Deep-Learning Traffic Flow Prediction for Forecasting Performance Measurement of Public Transportation Systems

Cyrus Shahabi
Yao Yi Chiang

University of Southern California
shahabi@usc.edu

Project Objective

Los Angeles is ranked the most congested city in the U.S. with a typical half-hour commute taking 81% longer during evening peak periods and 60% longer during the morning peak. These traffic congestions result in a large social and economic detriment and raise serious concern for drivers and transportation agencies. Therefore, increasing ridership of public transportations and hence reducing traffic congestions has been one of the primary objectives for transportation agencies and policymakers. The main objective of this project is to develop reliability analysis system using Deep Learning (DL) techniques that can process massive amounts of 1) GPS trajectories from public transportation vehicles and 2) real-world traffic sensor datasets archived in our data warehouse to predict traffic flow to then enable the forecasting of a variety of performance metrics of public transportation systems.

Problem Statement

Historical performance measurements of public transportation systems can help identify problems for improving ridership. For example, historical trends of bus travel-time reliability and on-time performance can help city transportation agencies to quickly identify potential problems with existing bus routes, such as quantifying the delays in bus lines caused by constructions in the city or making informed policy decisions including rearrange bus timetables. Beyond historical performance measurements, accurate predictive analysis of performance reliability helps to manage rider expectations (e.g., will the bus be on time in the next 30 minutes?) as well as provide a powerful tool for transportation agencies to manage the public transportation vehicles. However, predictive analysis on the performance reliability of public transportation vehicles is challenging because a major factor impacting the (near) future performance reliability of public transportation vehicles is traffic congestions in the future.

Research Methodology

We have a unique opportunity to use data-driven approaches to understand the factors causing traffic congestions and in turn, help forecasting the performance reliability of public transportation vehicles. In this project, we developed a reliability analysis system using Deep Learning (DL) techniques to forecast the future performances of the public bus system in Los Angeles. First, we developed a novel Graph Convolutional Recurrent Neural Network (GCRNN) to model and forecast traffic flows at different spatial (e.g., individual regions, road segments, or sensors) and the temporal (e.g., next 5 minutes and 30 minutes) resolutions. Our GCRNN model considers not only the location of traffic sensors but also their relationships (i.e., topological dependency) in space, which was critical to achieving the best performance for all forecasting horizons compared to the existing methods. Next, we implemented a Geo-Convolution Long Short-Term Memory (Geo-Conv LSTM) framework to model bus Estimated Time of Arrival (ETA) by incorporating the traffic flow predictions of our GCRNN (Figure 1). Lastly, we deployed both models as web applications so that users can access traffic prediction data and check bus arrival times to a destination location from a starting point (Figure 2).

Figure 1. Geo-Convolution Long Short Term Memory Network for Bus Travel Time Prediction

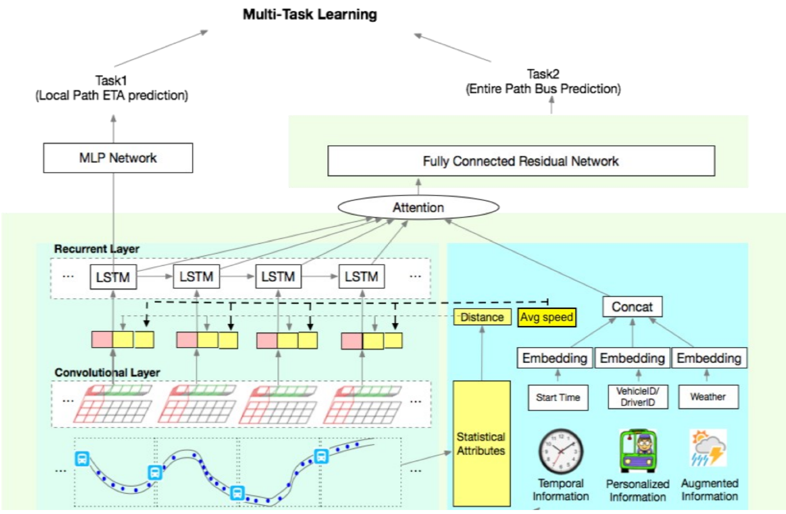
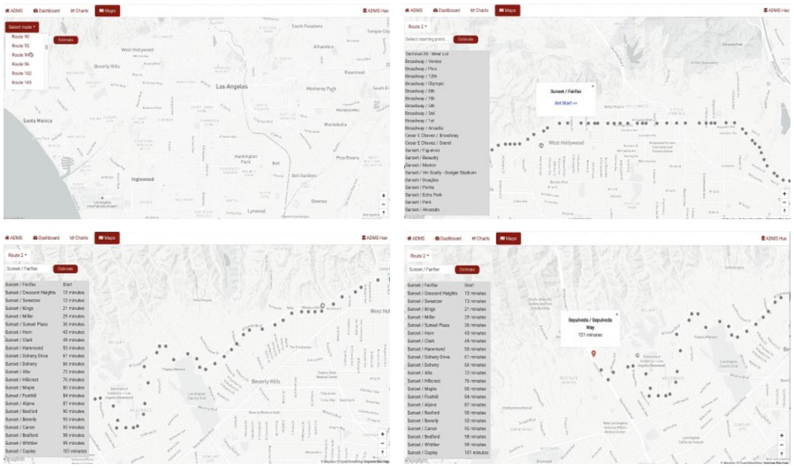


Figure 2. Prototype of Bus ETA Web Application Dashboard



Results

Using the real-world traffic sensor datasets archived in our data warehouse, we compared our bus ETA model with commonly used travel time estimation methods and showed that our proposed bus ETA model is more accurate than the existing method, Gradient Boosted Decision Tree (GBDT), by 27% in estimating bus travel times as shown in Table 1.

Table 1. MAPE of Bus ETA methods using LA dataset

Model	MAPE (%)
AVG	51
LR	79
SVR	39
GBDT	36
Geo Conv LSTM without traffic prediction	33
Geo Conv LSTM with traffic prediction	26