Mining Location-Based Social Networks: A Predictive Perspective

Defu Lian[†§], Xing Xie[§], Fuzheng Zhang[§], Nicholas J. Yuan[§], Tao Zhou[†], Yong Rui[§]
[†]Big Data Research Center, University of Electronic Science and Technology of China
[§]Microsoft Research, Beijing, China
dove@uestc.edu.cn, {xingx,nichy,v-fuz,yongrui}@microsoft.com, zhutou@ustc.edu

Abstract

With the development of location-based social networks, an increasing amount of individual mobility data accumulate over time. The more mobility data are collected, the better we can understand the mobility patterns of users. At the same time, we know a great deal about online social relationships between users, providing new opportunities for mobility prediction. This paper introduces a novelty-seeking driven predictive framework for mining location-based social networks that embraces not only a bunch of Markov-based predictors but also a series of location recommendation algorithms. The core of this predictive framework is the cooperation mechanism between these two distinct models, determining the propensity of seeking novel and interesting locations.

1 Introduction

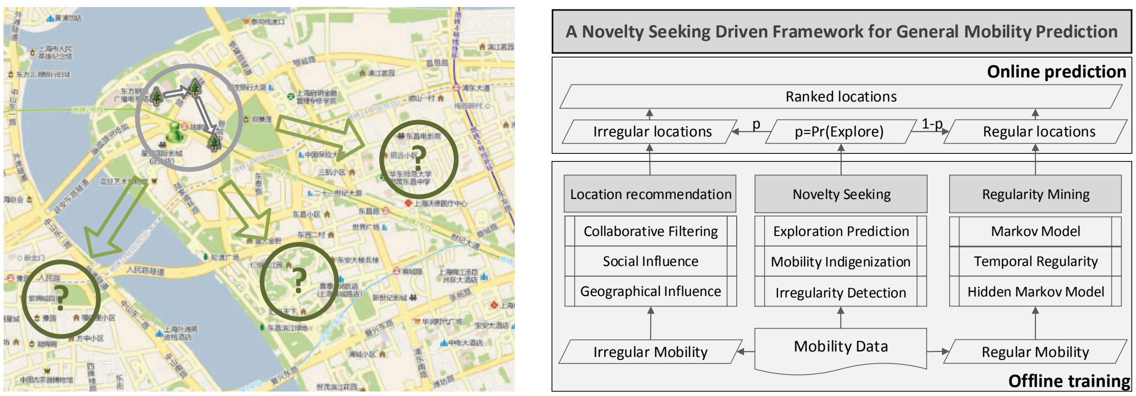
With the proliferation of smart phones and the advance in positioning technologies, location information can be acquired more easily than ever before. This development has led to the flourishing of a new kind of social network service, known as location-based social networks (LBSNs), such as Foursquare, Gowalla, and so on. In these LBSNs, people can not only track and share individual location-related information, but also learn collaborative social knowledge. Thus, a large amount of mobility data, such as check-ins (announcing a user's current location), have been collected, along with online social relationships between users. The more these data are collected, the better we can understand individual and crowd mobility patterns, and the more accurately we can predict future locations.

Mobility prediction plays important roles in urban planning [12], traffic forecasting [13], advertising, and recommendations [36], and has thus attracted lots of attention in the past decade. A typical scenario is shown in Fig 1(a). Past mobility data, such as GPS trajectories, sequences of Wifi access points, and cell tower traces, are either of coarse positioning granularity but passively recorded or only collected actively by a small number of volunteers. Thus, the collected data may be large scale, but redundant, so that the research for mobility prediction has mainly focused on frequent pattern mining. With the development of location-based social networks, mobility prediction is becoming a hot research topic once again. This is, on one hand, because mobility data are actively collected from a large number of users connected by online social networks; on the other hand, the introduction of social relationships provides new opportunities for mobility understanding and prediction since it has been observed that mobility behaviors, particularly long-distance travel, are more influenced by

Copyright 2015 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.
Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

social network ties [6]. At the same time, the locations are of extremely fine granularity (e.g., a room in an office) so that mobility patterns are much less redundant. Since users may not have an impetus to record their regular behaviors, some movement behaviors are missed. Due to these characteristics, mobility prediction on location-based social networks faces several challenges. First, mobility data are extremely sparse, so that only a small number of frequent patterns and only a portion of user preferences are implied. Second, more irregular behaviors are presented in the mobility data from LBSNs, increasing the difficulty of prediction and urgently requiring irregularity mobility prediction. Third, the collected check-ins tend to be noisy since check-ins don't necessarily imply a physical visit, so that mobility behaviors do not reveal an individual's full preferences.

