Information Sciences 474 (2019) 18–32

Contents lists available at ScienceDirect

Information Sciences

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

Discovering individual movement patterns from cell-id trajectory data by exploiting handoff features

Mingqi Lv[a], Ling Chen[b,*], Tieming Chen[a], Dajian Zeng[a], Bin Cao[a]

[a] College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China
[b] College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China

ARTICLE INFO

Article history:
Received 11 November 2017
Revised 10 September 2018
Accepted 17 September 2018
Available online 20 September 2018

Keywords:
Movement pattern mining
Cell-id trajectory
Handoff feature
Pattern loss problem
Mobile phone user

ABSTRACT

The discovery of movement patterns from trajectory data is crucial for supporting many location-based applications. Most existing methods require that the trajectories contain explicit location information. However, it is usually difficult to collect such kind of trajectories from mobile phone users. In this paper, we propose a method for mining movement patterns from cell-id trajectories (i.e., sequences of cell tower identifiers) without explicit location information. Specifically, we firstly estimate the spatial closeness between cell towers in a cell-id trajectory dataset by exploiting the handoff features. Then, we propose a novel sequential pattern mining algorithm to mine movement patterns from the cell-id trajectory dataset by taking into account the estimated spatial closeness. We evaluated the proposed method based on a real cell-id trajectory dataset. The experiment results show that the proposed method can adapt to the high uncertainty of cell-id trajectories and it outperforms state-of-the-art methods in terms of efficiency, completeness, and usefulness.
© 2018 Elsevier Inc. All rights reserved.

1. Introduction

People's daily trajectories show a high degree of temporal and spatial regularity [15], and thus mining movement patterns from individual's trajectories has become a hot research topic [26]. The movement patterns could provide in-depth knowledge to understand users and foster a number of location-based applications, e.g., future route prediction [17], trip planning [14], transportation planning [9], itinerary recommendation [40], etc.

The prerequisite of movement pattern mining is to collect trajectory data and most existing methods use GPS for trajectory data collection [26]. However, GPS has several limitations: First, GPS is extremely power hungry [30]. Second, the performance of GPS is unstable in urban environment and it suffers from prolonged lock-on periods when used on demand [37]. These limitations constrain the feasibility of using GPS for long-term trajectory data collection. On the other hand, every active mobile phone is connected to a cell tower, and the identifier of the connected cell tower (i.e., *cell-id*) can be obtained with ignorable cost. In addition, cellular signal is more robust in urban environment than GPS signal. Hence, collecting trajectory data using cellular signal is more feasible for mobile phone users. By using cellular signal, a mobile phone user's trajectory can be collected and represented as a sequence of cell-ids (i.e., *cell-id trajectory*).

Cell-id trajectory data can be collected either server-side or client-side. As for server-side collection, cellular network providers routinely collect users' call detail records (CDRs), which contain various attributes, e.g., call duration, call type, and cell-id [5]. However, CDRs have the following constraints: First, CDRs are very sparse, since they are collected only

* Corresponding author.
E-mail addresses: mingqilv@zjut.edu.cn (M. Lv), lingchen@cs.zju.edu.cn (L. Chen), tmchen@zjut.edu.cn (T. Chen), bincao@zjut.edu.cn (B. Cao).

https://doi.org/10.1016/j.ins.2018.09.033
0020-0255/© 2018 Elsevier Inc. All rights reserved.

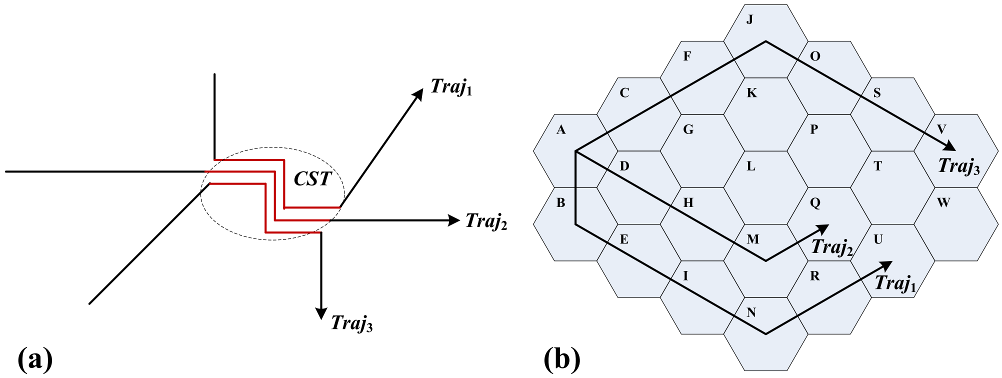

Fig. 1. (a) An illustration of the movement pattern definition; (b) an illustration of the pattern loss problem (the character inside a hexagon stands for the cell-id of the cell tower).

