Tourism Management 44 (2014) 67–81

Contents lists available at ScienceDirect

Tourism Management

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

Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium

Mathias Versichele [a,*], Liesbeth de Groote [a], Manuel Claeys Bouuaert [a], Tijs Neutens [a], Ingrid Moerman [b], Nico Van de Weghe [a]

[a] Department of Geography, Ghent University, Krijgslaan 281 (S8), B-9000 Ghent, Belgium
[b] Department of Information Technology, Ghent University, IBBT, Gaston Crommenlaan 8 Box 201, B-9000 Ghent, Belgium

HIGHLIGHTS

- Empirical case study demonstrating the use of Bluetooth tracking for non-participatory collection of tourist movement data.
- Tourists in Ghent, Belgium tracked over 14 tourist attractions, 14 hotels and tourist inquiry desk during 15 days.
- Visit pattern analysis through mining of association rules between tourist attractions.
- Visualization of discovered patterns by *visit pattern maps*.
- Non-participatory collection of tourist movement data using Bluetooth tracking.

ARTICLE INFO

Article history:
Received 3 July 2013
Accepted 21 February 2014

Keywords:
Data mining
Association rule learning
Bluetooth tracking
Tourist attraction visits

ABSTRACT

The rapid evolution of information and positioning technologies, and their increasing adoption in tourism management practices allows for new and challenging research avenues. This paper presents an empirical case study on the mining of association rules in tourist attraction visits, registered for 15 days by the Bluetooth tracking methodology. This way, this paper aims to be a methodological contribution to the field of spatiotemporal tourism behavior research by demonstrating the potential of ad-hoc sensing networks in the non-participatory measurement of small-scale movements. An extensive filtering procedure is followed by an exploratory analysis, analyzing the discovered associations for different visitor segments and additionally visualizing them in 'visit pattern maps'. Despite the limited duration of the tracking period, we were able to discover interesting associations and further identified a tendency of visitors to rarely combine visits in the center with visits outside of the city center. We conclude by discussing both the potential of the employed methodology as well as its further issues.

1. Introduction

Movement represents a key aspect of tourism, both in order to reach a certain destination from an individual's habitual environment and to move around within that tourist destination. As a consequence, many research efforts have focused on the spatiotemporal behavior of tourists in order to *inter alia* optimize tourist infrastructure, for marketing incentives, and to better manage the impacts of tourist mobility on the environment (Shoval & Isaacson, 2009). Due to the complex nature of tourism (McKercher, 1999), there is a growing need for empirical movement data to accompany theoretical models. Yet, empirical studies into tourist mobility have traditionally been rather scarce due to the labor-intensive and often expensive nature of traditional methods such as direct observation (Hartmann, 1988) or personal interviews (Kemperman, Borgers, & Timmermans, 2009). Space-time diaries (Connell & Page, 2008; Janelle, Goodchild, & Klinkenberg, 1988; Lau & McKercher, 2006) shift some of the weight away from the researchers but are often characterized by a low reliability as respondents tend to forget or neglect to register certain activities. Recently, however, tracking technologies offer a more scalable and objective way to capture spatiotemporal behavior in a detailed way (Shoval & Isaacson, 2009). The use of global navigation satellite systems – such as GPS – is currently the dominant approach and its adoption in tourism research through the distribution of logging devices is

* Corresponding author. Tel.: +32 (0)9 264 46 89.
E-mail addresses: Mathias.Versichele@ugent.be (M. Versichele), Liesbethm.DeGroote@ugent.be (L. de Groote), Manuel.ClaeysBouuaert@ugent.be (M. Claeys Bouuaert), Tijs.Neutens@ugent.be (T. Neutens), Ingrid.Moerman@intec.ugent.be (I. Moerman), Nico.VandeWeghe@ugent.be (N. Van de Weghe).

http://dx.doi.org/10.1016/j.tourman.2014.02.009
0261-5177/

well-documented (Shoval & Isaacson, 2007a, 2007b; Shoval, McKercher, Ng, & Birenboim, 2011; Tchetchik, Fleischer, & Shoval, 2009). An alternative approach is to track the movement of mobile phones through a cell tower network without the direct participation of the phone's owner (González, Hidalgo, & Barabási, 2008; Ratti, Pulselli, Williams, & Frenchman, 2006). Particularly in Estonia, this method has already been extensively used for studying regional movement patterns of tourists (Ahas, Aasa, Mark, Pae, & Kull, 2007; Ahas, Aasa, Roose, Mark, & Silm, 2008).

Despite the undeniably important contribution of both tracking methodologies to the research field, we argue that both approaches have certain limits. The distribution of logging devices necessitates the direct collaboration of the tracked individual. This makes it hard to scale up the methodology to large groups of individuals. Additionally, any participatory methodology presents a risk for self-selection bias where individuals with certain characteristics would show a higher degree of cooperation and thus be over-represented in the sample. While the use of smartphone apps for tracking tourist movements – e.g. through shared user-generated content such as pictures (Jankowski, Andrienko, Andrienko, & Kisilevich, 2010) – may decrease some of the intrusive nature in comparison with the use of logging devices, it still represents a participatory methodology with an inherent risk for bias and data sparseness in some locations. Finally, GPS technology is not applicable to indoor contexts. Cell phone tracking, on the other hand, encompasses other limitations. First, the spatial accuracy of the method is limited by the density of cell towers over the study area. In Estonia, for example, around 50% of measurements were correct to within only 400 m in urban areas and only 2600 m in rural areas (Ahas, Laineste, Aasa, & Mark, 2007). While this does not hinder the study of regional movements, it does pose a problem when studying movement within a certain tourist destination (e.g. a city). Second, these data sets are property of mobile operators and – as such – not freely available. In summary, it seems that small-scale spatiotemporal behavior cannot be measured without the direct involvement of the individual to be tracked. This hinders studying larger groups of individuals.

