Individual Trajectory Reconstruction from Mobile Network Data

Guangshuo Chen, Sahar Hoteit, Aline Carneiro Viana, Marco Fiore, Carlos Sarraute

Individual Trajectory Reconstruction from Mobile Network Data

Guangshuo Chen, Sahar Hoteit, Aline Carneiro Viana, Marco Fiore, Carlos Sarraute

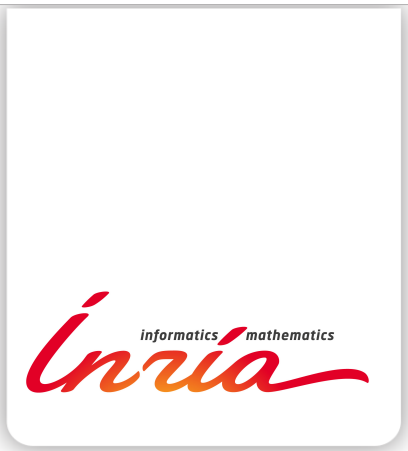
TECHNICAL REPORT

N° 495

Janvier 2018

Project-Teams INFINE

ISSN 0249-0803 ISRN INRIA/RT--495--FR+ENG



Individual Trajectory Reconstruction from Mobile Network Data

Guangshuo Chen*†, Sahar Hoteit‡, Aline Carneiro Viana†,

Marco Fiore§, Carlos Sarraute¶

Project-Teams INFINE

Technical Report n° 495 — Janvier 2018 — 20 pages

Abstract: Mobile phone datasets are a primary source of positioning information for human mobility studies across many disciplines. They provide individual trajectory footprints in the form of geo-referenced and time-stamped events recorded in the mobile network. The quality of the mobility information in mobile phone datasets depends on the nature of the network infrastructure and on the frequency of its interactions with mobile devices. Typically, geographical deployments of cellular networks are far from uniform, and the events triggered by each mobile device are sparse and irregular in time. As a result, individual trajectories inferred from mobile network data are often substantially incomplete.

Key-words: Call Detail Records, spatiotemporal trajectories, data sparsity, cellular networks, mobility, movement inference

This work was supported by the EU FP7 ERANET program under grant CHIST-ERA-2012 MACACO.

* École Polytechnique, Université Paris Saclay, 91128 Palaiseau, France
† INRIA Saclay-Île-de-France, Université Paris Saclay, 91120 Palaiseau, France
‡ Laboratoire des Signaux et Systèmes, Université Paris Sud-CNRS-CentraleSupélec, Université Paris-Saclay, 91192 Gif-sur-Yvette, France
§ CNR-IEIIT, 10129 Torino, Italy
¶ Grandata Labs, 550 15th Street, San Francisco, 94103 California, USA

RESEARCH CENTRE
SACLAY – ÎLE-DE-FRANCE

1 rue Honoré d'Estienne d'Orves
Bâtiment Alan Turing
Campus de l'École Polytechnique
91120 Palaiseau

Reconstruction individuelle de trajectoire à partir de données de réseau mobile

Résumé : Les ensembles de données sur les téléphones mobiles sont une source primaire d'informations de positionnement pour les études sur la mobilité humaine dans de nombreuses disciplines. Ils fournissent des empreintes de trajectoires individuelles sous la forme d'événements géoréférencés et horodatés enregistrés dans le réseau mobile. La qualité de l'information sur la mobilité dans les ensembles de données de téléphonie mobile dépend de la nature de l'infrastructure du réseau et de la fréquence de ses interactions avec les appareils mobiles. Typiquement, les déploiements géographiques des réseaux cellulaires sont loin d'être uniformes et les événements déclenchés par chaque appareil mobile sont rares et irréguliers dans le temps. En conséquence, les trajectoires individuelles déduites à partir des données du réseau mobile sont souvent substantiellement incomplètes.

Mots-clés : Enregistrements détaillés des appels, trajectoires spatio-temporelles, éparpillement des données, réseaux cellulaires, mobilité, inférence de mouvement

1 Introduction

Human mobility has been the subject of significant research efforts over the last decade. Seminal studies have demonstrated how individuals follow regular visiting patterns at specific locations [1] that are highly predictable [2], despite some spatiotemporal uncertainty [3] and uniqueness [4]. The main drivers for mobility were also identified [5], such as home-work commuting [6].

The deeper understanding of human movements has revealed a key enabler for the management and optimization of large systems. In the context of mobile networking, the characterization of individual movement patterns has proven critical to design efficient solutions for user terminal paging [7], location recommendation [8] and data offloading [9].

Datasets of spatiotemporal trajectories of individuals play a critical role in all these studies. In most cases, it is required that the positioning data encompasses a large (*e.g.*, citywide or nationwide) population and long time period (*e.g.*, weeks to months), so as to capture the full diversity of human mobility behaviors, and generalize conclusions. Mobile phone datasets meet such requirements: the success of mobile services and the ever higher penetration rate of mobile devices give operators the possibility to track very large populations at little cost [10]. Mobile phone trajectories have thus emerged as a primary source of information for human mobility research: all works mentioned above build on this class of data.

Among different types of mobile phone data, Call Detail Records (CDR) (also known as *charging data record* in the 3GPP lexicon) keep track of a variety of telecommunication events, including issued and received voice calls, sent and received text messages, and newly established data traffic sessions. Each CDR record contains information about the nature of the corresponding event, and include identifiers of the relevant mobile device. CDR events are also time-stamped and georeferenced with the location of the antenna the device is associated to at the moment when the activity takes place. CDR are collected by mobile network operators for billing purposes, but they also capture valuable digital footprints of millions of mobile subscribers. As such, CDR represent the most common type of mobile phone data used in the literature [10].

Motivation. The cellular network is far from being a perfect sensor of human mobility. CDR usually contain trajectory data that are coarse in space and sparse in time. The device location accuracy is limited by the antenna coverage, which can be as large as several km^2 in urban areas; the temporal distribution of CDR records of a single user is highly heterogeneous, with typical inter-event times of hours. The sparsity of CDR makes the vast majority of the trajectory data unusable: for instance, 100K, 50K and 700 individuals are retained for analysis from populations of 6M, 10M, and 1M in [1, 2, 11], respectively. This maps to percentages between 0.07% and 1.67% of the total user base, whereas the bulk of CDR information is dropped due to insufficient sampling frequency of a user's movements. We refer to this problem as *CDR sparsity*.

In this paper, we aim at mitigating CDR sparsity by reconstructing functional trajectories of individual users from the original incomplete CDR – a task we name *CDR completion*. A successful CDR completion would fill the gaps in the individual positioning information conveyed by CDR, and do so with spatioptemporal points that closely match the real-world movement patterns of each subscriber. CDR completion has the potential to increase the size of user populations considered in CDR-based mobility analyses by orders of magnitude, with straightforward benefits to the confidence and generality of results.

Existing approaches and limitations. The literature on CDR completion is fairly thin. Approaches based on interpolation only work in presence of trajectories composed of thousands of locations per day [11], which is hardly the case with CDR. A basic solution commonly adopted to complete CDR is to assume that subscribers remain at the locations where they are observed in the data for some fixed amount of time, *e.g.*, in the order of hours [12]. Extensions to this approach involve adaptive strategies tailored on each user [13, 14]. However, all these techniques

incur into non-negligible spatial errors, and reconstruct trajectories with inconsistent sampling frequency over time. It is also worth noting that travel reconstruction techniques from GPS traces [15] is a very different problem from CDR completion. GPS data are collected with a fixed periodicity and at high frequency (*e.g.*, at every second, or at every minute), hence do not need to be completed in time.

Contributions. In this paper, we make two contributions towards a solution to the CDR completion problem.

