

Statistical Modelling and Analysis of Sparse Bus Probe Data in Urban Areas

Andrei Iu. Bejan

Abstract

Congestion in urban areas causes financial loss to business and increased use of energy compared with free-flowing traffic. Providing citizens with accurate information on traffic conditions can encourage journeys at times of low congestion, and uptake of public transport. Metricating a city to provide this information is expensive and potentially invades privacy. Increasingly, public transport vehicles are equipped with sensors to provide real-time arrival time estimates, but these data are sparse. Our work shows that these data can be used to estimate journey times experienced by road users generally.

In this study we investigated how access to a large repository of bus trajectories combined with other publicly available data, such as OpenStreetMap (OSM) data and National Public Transport Access Node (NaPTAN) database, can be used in order to provide a detailed account of journey times and some of the important factors that affect them. Specifically, we (i) describe what a typical data set from a fleet of over 100 buses looks like; (ii) describe an algorithm to extract bus journeys and estimate their duration along a single route; (iii) show how to best visualise journey times and the influence of contextual factors; (iv) validate our approach for recovering speed information from the sparse movement data.

Several types of graphical displays, a nonparametric quantile regression and monotonic splines technique are used to visualise these effects on journey time and recover speed information from sparse data. We have also been able to study in some detail the sources of delay within individual journeys and have presented our findings using a novel and insightful technique based on a notion of local time profile which we define. Our techniques provide a detailed basis for understanding journey time behaviour in the urban environment and lead to further important research areas such as how to incorporate real-time data to improve prediction of journey times.

This is a joint work with Richard Gibbens, David Evans, Alastair Beresford, Jean Bacon, and Adrian Friday; presented at the STATWORKS workshop (Statistical modelling and inference for networks), 28 June–1 July 2010, Bristol, UK (<http://www.sustain.bris.ac.uk/ws-statworks/>).

Keywords: sensor networks, bus probe data, journey times, large scale data analysis, knowledge discovery, quantile regression, spline interpolation