To address these challenges, we start by analyzing the mobility data from location-based social networks in two ways to understand the distinct characteristics of mobility patterns. 1) *Spatial analysis*, is conducted on this mobility data to understand individual spatial distribution and the distance distribution between consecutively visited locations, given that regularly and irregularly visited locations coexist in the mobility data. 2) *Temporal analysis*, is achieved by delving into this mobility data, to determine the significance and strength of temporal regularity and Markov dependence.



(a) (b)

Figure 1: (a) A typical scenario for next check-in location prediction; (b) A novelty-seeking driven framework for general mobility prediction

Following the analysis of mobility data, we introduce a novelty-seeking driven predictive framework for mobility prediction, which consists of three components, as shown in Fig 1(b). 1) *Regularity mining for regular mobility prediction*, which includes a temporal-based regularity model and Markov-based predictors [16]. To address the sparsity challenge, we exploit kernel smoothing for regularity estimation and interpolation techniques for integrating different orders of Markov model. And we further analyze the limit of predictability by calculating the Kolmogorov entropy of trajectories, where the power of all Markov models from zero-order to infinity-order are taken into account [18]. 2) *Recommendation techniques for irregular mobility prediction*. Obviously, it is difficult for Markov-based models to predict the irregular mobility behaviors, such as visiting novel but appealing restaurants, but such behaviors are still subject to geographical restriction and are preference driven. Additionally, irregular mobility behaviors are probably affected by social influence since they may be more likely to involve distant travel. Thus, we introduce into the predictive framework the second component: a series of location recommendation algorithms that capture these three factors. In these algorithms, to alleviate the data sparsity when presenting individual preference, we resort to the histories of similar users and friends for collaboration and use geographical constraint to discover the highest possible negatively preferred locations. To reduce the effect of the noise when presenting user preference, we treat the data as an indication of positive and negative preference with vastly varying confidence. 3) *Mining propensity of novelty seeking*. In order to jointly predict both regular and irregular locations that a user will visit next, we introduce the core component, addressing the cooperation mechanism between these two distinct models by determining the propensity of seeking a novel and attractive location. When people have strong propensity for novelty seeking, more emphasis can be placed on irregular mobility prediction, but when people are more likely to behave regularly, regularity-based models are assigned larger importance.

2 Related Work

Mobility prediction has been widely studied in two independent fields. One field is statistical physics, by assuming human movement can be equivalent to particles and thus leveraging their well-studied motion model for mobility prediction. For example, statistical physicians analyzed mobile phone data, bank notes, GPS trajectories to understand users' individual mobility patterns at an aggregated level by studying the distribution of displacement and waiting time [2, 11, 25]. They then stimulated or predicted human movement based on the derived motion model, such as continuous time random walk and truncated levy flight. This aggregated scaling law can be analytically predicted by the mixed nature of human travel under the principle of maximum entropy, given the constraint on total traveling cost [31]. The other field is mobile communication and data mining in computer science, by directly modeling the mobility patterns based on the data. For example, in [1,9,23,28], the authors presented Markov models and a frequented pattern tree to capture sequential mobility patterns for mobility prediction. In [6, 8], time-aware regularity was modeled for mobility prediction. Furthermore, concomitant social relationships have brought new opportunities for mobility prediction and thus several novel prediction algorithms that incorporate social networks have been proposed [3, 6, 9, 24, 26]. All of this work has observed a small but significant effect of social relationships on mobility prediction. Although social influence is considered as a kind of collective wisdom, it neglects collaborative social knowledge, e.g., from users with similar mobility patterns. In contrast to this existing work, the proposed framework not only tries to fully capture collaborative social knowledge based on recommendation techniques, but also makes better use of the individual power of the regularity-based model and recommendation based on mining propensity of novelty seeking. Therefore, this framework prevents regularity (individual preference) from always playing a dominant role.

Although there are few research that suggest exploiting this knowledge for prediction, the learning of this collaborative social knowledge has been widely studied in *location recommendation*. For example, in [5, 10, 17, 20, 34], social influence, geographical restriction, and personalized user preference have been used for location recommendation. Since these authors all have observed the significant effect of geographical constraint, they have proposed different models, such as k-means clustering and kernel density estimation, for geographical modeling. In addition, the text content of locations, such as reviews and tips, has been used for further improvement [21, 32] of recommendation. In contrast to existing methods, the proposed framework not only takes into account the implicit feedback characteristics of mobility data but also presents a fully unified matrix factorization for jointly modeling user preference, geographical constraint, and social influence. Through this unified model, we have added more pseudo-negative (disliked) locations into the framework, thus alleviating the sparsity challenge.