First, we mine a real-world metropolitan-level CDR dataset, encompassing 1.8M users during 3 months, to assess the severity of the CDR sparsity problem. We provide empirical distribution estimates of relevant metrics, including the period of coverage, the sampling frequency, and the loss patterns. Our results show that legacy preprocessing techniques, used to select trajectories with a sufficient level of detail, in fact filter out users with substantial although irregular location information, thus wasting potentially serviceable data.

Second, we propose a novel approach for CDR completion that stems from the observation that human movement patterns are recurrent, and leverages tensor factorization methods. We evaluate the performance of our solution with real-world datasets that include ground-truth information about user locations. Our results show that the proposed strategy recovers, on average, 50% of the positions missing in the original CDR data of users who have trajectory are able to be precisely recovered by our approach when the completeness of trajectory is more than 10%. Also the distance error of estimated location is in the same degree as the location precision in the ground-truth datasets.

2 Trajectory incompleteness in CDR data

We start our study by investigating and quantifying the level of incompleteness of individual mobility information inferred from CDR data. To this end, we mine a real-world large-scale mobile dataset for sensible metrics.

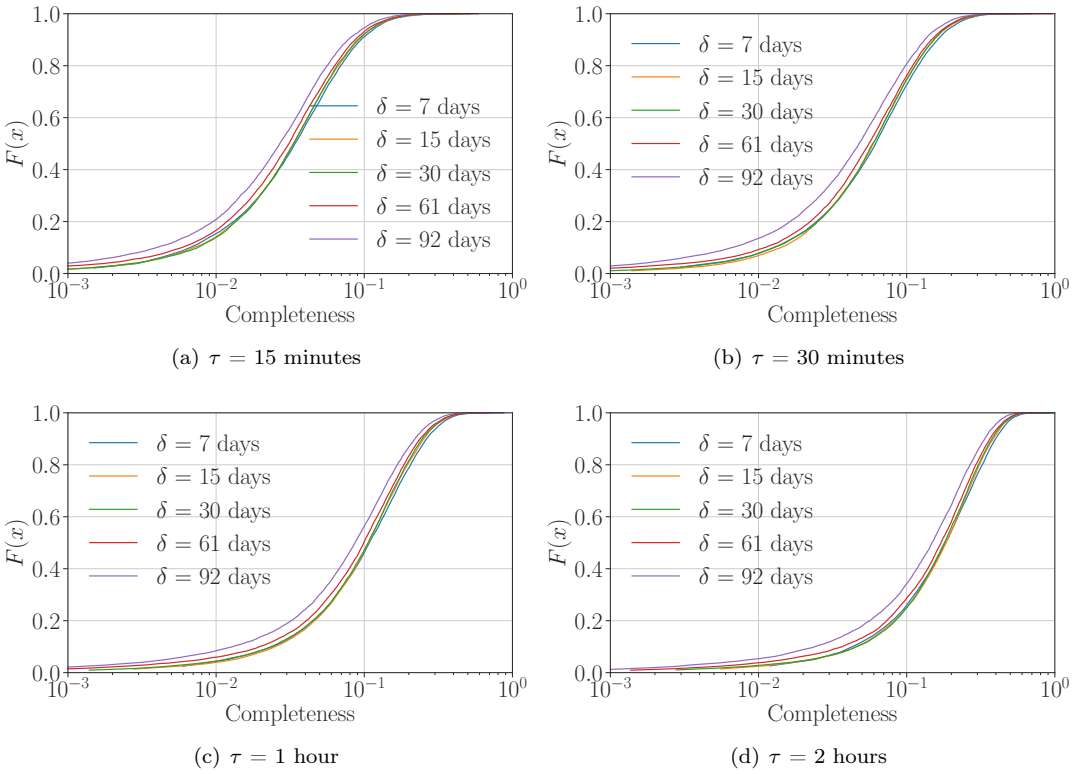
2.1 Dataset

We consisder a CDR dataset collected by a major mobile network operator in Mexico, and including data from 1.8M mobile phone users. The dataset consists of 778M CDR logs generated by the mobile subscribers during a 3-month period. Each dataset record is composed by a pseudonymised user identifier, a timestamp, the event duration, and the cell tower handling the event. The locations of the 4.2K cell towers present in the dataset are latitude and longitude pairs, and are also provided by the operator.

2.2 CDR-based trajectories and completeness

The sequence of time-stamped CDR records associated to a same mobile device provide an implicit and approximated sampling of the trajectory of the corresponding user. We refer to such sampling as a *CDR-based trajectory*. As anticipated in Sec. 1, CDR-based trajectories are typically fairly incomplete. The telecommunication activity of mobile subscribers is driven by endeavours and habits that are far from deterministic: ultimately, this leads to irregular, sparse samplings in CDR-based trajectory.

In the light of this situation, a relevant question is *what degree of (in)completeness can one expect in CDR-based trajectories inferred from real-world large-scale mobile phone datasets*. To answer such question in a structured manner, we first define the notion of *completeness* of a CDR-based trajectory as the fraction of time intervals during which the location of a mobile subscriber



(a) $\tau = 15$ minutes

(b) $\tau = 30$ minutes

(c) $\tau = 1$ hour

(d) $\tau = 2$ hours

Figure 1: Distributions of the completeness of the CDR-based trajectories of 1.8M users for combinations of dataset duration δ and temporal resolution τ.

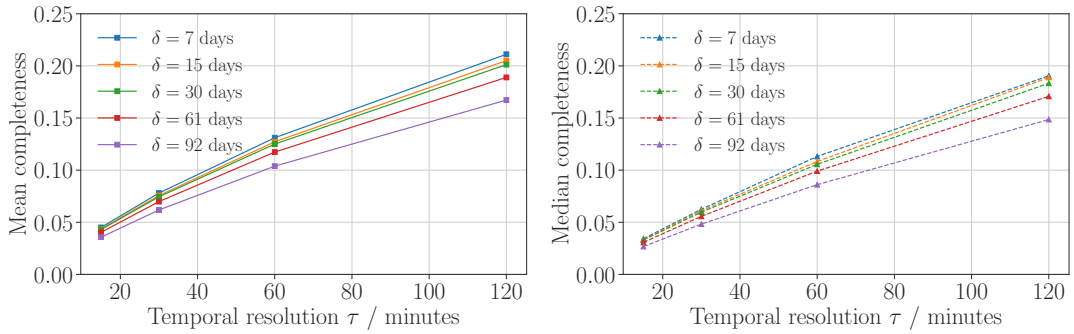
is recorded at least once. This definition implies that the overall period covered by the dataset, referred to as δ in the following, is discretized into time intervals, whose duration we denote as τ. For example, given a dataset with δ equal to 7 days and τ set to one hour, a CDR-based trajectory with samples in 80 different hours has a incompleteness $80/(7 \times 24) = 0.476$.

Intuitively, the completeness of CDR-based trajectories depends on both the duration of the dataset δ and the temporal resolution τ. We thus investigate how these parameters affect completeness. We consider all combinations of $\delta \in \{7, 15, 30, 61, 92\}$ days and $\tau \in \{15, 30, 60, 120\}$ minutes. For each combination of dataset duration and temporal resolution, we generate CDR-based trajectories for all 1.8M users in our dataset and compute their completeness.

Fig. 1 portrays the cumulative distribution function (CDF) of the resulting CDR-based trajectory completeness. Each plot refers to a different temporal resolution, and curves map to diverse dataset durations. We make the following observations.

- No matter the settings considered, fully complete trajectories are very hard to obtain from CDR directly. Not a single complete trajectory is inferred from our reference dataset, despite the large user population covered and even when considering $\delta = 7$ days and $\tau = 2$ hours.

- Interestingly, the completeness is only very slightly affected by the dataset duration. As one could expect, CDR-based trajectories that span a shorter time period tend to have higher completeness, however the difference of completeness remains fairly small, and never



(a) Mean (b) Median

Figure 2: (Mean and median) completeness versus temporal resolution τ with respect to duration δ.

exceeds 0.1 between $\delta = 7$ and $\delta = 92$. In fact, the completeness hardly varies at all when δ is equal to or less than 30d.

