DRISI
Caltrans Division of Research, Innovation and System Information

Research Results

Trajectory Data Mining for Performance Measurement of Public Transportation System

To develop a system that can handle massive amounts of GPS trajectories from public transportation vehicles to display different performance metrics.

Modal

MAY 2019

Project Title:
Trajectory Data Mining for Performance Measurement of Public Transportation System

Task Number: 3124

Start Date: January 1, 2017

Completion Date: March 30, 2018

Task Manager:
Stan Chow
Transportation Engineer
stan.chow@dot.ca.gov

WHAT IS THE NEED?

Improving the performance and reliability of public transportation vehicles has been one of the primary objectives for California Department of Transportation (Caltrans). The wealth of data collected from bus GPS trajectories can offer an unprecedented opportunity for analysis of public transportation systems towards reducing operation costs and increasing ridership.

WHAT WAS OUR GOAL?

Our goal was to provide an interactive web-based application that enables visualization, querying and analysis of performance metrics of buses in Los Angeles County based on real-time and historical GPS data. We created a suite of web services that implemented our statistical bus performance detection algorithms. A remote user could call these services to query data hosted on the University of Southern California (USC) Archived Traffic Data Management System (ADMS) Server.

WHAT DID WE DO?

This project developed a system that can process massive amounts of GPS trajectories from public transportation vehicles and implement statistical algorithms to analyze a variety of public transportation system performance metrics such as travel-time reliability, on-time performance, bus bunching and travel-time estimation. This project conducted fundamental research in mining and correlation of real-time and historical bus GPS trajectory datasets in LA County, which have been collected and

archived in a database over the past six years. Furthermore, to demonstrate the benefits of our research, the researchers developed a proof-of-concept web-based application to enable access and visualization of the performance metrics of public transportation vehicles. This research exploited the real-world Los Angeles traffic sensor and bus GPS datasets collected from Regional Integration of Intelligent Transportation Systems (RIITS) under ADMS project.

WHAT WAS THE OUTCOME?

The tangible outcome includes two types of assets. The application assets include 1) an interactive web-based application that enables visualization, querying, and analysis of performance metrics of buses in Los Angeles County based on real-time and historical GPS data, and 2) a suite of web-services that implements our statistical bus performance detection algorithms. A remote user can call these services to query data hosted on the servers. The second type of assets are the data mining algorithms to effectively analyze the performance of public transportation vehicles based on their GPS trajectories. The algorithms are summarized in the final report and will be disseminated in conference and journal publications.

Together the applications demonstrate how the algorithms developed from this project will help to increase the efficiency of public transportation systems. For example, the applications can be used by city transportation agencies to quickly identify problems with bus lines, such as delays possibly caused by driver behavior or bus technical problems or quantify the delays in bus lines caused by construction in the city. Even long-term policy decisions can be made to rearrange bus timetables. In addition, the web application will benefit riders to have a better understanding and access to travel-time delays and reliability.

Future directions include building on the current project outcomes to develop a comprehensive analytic framework on public transportation data. The current web and service applications, while useful, still provide limited types of analytics to the user. We envision a comprehensive analytic framework that will include 1) a web interface that the users can customize their queries to test scenarios and policy hypothesis as well as visualize the results and 2) the algorithms that will recommend possible improvements to the transportation system and employ user feedback in an active learning environment to advance the recommendation.

WHAT IS THE BENEFIT?

Building on a data management and trajectory analysis research will help develop a system that can process and analyze large GPS datasets. The input to the system is GPS datasets collected from buses in Los Angeles County and several other related datasets (e.g., time-tables, road network topology data, and traffic sensor data). In particular, the research focused on developing several novel components to i) clean and transform GPS datasets, ii) map-match and integrate the cleaned datasets to road network and time-tables, and iii) mining algorithms to compute performance and reliability metrics such as, travel-time reliability, bus bunching, and bus arrival-time estimation.