

Who you are is how you travel: A framework for transportation mode detection using individual and environmental characteristics

Thanos Bantis¹, James Haworth¹

¹University College London

Department of Civil Environmental and Geomatic Engineering
SpaceTime Lab

Abstract

With the increasing prevalence of geo-enabled mobile phone applications, researchers can collect mobility data at a relatively high spatial and temporal resolution. Such data, however, lack semantic information such as the interaction of individuals with the transportation modes available. On the other hand, traditional mobility surveys provide detailed snapshots of the relation between socio-demographic characteristics and choice of transportation modes. Transportation mode detection is currently approached using features such as speed, acceleration and direction either on their own or in combination with GIS data. Combining such information with socio-demographic characteristics of travellers has the potential of offering a richer modelling framework that could facilitate better transportation mode detection using variables such as age and disability. In this paper, we explore the possibility to include both elements of the environment and individual characteristics of travellers in the task of transportation mode detection. Using dynamic Bayesian Networks, we model the transition matrix to account for such auxiliary data by using an informative Dirichlet prior constructed using data from traditional mobility surveys. Results have shown that it is possible to achieve comparable accuracy with the most widely used classification algorithms while having a rich modelling framework, even in the case of sparse mobility data.

Keywords. Transportation mode detection; Dynamic Bayesian networks; mobility; disabilities; smartphones

1 Introduction

A fundamental aspect of an individual’s mobility is the interaction with the available modes of transport. Such semantic information can be used to inform decision making on a variety of transport related topics, such as the levels of accessibility of an area and the degree of transportation demand. The current decade has seen an ever-increasing body of literature exploring machine learning methods to infer transportation modes from mobility data. One of the main drivers for this is the availability of mobility data at a very detailed resolution which is the result of the widespread use of relatively cheap GPS sensors embedded in smart-phones. These methods vary to a great extent and can be used to capture different aspects of human mobility. Most commonly, they are used to decompose human mobility traces into travel modes and activities.

For the purposes of travel mode detection, the methods employed by the majority of studies are based on secondary quantities derived from the raw mobility data. For example, Bolbol et al. (2012) compared the effectiveness of quantities such as speed, acceleration, heading and difference in distance to classify raw GPS traces into transportation modes using a Support Vector Machine (SVM). Zheng, Liu, Wang & Xie (2008) have also used a SVM classification of raw GPS traces to transportation modes as part of a wider assessment of classification algorithms. Recognising the fact that boundaries between different modes of travel may not be crisp when using quantities such as speed, Wan & Lin (2016) used a fuzzy classification approach to transportation mode detection. Patterson et al. (2003) constructed a Bayesian network classifier which was enhanced using a combination of speed with GIS data such as bus routes. In a similar study, Stenneth et al. (2011) integrated GIS with GPS data before assessing the classification accuracy using different classification algorithms (Naive Bayes, Bayesian Networks, Decision Trees, Random Forests, Multilayer Perceptron). They have found a considerable increase in mode detection accuracy in the order of 20%. Similar classifiers to the above studies (SVM, Random Forests) were used by Shafique & Hato (2015) this time using only acceleration measurements. They found that accurate travel mode detection is possible using acceleration data only, thus by-passing the shortcomings of GPS measurements (loss of signal, battery consumption etc.).

Although the aforementioned studies have yielded improvements in classification accuracy at the aggregate level, an individual’s interaction with available transportation modes is treated in a rather simplistic way. Current methods assume that the travel mode depends only on features such as speed, acceleration and characteristics of the transport network and built environment in the form of GIS data. However, socio-demographic factors and personal characteristics of travellers (such as age and disability) are known to influence the choice of transportation modes to a large extent (Tyler 2006; Xie et al. 2016). For example, past research in London has shown that elderly and disabled people are more inclined to use the bus for their everyday journeys as opposed to the Underground (Transport for London 2012). This could be attributed to the fact that many Underground stations do not have step free access. As a result, most of the currently available methods will classify individuals who display more typical behaviour (i.e. able bodied) well, but perform poorly on those individuals whose usage of the transportation system differs from the average, such as elderly and disabled users.

Nevertheless, it is important to understand the mobility of these groups in order to make the transport system accessible to them. This necessitates a richer modelling framework that can account for such additional information. Such a modelling framework will benefit mobility modellers and policy makers alike, the former by providing a way to dis-aggregate

raw mobility traces, while the latter by offering a way to include marginalised, and often vulnerable, population groups when formulating mobility strategies.

This paper sets out to explore this relationship between personal characteristics, socio-demographic data and travel mode classification feature space. It attempts to achieve this goal by modelling individual mobility traces using a hierarchical model built within a dynamic Bayesian network framework. The performance of the model is tested using two individuals with mobility impairments, aiming to provide a proof of concept rather than a generalisation to population groups.

The structure of the paper is as follows: In the following section, a brief literature review on the most common methods of transportation mode detection is given. Section 3 introduces the data used in this study. Section 4 describes the methodology, which is followed by a description the results of the model using mobility traces from the two participants in section 5 as well as a performance evaluation relative to other commonly used classifiers. Section 6 provides a discussion of the results and finally Section 7 provides the conclusions and future directions.

2 Transportation mode detection methods: An overview

The task of transportation mode detection from high resolution mobility data is largely treated as a classification/clustering problem in the literature. A key characteristic of the classification process is that it has to be robust enough to compensate for the ambiguity of differentiating between similar travel modes. For example, in the case of using speed to determine different modes of travel, the classifier has to be able to cope with the uncertainty between travelling by bus on a congested road and walking. This problem becomes even more challenging given the irregular temporal frequency of data collected using data collected through mobile phone applications. Other issues to consider when choosing the overall classification approach is the classifier's ability to include a wide range of internal and external factors that could influence an individual's mode choice.

There are many studies in the literature comparing the efficiency of different classifiers in terms of predicting the correct travel mode from mobility data (Stenneth et al. 2011; Reddy et al. 2010; Lin & Hsu 2014). The following sections briefly describe the most used ones.

2.1 Discriminative models

Discriminative classification models use the conditional distribution of the class given the features to label the samples.

As has already been noted, a popular discriminative classifier within the transportation mode detection literature is SVM. A SVM is a supervised linear classifier that uses a kernel function to transform the original variables into a higher dimension feature space in order to tackle the problem of linear inseparability between different categories. This inseparability is commonly caused by non-linearities in the boundaries of the categories. As a method, it is guaranteed to provide a global solution to the classification problem, however, depending on the kernel specification, this is not the same for the problem of over-fitting.

A common approach to tackle the problem of inseparability between travel modes, is to expand the feature space by using more quantities eg. using both speed and acceleration. However, such measurements might not be available in the first place. Moreover, SVM

classification methods ignore the temporal structure of human mobility data, although there have been attempts to circumvent the problem (Bolbol et al. 2012).

Modifying SVM models to include a wider range of information in the classification problem can be done either by altering the kernel function, or by building the model within a regression framework. As SVMs are supervised classification models, they require a training set which might not be available beforehand. As SVMs are by definition non-probabilistic classifiers, it is difficult to assess the uncertainty in the estimates over the set of classes.

Another commonly used discriminative classification approach to transportation mode detection, often thought to be one of the best performing classifiers for this task (Jahangiri & Rakha 2015), is Decision Trees (Reddy et al. 2010;Griffin & Huang 2005;McGowen & McNally 2007;Zheng, Li, Chen, Xie & Ma 2008). These can appear alone or in combination with a multinomial logistic regression model. A Decision Tree classifier recursively segments the feature space in a binary fashion, based on the principle of minimising some loss function (eg. chi-squared, entropy etc.). An elementary example of a Decision Tree classifier for determining transportation modes using speed would be to "make decisions" on the mode based on how high or low the speed value is.

Decision Trees have the advantage of being direct and easily interpretable. However, they tend not to generalise well as they refer to a particular setting only. Another big disadvantage of Decision Trees is the large variance, especially in the case of correlated features (Janssens et al. 2006). An improvement over this, is the merging of a set of individual Decision Tree classifiers into a single one (Random Forests (RF)) to smooth the individual variances. The downside of using this method is the loss of interpret-ability.

A third family of discriminant models commonly used for transportation mode detection from mobility data is Neural Networks(Zhang et al. 2015;Stenneth et al. 2011;Shafique & Hato 2015). Neural networks are used to approximate complex functions by summing together weighted versions of simpler functions (neurons). These neurons can have a sigmoid response function in case of binary and categorical variables or linear response function in the case of continuous variables. For transportation mode detection, the complex function can represent the boundaries between different mode categories.

Advantages of Neural Network methods include the easiness of including a wide range of variables in the classification process in an straightforward way (Omrani 2015). A disadvantage is the loss of interpret-ability of the classification results due to the dense network of neurons. This fact can make the generalisation of a learned Neural Network to datasets of different spatial resolution difficult.

2.2 Generative Models

Generative models use the joint probability of all the variables in the feature space together with the class probabilities to solve the classification problem. Contrary to discriminative classifiers, they don't define the classification process using boundaries, but rather probability distributions that characterise the classes. This family of models include various versions of Bayesian Networks such as Naive Bayes, Hidden Markov Models (HMMs), Bayes Nets etc.

Bayesian networks impose a structure in the data characterised by a graph in which the nodes represent a conditional probability given the parents of the node following the rules of D-separation. The graph structure makes hierarchical modelling relatively straightforward, as researchers can represent the data generation process in a way that the nodes "tell a story". Figure 1 illustrates a simple Bayesian Network. In this case, information

from the A,B nodes propagates into C which in turn propagates into D. The joint distribution in this case based on D-separation independence assumption is $P(A; B; C; D) = P(D|C)P(C|A; B)P(A)P(B)$.

