Contextual Location Imputation for Confined WiFi Trajectories

Elham Naghizade[1], Jeffrey Chan[2], Yongli Ren[2], and Martin Tomko[1]

[1] Department of Infrastructure Engineering, University of Melbourne, Australia
{enaghi, mtomko}@unimelb.edu.au
[2] Department of Computer Science, RMIT University, Australia
{jeffrey.chan, yongli.ren}@rmit.edu.au

Abstract. The analysis of mobility patterns from large-scale spatio-temporal datasets is key to personalised location-based applications. Datasets capturing user location are, however, often incomplete due to temporary failures of sensors, deliberate interruptions or because of data privacy restrictions. Effective location imputation is thus a critical processing step enabling mobility pattern mining from sparse data. To date, most urban trajectory prediction studies focused on coarse location prediction at city scale. In this paper we aim to infer the missing location information of individuals tracked within structured, mostly *confined* spaces such as a university campus or a mall. Such datasets may be collected by sensing user presence via WiFi sensing and consist of timestamped associations with the network's access points (APs). Such coarse location information imposes unique challenges to the location imputation problem. We present a contextual model that combines the regularity of individuals' visits to enable accurate imputation of missing locations in sparse indoor trajectories. This model also considers implicit social ties to capture similarities between individuals, applying Graph-regularized Nonnegative Matrix Factorization (GNMF) techniques. Our findings suggest that people' movement in confined spaces is largely habitual and their social ties plays a role in their less frequently visited locations.

Keywords: Spatio-temporal trajectories, Data imputation, Matrix factorization

1 Introduction

The uptake in sensor-enabled smartphones and devices has enabled capturing human mobility data at a fine-grained scale. Such rich spatio-temporal datasets facilitate the (i) understanding of individuals' movement patterns and potentially their intention, (ii) delivery of tailored recommendations and services to individuals or groups as well as (iii) understanding the space use patterns and provision of public and social applications that benefit from people-centric sensing.

In data mining, *data quality* is a key predictor of the success of analyses. Large spatio-temporal datasets are, however, often incomplete due to sensor failure or, low sampling rates to maximise a sensor's life. Location data imputation has been suggested as a solution to tackle this issue [8].

Research has focused primarily on large-scale outdoor mobility datasets. Such trajectories cover a larger space and are the location is captured at fine precision using GPS measurement at fine intervals. These approaches are not suitable to coarse trajectories based on proximity sensing (symbolic trajectories) [5]. We explore location imputation for trajectories recorded using WiFi sensing across multiple indoor or well defined confined spaces. These trajectories capture a user's proximity to a single access point. This imposes unique challenges to trajectory sensing: the captured locations are coarse and may be sparse. A range of efforts has focused on using additional information such as Received Signal Strength Indicators to improve location points, or used crowd-sourced data to calibrate the location information and interpolate missing points [16, 7].

Social ties between visitors to these environments represent a significant source of information that remains largely unexplored in location prediction. Human physical behaviour is strongly mediated by social interactions. A key question is *how can social ties between individuals improve location imputation models in coarse trajectories?* Several studies explored the ability to predict social links between users given their trajectories [12, 15], primarily over large areas and using explicit social links (e.g., using Facebook data). Here, we explore physical co-presence to explore implicit social ties.

Factorization techniques has been successfully used in recommender systems – systems that estimate users' preferences (e.g., rankings) for unrated items. As such, they address a similar problem – the inference of missing values associated with a user's behaviour. Here, we utilise the graph-regularized non-negative matrix factorization technique (GNMF) to impute missing locations., based on the intuition that individuals with stronger social ties express higher trajectory similarity and are also more likely to visit specific locations together. Motivated by the approach proposed in [9], we build user profiles that capture users spatio-temporal behaviour (regularity, temporal entropy and gaps of visits). We further implicitly capture social ties between users. These are inferred based on physical co-presence, as actual information about users' relationships may not be known. A co-presence graph captures the strength of users' associations through edge weights (frequency, duration and location of co-presence between pairs of users). Finally, we build an affinity graph between locations, with edges capturing the number of times each pair of locations has been visited together in a trajectory.

