Transportation Research Part C 69 (2016) 291–312

Contents lists available at ScienceDirect

Transportation Research Part C

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

Updating origin–destination matrices with aggregated data of GPS traces

Qian Ge *, Daisuke Fukuda

Department of Civil Engineering, Tokyo Institute of Technology, 2-12-1 O-okayama, Meguro-ku 152-8552, Tokyo, Japan

ARTICLE INFO

Article history:
Received 17 August 2015
Received in revised form 30 May 2016
Accepted 2 June 2016
Available online 18 June 2016

Keywords:
Coarse-grained travel data
Mobile phone data
Travel survey
Origin–destination trip matrix estimation
Maximum entropy

ABSTRACT

The practice of estimating origin–destination (OD) demand usually requires large-scale travel surveys. To reduce the cost and time spent on surveys, individual trajectory data obtained from mobile devices has been used as an alternative dataset since the last two decades for OD estimation but also constrained in practice in some countries. To estimate OD matrices while protecting privacy, this study uses aggregated data of mobile phone traces to estimate work-related trips. The proposed approach is a sequential updater based on the maximum entropy principle. Trip production and attraction are firstly calculated by a non-linear programming problem followed by a matrix fitting problem to distribute trips to each OD pair. Numerical study shows that updated values are much closer to the synthesize real values than the referred ones. The case study in Tokyo further demonstrates that the proposed updating approach can track the change of travel pattern.

1. Introduction

1.1. Background

Travel demand, represented in the form of origin–destination (OD) matrices, is indispensable while developing traffic management and control schemes. In the past several decades, new methodologies for estimating OD matrices have been developed along with the availability of travel data. State-of-the-practice in travel demand estimation usually derives OD matrices through data obtained from large-scale travel surveys, while later development in this field has used a variety of new data sources to reduce the cost of data collection and make timely estimation.

Since the 1950s, the four-step model has become mainstream for transportation planning (e.g. Weiner, 1992). OD matrices are generated from trip generation and distribution models using surveys on travel behavior. Later on, the relationship between route choice and OD travel demand was observed by researchers, and OD estimation became an inverse problem of the assignment matrix and traffic counts on transport links. Optimization and statistical models were formulated with static/dynamic route assignment assumptions. Though the objective function is to minimize the difference between the estimated and target OD matrices as well as the assigned link flows and observations, the underlying assumptions on route assignment and the divergence between estimated values and observations can be different in models. These problems lead to discrepancies between estimation results.

* Corresponding author.
E-mail addresses: ge.qian@plan.cv.titech.ac.jp (Q. Ge), fukuda@plan.cv.titech.ac.jp (D. Fukuda).

http://dx.doi.org/10.1016/j.trc.2016.06.002
0968-090X/© 2016 Published by Elsevier Ltd.

In practice, household travel surveys are conducted in many countries. For example, the Person Trip survey is conducted every 10 years in large metropolitan areas of Japan. Usually, the sampling rate of travel survey tends to be quite low. Thus, sampling error cannot be neglected (Santos et al., 2011). Besides, there also exist non-sampling errors that bring considerable mistake in records (Groves, 2006). People tend to filter certain types of trips for reasons such as forgetfulness, soft non-response and rounding up time in surveys (Axhausen, 2014). Meanwhile, the dataset soon becomes outdated because a considerable amount of time is necessary for coding and pre-processing data. The interval between two consecutive surveys is extremely long, thus it is quite difficult to derive reliable estimates of travel demand during intermediate years.

The pervasive mobile device has become an alternative for travel surveys, especially for the past two decades. It can satisfy the requirement of detailed personal location data to know individual's travel behavior. There exist quite a few papers on developing tracking method (Asakura and Hato, 2004), field test (Murakami and Wagner, 1999) and using GPS data for helping travel survey (Wargelin et al., 2012). Data collection schemes have been meticulously designed for these datasets. Precision and spatial–temporal coverage of these *fine-grained* datasets are also very desirable but with considerable cost. GPS data can collect multiple day trips, handle short trips which is always missing in self-reported surveys, has good data quality on travel start and end times, total trip times and destination locations. But is still not perfect due to satellite signal loss and the error induced by GPS system especially when the frequency of recording is low (Murakami and Wagner, 1999). Furthermore, individual location data has also aroused public concern about the invasion of privacy. Therefore, data in coarse granularity, or *coarse-grained* datasets become promising alternative data sources for OD estimation.