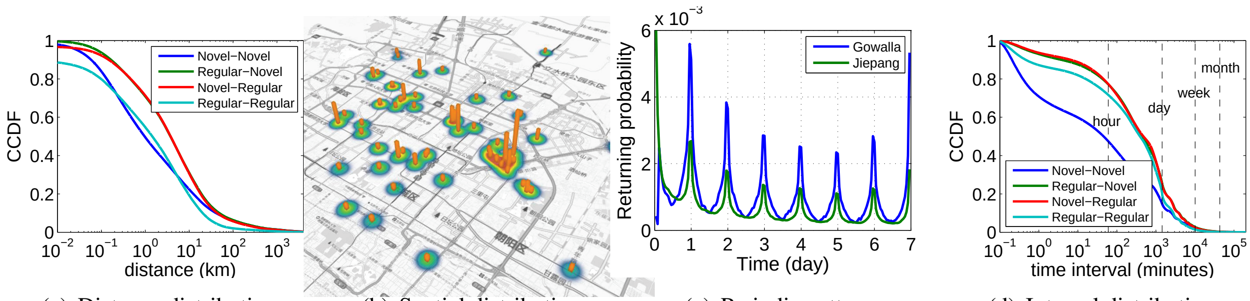
Similar ideas to *mining propensity of novelty seeking* have been proposed in [22, 27], where the probability of novelty seeking is empirically assumed to either be invariant or proportional to the number of distinct visited locations. If novelty seeking is considered to be a deviation from routine, it is related to the work in [29], where deviation from routine is detected by likelihood testing. In contrast, we have summarized our research from three perspectives, two of which are based on supervised learning, which can easily incorporate other features, and the third one is based on unsupervised learning but differentiates several levels of novelty seeking. Additionally, one method of them has a practical explanation, being directly related to the indigenization process of people.

3 Mobility Understanding on Location-based Social Networks

We first understand some basic mobility patterns on location-based social networks from the spatial and temporal perspectives.

3.1 Spatial Analysis

From the spatial perspective, first, we are interested in the distance distribution between consecutive mobility records given regular and irregular (novel) mobility behaviors coexisted, and show the distribution in Fig 2(a). Based on this, we find that 1) most check-ins (over 80%) are within 10 kilometers from the immediately preceding locations; 2) when we already know that users have checked in at regular locations, the next regular location is significantly nearer to them than next novel location; 3) users are more willing to explore continuously. This means that when a user has visited a new attraction, she may also try a nearby restaurant. These three characteristics indicate that spatial analysis can be useful for both regular and irregular location prediction and confirm the need to separate novel locations from regular ones. Second, we are interested in individual spatial density distribution. Thus, we randomly pick one user with more than 100 mobility records and plot her spatial distribution in Fig 2(b). This figure demonstrates that users usually have several major activity areas, such as home and working place, and implies that kernel density estimation is more appropriate for inferring the geographical preference.



(a) Distance distribution (b) Spatial distribution (c) Periodic patterns (d) Interval distribution

Figure 2: Illustration of spatial and temporal analysis

3.2 Temporal Analysis

From the temporal perspective, first, we are interested in periodicity, measured as returning probability [11], which is defined as the probability that a user will revisit a location t hours after her first visit. Its distribution is shown in Fig. 2(c), which indicates that the returning probability is characterized by peaks of each day, capturing a strong tendency to daily revisit regular locations. It thus confirms the existence of temporal regularity, which is thus necessarily introduced into the prediction model. Second, we study the distribution of the time interval between consecutive mobility records, and show the distribution in Fig. 2(d). This shows 1) when a user has visited a regular location, she is less inclined for exploration soon after; 2) users will be more likely to visit novel neighboring locations consecutively within a short interval (e.g., hour). Last, the existence of Markov dependence has been found in the mobility data by comparing the entropy of trajectories with randomly shuffled trajectories under the Markov assumption [30]. We do not elaborate on this here.

