Information Sciences 463–464 (2018) 110–128

Contents lists available at ScienceDirect

Information Sciences

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

Predicting attributes and friends of mobile users from AP-Trajectories☆

Pinghui Wang[a,b], Feiyang Sun[a,1,*], Di Wang[a], Jing Tao[a,c,*], Xiaohong Guan[b,a,d,*], Albert Bifet[e]

[a] NSKEYLAB, Xi'an Jiaotong University, China
[b] Shenzhen Research School, Xi'an Jiaotong University, China
[c] Zhejiang Research Institute, Xi'an Jiaotong University, China
[d] Department of Automation and NLIST Lab, Tsinghua University, China
[e] LTCI, Télécom ParisTech, Université Paris-Saclay, France

A R T I C L E I N F O

Article history:
Received 3 July 2017
Revised 11 June 2018
Accepted 12 June 2018
Available online 14 June 2018

Keywords:
User profiling
Spatiotemporal trajectories
Social network

A B S T R A C T

Exploring the demographic attributes and social networks of Internet users is widely employed by many applications, such as recommendation systems. The popularity of mobile devices (notably smartphones) and location-based Internet services (e.g., Google Maps) facilitates the collection of users' locations over time. There have been recent efforts to predict users' attributes (e.g., age and gender) from this data, and location-based social networks such as Foursquare and Gowalla are based on using the rich location context knowledge of points of interest (e.g., the name, type and description of restaurants and hotels) where users check-in online. However, little attention has been paid to inferring the attributes and social networks of mobile device users based on their spatiotemporal trajectories where there is little or no location context knowledge. In this paper, we collect logs of thousands of mobile devices' network connections to wireless access points (APs) of two campuses, and investigate whether one can infer mobile device users' demographic attributes and social networks solely from their spatiotemporal AP-trajectories. We develop a tensor factorization-based method ***Dinfer*** to infer mobile device users' attributes from their AP-trajectories by leveraging prior knowledge. Compared with our previous work, which only considered users' social networks, Dinfer further utilizes AP spatial information and achieves a 2% improvement. We also propose a novel method ***Sinfer*** to learn social networks between mobile device users by exploring patterns of their AP-trajectories, such as fine-grained co-occurrence events (e.g., co-coming, co-leaving, and co-presenting duration). Experimental results on real-world datasets demonstrate the effectiveness and efficiency of our methods.

☆ An earlier conference version appeared as a WWW'17 companion paper [27]. In this journal version, we extend the proposed Dinfer method to incorporate AP spatial closeness information to improve its accuracy. Moreover, we conduct more experiments to validate our proposed methods, include user department inference, grade inference, and class inference, and compare Dinfer with more classic methods, e.g., PCA and HOSVD.

* Corresponding authors.
E-mail addresses: fysun@sei.xjtu.edu.cn (F. Sun), xhguan@sei.xjtu.edu.cn (X. Guan).

[1] P. Wang and F. Sun contributed equally to this work.

https://doi.org/10.1016/j.ins.2018.06.029

1. Introduction

In today's era of digital big data, information on human activities has been collected and monitored in a variety of domains. For example, people's webpage browsing, shopping, video viewing and music listening behaviors on the Internet can easily be collected by providers of web services such as search engines and e-commerce, video and music websites. This data helps service providers to understand/profile their users, which is crucial for many applications such as targeted advertising and business recommendations, as well as for counter-criminal/terrorist strategies. Utilization of this data has revealed that service providers can infer their users' demographic attributes, including age and gender, even though users do not intend to reveal these attributes. The information is derived by the service providers from users' Internet browsing history [21], linguistics writing [11], mobile call/message records [8,9], music listening history [18], purchase data [26], and available demographic attributes of their friends on online social networks (OSNs) such as Facebook and LinkedIn [20].

