Belfast Mobility: Extracting Route Information from GPS Tracks

Davies, G.[*1], Huck, J.J.[†2], Whyatt, J.D.[*1], Dixon J. [‡3], Hocking, B.[‡3], Jarman, N.[§4], Sturgeon, B.[§4] and Bryan, D.[^5]
[1]Lancaster Environment Centre, Lancaster University
[2]Department of Geography, University of Manchester
[3]Department of Psychology, Open University
[4]Institute for Conflict Research, Belfast
[5]Institute of Irish Studies, Queen's University Belfast

January12th, 2017

Summary

Within highly segregated North Belfast 233 volunteers were recruited and using an app installed on their GPS enabled smart phones were asked to record their movements for a two week period. This created a unique and valuable dataset. Processing these data is however non-trial with over 25million track points recorded across a wide variety of devices. This paper presents how these recorded track points were subsequently cleaned, separated into tracks and assigned travel mode. Once prepared these data can be used investigate the impact of segregation on individual's movements and use of space, and be useful for informing local policy aimed at opening up more shared spaces.

KEYWORDS: GPS, Mobility, Segregation

1. Introduction

North Belfast is a highly segregated environment in which 'loyalist' and 'nationalist' communities live near each other yet inhabit very separate spaces, separated by both visible and invisible boundaries. GPS tracking has the potential to increase our understanding of the use of and movements within these segregated spaces (Roulston, Hansson, Cook, & McKenzie, 2016). In an attempt to further understand how this segregation affects the day to day movement within the city, 233 volunteers have been recruited from different communities to participate in the Belfast Mobility Project. Participants installed a tracking application on their GPS-enabled smartphones which captured their movements every 4 seconds. Participants were asked to use the application for two weeks before uninstalling it from their phones. The amount of recorded GPS points however varies considerably between participants from only a few points to 829,518 points. As of 10th January 2017 a total 24,572,763 GPS points had been recorded by all participants.

Once processed these data offer the potential to determine routes and stops and that can be analysed in order to identify: areas of segregation and shared spaces; routes taken by members of given communities to reach services such as shops, schools, parks and the extent to which segregation affects route choice; the proportion of time participants spend within areas of different community affiliation. There is also potential to explore the extent to which other factors such as time of day or gender affect the nature of movement through segregated space. In order to make use of the captured GPS locations, the data first needs to be cleaned, then tracks, stops and travel mode identified (Du & Aultman-Hall, 2007).

Many methods have been developed to identify tracks and stops where all data have been captured

* gemma.davies@lancaster.ac.uk; d.whyatt@lancaster.ac.uk
† jonathan.huck@manchester.ac.uk
‡ john.dixon@open.ac.uk; bryanna.hocking@open.ac.uk
§ n.jarman@conflictresearch.org.uk; b.sturgeon@conflictresearch.org.uk
^ d.bryan@qub.ac.uk

using the same GPS system (Thierry, Chaix, & Kestens, 2013; Wu, Jiang, Houston, Baker, & Delfino, 2011). Our data, however, was captured from a wide variety of devices and finding a method and threshold values that could be successfully applied in all cases was, therefore, more challenging (Du & Aultman-Hall, 2007). This paper highlights some of the challenges faced in preparing this potentially valuable data resource.

2. Methodology

2.1 Data sources

All data were captured using GPS-enabled Android smartphones belonging to participants. Data points where captured every 4 seconds and stored locally in a SQLite database. Once the smartphone connects to a Wi-Fi network the data was uploaded to a remote PostGIS database (Whyatt, et al., 2016). The fields available in the collected GPS data were: latitude, longitude, accuracy, and a timestamp. An additional field was created in the database recording the speed between track points.

The location of potential track destinations including supermarkets, schools and playing fields were captured as polygons using information from a variety of sources (OSNI maps, Google Maps, Google Street View, information from Belfast City Council and local knowledge). The polygons included the surrounding environment where appropriate (e.g. supermarkets and their car parks; schools and their playgrounds).