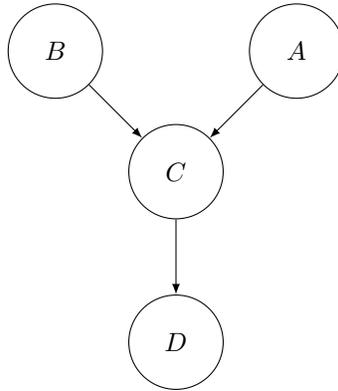


Figure 1: A simple Bayesian Network

The simplest classifier in this context is Naive Bayes. This method assumes complete independence over all variables in the feature space given the class, a condition which is difficult to defend in most cases. As a classifier, it has been found to have a reduced accuracy in the context of transportation mode inference compared to other classifiers when the feature space is limited to quantities such as speed, acceleration and heading (Stenneth et al. 2011; Reddy et al. 2010). On the other hand, in the case the feature space is broadened with variables such as distance to metro or bus lines, Naive Bayes has been found to perform better than any discriminant classifier (Feng & Timmermans 2016).

Bayes Nets offer an improvement over the conditional independence assumption of Naive Bayes and as a consequence, they are able to model more complex relationships between variables in the feature space. Such a model has been found to perform well in transportation mode classification by modelling the conditional relationships of acceleration, speed, trip distance and speed percentiles (Xiao et al. 2015).

The above described models tend to overlook the temporal dependence of mobility data. Dynamic models such as HMMs and dynamic Bayesian Networks attempt to address this issue.

HMMs are Bayesian Networks exploiting the sequential nature of timestamped data. The main assumption is that an unobserved "hidden" time dependent process is the driver behind the observations. HMM are memory-less models, in the sense that a node is dependent only on the preceding node and not on the previous ones. This assumption can be relaxed if a higher order HMM is employed, In this case, however, there is a risk of over-smoothing, making classes less distinguishable. Richer modelling specification frameworks, such as a combination of HMMs with Decision Trees, have been found to provide increased classification accuracy for different modes (Reddy et al. 2010).

Dynamic Bayesian Networks combine the graph structure of Bayesian Networks with the sequential structure of Markov models. By treating problems as time dependent stochastic processes, dynamic Bayesian networks can not only capture the associated uncertainty for each node, they can also reason about the way these evolve over time (Koller & Friedman 2009). This is due to the causal network structure of such models which allows researchers to "inject" domain knowledge in their models. Their flexibility and granularity made these models popular amongst a variety of disciplines such as speech recognition, automatic hand-

written character recognition and DNA sequencing. These models are being increasingly used within the human trajectory mining and activity recognition field within an unsupervised classification framework (Lin & Hsu 2014; Liao et al. 2007). A disadvantage is that, being unsupervised classification algorithms, with given additional data the parameters of the models have to be learned again (Lin & Hsu 2014).

Table 1 below summarises the advantages and disadvantages of some of the cited methodologies.

Table 1: Comparison of different transportation mode detection methodologies

Authors	Method	Accuracy	Sensor	External parameters	Advantages	Limitations
Bolbol et al. (2012)	SVM	88%	GPS	No	Includes a variety of transportation modes	Preprocessing required, inseparability issues between bus and underground mode
Janssens et al. (2006)	BN and DT	53.60%	Travel Diaries	Yes	Includes variety of information related to transportation habits	Combined method underperformed compared to BN and DT alone
Zhang et al. (2015)	ANN	62%	GPS	No	Fast computation even for large datasets	A preprocessing step is required as inference is based in a derivative of speed
Feng & Timmermans (2016)	Naïve Bayes	99.40%	GPS	Yes	Includes a variety of classification features and external parameters	Data obtained from a dedicated GPS logger with a variety of accuracy measures which are beyond the reach of low end GPS sensors
Xiao et al. (2015)	Bayes Nets	90%	GPS	No	Accounts for the inter-dependencies between classification feature space	Potential loss of information through discretisation of continuous variables
Reddy et al. (2010)	HMM	95.8%	Smart phone	No	No preprocessing step	Results were tested on one particular smart phone device only
Liao et al. (2007)	DBN	75%	GPS	No	No preprocessing step, no training set	Complex model

2.3 Summary

From the discussion in the above sections, one can conclude that there is no single classification method that can be used in all cases. The choice depends heavily on the researcher's needs in terms of interpret-ability of results, total number of travel mode categories, application domain as well as the available data specification in terms of spatial resolution, temporal frequency and feature space size. With big data-sets where the goal is inferring transportation modes alone, discriminant classifiers seem to perform best. A further advantage is the speed of inferences, making them suitable for near real-time applications. On the other hand, if the data-sets are relatively small and the goal of the researcher is to reason behind the observed mobility patterns, then generative models are a natural choice. An additional advantage is the natural extension of such models to include a temporal element in the analysis.

The problem of inseparability of transportation modes has been treated using different strategies in the methods reviewed. Some of them include broadening of feature space with additional quantities (Zhang et al. 2015; Bolbol et al. 2012. Xiao et al. 2015), others included secondary (usually GIS) data (Stenneth et al. 2011), while others have chosen to merge overlapping categories (eg. car and bus) (Reddy et al. 2010). In any case, as the number of different modes is increasing, so is the complexity of the model and the required assumptions.

3 Data

For this study, data from a bespoke developed geo-enabled smart-phone application was used. The use of smart-phone apps in mobility studies has seen an increased interest from researchers in the past years in the fields of activity and transportation mode detection (Montini et al. 2015; Kim et al. 2014; Widhalm et al. 2012). The main advantages of using such an approach are the ease and cost efficiency of data collection process as well as the relatively high spatial accuracy, as the majority of smart-phones are equipped with GPS receivers and accelerometers. Often, the most common disadvantage that has been reported is the increased battery demand on the user's devices (Xiao et al. 2015; Wu et al. 2016), especially when both GPS and accelerometer readings are logged. However, for any practical applications, the accuracy and precision of the collected data impose an additional challenge for modelling (Eftekhari & Ghatee 2016; Wu et al. 2016). This can be especially true for middle to low end smart-phones.

3.1 Data collection process and data idiosyncrasies

For this study, a smart-phone application was developed for both Android and iOS devices. The application makes use of android and iOS location APIs to log a user's coordinates in a non-intrusive way while simultaneously managing the trade-off between battery use and coordinate logging in an effective way.

The app was offered to download to two participants experiencing mobility difficulties. The first volunteer is 40-59 years old, female, full-time employed and a crutches user. The second volunteer is a 22-39 year old male, full-time employed and a mobility scooter user. Using the app, their coordinates were recorded at regular time intervals (2 minutes) if there is a significant distance between two subsequent position fixes. This distance was taken to be 30 meters. The raw locations are presented in Figure 2 below. For the crutches user,

the temporal window of observations was 7 days in total, while for the wheelchair user it was 3 days in total.



Figure 2: Raw mobile phone location data

The main disadvantage of pipe-lining the location logging process through the use of an API is that the researcher has limited control over location accuracy and temporal resolution. As the API is using different sensors to determine an individual’s location, the more sensors it utilises the greater the location accuracy. Depending on factors such as sensor availability at the moment of update and battery level, this can vary. For example, a fused GPS/Wi-Fi update can have accuracy in the order of tens of meters, while a GSM cell-tower update can have accuracy in the order of hundreds of metres. Moreover, the data-sets are characterised by data points with inconsistent temporal resolution, as the sampling interval is determined by the background apps that make use of location services. Figure 3 below shows the relationship between spatial accuracy and temporal resolution for the two participants. The spatial accuracy was obtained by querying the API for the estimated confidence interval of the location estimate.



Figure 3: Accuracy vs resolution for the two participants

Such artefacts can have an important effect in the overall classification task, by adding systematic (such as location “drift”) and non systematic noise (such as sudden “jumps” in location) that influence the regularity of point patterns and as a consequence and results in increased variability of the classification features such as speed (Figures 4, 5). This fact adds increased ambiguity in the classification procedure increasing the overlap between states,

especially for the ones that are characterised by more subtle changes such as walking or dwelling.

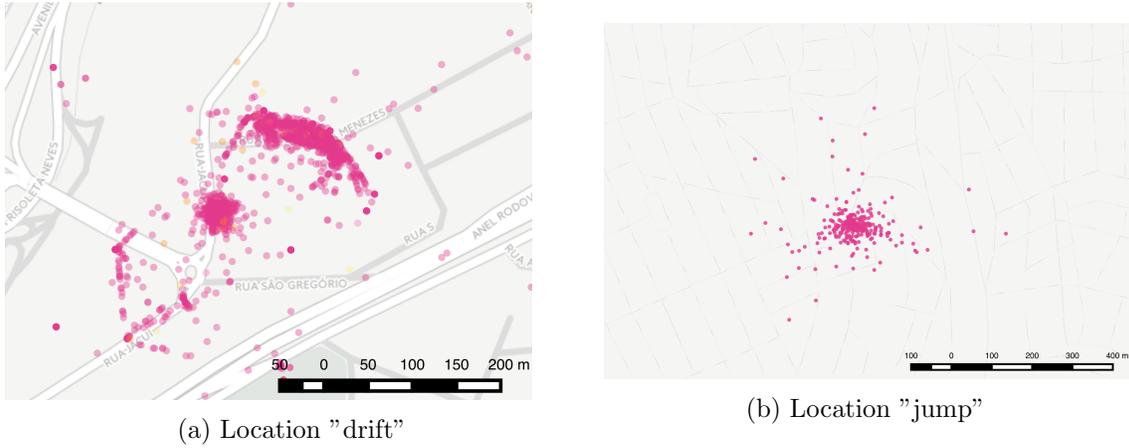


Figure 4: Examples of "jump" and "drift"

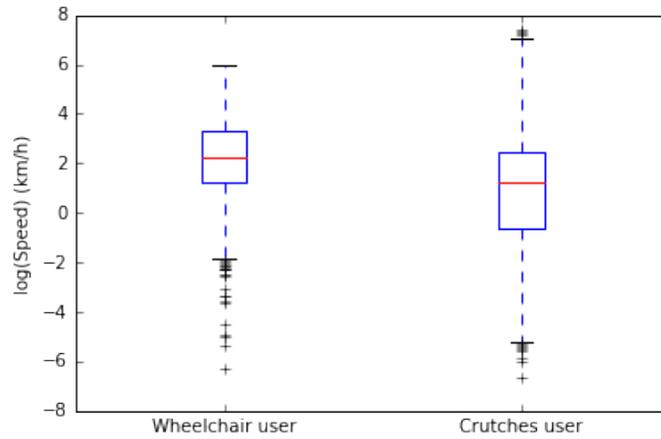


Figure 5: Boxplots showing the variability of the classification feature. For visualisation purposes, the speed was log transformed.