- Completeness is instead significantly affected by the temporal resolution τ. For instance, considering a dataset duration of 30 days, only 10% of the CDR-based trajectories display a completeness above 0.1 when the resolution is $\tau = 15$ minutes; this percentage grow to 75% when $\tau = 120$ minutes. To better exploit this observation, we plot in Fig. 2 the mean and median of completeness versus temporal resolution τ under all the durations, which clearly show that completeness and temporal resolution has a high degree of linear correlation. In fact, ignoring duration, the Pearson correlation coefficient between the two factors is still 0.983.

Overall, these results highlight that, although CDR never capture fully complete trajectories, the vast majority of CDR-based trajectories feature a non-negligible level of completeness (*e.g.*, $0.1 - 0.5$) given an appropriate resolution (*e.g.*, in the order of hours). These trajectories cannot be used as such for mobility analysis, but, as we will later see, they retain sufficient information to be salvaged by our CDR completion solution.

2.3 Modelling completeness

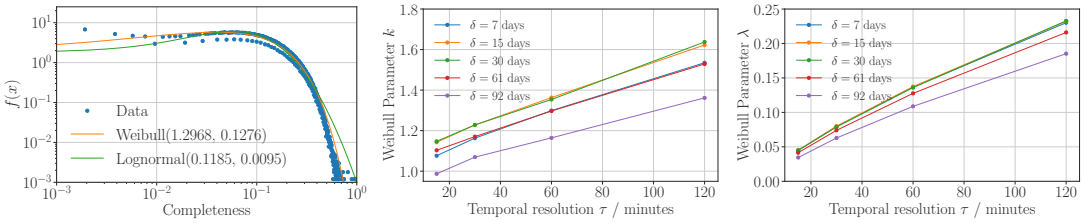
We also provide a data-driven model of the completeness of CDR-based trajectories. To this end, we derive the theoretical distributions that best fit the empirical CDF shown in Fig. 1, under all combinations of δ and τ.

Earlier studies have shown that the distribution of mobile phone activity across users tends to be long-tailed [16]. Since the completeness of CDR-based trajectories is determined by the user's level of activity, it makes sense to look at heavy-tailed distributions. We run a maximum likelihood estimation of the parameters of six standard long-tailed distributions (namely, Generalized Pareto, Lévy, Power-law, Lognormal, Gamma, and Weibull) on the empirical CDF of completeness. We evaluate the quality of the fitting by both computing the coefficient of determination (denoted as R^2) and running the Kolmogorov-Smirnov test (whose output, *i.e.*, the Kolmogorov-Smirnov statistic, is indicated as D_{KS}) on the empirical and estimated distributions.

Tab. 1 summarizes the results. Two distributions fit far better than the others as they occupy most of the maximum values of R^2 and D_{KS} among all the combinations of durations

Table 1: R^2 and D_{KS} of the best fits of the six standard long-railed distributions

		Weibull		Lognormal		Gamma		GenPareto		Levy		Powerlaw	
δ	τ	D_{KS}	R^2	D_{KS}	R^2	D_{KS}	R^2	D_{KS}	R^2	D_{KS}	R^2	D_{KS}	R^2
$7d$	$15m$	0.0334	**0.9990**	**0.0318**	0.9973	0.3710	0.4679	0.0495	0.9969	0.2509	0.8046	0.3538	0.5031
	$30m$	**0.0302**	**0.9993**	0.0345	0.9967	0.0548	0.9962	0.1716	0.8973	0.2649	0.7848	0.3587	0.4894
	$60m$	**0.0278**	**0.9990**	0.0372	0.9958	0.0475	0.9977	0.1198	0.9516	0.2888	0.7506	0.4162	0.2041
	$120m$	**0.0375**	**0.9972**	0.0443	0.9938	0.0629	0.9853	0.1043	0.9768	0.3238	0.6977	0.2456	0.7450
$15d$	$15m$	0.0264	**0.9986**	**0.0233**	0.9984	0.3594	0.5007	0.0627	0.9893	0.2620	0.7853	0.3949	0.4120
	$30m$	**0.0208**	**0.9995**	0.0264	0.9980	0.0271	0.9991	0.0724	0.9851	0.2765	0.7647	0.3576	0.4975
	$60m$	**0.0188**	**0.9997**	0.0279	0.9974	0.0257	0.9987	0.0935	0.9719	0.2997	0.7315	0.3176	0.6076
	$120m$	**0.0254**	**0.9987**	0.0332	0.9961	0.0913	0.9572	0.1880	0.8780	0.3349	0.6792	0.2721	0.7116
$30d$	$15m$	0.0239	**0.9985**	**0.0207**	**0.9985**	0.3514	0.5261	0.0700	0.9835	0.2619	0.7872	0.3829	0.4365
	$30m$	0.0205	**0.9992**	0.0216	0.9983	**0.0203**	0.9991	0.1528	0.8976	0.2763	0.7661	0.3912	0.4149
	$60m$	**0.0149**	**0.9996**	0.0263	0.9975	0.0289	0.9982	0.0895	0.9702	0.3003	0.7346	0.3131	0.6212
	$120m$	**0.0239**	**0.9984**	0.0315	0.9962	0.0995	0.9479	0.1069	0.9538	0.3337	0.6893	0.3154	0.6176
$61d$	$15m$	0.0266	**0.9980**	**0.0239**	0.9977	0.1990	0.8284	0.0458	0.9934	0.2527	0.8076	0.3850	0.4156
	$30m$	**0.0233**	**0.9985**	0.0234	0.9976	0.0286	0.9974	0.1944	0.8669	0.2649	0.7903	0.3897	0.4212
	$60m$	**0.0207**	**0.9985**	0.0264	0.9970	0.0310	0.9980	0.0772	0.9769	0.2883	0.7622	0.3451	0.5635
	$120m$	**0.0245**	**0.9975**	0.0316	0.9958	0.0754	0.9750	0.0946	0.9617	0.3209	0.7196	0.3491	0.5070
$92d$	$15m$	**0.0298**	**0.9954**	0.0329	**0.9954**	0.3619	0.5248	0.0322	0.9954	0.2371	0.8390	0.3866	0.4528
	$30m$	0.0336	**0.9958**	**0.0335**	0.9950	0.0583	0.9864	0.0398	0.9942	0.2487	0.8264	0.3819	0.4774
	$60m$	**0.0305**	**0.9960**	0.0355	0.9944	0.0454	0.9913	0.0611	0.9843	0.2705	0.8020	0.3231	0.6123
	$120m$	**0.0369**	0.9940	0.0379	0.9932	0.0486	0.9927	0.0760	0.9743	0.2998	0.7687	0.3359	0.5885



(a) Example of fitting, $\delta = 61$ days, $\tau = 60$ minutes

(b) PDF

(c) PDF

Figure 3: (a) Empirical PDF and Weibull, Lognormal theoretical fittings of the completeness of 1.8M CDR-based trajectories. (b,c) Linear correlation of the Weibull k and λ parameters with the temporal resolution τ, for varying dataset durations δ.

and temporal resolutions: they are the Lognormal, with probability density function (PDF)

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{\ln(x/\mu)^2}{2\sigma^2}}, \tag{1}$$

and Weibull, with PDF

$$f(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}. \tag{2}$$

An illustrative example, for $\delta = 61$ days and $\tau = 60$ minutes, is shown in Fig. 3(a). We also remark that, quite expectedly, the models are well aligned with those of mobile phone activity distributions found in the literature [16].

The theoretical modelling of completeness distributions paves the way to an in-depth analysis of how the system settings affect the model parameters. We focus on the Weibull model here, since is has a slightly better fittings than the Lognormal ones[1]. Plots (b) and (c) in Fig. 3 show an interesting phenomenon: both parameters of the Weibull model, k and λ, scale linearly with the temporal resolution τ, for any δ. More precisely, the Pearson correlation coefficient with respect to τ of k and λ are 0.89 and 0.98, respectively. The positive correlation correctly implies that higher value of τ shift the PDF mass towards larger completeness.

