Computer Communications 87 (2016) 19–36

Contents lists available at ScienceDirect

Computer Communications

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

On the properties of human mobility

Michela Papandrea[a,*], Karim Keramat Jahromi[b], Matteo Zignani[b], Sabrina Gaito[b], Silvia Giordano[a], Gian Paolo Rossi[b]

[a] NetLab, ISIN-DTI, SUPSI, Manno, Switzerland
[b] Computer Science Department, Universitá degli Studi di Milano, Italy

ARTICLE INFO

Article history:
Received 4 May 2015
Revised 15 March 2016
Accepted 19 March 2016
Available online 26 March 2016

Keywords:
Human mobility
Visited points
Pervasive computing

ABSTRACT

The current age of increased people mobility calls for a better understanding of how people move: how many places does an individual commonly visit, what are the semantics of these places, and how do people get from one place to another. We show that the number of places visited by each person (Points of Interest – PoIs) is regulated by some properties that are statistically similar among individuals. Subsequently, we present a PoIs classification in terms of their relevance on a per-user basis. In addition to the PoIs relevance, we also investigate the variables that describe the travel rules among PoIs in particular, the spatial and temporal distance. As regards the latter, existing works on mobility are mainly based on spatial distance. Here we argue, rather, that for human mobility the temporal distance and the PoIs relevance are the major driving factors. Moreover, we study the semantic of PoIs. This is useful for deriving statistics on people's habits without breaking their privacy. With the support of different datasets, our paper provides an in-depth analysis of PoIs distribution and semantics; it also shows that our results hold independently of the nature of the dataset in use. We illustrate that our approach is able to effectively extract a rich set of features describing human mobility and we argue that this can be seminal to novel mobility research.

1. Introduction

In recent years we have witnessed a rapid increase of people mobility as the world population has become more interconnected and has begun relying on faster transportation methods, simplified connections and shorter commuting times. Unveiling and understanding human mobility patterns has become a crucial issue in supporting decisions and prediction activities when managing the complexity of today's social organization. In this, novel mobile communications technologies play a fundamental role. With such mobile technologies it is now possible to collect data about human habits and behavior all day long. Nowadays, people always carry their mobile phone with them. So, either in the form of Call Detail Records (CDRs) or with specialized apps [22,25], people's mobility data can be collected from mobile phones. Therefore, in the recent years, researchers have devoted considerable effort to collecting and studying human mobility patterns [7] and have applied their understanding to a variety of critical problems ranging from disease spreading [2], urban planning, smart and green transportation to network infrastructure [14,37], economy and marketing [30], and mobile network services [13]. Nonetheless, despite the advances in communications technologies and other important achievements, human mobility still represents an open and challenging research issue. In practice, the mobility pattern of each individual consists of the sequence of locations s/he visited. These locations and their correlations represent the core block of any modeling research and any activity aimed at understanding human mobility. Even though visited locations underlie all works in this field, their features remain largely unknown. This is due mainly to the fact that they have been considered as points in an area and social aggregation places, without anchoring spatial features to the behavior of each single user.

This paper, which represents an extension of our previous works [31,44], aims to fill the gap by providing a general framework for dealing with modeling locations from a per-user perspective. Also, it paves the way towards enabling the semantic interpretation of locations to be overlaid on their spatial distribution.

First, we introduce the notion of user's Points of Interest (PoIs) along with the methodology to extract them from different types of data. Then we provide both a metric to measure the importance of PoIs for a person and a methodology to classify them in terms

* Corresponding author. Tel.: +41 586666502.
E-mail addresses: michela.papandrea@supsi.ch (M. Papandrea), karim.keramat@unimi.it (K.K. Jahromi), matteo.zignani@unimi.it (M. Zignani), sabrina.gaito@unimi.it (S. Gaito), silvia.giordano@supsi.ch (S. Giordano), gianpaolo.rossi@unimi.it (G.P. Rossi).

http://dx.doi.org/10.1016/j.comcom.2016.03.022
0140-3664/

of: *(i)* Most Visited Points (MVPs), the places that a person visits most regularly, e.g. home and work locations; *(ii)* Occasionally Visited Points (OVPs), locations of interest for the user but visited just occasionally; and *(iii)* Exceptionally[1] Visited Points (EVPs), which correspond to seldom visited locations. This classification allows us to define a human mobility profile where the number of locations per each class and the time spent there are the characterizing attributes. We further study how people move across PoIs and PoI classes, enriching the knowledge derived from classification with the spatial as well as the temporal dimensions of mobility. The proposed classification and the PoIs and user features provide the basis for understanding human behavior by extracting the semantics of visited places. In line with similar works [10,15,23,33], we used a heuristic approach for the semantic analysis and experimented it on a large dataset containing mobility patterns of hundred thousands of people in a metropolitan area.