4 Methods

This study uses a HMM as a framework for the modelling procedure for the following reasons:

- There is a requirement for a framework extensible enough to account for factors such as personal and socio-demographic characteristics, while accounting at the same time for the temporal dependency between subsequent location estimates.
- There is an interest in reasoning about the significance of those characteristics.
- There is an interest in quantifying uncertainties of the estimates. This is of great importance considering the variable accuracy of position fixes.

By modelling the transition probability matrix between travel modes, additional information such as age and disability can be included in an indirect way. External factors such as proximity to public transportation means are assumed to influence the mobility patterns in a more direct way and were included in the modelling of speed classification feature. These factors were included as stochastic quantities so that the levels of uncertainty can be assessed as well as their overall significance in the classification procedure. The following sections describe the modelling framework in more detail.

4.1 Model specification

The model is conceptually represented by the the graph of Figure 6. The classification feature space is comprised of speed readings only as determined through the location API's. As already been mentioned, this was done to conserve user's device battery as much as possible, and emulate at the same time the data that are usually available using such API's as a data collection method.

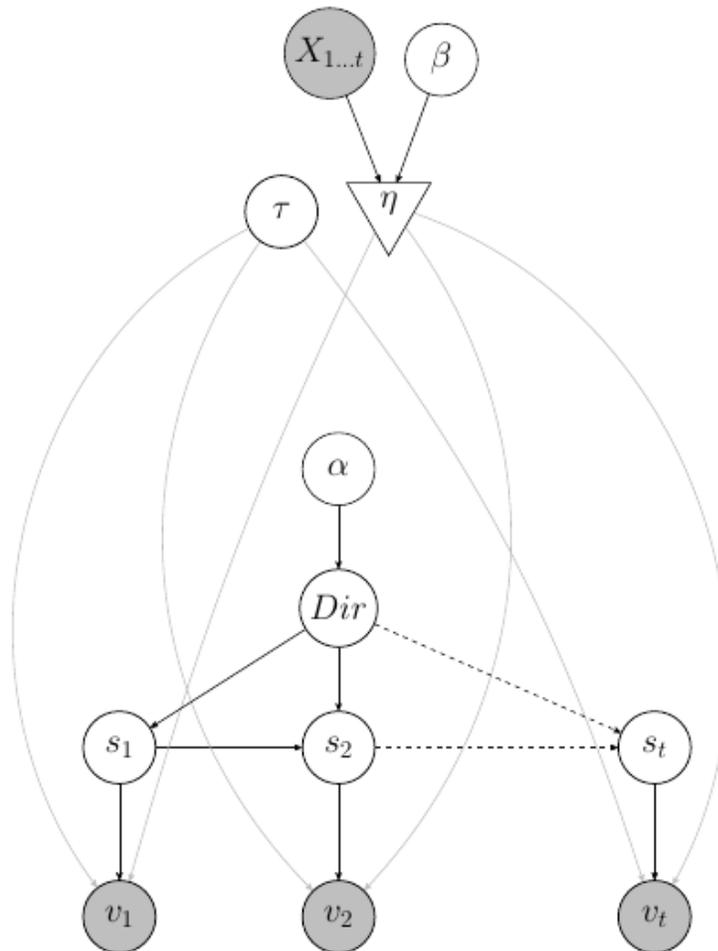


Figure 6: Schematic representation of the model

Table 2: Description of nodes in Figure 6

Node	Description
$v_{1...t}$	Speed
$s_{1...t}$	Transportation mode states
Dir	Dirichlet distribution
α	Concentration parameter vector for the personal preferences
τ	Precision vector of speed node
η	Deterministic function of mean for the speed node
β	Coefficient of external covariates
$X_{1...t}$	External covariates

The different transportation modes were modelled using a categorical probability distribution, having outcomes as described by a predetermined set of transportation modes such that $\sum s^\kappa = 1$ where s^κ are the event probabilities. For this research four different categories were used: being stationary, walking, riding the bus/driving a car and travelling by rail. The emission probabilities $P(v_t | s_t)$ were modelled as a mixture of Gaussian distributions representing the range of velocities each travel mode can take (Patterson et al. 2003; Liao et al. 2007).

A common problem encountered with the above approach during inference is related to the identifiability, or label-switching, between the candidate classes. This refers to permuting the subscripts of the mixture components without changing the likelihood in such a way that the interpretability of inferred classes is lost (Congdon 2010). Various strategies have been suggested in the literature to deal with this problem, from imposing an sorting structure (ascending or descending) on the Gaussian components (Zucchini & MacDonald 2009), to the use of informative priors (Congdon 2010). Due to its simplicity, a sorting structure has been applied in this study such that $\mu_t^1 < \mu_t^2 < \dots < \mu_t^\kappa$ for $\kappa \in \{stationary, walk, bus, rail\}$. The initial probability of using a particular transportation mode $P(s_0)$ was evaluated following the condition for a stationary Markov Chain following (Zucchini & MacDonald 2009). This states that the vector \mathbf{x} is the stationary distribution for the stochastic matrix \mathbf{P} if and only if:

$$\mathbf{x}(\mathbf{I} - \mathbf{P} + \mathbf{U}) = \mathbf{1} \quad (1)$$

where \mathbf{x} is the vector of non negative elements of the stationary distribution, \mathbf{I} is the $\kappa \times \kappa$ identity matrix, \mathbf{P} is the transition matrix and \mathbf{U} is an $\kappa \times \kappa$ matrix with all elements equal to one.

Two assumptions are made at this point:

1. Internal factors have a persistent effect on the ability of individuals to transition from one transportation mode to another
2. External factors have a non persistent effect on the ability of individuals to transition from one transportation mode to another

The first assumption is related to the personal preferences of an individual when switching between different transportation modes. For example, an individual with disabilities

might prefer to use a transportation mode that is more accessible compared with the other, and this preference is assumed to be constant regardless the data one is observing.

On the other hand, external factors, such as whether an individual is located within the catchment area of a bus or a rail station, are assumed to change throughout an individual's trajectory. For example, an individual moving within the radius of bus stops, is more likely to be using a bus.

4.1.1 Including external covariates

For this study, a 30 meter radius around bus stops and rail stations was taken as threshold to define the binary covariates depending on whether a person is located within, or outside this radius. This threshold corresponds to a compromise between the maximum achievable accuracy when the API is using a WiFi/Cell tower level accuracy and the minimum achievable accuracy when using the GPS sensor. Other spatially varying covariates that are assumed to influence an individual's mobility can be included. These could range from socio economic features such as crime levels, to features that characterise the aesthetic quality of a route (Evans 2009). For this study, the Index of Multiple Deprivation (IMD) was taken as a proxy for the level of attractiveness of an area. IMD is an index made up of seven sub-indices relating to features such as income level, employment, health, education skills, barriers to housing and services, crime and living environment. The index ranks the different UK census areas from most deprived to least deprived (UK Government 2015). For this study, the values were normalised to have zero mean and unit variance to assist inference as the scale difference between IMD and proximity covariates ranges from one to two orders of magnitude. Figure 7 below shows the location traces of the mobility scooter participant together with the levels of IMD for each census area.

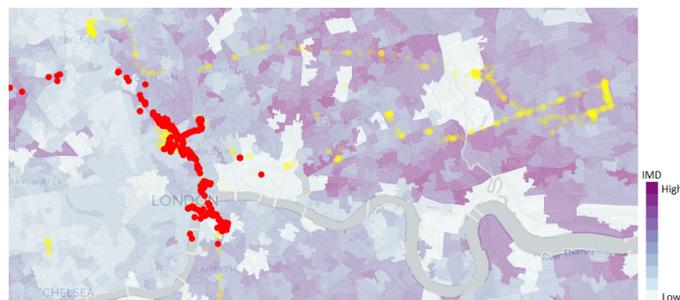


Figure 7: IMD overlaid with a participant's traces. The value of the covariate changes according to the census area he/she is located. The red traces correspond to the wheelchair participant, while the yellow to the crutches participant.

4.1.2 Including personal preferences

Personal preferences depending on age and disability were used to shape the prior belief of a person using one transportation mode over another. This prior belief was expressed by drawing samples from an asymmetric Dirichlet prior distribution during inference. The choice of a Dirichlet distribution prior is a natural choice for this problem given the fact that it is the conjugate prior of the categorical distribution of transportation states. This section describes the approach for determining the concentration parameters of the Dirichlet distribution.

Table 3: p-values of chi-squared test between transportation modes and sociodemographic variables using the LTDS dataset.

	Bus	Walk	Rail	Disability	Age	Income
Bus	NA					
Walk	2.13e-219	NA				
Rail	2.55e-242	2.13e-219	NA			
Disability	3.14e-13	0	6.39e-210	NA		
Age	2.21e-52	4.70e-117	5.049e-312	2.93e-284	NA	
Income	9.73e-121	0.1080	6.16e-158	6.63e-125	4.07e-167	NA
Sex	9.64e-11	0.1962	1.06e-29	0.0034	0.2626	9.97e-15

The vector of values of the concentration parameters was used to control the level of prior belief in the preferences of an individual towards the transportation modes. Smaller ($0 < \alpha < 1$) values of α express less uncertainty in the preference of a transportation mode over the other. On the other hand, bigger values ($\alpha > 1$) express more uncertainty on the preference of an individual for a transportation mode. A concentration parameter vector with unit values would represent complete ignorance over the preferences of a user, or a user with no particular preferences. Most commonly, values between 1-5 are used if the concentration parameters are assumed to be pre-set, or they can be assigned a prior distribution, most commonly a Gamma distribution (Congdon 2003).