Although the generality of these relationships needs to be validated on other CDR datasets, they open interesting perspectives on the possibility to characterize the level of completeness of CDR-based trajectories directly from the time resolution τ.

3 CDR Completion Problem

In this section, we formulate the CDR completion problem and discuss the rationale.

3.1 Problem statement

We first formulate the CDR-based trajectory. In the mobile dataset, telecommunication events are collected in the *period of study*. We divide the period of study into a set $\mathcal{T}$ of equivalent time slots, as $\mathcal{T} = \{1, \cdots, t\}$, and use one-hour slots considering the tradeoff between the temporal

[1]A possible reason is that the burstiness of CDR events pushes multiple samples in a same time slot; as these count as a single value in the completeness calculation, the distribution is shifted from Lognormal to Weibull.

resolution and the available population regarding the data completeness. For the user u from the population of study $\mathcal{U}$ ($u \in \mathcal{U}$), the CDR-based trajectory $L^u_{\mathcal{T}}$ is defined to be a time series of locations in terms of the slots $\mathcal{T}$:

$$L^u_{\mathcal{T}} = \{\mathbf{l}^u_i | i \in \mathcal{T}\}, \tag{3}$$

where $\mathbf{l}^u_i$ represents the sampled location of the user u in the i-th time slot, which is a vector of which the dimension depends on the coordinating system. Each user's CDR-based trajectory comes from the CDR-based trajectory set $\mathcal{L}^{\mathcal{U}}_{\mathcal{T}}$, *i.e.*, $L^u_{\mathcal{T}} \in \mathcal{L}_{\mathcal{T}},\ \forall u \in \mathcal{U}$.

We second formulate the data loss in the CDR-based traejectory. In general, $L^u_{\mathcal{T}}$ has certain missing locations because of the data loss in the mobile dataset. For the user u, we define the *observation set* as $\Omega_u = \{t_1, \cdots, t_k\} \subseteq \mathcal{T}$ consisting of the time slots in which the locations are known from the corresponding CDR and the *observation trajectory* $L^u_{\Omega_u} = \{\mathbf{l}^u_i | i \in \Omega_u\}$ to represent the existing part of $L^u_{\mathcal{T}}$.

The CDR completion is to fill temporal gaps in the CDR-based trajectory and particularly, is to infer the missing locations in the time slots that are not in Ω_u for the user u using the existing mobility information extracted from the mobile dataset. We formally define the CDR completion problem as the minimization problem in the following:

$$\min \sum_{i \in \Omega^C_u} \frac{|\mathbf{l}^u_i - \hat{\mathbf{l}}^u_i|}{|\Omega^C_u|}, \quad s.t.\ \mathcal{L}^{\mathcal{U}}_{\mathcal{T}}, \tag{4}$$

where $\Omega^C_u = \mathcal{T} - \Omega_u$ and $\hat{\mathbf{l}}^u_i$ is the estimation of the missing location in the i-th time slot.

3.2 Rationale of CDR completion

The previous results in Sec. 2 show that most of the user trajectories extracted from the CDR dataset are highly incomplete, which means these users have $|\Omega^C_u| \gg |\Omega_u|$ and the completion of their trajectories are challenging. The good news is that, it is still feasible to design techniques from the CDR completion despite of the high loss rate in the CDR-base trajectory because the human mobility is highly redundant [1]. The CDR completion is applicable due to the human mobility characteristics recently revealed in the state-of-the-art studies in the following.

- **Redundancy**: Barabási *et al.* [1, 2] have shown that human mobility is highly redundant in various aspects. In terms of the displacement, one returns frequently to a few highly visited locations [1]. Despite of the loss, these active locations are captured by CDR with a high probability. In terms of the temporal repetitiveness, the order of human visiting patterns contributes to the high degree of human mobility predictability [2]. It means that the missing location may be recovered with a high degree of potential from knowing the temporal order of visiting, which helps significantly the CDR completion.

- **Inter-call Stability**: One tends to spend a lot of time where one makes calls, and travels fast between consecutive communications [17]. Thus the locations captured by CDR tend to be the ones in which the user is stable. This supports the representativeness of the CDR-based trajectory.

- **Nighttime Stability**: Intuitively, one usually is stable during nighttime because of the human nature. Compared with the daytime, less CDR are expected to be captured in the nighttime but the nighttime stability can help identifying the home location and the period in which one is at home with a high degree of accuracy, as shown in [14].

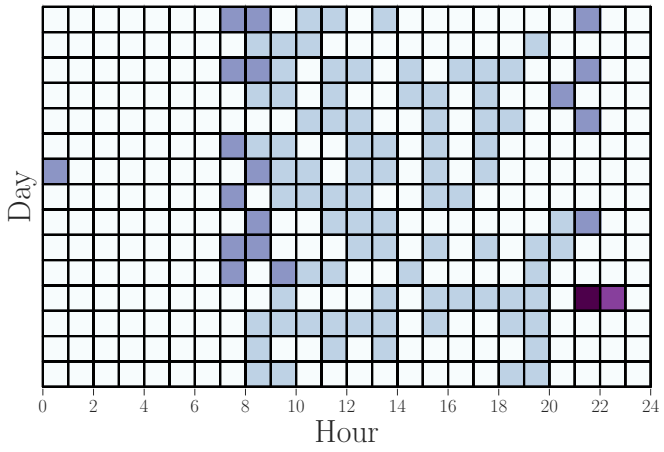


Figure 4: The location loss in an actual CDR-based trajectory having 30% of completeness. Each block shows the location of the corresponding hour, in which white means that the location is missing; other colors represents unique cell tower locations.

The above characteristics about human mobility are utilized in the design of our proposed technique in the next section. Moreover, recall that 50% of the users have at least 10% of the locations every hour, *i.e.*, tens of thousands of users. Completing their CDR-based trajectories is able to make opportunities to prevent them from being filtered out and put them back into the population of study. In summary, the completion of CDR data is challenging in terms of the data loss but the nature of human mobility helps overcoming such loss and achieving effective CDR completion.

4 CDR Completion Approach

In this section, we propose our novel approach to address the CDR completion problem. The approach, named as *temporal improved tensor factorization with home inference* (`TF-Home`), utilizes the characteristics of human mobility mentioned in Sec. 3. We give an overview the approach in Sec. 4.1 and present the approach in the rest of the section in detail.

4.1 Overview

A significant issue in the design of a CDR completion approach is to address the loss pattern of data appropriately. As revealed in Sec. 2, the CDR-based trajectory usually has a degree of data loss. Thus a CDR completion approach should be first able to handle such loss. The location loss is not random in the CDR-based trajectory because the telecommunication events, as the source of locations, are spatiotemporally correlated. Fig. 4 illustrates an actual CDR-based trajectory extracted from the CDR dataset and indicates the two significant features in this kind of trajectories, *i.e.*, *heterogeneity* and *redundancy*. For the former, more locations are captured by CDR during daytime than during nighttime. The good news is that the heterogeneity of data provides more plenty of daytime locations when the human trajectory is more complicated. The uncertainty of nighttime locations is low due to the nighttime stability of human mobility. For the latter, we see that the daily sub-trajectories in Fig. 4 are highly repetitive, which reduces the uncertainty and helps significantly in the data completion.

Based on the considerations above, we design the `TF-Home` approach Our approach receives the incomplete CDR-based trajectory as input and provides the complete estimated CDR-based trajectory as output via three steps in the following:

1. **Nighttime data enhancement**: This step is designed to deal with the heterogeneity of data. An adaptive strategy is proposed to fill nighttime temporal gaps in the CDR-based trajectory with the home location identified, which lightens the data loss during nighttime.

2. **Temporal improved tensor data completion**: This step is the core procedure in the proposed approach and adopts the *tensor factorization* to infer the missing locations using the redundancy of the CDR-based trajectory.

