Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 22 (2017) 144–153

Transportation Research Procedia

www.elsevier.com/locate/procedia

19th EURO Working Group on Transportation Meeting, EWGT2016, 5-7 September 2016, Istanbul, Turkey

Classification of Travel Data with Multiple Sensor Information using Random Forest

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

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

Abstract

Recently, a lot of studies have been focused on the use of smartphones for automatic detection of transportation mode. This task is made easy by the availability of sensors like accelerometer and GPS in modern smartphones. The advantages include the increased accuracy which was partially lost due to underreporting in case of conventional travel surveys. In this paper Probe Person data collected by 46 participants in three different cities of Japan, namely Niigata, Gifu and Matsuyama, was used. Although the data, comprising of acceleration and GPS information, was collected by a wearable device but the same can be achieved very easily with the help of smartphones. In order to address the most important problem of continuously changing position of smartphone during the trip, only resultant acceleration was taken. In addition, personal characteristics like age and gender were also included. Regarding GIS information, distance and time calculated by Google Maps for both driving and walking was introduced to increase the prediction accuracy. Random Forest was applied for the purpose of prediction. 70% of the data was randomly selected to train the algorithm and rest 30% was used to test it. Prediction was done among four different modes; walk, bicycle, car and train. The results are quite promising with an overall prediction accuracy of more than 99.6% for all three cities. A slight improvement in the prediction accuracy is achieved by selecting the best features for the classification purpose.

Peer-review under responsibility of the Scientific Committee of EWGT2016.

Keywords: Accelerometer; GPS; Random Forest; Travel Mode Detection

* Corresponding author. Tel.: +92-321-4223020; fax: +81-03-5841-8507.
E-mail address: awais@bin.t.u-tokyo.ac.jp

2214-241X
Peer-review under responsibility of the Scientific Committee of EWGT2016.
10.1016/j.trpro.2017.03.021

1. Introduction

Accelerometer and GPS are two of the many sensors integrated in smartphones nowadays. This addition of technology has greatly enhanced the tasks performed by smartphones. In addition to conventional uses of mobile phones, the phone bearer can have many other advantages like real time navigation, location of nearby/desired facilities, real-time weather forecast and early disaster warning alerts. New mobile games are utilizing more and more sensors to improve the playing experience. Hence an inexhaustible amount of possibilities are present.

Researchers have recently shifted their focus on detecting the mode of transportation used by the carrier of a smartphone or any custom-build device. For this purpose, GPS and accelerometer data is widely used. The need for automatic mode detection is crucial as it can decrease the error generated by conventional travel surveys. Travel surveys, in the form of travel dairies, questionnaires or telephone interviews, are used worldwide to capture the travel patterns of the people. Hence a continuous effort is done to understand and model the travel behavior of people. Consequently, a town planner or infrastructure designer can come up with a better plan or design. But conventional travel surveys have a major drawback. Mostly the respondents are asked to report their daily travelling at the end of the day or after some days. As the report depends on the memory, the respondent may not be able to correctly remember the exact starting or ending time of a particular trip. Moreover, the respondent might forget to report short trips. In addition to that the respondent might just get fed up with the huge number of questions present in the questionnaire and tend to answer them hurriedly without completely understanding them or just skip some. All these scenarios exhibit sources of error, hence resulting in reduced accuracy of the data collected. Devices embedded with GPS and accelerometer sensors can prove to be a useful solution to all these problems.

Moreover, mode detection using smartphones (or any other device) can assist in customer oriented advertisement. For example, if a mobile user is travelling by any means of transportation then the nearby facilities related to the detected mode of transportation can be advertised directly to the user of that mode. These facilities can be businesses directly related to the transportation vehicle e.g. cheap shop for bicycle accessories or nearby car parking area or discount tickets for railways. Or can be indirect businesses targeting customers on account of their route and mode e.g. cafes, restaurants and recreational areas.

In this paper, we have proposed a methodology to efficiently utilize the data collected by various sensors and sources, to detect the mode of transportation used by the device carrier, with good accuracy.

2. Related Work

Zheng et al. (2008), used GPS data to determine the mode of transportation. The authors made a comparison among different classification algorithms including decision tree, Support Vector Machine, Bayesian net and Conditional Random Field. Accelerometer embedded in iPhone was used in a study for mode prediction (Nham et al. 2008). Only 3 participants were employed for the collection of data. 70% of the data was used to train Support Vector Machine and the remaining 30% data was used to test the algorithm. The study showed good results but the accuracy varied from 88% to 97% among the participants.

In 2007 a large scale study was conducted in Netherlands in which 1104 respondents participated in a one week survey (Bohte and Maat 2009). The approach combined GPS logs and Geographic Information System (GIS) technology along with an interactive web-based application for validation. The aim was to predict trip purposes and travel modes. The study showed that when compared with the Dutch Travel Survey, more trips per tour were recorded, which indicated the ability to include short trips usually missed by travel diary methods. Other studies (Forrest and Pearson 2005; Wolf et al. 2003) also found significant underreporting in case of phone or paper surveys

Using GPS and accelerometer data, Figo et al. (2010) compared the various pre-processing techniques used in many previous studies. Prediction was done for three activities i.e. walking, running and jumping. The results suggested that for two-activity scenario the time-domain techniques performed better but for three-activity scenario the frequency-domain techniques showed comparable accuracy. Another study utilized acceleration data to distinguish between 3 modes; walk, car and train (Nick et al. 2010). A comparison was made between Naive Bayes Classifier and Support Vector Machine. 90% of the collected data was used to train the algorithms while rest 10% was used to test them. Results suggested that Support Vector Machine performed better than Naive Bayes Classifier with a prediction accuracy of 97.32%.