In this study, the calculation of different concentration parameter priors was based on the London Travel Demand Survey (LTDS) data-set. LTDS is a questionnaire survey combining socio-demographic data with information on the travel habits of participants with special focus on public transportation. The data-set provides very important insights between mobility and socio-demographic characteristics for marginalised population groups, however, it lacks specific location information on the daily mobility habits of individuals that could be used to better inform transportation planning. Examining the pairwise differences between the frequencies of journeys using different transportation mode and variables such as age, income, sex and disability using a Pearson's chi-squared test for the LTDS data (Table 3), one could see that for most variables there is a significant difference between the frequency of transportation modal use and socio-demographic characteristics. This suggests an overall strength of association between these variables. Exceptions are the variables income and sex in relation with walking.

The overall workflow of concentration parameter calculation is shown schematically in Figure 8.

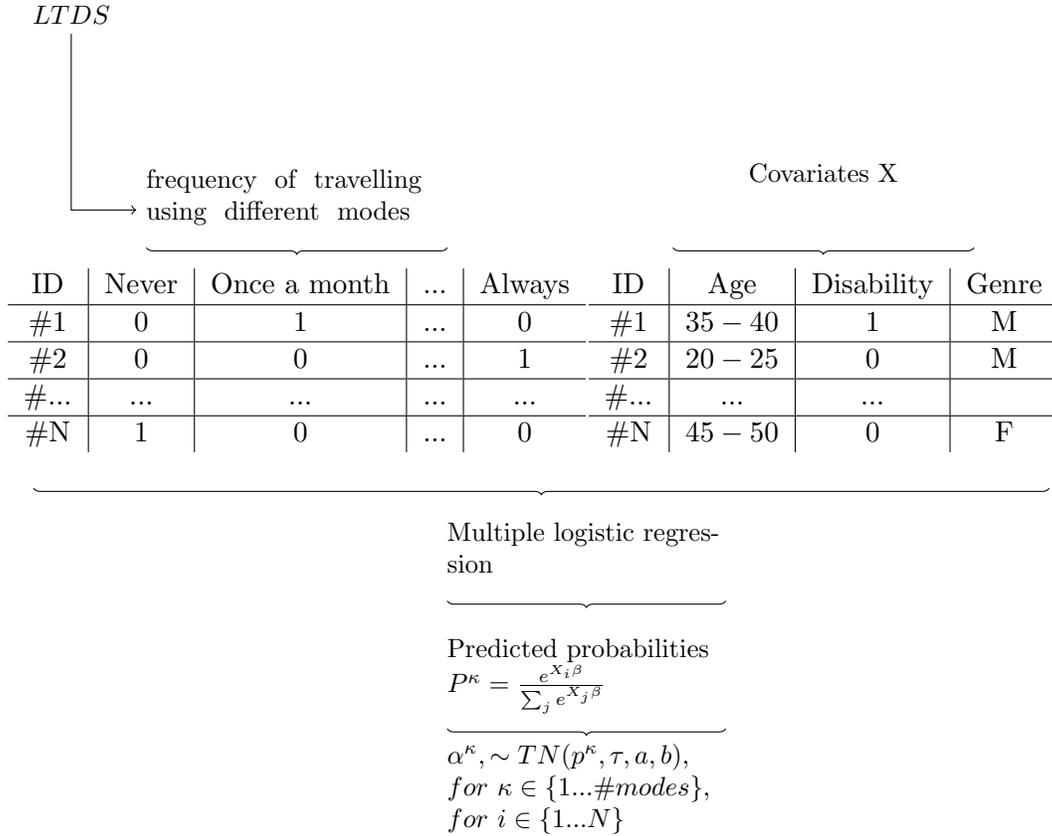
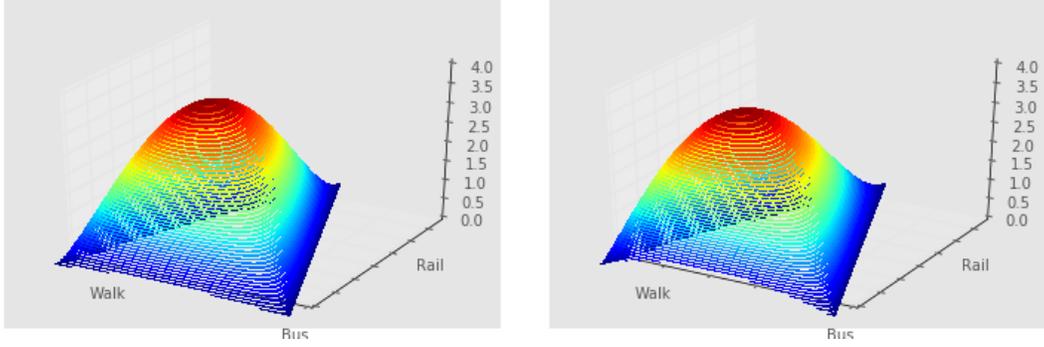


Figure 8: Concentration parameter calculation work-flow using LTDS data

The participants responses from LTDS datasets to walking, using the bus and rail transportation modes were dummy coded into multiple binary variables based on the frequency of use. The breakpoint condition in the coding procedure was the use of the respective transportation mode for more than once per month. The resulting data were then used in a multiple logistic regression model with independent variables being age and binary coded disability status and sex of the individuals. The predicted probabilities of using each transportation mode were then calculated using the actual age and disability status for each of the two participants in this study. The resulting values were then used as means in truncated normal distributions before including them as concentration parameters in the calculation of the Dirichlet prior. This was to allow for increased uncertainty between different transportation modes while ensuring that the values drawn were all positive. The stationary state was given a value of 1 for all participants reflecting lack of knowledge for this specific state.

The benefits of the above procedure are two-fold. First, by injecting prior knowledge in the model, the inference procedure becomes more robust as the posterior is weighted away from unlikely values as determined by past studies. Second, this procedure allows the determination of the extent of influence of socio-demographic characteristics shared amongst population groups when assessed at the individual level.

The Figure 9 below shows the resulting Dirichlet distributions for the two participants in the study. In this figure, each corner of the triangle corresponds to a potential transportation mode, while the z axis corresponds to the Dirichlet probability mass.



(a) Male aged between 20-39, disabled, $\alpha = [1.93, 1.75, 1.64]$ (b) Female aged between 40-59, disabled, $\alpha = [1.91, 1.75, 1.44]$

Figure 9: Dirichlet distribution results

More formally, the final model was:

$$\begin{aligned}
 P(s_t^\kappa | s_{t-1}^\kappa) &\sim \text{Cat}(\mathbf{p}), \\
 P(m^\kappa) &\sim N(0, 10^{-3}), \\
 \eta_t &= \boldsymbol{\beta}' \mathbf{X}, \\
 P(\mu^\kappa) &= \eta_t * m^\kappa, \\
 P(v_t^\kappa | s_t^\kappa) &\sim N(\mu^\kappa, \tau^\kappa) \text{ where } \mu^1 < \mu^2 < \mu^3 < \mu^4
 \end{aligned} \tag{2}$$

The hyperpriors in this model were:

$$\begin{aligned}
 P(\mathbf{p}) &\sim \text{Dir}(\boldsymbol{\alpha}), \\
 P(\alpha^\kappa) &\sim \text{TrN}(a^\kappa, 0.01, 0.1 < \text{bound} < +\infty), \\
 P(\beta_{1.. \#covariates}) &\sim N(0, 10), \\
 P(\tau^\kappa) &\sim \text{Gamma}(0.001, 0.001)
 \end{aligned} \tag{3}$$

where $\boldsymbol{\beta}$ is a vector of regression coefficients which is assumed to be distributed as a normal distribution with mean 0 and precision 10^{-5} , and \mathbf{X} is the matrix of external covariates.

5 Results

This section provides the results of travel mode detection using the specification described in the previous sections. All inferences were carried out within a Bayesian framework, using Markov Chain Monte Carlo (MCMC) methods. In particular, different MCMC sampling schemes were employed for the different stochastic nodes of the Bayesian network. Specifically, for the categorical nodes ($P(s_t^\kappa | s_{t-1}^\kappa)$) a discrete Metropolis sampling scheme, and for the continuous nodes ($P(m^\kappa), P(\mu^\kappa), P(\mathbf{p}), P(\alpha^\kappa), P(\beta_{1.. \#covariates}), P(\tau^\kappa)$) a combination of Metropolis, Adaptive Metropolis and Gibbs (Hit and Run sampler) (Brooks et al. 2011).

For all models, 5×10^5 iterations were used to approximate the unknown parameters. Convergence was assessed using visual methods and Geweke's convergence diagnostic. The first involves inspecting the MCMC chains for non-stationarity while the second employs

a Z-score test for significant differences between the first 10% and the last 50% of the Markov chain (Geweke et al. 1991) (Figure 10). The tests indicate that convergence has been achieved although additional samples would have improved the characterisation of posterior quantities, especially for the crutches user. This is particularly evident for the crutches user where the mixing of the samples was slower compared to the wheelchair user. Increasing the standard deviation for the latter to allow for more values to be rejected could improve mixing in the latter case. The acceptance ratio for all parameters for both users was found to be satisfactory, within the range of 0.2-0.245. In Figure 11 below, the convergence diagnostics are shown for the v node of Figure 6.