4 Mobility Prediction on Location-based Social Networks

Given regular and irregular mobility behaviors coexisting in mobility data, we propose a novelty-seeking driven predictive framework to jointly make use of regularity-based models for predicting regular behaviors and recommendation based algorithms for modeling irregular behaviors. The choice between them is based on people's propensity for novelty seeking, as shown in Fig. 1(b). To be more specific, when people have strong propensity for novelty seeking, recommendation-based algorithms can be relied on more, while when people are more likely to behave regularly, regularity-based models are assigned larger importance.

4.1 Regularity Mining for Regular Mobility Prediction

Regularity-based mining consists of Markov-based predictors for modeling the sequential dependence, temporal regularity for capturing periodical patterns, and a unified Hidden Markov Model for integrating these two models.

4.1.1 Markov-based Predictors

Learning the Markov model mainly depends on the estimation of location transition (due to the small amount of personal data, only first-order Markov models are taken into account). However, maximum likelihood estimation easily suffers from over-fitting due to the insufficiency of training data. Particularly, in most mobility datasets from LBSNs, the number of parameters in the Markov estimator is around 40× 40 since there are 40 POIs for each user on average, while there are only about 60 training instances (mobility records) on average. Although Laplace smoothing techniques can have some effect, they don't differentiate the events of the same observed frequency. Thus, we leverage the widely-used Kneser-Ney smoothing techniques [4], that is

$$P(l|k) = \frac{\max\{n(k,l)-\delta,0\}}{\sum_{l'} n(k,l')} + \frac{\delta \sum_{l'} \mathbf{1}_{\{n(k,l')>0\}}}{\sum_{l'} n(k,l')} \frac{\sum_p \mathbf{1}_{\{n(p,l)>0\}}}{\sum_{l'} \sum_p \mathbf{1}_{\{n(p,l')>0\}}} \tag{17}$$

where $\mathbf{1}_{\{\cdot\}}$ is an indication function and $0 \leq \delta \leq 1$ is a discounting parameter that can be set using the empirical formula $\delta = \frac{n_1}{n_1+2n_2}$ (n_1 and n_2 are the number of one-time transitions and two-time transitions across locations, respectively). Intuitively, this equation discounts the observed times of a transition and turns them over to the possibility that some locations cannot be transitioned from location k. Additionally, such an estimation ensures that zero-order distribution (the marginal of the first-order probability distribution) matches the marginals of the training data. Specifically,

$$\sum_k P(l|k) P_{ML}(k) = P_{ML}(l) \tag{18}$$

Thus $P_{ML}(l)$ is the stationary distribution of Markov process determined by the stochastic transition matrix $P(l|k)$.

4.1.2 Limit of Predictability

We only consider first-order Markov model above, but it is possible to use higher-order or even infinity-order Markov models. The benefit of using higher-order models can be studied by analyzing the limit of predictability [18]. Such analysis can be achieved by estimating the amount of information in terms of Kolmogorov entropy in mobility trajectories. Since it is difficult to estimate Kolmogorov entropy directly, Lempel-Ziv estimators in data compression [14] are often used for approximation, as they can converge to the real entropy of a time series when the length of trajectories is sufficiently large. One estimator of a trajectory of n points is defined as follows:

$$S \approx \frac{\ln n}{\frac{1}{n}\sum_{i=1}^{n} \Lambda_i^i} \tag{19}$$

where Λ_i^i is the length of the shortest substring starting at position i without appearing from position 1 to $i-1$. Without a sufficiently long mobility trace, the entropy will be overly estimated since some frequent patterns have not been observed yet. After estimating the entropy, we then resort to Fano's inequality [7] to transform it into the limit of predictability since this inequality connects the error probability of prediction with the sequential entropy. The overly estimated entropy will incur the lower predictability due to the concavity and monotonic decrease of the Fano function. The key problem of Fano's inequality should guarantee that the maximal prediction probability should be much higher than the random probability. The larger the difference between them is, the closer the upper bound is to the actual predictability. In other words, the more regular

