Enhancing survey methods and instruments to better capture personal travel behaviour

Sheila Ferrer¹, Tomás Ruiz²

¹Transport Department, Universitat Politècnica de València, e-mail: shferlo@upv.es
²Transport Department, Universitat Politècnica de València, e-mail: truizsa@tra.upv.es

Abstract

This paper describes a plan to passively collect personal activity-travel information using data from smartphones. First, an application that requires the respondent to introduce a reduced amount of information will be developed. This information will help to design and calibrate several algorithms to process the GPS and accelerometer data obtained. Then, a second application which requires a passive role of the respondent (except for its installation) will be developed. Using previously calibrated algorithms, it will be possible to study travel behaviour patterns of individuals, with emphasis on capturing short pedestrian trips.

Keywords: Smartphones, GPS and accelerometer data, travel behaviour

Acknowledgements: Spanish Ministry of Science and Innovation: Project reference TRA2011-27415

1. Introduction

Conventional paper-based methods for collecting data are often expensive, time consuming, have problems of missing trips and the level of accuracy is limited. In addition, pedestrian trip data collected is often underreported or inaccurate, and often lacks an adequate sample size to provide useful information. In the context of travel surveys, GPS devices have become a new tool for collecting travel data. GPS provide detailed information of routes and activity allocation in many types of outdoor spaces, allowing analysts to monitor quite accurately respondents’ travel behaviour and avoiding many of the problems of traditional paper surveys. Principal uses of such technology to date have been in the areas of validation of data collected by means of traditional travel and activity diaries, but its potential in characterizing travel behaviour is resulting in an increasingly number of GPS-based travel surveys.

In spite of all the advantages of using GPS data in travel surveys, signal losses or degradation in high-density cities and cold/warm start issues are common problems of these GPS devices. Accelerometers are a very useful complementary tool for tracking people travel behaviour in places with a weak or even no satellite signal. Since
accelerometers measure rotational and translatory movements, they are commonly used for activity-travel recognition. Currently, all modern mobile phones are embedded with various sensors such as GPS and accelerometer. Therefore, the study of people travel behavior using smartphones is feasible and likely to increase as such devices continually gain prominence in modern society. Recent contributions show initial research on the use of smartphones in this area: Manzoni et al. (2011), Deutsch et al., (2012), Nitsche et al. (2012), Parlak et al. (2012), Stenneth et al (2011, 2012), Fan, Y. et al. (2013).

This paper presents an on-going research with the aim to provide recommendations for enhancing survey methods and instruments to better capture personal travel behaviour. By means of a passive activity-travel survey using records from smartphones’ GPS and accelerometer, and the development of appropriate algorithms for automatically processing the output data, the estimation of individual’s mobility patterns will be achievable. Travel behaviour characteristics to be estimated include: starting times and duration of activities and travels; transport modes for each trip leg and trip chains; and activity type classification.

At present, a small sample of individuals is providing us detailed information about their travels and activities through their smartphones in the city of Valencia. This information is being used to calibrate algorithms to identify trips and transport modes characteristics from the GPS and accelerometer data (gyroscope and compass data as well, where available). In a few months, a second sample of respondents from the city of Valencia will participate in a completely passive activity-travel survey using smartphones. Then, the GPS and accelerometer data collected using smartphones will be analyzed implementing the selected algorithms to obtain detailed information about individual’s travel behaviour.

2. Data collection methodology

A data logging application has been developed in Android. It acquires data from the accelerometer of the smartphone, and synchronizes this data with the available location and speed information provided by the GPS. In addition, the phone calculates its position based on the available GSM Cell Ids and WLAN cells when the GPS signal is lost. The GPS records data at a sampling frequency of 10-12 seconds, as higher epoch rates lead to great battery consumption and lower epoch rates are insufficient for pedestrian tracking. In addition to each GPS record, information about the accuracy of the signal is stored, as the number of satellites in view and other parameters available. The GPS and accelerometer information is continuously stored in the smartphone, and it is periodically sent to a server using Internet connection and then deleted from the smartphone’s memory.

The user interface of the smartphone’s application to collect travel data was designed to be as simple as possible. The data collection application runs as a background task, and it is designed neither to distract users in their daily phone activities nor cause any limitations of the phone performance.
As probe data is necessary to calibrate and validate the algorithms for detecting starting times and duration of travels and activities, transport modes, and activity types, a first version of the data logging application was developed to obtain ground truth data from a small sample of respondents (Figure 1). This group of respondents has downloaded this version of the application to their smartphones; obviously, smartphones should be Android-based embedded with assisted-GPS and a triaxial accelerometer. Participants are instructed to manually register every single trip undertaken in outdoor public areas independently of the trip length, providing the following data to train the algorithms: when starting a trip, participants are instructed to thoroughly select the travel mode chosen (walk, bicycle, motorcycle, car, bus, electric tramway, metro, train, or wait if participant is transferring between transport modes) and, once the trip is finished, respondents have to end it in the application and to choose the main type of activity to be undertaken (to go home, work/study, have lunch/dinner, shopping, leisure time - sports, cinema, theatre, etc. - services - hairdresser, doctor, etc. - , pick up someone or other type of activity).

Participants are being selected considering the transport mode they usually utilize. The objective is to get sample data as varied as possible in terms of transport modes used. We are considering that more information may be necessary to be collected for those transport modes easily misclassified by the algorithms.