Using two real-world datasets (a university campus and a shopping mall) we explore two location imputation scenarios: i) the case of partially missing location information in individual trajectories (i.e., due to sensor failure), and ii) the case of location imputation for new users, never observed in the environment in the past (cold start). Our experiments show that the consideration of social co-presence improves the performance of location imputation for, in particular, less frequently visited locations of a user, e.g., top 10 to 20 locations. This is significant, as these are individually important locations that are not shared with others. We also observe that the history of users' visits plays an important role in predicting missing locations.

2 Related Work

2.1 Collaborative filtering in location recommendation

Collaborative filtering techniques have been successfully used for recommending movies, books, points of interest to users whose rating for a certain item is not known. This is why we focus on it for our location imputation problem.

Zheng et al. [19] provide an example of successfully using tensor factorization to decompose a user-location-activity tensor and provide suggestions on points of interest or activities using GPS trajectories of the individuals and their social network profiles. Yin et al. [17] use a user's location history as well as local preference to provide location recommendations. This approach particularly focus on recommending location/event to users with known profiles, but no location history. The model in [11], takes the the distance between regions, their popularity and user mobility patterns to build a geographical probabilistic latent model for location recommendation.

Similar to [6], this work assume an implicit feedback based on the frequency of users' visit to each places. The collaborative filtering model proposed in [9], leverages users' mobility to retrieve user similarities, however, it assumes an implicit positive feedback for locations that the user has visited rather than considering a negative feedback to the unvisited locations. It also uses the retrieved user profile or location features as the additional contextual information to improve the model. Further, the authors in [18] and [10] have focused on time-aware location recommendation problem and utilise the temporal affects of visiting places as well as spatial influences when providing location recommendations. Unlike the above-mentioned studies, we focus on a confined area and proximity-based trajectories rather than explicit location check-ins or GPS trajectories within large areas. Also, most of these studies assume the availability of other sources of data to extract additional information and boost the performance of the algorithm, while we do not use any information other than the mobility data.

2.2 Social Ties and mobility patterns

The study in [4] is one of the early attempts to find the relationship between individuals' movement patterns and their online social network. Similarly, the authors in [2] explore the spatial and social influences on individuals' mobility patterns. Mobility patterns has been used to predict users' social ties, including using an entropy-based measure of the diversity and frequency of their co-locations [13] and similarities between users' trajectories [14]. Similar to [14], we use how specific the location of co-presence is.

On the other hand, social ties have been consequently used to predict users' location. In [12], McGee et al build a network based on social ties to estimate the location of users. In particular, they used the links between users in Twitter as well as users' profile information to estimate the distance between any pair of users. The authors in [15], proposed a hybrid model to capture both the regularity of individuals' repetitive patterns as well as the influence of their social network, to predict their location. This approach also uses additional sources of information and ensures improved location information through heterogeneous data sources. These studies mainly focus on low resolution location predictions, e.g., city level resolutions.

3 Problem Description

Consider a mobility dataset D consisting of observations o of tracked individual users. Each observation o, $o = \{id, t, l\}$, is a tuple of a user, identified by a unique id, at (symbolic) location $l \in L$ at timestamp t. Assuming time to be discrete offsets from an arbitrary origin, the time of the observations belongs to a range between $[0, T_{max}]$, where T_{max} is the timestamp of the last observation in the dataset. Dividing the timestamps into intervals of window size w, results in a set of distinct trajectories of an individual:

For all users u_i from the set of users U ($\forall u_i \in U$) we have a set of sequences, s_j, in the form of $s_j = \{(l_1, t_1^d), (l_2, t_2^d), ..., (l_m, t_m^d)\}$ where $1 \leq i \leq M$ and $t^d \in \frac{T_{max}}{w}$. Assuming N_i is the number of u_i's trajectories in the dataset, we denote the set of distinct trajectories of u_i as $S_i = \{s_j | j = 1, 2, ...N_i\}$. In this paper we assume a day as the window size w, i.e., a user has at most one trajectory per day.

Table 1: Frequently used notations.