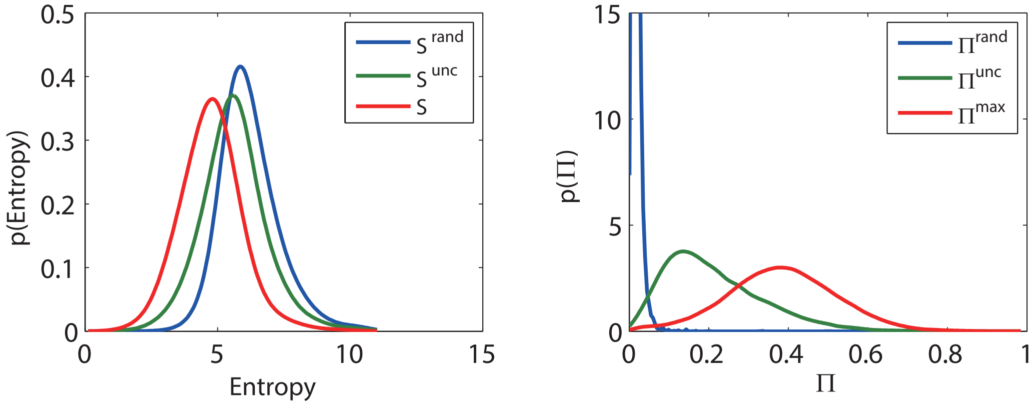


Figure 3: Left: The distribution of Kolmogorov entropy S, sequential uncorrelated entropy S^{unc} without and random entropy S^{rand} across user population. Right: The distribution of predictability of three entropies

the mobility behaviors are, the smaller the error between the upper bound and actual predictability is. Fig.3 shows examples of the distribution of estimated entropy and predictability on the Gowalla dataset [6], which only indicate 38% potential predictability.

4.1.3 Temporal Regularity

In temporal regularity, the conditional probability $P(l|d,h)$ must be estimated accurately, where d is the day of week and h is the hour of day. Assuming the conditional independence d and h given location l, this conditional probability can be transformed as

$$P(l|d,h) = \frac{P(d|l)P(h|l)P(l)}{\sum_l P(d|l)P(h|l)P(l)}. \tag{20}$$

The probability to estimate becomes $P(h|l)$ and $P(d|l)$. However, without sufficient training data, the MLE tends to be overfit. Also, the difference in the probability between neighbor hours of the day and between neighbor days of the week can not be guaranteed to be small. For example, assume a user has been to a Chinese restaurant at 6 p.m. only once. If this user returns to this restaurant in the near future, the distribution of the revisit time should be centered around 6 p.m. rather than at 6 p.m. exactly. Thus we exploit Gaussian kernel smoothing function for smoothing the MLE to the parameters.

$$\tilde{P}(h|l) = \frac{\sum_{g=0}^{23} K(\frac{d(h,g)}{\sigma_{g,l}})P_{ML}(g|l)}{\sum_{h'=0}^{23}\sum_{g=0}^{23} K(\frac{d(h',g)}{\sigma_{g,l}})P_{ML}(g|l)}, \quad \tilde{P}(d|l) \quad = \frac{\sum_{e=0}^{6} K(\frac{d(d,e)}{\sigma_{e,l}})P_{ML}(e|l)}{\sum_{d'=0}^{6}\sum_{e=0}^{6} K(\frac{d(d',e)}{\sigma_{e,l}})P_{ML}(e|l)} \tag{21}$$

where $d(h,g) = \min(|h-g|, 24-|h-g|)$ is the distance between the h^{th} and g^{th} hour of day and $d(d,e) = \min(|d-e|, 7-|d-e|)$ is the distance between the d^{th} and e^{th} day of week. The reason for defining distance in this way is that there is a cyclic property among them (the probability of 0 a.m. is close to 1 a.m. and 23 p.m. and the probability of Sunday is also close to Saturday and Monday). $K(x)$ is a truncated standard Gaussian distribution over $x \in [0, +\infty)$.

4.1.4 Hidden Markov Model

Temporal regularity and Markov model can be integrated in a unified Hidden Markov Model, where locations are considered as hidden states and the temporal information is considered as the observations of Hidden Markov Model. The supervised learning to estimate the parameters corresponds to the above estimation process, except the initial probability of the hidden state is not estimated. Actually, we can simply use MLE for the initial state probability. Note that we don't take social relationship into account since social network ties are more likely to influence long-distance travel according to [6] while long-distance travel may more probably involve irregular mobility behaviors we will introduce next.

4.2 Location Recommendation for Irregular Mobility Prediction

Obviously, regularity-based models will fail to predict irregular mobility behaviors, but such behaviors are still subject to geographical restriction, and are driven by both user preference and social influence. Below, we introduce how to leverage these factors for irregular behavior prediction.

4.2.1 User Preference Learning

Learning user preference mainly involves collaborative filtering techniques, which take the user-location matrix as input and mine the commonality between users. Each element of the matrix can either be visit frequency or a binary value indicating whether the visit has occurred or not. Below, we introduce two approaches for collaborative filtering that mines user commonality from different perspectives.

