Transportation Research Part C 44 (2014) 146–170

Contents lists available at ScienceDirect

Transportation Research Part C

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

A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures

Antonin Danalet [a,*], Bilal Farooq [b], Michel Bierlaire [a]

[a] Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne, Station 18, CH-1015 Lausanne, Switzerland

[b] Département des Génies Civil, Géologique et des Mines, École Polytechnique Montréal, 2500 Ch. Polytechnique Montréal, H3T 1J4 Montréal, Canada

A R T I C L E   I N F O

Article history:
Received 3 October 2013
Received in revised form 5 March 2014
Accepted 26 March 2014

Keywords:
Network traces
Activity choice modeling
Pedestrians
Semantically-enriched routing graph (SERG)
Potential attractivity measure
Activity-episode sequence

A B S T R A C T

In this paper, we propose a methodology to use the communication network infrastructure, in particular WiFi traces, to detect the sequence of activity episodes visited by pedestrians. Due to the poor quality of WiFi localization, a probabilistic method is proposed that infers activity-episode locations based on WiFi traces and calculates the likelihood of observing these traces in the pedestrian network, taking into account prior knowledge. The output of the method consists of candidates of activity-episodes sequences associated with the likelihood to be the true one. The methodology is validated on traces generated by a known sequence of activities, while the performance being evaluated on a set of anonymous users. Results show that it is possible to predict the number of episodes and the activity-episodes locations and durations, by merging information about the activity locations on the map, WiFi measurements and prior information about schedules and the attractivity in pedestrian infrastructure. The ambiguity of each activity episode in the sequence is explicitly measured.

1. Introduction

In recent years, interest in crowd dynamics and pedestrian modeling is reviving due to urban growth and its pressure on urban infrastructure (Bierlaire and Robin, 2009; Duives et al., 2013; Kasemsuppakorn and Karimi, 2013; Kneidl et al., 2013; Weidmann et al., 2014). Management of congestion is the main issue for pedestrian infrastructure. Crowd and pedestrian simulation is emerging as a tool for designing new infrastructures and optimizing the use of current infrastructures. Innovative data collection techniques and realistic experiments are vital in estimating the demand for these infrastructures.

In order to predict the total demand within a given area, activity choice models need to be developed at the scale of pedestrian infrastructure. Pedestrian demand is driven by a need to perform activities in different locations. The existence of time–space constraints in pedestrian infrastructures on the one hand and the spontaneous choice of en route destinations on the other hand ask for explicit modeling of activity scheduling decisions. Such models are traditionally used for car trips as an important source of information for strategic planning, and management or optimization of transportation networks (Ben-Akiva et al., 1996; Bowman and Ben-Akiva, 2001; Arentze and Timmermans, 2004; Balmer et al., 2006; Roorda et al., 2008, among others). For pedestrians, they will be useful in describing the congestion, for the efficient design of new facilities, and travel guidance and information systems.

* Corresponding author. Tel.: +41 21 6932532.
E-mail addresses: antonin.danalet@epfl.ch (A. Danalet), bilal.farooq@polymtl.ca (B. Farooq), michel.bierlaire@epfl.ch (M. Bierlaire).

http://dx.doi.org/10.1016/j.trc.2014.03.015
0968-090X/

Individual mobility traces are becoming available from pervasive systems, such as cellular networks (González et al., 2008) or WiFi hotspots (Section 2.1). In many cases, cost and privacy issues prohibit from installing high precision sensors such as cameras covering an entire pedestrian infrastructure. The large size of an airport or a railway station implies either precise sensors with incomplete coverage (e.g., cameras or bluetooth sensors in intersections), or full coverage with imprecise long range sensors (e.g., cellular network data, traces from WiFi infrastructures). As a result, localization data are either scarce, fuzzy, or both. We propose a methodology exploiting scarce data with an explicit modeling of the imprecision in the measure, and using prior knowledge of the infrastructure.