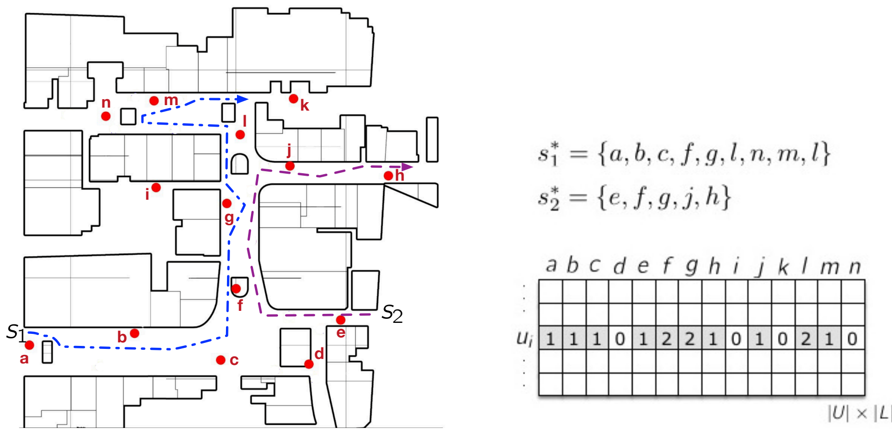
Symbol	Description	Symbol	Description
A	User/Location matrix	α	Confidence weighting function
C	User/Location confidence matrix	G_s, G_l	Implicit social and spatial(location) graphs
P	User profile matrix	λ_s, λ_l	Social and location graph weighting parameters
M	Number of users	ω	Social strength parameter
N	Number of locations	ρ	Regularization parameter
k	Number of latent features	K	top location query parameter

We assume some sessions are either partially or completely missing in D. We focus on the location information in the data and aim to impute users' location. One possible formalisation of this problem is to represent individuals' visited locations in a matrix:

Definition 1 - User-location Matrix (A): From the set of users' sequences, S_i, one can derive a matrix of size $M \times N$, where N is the cardinality of the location set and $A_{i,n}$ corresponds to the number of times u_i has been observed at l_n over all his/her trajectories. For any location n that is not visited by the i_{th} user, $A_{i,n} = 0$. It is noteworthy that elements in A do not consider the sequence of a user's movement, but rather the frequency of being seen by a certain access point during all visits.

Intuition: Individuals movement behaviour in an indoor space largely follows a frequent pattern. Intuitively, we would like to use the similarities between such movement pattern (reflected in the frequency of visits to certain locations) as well as user similarities that is potentially reflected in their temporal profile in order to impute the missing location frequencies. Furthermore, being socially connected to others can be reflected in an individual's movement pattern [14, 15, 13]. We aim to leverage the knowledge on potential social ties to improve our location imputation model.

Figure 1a depicts two sample trajectories of a user (dashed lines). The red circles represent the access points and Figure 1b shows the corresponding row in A for this scenario. Thus, assuming some location information of u_i in Figure 1b is missing and the respective row in A is then $[1, 1, ?, 0, 1, ?, 2, ?, 0, 1, 0, ?, 1, 0]$, the aim of our location imputation technique is to predict the missing values, shown as question marks.



(a) Space layout.

(b) User-location Matrix.

Fig. 1: Two sample trajectories of a user in an indoor space. The red circles indicate the location of access points. Note that $s^*_{1,2}$ represent the footprint of a session and the time of the observations is not captured in the user-location matrix.

4 Matrix Factorization for Location Imputation

4.1 Location Imputation with Implicit Feedback

One possible approach to infer unknown values in A is to apply matrix factorization (MF) techniques and decompose A into two matrices. The underlying idea is that MF uses the data of other users to discover the latent features that govern the interaction between uses and locations.

Given that matrix A has missing values, i.e., missing location information for some users, and assuming that we have k latent features, the aim of MF is to find two matrices $U_{M\times k}$ and $V_{k\times N}$ that approximate A in the following way:

$$U \times V = A' \approx A \quad where \quad a'_{ij} = \sum_{k=1}^{k} u_{ik} v_{kj}$$

U intuitively captures user similarities in the k-dimensional latent feature space and V reflects location similarities. The product of the obtained U and V provides the estimate values for missing location information.

