Regularity and Conformity: Location Prediction Using Heterogeneous Mobility Data

Yingzi Wang[⋆†], Nicholas Jing Yuan[†], Defu Lian[‡], Linli Xu[⋆],
Xing Xie[†], Enhong Chen[⋆], Yong Rui[†]
[⋆]School of Computer Science, University of Science and Technology of China
[†]Microsoft Research
[‡]Big Data Research Center, University of Electronic Science and Technology of China
{v-ywang, nicholas.yuan, xingx, yongrui} at microsoft.com,
dove.ustc at gmail.com, {linlixu,cheneh} at ustc.edu.cn

ABSTRACT

Mobility prediction enables appealing proactive experiences for location-aware services and offers essential intelligence to business and governments. Recent studies suggest that human mobility is highly regular and predictable. Additionally, social conformity theory indicates that people's movements are influenced by others. However, existing approaches for location prediction fail to organically combine both the *regularity* and *conformity* of human mobility in a unified model, and lack the capacity to incorporate *heterogeneous* mobility datasets to boost prediction performance. To address these challenges, in this paper we propose a hybrid predictive model integrating both the regularity and conformity of human mobility as well as their mutual reinforcement. In addition, we further elevate the predictive power of our model by learning *location profiles* from heterogeneous mobility datasets based on a gravity model. We evaluate the proposed model using several city-scale mobility datasets including location check-ins, GPS trajectories of taxis, and public transit data. The experimental results validate that our model significantly outperforms state-of-the-art approaches for mobility prediction in terms of multiple metrics such as accuracy and percentile rank. The results also suggest that the predictability of human mobility is time-varying, e.g., the overall predictability is higher on workdays than holidays while predicting users' unvisited locations is more challenging for workdays than holidays.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Data mining; H.2.8 [**Database Management**]: Spatial databases and GIS

General Terms

Algorithms, Experimentation, Performance

Keywords

location prediction, regularity, conformity, location profile, spatial influence, gravity model, collaborative filtering

KDD'15, August 10-13, 2015, Sydney, NSW, Australia.
© 2015 ACM. ISBN 978-1-4503-3664-2/15/08 ...$15.00.
DOI: http://dx.doi.org/10.1145/2783258.2783350.

1. INTRODUCTION

Over the past decade, an overwhelming number of location-aware services and apps have profoundly changed the way people live, from route planning to dining and even social networking. Understanding user mobility thus becomes an essential factor for improving service quality and user engagement.

While sensing a user's current location provides the user with timely reactive experiences, e.g., "searching for the closest subway station", predicting users' future locations can enable appealing proactive experiences in various applications. For example, recently emerging digital assistants such as Microsoft Cortana[1] and Google Now[2] aim to push relevant information to users or help users accomplish tasks without their querying, e.g., pre-heating (or cooling) the house when the user is on the way home [30]. An accurate prediction of user mobility is hence crucial for such proactive services. As another example, mobility prediction brings business intelligence to advertising and marketing. Given potential high-value customers' future locations, advertisers/marketers can better choose locations for organizing promotion events or distributing advertisements and coupons. Predicting future mobility patterns of crowds can help governments deal with public emergencies such as stampede prevention[3]. not about predicting mobility patterns

The recent development of sensing technology and smart devices makes various types of mobility data available to the industry and researchers, such as GPS trajectories [23, 41], cellular tower data [13], WiFi signals [27, 34], smart card transactions [39], and location check-ins from online social networks [28, 4, 19], all of which facilitate the exploration of mobility understanding and prediction. For instance, using cellular tower data, Song et al. [33] show that the predictability of human mobility has a limit of 93%, which demonstrates that human mobility is highly regular and predictable. However, the actual prediction performance heavily depends on many aspects including data types, sampling frequency, and granularity of predictions [32]. Even the best results reported by state-of-the-art approaches are far below this limit [6, 19, 27].

