DYNAMIC TRIP ATTRACTION ESTIMATION WITH LOCATION BASED SOCIAL NETWORK DATA BALANCING BETWEEN TIME OF DAY VARIATIONS AND ZONAL DIFFERENCES

Nicholas W. Hu [a], Peter J. Jin [b, *]

[a] Department of Civil and Environmental Engineering, Rutgers, the State University of New Jersey, CoRE 736, 96 Frelinghuysen Road, Piscataway, NJ 08854-8018 – nicholas.hu@rutgers.edu
[b] Department of Civil and Environmental Engineering, Rutgers, the State University of New Jersey, CoRE 613, 96 Frelinghuysen Road, Piscataway, NJ 08854-8018 - peter.j.jin@rutgers.edu

KEY WORDS: Dynamic Trip Attraction, Time of Day, Location Based Social Networking, Big Data

ABSTRACT

The emergence of location based social network (LBSN) services make it accessible and affordable to study individuals' mobility patterns in a fine-grained level. Via mobile devices, LBSN enables the availability of large-scale location-sensitive data with spatial and temporal context dimensions, which is capable of the potential to provide traffic patterns with significantly higher spatial and temporal resolution at a much lower cost than can be achieved by traditional methods. In this paper, the Foursquare LBSN data was applied to analyze the trip attraction for the urban area in Austin, Texas, USA. We explore one time-dependent function to validate the LBSN's data with the origin-destination matrix regarded as the ground truth data. The objective of this paper is to investigate one new validation method for trip distribution. The results illustrate the promising potential of studying the dynamic trip attraction estimation with LBSN data for urban trip pattern analysis and monitoring.

* Corresponding author

doi:10.5194/isprsannals-II-4-W2-193-2015

1 INTRODUCTION

Trip attraction estimates the number of incoming trips to a destination or zone based on land use and social-economic data. It is an essential part of the trip generation step in the classic four-step planning process (Meyer and Miller, 2001) and has been well-documented in standard planning manuals such as ITE trip generation manual (ITE, 2012). In activity-based travel demand models, analyzing the trip attracted to a traffic analysis zone or finer spatial units such parcel and census block levels also provides valuable input regarding destinations to assist the micro simulation on individual travel (Castiglione et al., 2014). One key limitation for the convention trip attraction models is their dependency on static land use and census data which limit their usability in dynamic travel demand estimation and prediction needed for the emerging Active Traffic and Demand Management (ATDM) solutions.

In recent years, the thriving development of wireless communication, social media technologies, positioning and computing technologies has presented new opportunities for transportation planners and engineers to develop effective solutions to collect travel demand data with high spatial and temporal resolution for dynamic travel demand analysis.

Current research has found four major limitations in emerging travel demand data collection technologies including sampling bias, privacy concern, positioning accuracy and the lack of trip purpose confirmation. The first problem concerns some emerging travel data survey methods require users to actively participate or use the application for the survey purposes. For example, in GPS-based survey, sampling bias varies based on income, education level, age and gender. Low-income, lower educated, and minority populations bring low participation rates in travel demand collection (Bricka et al., 2009). Secondly, especially in large-scale traffic data collection, legal and political issues about the privacy concern can make it difficult for transportation agencies, since most new technologies will request information to trace the identity of travelers. The third key limitation is positioning accuracy of survey data collection. The quality of data may be yielded due to the reliability of GPS signals, the distortion or block of signals in large scale urban area, or the requirement of high level of technical expertise (Wolf et al., 1999). The last key problem concerns is that the trip purpose confirmation might not be collected through emerging travel demand data collection technologies. Unlike the conventional household interview survey which records characteristics of personal trip ends, new primary and secondary data collection methods cannot passively identify the location confirmation of destination and trip purpose. In order to measure them, many technologies such as GPS, Bluetooth, and cellphone-based methods need to ask travelers to enter their trip information which creates concern of reducing the sample size, or use data mining models to determine those from repeated route and activity patterns which will have significantly repeating patterns such as commuting trips. Table 1 summarizes the characteristics of the traditional and emerging travel demand data source and related methods for trip attraction estimation.