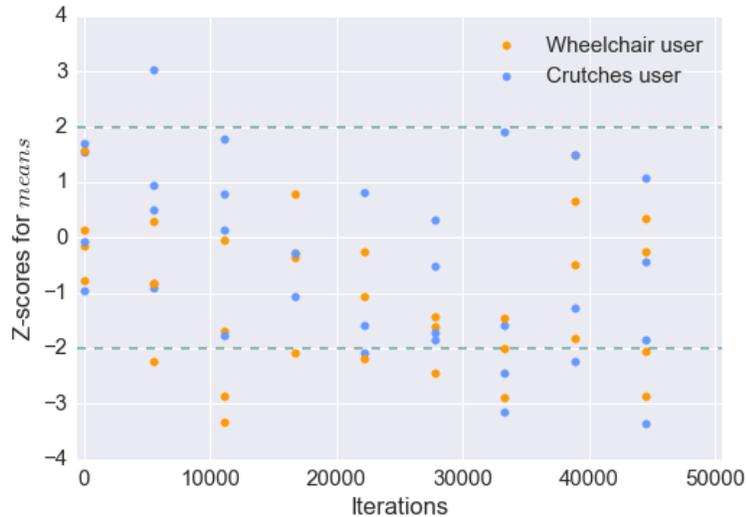


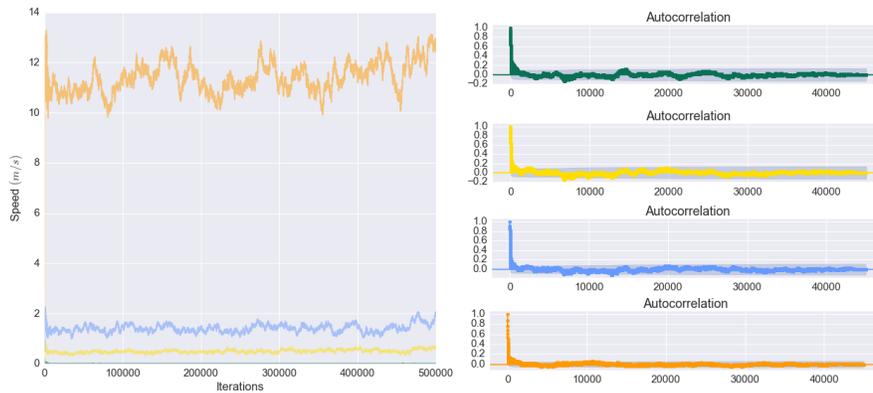
Figure 10: Geweke Z-scores for the speed nodes. The majority of the samples are within two standard deviations from the mean of the first and the last segment of the MCMC chain.

The posterior distributions of the inferred speeds for the two participants are shown in Figure 12 below. As it can be seen, the inferred speed is considerably different, especially for the walking mode. This is to be expected considering the fact that the two participants use different mobility aids when travelling without using car/public transport.

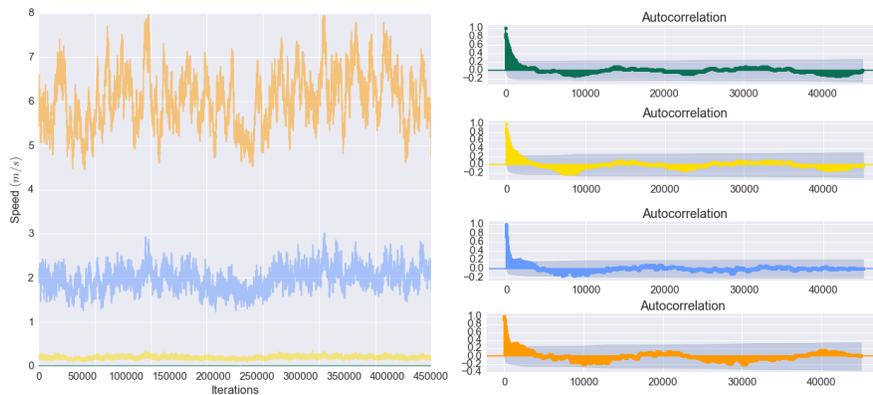
The dynamics of the participants' interaction with the different transportation modes was captured in the transition matrix. This is a stochastic matrix with each row representing a categorical distribution of switching between modes. This corresponds to the transition probability $P(s_t^k | s_{t-1}^k)$ of the model. Figure 13 below shows the posterior quantities of the transitions probabilities between different transportation modes.

For both participants, the effect of external factors on their movements was found to be either very small or statistically non-significant, given that the zero value is contained within the 95% credible intervals, and this was true for different standard deviations of β . The exception was the wheelchair user, where the IMD had a positive effect of 0.35 (Figure 14). These correspond to the β nodes of the model in Figure 6.

Looking at the internal effects as expressed by the concentration parameters of the Dirichlet distribution, (the α node of Figure 6), one observes that the values have shrunk towards values less than one, concentrating the Dirichlet distribution towards each individual categorical node and reflecting increased certainty of preferences of one transportation mode over the other (Figure 15).

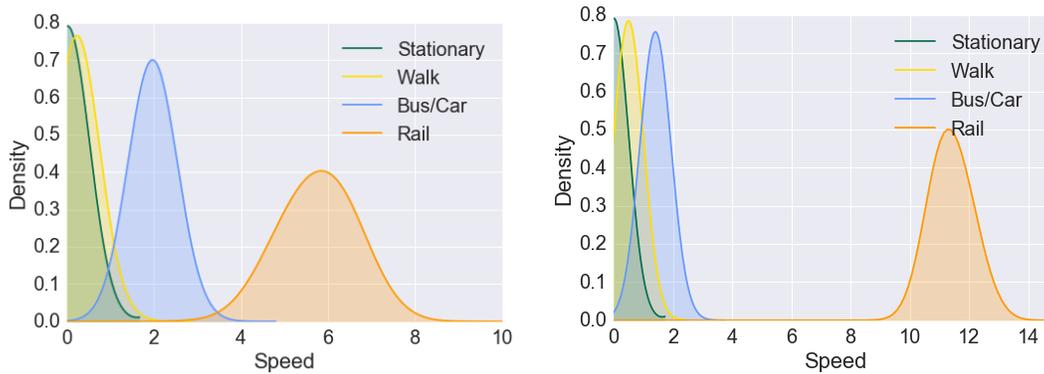


(a) MCMC traces and autocorrelation plots for the wheelchair user



(b) MCMC traces and autocorrelation plots for the crutches user

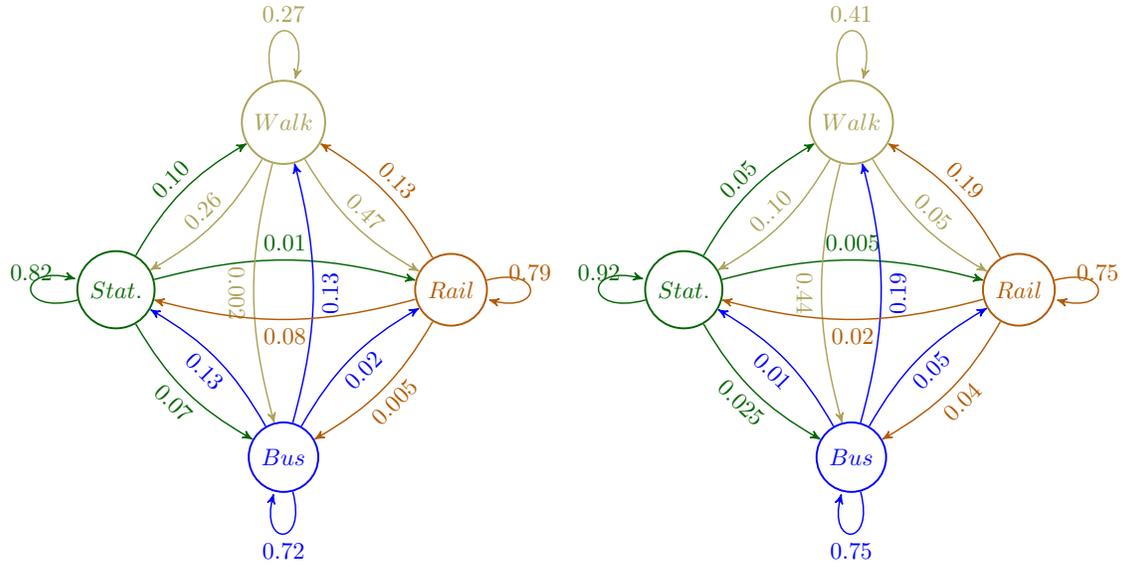
Figure 11: MCMC traces and autocorrelation plots for the two participants. The slow mixing of the crutches user can be seen from the tendency of the MCMC chain to make small jumps when proposing new speed values.



(a) Posterior speeds for the crutches user

(b) Posterior speeds for the wheelchair user

Figure 12: Posterior quantities of the speed node (v). The histograms are normalised. The unit of measurement is m/s.

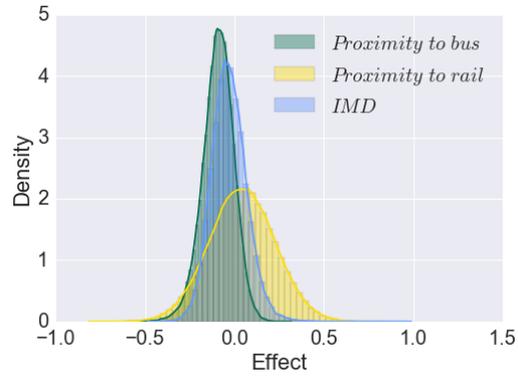


(a) Posterior mean transition probabilities for the crutches user

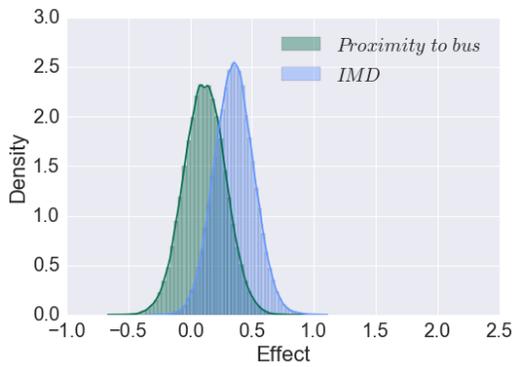
(b) Posterior mean transition probabilities for the wheelchair user

Figure 13: Transition probabilities for the two participants. The color coding corresponds to each row of the transition matrix with values as indicated from the corresponding arrows.