To bridge the gap between actual prediction performance and theoretical limit, many challenges still remain to be addressed: illogical

Regularity and Conformity. Several studies show that human mobility typically follows regular spatial-temporal patterns. In urban areas, people typically spend most of their time around several "ma-

[1]http://www.windowsphone.com/en-us/how-to/wp8/cortana/meet-cortana
[2]http://www.google.com/landing/now/
[3]http://en.wikipedia.org/wiki/2014_Shanghai_stampede

jor hubs", such as homes and workplaces [33], and periodically commute between them [18, 6]. Meanwhile, people frequently visit some "minor hubs" in a limited radius of their major hubs [9] at certain times, e.g., shopping malls, gyms, and restaurants. Nevertheless, human mobility is not only driven by *regularity*. People occasionally change their routines and visit some unfamiliar places, e.g., a bar recommended by friends or a popular restaurant on Yelp. Such irregular visits may be explicitly or implicitly influenced by others, usually a group of people who have similar social backgrounds, interests, and social statuses. This phenomenon is ~~the so~~-called social *conformity* [7].

paragraph does not flow smoothly

However, most existing approaches in location prediction typically fall into two categories: 1) developing individual mobility models, such as HMM [16], CRF [39], and periodic GMM [6], to capture users' regular behavioral patterns; 2) building collaborative models to leverage similar mobility patterns of different users [11, 26, 4]. Few studies have incorporated both the regularity and conformity of human mobility in predicting users' future locations. Although a few approaches have touched both factors ~~to a certain extent~~, the main endeavor of these methods still focuses on a single factor, while the other one is typically used as side information or a constraint [19]. Thus, the interdependency and mutual reinforcement of regularity and conformity are not fully exploited for location prediction.

touched on (?)

auxiliary

Sparsity and Heterogeneity. Continuous and precise tracking of users' long-term movements (e.g., using GPS) is often energy-intensive and costly, while mobility data captured by low-energy sensing technologies is typically sparse in terms of either granularity (e.g., cellular tower data) or sampling frequency (e.g., location check-ins). Besides, a user's actual mobility is usually delineated in different forms of mobility data, where any single type only partially reveals a user's mobility patterns. However, existing models for location prediction lack the capacity to boost prediction accuracy ~~with the help of~~ heterogeneous mobility datasets. The difficulty lies in how to integrate the mobility patterns mined from heterogeneous mobility datasets into a unified prediction model.

by utilizing

In this paper, we tackle the above challenges by proposing a hybrid model called **RCH**, combining both **R**egularity and **C**onformity, and employing **H**eterogeneous mobility data for location prediction. Specifically, we introduce a mobility model containing a regularity term and a conformity term, where the conformity term is represented by a time-aware factorization model, and the regularity term is represented as interactions between ~~users~~' hub visit patterns and spatial influence to ~~users~~' visited venues (detailed in Sec. 3.2). The regularity and conformity terms interplay and reinforce each other. In particular, the spatial influence to venues are learned through a Gravity model (detailed in Sec. 3.3). Our main contributions are summarized as follows:

too far apart

- We introduce a hybrid model for location prediction combing both the regularity and conformity of human mobility, which exploits the interdependent patterns of both routine visits and occasional visits.

utilizes

- We develop a method to learn a location's profile from heterogeneous mobility datasets based on a gravity model, and integrate the learned location profiles into a time-aware prediction model.

- We evaluate our model for predicting location check-ins based on a large dataset containing 7,355,962 check-ins of 161,794 users, where the location profiles used in our model are learned from several extra city-scale heterogeneous mobility datasets, such as GPS trajectories of taxis and public transit data. The experimental results validate that our model significantly outperforms state-of-the-art methods in terms of multiple metrics such as prediction accuracy and percentile rank.

misplaced

salesmanship

2. RELATED WORK

2.1 Predictability of Human Mobility

(?!)

The increasing availability of mobility data provides marvelous potential to study human mobility patterns. A considerable number of works have shown that human mobility is regular, predictable, and unique in both temporal and spatial spaces [35, 15, 8]. Observable regular movements among a few frequented locations, like home and work [9, 18], embody the regularity and predictability of human mobility. For example, using mobile phone logs of 100,000 users, Gonzalez et al. [13] showed a high degree of mobility regularity among several highly visited haunts. Song et al. [33] demonstrated ~~that~~ a 93% potential predictability of mobility patterns of mobile phone users. Besides, de Montjoye et al. [8] quantified ~~the~~ mobility uniqueness and demonstrated four distinct points are enough to distinguish 95% of users.

no comma

how is this related to the work in this paper?

In the past few years, several mobility prediction works have concentrated on human trajectory logs from personal mobile devices, smart cards, and vehicular digital records, like GPS data [23, 27, 1], wifi [23, 34] and bus-trip records [2, 39]. They have continual spatial and temporal mobility records and the conspicuous characteristic of periodical returning to some important places. Unlike these high frequency datasets, check-ins in LBSNs are usually sparse and sporadic [24]. Location prediction based on check-ins is more challenging than on dense datasets like GPS data [37].

Wifi on page 1.

