Assessing the Value of GPS Travel Data for Travel Forecasting Models

Jeffrey M. Casello
School of Planning and Department of Civil and Environmental Engineering
Xiameng Xu
Pedram Fard
Andrea Mikkila
School of Planning
University of Waterloo, ON Canada
e-mail: jcasello@uwaterloo.ca

Keywords: GPS data collection, travel forecasting, data validation

Introduction

A key component to travel modeling is trip generation – the expected number of trips to begin in a Traffic Analysis Zone (TAZ) (Ortúzar & Willumsen, 2014). Normally, generation rates are estimated using empirically gathered data on household and person trip activity over time. When sufficiently large samples are gathered, the generation rates can be estimated as a function of the travelers' socio-economic data (gender, age, income, household composition, number of vehicles) (McNally, 2000).

This study aims to develop, implement and evaluate a passive data collection method to inform trip generation for a regional travel forecasting model in the City of Edmonton, Alberta Canada. In this research, data are gathered on traveler behavior using two approaches. First, respondents self-report their travel activities using a web-based survey. In addition to the web-based survey, GPS data are gathered using a customized smartphone "app" – EdmoTrack – developed at the University of Waterloo. The specific foci of this research are to assess and improve the functionality of GPS data collection and analysis as a tool to improve the development of travel forecasting models.

Related Research

Paper-based surveys on travel behavior are known to produce low response rates and poor accuracy in those responses received (Hooper, 2017). Various, improved recruitment methods have been studied (Greaves, et al., 2014; Millar and Dillman, 2011). Studies such as Alsnih (2006) and Sills and Song (2002) have shown that compared to paper-based surveys, web-based surveys tend to have higher response trip rates at lower costs. Yet, there remains a tendency to underreport trips, most notably local walking, cycling and other short trips.

The widespread use of smartphones that are equipped with Global Position Systems (GPS) now offers a low cost, passive way to gather traveler information (Shen and Stopher, 2014). The primary benefits of GPS are the ease of data collection and detailed trip records produced. Possible challenges are both technological – the reliability of GPS signals and batter consumption, for example – and social – the potential sacrificing of traveler privacy through their participation in a GPS study.

Methods

This study attempts to answer three major sets of research questions:

1. Is GPS an effective way to validate the self-reported trip activities when building a travel forecasting model? Under what circumstances will the GPS data produced be of sufficient quality to add value to the travel forecasting model development?
2. Can the GPS data be used to correct errors in self-reported trips? More specifically, is it possible to compute the rate of underreported trips such that trip generation rates can be made more accurate?

To answer the first two questions, it is necessary to identify the number of trips made by a traveler. This observation leads to the third set of research questions:

3. What attributes of the GPS can be used in an automated algorithm to identify travel as opposed to conducting an activity? Under what circumstances will this algorithm work well?

We begin by defining a trip as travel occurring between two activities. The most obvious differentiating factor between activities and travel is very low travel speed. As such, the algorithm begins by computing speeds between consecutive points. Speeds that exceed a "stationary" threshold are initially identified as "travel" points, while speeds below the threshold are labeled as possible activities. This initial classification is insufficient to correctly sort travel from activities. Consider the case where an automobile traveler is stopped at a traffic signal. The speed data alone would incorrectly indicate an activity. On the other hand, sufficient "noise" exists in the GPS data that while stationary – for example sitting at one's desk in an office – the GPS data can reflect constant motion at speeds that may be construed as walking velocities.

