Consumer Data Research Centre

An ESRC Data Investment

movement strategies

An exploration of mobile data: towards proximity based passenger sensing on public transport

Christian Tonge[1], Dani Arribas-Bel[1] and Daniel Stockdale[2]
[1]University of Liverpool, [2]Movement Strategies

Project Background

Urban public transport systems have traditionally been planned with limited understanding of their customers travel patterns. Ticket gates and fareboxes yield disaggregate counts of passenger activity on specific stations, vehicles and individual passengers. However, greater information about origins, destinations and interchanges of individual passengers are typically acquired through much smaller samples of costly and therefore infrequent passenger surveys. Substantial research has been conducted on more intelligent systems for passenger monitoring, although little work has sought to capitalise on the wireless mobile sensors that transport customers carry. As such this study aims to evaluate the potential that a large geolocated mobile device data set has for gaining a better understanding of urban transit users.

Data and Methods

A methodology is developed for proximity based passenger sensing on the London bus network using location data collected from the Global Positioning System, and to a lesser extent Bluetooth. As the primary purpose of the data is for providing hyper-contextual adverts and not to estimate public transport patronage, a significant amount of pre-processing is required on the raw data.

The raw data is taken from 5 consecutive weekdays from Monday 3rd to Friday 7th July 2017, capturing a total of 66,179,367 data events composed of 393,360 distinct users. Initial investigation rendered Bluetooth and speed less useful determinants of ridership despite capturing daily means of 0.01% and 21.4% of total events as bus users respectively. Automated vehicle location data is used to enrich the GPS data, which is used to show 13,144 unique bus journeys on 3 London bus routes with origins and destinations for each.

Key Findings

This study has offered the data filtering and methodologies required to understand the boarding and alighting points using passenger proximity on the TfL network, offered at aggregations higher than previously permitted using traditional automated data collection systems alone. Pre-processing techniques have highlighted the issues of horizontal accuracy and users with multiple devices within the data.

100% of journeys using the combined data have origins and destinations, a figure higher than any previous GPS study in the literature. The busiest points of boarding and alighting are predominantly at stations which infers polymodal travel within the network. The distance travelled by users to the Oxford Street region can be used to indicate that the area attracts users from over 15km away, highlighting high levels of retail centre attractiveness. The successful cleaning processes and methodology developed in this study will allow more conclusive results to be drawn on OD matrices and other important insights. Such insights include temporal variability, daily and seasonal ridership patterns, whole network comparisons and data penetration levels. However, due to the novelty of the methodology and the spatiotemporal resolution at user level it is difficult to justify the success of the study against another form of data.

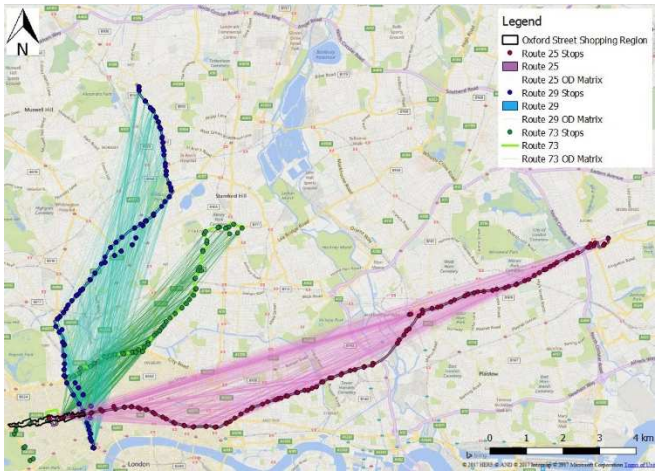

Figure 1. Origin Destination Matrix of Trajectories at Bus Stop Level

Value of the Research

As the first attempt at defining a methodology for this data, the time constraints prevent more interesting and certain insights from being gained than OD ridership levels. While the novelty of this methodology makes it difficult to compare to other data sources, it has been shown to infer aspects of public transport activity that are not immediately available from automated data collection systems. Implementing this methodology over a whole network for a longer period would help to provide more useful insights for transport planners and marketers alike. Despite this data being London centric, the pervasiveness and ubiquity of GPS and the relative ease of implementing SDK applications, makes this method reproducible wherever vehicle location and passenger GPS data are available.