2.2 Location Prediction Models

We summarize relevant mobility prediction models and their differences in Table 1, where we list the targeted mobility data (i.e., type of mobility to be predicted) and the features incorporated in the models. According to whether the prediction model is trained independently among all users (i.e., whether a user's mobility model is learned from the user's own historical mobility alone), we categorize existing models into two types: individual models and collaborative models.

Table 1: Comparison of location prediction methods

CI: check-in, SMP: spatial mobility pattern, TC: text content
IT: individual temporal patterns, SR: social relationship
CF: collaborative filtering, HT: heterogeneous mobility datasets

methods	target			feature					
	CI	GPS	Wifi	SMP	TC	IT	SR	CF	HT
PSMM [6]	√			√		√	√		
W^4 [40]	√			√	√	√			
M5Tree[25]	√					√	√		
CEPR [19]	√			√		√		√	
SHM [12]	√					√	√		
gSCorr [11]	√					√	√		
DBN [26]	√					√	√		
NextPlace [27]		√	√			√			
WhereNext [23]		√				√			
Markov [1]		√							
RCH(Our Model)	√			√		√		√	√

space

space

2.2.1 *Individual Models*

Historical spatial-temporal mobility patterns are fundamental factors for inferring users' future locations, given the regularity of human mobility [27, 23, 40, 10, 39, 16]. An approach based on nonlinear time series is applied for mobility prediction in [27], which focused on predicting the most important places. Using a GPS trajectory dataset generated by 17,000 cars, Monreale et al. [23] built a decision tree, named T-pattern tree, to find the best match path and predict future movements. Yuan et al. [40] proposed a probabilistic model W^4 (who, when, where, what) unifying spatial, temporal and activity topics to model users' behaviors. Using public transit records, Yuan et al. [39] provided a constraint Conditional Random Field model and ~~successfully~~ inferred ~~unknown~~ alighting/boarding stops given part of them.

The advantage of individual models is that the regularity of human mobility can be well captured. However, the similarity of mobility patterns between different users is not considered and utilized for predicting future locations. Instead, our hybrid approach excavates similar users' mobility patterns based on social conformity and collaborative filtering in addition to learning users' regular mobility patterns with a time-aware sparse group Lasso model. Furthermore, we collectively learn location profiles using several heterogeneous mobility datasets generated by city-scale populations and integrate the location profiles into our hybrid prediction model.

2.2.2 Collaborative Models

Different people may have similar location preferences. Social relationships of users have been taken into account for location prediction and recommendation to relieve data sparsity [6, 12, 25, 26, 4]. For example, Noulas et al. [25] developed a supervised learning model for next place prediction considering the location histories of users' friends. Cho et al. [6] introduced a time-aware Gaussian Mixture model considering both users' periodic mobility and social activities. Sadilek et al. [26] provided a Dynamic Bayesian Network model combining friends' temporal information for location prediction. Nevertheless, social relationships are reported to offer a limited predictive power for location check-ins due to the high sparsity [24].

inconsistent capitalization

Collaborative filtering methods are widely applied in recommendation systems including location recommendations, which assumes that similar users have similar behavioral patterns like rating or purchasing. This assumption is also in accordance with the conformity theory in social psychology [7]. For example, matrix factorization has evolved as a critical algorithm in location recommendation [5, 22, 20], where a user's preference of a venue is modeled as an inner product of latent factors. Lian et al. [20] introduced a location recommendation model considering both users' latent preferences and the geographical influence of locations, however, the influence is empirically determined, instead of learned from the data. Recently, probabilistic non-negative matrix factorization has also been adopted for location recommendations [21, 29], where the users' visited venues are considered observations of a generative process. However, these recommendation models cannot be directly applied for mobility prediction. Lian et al. [19] first employed collaborative filtering approaches for location prediction. Users' location visits are separated to explorations of novel or regular places based on a binary classification. They proposed a hidden Markov model for capturing regular mobility patterns and social-based collaborative filtering with 2D kernel density estimation to excavate novel mobility patterns. However, the antecedent division of mobility types through exploration prediction confronts the risk of two-layer errors.

In contrast to the above collaborative filtering approaches for location recommendation and location prediction, our method 1) simultaneously incorporates both regularity and conformity of human mobility in a unified prediction model, and utilizes the interplay between these two factors; 2) provides a time-aware collaborative model considering users' preference drifting at different time slots so as to enable time-aware location predictions; 3) learns spatial influence on venues using heterogeneous mobility datasets based on a gravity model and feeds it into the prediction model.

capitilization inconsistent with Table 1.

cited in main text?

capitalization inconsistent with table 1.