To further refine the classification, the algorithm next considers the trajectory of movement observed. During an activity, GPS data tend to be circuitous whereas GPS data during travel are typically more direct. To differentiate activities from travel, the directness of travel can be measured as the ratio of total cumulative travel distance to Euclidean distance between a subset of *n* points. Equation 1 expresses this calculation mathematically, while Figure 1 shows the calculation graphically.

$$Directness = \frac{\sum_{i=2}^{n} Distance(x_{i-1},y_{i-1};x_i,y_i)}{EuclideanDistance\ (x_1,y_1;x_n,y_n)} \quad \text{Equation 1}$$

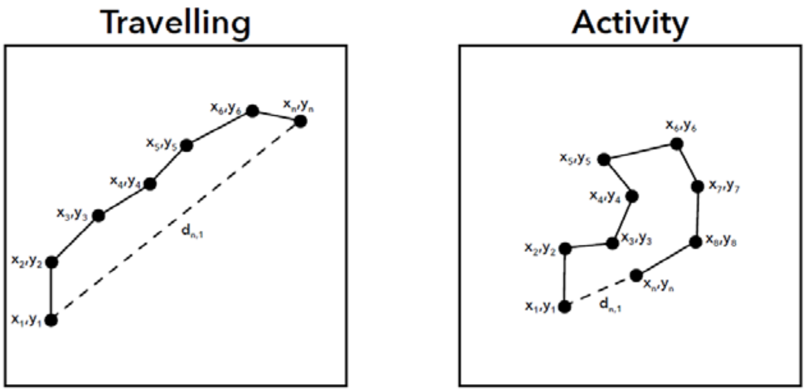
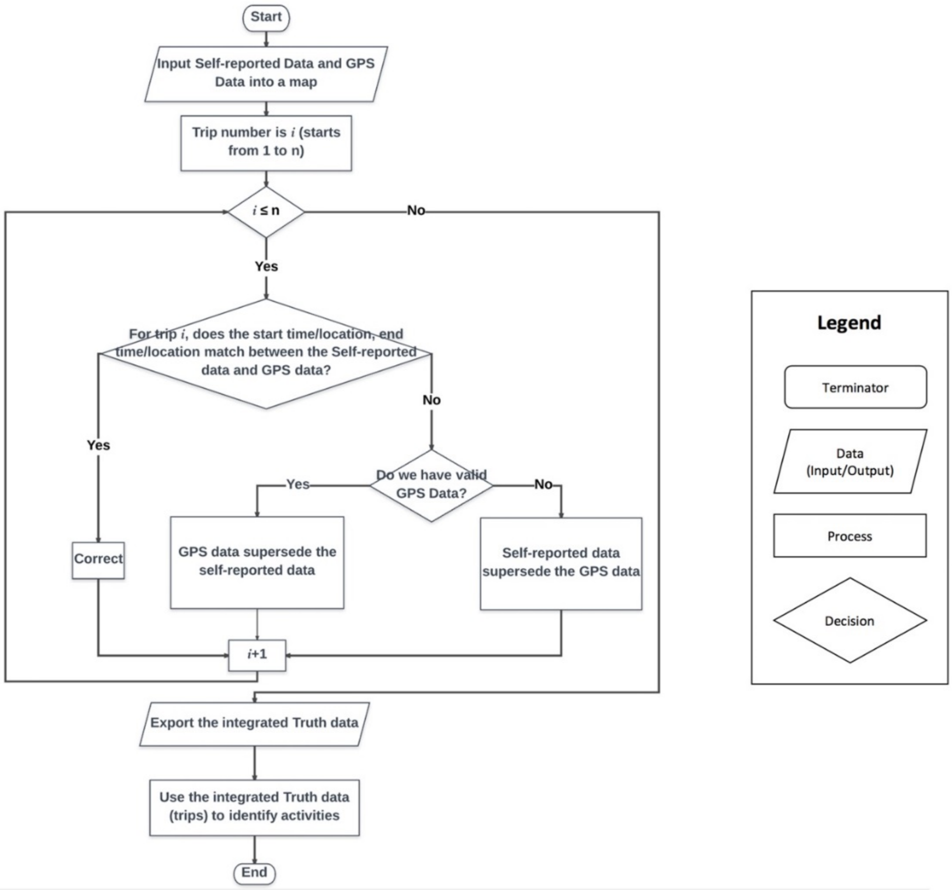


Figure 1 Contrasting GSP data records (trajectories) while travelling or conducting activities

When traveling, equation 1 generates values near to 1; while conducting an activity, results are near to 0. In our work, empirical approaches were taken to generate appropriate thresholds for equation 1, and for the value of *n* to produce the best results.

