Computers, Environment and Urban Systems 66 (2017) 38–52

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Spatial context mining approach for transport mode recognition from mobile sensed big data

Ivana Semanjski [a,*], Sidharta Gautama [a], Rein Ahas [b], Frank Witlox [c,b]

[a] Department of Telecommunications and Information Processing, Ghent University, Ghent, Belgium
[b] Department of Geography, University of Tartu, Tartu, Estonia
[c] Department of Geography, Ghent University, Ghent, Belgium

A R T I C L E   I N F O

Article history:
Received 21 December 2016
Received in revised form 24 July 2017
Accepted 26 July 2017
Available online xxxx

Keywords:
Transport mode recognition
Mobile sensed big data
Spatial awareness
Geographic information systems
Smart city
Support vector machines
Context mining
Urban data

A B S T R A C T

Knowledge about what transport mode people use is important information of any mobility or travel behaviour research. With ubiquitous presence of smartphones, and its sensing possibilities, new opportunities to infer transport mode from movement data are appearing. In this paper we investigate the role of spatial context of human movements in inferring transport mode from mobile sensed data. For this we use data collected from more than 8000 participants over a period of four months, in combination with freely available geographical information. We develop a support vectors machines-based model to infer five transport modes and achieve success rate of 94%. The developed model is applicable across different mobile sensed data, as it is independent on the integration of additional sensors in the device itself. Furthermore, suggested approach is robust, as it strongly relies on pre-processed data, which makes it applicable for big data implementations in (smart) cities and other data-driven mobility platforms.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

To know the transport mode people use for their travel is a key element in any mobility study (Mc Fadden, 1978). Next to trip purpose, travel frequency and origin-destination pairs, transport mode information allows us to better understand people's travel behaviour (Bohte & Maat, 2009; Chen, Ma, Susilo, Liu, & Wang, 2016), manage traffic flows (Asakura, Tanabe, & Lee, 2000), and ensure a better transport and urban planning strategy (Magnanti & Wong, 1984). In general, transport modes can be seen as competing with or complementing one another in terms of cost, travel time, accessibility or comfort. It is not uncommon that mobility systems are compared, in regard to sustainability and efficiency of transport policy measures, based on achieved modal shifts (Banister, 2008; Chapman, 2007; Davison & Knowles, 2006) or accomplished modal splits of overall population in a defined area (Banister, 2008; Nabais, Negenborn, Carmona Benítez, & Ayala Botto, 2015; Steininger, Vogl, & Zettl, 1996; Tabuchi, 1993). On a more localized level information on the transport mode used forms a basis for development of location-based mobility services, and/or delivery of targeted mobility messages such as route information (Raper, Gartner, Karimi, & Rizos, 2007; Semanjski & Gautama, 2016a; Semanjski & Gautama, 2015).

Clearly, transport mode data collection is a key issue. Traditionally, this information is collected based on the questionnaires, travel diaries, and/or interviews. These data sources are also often referred to, in recent literature, as small data (Chen et al., 2016; Semanjski & Gautama, 2016b). Alternatively, big data based methods and techniques include extraction of transport mode information from Global Navigation Satellite Systems (GNSS), Bluetooth and/or mobile sensing techniques or by inferring it from mobile sensed big data (Semanjski & Gautama, 2016b; Toole et al., 2015; Vlahogianni, Park, & Van Lint, 2015). In this context, big data is often typified by the so-called 3Vs definition (Chen, Mao, & Liu, 2014; Semanjski et al, 2016; Witlox, 2015) where the three Vs stand for increase of volume (data scale becomes increasingly big when compared with small data), variety (data come in different formats, as structured, semi-structured and unstructured data) and velocity (data generation frequency strongly increases resulting in need for timely conduction of data collection-processing chain). In this paper, we explore potential to extract transport mode information from big data based on the trip's spatial context. Suggested approach relays on machine learning technique and complements existing approaches that mainly relay on variables that describe the moving object itself (as acceleration, speed etc.). Concretely, we aim to investigate the potential of implementing freely available geographical information and location information in order to infer and/or recognize transport modes from mobile sensed big data. To this end, we develop a new method, test this method on an extensive dataset collected over a four