The paper supports its findings by extensively validating results on four different datasets. The first two datasets contain Call Detail Records of phone activities of a large mobile operator. The third dataset is mainly composed of trajectories (parts of a continuous mobility trace), while the last one consists of continuously sampled location data. The first two datasets have different characteristics in terms of spatial and temporal distribution of the visited places w.r.t the other two databases. By showing the validity of our approach throughout datasets with sometimes antithetical properties, we demonstrate the independence of our results w.r.t. a specific setting, and we are able to extract a deeper understanding of human mobility.

As a result of this work, some interesting properties about human mobility emerge. In fact, it turns out that people visit many locations in their life, but they have a *very small number of preferred locations (MVPs)* which are visited daily (e.g., home, work place), and *a higher, but still limited, number of locations of interest (OVPs)* which are visited with a lower frequency (e.g., gym, favorite restaurant, parent's house). We spend more than 50% of our time in MVPs. This indicates that those points are *the ones that best represent and characterize our lives.* On this basis, we propose an algorithm to identify home and work places which leverages the relevance of a place for a specific person and outperforms other algorithms in terms of semantic accuracy.

By analyzing the transition rules between PoIs, we find that, in *contrast with commonly accepted assumptions, the decision to move between two places is not taken on the basis of the geographical distance, but according to the relevance individuals ascribe to them and to the travel time between places.* Also, we show that the transition rule based on relevance follows the same distribution law independently of the mobility scenario.

The key contributions of our mobility framework can be summarized as follows:

- a novel per-user mobility analysis that highlights the following key properties:
  - people visit regularly just few places where they spend most of their time;
  - people also spend a significant amount of time in places they only visit once;
  - people commute between places based on their temporal distance and not the spatial distance;
  - HOME and WORK places are in the set of few places mostly visited, and, as such, the relevance R is a fundamental feature for their semantic identification;
- a classification of visited locations (PoIs) that enables the above mentioned analysis;
- a classification of users, based on how people move across PoIs and PoIs classes, derived from our mobility analysis;
- a semantic understanding of human behavior based on our mobility analysis;
- a thorough experimental validation on datasets with different properties.

The comprehension and the modeling of human mobility patterns play a key role in the design of protocols and forwarding strategies in contact-centric network infrastructure. These novel results can change how mobility is analyzed and modeled. Indeed, we argue that, to produce more realistic mobility traces, a mobility model needs to consider *i)* the new classifications introduced herein, and *ii)* the new features, their relationships and their different laws. This work could impact several computer and communications areas such as: localization [28,29], where our results indicate that a person's location can be predicted in the set of MVPs with a probability higher than 0.7; social interaction studies and data offloading [32] [16], as people tend to meet more frequently people with some MVPs in common and the latter characterize the single individual's mobility; human mobility modeling [41], as mobility can be described in terms of regular movement among MVPs and OVPs and extemporarily EVPs; recommendations [26] as people can get recommended places close to their MVP and not far in time from their current location.

2. Related work

Nowadays smartphones have an important role in capturing various behavioral aspects of users, ranging from how the device is used across different contexts to analyzing the spatial, temporal and social dimensions of everyday life through sources such as GPS, call and text logs, Internet access and Bluetooth logs. These data can be used in many areas, from urban planning, predicting and controlling epidemic infection diseases to planning and optimization of wireless and infrastructure-less communication systems. Fundamentally, these applications require the comprehension and recognition of predictable mobility patterns. To gain a better understanding of the dynamics involved in mobility, many experiments, based on different detecting technologies and performed in various locations, have been conducted. Most of them have been made available in the public repository CRAWDAD [1]. Among these datasets we focus on GPS-based traces as they allow us to precisely determine the geographical positions of users. In this study we also compare mobility data from cellular network towers with the GPS positioning. We made this choice to highlight similarities and differences among the mobility habits, due to the different detecting technologies usually adopted to study them. That results in a heterogeneous set of data which require different pre-processing techniques to get a uniform representation through which we deal with the analysis. For the above reasons this work relates to different research topics.

2.1. Significant location extraction

Part of our work, which involves GPS data, has been devoted to detecting the significant locations of a user. Many authors have suggested different extraction methods [8,18,20,38–40] based on clustering algorithms. Ashbrook and Starner [8] have proposed a two-step method to infer the significant locations. In the first step, the loss of the GPS signal is used as an indicator of interesting locations because it likely corresponds to buildings or indoor points. In the second step these points are clustered into locations using a variant of the k-means algorithm. In the clustering procedure, round clusters with a given radius are initially placed at k chosen points, and iteratively they move to a denser area, until no further

[1] We use the adverb 'exceptionally' as a synonym for rarely, seldom.