Neurocomputing 278 (2018) 99–109

Contents lists available at ScienceDirect

Neurocomputing

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

A hybrid Markov-based model for human mobility prediction

Yuanyuan Qiao [a,*], Zhongwei Si [a], Yanting Zhang [a], Fehmi Ben Abdesslem [b], Xinyu Zhang [a], Jie Yang [a]

[a] *Beijing Key Laboratory of Network System Architecture and Convergence, and Beijing Laboratory of Advanced Information Networks, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[b] *Decisions, Networks and Analytics Laboratory, SICS Swedish ICT AB, Kista SE-164 29, Sweden*

ARTICLE INFO

Article history:
Received 1 December 2016
Revised 9 May 2017
Accepted 21 May 2017
Available online 5 September 2017

Keywords:
Non-Gaussian mobility data
Hybrid Markov-based model
Human mobility
Mobility prediction
Spatio-temporal regularity

ABSTRACT

Human mobility behavior is far from random, and its indicators follow non-Gaussian distributions. Predicting human mobility has the potential to enhance location-based services, intelligent transportation systems, urban computing, and so forth. In this paper, we focus on improving the prediction accuracy of non-Gaussian mobility data by constructing a hybrid Markov-based model, which takes the non-Gaussian and spatio-temporal characteristics of real human mobility data into account. More specifically, we (1) estimate the order of the Markov chain predictor by adapting it to the length of frequent individual mobility patterns, instead of using a fixed order, (2) consider the time distribution of mobility patterns occurrences when calculating the transition probability for the next location, and (3) employ the prediction results of users with similar trajectories if the recent context has not been previously seen. We have conducted extensive experiments on real human trajectories collected during 21 days from 3474 individuals in an urban Long Term Evolution (LTE) network, and the results demonstrate that the proposed model for non-Gaussian mobility data can help predicting people's future movements with more than 56% accuracy.

1. Introduction

Analyzing the characteristics of human mobility reveals that human trajectories are predictable. They often exhibit a high degree of temporal and spatial regularity. In order to find the basic rules governing human dynamics and build a model to predict human mobility, various studies of human mobility have been conducted in recent years. With the emergence of smartphones and location-based services, since most of location-based services require accurate or approximate position of user, predicting user's next locations shows a great potential for service providers to improve the user experience. In particular, it has become a critical enabler for a wide range of applications, such as location-based advertising, early warning systems, and city-wide traffic planning [1].

Over the previous years, different methods have been proposed in the literature, using varied types of data and aiming at predicting distinct aspects of human mobility. Real traces are crucial to train and evaluate prediction models. Nowadays, people's movements can be easily sensed with mobile phones, which are generating large volumes of mobility data, such as Call Detail Records (CDRs) [2,3], Global Positioning System (GPS) tracks [4–7], data traffic of mobile networks [8–11], and Wi-Fi access points data [12]. Recently, researchers found that data traffic from 2G/3G/4G cellular networks is extremely useful for studying human dynamics [8,9,11,13–17], and can provide people's trajectories in a large scale. Passively collecting human movement trajectories while they access the Internet with their smartphone presents many advantages: high cost efficiency, low energy consumption, covering a wide range and a large number of people. Moreover, this can be done with a fine time granularity, as people tend to surf the Internet frequently on their smartphone while commuting. Also, many apps send or receive network traffic packets, even when running in the background [18]. Collected datasets, despite having different collection methods, and covering different populations with various accuracy or time granularity, always present a similar characteristic: human trajectories described by the data all follow a non-Gaussian spatio-temporal distribution. [18–22].

Existing research works on mobility prediction are exploiting diverse types of data [1]. The most commonly used algorithms to predict mobility include machine learning algorithms (clustering techniques [4–6,8,9,23–25], Bayesian models [26–30], neural networks [2,31]), state-based techniques (Markov models [7,12,14,15,32], LeZi family [11,13], hidden Markov models [4,5,10,25,33]), and pattern matching algorithms (prediction by

* Corresponding author.
E-mail addresses: yyqiao@bupt.edu.cn (Y. Qiao), fehmi@sics.se (F.B. Abdesslem).

http://dx.doi.org/10.1016/j.neucom.2017.05.101

Partial Matching [34,35]). As for predicting mobility in a large population through the cellular network, Markov-based algorithms are more suitable [36] and out-perform other methods when applied to short trajectories, and when considering the temporal factor [37]. They also prove to be very appropriate for future generation mobile networks [38].

