Computers, Environment and Urban Systems 48 (2014) 124–137

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Inferring building functions from a probabilistic model using public transportation data

Chen Zhong [a,*], Xianfeng Huang [a,b], Stefan Müller Arisona [c], Gerhard Schmitt [a], Michael Batty [d]

[a] Future Cities Laboratory, Department of Architecture, ETH Zurich, 8092 Zurich, Switzerland
[b] State Key Lab of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 430079 Wuhan, China
[c] Institute of 4D Technologies, University of Applied Sciences and Arts Northwestern Switzerland FHNW, 5210 Windisch, Switzerland
[d] Centre for Advanced Spatial Analysis, University College London, 90 Tottenham Court Road, W1N 6TR London, England, United Kingdom

ARTICLE INFO

Article history:
Received 19 March 2013
Received in revised form 14 July 2014
Accepted 16 July 2014
Available online 24 August 2014

Keywords:
Bayesian model
Spatial statistics
Building function
Activity
Smart card data

ABSTRACT

Cities are complex systems. They contain different functional areas originally defined by planning and then reshaped by actual needs and use by the inhabitants. Estimating the functions of urban space is of significant importance for detecting urban problems, evaluating planning strategies, and supporting policy making. In light of the potential of data mining and spatial analysis techniques for urban analysis, this paper proposes a method to infer urban functions at the building level using transportation data obtained from surveys and smart card systems. Specifically, we establish a two-step framework making use of the spatial relationships between trips, stops, and buildings. Firstly, information about the travel purposes for daily activities is deduced using passengers' mobility patterns based on a probabilistic Bayesian model. Secondly, building functions are inferred by linking daily activities to the buildings surrounding the stops based on spatial statistics. We demonstrate the proposed method using large-scale public transportation data from two areas of Singapore. Our method is applied to identify building functions at building level. The result is verified with master plan, street view, and investigated data, and limitations are identified. Our work shows that the presented method is applicable in practice with a good accuracy. In a broader context, it shows the effectiveness of applying integrated techniques to combine multi-source data in order to make insights about social activities and complex urban space.

1. Introduction

Urban systems are composed of many different forms of functional areas, which interact with one another to generate the complexity that defines a city. These functional areas are historically associated with many urban processes, some related to the institutions that are used to support planning but most being shaped by individuals' actual needs through processes of bottom-up change. In this spirit, Jane (1961) described cities as 'problems of organized complexity'. Taking a small park as an example, she argued that "... even this partial influence of the park's design upon the park's use depends, in turn, on who is around to use the park and when, and this in turn depends on uses of the city outside the park itself...". Similarly, in the book by Rodrigue (2013), land uses are defined in two ways. Formal land use refers to its form, pattern, and aspect, while functional land use refers to its socioeconomic description in space. The latter aspect is likely to imply higher levels of dynamic temporal change compared to the former as activities change faster than the physical locations and land uses that contain them. As discussed in Green (2007), functional changes in cities are not tied to morphological changes. In places such as Singapore, it is crucial to understand urban functions and their compatibility with the original Master Plan, which is very important to the development of the urban system, and the current push in understanding the dynamics of urban areas requires costly cross-sectional survey data, which in principle should be used to dynamically update information. As a potential solution to these problems, only recently has the availability of multiple location data sources, such as GSM traces, Wi-Fi data, GPS traces from taxis and smart-card data, emerged, and this is, for the first time, greatly stimulating the use of these "big" data sets for urban analysis. As it implies in Yuan, Zheng, and Xie (2012) that regions of different functions in a city can be detected using human mobility data and points of interest data. In Roth and et al. (2011), the characteristics of a polycentric urban form are defined from the analysis of large-scale, real-time smart-card data from which individuals' movement patterns can be inferred.

* Corresponding author.

http://dx.doi.org/10.1016/j.compenvurbsys.2014.07.004
0198-9715/

The motivation for our study is to measure the structure and form of urban spaces in terms of real functions associated with urban land use using newly available ‘big data’ and, in this way, to explore the potential of using smart-card data to infer urban functions. To do this, we deduce information about individuals’ travel purposes for daily activities from mobility patterns so that we can link these activities to specific locations to detect building functions. A two-step framework that makes use of the spatial relations between trips, bus stops, and building plots is presented. In this framework, we first analyze the survey data to find the mobility patterns of typical travel purposes based on travel time, activity time, and travel frequency. The analyzed results are then used for trip classification of the smart-card data using a probabilistic model. By analyzing the distribution of bus trips from each stop to the surrounding buildings, we can infer the most likely function of the building using a standard inverse distance weight function. The data used in this study are the Household Interview Travel Survey (HITS) and seven-day smart-card data obtained in Singapore from the Land Transport Authority, which pertain to all public transport usage. In an exploratory manner, we apply the method at the building level as the main focus of this paper. However, a more reasonable result is achieved at the block level due to the volume and resolution of the data, and we discuss the implications of this issue here. The inferred functions reflect the real use of urban space, which can be used to verify independent observations from various original plans. Moreover, corresponding results from different time series can be further used to detect the changes in activity location choice.