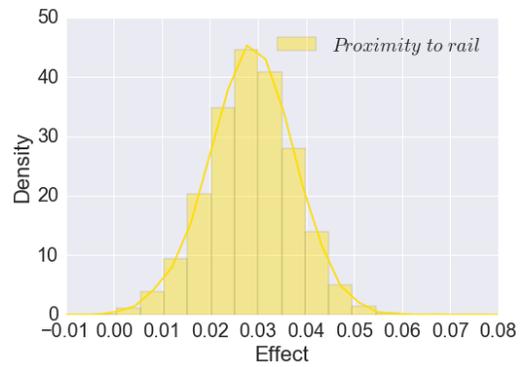
For the s node in Figure 6 model the posterior classification travel mode detection accuracy was assessed using the participant's own travel mode labelling. Figure 16 below shows the posterior median quantities for all data points, categorised by day.



(a) External factors effect for the crutches user

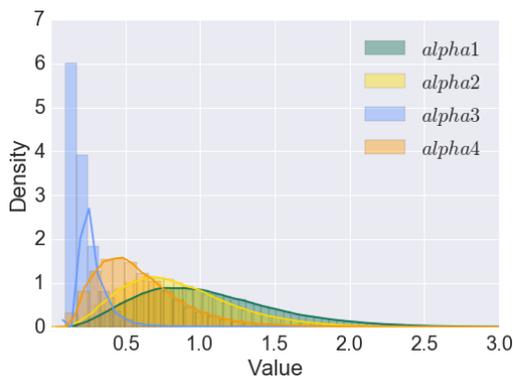


(b) External factors effect for the wheelchair user

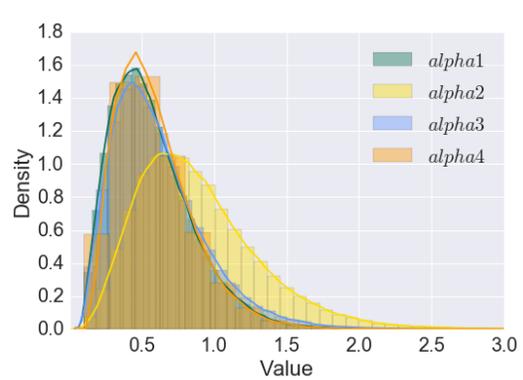


(c) External factors effect for the wheelchair user

Figure 14: Posterior quantities for the IMD, proximity to bus stops and proximity to rail stations.



(a) Concentration parameters for the crutches user



(b) Concentration parameters for the wheelchair user

Figure 15: Posterior quantities for the truncated normal priors of the Dirichlet concentration parameters. The color coding represents the corresponding categorical distributions and is the same as Figure 13.

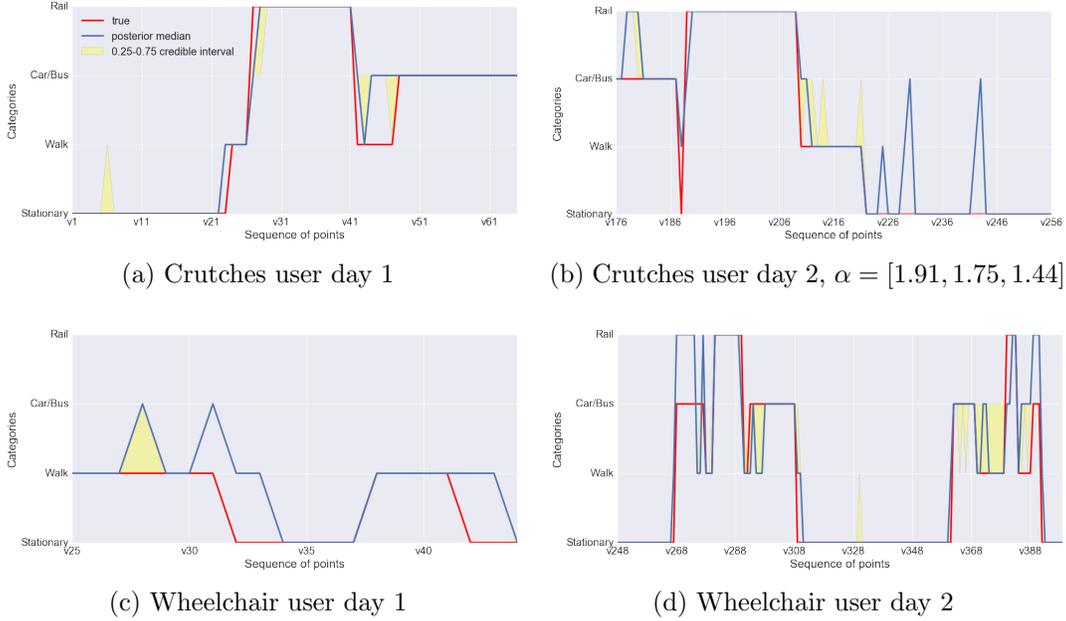


Figure 16: Posterior medians of the categorical node (s) of the model. The red line correspond to the self-labeled value while the blue line corresponds to the inferred quantities. The yellow faces are the 95% credible intervals of the MCMC simulation.

5.1 Performance evaluation

The performance of the proposed method was compared to other popular classification algorithms, namely Random Forests (RF), Support Vector Machines (SVM's) and Multilayer (MPL). It is important to notice that, contrary to the aforementioned classifiers, the proposed method is essentially an unsupervised classification procedure, and as such a training step is not needed. For this task, 60% of the data-sets were used during the training procedure, and 40% for testing. The benchmark for comparison was the participants self-labeled true states.

The overall classification accuracy using the proposed method is illustrated in the confusion matrix (Figure 17). It can be seen that mis-classification mostly occurred between the walk and the bus travel modes. This can be explained by considering the lower accuracy of data generated by mobile phone API's, together with the low mean speed of travel for buses in peak hours in London, which could be as low as 6km/h (Transport for London — Every Journey Matters 2017). The same holds true for mis-classification artifacts between walking and stationary states, which could be attributed to the effects of “sudden jumps” and “drift” in location. The overall accuracy was 71% for the wheelchair user and 78% for the crutches user.

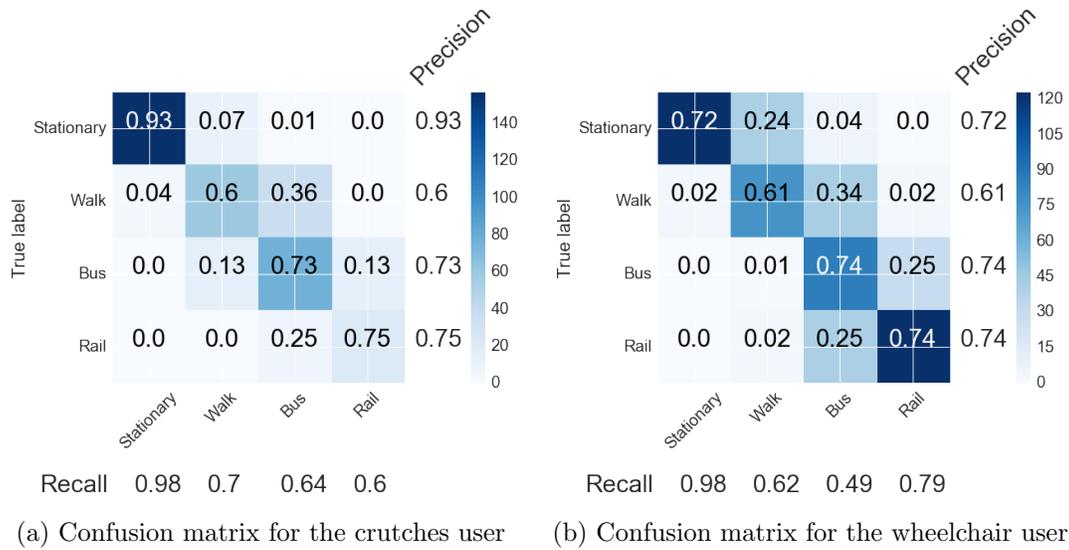


Figure 17: Confusion matrices between the self labeled data and the inferred transportation modes for the proposed method. The color-bar corresponds to the number of data points.

Next, a RF classifier was employed. Maximum accuracy was achieved for training 10 trees in the forest. No restrictions were placed for the maximum number of features for each individual tree. Looking at the results of the RF classifier, miss-classification is more profound for the walking mode. This is especially true for the wheelchair user data-set which proved to be challenging in terms of classification performance. The overall accuracy was 79% for the crutches user and 67% for the wheelchair user.

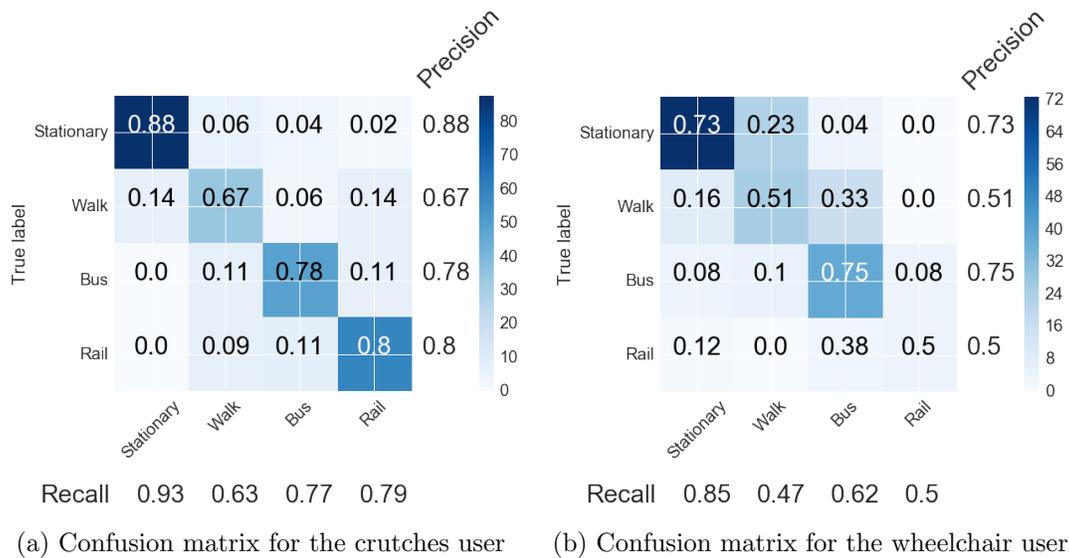


Figure 18: Confusion matrices between the self labeled data and the inferred transportation modes for the RF classifier.

