Pervasive and Mobile Computing 42 (2017) 77–92

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

An approach for the discovery and validation of urban mobility patterns

Eugenio Cesario [a,*], Carmela Comito [a], Domenico Talia [b]

[a] ICAR-CNR, Italy
[b] DIMES, UNICAL, Italy

ARTICLE INFO

Article history:
Received 16 November 2016
Received in revised form 14 July 2017
Accepted 20 September 2017
Available online 7 October 2017

Keywords:
Trajectory pattern mining
Trajectory similarity
Urban computing

ABSTRACT

The increasing pervasiveness of mobile devices favors the collection of large amounts of movement data that can be analyzed to extract knowledge, i.e. patterns, rules and regularities, from user trajectories. In this paper we present TPM, an integrated algorithm which supports the overall trajectory pattern discovery process for detecting user's mobility behaviors. Specifically, the algorithm includes two main phases: (i) finding dense regions, more densely passed through ones; (ii) extracting trajectory patterns from those regions. Another contribution of the paper is a validation methodology for assessing the effectiveness of the TPM algorithm, e.g., evaluating how the discovered knowledge model fits to the input data it is discovered from. Such methodology represents a general solution that can be used to evaluate the accuracy of any algorithm aiming at extracting dense regions and trajectory patterns from GPS data. Furthermore, we propose novel trajectory similarity measures to evaluate the quality of the extracted patterns. A detailed experimental evaluation, performed by exploiting the proposed validation process, proves the efficiency and effectiveness of TPM.

1. Introduction

The widespread presence of wireless technologies, as well as the rapid development of satellite-enabled Global Positioning System (GPS) and mobile computing technologies, are providing large amounts of data pertaining to the mobility of people, cars, buses, trucks, etc. in urban areas. In particular, these networking infrastructures allow for sensing and collecting massive amount of spatio-temporal data. For instance, mobile phones leave positioning logs, which specify the cell they are connected in a cellular network. Analogously, GPS-equipped portable devices can record the latitude–longitude position everytime they are exposed to GPS satellites. These scenarios lead to the generation of a large number of trajectories drawn by mobile users during their daily activities allowing, thus, the collections of large amounts of movement data. The use of this data opens the opportunity of making discoveries about movement habits expressed as patterns, rules and regularities.

Analysis of people trajectories is a typical application of *urban computing* [1], an emerging multi-disciplinary research field that focuses on computing and digital networks in urban landscapes (e.g., cities, parks suburbs, etc.). The final goal consists in having smarter cities, by the application of ubiquitous and pervasive computing paradigms to urban spaces focusing on developing innovative services for citizens. In particular, traffic management is considered as one of the most challenging issues in urban areas and a great effort is being made to create more robust and intelligent systems that can be used to meet traffic flow needs.

* Corresponding author.
E-mail address: eugenio.cesario@icar.cnr.it (E. Cesario).

https://doi.org/10.1016/j.pmcj.2017.09.006
1574-1192/

1.1. Motivations and contributions

Several approaches for trajectory pattern mining have been proposed in literature. Most of them are based on a combination of density-based clustering and sequential pattern discovery concepts, which are extensively used for detecting patterns from mobility data [2–7]. The general approach used to retrieve trajectory patterns consists of (i) discovering hotspots (e.g., single points or bounded areas) of interest and (ii) extracting mobility patterns among them. We refer to such an approach as *Density-based Sequential Pattern Mining* (DSPM).

The main contributions of the work proposed in this paper are summarized as follows:

- The *Trajectory Pattern Miner* algorithm, TPM. TPM implements the DSPM approach through an integrated algorithm which supports the overall trajectory patterns discovery process. The two main steps of the approach are: (i) finding frequent regions, and (ii) extracting sequential trajectory patterns among those regions.
- *A comprehensive validation methodology* for assessing accuracy and quality of detected dense regions and trajectory patterns, which consists of three steps: (i) generating synthetic trajectory data, adhering to some pre-defined mobility patterns and tuned by several input parameters, (ii) running a DSPM algorithm on such data to discover mobility models and (iii) finally computing how much the discovered models (i.e., dense regions and trajectory patterns) adhere to the characteristics of the input dataset they have been generated from.
- *Novel trajectory similarity measures* to evaluate the quality of the extracted patterns. Differently from the approaches in literature, such metrics are specifically tailored to trajectories among regions and not just GPS points. Thus, they are effective in providing the real geographic overlap among the regions involved in the mined trajectory patterns.

The proposed validation methodology represents a general solution that can be used to evaluate the accuracy of any of the algorithms implementing the DSPM approach introduced above, also including other kind of geo-referenced data like geo-tagged data from social networks (e.g., Twitter and Facebook). To the best of our knowledge, there are no other works in the literature proposing a comprehensive approach to validate both trajectory patterns and dense regions extracted from a GPS dataset.

For the sake of clarity, this paper largely extends the work presented in [8] and it provides several original contributions with respect to the previous one, as summarized in the following. First, the description of the trajectory pattern mining algorithm has been extended by enhancing it with formal definitions and a meta-coding of the algorithm (Section 3). Moreover, it describes a real-case study and shows the results of the algorithm on a real-world dataset. Then, the validation methodology (Section 4) has been extended by introducing further trajectory pattern similarity measures and metrics. Finally, several original contribution concerning the experimental evaluation (Section 6) have been reported in this paper.

1.2. Plan of the paper

The paper is structured as follows. Section 2 reviews related work on sequential pattern mining algorithms. Section 3 describes TPM, the proposed algorithm to discover trajectory patterns from large collections of moving object data, and the results of the analysis carried on T-Drive which we exploit as case study. The methodology introduced to validate the quality of the discovered patterns is detailed in Section 4. Section 5 presents novel similarity metrics designed to measure trajectory pattern similarity. Section 6 provides a detailed evaluation of the accuracy and effectiveness of the TPM algorithm, performed on several synthetic datasets. Finally, Section 7 concludes the paper and proposes some further research issues.

2. Related work: Density-based sequential pattern mining

Discovering patterns from historical object movements is a very challenging task and several algorithms tackle the problem by adopting sequential pattern mining approaches. Among all, particularly relevant are those based on a common inspiring idea that first detects geographic dense regions and then extracts sequential mobility patterns among such regions. As aforementioned in Section 1, we refer to such an approach as *Density-based Sequential Pattern Mining* (DSPM). In the following we review some representative algorithms of the DSPM approach, also sketching the main differences and similarities with the proposed TPM algorithm.

Mamoulis et al. [5] address the problem of mining sequential patterns from spatio-temporal data by considering the patterns in the form of trajectory segments. They first decompose the original trajectories into segments, then group them according to their shape and closeness. In particular, authors define the spatio-temporal periodic pattern mining problem and propose an algorithm for retrieving maximal periodic patterns based on a tree structure and an Apriori paradigm. In further extensions [6] and [9] of this work, authors study how to mine frequent sequences of trajectory segments and search approximate instances in the data representing approximate object movements over time.

In [2] authors extend the sequential pattern mining methodology to analyze moving objects, by proposing a method for extracting patterns containing both spatial and temporal information. In particular, a trajectory pattern describes movement trends in both spatial and temporal contexts, based on RoI (Region of Interest) detections. The approach first identifies RoIs, which are mostly visited regions, then, find frequent patterns from sequences of regions of interest. The sequences of RoIs not only represent the spatial movements but also the travel time of the moving objects. A similar approach has been presented in [3,4], where a trajectory pattern mining task is modeled as an extension of the association rules concept. In this case