when a mobile phone engages in a voice call or text message exchange. Second, CDRs are possessed by cellular network providers and not available to mobile phone users. Thus, CDRs are usually used for coarse-grained city-scale analysis, and not suitable for mining individual users' movement patterns. In contrast, client-side collection uses mobile phones to collect users' own cell-trajectory data [6], so it is free of these problems: First, mobile phones can record cell-id trajectory data in a periodic manner, no matter voice calls or text message exchanges occur or not. Second, cell-id trajectory data can be stored and analyzed locally on mobile phones. Therefore, it is more feasible to mine individual users' movement patterns based on client-side collected cell-id trajectory data. Then, the users could choose whether or not to reveal the results to third-party service providers. A potential application for movement pattern mining based on client-side collected cell-id trajectory data is intelligent personal assistant (e.g., Google Now) [34], which provides proactive services by understanding users' personal activity patterns from long-term client-side collected data.

Based on this motivation, some methods have been dedicated to mining individual movement patterns from client-side collected cell-id trajectory data [3,16,22,27,28]. According to the definition of movement pattern, these methods could be categorized into two types: clustering based methods and sequential pattern mining based methods. Clustering based methods define movement patterns as trajectory clusters [16,22,28], and sequential pattern mining based methods define movement patterns as common sub-trajectories shared by multiple trajectories [3,27]. The latter definition could lead to discovering more movement patterns. As shown in Fig. 1(a), clustering $Traj_1$, $Traj_2$, and $Traj_3$ as a whole could not detect movement patterns. However, the three trajectories share a common behavior (i.e., *CST*), which could be potentially discovered by sequential pattern mining. In this paper, we use common sub-trajectories as the movement pattern definition.

The major challenge of mining individual movement patterns from client-side collected cell-id trajectory data is how to process trajectories without explicit location information. Someone may argue that the explicit location information of a cell-id trajectory could be roughly estimated based on the location information of the cell towers. However, it is usually difficult for individual users to access to the location information of cell towers in practice: The location information of cell towers is possessed by cellular network providers and not freely available to the public. Although there are a few open databases (e.g., OpenCellID), the coverage is highly limited, especially in the developing countries.

Aiming at this challenge, a cell-id trajectory could be treated as a string (i.e., each unique cell-id is treated as a distinct character), and then individual movement patterns could be extracted as substrings or subsequences [27,38]. However, this strategy has a problem: The spatial closeness between cell towers is ignored, so it would cause severe pattern loss problem. As shown in Fig. 1(b), $Traj_1$, $Traj_2$, and $Traj_3$ move in an area covered by multiple cell towers. The three cell-id trajectories have almost totally different string representations: $Traj_1$ corresponds to "ABEINRU", $Traj_2$ corresponds to "ADHMQ", and $Traj_3$ corresponds to "ACFJOSV". Thus, existing methods could not find any movement pattern from them. However, since $Traj_1$ and $Traj_2$ share a common route covered by cell towers that are spatially close to each other, they should imply a movement pattern moving from the left area to the bottom area. With more cell towers deployed in urban area, the problem would become even more serious.

To address this problem, we propose a method for mining individual movement patterns from cell-id trajectories by taking the spatial closeness between cell towers into account. First, the spatial closeness cannot be calculated directly, as the explicit location information of cell towers is unavailable. For this problem, we exploit the handoff features to estimate the spatial closeness by using a regression based method. A handoff indicates the situation that the mobile phone's connected cell tower switches from one to another, and we have found that the handoff features between a cell tower pair could indirectly reflect their spatial closeness. Second, even if the spatial closeness is known, the existing sequential pattern mining algorithms (e.g., PrefixSpan [31] and Apriori [1]) are unable to leverage this information, because they work upon sequential data with nominal elements, and cannot adapt to sequential data with elements that have mutual correlation measured by a continuous value. Thus, by integrating the estimated spatial closeness into a database projection based approach, we design a new sequential pattern mining algorithm, called SCPG (spatial closeness based pattern generation), and utilize the

Download English Version:

https://daneshyari.com/en/article/11021153

Download Persian Version:

https://daneshyari.com/article/11021153

Daneshyari.com