Tracks can be visualized on a Google Maps-based website. This website can be helpful to evaluate the quality of the data obtained from the GPS and to test the results of the algorithm to remove outliers from the data. In addition, it can help to estimate timing and duration of activities and travels (Figure 2).
Figure 1: Active version of the application
Figure 2: Examples of routes from the active version of the application as visualized on the website
3. Data processing

As the amount of data generated by the GPS and accelerometer is very large, the development of a methodology to automatically process the output data is essential to avoid time consuming. Thus, algorithms to automatically reconstruct individual trips including the mode choice are being developed. First, the models being developed identify starting times and duration of trips. Second, for each detected trip, the travel mode inference recognizes whether the user is stationary, walking, running, biking or in motorized transport (bus, tram, metro or train).

The method developed to date involves filtering the data and extracting suitable features, using statistical models to calculate probabilistic classifications and finally, reconstructing the most likely sequence of transport modes used. Each of these steps is explained in more detail in this section.

In the first step, the information is thoroughly cleaned and smoothed to identify erroneous data points and delete them. The number of satellites in view and the Horizontal Dilution of Precision Value (HDOP) are fairly efficient in determining systematic errors: all data points with too few (less than three) satellites in view and/or a value of HDOP of five or greater are removed.

In the second step, periods of activity (stops) and trips (periods of motion) are determined. Stops are defined as moving slower than 0.5 m/s for a period of 120 seconds. The second criteria to detect activities is based on point density or so-called bundle of GPS points (Stopher et al., 2005), a bundle is a sequence of GPS points positioned very close to each other. Each of the two criteria is used individually to determine potential activity start and potential activity end points. For stop detection and separation of the distinct phases of the trajectory when the GPS signal is weak, activities are detected by means of the accelerometer. As most pedestrian trips are short, GPS signal often fails to capture the complete walking trip. In these cases, the accelerometer data and the implementation of the map matching algorithms are helpful resources to reconstruct these short journeys.

Third, different features are extracted from the GPS and accelerometer signals with a sliding window technique. A size of the time window is chosen as the period of classification, nevertheless, this size is continuously adapted to be large enough to provide meaningful information about the transport mode. Features of the GPS signal to be extracted include: variance of latitude and longitude values, speed and acceleration statistics. To remove the redundant features from the feature set that do not provide additional information, the Correlation-based Feature Selection Method is considered, allowing the identification of the most useful feature subset.

Similarly, accelerometer features are extracted from the acceleration signal. Various features including the mean, variance and frequency domain features (the sum of Fourier transform coefficients at several subbands, energy, etc.) are extracted from the acceleration magnitude signal and by applying the Correlation-based Feature Selection
Method algorithm on the set of accelerometer features, the optimal feature subset is obtained.

Once selected the useful accelerometer and GPS-based features, a feature vector extracted within a time window is formed with both of these features. In the classification step, the feature vector is passed to a classifier to infer the corresponding transport mode.

Next step involves the classification of the feature vectors into transport modes. As other studies have shown, the Decision Tree-based classification is a simple and effective system for identifying transport modes. In this step, about 75 percent of the ground truth data is used to train the models and the remaining 25 percent as a test set for evaluation. After the classifier, the resulting sequence of classifications may show unreasonably trip chains or short periods of using one mode before changing to another. To remove these short term transitions, transition probabilities are estimated from the distribution of the time between mode changes in the training data and the Viterbi algorithm or a Hidden Markov model is applied on the classification output to identify the most likely sequence of classifications.

Later, transport mode determination performance of the model is evaluated in terms of classification accuracy. The confusion matrix is calculated for the complete test set to identify the percentage of predicted class labels with respect to the actual class.

4. On-going and future work

Several algorithms will be trained and tested in the next months, including more refined algorithms. A comparative analysis of results of each algorithm will be carried considering prediction accuracy, computer processing time and the number of pedestrian trips detected. Special attention will be given to the ability of the map matching algorithm and the usefulness of the accelerometer data in reconstructing missed walking trips.

As the quality of the data collected varies to a great extent depending on the smartphone type and the Android version, several tests are being carried out using different smartphone models and version of the operating system. In this respect, latest Android versions are having compatibility problems with GPS receivers. On the other hand, Samsung smartphones collect, in general, better GPS data than other smartphone types.

A second version of the smartphone application has been developed to collect passive travel data to a sample of about 1,000 respondents in the city of Valencia. This sample will be carefully selected to be representative in terms of their socio-demographic characteristics and the transport modes used. As there might be a problem of participation bias towards younger people or smartphones users with a personal technical interest, special incentives will be considered to be given to the underrepresented groups to increase their participation.
In this version of the application, GPS and accelerometer data (gyroscope and compass data as well, where available) is automatically collected from respondents’ smartphones without requiring manual entries or efforts from them by using a passive travel survey collection method. Respondents participating in the survey will download the second version of the application to their smartphones and will be instructed to use the application during one week period, turning it on before leaving home in the morning and turning it off when coming home at night. Data collected will be analysed using the algorithm previously selected to characterize mobility patterns and travel behaviour of Valencia’s citizens.

References


Stopher, P. R., Q. Jiang and C. FitzGerald (2005) Processing GPS data from travel surveys, paper presented at the 2nd International Colloquium on the Behavioural