Spatiotemporal Identification of Trip Stops from Smartphone Data

Adrian B. Ellison[*1], Richard B. Ellison[†1],
Asif Ahmed[‡1], and Stephen P. Greaves[§1]

[1]Institute of Transport and Logistics Studies, The University of Sydney, Australia

09 January, 2015

Summary

As part of a three-year study on cycling infrastructure, a smartphone app was used to passively collect location information resulting in 54 million observations. These data are then used to identify trip stops using a new method that employs a moving average position. In total 12,849 stops are identified with a median time of one hour and a spatial distribution consistent with the travel diary data collected as part of the same study.

KEYWORDS: Transport, Stop Detection, Algorithm, Smartphones, Cycling

1. Introduction

Collecting travel data using smartphones is gaining increasing attention due to the ability to (largely) unobtrusively collect data over time and space. These methods facilitate the collection of ever larger datasets providing geographic data at a detailed and highly disaggregate level across time and space. This being the case, these datasets have potential to be used for analyses of travel behaviour that incorporates both the spatial and temporal variation in behaviour. However, as a direct consequence of this these datasets tend to be very large and, therefore, become too labour intensive to prepare and analyse through existing labour-intensive methods designed largely for traditional travel surveys. In many cases, researchers resort to aggregating the data such that many of the spatial characteristics are lost.

This paper describes a method of identifying trip stops using location data collected from smartphones as part of a study on the impact of cycling infrastructure (Rissel et al. 2013) in Sydney, Australia. In total approximately 54 million observations were recorded over two data collection periods in 2013 and 2014. The aim here is two-fold. The first of these is to develop a method that accurately and consistently identifies the locations and durations of stops between and during travel from smartphone data. That is to say, stops that occur at a destination (or activity), intermediate stops and stops while waiting for a train, bus or other transport mode. This enables continuous analysis of travel data on a large scale. The second objective is to use the latitude and longitude of these detected stops as a basis for identifying the spatiotemporal characteristics of these stops by combining the smartphone data with existing geographic data.

1.1. Context

[*] adrian.ellison@sydney.edu.au
[†] richard.ellison@sydney.edu.au
[‡] asif.ahmed@sydney.edu.au
[§] stephen.greaves@sydney.edu.au

While transport research contains a long history of using personal GPS devices to measure travel (for example, Murakami and Wagner 1999; Stopher, FitzGerald, and Zhang 2008) alone or in combination with a travel diary, not all of the methods employed are transferrable or scalable to smartphone data collection. For instance, many studies using GPS employ methods that are dependent on the availability of reliable Doppler speed (Stopher, FitzGerald, and Xu 2007; Doherty, Papinski, and Lee-Gosselin 2006); a measure that is not available in data collected using smartphones without the aid of the (battery-draining) built-in GPS. In any case, even with speed data available many studies rely on manual map-editing to identify any false-positives and false-negatives of which there are many. This also potentially introduces the issue of inconsistency of the analyst or analysts involved. Although this may be practical when there are a finite number of participants carrying these devices for a small number of days (typically one to three) with the potential increase in scale this becomes infeasible. The time lag necessary to accomplish this also reduces the ability for researchers to ask participants for additional information due to the limitations of memory recall.

2. Data Collection

The data were collected as part of a three-year study on the impact on cycling of new dedicated bicycle paths. The study combined a travel and health questionnaire, an online seven-day travel diary and a smartphone app to passively record travel at five-second intervals in three data collection periods in the (Australian) spring of 2013 (baseline), 2014 and 2015. Participants were recruited from two inner-city areas shown in Figure 1, an intervention area in which the bicycle infrastructure was being built and the control area in which there was no change in the provision of bicycle infrastructure (Greaves et al. 2014). To simplify management, access and analysis, all data collected during the study is stored in a single relational database that can be queried as necessary on an *ad-hoc* basis or using algorithms such as the one described in this paper.