Key Limitation	**GPS**	**Blue tooth**	**Smart Phone**	**Cell Phone**	**Social Media**
Sampling bias	M	Y	M	M	Y
Privacy concern	Mediu m	No	No/M	No	N
Positioning accuracy	Low	Low	High	High	High
Trip purpose confirmation	M	M	M	M	Y

* Characteristics are based on NCHRP report 735 (Schiffer 2012)

Table 1. Comparison among new and emerging travel demand data collection technologies

This paper is to investigate the model to validate trip attraction with the social networking data and to capture the temporal and spatial characteristic of zonal trip attraction pattern. Compared to the typical methods of static trip attraction estimation, we explore one time-dependent function to make dynamic trip attraction using the LBSNs data.

The rest of paper will be organized as follows. Section 2 provides the literature review of research topic and method. The methodology and procedure will be introduced in Section 3. Next, Section 4 introduces details on the experimental design as well as results from the proposed algorithm. Finally, Section 5 concludes the paper and provides some areas for the continuation of this research effort.

2 LITERATURE REVIEW

2.1 Location Based Social Network (LBSN) Data

As one of above potential new data sources, the emergence of location based social network (LBSN) services make it accessible and affordable to study individuals' mobility patterns in a fine-grained level and to estimate trip attraction for each venue in the future. Location based social network refers to special social networking services that use LBS (location-based service (Quercia et al., 2010), replying on GPS to locate users, and allow members of the communities to broadcast their locations and activities through their mobile devices. LBSN does not only mean adding a location to an existing social network so that people in the social structure can share location-embedded information, but also consists of the new social structure made up of individual connected by the interdependence derived from their location in the physical world as well as their location-tagged media content (Zheng, 2011). As a source of input data, it recorded in "check-in" or tweeting activities of massive users at different points of interests (POIs) named "venue" such as ground transportation center, popular restaurant, bar, club, sports stadium, and even a bus. In LBSN services such as Foursquare and Twitter has enabled users to share their location or the venues they have visited in the past with their social communities. A user will make one check-in or sending one tweet by using a smartphone or tablet to choose nearby venue which will be recorded in LBSN server with the geographical location (i.e., latitude and longitude coordinates). This novel market has attracted much attention from worldwide companies and the estimated number of LBSN websites has reached more than a hundred by 2011 (Schapsis, 2011). Foursquare is the leading LBSN provider in the US, attracting 10 million registered users by June 2011, with about 3 million check-ins per day (Tsotsis, 2011).

With respect to emerging travel demand data collection technologies, location based social network data shows unique advantages over the GPS, cell phone, and Bluetooth data in term of the four major limitation in data collection. The influence of social impact from users' LBSN group or monetary discount such as a coupon for the visiting venue provide the great incentives for the data collection work. Additionally, with the rapid development of smartphone, the LBSN application can be easily built in personal mobile and tablet without concerning the maintenance and updates issue in the traditional traffic monitor infrastructure. The sample size can be much larger than other methods due to the penetration rate of social networking service growing at a rapid pace. When it comes to the major concerns of travel demand data collection technologies, for the privacy concern of LBSN data, the user-side data contain the detailed demographic information of each user and their detailed check-in log at every venue is only available after signing the data-sharing agreements between the user and Foursqaure application provides. Thirdly, to take full advantage of the social networking service provider such as Foursquare, the position accuracy is based on the venue-side data which the type and location of a venue is pre-defined and well-maintained. Such high spatial and temporal resolution enabled researchers to perform fine-grained analysis of users' mobility patterns and their impact on social interactions. For the last key consideration of the trip purpose measurement, through the "venue" type it is clearly to identify the location of trip activity such as food, entertainment, work, shop and other trip purpose. Moreover, compared to the tradition alternative source, analysis of LBSN data gives the researchers the latent characteristic of human activity and mobility patterns, such as trip repetition and temporal clustering.