* Corresponding author.
E-mail address: ivana.semanjski@ugent.be (I. Semanjski).

http://dx.doi.org/10.1016/j.compenvurbsys.2017.07.004
0198-9715/© 2017 Elsevier Ltd. All rights reserved.

months' timeframe involving more than 8000 individuals for the city of Leuven (Flanders, Belgium), and evaluate our approach. The paper is organized as follows. In Section 2, we present extensive literature review on the matter and position our approach in regard to the state of the art. In Section 3 we describe necessary data input. This relates to the collected mobile sensed data, and the spatial context awareness to determine the transport mode. In Section 4 we develop our model. The details of implemented model are given in the Section 5 and followed by the presentation of the results. The final sections include discussion and the conclusion remarks.

2. Literature review

As briefly mentioned in the introduction section, several methods and techniques exist that are used to extract information on the transport mode one utilizes for his/her travels: (i) traditional questionnaires, travel diaries, and/or interviews; (ii) with the help of GNSS, Bluetooth and/or mobile sensing techniques; or (iii) by inferring it from mobile sensed big data.

2.1. Surveys and interviews

Surveys and interviews are traditional and a rather straightforward way to collect data on transport mode people utilizes for their travels. They are conducted by means of paper or phone household surveys or interviews during which one is asked to record, or state, his or her travel behaviour on an average weekday. However, many studies (Ettema, Timmermans, & van Veghel, 1996; Stopher & Greaves, 2007) have shown that data collected in this manner deviated systematically from actual travel behaviour. These deviations include, among others, respondents' tendencies to underreport small trips (Itoh & Hato, 2013), car drivers to underestimate and public transport users to overestimate their travel times (Clifton & Muhs, 2012). In order to avoid these pitfalls, paper travel diaries were introduced (Stopher & Wilmot, 2000). Here, people are asked to systematically note their travel behaviour details with respect to travel times, transport modes, trip purposes, and frequencies. The data collection time span usually covers a period of one full typical week during a non-holiday seasons. Arentze et al. (2001) and Groves (2006) report that participants tend to postpone filling in these diaries which resulted in obtaining incomplete and inconsistent information. Quite often this reflects in forgetting to mention some smaller trips (e.g. walking to nearby restaurant during the lunch break), and rounding time and distances (Witlox, 2007) or having difficulties in defining the exact locations of places they have visited. As smaller trips were usually made by active transport modes, like cycling or walking (Declercq, Janssens, & Wets, 2013; Saelens, Sallis, & Frank, 2003), such data collection practice resulted in bias observed modal splits and further underpinned evolution of car-oriented transport planning.

2.2. GNSS and mobile sensing

Global navigation satellite systems opened new horizons in the collection of travel data (Feng & Timmermans, 2014; Wolf, Bricka, Ashby, & Gorugantua, 2004; Wolf, Guensler, & Bachman, 2001). By providing high resolution tracking data, GNSS collected data showed potential to overcome some of the disadvantages of the more traditional, above-mentioned, approaches (Bohte & Maat, 2009; Bricka, Sen, Paleti, & Bhat, 2012; Montini, Prost, Schrammel, Rieser-Schüssler, & Axhausen, 2015). However, GNSS-logging exhibited a number of major restrictions as devices were typically installed in vehicles (Turner, Eisele, Benz, & Holdener, 1998). Hence, they only tracked a small portion of mobility behaviour (i.e., vehicle trips). Alternatively, when using portable handheld GNSS-devices, effort and discipline from the respondent to continuously carry the device with him/her are required, as forgetting the device results in unreported gaps in the trip data. Further developments in the field of GNSS chipsets enabled their integration in mobile phones allowing new possibilities for tracking. Today, smartphones have the same capabilities as the portable GNSS-device. However, carrying a smartphone has also become a habit and is therefore considered less of a burden, reducing the risk of non-reported trips. Furthermore, while vehicles were potentially shared among family members, this is less of a case for smartphones. Thus, this way sensed mobility behaviour corresponds more to ones' true travelling patterns.