This paper develops three main contributions. First, we apply a probabilistic model to infer the travel purpose of a trip using spatiotemporal as well as socially related information. This enables us to explore a method of studying urban spaces through peoples’ mobility patterns. Second, we propose a framework to infer building functions by combining multi-source data, namely, survey data and smart-card data, while integrating data mining techniques using a spatial statistical method. Third, the proposed method is demonstrated with real data collected in Singapore. Consequently, this study is focused on investigating the potential of using big data to infer the dynamics of various space functions, and we believe that this is one of the first attempts anywhere to extract activity and land usage data in this manner. The remainder of the paper is organized as follows. In the next section, related works are reviewed. In Section 3, the proposed methodology is presented, including the terminology, framework, and details of the probabilistic model. Section 4 discusses the experiments demonstrating the proposed method using the Household Interview Transportation Study (HITS) and the smart-card data. Section 5 concludes the paper and discusses further research.

2. Related work: Discovering functional areas in cities from movement data

Assessing the functions of urban spaces in terms of land use types is of significant importance for understanding urban problems (Taleai et al., 2007) and for evaluating planning strategies (Kajtazi, 2010). However, assessing urban functionality requires costly survey methods such as field investigation and interview questionnaires. In addition to the amount of manpower and time involved, the reliability of information is heavily influenced by subjective factors such as the time, place, and investigator’s personal experience. The development of techniques based on Geographic Information Systems (GIS) and the availability of multiple data sources, such as GSM traces on cars, trains and taxis; mobile phone calls; Wi-Fi data; and smart-card systems, provide us with alternative solutions and change the way we can approach urban analysis.

Valuable insights have been gained regarding social activity and the complexity of urban space through analyses of movement data because urban travel is a good proxy for the transfer of urban flows, such as people, material products, and energy, thus revealing the importance of temporal dynamics in cities. Findings have been achieved from exploring such dynamics, for instance, using mobile telephone position data to analyze daily activity patterns (Phithakkitnukoon et al., 2010; Ratti et al., 2006), comparing the differences in temporal patterns with respect to the consumption of space (Ahas et al., 2010), studying spatiotemporal structures of urban mobility at a large scale (Sun et al., 2011) and classifying land uses based on aggregated data (Pei et al., 2014). The GPS trajectories of taxi cabs traveling in urban areas provide detailed location information, and in (Qi et al., 2011), the on-entrance/off-exit frequencies of taxi passengers were used to depict social activities in a region. Similarly, temporal patterns of pick-ups and drop-offs have been analyzed and associated with different land-use features in Liu and et al. (2012). In Yuan et al. (2012), taxi data combined with points of interests (POIs) were used to discover regions with different functionalities in a city. Some discussions regarding the opportunities and challenges of using various location data can be found in Lu and Liu (2012).

As smart-card payment systems are rapidly being adopted in cities around the world, they have also become an important data source that produces large quantities of very detailed data about an individual’s daily travel (Pelletier, Trépanier, & Morency, 2009). In Quebec, data mining methods and public transport planning models have been used to obtain an improved portrait of users’ travel behavior, and this has been tested using twelve one-week records (Agard, Morency, & Trépanier, 2006). In Seoul, a study was conducted by Park, Kim, and Lim (2008), in which boarding times and disembarking times were mapped and analyzed to prove the reliability of smart-card data. Liang and et al. (2009) investigated spatiotemporal human mobility patterns using smart-card data in Shenzhen, China, while (Munizaga & Palma, 2012) estimated a public transport O-D matrix from smart-card and GPS data in Santiago, Chile for transport systems analysis. In a study by Roth et al. (2011), data were collected from the smart-card system in the London tube, which is based on the Oyster card system, and were used to infer the statistical properties of individual movement patterns and to identify the polycentric nature of the various transport hubs in central London.

A clear trend that is revealed from this brief survey is in the exploration of the potential of using ‘big’ positional (geospatial) data for the analysis of urban forms, as proposed in Ratti and et al. (2006). In line with such trend, there are new methods of urban analysis emerging. For instance, discrete choice models can be used to estimate dynamic workplace capacities (Ordóñez Medina & Erath, 2013), identify urban activities from a synthesis of smart-card and survey data (Chakirov & Erath, 2012) and discover different functional regions within a city using floating car and point of interest data (Yuan et al., 2012). Machine learning methods are also being introduced to infer land use from mobile phone activity records and from zoning regulations (Toole et al., 2012). Spatial network analysis has been applied to the same set of smart-card data used in this paper to trace the urban transformation of decentralization in Singapore (Zhong et al., 2014).

In light of these potential uses of new data sources and analysis methods, this paper proposes a two-step integrated spatial data mining method aiming at inferring building functions using smart-card and survey data. A probabilistic model based on a Bayesian classifier and its related spatial statistics is used (1) to integrate considerably more attributes compared to when simply using spatiotemporal information and (2) to infer additional types of activity places instead of detecting only residential and workplaces, which has occurred in most case studies to date. Bayesian