User-based collaborative filtering [16], directly measures user's commonality in terms of similarity on behavior data. According to our analysis, considering the element of matrix as a binary value to define the similarity is empirically optimal for recommendation. In this case, a user u is represented as $\mathbf{r}_u \in \{0,1\}^N$, where there are N locations in total and her similarity with another user v is defined as follows,

$$s_{u,v} = \frac{\mathbf{r}_u^T \mathbf{r}_v}{\|\mathbf{r}_u\| \|\mathbf{r}_v\|}. \tag{22}$$

The scoring function of user u to location i is in proportion to $\mathbf{s}_u^T \mathbf{r}_i$.

Matrix factorization is a dimension reduction technique such that the dot product between users, between items, and between user and item in the reduced latent space can measure the commonality. However, since mobility data only include the locations where users have been and are likely to prefer, while unattractive locations and undiscovered but potentially appealing ones are mixed in unvisited locations, mobility data are actually a kind of implicit feedback. In this case, we need to use a special class of matrix factorization algorithms, which treat all unvisited locations as pseudo-negative and assign them a significantly lower confidence. User preference is thus learned by solving the following optimization problem,

$$\min_{P,Q} \sum_{u,i} w_{u,i}(r_{u,i} - p_u^T q_i)^2 + \lambda(\|P\|_F^2 + \|Q\|_F^2)$$

where $p_u \in \mathbb{R}^K$ and $q_i \in \mathbb{R}^K$ represent the preferences of user u and POI i. The weight $w_{u,i}$ is empirically set as $\alpha(c_{u,i}) + 1$ if $c_{u,i} > 0$; and 1 otherwise, where $\alpha(c_{u,i})$ is monotonic increasing w.r.t visit frequency $c_{u,i}$, indicating the visit frequency reflect confidence that the users are fond of them.

4.2.2 Geographical Constraint

Kernel density estimation. The geographical information of location requires physical interactions with users to foster the universality of Tobler's First Law of Geography: "Everything is related to everything else, but near things are more related than distant things." The key for capturing this phenomenon is geographical modeling. We use two-dimensional kernel density estimation, which infers the probability a user will show up around location l_j, i.e.,

$$P(l_j) = \frac{1}{|L_i| h} \sum_{l_k \in L_i} K(\frac{d_{j,k}}{h}),$$

where $K(\cdot)$ is a kernel function. The setting of bandwidth h in the kernel function is determined by the requirement that the influence of candidate locations on the border of the influence circle is close to zero. If the probability on the border is at most ϵ times smaller than the maximum possible check-in probability, it is subject to $K(\frac{d}{h}) < \epsilon K(0)$.

Learning-based geographical inference is proposed for the sake of seamlessly integrating geographical modeling with matrix factorization based user preference learning. This is achieved by splitting the whole world into grids of approximately the same size and pre-computing the received influence of each grid from all the locations, and then converting kernel density estimation to the following optimization problem,

$$\min_{x_u} \sum_i w_{u,i}(x_u^T y_i - r_{u,i})^2 + \lambda\Omega(x_u), \text{subject to } x_u \geq 0$$

In this objective function, y_i is an influence vector of a location i, and each element corresponds to a grid's influence received from this location; and x_u is an activity area vector of user u, in which every element represents the possibility that this user will appear in a certain grid. Thus, the dot product between them can be considered to be the possibility that user u will show up around location i. $\Omega(x_u)$ is a regularized term for avoiding over-fitting.

4.2.3 Social Influence

Social-based filtering [16] is similar to user-based collaborative filtering, except it captures user commonality based on social network information. The simplest commonality/similarity between two users is defined as 1 if they are friends and 0 otherwise. In this case, a user's preference score for a location can be expressed as the number of her friends who have visited. To more accurately distinguish the importance of friends based on their closeness, we exploit another strategy, which is in proportion to the number of common friends, i.e.,

$$s_{i,l} = \frac{|F_i \cap F_l|}{|F_i \cup F_l|},$$

where F_i and F_l represent the friend sets of user u_i and u_l, respectively.

Graph Laplacian regularization [15] is more often exploited for capturing social influence for the sake of seamless integration with matrix factorization based preference learning, although social-based filtering tends to be more intuitive. Given all users' symmetric similarities S based on social network ties, such as the ratio of common friends [19], this regularizer can be defined as follows:

$$\Omega_S(P) = \frac{1}{2} \sum_{i,l} s_{i,l} \|p_i - p_l\|^2 = tr(P^T L P)$$

where $D_{i,i} = \sum_l s_{i,l}$ and $L = D - S$ is a Laplacian matrix.