As discussed in [9], in a mobility dataset, a visit to a certain location can reflect an individual's positive feedback to that location. However, not visiting a location does not necessarily imply negative feedback. In our setting, an individual can actually be present at a location, but not accessing the WiFi network would exacerbate this issue, e.g., having the mobile phone off. Hence, similar to [6], we define a confidence matrix, C, whose elements are computed as follows:

$$c_{i,j} = \begin{cases} \alpha(a_{i,j}) & \text{if } a_{i,j} > 0 \\ 1 & \text{otherwise,} \end{cases}$$

where α is a monotonically increasing function. Having a binary matrix R, where $r_{u,l} = 1$ if u_i has visited l_j at least once during his/her visits, the objective function can be expressed as minimizing the following error (ρ is the regularization parameter):

$$\mathscr{O} = \sum_{i,j} c_{i,j}(r_{ij} - u_{ik}v_{kj})^2 + \frac{\rho}{2}(\|U\|_F^2 + \|V\|_F^2)$$

4.2 Contextual Imputation

The direct factorization does not use the contextual information to improve the location predictions. In this paper we assume no additional source of information about the users is available and propose to create a user profile matrix, P, which mainly incorporates temporal features of users' visits. We focus on the following factors:

- *The average duration* of each visit, which is determined as $1/N_i \sum_{j=1}^{N_i}(t_{end\,j} - t_{start\,j})$, where $t_{start\,j}$ and $t_{end\,j}$ are the first and last point of trajectory s_j of user i.
- *Regularity* of the visits captures the predictability of an individual's visits with respect to time. We compute the entropy of a user's visits given specific days of the week, hour of the day as well as the combination of both, i.e., $\{D\}, \{H\}, \{D,H\}$ to estimate whether or not they follow some repetitive temporal pattern.
- *Temporal Gap* between the visits reflects how often a user is observed in the dataset. The regularity measure cannot differentiate regular visits that happen daily from those that happen monthly, hence we compute the temporal gap, in terms of the intervening number of days between a user's consecutive visit.

The above-mentioned factors are estimated for each individual and form a $M \times k'$ matrix, where k' is the number of derived factors in P. We augment U with this information and create $\hat{U}_{(M+k')\times k}$ matrix. Following the proposed algorithm in [9], we can rewrite the objective function as:

$$\mathscr{O} = \sum_{i,j} c_{i,j}(r_{ij} - \hat{u}_{ik}v_{kj})^2 + \frac{\rho}{2}(\|\hat{U}\|_F^2 + \|V\|_F^2)$$

4.3 Implicit Social Ties and GNMF

The Contextual Imputation approach aims to capture users' similarity to assign locations to users, however, the features discussed in the previous section does not consider potential social ties. As discussed in Section 2, mobility patterns of individuals can be strongly correlated to their social dependencies. As a result, we propose to leverage such connections to improve our location imputation task. Similar to the previous section, we assume that no additional source of data, e.g., social networks, is available and aim to build an implicit social graph using the available movement data.

To create our implicit social graph, G_s, we build a co-presence graph where there is an edge between any pair of users who have been present at a given location at least for one timestamp. Given that such graph may be noisy (since it adds an edge between random users who are at a popular location at the same time), we use the following parameters in addition to the co-presence frequency to estimate the social strength between any pair of individual:

- *Relative Average Duration:* The duration of co-occurence at each session can be an indicator of the strength of the relationship, however, considering the absolute duration of the co-presence has two weaknesses as i) it does not consider the overall length of users' trips, i.e., a co-presence of for instance 10 minutes for a 10-minute trip is much more intense compared to a 60-minute trip and ii) it has not an absolute upper-bound. Hence, the relative average duration computes the duration of co-presence divided by the duration of users' trips:

$$d_{i,j} = 1/N \sum_{n=1}^{f_{i,j}} \frac{d_{ij,n}}{min(d_{i,n}, d_{j,n})}$$

where $1 \leq n \leq f_{i,j}$ is the frequency of co-presence between u_i and u_j and $d_{i,n}$,$d_{j,n}$, $d_{ij,n}$ denote the duration of u_ith, duration of u_jth and the duration of their co-presence, at the nth session respectively.
- *Specificity of the co-presence:* Specificity aims to capture the popularity of the location where the co-presence has occurred. This aims to differentiate between popular locations such as main entrances or food courts, where random users may be seen at the same time, compared to less popular locations that can better reflect potential social ties. For any pair of users, (u_i, u_j) we compute the location specificity, i.e., $z_{i,j}$, at location l as the number of times u_i and u_j co-presence has been observed at l divided by the total number of co-presences occurred at l. We further average the location specificity for all the co-presences of u_i and u_j.