To differentiate travel stops from activities, a final step is conducted in a GIS. Potential activities – subsets of points with low speeds and circuitous travel – are analyzed in a spatial context. When these sets of points occur within a buffer distance to a roadway or travel facility, the points are re-labeled as travel. The results of these three steps are a set of activities and trips for a participant that can then be compared to the self-reported data.

We now have two records for the number of trips made by respondent – the self-reported data and the GPS analysis data. To determine the actual number of trips taken, we take the following approach. For a given participant, we begin with the first self-reported trip. We compare the reported origin, destination and time to the first record in the GPS file. If these two trips match (within acceptable error tolerances), we consider the trip to be valid and we iterate to the next trip on the respondent's list. If the two data sources do not match and valid GPS data exist, then we assume that the GPS data are correct and the traveler made an error in recording a trip, or failed to report a trip taken. If no valid GPS data exist for the time period when the trip began, then we assume that the traveler made the reported trip, but without gathering the concurrent GPS data. We again iterate through this process until all the self-reported trips have been tested. The result of this step is a record for each participant with the three data points on trip activity: self-reported, GPS identified, and truth. Figure 2 shows the logic.



Results

The analysis demonstrates that about 24% of respondents underreported their trips. Only 8% of respondents overreported their trips. The integrated data set also indicates that the GPS analysis alone underestimates the number of trips made by 65% of respondents, while the GPS analysis generates an overestimate of trip activity for 14% of respondents. The majority of the GPS errors are a result of participants failing to keep the app running on their smartphones. In many cases, respondents turned off the app while beginning an activity and never restarted the program to capture travel. Beyond user error, other sources of GPS error include entry into tunnels where signal obstructions can prevent GPS signals reaching the smartphone; travel by Light Rail Transit, where presumably the electric powered vehicles produced interference which precluded reliable GPS. The GPS algorithm developed also produced erroneous results when travelers engaged in a uniquely North

American travel activity – entering “drive thrus” – as these activities were commonly misclassified as travel stops.

Conclusions

The results of this research suggest that additional training and education may need to be done in order for GPS to become the sole data source for a large-scale, professional studies. But, this work suggests that GPS can be used effectively to validate the self-reported trips, and adjust the generation rates to better inform travel models.

References

Ortúzar, J. D., & Willumsen, L. G. (2014). Modelling transport. Chichester: Wiley

Alsnih, R., 2006. Characteristics of Web Based Surveys and Applications in Travel Research. Travel Survey Methods. Quality and Future Directions, pp. 569-592.

Greaves, S. et al., 2014. A Web-Based Diary and Companion Smartphone App for Travel Activity Surveys. Leura, Australia, s.n.

Hooper, K. G. (2017). Trip generation handbook: an ITE recommended practice. Washington, D.C.: Institute of Transportation Engineers.

McNally, M., 2000. The four-step model (ch. 3). In: Hensher, D., Button, K. (Eds.), Handbook of Transport Modeling. Pergamon, Amsterdam, pp. 35–69.

Millar, M. M. & Dillman, D. A., 2011. Improving Response to Web and Mixed-Mode Surveys. Public Opinion Quarterly, 75(2), pp. 249-269.

Ortúzar, J. D., & Willumsen, L. G. (2014). *Modelling transport*. Chichester: Wiley.

Parsons, C., 2007. Web-based Surveys – Best Practices Based on the Research Literature. Visitor Studies, pp.13-33.

Shen, L., & Stopher, P. R. (2014). Review of GPS Travel Survey and GPS Data-Processing Methods. Transport Reviews,34(3), 316-334. doi:10.1080/01441647.2014.903530

Sills, S. J., & Song, C. (2002). Innovations in survey research: An application of web-based surveys. Social Science Computer Review, 20(1), 22–30.