2.2 Trip Attraction Estimation

Traditionally, in first part of the four-step model, in trip generation, measures of trip frequency are developed providing the propensity to travel. Trips are represented as trip ends, productions and attractions, which are estimated separately. Trips can be modeled at the zonal, household, or person level, with household level models most common for trip productions and zonal level models most common for trip attractions (McNally, 2008). One limitation of the traditional four-step scheme is the absence of temporal scale that trips are not specified in any time reference. Furthermore, as noted by Boyce (Boyce, 1998), the feedback loop between the upper and lower levels is often neglected or at best implemented using ad hoc rules since the dynamics of the decision update is not obvious. Activity-based demand models are a promising replacement solution (Damm, 1983). Currently, the most common trip attraction method in practice is the ITE trip generation procedures (ITE, 2012). For the parameter optimization during the model simulation, a genetic algorithm was implemented. The genetic algorithm (GA) is a search heuristic that mimics the process of natural selection. Such heuristic is routinely used to generate useful solutions to optimization problems. The search strategy was based on the selection of the improved chances of finding a global solution, using the concept that "individuals" are randomly selected from the current "population" as "ancestor" of the "offspring" for the next generation. Figure 1 describes the modeling flowchart of genetic algorithm as follows.

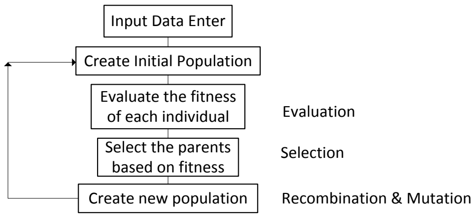

Input Data Enter
Create Initial Population
Evaluate the fitness of each individual
Evaluation
Select the parents based on fitness
Selection
Create new population
Recombination & Mutation

Figure 2. Genetic algorithm optimization procedure

3 METHODOLOGY

The proposed model used for trip attraction estimation is as the following:

$$\hat{X}_{i,n,t} = \sum_n \sigma_{n,t} Y_{i,n}, i = 1,2,\dots,730; t = 1,2,\dots,24 \quad (1)$$

Where
$\hat{X}_{i,n}$: Trip attraction for venue type n in zone i
$\mathrm{Y}_{\mathrm{i,n}}$: Check-ins statistics for venue type n in zone i
$\sigma_{n,t}$: The ratio of trip attraction to Foursquare check-ins for venue type n with respect to t which falls in 24 TOD regimes

Here the trip attraction estimation by various venue types n are in consonance with the different trip purposes contained by the CAMPO matrices of the daily trip tables. Similar to the definition of passenger car equivalent, in order to explore the essential impact that one type of venue has on check-ins occurrence variables (such as land use, venue location and the demographic characteristics of Foursquare users) compared to one specific venue type, we explore the different trip arrival pattern for different venue types through the proposed model parameters. For this study, zonal trip attraction are aggregated in the condition of general trip purposes which relates to all trip purposes.
The analysis needs to solve the following equations for balancing the social media activities and trip arrivals between time of day variations and zonal differences:

$$Min. \sum_i \left| \left(\sum_t \hat{X}_{i,n,t} \right) - X_{i,n} \right| + \sum_t \left| \left(\sum_i \hat{X}_{i,n,t} \right) - X_{n,t} \right| \quad (2)$$

Where
$X_{i,n}$: Daily ground truth trip arrival for venue type n in zone i
$X_{n,t}$: Time of day ground truth trip arrival for venue type n in zone i

Using the dynamic LBSN statistics, the parameters σ_{nt} were optimized through the genetic optimization algorithm.