We define a decay function, $score(u_i, u_j) = \omega e^{f_{i,j} * d_{i,j} * log(\hat{z}_{i,j})}$, to compute the social strength between any pair of users. This function assigns lower weights to the edges that are likely to be noisy.

We use the GNMF technique to incorporate potential social ties in our model. GNMF builds a nearest neighbour grap, W, using $score(.)$ to determine the neighbours, and aims to decompose the original matrix in a way that connected points in the graph are closer in the latent space. Hence, similar to [1], we define our objective function as:

$$\mathscr{O} = \|A - UV\|_F^2 + \lambda Tr(U^T L U) + \frac{\rho}{2}(\|U\|_F^2 + \|V\|_F^2)$$

where L is the graph Laplacian [3] of the nearest neighbour graph and $Tr(.)$ is the trace of the matrix. L can be estimated as $D - W$, where D is a diagonal matrix whose entries are column sums of W.

Furthermore, we can build a graph, G_l to capture the similarity between the location points, where the edges indicates how frequently any pair of locations are visited together. The objective function can then be expressed as:

$$\mathscr{O} = \|A - UV\|_F^2 + \lambda_s Tr(U^T L_s U) + \lambda_l Tr(V^T L_l V) + \frac{\rho}{2}(\|U\|_F^2 + \|V\|_F^2)$$

and the update rules to minimize $\mathscr{O}$ are as follows:

$$u_{ij} \leftarrow u_{ij} \frac{(AV + \lambda_s W_s U)_{ij}}{(UV^T V + \lambda_s D_s U)_{ij}} \quad v_{ij} \leftarrow v_{ij} \frac{(A^T U + \lambda_l W_l V)_{ij}}{(VU^T U + \lambda_l D_l V)_{ij}}$$

5 Experiments

5.1 Dataset and Experimental Setup

Dataset: In this work we evaluate the performance of our location imputation algorithm using the following two mobility datasets:

D_1: This dataset is the collection of users connecting to the WiFi network of a University over a period of 22 days (10 days in May and 22 days in September) in 2016. We focused on a coarser location granularity (building level) to address the sparsity issue. We focused on individuals that have at least 2 sessions in the dataset and further filtered users whose overall footprint perfectly matches their daily visited locations (As shown in Figure 2c, we compute the mutual information between daily sessions of each user and the set of all their visited locations. For more than 18% of users in the dataset, the average NMI is equal to 1, i.e., they visited the same set of locations in every session). Hence, we obtained a dataset of $\approx 120,000$ users visiting 118 buildings.

D_2: This dataset has been collected from over 120,000 anonymized users connecting to the operating Wi-Fi network of a shopping mall in the city of X between September 2012 to October 2013. There are 67 Wi-Fi access points in an area of around 90,000 square meters. We focus on users who have visited the mall at least 5 times (5% of the users).

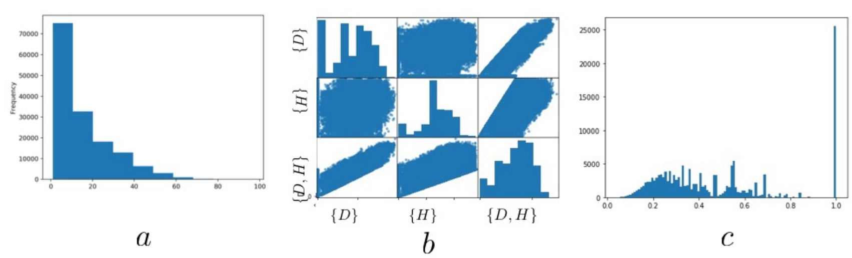


Fig. 2: Properties of the sessions in D_1. Figure 2a shows the distribution of number of unique locations visited by users. Figure 2b shows the correlation between Day, Hour and Day/Hour entropy. Figure 2c shows the distribution of normalised mutual information between users' daily trips and the overall footprint.