Section 2 reviews existing works about traces from communication infrastructure, pedestrian maps and activity-based modeling. Section 3 describes the necessary data for detecting pedestrians, while Section 4 describes the methodology to merge these data. A case study on the Ecole Polytechnique Fédérale de Lausanne (EPFL) campus is described in Section 5, with results of this case study, together with validation and sensitivity analysis. Finally, we conclude and discuss future work in Section 6.

2. Literature review

This paper focuses on using existing localization data for modeling macroscopic behavior such as destination or activity choice. The following literature review is divided in three parts, corresponding to the three challenges we meet in pedestrian demand modeling: data collection (Section 2.1), representation of space (Section 2.2) and modeling (Section 2.3).

2.1. Collecting data from digital footprints

The recent developments in detection technologies open doors to new researches about pedestrian behavior. In the field of trajectory detection, Taniguchi et al. (2013) are using Bayesian estimation on binary sensors located at the border of a cell, while Alahi (2011) and Alahi et al. (2014) use networks of cameras to track and analyze pedestrian trajectories. Alahi's main motivation is the number of already installed cameras generating large datasets. Smartphones are sharing this characteristic: a majority of people are carrying a mobile device such as a smartphone, and they generate data. Several data collection techniques about smartphones are device-centric. We focus here on data from communication network infrastructure ("network traces").

Using traces from communication network infrastructure has several advantages on data from the smartphone. First, full coverage of the facility is usually cheap and allows for an estimation of the overall demand. The communication infrastructure sometimes already exists, and increasing its density has a positive side effect. Smartphone users do not need to install anything on their device, and so the access to sensitive information such as emails or address book is limited for the analyst, which ensure privacy for the users. Finally, traces from communication network infrastructure are related to the infrastructure and not to the individual: we are tracking all individual smartphones going through a facility and not all places visited by the same individuals. It allows the analyst to focus on the pedestrian facility covered by the communication network.

There are few drawbacks to network traces as well. Socio-economic and demographic attributes are difficult to collect due to both privacy concerns (if the data already exist) and to the difficulty to survey the tracked person from the infrastructure side (if the data does not exist). Additionally, smartphone users are not necessarily representative of the full population.

Several applications using data from communication infrastructure, both with WiFi and GSM traces (Bekhor et al., 2013; Calabrese et al., 2011), have been developed to study mobility behavior. These new data collections are motivated by the needs for calibrated agent-based models. Post-processing methods are needed to transform these raw observations into data adapted for modeling purpose to overcome imprecision and missing observations in the data: detection of stops points, activity purpose detection through land-use information and spatial matching (Rieser-Schüssler, 2012). With GSM traces, Bekhor et al. (2013) mention the elimination of "unreasonable movements performed in short time periods between antennas located far apart" without more details. Calabrese et al. (2011) does not consider the underlying transportation network to correct for anisotropy.

A large literature exists about WiFi traces from a computer communication point of view. A complete review can be found in Aschenbruck et al. (2011). All references in this paper define mobility trace-based models as a tool to improve the quality of the WiFi. Field studies have been done (Tang and Baker, 2000; Balachandran et al., 2002; Balazinska and Castro, 2003; Yoon et al., 2006; Sevtsuk et al., 2009; Zola and Barcelo-Arroyo, 2011; Wanalertlak et al., 2011; Meneses and Moreira, 2012). The main results are the prediction of changes in access points (APs). The main problem reported in these articles is the *ping pong* effect, when the device has similar signal strengths from different APs and changes regularly from one to another. This is a problem from a network viewpoint, and also for modeling pedestrian origins and destinations. Yoon et al. (2006) propose to use a moving average weighted by time spent at destination to remove the extra AP logs. A general solution presented in Aschenbruck et al. (2011) consists in aggregation of data over different APs. Most studies about WiFi are focusing on network performance and management and not on human mobility. In Yoon et al. (2006), contrarily to all other papers cited here, an OD matrix is estimated at the building level in Dartmouth college. Variations in time/day are not considered, as Aschenbruck et al. (2011) noticed.

In the literature about mobility models for WiFi infrastructure from a computer communication point of view, the most common model, Random Waypoint model (RWP), is often criticized as not representing real human mobility (Conti and