Large-scale and Long-term Forecasting of Performance Measurement of Public Transportation Systems

PI: Cyrus Shahabi <shahabi@usc.edu>
University of Southern California

Project Objective

The project objective is to build a machine learning approach and a system to achieve accurate estimations of public bus arrival times by forecasting long-term traffic flows for the entire Los Angeles Metropolitan Area (LAMA).

Problem Statement

Estimating bus arrival times requires 1) long-term and accurate traffic forecasting that applies to the scale of a city, 2) an end-to-end approach that effectively handles missing data and leverages the traffic forecasting to estimate bus arrival times, and 3) a system and an easy-to-use dashboard to communicate the results to various stakeholders.

Research Methodology

First, we explore both spatial statistical methods and machine learning methods to estimate traffic flows for the road segments that do not have traffic sensors. Second, we develop methods to enable traffic forecasting with a deep learning model designed for small networks for the entire LAMA road network. Our methods split the large road network into multiple sub-networks and train a deep learning model for each sub-network. We also study various training strategies (e.g., teacher forcing) and deep learning architectures to enable accurate long-term forecasting of traffic flows and bus arrival times. Lastly, we develop an end-to-end deep learning approach that combines the estimation and forecasting of traffic flows with data imputation methods for estimating bus arrival time for each stop in individual bus routes in LAMA.

Results

Using the real-world freeway and arterial traffic data and bus location data in the USC Archived Transportation Data Management System (ADMS), we show that the proposed approach and system are capable of predicting bus arrival times with a city-level spatial coverage and a route-level temporal forecasting horizon. Our distributed traffic forecasting model is capable of making city-level predictions on more than 10,000 road sensors. Figure 1 shows our traffic forecasting experimental results compared with a baseline approach called “Historical Average” (taking the average speed at the same time from historical data). The comparison shows that our approach significantly outperformed the baseline approach. Our approach had small prediction errors for traffic forecasting at 15, 30, 45, and 60 minutes into the future, which enables accurate estimates of bus arrival times.

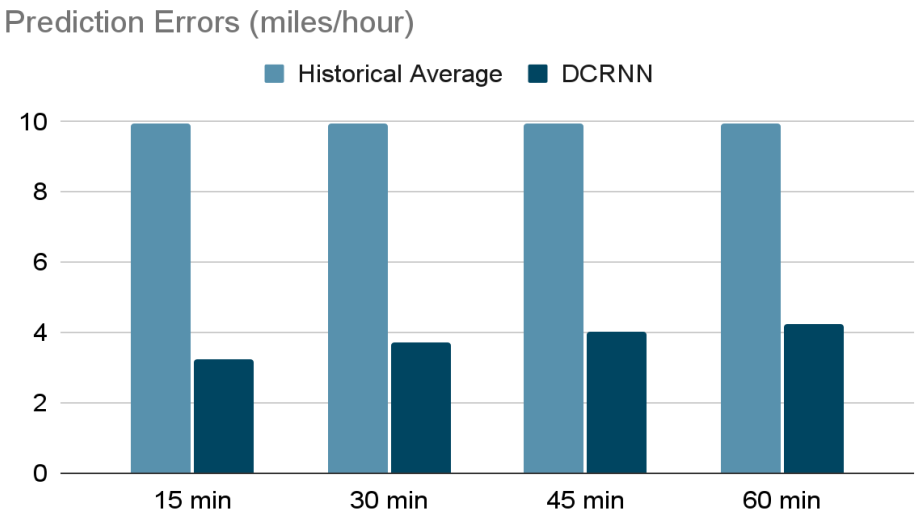

Figure 1. Traffic Forecasting Performance

We also demonstrate the overall result of the bus arrival time estimation in a web dashboard (Figure 2). Using the dashboard, a user can select a bus route, the starting and ending bus stops, and the time to leave to the bus arrival time estimation at each bus stop along the route. This dashboard enables users at all levels of technical skills to benefit from the developed machine learning approach and access to valuable information for trip planning, vehicle management, and policymaking.

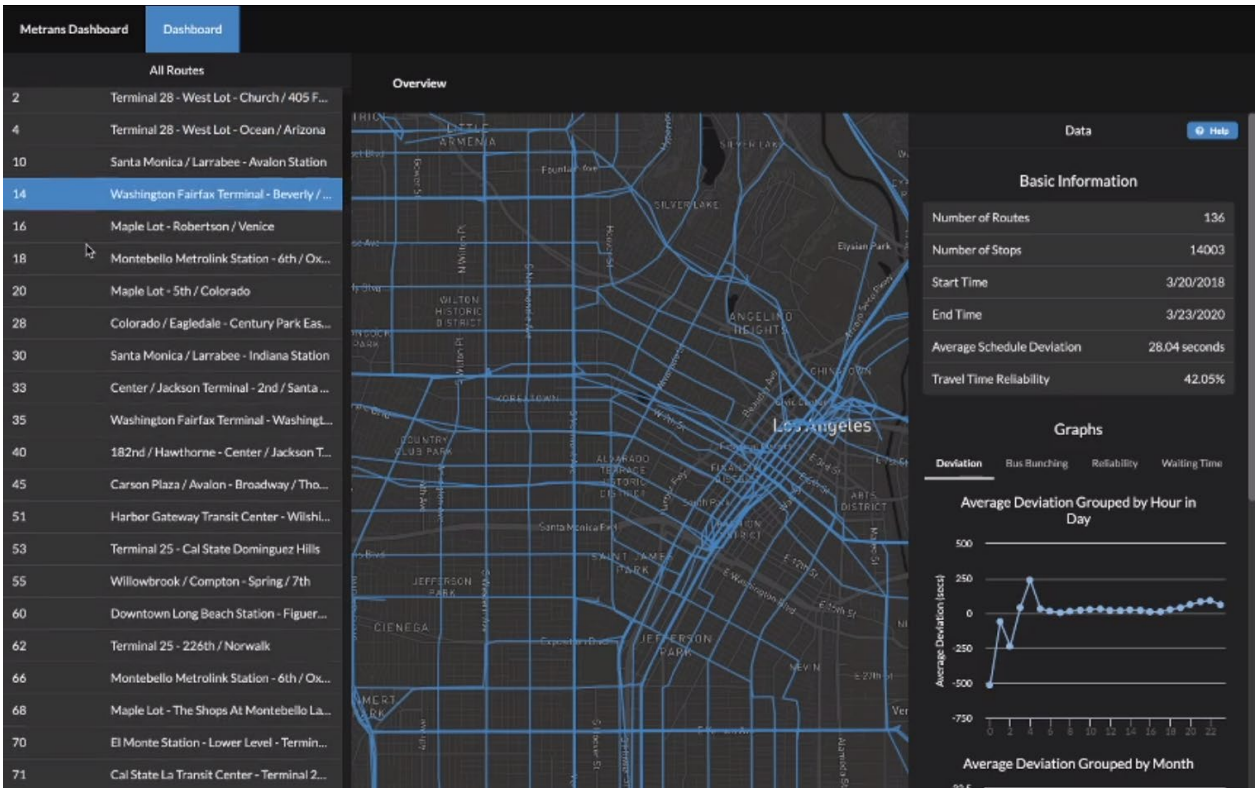

Figure 2. The web dashboard for bus arrival time estimation