$$P(\sigma_{n,t}, g+1) = f(\sigma_{n,t}, g)[1-e]P(\sigma_{n,t}, g) \quad (3)$$

Where
$f(\sigma_{n,t}, g)$: The fitness function of the population contain $\sigma_{n,t}$ at gth generation
e: The overall probability that the new population will be created by mutation and recombination
$P(\sigma_{n,t}, g)$: The population matrices contain $\sigma_{n,t}$ at g th generation

4 DATASET AND PRELIMINARY ANALYSIS

4.1 Review of the Dataset

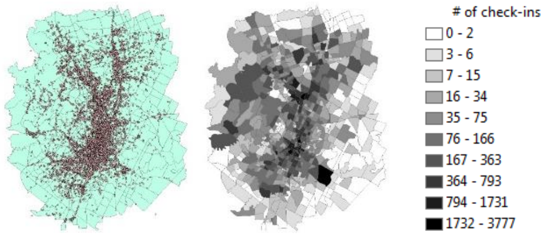
We selected the city of Austin, Texas as the study area. Austin is a diverse city that had a July 1, 2013 population of 885,400 people (U.S. Census Bureau estimate) and encompasses an area of 272 mi^2. The data used in this paper can be categorized into three parts: the Geographic Information System (GIS) data of the Austin Central area, the personal trip data from the Capital Area Metropolitan Planning Organization (CAMPO) and the check-ins statistics from Foursquare. The GIS data is used to define the boundary of the traffic analysis zone in term of the spatial pattern analysis of trip arrival and check-ins occurrence in the area of the city of Austin. There are 730 identified traffic analysis zones (TAZ) by CAMPO within the city of Austin's jurisdiction, which will serve as the study area for this paper. CAMPO's 2005 Travel Demand Model (TDM) serves as the ground truth analysis used for comparison. Furthermore, the zonal OD matrix data from CAMPO contain the modeled daily trip tables for 17 detailed trip purpose, here we selected 10 categories of trip purposes which are listed as follows:

- Home Based Work Person Trips Direct
- Home Based Work Person Trips Strategic
- Home Based Work Person Trips Complex
- Home Based Non-work Retail Person Trips
- Home Based Non-work Other Person Trips
- Home Based Non-work Primary Education Person Trips
- Home Based Non-work University/College Person Trips
- Home Based Non-work UT-Austin Education Person Trips
- Non-home Based Work-related Person Trips
- Non-home Based Other Person Trips

The calculated TOD factors for selected trip purposed will be used as the ground truth dataset in the section of model calibration. The zonal OD matrix data from CAMPO will be applied for developing zonal attraction table in the section of model evaluation and model application.

4.2 Preliminary Analysis of the Check-ins Data

A preliminary analysis is conducted on the characteristics of the check-ins occurrence by investigating both the spatial and temporal pattern of the check-ins data created by Foursquare LBSN service users. The location of the 124,611 check-ins are represented using a dot in Figure 2(a). As shown in Figure 2(b), a heat map also represents the geographic density of check-ins features on study area by using graduated color areas to represent the quantities of those points. The check-ins are more densely distributed in downtown and north central area of the city.



(a) (b)

Figure 3. Foursquare check-ins locations and their spatial distribution among TAZs

To investigate the temporal characteristics of when people use the Foursquare LBSN services, the number of check-ins are aggregated for every one hour of different hours of day. We show, in Figure 2 the empirical occurrence rates of check-ins, where the x-axis represents the time since 0:00 to 23:59 while the y-axis stands for the occurrence rate of check-ins per hour. Figure 2 clearly indicates that the normal check-ins flow pattern is different with the conventional traffic pattern which usually has AM peak, PM peak and lowest trip arrivals during the night time.

