Pervasive and Mobile Computing 9 (2013) 616–628

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

Identifying and understanding urban sport areas using Nokia Sports Tracker

Laura Ferrari *, Marco Mamei

Dipartimento di Scienze e Metodi dell'Ingegneria, University of Modena and Reggio Emilia, Italy

ARTICLE INFO

Article history:
Available online 2 November 2012

Keywords:
Mobility pattern
Urban and environmental planning
Social dynamics
Web mining
Mobile applications

ABSTRACT

Current advancements in pervasive technologies allow users to create and share an increasing amount of whereabouts data. Thus, some rich datasets on human mobility are becoming available on the web. In this paper we extracted approximately 790,000 mobility traces from a web-based repository of GPS tracks—the Nokia Sports Tracker Service. Using data mining mechanisms, we show that this data can be analyzed to uncover daily routines and interesting schemes in the use of public spaces. We first show that our approach supports large-scale analysis of people's whereabouts by comparing behavioral patterns across cities. Then, using Kernel Density Estimation, we present a mechanism to identify popular sport areas in individual cities. This kind of analysis allows us to highlight human-centered geographies that can support a wide range of applications ranging from location-based services to urban planning.

1. Introduction

The recent diffusion of smart phones equipped with localization capabilities provides a practical way to collect geo-located information from a large user population [1]. The analysis of such collected data is a fundamental asset in the development of pervasive and mobile computing applications [2].

In this paper we analyze a large dataset of mobility tracks related to sport activities. In particular, we extracted approximately 790,000 mobility traces from a web-based repository of GPS tracks—the Nokia Sports Tracker service. User-generated mobility data, labeled with the sport activity the user was performing, allows us to identify both the areas of the city most "used" for a given kind of activity, and the temporal patterns and routines in which people use the above areas. The extracted areas and information could be accessed from a mobile application to support a variety of location-based applications: where are my sport pals? What are the best areas in the city to practice sport? In addition, since this kind of data describes how the city is used by some people it can also provide useful information to urban planners and local governments: are bike routes well positioned and used? At what time do people practice sport? Is it compatible with the opening time of sport-related facilities?

In order to concretize the depicted scenario, it is fundamental to develop data analysis tools to extract high-level meaningful information from raw mobility traces. In this direction, we make the following contributions:

1. We show that simple statistical analysis on people mobility allows us to highlight cultural and climate-related differences among cities. This information can be derived from the different temporal patterns with which users perform sport activities at different time scales. While other works focus on sport-related datasets on selected cities [3–7] (see the Related Work section), our analysis allows us to compare multiple cities on the large scale.

* Corresponding author.
E-mail address: laura.ferrari@unimore.it (L. Ferrari).

1574-1192/$ – see front matter
doi:10.1016/j.pmcj.2012.10.006

2. We show that data analysis at the city-level unveils both temporal and spatial city patterns. For this purpose, we developed a set of novel mechanisms to apply Kernel Density Estimation to our data. This allows us to efficiently analyze city changes over time and space. Areas being identified describe the human-centered geography of the city with regard to sport activities. In contrast with related works [8,9], our approach does not rely on hard-coded parameters. Thus it can be seamlessly applied to cities of different size and geography.
3. We present some novel applications of the above technique. On the one hand, we show that the areas identified by the kernel density estimation can be analyzed in comparison with the city infrastructures and the main city landmarks (e.g., parks, cycle routes) to assess and highlight their impact on the way in which people "use" the city. On the other hand, we show that such areas can be partitioned among groups of users to highlight differences in the routine behavior of various demographic/social communities. While the state of the art for this kind of analysis relies on survey data, the proposed approach is fully automatic and data-driven.

Although our discussion is focused on sport activity, we believe that the proposed mechanisms and analysis are rather general and could be fruitfully applied to other data to discover other kinds of human-centered areas and routine behaviors in cities.

The remainder of this paper describes the methods and the results obtained w.r.t. the above contributions. In particular: in Section 2 we illustrate related works in the area and discuss how the presented work compares with them. Section 3 describes the Nokia Sports Tracker dataset and presents mechanisms to highlight cultural and climate-related differences across cities. Section 4 discusses the proposed Kernel Density Estimation approach that allows us to identify areas in cities related to sport activities. Moreover, we describe the design of our algorithms via the MapReduce framework in order to parallelize computations. Section 5 presents some analysis and experiments we conducted to validate our data and results. Section 6 presents some exemplary applications of the proposed approach. Finally, Section 7 concludes.

2. Related work

The problem of discovering and understanding human-centric city dynamics has recently been addressed in the research community, mostly through the use of mobile phone tracking, and of tagged and geo-referenced user-contributed content. Mainstream approaches can be divided into two groups on the basis of which kind of data they use. Some works are based on *dedicated* datasets collecting data from a number of individuals recruited for the study. Other works are based on *third-party* repositories that were collected for different purposes (like the NST dataset used in this paper). The main problem with the first kind of dataset is that it is difficult and expensive to collect enough data to study large-scale mobility patterns. The main problem with the second kind of dataset is the lack of control on data acquisition and the fact that groundtruth information is often missing.

Dedicated repositories. Among a wide range of small datasets collected for specific research studies, some recent works stand out for acquiring data from a large number of people over a long period of time.

The Geolife project [10] is based on a large set of GPS trajectories collected in the (Microsoft Research Asia) Geolife project by 165 users in a period of over two years. Using such traces and users' annotations, it is possible to discover interesting locations and possible activities that can be performed there.

The Reality Mining project (http://reality.media.mit.edu) is based on a large dataset of GSM-based localization collected from 100 subjects at MIT over the course of one year. Several researches have been conducted over this data to discover routine behaviors on users' whereabouts [11].

Similar to our work, these projects analyze mobility tracks to identify relevant areas of the city and to discover recurrent mobility patterns. Relying on a wider web-based repository, our analysis is able to capture human-centered patterns and geographies on a wider scale.

Third-party repositories. An exemplary project in this category is represented by the works on Telecoms' data by the MIT Senseable city group (http://senseable.mit.edu). This project employs extremely large anonymized GSM-based localization data coming from Telecom operators [12]. The goal is to extract the spatio-temporal dynamics of the city, highlighting where people usually go during the day. In another work from the same group [13], the authors were able to identify the most visited areas by tourists and the typical time of the visit.

Similar to our work, these projects analyze mobility tracks to extract the human-centered geography of the city and to infer patterns and routine behaviors. These works tap into extremely large datasets, however none of them get the precision of GPS localization.

Another group of works is based on photo-sharing sites which contain billions of publicly accessible images taken virtually everywhere on earth. These photos are annotated with various forms of information including geo-location that implicitly identifies the location of the user. Researchers have been able to analyze a global collection of geo-referenced photographs, and evaluate them on nearly 35 million images taken from Flickr [14] with the goal of identifying hot spots and tourist routine behavior. A similar work on the basis of geo-localized social networks is presented in [15].

Sport-based third-party repositories. The works that are most directly compared to our proposal are [3–7]. The work in [3] analyzes traces gathered from the NST repository to infer patterns and sport-related areas in a few selected cities. A similar analysis of collective mobility patterns in an urban environment is done for instance in [4–7] by tracking bicycles at the