3. **Cell tower estimation**: This step aims at locating the very cell where the user is from the location estimated. Instead of using the nearest cell tower, we use a strategy combining the distance and the historical appearance.

4.2 Nighttime data enhancement

The first step adopts an adaptive strategy for adding more location between evening and morning with minor affecting the localization precision. The objective is to lighten the data loss, improve the accuracy of user location during nighttime, and ensure the performance of applying the tensor factorization technique in the next step. Given a user u and his observed locations $L^u_{\Omega_u} = \{\mathbf{l}^u_i | i \in \Omega_u\}$, we proceed as follows:

1. Identify the *home location* $\mathbf{l}^u_{\mathrm{H}}$ as the most frequent location in the observed location set $L^u_{\Omega_u}$ during the nighttime period $(22h, 9h)$. To affect the precision as slightly as possible, go to the next operation only if $\mathbf{l}^u_{\mathrm{H}}$ contains $\geq 80\%$ of the telecommunication events during the nighttime period.

2. Compute the *home period* as a set of nighttime time slots Ω^{H}_u. For that, we find the most probable period $t^u_{\mathrm{H}} \subseteq (22h, 9h)$ and examine $L^u_{\Omega_u}$ on a daily basis: time slots of a day are added into Ω^{H}_u if no location or only $\mathbf{l}^u_{\mathrm{H}}$ is found during the period t^u_{H} of that day.

3. Use $\mathbf{l}^u_{\mathrm{H}}$ as the estimated location for the time slots in Ω^{H}_u. Fill these time slots with $\mathbf{l}^u_{\mathrm{H}}$ in the CDR-based trajectory.

After the operations above, some missing locations in the CDR-based trajectory are estimated. The rest is estimated in the next step.

4.3 Temporal improved tensor data completion

In this step, all the remaining missing locations are inferred via the tensor factorization technique. As discussed in Sec. 3, the human mobility is redundant. By tenserizing the CDR-based trajectory, the redundancy of trajectory is reflected by the hyperparameters in the decomposition of the obtained tensor. Inferring these hyperparameters by the tensor factorization technique is equivalent to inferring the missing locations. Hereby the tensor factorization is adopt because it is proved to effective in recovering highly incomplete sensory environmental data [18, 19].

We first present the bootstrap approach named low-rank matrix factorization (`MF`). Because human footprints has daily redundancy [1, 5], we convert the CDR-based trajectory L^u into the

2D-tensor (matrix) $X^u \in \mathbb{R}^{N_D \times N_H}$ as follows:

$$X^u = \begin{bmatrix} \mathbf{l}^u_0 & \mathbf{l}^u_1 & \cdots & \mathbf{l}^u_{23} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{l}^u_{n_\mathrm{D}+0} & \mathbf{l}^u_{n_\mathrm{D}+1} & \cdots & \mathbf{l}^u_{n_\mathrm{D}+23} \end{bmatrix} \tag{5}$$

where N_D is the number of days in the period of study. Each row in X^u is a sub-trajectory representing the locations in a single day and thus $N_H = 48$. For simplicity, we use X instead of X^u in the remaining of this section.

We assume that X not have a full rank because of the redundancy in the CDR-based trajectory L^u. In this case, we use two matrices of hyper parameters, *i.e.*, $P \in \mathbb{R}^{N_D \times d}$ and $Q \in \mathbb{R}^{N_H \times d}$, as the d-rank approximation of X. An estimation of X as $\hat{X}$ can be obtained from P and Q, *i.e.*, $\hat{X} = PQ^T$ by solving the following optimization problem:

$$\arg\min_{L,R} ||\hat{X}_{\Omega_u} - X_{\Omega_u}||^2_F + \lambda(||L||^2_F + ||R||^2_F) \tag{6}$$

where $||\cdot||_F$ is the Frobenius norm. Note that this is a convex optimization problem [11] and has several existing solvers [18].

We see that the `MF` approach and the conversion from L^u to X^u make it possible to utilize the redundancy of one day period and infer the missing locations. Such redundancy also exists in the human trajectory for a longer period (*e.g.*, one week) but is weaker [1]. To reasonably utilize such redundancy as well as the redundancy of one day, we hereby present the temporal improved tensor data completion step. We use a time window and divide the period of study into several subperiods. Hence the 3D-tensor $\mathcal{X}^u \in \mathbb{R}^{N_\mathrm{WD} \times N_H \times N_W}$ is converted from L^u similarly as follows:

$$\mathcal{X}^u = [X^u_1, X^u_2, \cdots, X^u_{N_\mathrm{W}}], \tag{7}$$

where N_W is the number of subperiods, N_WD is the number of days in the time window, and X^u_i is the location matrix in the i-th subperiod. For simplicity, $\mathcal{X}$ is used instead of $\mathcal{X}^u$ hereafter.

As in the `MF` approach, we use tensors of hyperparameters to infer the missing values in $\mathcal{X}$. There are several ways to decompose a tensor [20] and we use the *Tucker* decomposition in this paper. $\mathcal{X}$ is factorized into a core tensor $\mathcal{C} \in \mathbb{R}^{d_{N_\mathrm{WD}} \times d_{N_H} \times d_{N_W}}$ and three matrices $P \in \mathbb{R}^{N_\mathrm{WD} \times d_{N_\mathrm{WD}}}$, $Q \in \mathbb{R}^{N_H \times d_{N_H}}$, and $R \in \mathbb{R}^{N_W \times d_{N_W}}$ where $d_{N_\mathrm{WD}} > d_{N_W}$ in order to emphasize the redundancy of one day. An estimation of $\mathcal{X}$ as $\hat{\mathcal{X}}$ can be obtained from these hyperparameters as follows:

$$\hat{\mathcal{X}} = \mathcal{C} \times_P P \times_Q Q \times_R R \tag{8}$$

where $\times_*$ is the *tensor-matrix multiplication* [20], for instance, $(\mathcal{C} \times_P P)_{pjk} = \sum_i \mathcal{C}_{ijk} U_{ip}$.

Before we build an convex optimization problem based on Eqn. 8, we further enhance the redundancy of one day in the estimation by adding a temporal constrain. For that, we import a constrain matrix $T = \mathrm{Toeplitz}(0, 1, -1)_{N_\mathrm{WD} \times N_\mathrm{WD}}$. TX^u_i contains the differences between positions of the same hour in two consecutive days in the i-th subperiod and $\mathcal{X}_P \times T$ contains the differences in all subperiods.

Combining the constrain and the parameters, we build the optimization problem to obtain the estimation $\hat{\mathcal{X}}$ in the following:

$$\begin{aligned} \arg\min_{P,Q,R} \quad & ||\hat{\mathcal{X}}_{\Omega_u} - \mathcal{X}_{\Omega_u}||^2_F \\ & + \lambda(||P||^2_F + ||Q||^2_F + ||R||^2_F) \cdot \\ & + \eta||\hat{\mathcal{X}} \times_P T||^2_F \end{aligned} \tag{9}$$

To solve this problem, we apply the *stochastic gradient descent* algorithm on the existing locations in L^u as in [19].

Overall, the tensor factorization infers all the remaining missing locations. The last operation in this step is to vectorize the completed tensor and obtain the estimated CDR-based trajectory that is complete. We use $\hat{L}^u_{\Omega^C_u}$ to represent the set of estimated locations for the missing time slots, *i.e.*, $\hat{L}^u_{\Omega^c_u} = \{\hat{\mathbf{l}}^u_i | i \in \Omega^c_u\}$.