We can distinguish three general ways mobile phone data are sensed for mobility studies:

First, we have call detail record (CDR) and network signalization data. This represents standardized data, collected by mobile network operators for billing purposes. Such data include records of all activities such as calls, SMSs, internet and data services where each record includes spatial and temporal parameters. Their applicability in the scope of mobility studies has been investigated for rush hour analysis (Bar-Gera, 2007; Järv, Ahas, Saluveer, Derudder, & Witlox, 2012), detection of variability in human activity spaces (González & Hidalgo, 2008; Hoteit, Secci, Sobolevsky, Ratti, & Pujolle, 2014; Järv, Ahas, & Witlox, 2014; Noulas, 2013; Williams, Thomas, Dunbar, Eagle, & Dobra, 2015; Xu et al., 2016), correlation of mobility behaviour with land use (Toole, Ulm, González, & Bauer, 2012), detection of origin-destination pairs and traffic zones (Alexander, Jiang, Murga, & Gonzalez, 2015; Dong et al., 2015; Iqbal, Choudhury, Wang, & Gonzalez, M. C., 2014; Toole et al., 2015). Kang et al. (2010) tackled complexity of such data pre-processing and related spatiotemporal analysis. They implemented geographical mapping and statistical analysis to gain deeper insight into ones' personal mobility patterns over different days and times of a day. Furthermore, Xu et al. (2015) studied a home-based approach to understand differences between human mobility patterns across diverse urban areas based on hierarchical clustering algorithm. Although CDR and network signalisation data are collected by all network operators, require no additional effort by users, no additional financial resources for their collection, cover wide areas and large populations their usage for mobility, and other, studies is still hindered by a number of privacy and regulatory issues as well as some technological, business related and methodological ones (Ahas et al., 2014; Calabrese, Ferrari, & Blondel, 2014; Seidl, Jankowski, & Tsou, 2016; Vij & Shankari, 2015). Among the problems mentioned, potentially the most significant ones are those related to location precision (limited to cellular network base station locations) and time resolution (limited to users' activity or regular network location updates dependable upon the type/generation of the network).

Second, we have so-called 'passive' mobile phone application based logging. This refers to the use of dedicated applications that run as a GNSS-based data logger in the background on the smartphone. Use of such sensed data is examined for the purposes of investigating individual mobility patterns (Calabrese, Diao, Di Lorenzo, Ferreira, & Ratti, 2013; Shin et al., 2015), speed analysis (Huss, Beekhuizen, Kromhout, & Vermeulen, 2014) or traffic monitoring (Herrera et al., 2010). The main advantage of this approach comes from higher spatial and temporal resolution of collected data than it is the case for mobile network call detail record, which are also considered to be 'passively' generated, as they also do not require any effort from the user. However, constantly active GNSS sensor tends to drain smartphone battery quite fast. This results in increased burden for users who, consequently, need to charge their phones more frequently. Compared to 'active' logging, the main advantage is that there is no need for interaction by respondents. That said, data collected this way require demanding data processing and interpretation efforts when compared to 'active' logging.

Third, and finally, we have 'active' and/or 'interactive' mobile phone application based logging. It represents the use of interactive mobile applications where respondents can report additional trip data as start of the trip or transport mode. In its essence, it can be considered as detailed mobile travel diary with the GNSS logging. Such reporting was, for instance, used to investigate the influence of carbon dioxide emissions

Download English Version:

https://daneshyari.com/en/article/4965199

Download Persian Version:

https://daneshyari.com/article/4965199

Daneshyari.com