Next, a SVM classifier was employed with an exponential kernel. Maximum accuracy was achieved with a penalty parameter of 1 and a γ value of $\frac{1}{\#features}$. The algorithm was found to perform comparatively well to both RF and the proposed method. The accuracy for the wheelchair user was 62% while for the crutches user was 71%.

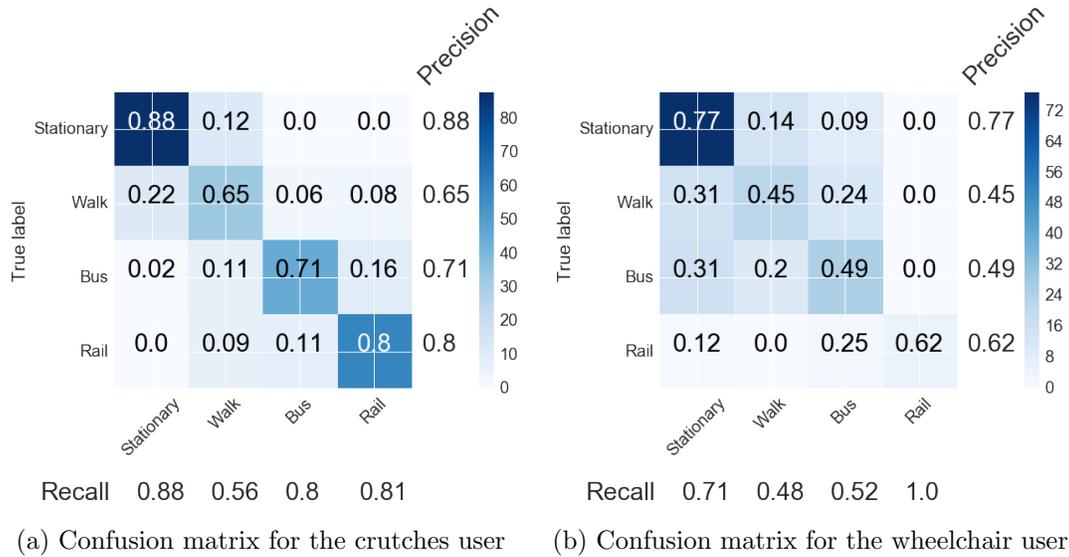


Figure 19: Confusion matrices between the self labeled data and the inferred transportation modes for the SVM classifier.

Finally, a Multilayer Perceptron (ANN) was employed using backpropagation for the training procedure. The total number of hidden layers that yielded maximum accuracy was 15. The algorithm performed comparably to both the proposed method and RF and overperformed the SVM. For this classifier the accuracy for the wheelchair user was 69% while for the crutches user was 70%.

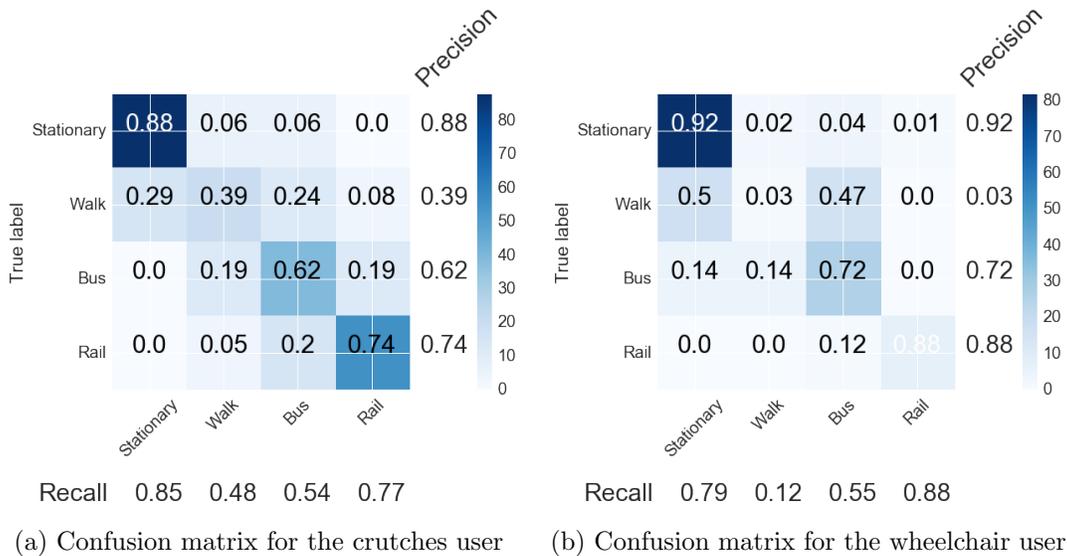


Figure 20: Confusion matrices between the self labeled data and the inferred transportation modes for the MLP classifier.

To assess whether the classification results between the different methods were statistically significant, a chi-squared test was carried out between the classification results of the proposed method and the results of RF, SVM and MLP classifiers. The null hypothesis is that the difference in the classification results could be generated by chance alone. Looking at the p-values, the proposed method produced statistically significant results for all of the transportation modes for the wheelchair user and for nearly half the transportation modes for the crutches user. The corresponding chi-squared statistics and p-values are presented in Tables 4 and 5 below:

Table 4: Chi-squared statistic and p-values for the crutches user.

Mode	Statistic	RF	SVM	MLP
Stationary	chi-sq	7.4976	2.9728	7.2529
	p-value	0.0576	0.2261	0.0266
Walk	chi-sq	29.6063	29.6905	28.929
	p-value	1.67E-06	1.60E-06	2.32E-06
Bus	chi-sq	0.4605	1.9966	2.28382
	p-value	0.7943	0.57310	0.31920
Rail	chi-sq	5.48766	5.48766	1.72545
	p-value	0.06432	0.0643	0.4220

Table 5: Chi-squared statistic and p-values for the wheelchair user.

Mode	Statistic	RF	SVM	MLP
Stationary	chi-sq	10.2118	6.7393	21.7692
	p-value	0.0060	0.0344	7.29E-05
Walk	chi-sq	30.2307	30.4187	67.8636
	p-value	1.23E-06	1.13E-06	1.22E-14
Bus	chi-sq	13.0388	71.4039	36.9357
	p-value	0.0045	2.14E-15	4.75E-08
Rail	chi-sq	12.2690	21.1491	0.8027
	p-value	0.0021	9.80E-05	0.6694

Finally, to assess the generalisation of the model to different datasets, the proposed method was employed to a GPS dataset of 5 individuals. The temporal resolution of this dataset was 60 seconds of each subsequent GPS point, while the spatial accuracy according to the horizontal dilution of precision value was within the range of ≤ 2.5 which translates to a good accuracy level for most applications (Langley et al. 1999). The temporal domain of this dataset spanned for over a week for all participants. Since information about the individual socio-demographic characteristics of participants was unavailable for this sample, an uninformative Dirichlet prior distribution was used for modelling the effect of personal preferences. The performance of classifiers was tested against the individuals self-labelled data. The hyper parameters of SVM, RF and MLP were tuned for different values using the exhaustive grid search method. The results are presented in the table below along with the corresponding results of SVM, RF and MLP classifiers. As it can be seen, the proposed method performed comparably when compared to SVM, RF and MLP classifiers.

Table 6: Performance evaluation using a GPS sample of 5 individuals.

Classifier	Stationary	Walk	Bus/Car	Rail	Recall
Pr. method	0.94	0.05	0.0	0.0	0.94
	0.28	0.64	0.03	0.05	0.65
	0.03	0.04	0.92	0.01	0.92
	0.02	0.04	0.01	0.94	0.94
<i>Precision</i>	0.94	0.65	0.88	0.91	<i>Acc: 0.90</i>
RF	0.91	0.05	0.03	0.01	0.91
	0.36	0.49	0.12	0.02	0.49
	0.18	0.08	0.65	0.08	0.65
	0.07	0.05	0.16	0.72	0.72
<i>Precision</i>	0.92	0.53	0.44	0.84	<i>Acc: 0.82</i>
SVM	0.89	0.09	0.02	0.01	0.89
	0.29	0.61	0.03	0.08	0.61
	0.04	0.08	0.77	0.11	0.77
	0.09	0.02	0.09	0.81	0.81
<i>Precision</i>	0.95	0.43	0.53	0.79	<i>Acc: 0.85</i>
MLP	0.91	0.05	0.03	0.01	0.91
	0.36	0.49	0.12	0.02	0.49
	0.18	0.08	0.65	0.08	0.65
	0.07	0.05	0.16	0.72	0.72
<i>Precision</i>	0.92	0.53	0.44	0.84	<i>Acc: 0.82</i>

6 Discussion

The posterior quantities for the speed of the two participants in the study were found to be different when compared with each other. This can be attributed to numerous reasons. As already mentioned, the mobility aids each participant is using are different, influencing the walking speed in different ways. In addition, each participant is using different rail transportation modes, namely the London Overground rail service for the crutches user and the National rail services for the wheelchair user. These two services have very distinct speed signatures, the first one being a city wide transportation mode with more frequent stops while the latter being used for intercity travels with fewer intermediate stops. Looking at the state transition probabilities, a detailed insight on the way the participants use the different transportation modes can be made. The high probabilities for staying at each node, are the result of the coordinate by coordinate classification process. The relatively low transition probabilities of the wheelchair user reflect the fact that this individual uses the public transportation fewer times in the weekly sample of the analysis. This is contrasted with the crutches user that interacts with the public transportation network in a more regular way. With regards to the overall strength of the travel mode preferences, the crutches user appears to have a preference over traveling by bus. This is due to the relatively small posterior values of the concentration parameter of the Dirichlet prior for the bus travel mode. This fact is in line with the initial assumptions as shaped by the LTDS dataset. However, this is not the case for the wheelchair user where a mix of modes is used. This is due to the overlap of the posterior quantities of the concentration parameters.