2.2 Track and stop separation

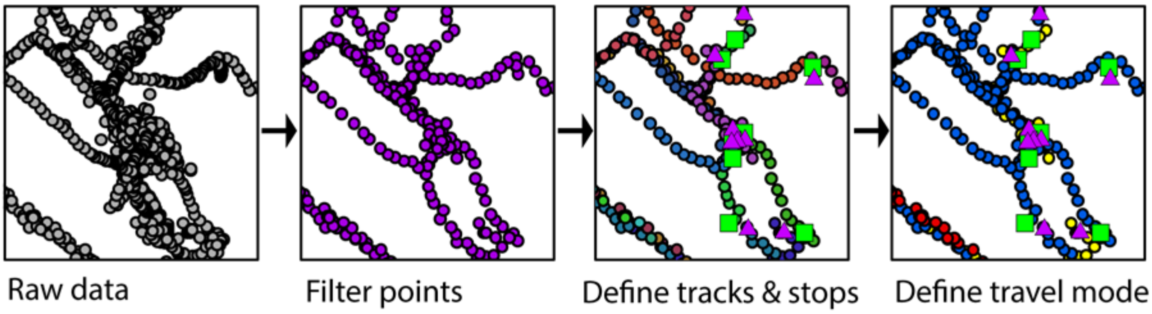

Figure 1 The key stages in GPS data processing to define tracks and stops

Figure 1 demonstrates the key steps in the data cleaning process. Points first need to be filtered to remove those of low accuracy (e.g. scatter created while GPS is indoors) and points where in close proximity to others (<10m apart). Given differences in how devices capture point and record accuracy, finding a method that helps clean noise from the data while minimising the loss of relevant track information is a non-trivial task. The next step is to defined stops and separate points into tracks where there is a time gap of greater than 2 minutes between filtered track points. From this individual journeys can be identified. Tracks were then filtered to remove short tracks wrongly defined from remaining noise in the data, these occurred particularly around home locations where the device was switched on for long periods indoors. As home addresses were available for all participants, these were able to be used to filter all tracks where half of the track points were within 50m of home. Home locations were geo-coded using https://www.doogal.co.uk/BatchGeocoding.php, then manually checked and corrected using a detailed OSNI map to ensure they were located as accurately as possible. Removing all routes with a mean heading change greater than 60° further removed many small tracks where there were unrealistic changes in direction suggesting that the track was more likely to be noise than an actual journey. Finally travel mode was estimated, primarily to determine the likelihood that journeys were on foot or in vehicle. The track and stop separation method we adopted was adapted from that developed by Bohte and Matt (2009), designed to work for data from multiple GPS devices. Details of the sequence of filters applied to the data can be found in Table 1.

Table 1 Workflow for data cleaning and separating of stops and tracks. Process marked with * where taken directly from Bohte and Maat (2009).

Step	Process	Outcome for example user
0	Input raw data	155,936 points
1	Remove a point if its recorded accuracy is greater than 20	97,922 points
2	Remove a point if its distance from the previous point is less than 10m*	2,911 points
3	Remove a point if the speed is greater than 150kmph	2,910 points
4	Define a stop and split track where stop duration is greater than 2 minutes*	117 tracks
5	Remove a track where one third of the points have a speed less than 5kmph and the duration is greater than 1 minutes*	117 tracks
6	Remove tracks where the diagonal length of the track's bounding box divided by the trip distance is less than 0.3*	105 tracks
7	Remove short tracks with less than 4 points*	50 tracks
8	Remove tracks where 50% of the points are within 50m of defined home address	50 tracks
9	Remove tracks where the mean heading change is greater than 60°	44 tracks
10	Merge adjacent tracks where greater than 50% of the points fall within the same destination polygon*	44 tracks

In order to define travel mode we adopted the speed based definition used by Bohte and Maat (2009), using this to define tracks likely to be walking or in vehicle and also those with greater uncertainty (classified by Bohte and Maat as cycling). Travel mode was therefore defined with the following rules:

1. If average speed < 10kmph and maximum speed < 14kmph, define mode as 'walk'
2. Otherwise if average speed < 25kmph and maximum speed < 45kmph, define mode as 'uncertain'
3. Otherwise if average speed < 200kmph define mode as 'in vehicle'

3. Further work

With tracks and stops clearly defined this large and unique data resource provides the opportunity to explore the extent to which segregation affects people's movements through space, as well as identifying shops and services used by individuals and communities. Importantly we can also begin to identify favoured routes, shared spaces and neighbourhood areas with particularly high levels of segregation. The ultimate aim being to help inform policy makers within the city, identifying potential areas that could be opened up to increase the amount of shared space within the city.

4. Acknowledgements

This work is being funded by the ESRC under grant ES/L016583/1.

5. Biography

Gemma Davies is the GIS Officer for the Lancaster Environment Centre, providing support for teaching and research throughout the department.

Jonny Huck is a Lecturer in GIS at the University of Manchester with research interests in the representation of vague geographical entities in geographical information science, novel approaches to cartography, and the application of new technologies to geographical problems.

Duncan Whyatt is a Senior Lecturer in GIS at Lancaster University with research interests in both the natural and social sciences.

References

Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies*, *17* (3), 285-297.

Du, J., & Aultman-Hall, L. (2007). Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues. *Transportation Research Part A: Policy and Practice, 41* (3), 220-232.

Roulston, S., Hansson, U., Cook, S., & McKenzie, P. (2016). If you are not one of them you feel out of place: understanding divisions in a Northern Irish town. *Children's Geographies*, 1-14.

Thierry, B., Chaix, B., & Kestens, Y. (2013). Detecting activity locations from raw GPS data: a novel kernel-based algorithm. *International Journal of Health Geographics, 12* (1), 1-10.

Whyatt, J., Davies, G., Huck, J., Dixon, J., Hocking, B., Jarman, N., Sturgeon, B., & Bryan, D. (2016). Belfast Pathways: Space-Time Analysis of Everyday Activities in a Segregated Environment. In *GISRUK*, Greenwich, UK.

Wu, J., Jiang, C., Houston, D., Baker, D., & Delfino, R. (2011). Automated time activity classification based on global positioning system (GPS) tracking data. *Environmental Health*, *10* (1), 1-13.