Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 10 (2015) 236 – 244

Transportation Research Procedia

www.elsevier.com/locate/procedia

18th Euro Working Group on Transportation, EWGT 2015, 14-16 July 2015, Delft, The Netherlands

Modelling of accelerometer data for travel mode detection by hierarchical application of binomial logistic regression

Muhammad Awais Shafique [a,b,*], Eiji Hato [b]

[a] *University of Engineering and Technology, GT road Lahore 54890, Pakistan*
[b] *The University of Tokyo, 3-1-7, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

Abstract

Household trip data collection is essential for design and construction of transportation infrastructure. Conventionally, this information is collected by travel surveys, which require the respondents to answer a list of questions targeting their daily travelling. As the responses depend on the memory of the respondents, inaccuracies usually occur during the reporting process. To improve the accuracy of the collected data, a lot of research is currently being focused on inferring the important information from data collected automatically with the help of devices like smartphones. The current study proposes a new method for identifying the travel mode, by applying the binomial logistic regression in a hierarchical manner, using the data collected by the accelerometer of the smartphone. Three methods of application are discussed, namely ranking, one against rest and one against all. Apart from train, all the other modes are successfully modelled with goodness of fit approaching to 1. Low goodness of fit in case of train is due to the wide range of accelerations recorded. Although, all the three methods exhibit good outcomes, one against all method provides relatively better results.

Peer-review under responsibility of Delft University of Technology

Keywords: Acceleration; Binomial Logistic Regression; Smartphone; Travel Mode

1. Introduction

Household travel data collection is an important and integral part of transportation infrastructure planning and management. Currently, two methods are used for this purpose, the conventional method and the recent method. The conventional method is to collect the data with the help of surveys. These can be paper questionnaires, mail-back

* Corresponding author. Tel.: +81-03-5841-6118; fax: +81-03-5841-8507.
E-mail address: awais@trip.t.u-tokyo.ac.jp

Peer-review under responsibility of Delft University of Technology
doi:10.1016/j.trpro.2015.09.073

surveys, telephone surveys and internet surveys. They all require the respondent to recall the daily travelling and then provide that information to the surveyor. This reliance on the memory of the respondent causes a lot of drawbacks to the collection methodology. The participants are usually unable to provide accurate starting and ending times of the trips they made during that day, or worst, the previous day. They also tend to forget reporting short trips. This can lead to erroneous data being collected. Moreover, these surveys have low response rate as the respondents are expected to answer a huge number of questions, which is a tedious job.

The recent method of data collection is to address the issues of conventional method by eliminating the reliance on the memory of the respondents. This can be achieved by automatically collecting the travel data, without any input from the participants. For this purpose, a lot of research is focused on the utilization of sensors. Experiments have been conducted using Global System for Mobile (GSM) communications (Anderson and Muller, 2006; Sohn et al., 2006), local area wireless technology (Wi-Fi) (Mun et al., 2008), Global Positioning System (GPS) (Tsui and Shalaby, 2006), accelerometer (Shafique and Hato, 2015) and smartphones (Nham et al., 2008; Reddy et al., 2010). The data from the sensors can be employed to identify the mode of transportation used by the respondent. So without any input required on part of the respondents, their daily trips, which includes the origin, destination, starting time, ending time, mode used and route taken, can be recorded with high accuracy. Usually, a supervised learning algorithm is used to classify the data into various travel modes.

Accelerometer and GPS data are mostly used in the detection of travel mode. Accelerometer provides the accelerations along the 3 axes after pre-defined intervals whereas GPS records the location of the device at every instance. Both types of data are individually or collectively used to successfully classify the transportation modes (Gonzalez et al., 2008; Kwapisz et al., 2011; Nham et al., 2008; Nick et al., 2010; Reddy et al., 2008; Reddy et al., 2010; Shin et al., 2014; Xia et al., 2014; Yang, 2009).

This study is unique in sense that it explores the possibility of applying the binomial logistic model using the accelerometer data. Although, multinomial logit models have been adopted for assessing the effect of a policy change on mode shift or the possibilities of introducing a new travel mode, but using the logit model to identify the travel mode from only the data collected by accelerometer is a novel approach. The binomial logit model is applied in a hierarchical manner to separate six modes namely walk, bicycle, car, bus, train and subway.

2. Binomial Logistic Regression

Binary logistic regression is a type of generalized linear models (GLM), which models how a binary response is dependent on a set of explanatory variables. The explanatory variables can be discrete, continuous or a combination. Binary response means that there can be only two possible outcomes, either success or failure. For example, a doctor wants to figure out the proportion of breast cancer patients in a given population. Naturally, every person's risk of being a patient of breast cancer will vary, depending on a number of factors including age, lifestyle and eating habits. Consider these factors or predictor variables be represented by $X = (X_1, X_2, \dots X_k)$ with observed value $x_i = (x_{i1}, x_{i2}, \dots x_{ik})$ for a person i. Let Y be the binary response variable where $Y_i = 1$ if person i is a patient and $Y_i = 0$ if otherwise. The probability (π) that the person i is a patient can be formulated as follows

$$\pi_i = \Pr(Y_i = 1 | X_i = x_i) = \frac{\exp(\beta_0 + \beta_i x_i)}{1 + \exp(\beta_0 + \beta_i x_i)} \quad (1)$$

Or

$$\begin{aligned} logit(\pi_i) &= log\left(\frac{\pi_i}{1-\pi_i}\right) \\ &= \beta_0 + \beta_i x_i \\ &= \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} \end{aligned} \quad (2)$$