Posterior inferences of external covariates has shown that their influence on the travel mode classification process varies between the two participants. In particular, proximity to available transportation modes and IMD had a reduced effect on participants' interaction with the transportation modes, the magnitude of which is different between them. This magnitude varied between being statistically non significant and having a significant, but small effect. The former was the case for proximity to available transportation means, while the latter was relevant for IMD. This could reflect the fact that the different use of transportation modes, as expressed by the different speed values, is not influenced significantly by the chosen covariates. However, given the low accuracy of traces, this is yet to be verified by the use of more accurate position data.

7 Conclusions and future research

This study proposes a framework for modelling human mobility patterns at the individual level, with particular focus on public transport. Such a framework has the potential to capture the subtleties of the overall interaction of individuals with the available transportation modes more comprehensively than existing studies that use GPS tracking and GIS data alone, even in the face of low accuracy data. This is achieved by including secondary data in the analysis in the form of environmental covariates as well as past transportation demand studies.

A hierarchical dynamic Bayesian network was used to build the model. Such modelling structure offers the advantage of providing information about the degree of interaction of individuals with the available transportation modes, together with the extensibility required to include a wide range of variables influencing it. Judging from the posterior densities of the concentration parameters, internal factors inherent to a person's capabilities, as expressed

by the shape of the Dirichlet prior, play an important role in the interaction with the different transportation modes. On the other hand, external covariates have a reduced effect on the inferred modes. The proposed approach is also able to characterise the transition dynamics of an individual through the use of the transition matrix. Such information can provide valuable insights in the preferences of the user, disaggregated by their individual capabilities.

Approaching mobility modelling using this bottom-up framework offers the potential of several advantages if transferred to a larger scale mobility modelling. By including sources of secondary socio-demographic information, it is possible to formulate assumptions about individual population groups from untagged mobility data. Although such information may not be available beforehand at such an individual level, one could use existing aggregated reports on mobility preferences of population groups and incorporate them in the analysis through the prior distribution of the concentration parameters. The level of correspondence between the prior information and the posterior quantities can then be used to evaluate the original assumptions. Similar benefits can be gained by using the posterior quantities of the transition matrix. Particularly, characteristic mobility patterns can be extracted using the transition probabilities of interacting with different transportation modes given the capabilities of groups of individuals. These patterns can then be used to model similar population groups within an agent based model.

Finally, adopting a more disaggregated approach to modelling mobility patterns would inform policy makers into actions targeted at marginalised population groups, allowing for a more complete formulation of mobility strategies. This is of particular importance when formulating contingency plans during mobility disruptions.

Future research involves scaling the analysis to a larger participant sample size as well as using its output to reason behind peoples transportation patterns. As an example, a sensitivity analysis on the overall effect of the internal factors prior distribution could be used to assess the effect of a variable to the shape of the posterior distribution of the transportation mode categories. Moreover, it will experiment using the transition matrix to define new population classes based on similarity characteristics between its row categorical distributions.

References

- Bolbol, A., Cheng, T., Tsapakis, I. & Haworth, J. (2012), 'Inferring hybrid transportation modes from sparse gps data using a moving window svm classification', *Computers, Environment and Urban Systems* **36**(6), 526–537.
- Brooks, S., Gelman, A., Jones, G. L. & Meng, X.-L. (2011), *Handbook of markov chain monte carlo*, Chapman and Hall/CRC.
- Congdon, P. (2003), *Applied bayesian modelling*, John Wiley & Sons.
- Congdon, P. D. (2010), *Applied Bayesian hierarchical methods*, CRC Press.
- Eftekhari, H. R. & Ghatee, M. (2016), 'An inference engine for smartphones to preprocess data and detect stationary and transportation modes', *Transportation Research Part C: Emerging Technologies* **69**, 313–327.
- Evans, G. (2009), 'Accessibility, urban design and the whole journey environment', *Built environment* **35**(3), 366–385.

- Feng, T. & Timmermans, H. J. (2016), 'Comparison of advanced imputation algorithms for detection of transportation mode and activity episode using gps data', *Transportation Planning and Technology* pp. 1–15.
- Geweke, J. et al. (1991), *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, Vol. 196, Federal Reserve Bank of Minneapolis, Research Department.
- Griffin, T. & Huang, Y. (2005), A decision tree classification model to automate trip purpose derivation, in 'Proceedings of the ISCA 18th international conference on computer applications in industry and engineering, Sheraton Moana Surfrider, Honolulu, HI', Citeseer, pp. 44–49.
- Jahangiri, A. & Rakha, H. A. (2015), 'Applying machine learning techniques to transportation mode recognition using mobile phone sensor data', *Intelligent Transportation Systems, IEEE Transactions on* **16**(5), 2406–2417.
- Janssens, D., Wets, G., Brijs, T., Vanhoof, K., Arentze, T. & Timmermans, H. (2006), 'Integrating bayesian networks and decision trees in a sequential rule-based transportation model', *European Journal of operational research* **175**(1), 16–34.
- Kim, Y., Pereira, F. C., Zhao, F., Ghorpade, A., Zegras, P. C. & Ben-Akiva, M. (2014), Activity recognition for a smartphone based travel survey based on cross-user history data, in 'Pattern Recognition (ICPR), 2014 22nd International Conference on', IEEE, pp. 432–437.
- Koller, D. & Friedman, N. (2009), *Probabilistic graphical models: principles and techniques*, MIT press.
- Langley, R. B. et al. (1999), 'Dilution of precision', *GPS world* **10**(5), 52–59.
- Liao, L., Patterson, D. J., Fox, D. & Kautz, H. (2007), 'Learning and inferring transportation routines', *Artificial Intelligence* **171**(5), 311–331.
- Lin, M. & Hsu, W.-J. (2014), 'Mining gps data for mobility patterns: A survey', *Pervasive and Mobile Computing* **12**, 1–16.
- McGowen, P. & McNally, M. (2007), Evaluating the potential to predict activity types from gps and gis data, in 'Transportation Research Board 86th Annual Meeting, Washington', Citeseer.
- Montini, L., Prost, S., Schrammel, J., Rieser-Schüssler, N. & Axhausen, K. W. (2015), 'Comparison of travel diaries generated from smartphone data and dedicated gps devices', *Transportation Research Procedia* **11**, 227–241.
- Omrani, H. (2015), 'Predicting travel mode of individuals by machine learning', *Transportation Research Procedia* **10**, 840–849.
- Patterson, D. J., Liao, L., Fox, D. & Kautz, H. (2003), Inferring high-level behavior from low-level sensors, in 'UbiComp 2003: Ubiquitous Computing', Springer, pp. 73–89.
- Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M. & Srivastava, M. (2010), 'Using mobile phones to determine transportation modes', *ACM Transactions on Sensor Networks (TOSN)* **6**(2), 13.

- Shafique, M. A. & Hato, E. (2015), 'Use of acceleration data for transportation mode prediction', *Transportation* **42**(1), 163–188.
- Stenneth, L., Wolfson, O., Yu, P. S. & Xu, B. (2011), Transportation mode detection using mobile phones and gis information, in 'Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems', ACM, pp. 54–63.
- Transport for London (2012), 'Understanding the travel needs of london's diverse communities: Disabled people'. Online; Accessed 01/08/2014.
URL: <https://www.tfl.gov.uk/cdn/static/cms/documents/understanding-the-travel-needs-of-london-diverse-communities.pdf>
- Tyler, N. (2006), 'Capabilities and radicalism: Engineering accessibility in the 21st century', *Transportation planning and technology* **29**(5), 331–358.
- UK Government (2015), 'English indices of deprivation 2015'. Online; Accessed 24/09/2016.
URL: <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015>
- Wan, N. & Lin, G. (2016), 'Classifying human activity patterns from smartphone collected gps data: A fuzzy classification and aggregation approach', *Transactions in GIS* .
- Widhalm, P., Nitsche, P. & Brändie, N. (2012), Transport mode detection with realistic smartphone sensor data, in 'Pattern Recognition (ICPR), 2012 21st International Conference on', IEEE, pp. 573–576.
- Wu, L., Yang, B. & Jing, P. (2016), 'Travel mode detection based on gps raw data collected by smartphones: a systematic review of the existing methodologies', *Information* **7**(4), 67.
- Xiao, G., Juan, Z. & Zhang, C. (2015), 'Travel mode detection based on gps track data and bayesian networks', *Computers, Environment and Urban Systems* **54**, 14–22.
- Xie, K., Xiong, H. & Li, C. (2016), The correlation between human mobility and socio-demographic in megacity, in 'Smart Cities Conference (ISC2), 2016 IEEE International', IEEE, pp. 1–6.
- Zhang, L., Liu, L., Bao, S., Qiang, M. & Zou, X. (2015), 'Transportation mode detection based on permutation entropy and extreme learning machine', *Mathematical Problems in Engineering* **2015**.
- Zheng, Y., Li, Q., Chen, Y., Xie, X. & Ma, W.-Y. (2008), Understanding mobility based on gps data, in 'Proceedings of the 10th international conference on Ubiquitous computing', ACM, pp. 312–321.
- Zheng, Y., Liu, L., Wang, L. & Xie, X. (2008), Learning transportation mode from raw gps data for geographic applications on the web, in 'Proceedings of the 17th international conference on World Wide Web', ACM, pp. 247–256.
- Zucchini, W. & MacDonald, I. L. (2009), *Hidden Markov models for time series: an introduction using R*, CRC Press.