A recent alternative in the non-participatory tracking of mobile phones is the use of ad-hoc sensor networks distributed over a study area. Bluetooth technology, for example, has already been employed for studying pedestrian flows at mass events (Delafontaine, Versichele, Neutens, & Van de Weghe, 2012; Stange, Liebig, Hecker, Andrienko, & Andrienko, 2011; Versichele, Neutens, Delafontaine, & Van de Weghe, 2012; Versichele, Neutens, Goudeseune, Van Bossche, & Van de Weghe, 2012) and in social studies (Eagle & Pentland, 2005). WiFi (Bonné, Barzan, Quax, & Lamotte, 2013) and RFID (Öztayşi, Baysan, & Akpinar, 2009) technology provide similar possibilities. Due to the limited coverage of each sensor, a careful deployment of sensors may thus provide movement records with a granularity that is much smaller than the accuracy level of cell phone tracking data. By deploying sensors with these wireless technologies at a set of pre-defined tourist locations, one is able to study the spatiotemporal behavior at and between these locations. Despite this potential in the non-participatory registration of small-scale movements, we have as yet no indication of the application of the methodology for tourism management purposes.

This paper aims to address this issue by presenting a case study where visitors to tourist attractions in Ghent, Belgium were registered through an ad-hoc Bluetooth sensor network. Due to the novelty of Bluetooth technology – and the use of ad-hoc sensing networks in general for that matter – we will not only elaborate extensively on the working principle of the methodology, but also on the analytical potential of such tracking data. Ad-hoc sensor network data lack the typical socio-demographic or psychographic variables used as explanatory factors in various studies related to tourism behavior. In contrast with hypothesis testing procedures, sensor network data often need to be investigated without any a priori assumptions. The collection of such methods that can be used to discover (non-trivial) patterns and knowledge from large data sets is called data mining (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Several data mining techniques have already been frequently applied to tourism data, including r*egression* techniques (Song & Li, 2008; Witt & Witt, 1995), *clustering* (Bloom, 2005; Cini, Leone, & Passafaro, 2010; Dolničar, 2004; Dolničar & Leisch, 2003; Tchetchik et al., 2009), *sequential pattern mining* (Orellana, Bregt, Ligtenberg, & Wachowicz, 2012; Shoval & Isaacson, 2007a), and *classification* (Law & Au, 2000; Law, Bauer, Weber, & Tse, 2006). *Association rule learning* is concerned with discovering associations between variables without fixing the output variable, as is the case in classification. In comparison with the other techniques, implementations of association rule learning in tourism research are rather scarce. Documented applications found in literature include tourism product development (Al-Salim, 2008; Liao, Chen, & Deng, 2010), domestic tourist profiling (Emel, Taskin, & Akat, 2007), sharers and browsers of touristic websites (Rong, Vu, Law, & Li, 2012), and change and trend identification in Hong Kong outbound tourism (Law, Rong, Vu, Li, & Lee, 2011).

This paper aims to be a methodological contribution to the field of spatiotemporal tourism behavior research by demonstrating the potential of ad-hoc sensing networks in the non-participatory measurement of small-scale movements. We describe a case study where visitors to 14 tourist attractions were registered through Bluetooth technology sensing the mobile devices they were carrying around. In an attempt to investigate the analytical potential of the resulting data, we employ an association rule learning algorithm to mine for 'interesting' patterns in the combinations of visits to different attractions (in the sense that they represent potentially valuable information which would be hard to discover in another more straightforward way). As the tracking data are completely anonymous, it is impossible to directly distinguish between local visitors and actual tourists as defined by the World Tourism Organization: people "traveling to and staying in places outside their usual environment for not more than one consecutive year for leisure, business and other purposes" (World Tourism Organization, 1995). By deploying sensors in 14 hotels, however, some visitors will be identified as hotel guests therefore giving a strong suggestion that they are indeed tourists. Extra context is added by tracking visitors at the tourist inquiry desk as well. Combining the tracking data with these contextual assumptions, we will investigate patterns for different visitor *segments* (e.g. those that were only detected on one day, those that were identified as hotel guests, etc.). For the sake of clarity, we will always use the term *visitors* instead of further labeling them as tourists.

The remainder of the paper is organized as follows. In Section 2, we first discuss the Bluetooth tracking methodology and its specific implementation in the case study (Section 2.1). Next, we describe association rule learning in more detail (Section 2.2) and how the information it generates can be summarized in *visit pattern maps* (Section 2.3). Section 3 outlines the filtering of the raw tracking data in detail, and Section 4 presents a first data exploration. The actual association rule mining is performed for the different visitor segments in Section 5. We finish with a discussion and conclusion (Section 6).

2. Methods and data

2.1. Bluetooth tracking

For this study, scanners with Bluetooth sensors were deployed at 29 locations in and around the historical center and the 'arts