To predict the next location using large-scale non-Gaussian mobility data, this paper aims to improve the prediction accuracy of Markov-based algorithms. In order to achieve this goal, the key is to enrich the states in the underlying Markov model by considering relevant external information, increasing time and spatial complexity. In this paper, a hybrid Markov-based prediction model, which considers non-Gaussian and spatio-temporal characteristics of real human mobility data while predicting, is constructed. The prediction accuracy of a simple Markov algorithm on our real dataset, can be improved from 44.02% to 56.39% in our experiments. Overall, the contributions of this paper can be summarized as follows:

- We examine the characteristics of real mobility data extracted from user's data traffic in an LTE network, which provide properties we should consider while predicting user's future movement. The human mobility represented in the dataset shows two main characteristics that follow a non-Gaussian distribution, namely the trip distance and the radius of gyration. This means that (1) displacement within short distance is frequently seen in the dataset and (2) frequent travels occur in a limited range in individuals' daily life. In other words, people generally move within a bounded region and only occasionally travel long distance. In addition, we further study the probability of finding a user at different locations, and returning to the same location. Analysis results show that people visit some primary locations periodically with high probability, which confirms the intuitive existence of mobility patterns for each user. The periodicity of this pattern hints that temporal factors can contribute to predicting the next location of individuals.
- We propose a hybrid Markov-based model to predict users' future movements. It uses different methods consecutively to discover spatio-temporal pattern of each individual's trajectory. More specifically, the model determines the order of Markov algorithm by discovering the regular mobility patterns for each individual, and takes into account the time of the day where locations are visited. This way, non-Gaussian and spatio-temporal characteristics of users' trajectories are fully considered, which contributes a lot to getting a better prediction result.
- Markov-based algorithms fail to correctly predict future movements if the new location has never been visited by an individual. To alleviate this issue, we consider the trajectories of geo-friends: users sharing similar trajectories and mobility patterns. We then employ a user-based recommendation method with Collaborative Filtering to predict users' future movements when their own mobility pattern cannot contribute to the prediction.

The remainder of this paper is organized as follows. In Section 2, we provide a survey of the related works. Section 3 examines the non-Gaussian and spatio-temporal characteristics of users' trajectories extracted from our dataset. In Section 4 we describe the hybrid Markov-based mobility prediction model, before presenting in Section 5 the methods employed in the mobility prediction model, including the hotspot detection method, the mobility pattern discovery algorithm, the variable-order Markov prediction algorithm with temporal factors, and finally the algorithm enhancement using similarity of geo-friends' movements. The performance of the proposed mobility prediction model is then evaluated in Section 6, and conclusions are drawn in Section 7.

2. Related work

Many datasets collected from real world applications and services have been found to show non-Gaussian characteristics, such as gene networks [39], hyper-spectral images [40], or climate extremes [41], for example. As for datasets describing human mobility, despite some characteristics showing Gaussian distributions [42,43], many other characteristics show non-Gaussian distribution [44,45] and have been studied in the literature. In particular, the locations visited by people in their daily life show a non-Gaussian distribution [19], and predicting those locations is a challenging objective addressed by recent research works to enable a wide range of applications, such as location-based advertising, early warning systems, and city-wide traffic planning.

Markov models are widely used in prediction algorithms, due to their efficiency, simplicity, and low computing costs [38]. For example, a Markov chain prediction model considers the sequence of locations last visited by a user to predict the next location. The length k of that sequence of locations represents the order of the Markov chain, and we refer to this model as an order-k Markov-based model. Markov-based prediction methods fall into one of four categories: (1) order-k Markov-based methods that only use the historical locations to discover individual movement patterns [7,46,47], (2) order-k Markov-based methods that also consider external information in addition to historical trajectories [48–50], (3) hybrid Markov-based methods that are enhanced with other prediction methods [32,51,52], (4) evolved algorithms based on Markov models, such as the LeZi algorithm [11,13] or algorithms based on Hidden Markov Models [4,5,10,25,33].

Over a decade ago, classical Markov models of low order (generally, 1 or 2) have already been used to predict future movements [7,46,47]. It has been found that low-order Markov-based predictors performed as well as, or better than the more complex and more space-consuming compression-based predictors such as Prediction by Partial Matching (PPM), or Sampled Pattern Matching Algorithm (SPM) [46]. Markov models of higher order, combined with external information about people's schedule, have also been used to improve the prediction accuracy. An order-k Markov-based predictor with fall-back was proposed in [48]. The predictor proposed by the authors falls back to a lower order of k if a certain order-k predictor was unsuccessful. In addition, they also incorporate external information, such as the event notes on Microsoft Outlook and Google Calendar, into the context of an $O(k)$ Markov predictor, enriching the states in the underlying Markov model. In other works, the distances of trajectories, travel times, driving habits, and social context have been incorporated into corresponding models to compute the probabilities of future locations in [49,50,53,54]. Recently, it has been found that different predictors should be applied according to the application scenario, to the mobility characteristics of the data, and to individuals [55]. To this aim, hybrid methods have been proposed, which uses different prediction methods consecutively [51,52,56,57], for different datasets [32], for each individuals [14,55], or even for different times of the day [58].

Authors in [51] present a prediction model in LTE networks that combines two complementary algorithms: the global profiles-based and the local profiles-based algorithm. The former is implemented in the enhanced Node B and the home enhanced Node B, and the latter works at the user terminal level. By using a real vehicle passage record dataset, the authors in [32] present the Global, Personal, and Regional Markov Model for a dataset with different granularity to tackle the problem of predicting next locations for vehicles' trajectories. In [58], the authors proposed a prediction methods to capture the relationship between movement patterns in different time periods, which are based on the important observation that movement patterns often change over time. In addition,