SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Assessing GTFS Accuracy

Gregory L. Newmark

Project 2017
August 2024

Introduction

Millions of transit users in the United States rely on the trip planning applications powered by the General Transit Feed Specification (GTFS) standard every single day. These applications allow users to make informed choices to best meet their travel needs. This standard has two core components, GTFS Static, which provides the scheduled route and stop information, and GTFS Realtime, which tracks the actual location of transit vehicles and offers predicted deviations from the static schedule. The emergence of the GTFS standard has tremendous potential for improving the public transit user experience, but only if the underlying GTFS information is accurate. This research presents several innovative techniques to assess the temporal accuracy of the vehicle arrival predictions in GTFS Realtime and the spatial accuracy of the transit paths and stops in GTFS Static. The metrics presented in this research can help transit agencies continually evaluate and improve the accuracy of the trip planning information they share with the public.

Study Methods

This research collected five full days of the GTFS Realtime (and the associated Static) data from five transit agencies across California. These data were carefully cleaned to provide a consistent baseline for metric development.

To assess the temporal accuracy, predicted and actual arrival values in the TripUpdate messages from the GTFS Realtime feed were combined with easily derived values capturing the time to prediction, prediction error, time to stop, and prediction change to generate a series of metrics. These metrics include the share of trip minutes for which an update is available, plots of prediction error percentiles, interquartile range of predictions scaled by the time to prediction, the likelihood of catching a bus given the prediction, the expected wait time (including to the next bus) given the prediction, the amount of padding necessary to not miss the desired bus, and the prediction inconsistency. The report presents these metrics as numbers and visualizes them into charts (at either the systemwide or route level) to demonstrate how they might be effectively employed to diagnose prediction accuracy.

To assess the spatial accuracy, timestamped geolocations (i.e., pings) in the VehiclePosition messages from the GTFS Realtime feed were combined with the scheduled path and stop locations in the GTFS Static feed. The straight-line distances between the pings and the paths and the pings and the stops were calculated. The distribution of these distances was examined to, along with professional judgment and local knowledge, determine an arbitrary threshold for flagging pings as 'exceedances' that represent inaccuracy. A series of metrics were proposed to identify the share and magnitude of these exceedances from the stop to the route level. These metrics were then used to create maps and charts that visualize spatial discrepancies between the scheduled paths and stops and where the transit vehicles actually go and board/alight passengers. Three conditions were presented to explain these discrepancies: the GTFS Static is inaccurate, the GTFS Realtime is inaccurate, or the transit vehicle is not following the scheduled path or stop locations. The report illustrates these conditions with examples drawn from the studied transit systems and provides guidance for transit agencies seeking to improve the spatial accuracy of their GTFS products.

Findings

The temporal analysis finds that GTFS Realtime TripUpdate messages contain surprising data errors, such as more than one prediction for the same stop with the same timestamp (timestamp conflicts) and predictions for upcoming stops that are for times prior to the timestamp (continuity errors). The shares of these errant values to the clean data represent a useful accuracy metric.

The study also finds that the availability of TripUpdate messages varies substantially by system and route. Many routes approach at least one update per minute while others provide updates for less than 80% of operational minutes hampering their riders' access to information. Similarly, the bus catch likelihood, the percentage of time that a user following the arrival predictions exactly will actually catch the desired bus, is less than three-quarters of the time for the studied systems—with some systems less than half of the time. This inaccuracy requires users who want to catch the bus 95% of the time to substantially pad the predicted arrival times. Finally, this research finds that predictions are rather inconsistent with the successive predicted arrival times for an upcoming stop often rising and falling to the annoyance of customers.

The spatial analysis finds that GTFS Static data is often poorly coded, either not on the road at all or on the wrong road. On the flip side, the GTFS Realtime data shows that drivers often deviate consistently from scheduled paths and stops – hinting perhaps at the need to revise the schedules accordingly.

Policy/Practice Recommendations

This research provides an array of techniques to assess the accuracy of the GTFS products shared with the public. Since the only data necessary to conduct these analyses are the standardized data provided in the GTFS feeds, it is very reasonable for transit agencies (or local governments or advocacy organizations) to set up monitoring programs to continually collect the GTFS data and generate accuracy metrics. Such a program would ideally encourage transit agencies to improve their operations and GTFS protocols to provide ever more accurate data to the traveling public.

About the Author

Dr. Gregory L. Newmark is an Associate Professor in the Department of Design and Planning in the School of Architecture and Planning at Morgan State University.

To Learn More

For more details about the study, download the full report at transweb.sjsu.edu/research/2017

MINETA TRANSPORTATION INSTITUTE

MTI is a University Transportation Center sponsored by the U.S. Department of Transportation's Office of the Assistant Secretary for Research and Technology and by Caltrans. The Institute is located within San José State University's Lucas Graduate School of Business.