4.4 Cell tower estimation

This step is to further enhance the accuracy of location in the estimated CDR-based trajectory. In most of the mobile phone datasets, the locations are the cell towers that handle telecommunication events. Nevertheless, the inferred locations in the set $\hat{L}^u_{\Omega^c_u}$ do not precisely match the corresponding cell towers because the tensor factorization step is not aware of any information about the cell towers. We design this cell tower estimation step to locate the cell tower for each inferred location. It is worth noting that this step is adopted only if the deployment of cell towers is a prior knowledge. Usually, a mobile dataset contains millions of telecommunication events attached to cell tower locations, and these events should cover all the cell towers. Hereby the set of cell towers is defined as $C = \{\mathbf{c}_j | 1 \leq j \leq N_{bs}\}$. $\mathbf{c}_j$ is a location vector and represents the j-th cell in the area; N_{bs} is the total number of cell towers in the area of study.

Given an inferred location $\hat{\mathbf{l}}^u_i \in \hat{L}^u_{\Omega^c_u}$ from the tensor factorization step, the goal is to find a cell tower to replace $\hat{\mathbf{l}}^u_i$ as the new inferred location $\hat{\mathbf{l}}^u_{i,cell}$. Mathematically, we need a metric function $f(\mathbf{c}, \mathbf{l})$ to measure the correlation between the cell tower and the inferred location, and then solve the following problem for the cell tower estimation:

$$\hat{\mathbf{l}}^u_{i,cell} = \arg\max_{\forall \mathbf{c}_j \in C} f(\mathbf{c}_j, \hat{\mathbf{l}}^u_i). \tag{10}$$

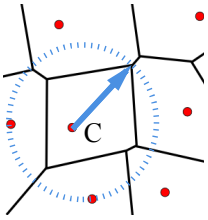


Figure 5: An example of computing the cell radius. For the cell C in the center of the figure, its radius is computed as the largest distance from the cell tower to the edge.

For the metric function $f(\mathbf{c}, \mathbf{l})$, the simple idea is to use the distance, *i.e.*, $|\mathbf{c} - \mathbf{l}|$, but is not good because it is known that the telecommunication event sometimes is not handled by the nearest cell tower to the device. For that, we propose a novel metric function combining the distance and the possible handover. We first estimate cell coverage by computing a Voronoi tessellation [21], and then compute for each cell tower the *cell radius* as the largest distance for the cell tower to its Voronoi polygon contour as illustrated in Fig. 5. We define the metric function as follows:

$$f(\mathbf{c}, \mathbf{l}) = \begin{cases} n^{\mathbf{c}}_e / n^{\text{total}}_e & \text{if } |\mathbf{c} - \mathbf{l}| \leq r_{\mathbf{c}} \\ 0 & \text{otherwise} \end{cases}, \tag{11}$$

where $r_{\mathbf{c}}$ represents the cell radius, $n^{\mathbf{c}}_e$ is the number of telecommunication events observed in the cell tower $\mathbf{c}$, and n^{total}_e is the total number of telecommunication events. Our metric function

ensure that the cell towers are selected as candidates, if their distances to the inferred location are less than their own cell radii, and the candidate with the highest probability of appearance is chosen as the new inferred location.
Conclusion: In this section, we propose the `TF-Home` approach to address the CDR completion problem. The approach is designed to deal with high data loss, heterogeneity, and redundancy seen in the CDR-based trajectory. We evaluate the performance of our approach in the next section, .

5 Performance Evaluation

5.1 Methods

We compare our proposed `TF-Home` approach with existing techniques for missing data inference. To verify the effectiveness of our approach, three classical techniques are adopted into comparison:

- `LastSeen`: this is the simplest solution for filling gaps of data. Particularly, each time slot without a location is filled by the nearest previous location in the CDR-based trajectory.

- `Interpolation`: this technique is proposed by Hoteit *et al.* [11] designed for completing the trajectory of telecommunication events with a high sampling rate. Particularly, the missing locations in the trajectory are estimated by the linear or cubic interpolation via the existing locations regarding the user's radius of gyration.

- `MF`: this is a typical solution for missing data interpolation. The CDR-based trajectory is tenderized to the matrix in the shape of (days, hours) and the missing values are estimated via the matrix factorization as introduced in Sec. 4.3.

5.2 Ground-truth

We leverage two fine-grained mobile phone datasets as our ground-truth because the CDR dataset introduced in Sec. 2 does not have enough completeness. The ground-truth datasets are as follows:

1. **Flow Dataset**. This dataset comes from the same source as the CDR dataset in Sec. 2 but has a far smaller population ($1,450$ users). These users have more location samples than their counterparts have in the CDR dataset, because they have all telecommunication events collected from call, sms and data traffic. Each user has a one-week trajectory in the period of July 2015, and each trajectory has locations observed in at least 18 hours per day. The location is the cell tower position as in the CDR dataset and thus the cell tower deployment is obtainable.

2. **Shanghai Dataset**. This dataset consists of 5K mobile phone subscribers in one of the major telecommunication operators in China. Each user has an two-week trajectory in the period of December 2016, and each trajectory has known locations in at least 21 hours per day. Each location is given directly as the estimated position with the precision of 200 meters but the cell tower deployment is unobtainable.

Note that we only use the weekday trajectory into the evaluation. Hence each user has a 5-day trajectory or a 10-day trajectory as ground-truth in the Mexico or Shanghai dataset, respectively.

5.3 Methodology

The procedure of simulation is designed as follows:

1. Expand duplicately the trajectory in the ground-truth datasets to 15/30/60/90-day periods to simulate the CDR-based trajectory, because the latter is usually provided in a long observing period.

2. Generate the CDR-trajectory with the completeness of 5/10/15/20/25/30%. For that, we manually eliminate some locations in the expanded ground-truth trajectory so as to create the *mimic* CDR-based trajectory, and we use the actual distribution of CDR in the elimination. For instance, to generate trajectories having a 10% completeness, we select all trajectories having the same degree of completeness in the CDR dataset, compute the hourly distribution of completeness, and then use this hourly distribution to determine the probability of eliminating the location in each hour. We generated one hundred mimic CDR-based trajectories from each expanded ground-truth trajectory.

3. Apply the methods for comparison and the `TF-Home` method to complete all the mimic CDR-based trajectory. Note that, on the trajectories in the Shanghai dataset, only the first two steps of `TF-Home` are applied.

5.4 Results

We evaluate the performance regarding two metrics: *distance error* and *cell estimation accuracy*. The distance error is computed as the average distance between the estimated and the actual location on all time slots having unknown locations. Mathematically, given a CDR-based trajectory L with a loss set Ω, the distance error of the trajectory is computed as follows:

$$\text{error}(\Omega, L) = \frac{1}{|\Omega|} \sum_{i \in \Omega} ||\mathbf{l}_i - \hat{\mathbf{l}}_i||_{\text{geo}} \tag{12}$$

where $\mathbf{l}_i$ and $\hat{\mathbf{l}}_i$ represents the actual and estimated location at the i-th time slot respectively. Fig. 6 show the mean distance error of all completed trajectories in each ground-truth data set. We can clearly observe the following:

- The distance error of the `TF-Home` approach is less than the ones of other comparison techniques. When the trajectory completeness $\geq 10\%$, the `TF-Home` approach can almost have the distance error below 2.6 km in the Mexico flow dataset and below 0.75 km in the Shanghai data set. It is worth noting that the size of area that the cell tower covers around 2 km^2 in the former and the location represents $200m \times 200m$ in the latter. Such distance error is relatively good regarding the location precision of the ground-truth datasets.

- The distance error decreases with the increasing of data completeness, and moreover, the differences among the techniques become smaller with the same increasing. This indicates that the increasing of mobility information contributes to all the methods. Still, the distance errors of the interpolation and last seen approaches are higher than the ones of the rest, indicating that utilizing the redundancy of human mobility helps to the completion a lot, particularly when the completeness is low.

- The `MF` approach has almost equivalent performance compared with the `TF-Home` approach when the completeness $\geq 15\%$, which indicates that, when the completeness is high, the

redundancy of human mobility makes major contribution in the data inference. Nevertheless, recall that in the CDR data set in Sec. 2, approximately 50% of the CDR-based trajectories have the completeness of 10% but only 10% have the completeness of 30%. We conclude that the `TF-Home` approach performs better and is more applicable considering the potential population in the mobile phone data set.