Recently, mining people's locations over time has attracted a lot of attention due to the popularity of mobile devices along with location-based Internet services that include navigation applications (e.g., Google Maps and Uber) and location-based online social networks (location-based OSNs, e.g., Instagram and Foursquare). Smartphones, smartwatches and other mobile devices can be precisely located outside or inside buildings by current and near-future positioning techniques, which precisely infer the locations of these devices using multiple information sources that include built-in sensors (e.g., GPS and barometers) and the location and signal strength of connected wireless access points (APs) and cell-phone towers. These techniques lead to mobile device users' spatiotemporal trajectories being explicitly or implicitly exposed to third parties. Explicit exposure occurs when people enable map services to retrieve their GPS locations automatically, distribute geo-tagged posts (e.g., tweets and photos) or check-in at points of interest (POIs) on OSNs. Implicit exposure of users' trajectories can take place even though users do not intend to reveal their locations; for example, banks and mobile phone service providers can learn users' spatiotemporal trajectories from collected credit card and telecommunication transactions respectively. The collection of a large number of users' spatiotemporal trajectories facilitates the study of research topics such as urban planning, congestion prediction and POI recommendations in smart cities.

In this paper, we are interested in testing whether one can infer mobile device users' demographic attributes and social networks from their spatiotemporal trajectories. Existing work [30] derives users' attributes from their check-in POIs on location-based OSNs such as Facebook, Foursquare and Yelp. The method relies on POIs' rich semantic features such as categories, user reviews and descriptions; however, in practice, the semantic features of locations may not be publicly available. For example, the daily activities of most college students and faculty staff may be concentrated in campuses. Third parties now, or in the near future, may be able to collect fine-grained ***indoor*** spatiotemporal trajectories of persons on campus but fail to get the fine-grained context (e.g., library or gym) of a place on campus, because a building may have different functional areas (e.g., research lab and restaurant) and the context of each building or place on campus may not be publicly available on the Internet. In addition to demographic attributes, we also studied the problem of predicting who mobile device users' close friends are from their spatiotemporal trajectories. Existing methods [4–6,23,25,29] mine friendships and infer social strengths mainly from co-coming events between users, e.g., check-in records on OSNs, because most location-based OSNs record the time of users' check-ins but not check-outs (in practice, users check-in at POIs on location-based OSNs but never check-out).

To the best of our knowledge, our work is the first attempt to infer users' attributes by utilizing AP-trajectories with AP spatial information instead of location context knowledge. Compared with our previous work [27], introducing locations' closeness makes the model more accurate. We studied three types of fine-grained co-occurrence events: co-coming, co-leaving, and co-presenting duration, and observed that these fine-grained events are effective and complementary to each other for predicting friendships between mobile device users.

Our contributions are summarized as:

- We conducted an in-depth measurement study on spatiotemporal trajectories of 52,000 mobile devices on two campuses, which were revealed by their network connections to wireless APs. To facilitate further research in this field, we made these trajectories publicly available[2].
- We developed a tensor factorization-based learning method *Dinfer* to infer mobile device users' attributes from their AP-trajectories by leveraging AP spatial information and user social networks that are learned from users' spatiotemporal trajectories.
- We used pointwise mutual information (PMI) to evaluate whether a fine-grained co-occurrence happens by chance or is a social event, and propose an effective method *Sinfer* to learn social networks of mobile device users by exploring their fine-grained co-occurrence events.

The rest of this paper is organized as follows. Section 2 describes our datasets, consisting of AP-trajectories of mobile devices on campus and the ground truth of mobile device users' social networks and demographic attributes. Section 3 presents our measurement study of micro- and macro-statistics of mobile devices' AP-trajectories. Section 4 formulates the problem and the overview of our methods. Sections 5 and 6 present our methods Dinfer and Sinfer for learning demographic attributes and social networks of mobile device users from their spatiotemporal AP-trajectories respectively.

[2] http://nskeylab.xjtu.edu.cn/dataset/phwang/data/APtraj.zip.