In summary, although the availability of data on travel demand has improved along with technological development in the past several decades, cost and privacy issues impose progressively more severe constraints. There is still a great gap between practice and art. In addition, individual location data are not available in most countries because of privacy and cost issues. This study is dedicated to developing a calibration method for OD matrices using low cost mobile phone datasets that do not invade privacy, i.e., aggregated mobile phone data.

1.2. Literature review

The problem of OD travel demand estimation has been studied extensively in the past several decades. Considerable number of studies in this line focus on the inverse problem of traffic assignment and OD demand on using link counts. These models usually estimate OD matrices by two sub-models: determining OD trip flows (or path flow) by the upper level model, and determining the most probable route choice of travelers given the traffic counts by the lower level model. These two problems are solved iteratively (e.g. Yang et al., 1992) or relaxed to be a decoupled form (e.g. Nie et al., 2005). Studies on OD estimation problem have developed from estimating OD trips in a static scenario to a dynamics scenario within a day, and then further to incorporating different data sources to obtain filtered calibration of the OD matrices (e.g. Zhou and Mahmassani, 2007). Effective methods are also developed to reduce computation load (Cipriani et al., 2011).

For the master problem of OD estimation, a variety of assumptions are made to measure the divergence between the estimated OD demand and the reference one. Maximum entropy (information minimization) principle is firstly used to distribute OD demand by the trip attraction and trip production (Wilson, 1971). Later, constraints on the traffic counts and assignment map are added to this modeling framework (e.g. Van Zuylen and Willumsen, 1980) enabling it to deal with OD estimation in a more modern structure of mathematical programming with equilibrium constraint (MPEC) (Yang et al., 1992). Recently, the maximum entropy method is applied to subnetworks to estimate elastic travel demand (Xie et al., 2011). Constrained Generalized Least Square (CGLS) are also common assumptions in estimating OD demand (Hendrickson and McNeil, 1984; Bell, 1991). The objective function of CGLS is flexible and thus can be used adapted to new datasets, for instance, Zhou and Mahmassani estimate OD demand with automatic vehicle identification data (AVI) by this structure. The CGLS can be further relaxed to one-level estimators (Nie et al., 2005). Bayesian statistical model is a nature choice when studying the OD estimation problem from statistical perspective. One may treat the link count (Maher, 1983), link choice (Lo et al., 1996), routing choice (Parry and Hazelton, 2012) or other observations of traffic flow such as plating scanning as random variables (Castillo et al., 2013) to form a statistical model for OD estimation.

The diversity of underlying assumptions in these models will also admit different estimations for one problem. Advances in data collection methods, especially ubiquitous data have helped to relieve the problems. Automated vehicle identification (Zhou and Mahmassani, 2006), probe vehicle (Cao et al., 2013), billing data (White and Wells, 2002) and GPS trajectory data (e.g. Pan et al., 2006) are used for estimating OD demand through either simulation or field tests to name a few. Despite the relative accuracy and soundness of these fine-grained data sets, the installation and maintenance cost of GPS devices and license plate readers can be expensive. These are the driving motivation for using passive data.

Coarse-grained data are datasets that have coarse granularity. These data sets are usually collected over long intervals and have low precision for positioning. One source of coarse-grained data is passive data, i.e., datasets from surveys or records that are not intentionally designed for travel behavior studies. A pioneer study of coarse-grained mobile phone data is by Gonzalez et al. (2008), who investigated the mobility pattern of people using call dial record (CDR) data. In Gonzalez et al.'s study, the dataset records each person's location when he/she sends a message, makes a phone call or connects to the Internet, all activities that use a local cell tower. Similar dataset has also been used to estimate OD matrices based on the fact that mobile phone moving along a route always tends to change the base stations near one's position (Caceres et al., 2007). Using mobile phone data, several rules of thumb are established to infer each individual's location from one's