Developing Trip Generation Models Using Mobile Phone Data

Charisma Choudhury, Institute for Transport Studies, University of Leeds, UK
Md. Shahadat Iqbal, Department of Civil Engineering, Florida International University, USA

Background

Estimating reliable trip generation rates is an important component of the transportation planning process. Traditional approaches of estimating trip generation rates rely on household surveys which involve expensive data collection and thereby have limited sample sizes and lower update frequencies. Moreover, they are often prone to sampling biases and reporting errors.

Mobile phone users on the other hand leave digital footprints of their approximate locations whenever they make a call or send an SMS. Over the last decade, mobile phone penetration rates have increased manifold both in developed and developing countries: the current penetration rates being 128% and 89% in developed and developing countries respectively (International Telecommunication Union, 2013). Subsequently, mobile phone data has emerged as a very promising source of data for transportation researchers. In recent years, mobile phone data have been used for human travel pattern visualization, mobility pattern extraction, route choice modelling, traffic model calibration, traffic flow estimation and origin-destination flow estimation, to name a few. Mobile phone data however has penetration, location and usage bias (i.e. not everyone has a mobile phone and the rate of using it is not uniform in all locations and by all users) and the spatial resolution of the data is often coarse. Moreover, due to privacy concerns, very often these data are anonymized and socio-demographic information of the users are not available. These limitations make it unsuitable to use it directly for developing trip generation models.

In this research, we propose a trip generation modelling framework combining Mobile phone Call Detail Records (CDR), Traffic counts and Census data. The framework is demonstrated using data from Dhaka, Bangladesh and the developed model is compared with a trip generation model developed using household survey data.

Methodology

CDR consists of time stamped tower locations with anonymized caller IDs and reveals an approximate location of the user whenever he makes a phone call or sends a text message. The data therefore provides useful information about the trip patterns of the users. These trips patterns are analyzed first and trips occurring within certain time windows are used to generate tower-to-tower OD matrices for different time periods. These are then associated with corresponding nodes of the traffic network and converted to node-to-node OD matrices. The actual OD matrices are derived by scaling up these node-to-node OD matrices. An optimization based approach, in conjunction with a microscopic traffic simulation platform MITSIMLab, is used to determine the scaling factors that result best matches with the observed traffic counts:

$$OD_{ij} = \sum_{ij} \left(m - OD_{ij}\right) * \beta_{ij}$$

Where, $OD_{ij}\ and\ m - OD_{ij}$ are actual and mobile phone ODs between zones i and j respectively and β_{ij} are the corresponding scaling factors.

The total trip generation from each zone (T_i) is determined by summing up the ODs originating from that zone.

$$T_i = \sum_j OD_{ij}$$

The census data is used to generate a synthetic population which in conjunction with the trip generation rates derived from CDR and traffic count data, is used for estimating discrete choice models for trip generation.

Data

The data used for the research includes the following:

- CDR data which consists of calls from 6.9 million users (which are more than 65% of the population of the study area) over a month. This comprises of 971.33 million anonymized call records in total made in between June 19, 2012 and July 18, 2012. It may be noted that no demographic data related to the phone users are available.
- Video data collected from 13 key locations of Dhaka city network over 3 days (12th, 15th and 17th July 2012) have been used in this study to extract the traffic counts . The data has been collected for 8 hrs (8.00 am to 12.00 noon and 3.00 pm to 7.00pm) and analyzed using the software TRAZER (e.g.48,49) to generate classified vehicle counts.
- Socio-economic data collected as part of the Household Income and Expenditure Survey (BBS 2012) and National Census (BBS 2010). It may be noted that this data does not include travel information.
- Household travel diary data collected as part of the Dhaka Urban Transport Study (DHUTS 2010) which includes individual level socio-demographic information along with trip rates for 10,000 households (less than 0.09% of the population). Moreover it is acknowledged that the trip rates gathered from the survey data is likely to have reporting errors.

Results and Application

Preliminary results indicate similarities in trip patterns generated by the two approaches though there is marked scale difference between the two. The reporting errors in the household surveys may be a potential cause of the discrepancy and needs further investigation. However, the developed approach holds significant promise of validating and/or supplementing the models generated using household survey data. It can also be a very economic alternative for updating trip generation models in developing countries where data and resources are limited.