Transportation Research Part C 80 (2017) 286–309

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

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

Thanos Bantis *, James Haworth

University College London, Department of Civil Environmental and Geomatic Engineering, SpaceTime Lab, United Kingdom

ARTICLE INFO

Article history:
Received 19 December 2016
Received in revised form 6 May 2017
Accepted 6 May 2017

Keywords:
Transportation mode detection
Dynamic Bayesian networks
Mobility
Disabilities
Smartphones

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.

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

* Corresponding author.
E-mail address: thanos.bantis.13@ucl.ac.uk (T. Bantis).

http://dx.doi.org/10.1016/j.trc.2017.05.003

such as speed, acceleration, heading and difference in distance to classify raw GPS traces into transportation modes using a Support Vector Machine (SVM). Zheng et al. (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 and 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 and 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 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 classier 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 and 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 an global solution to the classification problem, however, depending on the kernel specification, this is not the same for the problem of over-fitting.