Exploiting OpenStreetMap topology to aggregate and visualise public transport demand

Mark Dimond*, David Brenig-Jones and Neil Taylor

Integrated Transport Planning Ltd.

January 13, 2017

Summary

Despite a profusion of mobile computing devices and positioning sensors, new technology has yet to fully realise the benefits of better data collection in the field of transport modelling. This paper presents a method for surveying and measuring public transport demand, and equally importantly, a scalable way of collating the resulting data. Data recorded manually by on-vehicle surveyors through a smartphone 'app' can be stored against OpenStreetMap road networks using map matching. The benefit of this is a very flexible data model for recording of demand data, which allows interrogation of network statistics about demand and supply.

KEYWORDS: topology, public transport, OpenStreetMap.

1 Introduction

Public transport routes must be continually managed to ensure they meet the needs of the populations they serve and maintain operational efficiency. This means adding, removing or re-routing services when necessary to account for changes in demand. But to achieve this requires detailed knowledge of the routes served, and the extent to which the services that run are over- or under-utilised by passengers relative to capacity (and vehicle traffic in general). Busy services can be upgraded, whilst services with spare capacity can be removed or rerouted. Increasingly, data needed to support this intelligence can be derived from electronic sources: from ticket machines, from vehicle telemetry, and from app-based manual surveys of passenger numbers.

Aggregating this data showing public transport demand can be best performed by corroborating with existing timetable data, as this allows comparison of related services and understanding performance against timetables. The General Transit Feed Specification (GTFS), a simple text-based format that captures these timetables and the locations visited, is emerging as an international standard for transport supply[1]. In this paper we describe an approach for collating demand data

*dimond@itpworld.net

[1] https://developers.google.com/transit/gtfs/reference/

using this knowledge of supply. We also outline the working of a prototype visualisation tool for exploring this data, reducing the need for knowledge of desktop GIS packages. This aggregation relies heavily upon the street network topology implicit in OpenStreetMap, which acts as a central means of identifying trip data. This permits greatly improved management and visualisation of recorded demand data. For example, in Figure 1, the number of passengers on each surveyed segment of a road network are shown (across all services), with areas of high demand highlighted in orange and red.

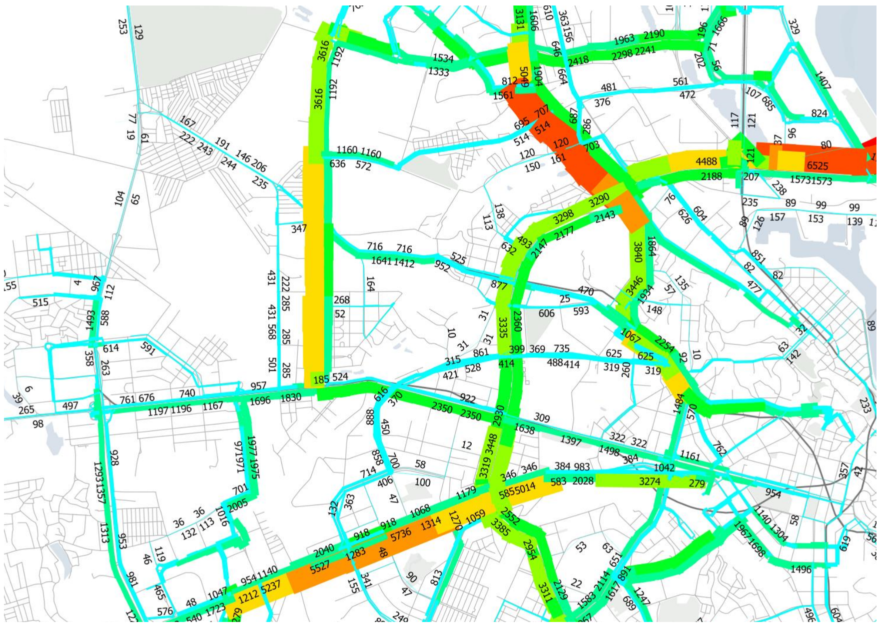

Figure 1: Passenger counts from different services aggregated and visualised in GIS. Data shown is from a large survey project in Kiev, Ukraine, which captured 2,816 surveys of 300 routes.

2 Background

The General Transit Feed Specification (GTFS) has become the international standard for managing public transport network data. However, the geography in a GTFS file is not topologically linked, meaning that it does not include data about which road segments are used by services or even which services share sections of road. The relationships between similar transport routes are therefore not trivial to determine. Figure 2 shows an example: routes captured by transit operators only approximate the geography of the road network, often due to having been recorded by GPS or map-based digitisation. In previous work (Dimond et al., 2016), we present a method for map-matching GTFS routes (which are described by stop locations) to a given city's road network topology as derived from OpenStreetMap. This greatly improves querying of public transport *supply*, meaning that routes that share road segments can be checked for eg. congestion or duplication of service.

