Transportation Research Part C 74 (2017) 81–96

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Bus arrival time calculation model based on smart card data

Yuyang Zhou *, Lin Yao, Yanyan Chen, Yi Gong, Jianhui Lai

Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China

ARTICLE INFO

Article history:
Received 13 November 2015
Received in revised form 26 October 2016
Accepted 12 November 2016
Available online 19 November 2016

Keywords:
Public transit
Bus arrival time
Smart card data
Ridership
Seating capacity

ABSTRACT

Bus arrival time is usually estimated using the boarding time of the first passenger at each station. However, boarding time data are not recorded in certain double-ticket smart card systems. As many passengers usually swipe the card much before their alighting, the first or the average alighting time cannot represent the actual bus arrival time, either. This lack of data creates difficulties in correcting bus arrival times. This paper focused on developing a model to calculate bus arrival time that combined the alighting swiping time from smart card data with the actual bus arrival time by the manual survey data. The model was built on the basis of the frequency distribution and the regression analysis. The swiping time distribution, the occupancy and the seating capacity were considered as the key factors in creating a method to calculate bus arrival times. With 1011 groups of smart card data and 360 corresponding records from a manual survey of bus arrival times, the research data were divided into two parts stochastically, a training set and a test set. The training set was used for the parameter determination, and the test set was used to verify the model's precision. Furthermore, the regularity of the time differences between the bus arrival times and the card swiping times was analyzed using the "trend line" of the last swiping time distribution. Results from the test set achieved mean and standard error rate deviations of 0.6% and 3.8%, respectively. The proposed model established in this study can improve bus arrival time calculations and potentially support state prediction and service level evaluations for bus operations.

1. Introduction

With increased urbanization, the rapid growth in the demand for travel has put enormous pressure on public transit systems. Complex traffic conditions increase the uncertainty of the times at which buses actually run. However, passengers need accurate bus arrival time information to decide their choice and the transit travel plans. Moreover, accurate timekeeping is also a key factor in evaluating the bus operation efficiency by both bus companies and public transit management departments. The bus arrival time calculation has become a basic problem in the intelligent public traffic system. The objectives of this study are to (1) develop a novel bus arrival time calculation model based on the characteristics of the passengers' unpunctually swiping behavior; (2) analyze several bus operation key factors affecting the correlation of the swiping behavior and the alighting for the parameter calibration; and (3) contribute to accuracy improvements in bus arrival time calculations based on smart card data (SCD).

Travel time is a critical factor when bus riders choose a route. The researches in this field focus on the waiting time, the inter-station bus travel time and the bus dwell time. The accurate bus arrival time information is of great importance to the

* Corresponding author.
E-mail address: zyy@bjut.edu.cn (Y. Zhou).

http://dx.doi.org/10.1016/j.trc.2016.11.014

passengers because it reduces waiting time and the anxiety at bus stations. Using real-time bus operation information, bus route design and scheduling can be optimized for public transit networks. Tirachini (2013) identified several important variables that could influence bus travel time and provided an assessment of speed that aimed at reducing bus travel time to improve bus operations. A real-time bus information system was proposed to evaluate the ridership effects and provide route optimization information in Chicago (Tang and Thakuriah, 2012). An empirical analysis of real-time bus ridership information had implications for improving public transit systems when faced with increased ridership. In addition, it helped minimize ridership waiting time and improve transit network reliability (Kusakabe and Asakura, 2014; Brakewood et al., 2015; Berrebi et al., 2015).

Studies that focus on the bus arrival time calculation, to a large degree, are based on bus route data collected using global positioning system (GPS). Based on historical inter-stop travel time and flex-schedules, the travel time determined by GPS was used to estimate bus arrival time and assess on-time reliability. The standard deviation and interquartile range were explored to analyze ridership route choices of public transit when travel time reliability improved (Hernandez, 2014; Carrion and Levinson, 2013). According to SCD and GPS data, the first card-swiping time in a cluster is considered as the bus arrival time to systematically monitor the average bus speed (Chen et al., 2012; Cortés et al., 2011). However, a result calculated by the BP artificial network showed that the percentile-based dynamic and median link time could estimate bus arrival time more precisely (Niu, 2010). To study ridership behavior at destinations, Cantis et al. (2016) proposed a group of indicators to analyze ridership mobility based on traditional survey instruments and GPS technologies. However, in many developing countries such as China, Brazil, and India, GPS is not widely used in public transit research, due to its cost, while SCD is commonly used due to its lower cost and more extensive application.