Methods: To evaluate our approach consider the following scenarios:

f-fold missing: We adopt an f fold cross validation framework where for each user we randomly split their mobility data into f folds and for each fold we train our model based on $f-1$ folds and test its performance to predict the visited locations on the remaining fold. Our default is a 2 fold cross validation which assumes half of the location information of all users is missing. For D_1, we train our model based on the sessions in one period and test it on the other.

cold start: In this scenario we evaluate the perfomance of our model in finding the top locations of new users. We use a 10 fold cross validation framework where the model is trained using the entire location history of 90% fold of the users and is evaluated based on its success in inferring the location of the remaining users.

We consider four varieties of the GNMF-based models: a model that only uses the location graph ($\lambda_s = 0$), L, a model that only considers the users' social graph, called U, a model that uses both graphs to make the prediction, called LU, and a model that uses users' profile when building the users' graph, called $LU + P$.

We compare the performance of the GNMF-based model with approaches proposed in [6] and [9], denoted as WRMF and ICCF respectively. For the f-fold missing scenario, we assume a model, called $HIST$, that returns the observed set of locations in the train set as the estimates of the missing locations, i.e., it relies on the history of users' location. We also implemented a Frequency-based Imputation technique (FI) for the cold start scenario that selects the top K locations visited by all users who have complete location information as the result of the top K query for any new user.

Metric: Similar to [6, 9] we used the precision at K and recall at K to evaluate the performance of our algorithm. For each user, we sort the score for each location and select the top K predicted locations. The precision at K and recall at K, denoted as $p@K$ and $r@K$ respectively are therefore computed as:

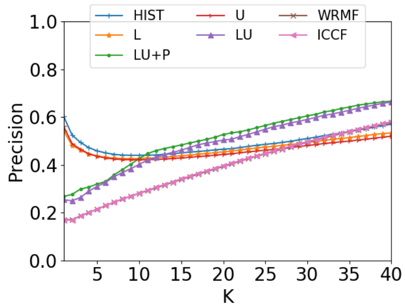
$$p@K = \frac{1}{M}\sum_{i=1}^{M}\frac{|T_i(K)\cap V_i|}{K}, \quad r@K = \frac{1}{M}\sum_{i=1}^{M}\frac{|T_i(K)\cap V_i|}{|V_i|}$$

where $T_i(K)$ is the set of top K predicted locations and V_i is the set of actual visited locations for u_i. Note that $r@K$ does not necessarily reach 1.

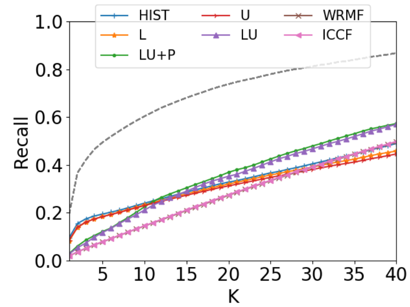
Setting: Our default number of latent features, k is set to 40 for D_1 and 20 for D_2 since the location set in D_2 is much smaller. Also, the number of factors that are considered when building P is 6 in our work since we add the average and maximum gap between users' visits to our feature set. Prior to feeding P to the model, we perform PCA to reduce the dimensionality of P to 3 since we observed that the derived temporal features in our datasets are moderately correlated (Figure 3b). When building the affinity graph, we set $n = 20\% * M$ and $\omega = 10$. We also used a similar function suggested in [9] to build our confidence matrix: $\alpha(c_{i,j}) = 1 + log(1 + c_{i,j} * 10^{\varepsilon})$, where $\varepsilon = 300$. We also set λ_s and λ_l to 0.5, 100 for D_1 and 0.05, 0.1 for D_2. Also, $\rho = 0.05$.

5.2 Results

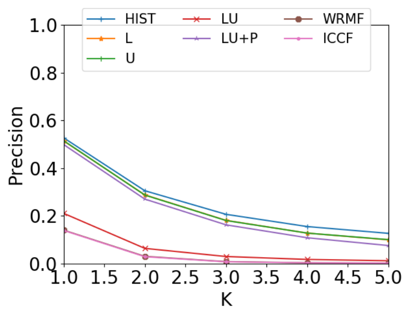
Figure 3 shows the performance of our proposed model with regard to the first scenario. Figure 3a and 3b consider the case where the imputation models are trained based on sessions in May and are then tested on sessions in September (D_1). As can be seen, the naive baseline outperforms all varieties of our proposed model as well as ICCF and WRMF for smaller Ks. However, with an increase in K, i.e., more specific or less frequent locations, the model that uses both location and user information outperforms the other methods. This suggests that in confined spaces, individuals' habits plays the most important role in predicting their top visited locations. Another interesting observation



