Transportation Research Part C 93 (2018) 410–424

Contents lists available at ScienceDirect

Transportation Research Part C

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

Development of destination choice model with pairwise district-level constants using taxi GPS data

Jiayu Zhu, Xin Ye*

Key Laboratory of Road and Traffic Engineering of Ministry of Education, College of Transportation Engineering, Tongji University, Shanghai, China

ARTICLE INFO

Keywords:
Destination choice model
Pairwise district-level constants
Iterative algorithm
Taxi GPS data
Trip distribution

ABSTRACT

In this paper, a destination choice model with pairwise district-level constants is proposed for trip distribution based on a nearly complete OD trip matrix in a region. It is found that the coefficients are weakly identified in a destination choice model with pairwise zone-level constants. Thus, a destination choice model with pairwise district-level constants is then proposed and an iterative algorithm is developed for model estimation. Herein, the "district" means a spatial aggregation of a number of zones. The proposed model is demonstrated through simulation experiments. Then, destination choice models with and without pairwise district-level constants are estimated based on GPS data of taxi passenger trips collected during morning peak hours within the Inner Ring Road of Shanghai, China. The datasets comprise 504,187 trip records and a sample of 10,000 taxi trips for model development. The zones used in the study are actually 961 residents' committees while the districts are 52 residential districts that are spatial aggregations and upper-level administrative units of residents' committees. It is found that the estimated value of time dramatically drops after the involvement of district-level constants, indicating that the traditional model tends to overestimate the value of time when ignoring pairwise associations between two zones in trip distribution. The proposed destination choice model can ensure its predicted trip OD matrix to match the observed one at district level. Thus, the proposed model has potential to be widely applied for trip distribution under the situation where a complete OD trip matrix can be observed.

1. Introduction

With the rapid development and application of computer science, communication technology, wireless sensor network and Global Positioning System (GPS), etc., current advances in the internet, mobile internet, Internet of Things (IoT) and Internet of Vehicles, etc. are leading to an exponential growth in the amount of available mobility big data produced continuously at high speed (Galić et al., 2016). Big data are characterized by several typical features: huge volume, various data types, fast processing speed, automated collection methods, wide scope of coverage, etc. (Yuan et al., 2016). Currently, transportation big data have been widely used including mobile phone data (Chen et al., 2016; Fang et al., 2016; Bantis and Haworth, 2017), floating car data (Knapen et al., 2016; Tu et al., 2016; Li et al., 2016; Günther et al., 2017; Knapen et al., 2016), transit smart card data (Hong et al., 2017; Ma et al., 2017), social media data (Rashidi et al., 2017; Gkiotsalitis et al., 2016), pass-recording data (Yuan et al., 2016), loop detector and remote sensor data (Tang et al., 2017; Zou et al., 2017), etc. Big data record almost all the activities of systems, which provides a rich source of data and leads to possibilities of system modeling, predicting and optimizing (Jiang et al., 2017). In the era of big data, the

* Corresponding author at: Key Laboratory of Road and Traffic Engineering of Ministry of Education, College of Transportation Engineering, Tongji University, Tongda Building, Room 549, 4800 Cao'an Road, Shanghai 201804, China.

E-mail addresses: zhujiayu@tongji.edu.cn (J. Zhu), xye@tongji.edu.cn (X. Ye).

https://doi.org/10.1016/j.trc.2018.06.016
Received 18 December 2017; Received in revised form 28 May 2018; Accepted 26 June 2018
0968-090X/

acquisition of complete OD (Origin-Destination) trip matrices of different modes may provide better opportunities to refine transportation models.

Travel demand models are the commonly-used approach to predict traffic volumes, transit ridership, walking and biking market shares across transportation networks (Zarwi et al., 2017). Conventional travel demand models include four major steps: trip generation, trip distribution, mode split and network assignment. The second step, trip distribution, is a critical step, through which zone-to-zone trip matrix is initially generated (Ye et al., 2012). Trip distribution models can be classified into aggregate and disaggregate models. Aggregate models analyze the total amount of traffic flow between the TAZs (Traffic Analysis Zones) while disaggregate models attempt to explain individuals' behaviors in selecting destinations of their spatial movements (Celik et al., 2010). Aggregate models may be further classified into growth-factor methods, gravity models (entropy models) and intervening opportunities models while random utility models or destination choice models (e.g. Logit, Nested Logit, Cross-nested logit, Mixed Logit Models, etc.) are considered as disaggregate models (Cascetta et al., 2007; Outwater et al., 2014; Lamondiaet al., 2010).