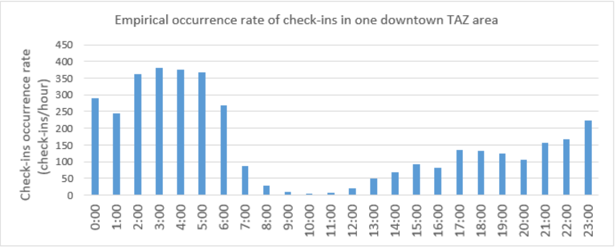


Figure 4. Empirical occurrence rates of check-ins in one downtown TAZ area

5 MODEL CALIBRATION AND EVALUATION

5.1 Model Calibration

For the section of model calibration, given by the daily personal trip tables disaggregated into 10 selected categories of trip purposes confirmation by CAMPO's 2005 TMD, the proposed model was to be calibrated using the genetic algorithm to obtain the parameters for each set of parameters for the two corresponding venue types, and the objective function is to minimize the MAE (Mean Absolute Error) between the modeled TOD percentages and the ground truth trip attraction TOD percentage. For the CAMPO daily traffic survey data, the rescaling work was applied through balancing total trip attraction volume in general map between the predicted trip attraction matrices and ground truth CAMPO trip attraction matrices. The Time of Day (TOD) regime were developed to allocate hourly trips in each TAZ area in order to explore the temporal and spatial characteristics of the mobility pattern in the study area.

5.2 Experimental Evaluation and Result Analysis

A comparison between the calibrated formulation for dynamic trip attraction and CAMPO trip attraction matrix was done by examining the spatial and temporal distribution of trip attraction rate.

Temporal Distribution We use the percentage of the predicted trip arrivals pattern in the various TOD regimes from proposed model to describe the temporal distribution of trip attraction estimation. Figure 5 shows the calibration results from genetic algorithm. In this paper, we discuss the condition of all trip purposes. While the temporal distribution of trip attraction may become various in term of different trip purposes, the LBSN dataset performs the ability to verify the categories of venue type of check-ins data which can indicates the trip purposes of individual check-ins.

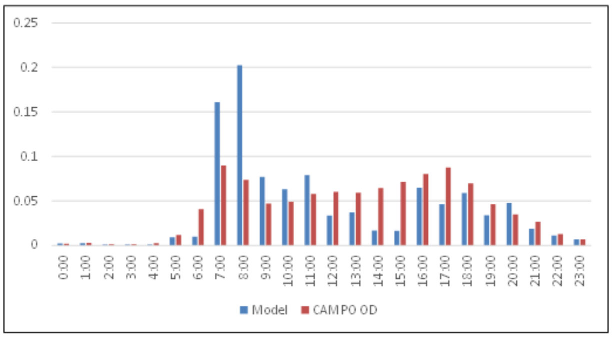


Figure 5. Temporal distribution of the proposed model v.s. CAMPO ground truth data in one downtown TAZ area

As shown by Figure 5, due to the consideration of time of day variation, the predicted temporal characteristic of the human trip arrivals pattern shares the similarity of the percentage of trip attraction with the ground truth data. Meanwhile, in certain section of TOD regimes including the beginning of AM peak and Mid-Day period, the proposed model introduced a relative higher difference due to the high compensation rate for lowest check-in activities arrival periods.

Spatial Trip Attraction flow Pattern Comparison The zonal trip attraction estimation model evaluation is the same as those defined for calculating the swap ratio, shown in Figure 5, whose angle between the trend line of result and line "y=x" can illustrate how well the model output matches the ground truth data. In Figure 5, the horizontal axis represents both the design value of trip attraction for each zone which we used the sorted Zone ID as the value, and the vertical axis is the calibrated trip attraction estimation for each zone based on the design value. Each grid in the diagram displays the calibrated trip attraction I_i for zone i defined as the following:

$$I_i = i * (\frac{TA_i^o}{TA_i^h}) \quad (4)$$

Where
TA_i^o: Trip attraction estimation from CAMPO matrix for zone i
TA_i^h: Ground truth trip attraction from Foursquare matrix for zone i