Figure 2: Typical capture of GTFS routes (coloured lines), and topology in OpenStreetMap

In this work, we present an extension which allows understanding of an equally important element of transport planning: population *demand* for public transport services. Unlike supply, the level of demand experienced by services is not available in a common data standard. Methods for estimating demand may include on-vehicle passenger counts or analysis of ticketing data. Better visualisation of demand permits valuable and actionable insight to the transport network. For example, less well used services can be combined or removed, whilst those that are busy can be targeted with more vehicles.

Modelling of demand on a precise scale such as this has received limited attention to date. Work by (Horbury, 1999) describes a method for deriving coarse estimates of passenger numbers from timings of vehicle arrivals and departures relative to timetables. (Young, 2016) presents an architecture for deriving specific transport modelling statistics from diverse data such as OpenStreetMap and GTFS (and even describes how trip demand can be incorporated) but this focuses on regional and citywide statistics rather than specific route surveys. In (Bejan et al., 2010), the authors present a framework for estimating realistic public transport journey timings, but this does not account for the level of crowding experienced by passengers or potential for combining services. (Nanni et al., 2014) present an impressive method for collecting journey data for the general public from cellular telephones, and estimating routes taken. But they note (p23) that the transport modes used cannot generally be determined, and so further analysis or post-hoc processing is necessary to determine which data refers to public transport.

3 Proposed Solution

Recording of public transport demand can take place at varying scales, from individual vehicle trips to city-wide population preferences. Our work focuses on the most detailed of these scales, capturing passenger numbers of specific vehicle journeys at every stop along the routes taken. Data regarding passenger movement is captured through an app, TransitWand[2], which is operated by a surveyor whilst on the vehicle. The surveyor counts passengers boarding and leaving the service each time the vehicle (eg. bus or tram) stops. The app records this passenger count data against GPS and the 'stop' points denoted by the surveyor. Though this is a labour-intensive approach, our visualisation tool aims to be agnostic about data sources, and is designed to also accept spatially located demand data from automatic sources such as ticket machines.

3.1 Trip matching

Given a set of these surveys (or other demand data) recorded across a transport network, our aim is then to reconcile these with the expected trips published in a GTFS feed by a transport operator. Traffic variation, GPS error, and surveyor mistakes can all mean the stopping points recorded do not precisely match a known trip pattern. Nevertheless it is possible to estimate the closest using a curve-matching algorithm such as the Fréchet distance. We use the Fréchet distance to determine the closest GTFS trip to the route surveyed, and record a set of 'survey points' capturing travel time, stop time, and the number of passengers boarding and alighting. Note that this matching process would not always be fully necessary: a better survey methodology would require surveyors to select the exact trip surveyed. Nevertheless, this would still be subject to mistakes, unexpected service diversions, and GPS error.

Recorded demand is attached to the graph of public transport services already captured in GTFS. In our previous work (Dimond et al., 2016) we augment the GTFS graph by describing a simple map-matching process that links transport services to the OpenStreetMap (OSM) road network using a routing algorithm. Using this workflow here, we capture public transport *demand* against individual road segments (graph edges), having matched trips using the process above. Passenger numbers can then be calculated for the entire route travelled by a particular service, even inferring passenger numbers at road segments that fall between bus stops.

3.2 Querying, aggregation and visualisation

Given a set of surveys matched against a GTFS transport network in this fashion, there are several types of query possible relating to both public transport and the road network more generally:

- by route, showing the passenger numbers observed on particular vehicle routes over multiple surveys
- by road segment, collating the observed usage of different services traversing a particular highway network. Total numbers of passengers (on all services) can also be calculated for a specific segment.
- by area, allowing estimation of demand across all road segments in a given neighbourhood or for a whole city.

[2]http://www.transitwand.com/

Each of these types of queries requires full enumeration of the demand on each road link of a surveyed transport route. This stage is necessary because passenger movements are recorded against stoppings of a service. A service may not have stopped at a particular road link, but could nevertheless have been transporting passengers at that point[3]. Our software iterates over the road segments recorded for a particular trip, and tracks the surveyed boarding and alighting between surveyed stop locations. This means that the numbers of passengers using the service can be determined for any point on the route, as illustrated in Figure 3. Once the road links visited by a surveyed service have been enumerated, they can be aggregated to resolve the queries above.