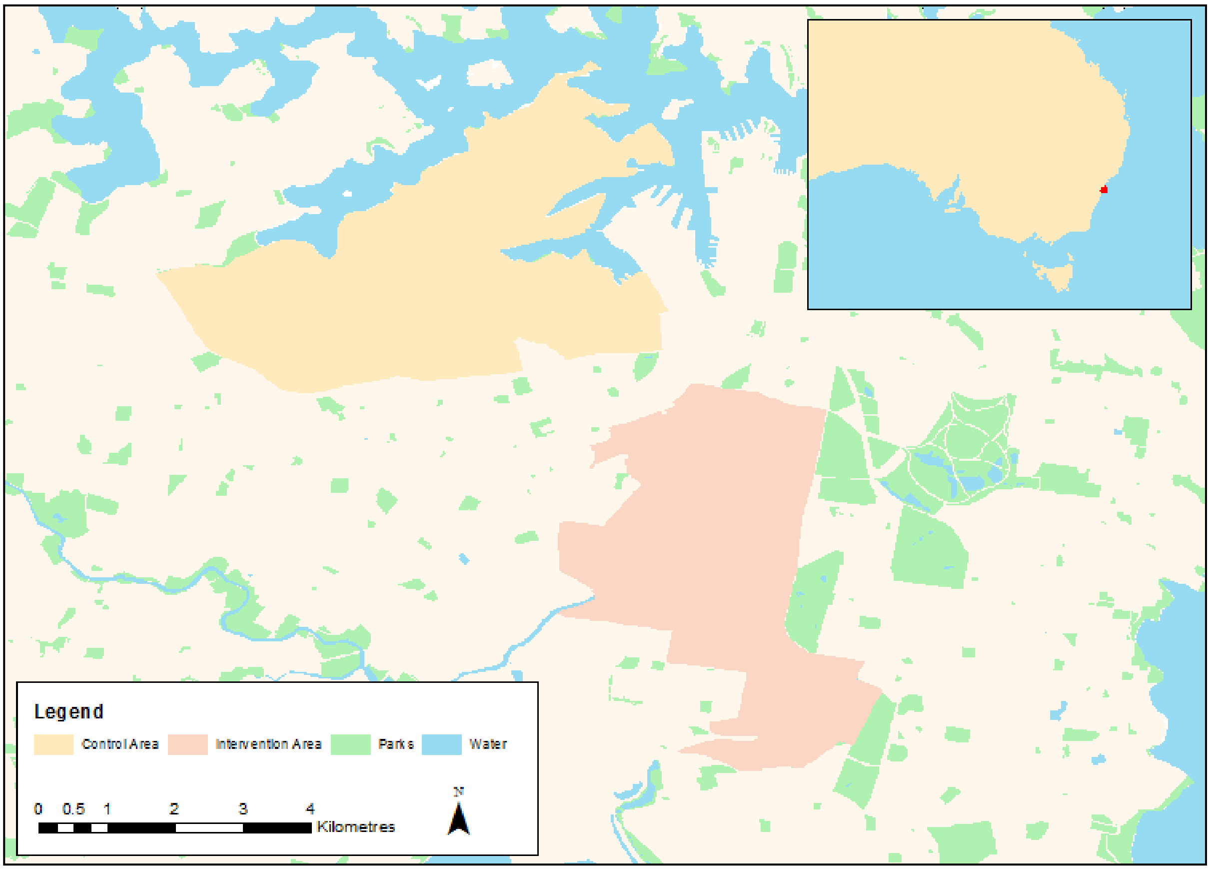

Figure 1 Location of Control and Intervention Areas

The entire smartphone dataset contains almost 54 million observations collected from 469 study participants (many of whom over two years) and an additional 68 unidentified participants who likely downloaded the (free) smartphone app despite not being involved in the study. This represents 12,384 person-days of data albeit with an unequal distribution of days by user.

To verify the accuracy of the stop detection method an alternative source of data is needed. In this case, the seven-day travel diary is the best available source of comparison data. As such, this paper excludes smartphone data collected from those without a valid and complete travel diary leaving 401 unique participants with 38 million observations during 9,552 person-days. For the same reason, data collected outside the seven-day diary period was excluded from this particular analysis.

3. Stop Identification Method

The stop identification method was designed to identify the times and locations in which users stopped from their smartphone data alone. Crucially, stops in this case are not intended to delineate trips and therefore stops to change mode or to pick up something or someone are considered to be valid stops. Similarly, a day spent entirely at home would generate one 'stop'. Simultaneously, the algorithm was designed to move away from the rule-of-thumb time duration methods common in GPS trip identification by focusing on distance.