Table 2: Important Notations

Notation	Size	Description
$\mathbf{R}(t)$	$M \times N$	user-venue preference matrix at time slot t
$\mathcal{T}$	$1 \times T$	time slot set; t is a time slot in $\mathcal{T}$
$\mathcal{C}$	$1 \times I$	grid set; d is a grid in $\mathcal{C}$ with length I
$\mathcal{G}$	$1 \times G$	group set; g is a group in $\mathcal{G}$
$\mathcal{P}$	1×3	mobility type $=\{B(\text{bus}), A(\text{taxi}), C(\text{check-in})\}$
$\mathcal{O}^*$	$1 \times I$	out-going flows of all grids in $\mathcal{C}$ w.r.t. $* \in \mathcal{P}$
$\mathcal{D}^*$	$1 \times I$	in-coming flows of all grids in $\mathcal{C}$ w.r.t. $* \in \mathcal{P}$
$\mathbf{T}^*$	$I \times I$	transition matrix of grids in $\mathcal{C}$ w.r.t. $* \in \mathcal{P}$?
$\mathbf{U}$	$M \times K$	user stationary latent factor
$\mathbf{U}(t)$	$M \times K$	user changing latent factor of time ~~slot~~ t
$\mathbf{V}$	$N \times K$	venue latent factor
$\mathbf{H}(t)$	$M \times I$	hub matrix at time ~~slot~~ t
$\mathbf{Q}^*$	$N \times I$	sptial influence matrix w.r.t. $* \in \mathcal{P}$
$\mathbf{H}^{(g)}(t)$	$M \times L^{(g)}$	$\mathbf{H}(t)$'s sub-matrix of group g
$\mathbf{Q}^{*(g)}$	$N \times L^{(g)}$	$\mathbf{Q}^*$'s sub-matrix of group g

3. MODEL

3.1 Overview

Given visited venues of a group of users, our goal is to predict their future locations at a certain time. Let $\mathcal{U} = \{u_1, u_2, ..., u_M\}$ be M users and $\mathcal{V} = \{v_1, v_2, ..., v_N\}$ denote N venues. Note that here $\mathcal{V}$ may contain unvisited venues of users in U. We categorize days into two types, workdays and holidays, and let $\mathcal{T} = \{t_1, t_2, ..., t_T\}$ represent the T time slots in the two classes of days. Given a specific time slot t, we predict user u_i's location by calculating the mobility preferences of u_i to v_j for $j = 1, 2, \ldots, N$ at t, and returning the v_j that has the maximum mobility preference. Let $\mathbf{R}(t) \in \mathbb{R}^{M \times N}$ be the preference matrix of $\mathcal{U}$ to $\mathcal{V}$ at time t, i.e., $R_{ij}(t)$ indicates u_i's preference to v_j at t. As mentioned earlier, a user u_i's visit to a venue v_j can be driven by either regularity or conformity, i.e.,

$$R_{ij}(t) = R_{ij}^{(r)}(t) + R_{ij}^{(c)}(t), \tag{1}$$

where $R_{ij}^{(r)}(t)$ is the regularity term, indicating that v_j is a regular venue of u_i at time t; and $R_{ij}^{(c)}(t)$ is the conformity term, indicating that v_j is frequently visited by users who are similar to u_i at time t. These two factors can interplay and reinforce each other to drive u_i's visit to v_j. Next, we introduce both terms respectively as follows.

3.1.1 Regularity Term $\mathbf{R}^{(r)}$

For simplicity, in the rest of Sec. 3.1.1, we restrict our notations and description of the model to a specific time (without considering the time varying effect), and later in Sec 3.3, we will formulate the time-aware model.

To learn users' regular mobility patterns, we map users' visited venues to the geospatial space. Let $\mathcal{C} = \{d_1, d_2, ..., d_I\}$ be the C geographical grid cells (e.g., 100m$\times$ 100m) discretizing the whole geospatial space of a city. Each venue v_j is associated with a geo-coordinate $\{lat_j, lon_j\}$ belonging to a certain grid d_{k_j}, where lat_j, lon_j are the latitude and longitude of v_j. As shown in Fig. 1 a), u_i's visited venues are mapped to the grids shown as the ones with plus signs.

Consider the probability that u_i visits v_j in terms of regularity, denoted as $\Pr(v_j|u_i)$. We assume that v_j belongs to a grid d_{k_j}, and u_i travels from a grid d_k to v_j (note that it is possible that