The cell estimation error is defined as the ratio of time slots having mistakes in estimating the cell tower, which mathematically defined as follows:

$$\text{cell-error}(\Omega, L, C) = 1 - \frac{1}{|\Omega|} \sum_{i \in \Omega} \mathbb{1}(\mathbf{l}_i = \mathbf{c}_i), \tag{13}$$

where $\mathbf{l}_i$ and $\mathbf{c}_i$ are the estimated and actual cell tower in the i-th time slot. Clearly, this metric is more strict than the distance error because it requires absolute accurate estimation. We plot the cell estimation error in Fig. 7 only for the Mexico dataset. The results show that, unless the completeness is extremely low (5%), the `TF-Home` approach outperforms the other ones regarding the cell estimation error. We see that when the completeness $\geq 10\%$, more than 60% of the cell towers can be accurately recovered by the `TF-Home` approach. Nevertheless, the simple `lastseen` approach is enough when the completeness $\geq 25\%$.

6 Related works

The incompleteness of human mobility information in the mobile dataset is a common challenge. In the literature, three major solutions are capable of overcoming such incompleteness: *(i)* estimating mobility features by indirect algorithms instead of direct ones that using the whole trajectory as in [2]; *(ii)* applying heavy filters on the data completeness when selecting the population of study as in [1, 2, 11], in which usually a small portion of the dataset population survives; *(iii)* focusing on mobility features that are not sensitive to the data completeness, such as temporal-uncorrelated features (*e.g.*, home/work locations [6] or trajectory uniqueness [4]). These solutions either shrink heavily the population of study or works only in several specific limited scenarios.

Several recent attentions have focused on the inference of human mobility information in the mobile dataset, which show potential in the CDR completion to recover fully or partially the mobility path. Ficek *et al.* [17] proposed a probabilistic inter-call mobility model based on Gaussian mixture, which was able to determine the human position between two consecutive communication events (calls or sms). Sahar *et al.* [11] proposed techniques for estimating complete human trajectories via a mixture of interpolation methods on CDR and validated the availability of their techniques on highly active users ($>$ 1000 samples/day). [13, 14] proposed adaptive techniques aiming at the CDR completion. Further, [14] showed that the interpolation method in [11] were less effective than their other techniques on relatively less active users ($>$ 30 samples/day). Yet the techniques in [13,14] only produced incomplete trajectories: temporal gaps with unidentified locations still exist.

It is worth noting that path reconstruction techniques for GPS samples, *e.g.* [15], could be available for the CDR completion on the user who has a large number of location samples. Nevertheless, the sampling rate of CDR (via voice calls) is far less than that of GPS surveys in general. Hence techniques designed for GPS surveys are not considered in this paper.

Unlike these works, we propose a novel CDR completion solution, which is designed to fully recover the human trajectory by filling all temporal gaps for common users who have a relatively high loss rate in their mobility information. Also, the data loss in the mobile dataset is also

investigated in this paper. This part is inspired by the work of Bolot *et al.* [16]. They proposed the distributions fitting the individual number of voice calls, number of partners, and voice call duration.

7 Conclusions

In this paper, we studied the dat completion problem in the mobile phone dataset. We investigated the data loss in the mobile phone dataset in terms of locations. Then by utilizing several human mobility characteristics, we proposed a novel CDR completion approach to infer the missing locations in the CDR-based trajectory. The proposed approach combined the power of data factorization and human routines. Experiments driven by actual mobility datasets illustrated that our approach outperformed the competitors.

References

[1] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, June 2008.

[2] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[3] C. Iovan, A.-M. Olteanu-Raimond, T. Couronné, and Z. Smoreda, "Moving and calling: Mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies," in *Geographic Information Science at the Heart of Europe*, pp. 247–265, Springer, 2013.

[4] Y. A. De Montjoye, C. A. Hidalgo, and M. Verleysen, "Unique in the crowd: The privacy bounds of human mobility," *Scientific reports*, 2013.

[5] C. M. Schneider, V. Belik, T. Couronné, Z. Smoreda, and M. C. González, "Unravelling daily human mobility motifs," *Journal of The Royal Society Interface*, vol. 10, pp. 20130246–20130246, July 2013.

[6] R. Ahas, S. Silm, E. Saluveer, and O. Järv, "Modelling Home and Work Locations of Populations Using Passive Mobile Positioning Data," in *Location Based Services and TeleCartography II*, pp. 301–315, Berlin, Heidelberg: Springer, Berlin, Heidelberg, 2009.

[7] H. Zang and J. C. Bolot, "Mining call and mobility data to improve paging efficiency in cellular networks," in *MobiCom '07: Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, (New York, New York, USA), pp. 123–134, ACM, Sept. 2007.

[8] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *Proceedings of the 18th international conference on World wide web*, pp. 791–800, ACM, 2009.

[9] K. Y. Lai, Z. Tari, and P. Bertok, "Supporting user mobility through cache relocation," *Mobile Information Systems*, vol. 1, no. 4, pp. 275–307, 2005.

[10] D. Naboulsi, M. Fiore, S. Ribot, and R. Stanica, "Large-scale mobile traffic analysis: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 124–161, 2016.

[11] S. Hoteit, S. Secci, S. Sobolevsky, C. Ratti, and G. Pujolle, "Estimating human trajectories and hotspots through mobile phone data," *Computer Networks*, vol. 64, pp. 296–307, 2014.

[12] H.-H. Jo, M. Karsai, J. Karikoski, and K. Kaski, "Spatiotemporal correlations of handset-based service usages," *EPJ Data Science*, vol. 1, p. 10, Nov 2012.

[13] G. Chen, A. C. Viana, and C. Sarraute, "Towards an adaptive completion of sparse call detail records for mobility analysis," in *Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on*, pp. 302–305, IEEE, 2017.

[14] S. Hoteit, G. Chen, A. Viana, and M. Fiore, "Filling the gaps: On the completion of sparse call detail records for mobility analysis," in *Proceedings of the Eleventh ACM Workshop on Challenged Networks*, CHANTS '16, (New York, NY, USA), pp. 45–50, ACM, 2016.

[15] E.-H. Chung and A. Shalaby, "A trip reconstruction tool for gps-based personal travel surveys," *Transportation Planning and Technology*, vol. 28, no. 5, pp. 381–401, 2005.

[16] M. Seshadri, S. Machiraju, A. Sridharan, J. Bolot, C. Faloutsos, and J. Leskove, "Mobile call graphs: beyond power-law and lognormal distributions," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 596–604, ACM, 2008.