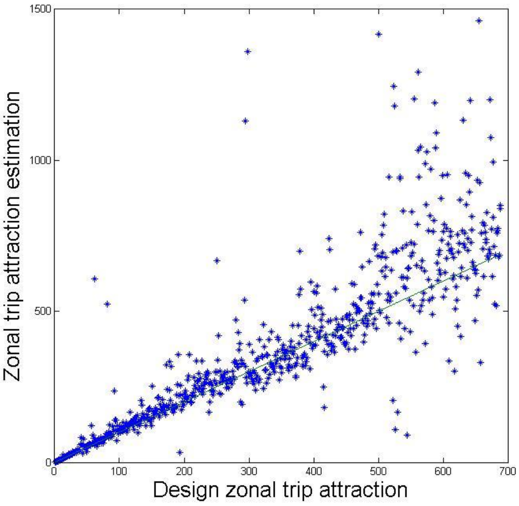
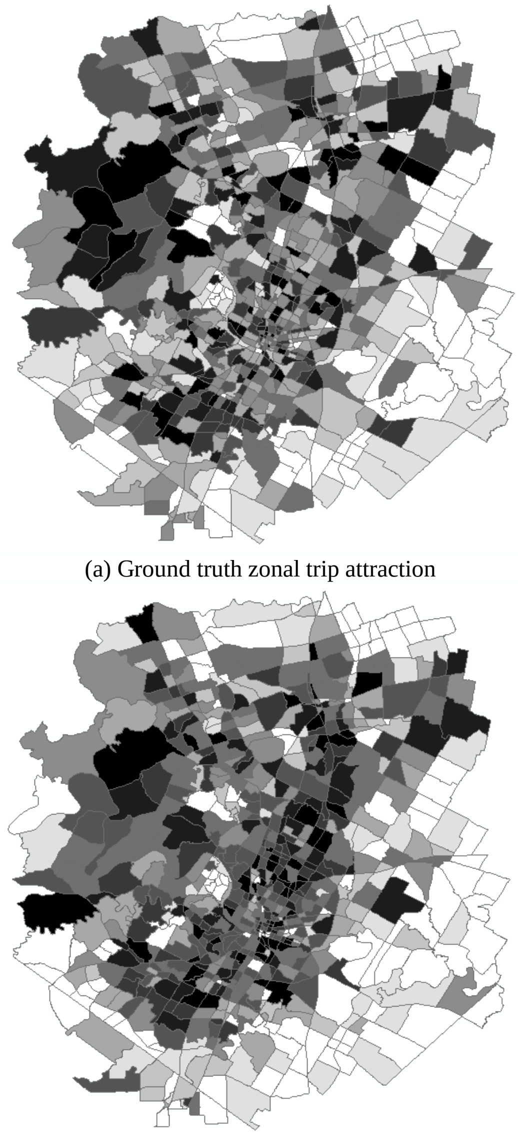


Figure 6. Zonal trip attraction pattern comparison

The nearby position of points to the perfect line represent accurate estimation of zonal trip attraction, and the remote position of points suggest inaccurate estimation. As shown in Figure 6, the trip attraction pattern from the proposed model and the CAMPO ground truth trip attraction pattern exhibit more similar characteristics than the baseline model. Such similarity is in consistence with the small swap ratio value. Some slight inconsistency can be found for trip attraction where the proposed model underestimated or overestimated the trip attraction, especially on some zones most of which are those residential areas may indicate less check-ins activities than the others.

The zonal trip attraction obtained from the CAMPO OD matrix and the modeled matrix are color coded in Figure 7. Larger values were represented by darker color and smaller values by lighter colors. Trip attraction with zero trip counts in CAMPO data are identified as blank areas in the trip attraction pattern figure. Analysis between the ground truth data and the modeled data suggest a consistency between the CAMPO trip attraction spatial pattern and modeled spatial pattern. The proposed model show similar spatial characteristic toward the CAMPO data. However, some inconsistencies can be found within the attraction heat maps. These inconsistencies can be attributed to the relatively lower number of check-in of Foursquare data residences, which may also explain the slight inconsistencies in Figure 7. Such characteristic may cause blank color in some area especially in the residential area which contains few LBSN check-in activities. Despite the inconsistencies, in general, the model shares significant similarity between the trip attraction matrix generated from the model and the CAMPO OD matrix.