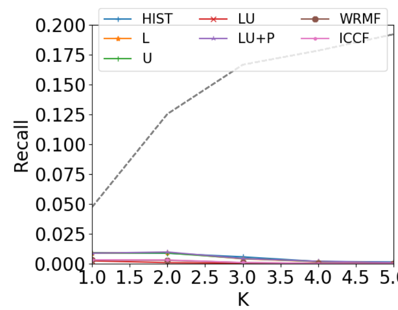
(a) $p@K$ - monthly split for D_1.



(b) $r@K$ - monthly split for D_1.



(c) $p@K$ - random split for D_2.

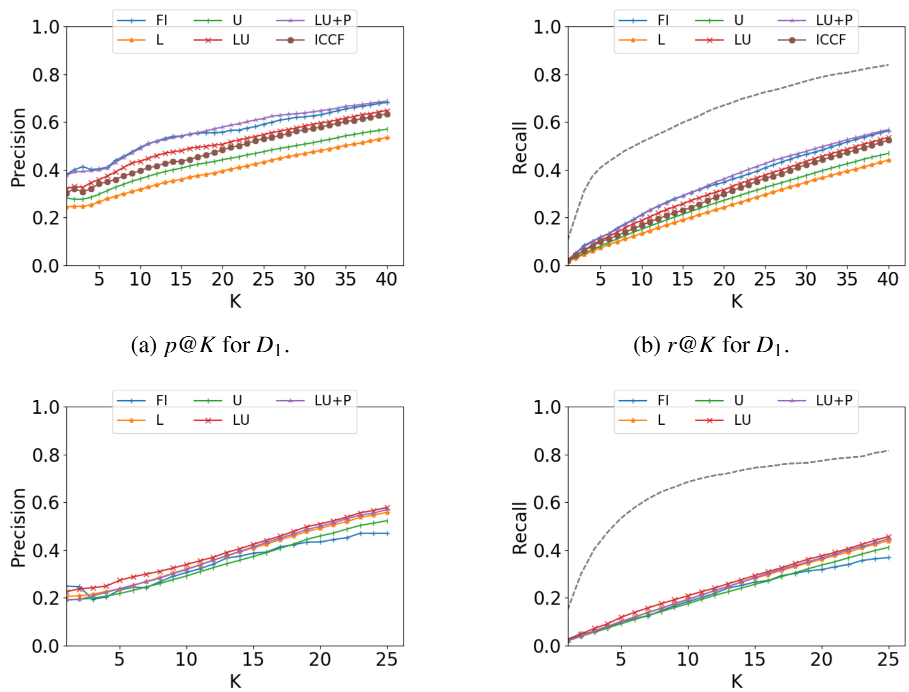


(d) $r@K$ - random split for D_2.

Fig. 3: The performance of our location imputation models in the case that 50% of the location data for all users is missing. The dashed grey line shows the highest achievable recall@K. Also, note that we changed the y-axis scale in Figure3d for visibility.

is that when user profile is used as the extra side information is a similarity graph, it can actually help to improve the performance of the GNMF-based model. However, the same trend is not observed when it is added as extra feature columns in the matrix, and is mixed with the location data (ICCF versus WRMF). We do not include the performance of WRMF in Figure 4 since the same observation holds in that scenario. As can be seen in Figure 3c and 3d, the same trend is not observed for D_2, which may be due to the limited number of options in a shopping mall.

Figure 4 depicts the performance of our proposed model in the cold case scenario. It can be observed that considering spatial and social influences as well as users' profile, i.e., $LU+P$, improves the performance of the model, however, FI performs almost as good as $LU+P$. This may be due to the fact that the new users are predicted to visit the most popular places, such as libraries in a campus and food court in a shopping mall. This assumption seems to be largely true in confined spaces.



(a) $p@K$ for D_1.

(b) $r@K$ for D_1.

(c) $p@K$ for D_2.

(d) $r@K$ for D_2.

