

## Analysing taxi GPS data for mobility management

Peer-reviewed author version

LIU, Feng & HERMANS, Elke (2016) Analysing taxi GPS data for mobility management. In: Proceedings of Big Data & Business Intelligence in the Travel & Tourism Domain, p. 59-62.

Handle: <http://hdl.handle.net/1942/21508>

# Analysing taxi GPS data for mobility management

Feng Liu & Elke Hermans

Transportation Research Institute (IMOB)  
Hasselt University, Belgium  
elke.hermans@uhasselt.be

## 1 Problem Definition

This study investigates the pressure of festivals on the urban transport network in the city of Harbin in China during the annual international ice festival in January. Within this city (as is the case in other emerging cities), the population size significantly increased over the past decade resulting in additional vehicles joining the daily urban traffic flow. Apart from daily life, successful events attract extra tourists and thus users of the network. During such events, transport problems aggravate, leading to significant negative local externalities (e.g. Albalade & Bel, 2010). The flow of tourists along with a growing number of trips by locals puts a serious challenge to the city's transport system.

## 2 Related Literature

Several types of data are currently used for conducting studies on tourist travel behaviour and transport studies. Survey data, in some cases complemented with detailed movement traces (e.g. by GPS devices) are often used (e.g. Orellana et al., 2012). This offers information on tourist experiences and movement patterns. However, privacy concerns, the burden imposed on respondents and the relatively high data collection cost lead to rather small samples of tourists being involved in the survey over a very short time interval. Recently, advances in multimedia and mobile technologies have allowed large volumes of user-generated travel photos to be created and shared, and subsequently analyzed (see e.g. Vu et al., 2015). Geo-tagged photos with spatial-temporal signatures allow inferring the movement trajectories of the corresponding individuals but the data is only limited to the level of locations, the detailed travel routes concerning road links as well as the actual traffic conditions along the routes are not revealed by the data. This research uses data collected by means of GPS devices installed in taxis. This data collection method implies no additional costs nor respondent burden and are an interesting way of complementing traditional data collection in transportation research, and in the analysis of travel demand and transport conditions at festivals in particular. In fact, the value and effectiveness of taxi GPS data for the estimation of traffic conditions and the modelling of travel demand can be shown (see e.g. Lu & Li, 2014 and many more). However, the data has so far not been explored for the city-wide analysis of the additional travel demand patterns generated during a specific festival, nor for the examination of the particular transport conditions in this period.

### **3 Methodological Approach**

Taxi passenger travel patterns on the urban road network are first modelled based on Global Positioning System (GPS) data collected from taxis, both during the festival and in a ‘normal’ period. From this, accommodation and visit zones that originate and/or absorb substantially more trips during the festival are identified. Next, zone pairs are detected characterized by high levels of additional travel demand but a low travel speed. Finally, the specific road conditions in the detected areas are further examined and optimal travel routes are determined.

A preliminary step before starting the analysis consists of data cleaning and passenger trip identification. For each of the obtained passenger trips, four key variables are computed, i.e. the travel time, travel distance, travel speed and route directness. Next, the urban area is divided into zones along the latitude and longitude directions using a grid-based partition method. In addition to the spatial partition, the temporal dimension of the trips is classified into different time slots of the day, the specific day, the day of the week and the type of the day. Based on this, a passenger travel pattern matrix is constructed (showing the total number of trips between 2 zones (A and B), as well as the number of trips that either start in zone A or end in zone B). In a next step, the accommodation and visit zones with more trips during the festival than in the normal period (i.e. additional travel demand for several days higher than the threshold) are identified.

Problematic zone pairs, characterized by high levels of additional travel demand during the festival but suffering from poor transport conditions and low travel speeds are detected. Finally, specific road conditions are examined and optimal routes discovered. To this end, map-matching occurs (based on the algorithm of Qudus & Washington, 2015 for low-resolution data). The method assumes that vehicles follow the shortest path between a current GPS point and its previous point and a path searching algorithm is employed. The link with the highest score is chosen as the matched road to the current GPS point. For each of the obtained links, the average driving speed and average travel time along the link are computed. That way, the network consists of congested roads and less-used-roads (the latter constituting the partial network). To identify precise trip start or end locations in the study zone pair, a clustering process is applied. Finally, the optimal routes which have the shortest travel times (or distances) under the partial network are determined.

### **4 Results**

The GPS data is collected from all licensed taxis in Harbin. The positions of the vehicles are recorded every 30 seconds during the day (every 2 minutes at night) generating data of 1.6 gigabytes in size and around 24 million GPS points each day. Compared to a ‘normal’ month, in January, the total volume of taxi passenger travel demand rose considerably (15%) leading to a total of extra taxi trips of 1.96M across the whole festival period.

In the data cleaning phase, 0.08% of the total points were eliminated (having coordinates equal to zero). The spatial partition resulted in 1600 regions (each being 1.87 km<sup>2</sup> in size), and 5 time slots were used. 19 and 34 zones were identified as

festival-accommodation-zones respectively festival-visitor-zones. These areas undertook more than 100 additional passenger trips per day for at least 91% of all the festival weekdays. Next, 55 zone pairs were identified as festival-travel-demand-zone-pairs. 48 from the 55 zone pairs undertake high additional travel flows but suffer from travel speeds lower than 20km/h on 91% of all the festival weekdays. They are thus diagnosed as festival-transport-problem-zone-pairs. In the paper, the pair with the highest level of additional travel demand (+49.5% passenger trips) during the festival is analysed in more detail to propose an optimal route (i.e. a shorter travel distance and time).

## **5 Research Outlook**

From a theoretical perspective, common and interesting techniques used to analyse traditional travel data (e.g. diaries or road sensor data), such as travel demand modelling, transport network analysis, map-matching, clustering and shortest path searching are integrated and applied to a large set of GPS data. In the future, apart from the detection of optimal routes that connect the festival-accommodation-zones and festival-visit-zones, the method can be applied to search for efficient paths between the festival-visit-zones and the hotel-zones characterised with low increase-in-hotel-occupancy-rates, attracting more tourists to lodge there.

## References

- Albalade, D. & Bel, G. (2010). Tourism and Urban Public Transport: Holding Demand Pressure under Supply Constraints. *Tourism Management* 31: 425-433.
- Lu, Y. & Li, S. (2014). An Empirical Study of With-in Day OD Prediction Using Taxi GPS Data in Singapore. *Transportation Research Board 93<sup>rd</sup> Annual Meeting*.
- Orellana, D., Bregt, A.K., Ligtenberg, A. & Wachowicz, M. (2012). Exploring Visitor Movement Patterns in Natural Recreational Areas. *Tourism Management* 33: 672-682.
- Quddus, M. & Washington, S. (2015). Shortest Path and Vehicle Trajectory aided Map-matching for Low Frequency GPS Data. *Transportation Research Part C: Emerging Technologies* 55: 328-339.
- Vu, H.Q., Li, G., Law, R. & Ye, B.H. (2015). Exploring the Travel Behaviors of Inbound Tourists to Hong Kong using Geotagged Photos. *Tourism Management* 46: 222-232.