(a) Ground truth zonal trip attraction

(b) Modeled zonal trip attraction estimation

Figure 7. Zonal daily trip attraction heat maps

6 Conclusion

This paper investigate the feasibility of using the location-based social networking (LBSN) data to analyze the urban travel demand pattern using a time-dependent model. Given by the check-ins statistics by Foursquare, LBSN data was used to provide zonal trip attraction in the city of Austin Area, which use CAMPO ground truth data to evaluate the performance of the proposed methodology. With respect to traditional and emerging travel demand data collection technologies, LBSN data shows unique potential to investigate better spatial and temporal coverage, overcomes the major limitation of sampling bias, privacy concern and provide trip purpose confirmation by distinguishing the categories of venue types. Compared to the previous study of LBSN data, we explore the feasibility of finding the hidden trip attraction rate based on current check-ins rates and TOD factors. Future research should continue in three areas. Firstly, the discussion of various trip purposes should be examined in temporal and spatial patterns. Typically, LBSN's data was collected by first identifying the venues with the venue ID, venue name, category, latitude, and longitude. An initial analysis of the check-ins should be performed to verify that categories were assigned to the venues such as Professional Places, Food, Shops & Services, and Colleges & Universities. With a categorical breakdown of the venues, the trip purposes were assigned to the corresponding check-ins such as Home-based Work trips, Home-based Non-work Retail trips, and Home-based Non-work School trips identified with the CAMPO study. Secondly, the adaptability of the proposed methodology in the area such as the residential venues whose may contains less LSBN service users should also be examined. Moreover, due to the thriving development of various LBSN like Foursquare, a leading LBSN data provider, the technology and solution of existing social media data integration for transportation agencies should be reviewed to improve the efficiency of data collection, archival, analysis and dissemination. Such examination should be researched to further address the issue of potential lower penetration rate of individual social media platform in case of changing trend in LBSN users' preferences which can be told by app usage and download rate in mobile app stores.

ACKNOWLEDGEMENT

The authors thank Foursquare for allowing the research team to obtain data through the programming interface of their developer application and the Capital Area Metropolitan Planning Organization for providing the 2010 Origin-Destination data used in this study.

REFERENCE

Boyce D. Long-term advances in the state of the art of travel forecasting methods. Equilibrium and advanced transportation modelling: Springer; 1998. p. 73-86.

Bricka S, Zmud J, Wolf J, Freedman J. Household travel surveys with GPS. Transportation Research Record: Journal of the Transportation Research Board. 2009;2105(1):51-6.

Castiglione J, Bradley M, Gliebe J. Activity-Based Travel Demand Models: A Primer. 2014.

Damm D. Theory and empirical results: a comparison of recent activity-based research. Recent advances in travel demand analysis. 1983;3:33.

ITE. Trip Generation Manual. Washington, D.C.: Institute of Transportation Engineers, 2012.

Meyer MD, Miller EJ. Urban transportation planning: a decision-oriented approach2001.

McNally MG. The four step model. Center for Activity Systems Analysis. 2008.

Quercia D, Lathia N, Calabrese F, Di Lorenzo G, Crowcroft J, editors. Recommending social events from mobile phone location data. Data Mining (ICDM), 2010 IEEE 10th International Conference on; 2010: IEEE.

Schapsis C. Location Based Social Networks, Location Based Social apps and games. Links. 2011.

Tsotsis A. A Week on Foursquare. The Wall Street Journal. 2011.

Wolf J, Hallmark S, Oliveira M, Guensler R, Sarasua W. Accuracy issues with route choice data collection by using global positioning system. Transportation Research Record: Journal of the Transportation Research Board. 1999;1660(1):66-74.

Zheng Y. Location-based social networks: Users. Computing with Spatial Trajectories: Springer; 2011. p. 243-76.