and correct traffic flow data in addition to identifying malfunctioning sensors, while most previous sensor health studies concerned only the latter. Second, the method uses the entire network structure of the traffic monitoring system, while many previous studies that focused on spatial relation gave attention only to those immediately neighboring sensors on a corridor. Compared to previous research that also exploited the network feature, the method of this research lessens the requirement of flow balance on the entire network, which may take several hours to establish. Instead, the way flow balance equations are used in the method of this research only concerns the adjacent sensors at one time and thus requires much less time to establish. Therefore, it is possible for users to choose much shorter time intervals and obtain larger samples within a fixed total observation time.

IMAGES

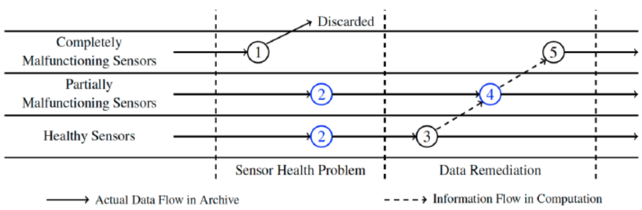

Image 1: Vehicle Detection Station Data Processing Chart

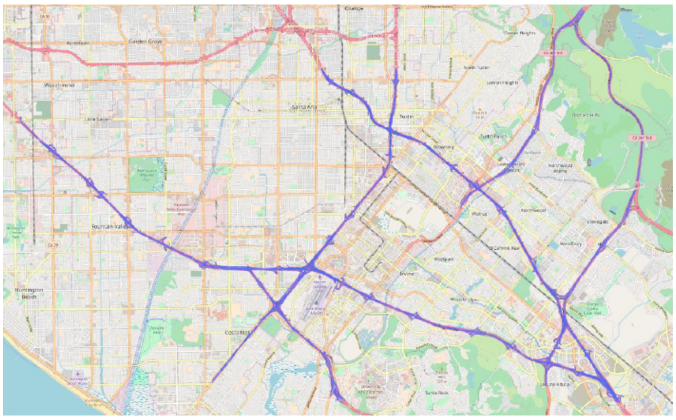

Image 2: North Orange County Freeway Network