4.2.4 Hybrid Recommendation

Given the factors affecting the prediction of irregular behaviors, there are many methods for empirical integration. Since geographical modeling is converted into an optimization problem, it can be seamlessly incorporated into user preference learning in terms of matrix factorization, as shown in Fig. 4. In this model, the influence areas of a POI are considered as an extra part of the POI's latent factors and the activity areas of a user are considered as an extra part of the user's latent factors. Since they are aligned in position, the dot product between them indicates two-dimensional kernel density estimation. At this moment, because unvisited locations around visited ones share similar geographical influence, user preference for them needs to offset the geographical influence. Thus, such an integration allows us to find more potential disliked locations and plays an important role in

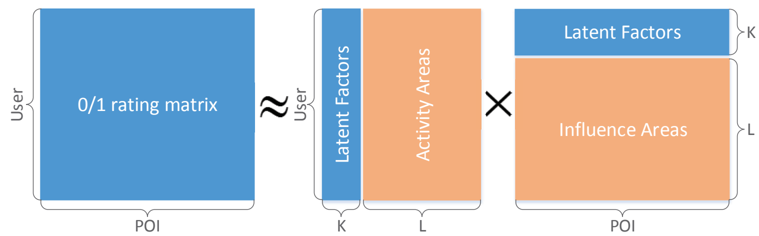


Figure 4: The augmented model for matrix factorization, where the dimension of the latent space is K and the number of grids is L.

alleviating the data sparsity. Furthermore, combining this with graph Laplacian regularization for incorporating social relationships, the overall objective function becomes as follow:

$$\min_{P,Q,X} \|W \odot (R - PQ^T - XY^T)\|_F^2 + \gamma(\|P\|_F^2 + \|Q\|_F^2) + \eta\Omega_S(P) + \lambda\|X\|_1, \text{subject to } X \geq 0$$

where X is a matrix stacking a user's activity area by columns and Y is a matrix stacking the items' influence areas vector by columns. ℓ_1 norm of matrix X, $\|X\|_1$, constrains that users usually stay around several important locations, such as home and working places.

4.3 Mining Propensity of Novelty Seeking

Mining individual propensity of novelty seeking is conducted from three perspectives: exploration prediction, mobility indigenization, and irregularity detection. Exploration prediction is spatially and temporally dependent while mobility indigenization is only with respect to cities. However, irregularity detection is independent to both spatial and temporal contexts.

4.3.1 Exploration Prediction

Exploration prediction determines whether people will seek novel (irregular) locations next. Given mobility data, whether a visit to a location is regular or not can be determined by searching the mobility history of the user. If the visit location has already been visited earlier, the visit is considered as regular; otherwise, it is irregular. Exploration prediction is thus boiled down to a binary classification problem, which can output a classification result (regular or not) or exploration tendency (e.g., a probability of classifying the next location as irregular). In the classifiers, we consider the following three types of features.

Historical features not only summarize the personality traits of novelty seeking, i.e, how often they check in, but also reflect a user's current status of neophilia, including whether a user is currently doing exploration and how many opportunities a user has left to seek novel locations. The more locations near her activity area are visited, the smaller the number of opportunities are left, and the smaller the propensity of seeking novel locations is becoming.

Temporal features are introduced to consider the effect of this temporal information since users usually have distinct degrees of novelty seeking at different times. As we have discovered, 1) users may prefer to do exploration during weekends; 2) the time interval between consecutive records also affects novelty seeking.

Spatial features are also taken into account for exploration prediction because users also exhibit different propensity of novelty seeking at locations with distinct degrees of familiarity. For example, if a user has arrived in an unfamiliar location (e.g., city), her propensity for novelty seeking will increase.

4.3.2 Mobility Indigenization

When considering irregular mobility behaviors as mainly occurring out of town, we can use a more interesting index, i.e., indigenization coefficients, for integration [33]. This index quantifies what extent an individual behaves like a native. Therefore, this index is opposite to the propensity of novelty seeking. The smaller the index of a user in a city is, the more likely she is non-native to the city, so that irregular-based models should be given higher emphasis.