The gravity model is analogous to the Newton's theory of gravity, which assumes that the trips produced at an origin and attracted to a destination are directly proportional to the total trip productions at the origin and the total trip attractions at the destination and inversely proportional to travel distance/time (or separation) between the origin and destination. With the well-known theoretical basis, the gravity-type spatial interaction model has been the most commonly used trip distribution model (Celik et al., 2013; Thomas and Tutert, 2013; Abdel-Aal, 2014; Lenormand et al., 2016). However, the gravity model has also been criticized in literature. It assumes that all the information lies in the constraints from trip production, attraction, length distribution and it uses an aggregate calibration procedure (Cascetta et al., 2007). Prior to the application of a gravity model, a regression-based trip attraction model needs to be estimated but it does not provide reasonable coefficients and therefore cannot correctly predict trip attractions. The biased trip attraction will result in considerable errors in a predicted trip matrix (Ye et al., 2012).

Wilson (1967) showed that the gravity model and logit-based destination choice model have the same mathematical form as an entropy maximization model used in statistical mechanics. Some distribution models use k-factors or socioeconomic factors to modify the results of the gravity model to more closely match real trip characteristics (Martion and Mcguckin, 1998). The k-factors in the gravity model and the constants in the destination choice model have been shown to be of essentially the same form (Zhang, 2016).

As disaggregate models, a destination choice model can involve multiple variables and simultaneously take account of zone attraction and accessibility, which does not only improve the accuracy of the forecast but also addresses problems in traditional aggregate models. It can involve policy-sensitive variables in a flexible way so as to predict and evaluate policy impacts. Thus, the destination choice model, which is more flexible than the gravity model, becomes increasingly popular (Clifton et al., 2016; Ding et al., 2014; Yang et al., 2010, 2009; Bekhor and Prashker, 2008; Bhat and Guo, 2004; Wang et al., 2016). Research in destination choice model development with big data has already led to some pioneering work in this area (e.g. Huang and Levinson, 2015). In addition, Zhu and Ye (2017) pointed out that passively collected big data can avoid possible biases in traditional travel surveys limited by the sampling process and potential discrepancies between respondents' actual behaviors and their responses. The taxi GPS data used in this study is essentially a large sample of RP (Revealed Preference) data with trip origin/destination locations and departure/arrival times.

Motivated by the need to better predict how people travel in space, the main objective of this paper is to improve the destination choice model with pairwise constants for trip distribution. As is known to all, when a destination choice model is compared with a mode choice model, one of major differences is in the choice set with a large number of alternative destinations. In the four-step model, each of all the TAZs can be considered as a potential destination, which typically brings thousands of alternatives into the choice set (Pozsgay and Bhat, 2001). Since the small sample size in a traditional survey covering a very limited number of trip records between zones prevents researchers from giving a specific utility function to each zone pair, a traditional destination choice model can only have a uniform utility function for all the zone pairs. If pairwise zone-level constants are specified into utility functions for specific OD pairs, they will not be identified. Thus, there is no constant term in the utility function of a traditional destination choice model. Why do we think of adding constants into utility functions to improve a destination choice model? The constant in the utility function of a logit model is the expectation of the overall effect from excluded influential factors, which can play an important role in adjusting inconsistencies between the trip matrix predicted from a model and the observed counterpart. Since there is a potential to observe a complete OD trip matrix in the era of big data, it is a good timing to consider a destination choice model with pairwise constants to further improve the reliability of a trip distribution model. The added constant terms can capture differences between different OD pairs, which is reflected by the trip distribution between each pair of zones. This kind of new modeling approaches can only be developed with emerging technologies to passively collect travel big data that almost cover the entire population in a region. Otherwise, those constants would not be identifiable.

The rest of this paper is organized as follows. The second section proposes the modeling approach. In the third section, simulation experiments are conducted to validate the model derivation and estimation method. The fourth section presents a case study and shows comparisons between destination choice models with and without pairwise district-level constants. Finally, the last section provides a summary and conclusions.

2. Modeling approaches

As an important part of travel demand forecast system, a destination choice model is usually developed at individual level and therefore considered as a disaggregate model.