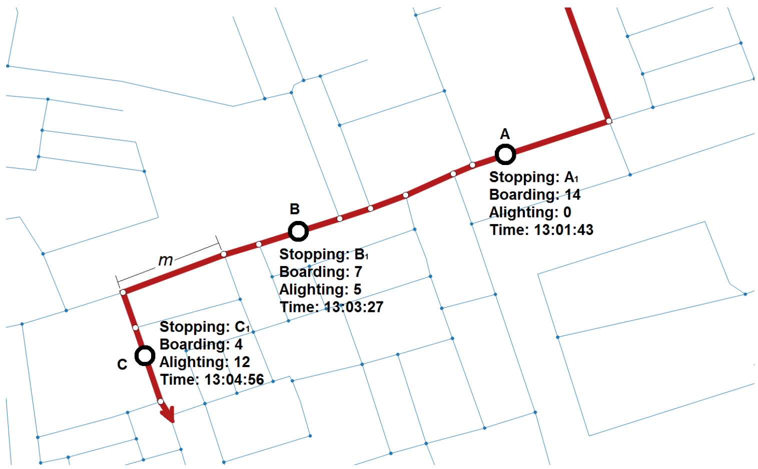

Figure 3: Though on-vehicle passenger numbers are only updated at stops, it is possible to calculate the number on a vehicle at any part of a route. For example, the passenger count at segment m is derived by adding data from stops A and B. Note also that timestamps allow calculation of traffic speed on each segment.

Aggregation of the surveyed passenger counts depends upon the type of query. Intuitively, the time at which a survey occurred is directly relevant to whether it satisfies a particular high-level 'question' related to parts of the transport road network. For example, transport planners may be interested in the total number of bus passengers at a road link in a given peak travelling period (usually around 7am-9am); they may require a different time window; or they may wish to undertand off-peak or weekend travel loads.

Querying software must therefore allow selection of the surveys that satisfy such requirements, by picking only surveys that were performed at specific times of day and on the same days (or sets of days) of the week. We are currently finalising a web application which allows users to specify timings of queries, and geographically interrogate demand based upon service route, road segment, or neighbourhood (as described above). This app will streamline the production of demand visualisation such as that in Figure 1, reducing the need for desktop GIS expertise and improving

[3]An alternative approach would be to perform map-matching in real time using the survey device. We avoid this method, as it would require each device to have accesss to a stored road network. Though this may be feasible for smartphone-based app surveys, we want the survey to be extensible to other sources of data such as ticket machines and onboard Wifi logs.

the timeliness of planning intervention.

4 Conclusions

This paper presents a method for collating on-vehicle passenger numbers for public transport services against a city's OpenStreetMap road network. Using topology from OSM enables sophisticated querying of public transport supply and demand data, enabling queries about the network not supported by current systems. In future work we will continue to develop the interactive visualisations necessary for demand data in particular, presenting aggregated results of surveys as detailed intelligence about transport networks. We will also further investigate 'incidental' sources of public transport passenger numbers: smartcard- and smartphone-based ticketing, onboard WiFi access logs, and machine-learning over CCTV streams, to begin to remove the survey requirement for timely transport network intelligence.

5 Acknowledgements

This research was supported in part by the UK Department for Transport's 'Technology in Transport Innovation Grant' 2016. Data in Figure 1 thanks to a World Bank funded project.

6 Biography

Mark Dimond is a software developer and researcher with an interest in spatial data mining, particularly for infrastructure applications. His work at ITP focuses on the production of tools for working with transport-related open data formats.

David Brenig-Jones is a data modelling expert with particular experience deploying big data to analyse and optimise transport systems. He has developed innovative techniques that allow transport markets in complex cities to be understood in much greater detail. Using these analyses he has developed citywide public transport optimisation plans for Metro-Manila and Kiev.

Neil Taylor brings a people-centred focus to his work which allows him to understand transport issues from the perspectives of the user. Neil leads much of ITP's work in the emerging field of intelligent mobility and has delivered highly innovative crowd-sourcing and open data projects in cities like Manila, Mexico City and Dhaka.

References

Bejan, A. I., Gibbens, R. J., Evans, D., Beresford, A. R., Bacon, J., and Friday, A. (2010). Statistical modelling and analysis of sparse bus probe data in urban areas. In *13th International IEEE Conference on Intelligent Transportation Systems*, pages 1256–1263.

Dimond, M., Taylor, N., and Houghton, R. (2016). Estimating and editing transit topology over the road graph using supply data feeds. In *Proceedings of the Association of Geospatial Information Laboratories Europe (AGILE)*.

Horbury, A. X. (1999). Using non-real-time automatic vehicle location data to improve bus services. *Transportation Research Part B: Methodological*, 33(8):559 – 579.

Nanni, M., Trasarti, R., Furletti, B., Gabrielli, L., Van Der Mede, P., De Bruijn, J., De Romph, E., and Bruil, G. (2014). Transportation planning based on gsm traces: A case study on ivory coast. In *Citizen in Sensor Networks*, pages 15–25. Springer.

Young, M. (2016). An automated framework to derive model variables from open transport data using r, postgresql and opentripplanner. In *Proceedings of 24th GIS Research UK (GISRUK)*.