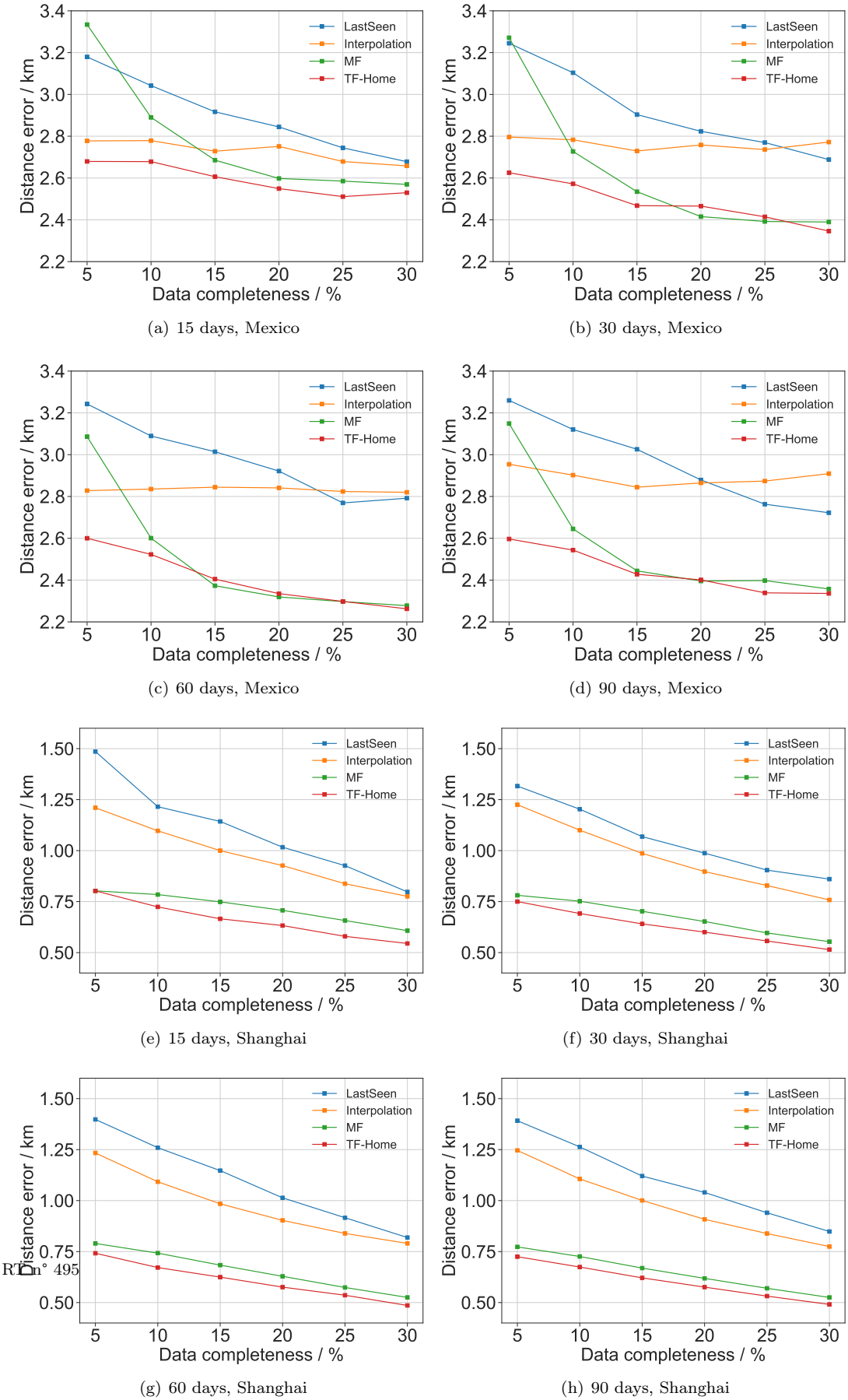
[17] M. Ficek and L. Kencl, "Inter-call mobility model: A spatio-temporal refinement of call data records using a gaussian mixture model," in *INFOCOM, 2012 Proceedings IEEE*, pp. 469–477, IEEE, 2012.

[18] L. Kong, M. Xia, X.-Y. Liu, G. Chen, Y. Gu, M.-Y. Wu, and X. Liu, "Data loss and reconstruction in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 11, pp. 2818–2828, 2014.

[19] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver, "Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering," in *Proceedings of the fourth ACM conference on Recommender systems*, pp. 79–86, ACM, 2010.

[20] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[21] J. Portela and M. Alencar, "Cellular network as a multiplicatively weighted voronoi diagram," in *Consumer Communications and Networking Conference, 2006. CCNC 2006. 3rd IEEE*, vol. 2, pp. 913–917, IEEE, 2006.



(a) 15 days, Mexico

(b) 30 days, Mexico

(c) 60 days, Mexico

(d) 90 days, Mexico

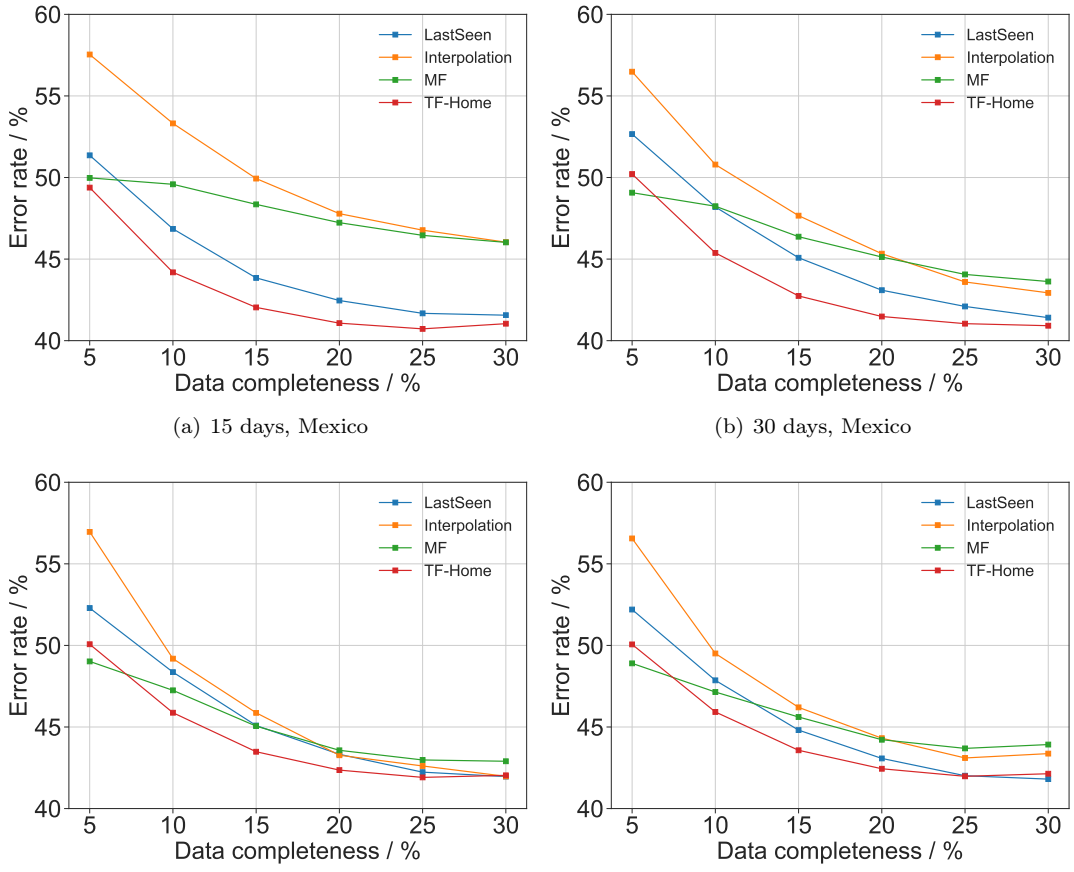
(e) 15 days, Shanghai

(f) 30 days, Shanghai

(g) 60 days, Shanghai

(h) 90 days, Shanghai

Figure 6: Performance on distance error



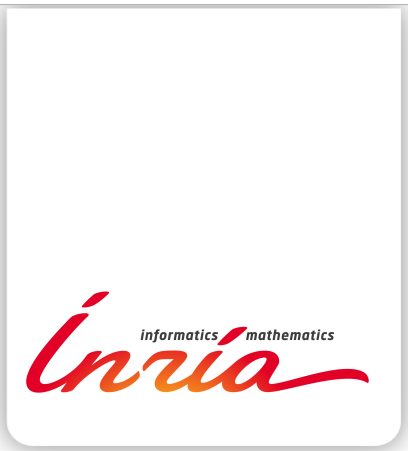
(a) 15 days, Mexico

(b) 30 days, Mexico

(c) 60 days, Mexico

(d) 90 days, Mexico

Figure 7: Performance on cell estimation error



RESEARCH CENTRE
SACLAY – ÎLE-DE-FRANCE

1 rue Honoré d'Estienne d'Orves
Bâtiment Alan Turing
Campus de l'École Polytechnique
91120 Palaiseau

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-0803