Fig. 4: The performance of our location imputation models in the cold start scenario.

6 Conclusion

In this paper, we proposed a data imputation model for sparse spatio temporal trajectories. In contrast to most previous research, we have focused on constrained and indoor environments, such as a large, multi-building campus environment or a complex shopping mall. The highly regular user behaviour and sparsity of the data capturing their behaviour poses a challenge to sophisticated imputation algorithms. First, the highly regular nature of the movements renders the prediction of the top K locations (here, up to $K = 10$) trivial. We see that the naïve baseline matching the users to the most frequently visited locatins in the dataset performs surprisingly well, and significantly outperforms previous state-of-the-art-models, WRMF and ICCF. Yet, the inclusion of physical associations between visitors, captured by their physical co-occurrences outperforms all models for the less frequently visited locations. Our findings is consistent with the observations of the authors in [2]: 'social relationships explain about 10% to 30% of all human movement, while periodic behaviour explains substantially more, 50% to 70% of the behaviour.' As the value in sophisticated location prediction lies in the ability to reflect on the less frequently visited locations, our $LU + P$ model improves on the state-of-the-art and can support personalised services.

7 Acknowledgement

This research is supported by a Linkage Project grant of the Australian Research Council (LP120200413).

References

1. D. Cai, X. He, J. Han, and T. S. Huang. Graph regularized nonnegative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.
2. E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: User movement in location-based social networks. In *KDD*, 2011.
3. F. R. Chung. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.
4. J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh. Bridging the gap between physical location and online social networks. In *UBICOMP*, 2010.
5. R. H. Guting, F. Valdés, and M. L. Damiani. Symbolic trajectories. *ACM Transactions on Spatial Algorithms and Systems*, 2015.
6. Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. In *Procceedings of ICDM*, pages 263–272. Ieee, 2008.
7. Y. Kim, H. Shin, and H. Cha. Smartphone-based wi-fi pedestrian-tracking system tolerating the rss variance problem. In *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*, 2012.
8. W. Li, Y. Hu, X. Fu, S. Lu, and D. Chen. Cooperative positioning and tracking in disruption tolerant networks. *IEEE Transactions on Parallel and Distributed Systems*, 2015.
9. D. Lian, Y. Ge, F. Zhang, N. J. Yuan, X. Xie, T. Zhou, and Y. Rui. Content-aware collaborative filtering for location recommendation based on human mobility data. In *Proceedings of ICDM*, pages 261–270. IEEE, 2015.
10. D. Lian, Z. Zhang, Y. Ge, F. Zhang, N. J. Yuan, and X. Xie. Regularized content-aware tensor factorization meets temporal-aware location recommendation. In *ICDM*, 2016.
11. B. Liu, Y. Fu, Z. Yao, and H. Xiong. Learning geographical preferences for point-of-interest recommendation. In *Proceedings of KDD*, pages 1043–1051, New York, NY, USA, 2013.
12. J. McGee, J. Caverlee, and Z. Cheng. Location prediction in social media based on tie strength. In *Proceedings of CIKM*, 2013.
13. H. Pham, C. Shahabi, and Y. Liu. Inferring social strength from spatiotemporal data. *ACM Trans. Database Syst.*, 2016.
14. D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi. Human mobility, social ties, and link prediction. In *ACM SIGKDD*, 2011.
15. Y. Wang, N. J. Yuan, D. Lian, L. Xu, X. Xie, E. Chen, and Y. Rui. Regularity and conformity: Location prediction using heterogeneous mobility data. In *Proceedings of KDD*, 2015.
16. A. Ye, J. Sao, and Q. Jian. A robust location fingerprint based on differential signal strength and dynamic linear interpolation. *Security and Communication Networks*, 9(16):3618–3626, 2016. SCN-15-0656.R2.
17. H. Yin, Y. Sun, B. Cui, Z. Hu, and L. Chen. Lcars: A location-content-aware recommender system. In *Proceedings of KDD*, 2013.
18. Q. Yuan, G. Cong, and A. Sun. Graph-based point-of-interest recommendation with geographical and temporal influences. In *Proceedings of CIKM*, 2014.
19. V. W. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang. Collaborative filtering meets mobile recommendation: A user-centered approach. In *Proceedings of AAAI*, 2010.