For each eligible participant, the algorithm starts by retrieving the observations in chronological order from one participant and one day at a time. Subsequently, the algorithm loops through each observation with a recorded accuracy of less than 1,000 metres simultaneously using the latitude and longitude to maintain a rolling average position and a total time within a 150 metre radius of the average position. If an observation is within 150 metres of the average position then the average position and the amount of time within the 150 metre radius is recalculated including the observation. If this is not the case (i.e. the observations is located greater than 150 metres away from the rolling average position) then the time between the current and previous observation is removed from the time spent inside the 150 metre radius.

Once the total time spent within the 150 metre rolling average position exceeds five minutes, then a stop is deemed to have occurred at that location. At that point the algorithm continues but reverses the situation in which time spent outside the 150 metre rolling average position is added to a time spent moving and time spent inside the 150 metre rolling average position is subtracted from the same variable until zero is reached. If the time spent outside the rolling average position reaches 50 seconds then the previous stop is considered to have ended.

A final step is to loop through each of the detected stops and check that consecutive stops are at least 300 metres apart. This is done to as it was found that when indoors, moving from one side of the building to another could create spurious stops in the data due to (on average) less accurate positions.

This process ensures that the algorithm is not as susceptible to spurious data which would otherwise suggest there is movement where there is none. This is particularly problematic when somebody is located inside as this tends to increase the proportion of spurious locations observed in the dataset. The methodology is also computationally efficient allowing it to be used while data is being collected in addition to as a post-processing tool.

4. Identified Stops

Running the algorithm resulted in 12,849 stops (shown in Figure 2) being detected from 382 participants. This compares to 16,660 trip 'legs' reported in the travel diary by the same participants. The average number of stops per participants was 34, the median was 31 and the highest was 99. Unsurprisingly given the study area, most stops occurred within the control area, the intervention area or the Central Business District (CBD).

In terms of time, the minimum stop duration was 75 seconds as measured from the first observation within a stop to the last observation within the stop. The maximum stop time was 24 hours due to the requirement for at least one stop to be identified each day. As such, where only one stop was detected this was equivalent to staying in the same location for one calendar day. The average and median stop durations were three hours and one hour respectively illustrating the extent to which (in absolute terms) most stops are short. The second highest concentration of stop durations was observed at eight hours of duration corresponding to the time between midnight (i.e. the start of a new day) and 08:00 in the morning and departures to work.