SCD has become increasingly popular in traffic research because the systems record large amounts of public transit data. Smart card data reflect real traffic situations better than manual survey data. Sun et al. (2014) explored passenger boarding and alighting dynamics to study their effects on bus dwell time. A model for estimating ridership activity time was established based on the critical occupancy from eight vehicle models using SCD. The results indicated that critical occupancy determined not only the regime of boarding and alighting activity but also the overall activity time. In South Korea, a distance-based fare system came into service in 2008, in which the boarding station, alighting station and transfer time were recorded. SCD is used for bus arrival time estimation (ATE) and travel time distribution analysis (Jin and Dong, 2008). To estimate the trip purpose and bus passenger arrival time, a data fusion method applied to regularize the data was proposed for the analysis of bus arrival times. The boarding location was determined from the GPS data, while the bus arrival time was computed from GPS and SCD (Chu and Chapleau, 2008; Shi and Lin, 2014). Long et al. (2012) proposed a method to visualize commuting patterns in Beijing, China. Bus arrival time was calculated to generate multi-modal trip routes using SCD matching. In the above approaches, bus arrival time was generally regarded as the card swiping time after the ridership had alighted.

Typically, passengers move to the bus exit door and swipe their smart cards when preparing to get off—often a couple of minutes before or just after the bus stops. Consequently, the card swiping time recorded in the SCD system is not necessarily precisely when the bus stopped. When using either the first swiping time or the average swiping time when alighting (which is always considered the bus arrival time), there is an unknown error between the time recorded in the SCD system and the actual bus arrival time.

Although the GPS coverage rate on buses was only 11.4% in Beijing, China's capital city, the automatic fare collection system was used on nearly all buses by the end of 2014. Previously, the major smart card system was the one-ticket fare collection system, in which the ridership swiped their smart cards only once when boarding but not when alighting. However, since December 28, 2014, a new smart card system called the double-ticket system has been implemented under the distance-based fare rule. In the double-ticket system, riders swipe their cards both when boarding and when alighting; however, the swiping time is recorded only for fare collection when alighting.

The SCD from the novel double-ticket system need to be deeply analyzed. Because the collected data are uploaded and stored to the double-ticket system every day, the record structure is minimal to save storage space in the system. The records include the station location and alighting time, and the boarding station is recorded in order to calculate the travel distance and, thus, the distance-based fare. Unfortunately, the boarding time, which is of no use for fare collection, is not recorded in the system. Therefore, it is still important and applicable to determine a novel method to calculate bus arrival time using SCD—especially, to correct the ATE based on the alighting swiping time as riders exit.

Approximately 85% of bus riders use smart cards to pay their bus fares, while the other 15% use cash. These data were collected by the China National Bureau of Statistics from the Beijing Transport Operation Coordinate Center (TOCC). Currently, there are approximately three million pairs (board/exit) of smart card data recorded for 1085 bus lines by different bus companies every day in the double-ticket system. Smart cards offer several advantages.

(1) Smart card payment is cheaper. There is no discount for cash payment. In the single-ticket system (before Dec. 28, 2014), a ticket was ¥ 1, while smart card and student smart card user fares were discounted by 60% and 80%, respectively. These large discounts are the reason that smart cards play a dominant role in bus fare payments. In the double-ticket system, the base fare is ¥2 for 10 km, with a surcharge of ¥1 for every additional 5 km. The discounts for the two types of smart cards are 50% and 75%, respectively. Moreover, there is an extra discount of 20% when fare expenditures exceed ¥100 and a 50% discount for those between ¥150 and ¥400. For example, suppose a rider commutes 22 days (44 trips) in a month; if the bus fare is ¥8 for one trip, a smart card user can obtain an extra 20% discount starting from