Pervasive and Mobile Computing 20 (2015) 13–28

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

A probabilistic kernel method for human mobility prediction with smartphones

Trinh Minh Tri Do [a,*], Olivier Dousse [b], Markus Miettinen [b], Daniel Gatica-Perez [a,c]

[a] Idiap Research Institute, Martigny, Switzerland
[b] Nokia Research Center, Lausanne, Switzerland
[c] EPFL, Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 13 January 2014
Received in revised form 27 August 2014
Accepted 6 September 2014
Available online 16 September 2014

Keywords:
Smartphone
Human mobility
Prediction
Personalized service

ABSTRACT

Human mobility prediction is an important problem that has a large number of applications, especially in context-aware services. This paper presents a study on location prediction using smartphone data, in which we address modeling and application aspects. Building personalized location prediction models from smartphone data remains a technical challenge due to data sparsity, which comes from the complexity of human behavior and the typically limited amount of data available for individual users. To address this problem, we propose an approach based on kernel density estimation, a popular smoothing technique for sparse data. Our approach contributes to existing work in two ways. First, our proposed model can estimate the probability that a user will be at a given location at *a specific time in the future*, by using both spatial and temporal information via multiple kernel functions. Second, we also show how our probabilistic framework extends to a more practical task of location prediction for *a time window in the future*. Our approach is validated on an everyday life location dataset consisting of 133 smartphone users. Our method reaches an accuracy of 84% for the next hour, and an accuracy of 77% for the next three hours.

1. Introduction

The advances in mobile sensing and computing have enabled the integration of machine learning into personal mobile devices. In particular, smartphones emerge as all-purpose devices with personalized services, where the personalization is based on what the smartphone knows about the user. Smartphones can unobtrusively collect data about where users go and what they do, and build a detailed understanding of the user. First, the recorded data can be used to characterize multiple aspects of the user including demographic information [1,2] or personality [3]. Second, clustering methods can be applied to extract recurrent user contexts such as commonly visited places [4], providing a high level representation of context (instead of raw measurements). Finally, along with extracting and organizing information from the past, the phone can also learn a behavior model that can predict future activities and venues.

Location prediction can benefit mobile applications and services by letting the applications adapt to possible movements of the user. This can help a mobile device, for instance, to adapt its user interface based on the anticipated locations that the user will visit during the course of a day. As one example, it can prefetch and display relevant information related to the

* Corresponding author. Tel.: +41 27 721 77 51.
E-mail addresses: tri.do@idiap.ch, do@idiap.ch (T.M.T. Do), olivier.dousse@here.com (O. Dousse), markus.miettinen@trust.cased.de (M. Miettinen), gatica@idiap.ch (D. Gatica-Perez).

http://dx.doi.org/10.1016/j.pmcj.2014.09.001
1574-1192/

predicted target locations. Note that, for such scenarios, personalization is key as the interest does not lie in predicting the places that people are likely to visit, but rather in anticipating the movements of a *single user*. Also, since location traces of users are highly privacy-sensitive, it is not desirable to rely on a solution that requires location traces to be aggregated to a central data storage. Therefore, the prediction method has to be such that it relies only on the context history of a specific user for whom the prediction will be made.

Previous studies on mobility prediction have usually focused on predicting the next place where a user goes [5]. However, in practice, the prediction capability needs to go beyond the anticipated next place of the user, and instead provide predictions for different look-ahead periods. This is because applications might want to provide different kinds of information, depending on how imminent the user's visit to a location will be. For example, an application may want to prefetch traffic information for the route to a place that the user is predicted to visit during the next few hours, but for visits farther away in the future (during the next 24 h) other information like a weather forecast for the target location might be more appropriate. With these requirements in mind, we develop a flexible prediction method which can predict user location for a given timestamp or for different look-ahead time windows.

The mobility prediction problem can be formalized as a contextual prediction problem where the future movements are assumed to depend only on the user context, which is characterized by space and time in this paper. The assumption is based on the repetitive nature of human mobility: similar contexts might imply similar movements in the future. For example, from the mobility traces of a given user, one might observe that if he is at a given train station around 8:00 AM on Monday then he will likely be at work around 8:25 AM. Under a probabilistic framework, the location prediction task consists on estimating the conditional distribution over the set of future location candidates for a specific context, based on mobility history. This can be modeled by representing the user context as a combination of discrete states (e.g., at place X at hour Y on day Z), and so the conditional distribution is proportional to the counts of possible outputs for the considered context. This approach, however, suffers from a major issue with discretization: the relationship between states are lost. For example, if we discretize the time of day into 24 time slots by hour, then 7:59 AM and 8:00 AM belong to two completely different time slots, while they are actually very close. We resolve this problem by using kernel density estimation (KDE), a non-parametric approach for the estimation of the conditional probabilities. The idea is to use kernel functions to measure the similarity between the current context and data points in the location history. Data points with the highest similarity scores will have significant impact on the outputs. This approach is advantageous for dealing with sparse data, which happens when the amount of data is limited or when the user is in an infrequent context.

Our paper makes two contributions. First, we propose a non-parametric approach for location prediction based on kernel density estimation, for which we introduced several kernels to capture different aspects of spatio-temporal context. Our probabilistic framework can make predictions for a specific time or for a look-ahead time-window without any heuristics. Second, we present a thorough application-oriented study of location prediction, which considers the look-ahead time interval as a key aspect. Our analysis is conducted on a real life dataset with state-of-the-art spatial resolution and longitudinal recording period. Our experimental results show how the prediction performance is affected by various factors such as the time of the day or the look-ahead time window.

The paper is organized as follows. The next section discusses related work on mobility prediction in the context of mobile computing and compares our contributions with respect to the existing literature. Section 3 presents our prediction frameworks, with formal descriptions of the data representation and the prediction task, that is to predict user location at a given time in the future. Our analysis starts in Section 4, which introduces the location dataset. We report baseline results in Section 5. The results reveal the contexts for which the baseline performance is low, and motivates our proposed KDE approach presented in Section 6, a probabilistic model which uses spatio-temporal context. While the proposed method improves the accuracy on difficult settings such as large look-ahead time, we also found situations in which a simple baseline works best. Section 7 thus presents our final solution, which is a combination of the proposed model and a probabilistic version of the baseline method. In Section 8, we generalize the framework from predictions for a specific time, to predictions for a time window, reporting experimental results with information retrieval measures appropriate for this new task. Finally, Section 9 provides concluding remarks.

2. Related work

Human mobility analysis has become an active research topic thanks to the development of location tracking techniques [6,7]. Song et al. [8] presented a study on predictability of human mobility by analyzing the entropy of location traces. The analysis of entropy shows that the limit of predictability is around 93% for hourly sequences of GSM cell IDs, where the average size of a cell's area is about 3 km^2. Jensen et al. [9] applied the same methodology for analyzing predictability of discrete time series coming from several sources including GSM, WLAN, Bluetooth, and accelerometer. Recently, Lin et al. [10] extended the original work by studying the effect of spatio-temporal scales on predictability, showing that predictability increases with spatial scale and decreases with temporal scale.

Several prediction methods have been proposed for human mobility in different contexts (i.e., using different devices and sensors) and with different definitions of the prediction task. Some notable works are listed in Table 1. In transportation, Krumm et al. [12] consider the problem of inferring the destination based on partial paths which could be applied in navigation assistance systems. For example, context-aware trip recommendations can be produced by combining user specific needs (e.g., finding a gas station) with the inferred primary destination [19]. At a higher level, the prediction task