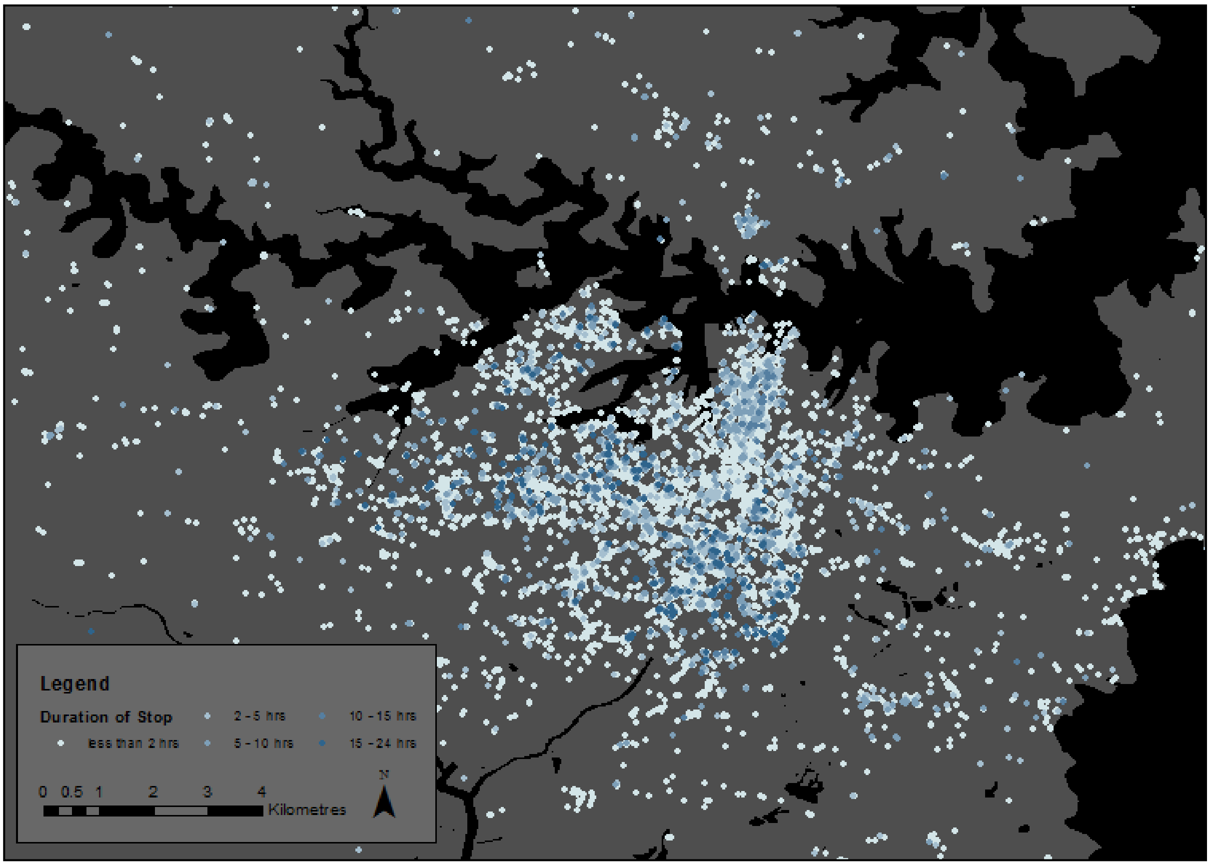

Figure 2: Location and Duration of Stops

5. Conclusions

This paper describes a new method of identifying stops dynamically from large location datasets collected using smartphones. It examines the way in which clusters of observations within 150 metres of a rolling average position can be used to indicate that a stop has occurred in that location. This can then be compared to other sources of data or combined with geographic data to allow for spatiotemporal analyses of travel (or lack thereof). Most importantly, it permits the analysis of very large datasets in a manageable and scalable manner increasing the potential to extract useful information.

6. Biography

Adrian B. Ellison is a Research Fellow at the Institute of Transport and Logistics Studies (ITLS), The University of Sydney. Adrian's main research interests are in road safety, active travel and the use of GPS and smartphones to collect spatially aware data.

Richard B. Ellison is a Research Fellow at ITLS. His current research interests include modelling of freight transport and its environmental effects. He is also involved in several projects on cycling as well as broader research on the interaction between transport infrastructure investments and other wider benefits.

Asif Ahmed is a PhD student at ITLS. His thesis title is Analysis of travel time expenditure and budgets from multi-day multi-year GPS data. This study aims to undertake a completely new exploration of the concept of stable travel-time budgets using multi-day and multi-year data collected by personalised GPS devices.

Stephen P. Greaves is a Professor of Transport Management at ITLS. Stephen's current research is focused around the health/environmental/safety impacts of transport, active travel including cycling, and innovative travel data collection methods using the latest technologies.

References

Doherty, Sean T, Dominik Papinski, and Martin Lee-Gosselin. 2006. "An Internet-Based Prompted Recall Diary with Automated GPS Activity-Trip Detection: System Design." In *Annual Meeting of the Transportation Research Board*. Washington D.C.

Greaves, Stephen P, Adrian B. Ellison, Richard B Ellison, Dean Rance, Chris Standen, Chris Rissel, and Melanie Crane. 2014. "A Web-Based Diary and Companion Smartphone App for Travel/Activity Surveys." In *International Conference on Transport Survey Methods*. Leura, Australia.

Murakami, E, and D P Wagner. 1999. "Can Using Global Positioning System (GPS) Improve Trip Reporting?" *Transportation Research Part C* 7: 149–65.

Rissel, Chris, Stephen P Greaves, Li Ming Wen, Anthony Capon, Melanie Crane, and Chris Standen. 2013. "Evaluating the Transport, Health and Economic Impacts of New Urban Cycling Infrastructure in Sydney, Australia - Protocol Paper." *BMC Public Health* 13 (January). BMC Public Health: 963–70. doi:10.1186/1471-2458-13-963.

Stopher, Peter, Camden FitzGerald, and Min Xu. 2007. "Assessing the Accuracy of the Sydney Household Travel Survey with GPS." *Transportation* 34 (6): 723–41. doi:10.1007/s11116-007-9126-8.

Stopher, Peter, Camden FitzGerald, and Jun Zhang. 2008. "Search for a Global Positioning System Device to Measure Person Travel." *Transportation Research Part C: Emerging Technologies* 16 (3): 350–69. doi:10.1016/j.trc.2007.10.002.