ASSESSING THE ACCURACY OF BUS ARRIVAL TIME PREDICTIONS

David Czerwinski, College of Business, San Jose State University, One Washington Square, San Jose, CA 95192, 408-924-3528, david.czerwinski@sjsu.edu

ABSTRACT

Many public transit agencies now provide bus arrival time predictions for their customers. In this project, we assess the accuracy of the predictions provided by the Chicago Transit Authority (CTA). Overall, we found the predictions to be quite accurate but with some room for improvement. To our knowledge, this is the first study to use actual bus location data to assess a real-world bus arrival prediction system.

INTRODUCTION

With the ready availability of GPS tracking, many public transit agencies now provide bus arrival time predictions for their customers. In this project, we assess the accuracy of the predictions provided by the Chicago Transit Authority (CTA). To our knowledge, this is the first study to use actual bus location data to assess a real-world bus arrival prediction system. Below, we describe the existing bus arrival prediction methodology, describe the setting in which we collected data, detail the steps taken to collect, clean, and prepare the data, and report on our findings. We conclude with a discussion of how information about prediction uncertainty can be communicated to customers and an outline of future research directions.

Bus arrival time predictions rely on several pieces of technology on the bus itself and at a central processing center. The bus's location is tracked by an onboard GPS receiver, which is augmented with odometer readings for "dead-reckoning" when GPS signal reception is poor (such as in an "urban canyon" of tall buildings). The GPS and odometer readings are transmitted by radio from the bus to a central server where they are stored and processed. The server uses the location information, combined with information from the bus's schedule, to provide arrival time predictions at individual bus stops. Customers can access the predictions in a number of ways, including through a web browser, a smartphone app, or even by text message.

The CTA makes its bus arrival time predictions available on their Bus Tracker website (www.ctabustracker.com). Current locations of all buses on the route are also available through the site. The Bus Tracker predictions are made by calculating the bus's remaining distance to the bus stop and dividing by the scheduled speed of the bus. An inaccurate prediction will result if the bus's actual speed deviates from the scheduled speed. The possible causes of an inaccurate prediction are nearly limitless. Unexpected traffic congestion, slow boardings, a mechanical problem, inclement weather, etc. can all lead to an inaccurate prediction.

DATA COLLECTION

For our study, we selected the 8 Halsted bus route. The route operates 22 hours a day, running for approximately 14 miles between the South and North sides of Chicago, skirting the edge of the Central Business District (CBD) along the way. Because it doesn't go directly into the CBD, a reliable GPS signal is available along the entire route. Yet the route passes near enough to the CBD that buses

encounter a variety of traffic conditions, including heavily congested areas. On an average weekday, 24,000 passengers board the 8 Halsted. Headways vary from four minutes during the peak commute to 24 minutes off-peak.

We collected data at five bus stops spaced out along the route, in both directions. Data was collected in January and February of 2012. Bus locations and arrival time predictions were downloaded once every minute by an automated script accessing the CTA Bus Tracker's public API. During the collection period, 1,076,830 predictions were downloaded and 36,841 arrivals were recorded at the five stops. (The CTA Bus Tracker starts providing predictions when a bus is 30 minutes from a stop, and then continuously updates the prediction, so multiple predictions are made for each actual arrival.)

To assess the accuracy of the predictions, a bus's actual arrival time at a stop had to be derived from the bus location data and then matched up with the prediction. It was unlikely that the bus arrived at the exact moment that its location was recorded by our script. Therefore, linear interpolation was used to determine the arrival time, interpolating between the bus's last recorded location prior to reaching the stop and the first location recorded after the stop. In all, 926,688 (predicted, actual) arrival time pairs were collected and formed the basis for our analysis.

FINDINGS

Overall, we found the CTA Bus Tracker predictions to be quite accurate. The average absolute prediction error was approximately one minute. There was some bias in the predictions, with a bus more likely to arrive later than predicted, not earlier. Errors were more variable the further a bus was from the stop, as would be expected. The standard deviation of the errors ranged from 2.4 minutes when the bus was 10 minutes from the stop, to 4.2 minutes when the bus was 30 minutes from the stop. Errors varied somewhat based on the stop's location on the route. Somewhat surprisingly, time of day had little influence on the magnitude of the prediction errors.

The CTA Bus Tracker website could incorporate information about the level of uncertainty in its predictions in a number of ways, such as graphically, using a color coding, or numerically. Future research directions include used collected data to improve the accuracy of predictions, incorporating weather conditions into the analysis, and extending the analysis to other transit systems or other modes of transportation.