We have proposed two coefficients for this indigenization index. The first one is an individual behavioral index, $I_i(u)$, which counts the ratio of repeated mobility records in a city, inspired by the fact that a native is more likely to visit some locations many times than a non-native. That is, for a user, N_T indicates the total number of her mobility records and N_D the number of different locations visited by her. The index is then defined as

$$I_i = 1 - \frac{N_D}{N_T}. \tag{23}$$

The second one is a collaborative behavioral index I_c, measured as the average normalized popularity of a user's visit locations, which is inspired by the fact that a native is less likely to visit popular locations than a non-native. Given that $R(l_k)$ is the normalized rank of location l_k (dividing the rank by the total number of locations in a city), this index is formally defined as

$$I_c = \frac{1}{N_T} \sum_{k=1}^{N_T} R(l_k). \tag{24}$$

These two indigenization coefficients can be used to define an integrated coefficient

$$I = \frac{1}{1 + \exp(-w_i I_i - w_c I_c)}, \tag{25}$$

where the parameters w_i and w_c can be learned from the logistic regression that best classifies natives and non-natives. In other words, these two coefficients are taken as features for classifying people as native and non-natives. After learning these two parameters, we obtain a probabilistic value for the indigenization level and thus obtain a probability (i.e.,$1 - I$) for novelty seeking.

4.3.3 Irregularity Detection

Irregularity detection [35] further distinguishes several levels of propensity of novelty seeking, and detects the level of novelty seeking by measuring the popularity of the visit locations and the transition frequency to visiting location with respect to individual mobility history before the visit time. When both the popularity and transition frequency are smaller at the same time, the level of novelty seeking tends to be higher. After determining the level of novelty seeking for each visit in the mobility data, we can measure the novelty seeking trait for each user. For example, we can use the average level of novelty seeking. In other words, such an algorithm will give each user the same but distinct propensity of novelty seeking at any time and any location. In order to leverage it in the general mobility prediction framework, we can normalize it by dividing the maximum level of novelty seeking to get a pseudo probability value. A larger value indicates a higher possibility of novelty seeking.

4.4 A Novelty-Seeking Driven Framework for General Mobility Prediction

Provided the probabilistic output of the regularity mining algorithm $P_r(l)$ (r indicates regular) and recommendation algorithm $P_n(l)$ (n indicates novel), we exploit novelty seeking to combine them based on the probability of exploration $Pr(Explore)$ as follows:

$$P(l) = Pr(Explore)P_n(l) + (1 - Pr(Explore))P_r(l), \tag{26}$$

If $Pr(Explore) \in \{0, 1\}$, i.e., novelty seeking just classifies the next location as novel or not, we can switch between location recommendation and the regularity-based model. Due to the discrete value of $Pr(Explore)$, we denote this case as "hard" integration. If $Pr(Explore) \in [0, 1]$, i.e., representing the propensity of novelty seeking, we can interpolate the regularity-based model with location recommendation. In other words, both novel and regular locations are ranked together in this case for the final location prediction. Due to the continuous value of $Pr(Explore)$, we denote this case as "soft" integration.

5 Conclusions

In this paper, we have introduced a novelty-seeking driven framework for incorporating regularity-based prediction algorithms and recommendation algorithms for predicting irregular mobility behaviors. In regularity-based prediction, we exploit Hidden Markov model for modeling location transition and temporal dependence. For recommendation algorithms, we propose a unified recommendation framework to integrate social influence, geographical restriction, and user preference based on the implicit feedback characteristics of mobility data. And the central idea of this predictive framework is the mechanism of cooperation between these two distinct models, by exploiting exploration prediction, indigenization coefficient and irregularity detection to characterize the propensity of seeking a novel and appealing location.

References

[1] D. Ashbrook and T. Starner. Learning significant locations and predicting user movement with gps. In *Proceedings of the 6th IEEE International Symposium on Wearable Computers(ISWC'02)*, pages 101–108. IEEE, 2002.

[2] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.

[3] J. Chang and E. Sun. Location3: How users share and respond to location-based data on social. In *Proceedings of ICWSM'11*, 2011.

[4] S. F. Chen and J. Goodman. An empirical study of smoothing techniques for language modeling. In *Proceedings of ACL'96*, pages 310–318. ACL, 1996.

[5] C. Cheng, H. Yang, I. King, and M. Lyu. Fused matrix factorization with geographical and social influence in location-based social networks. In *Proceedings of AAAI'12*, 2012.

[6] E. Cho, S. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of KDD'11*, pages 1082–1090, 2011.

[7] R. Fano. *Transmission of information: a statistical theory of communications*. M.I.T. Press, 1961.

[8] H. Gao, J. Tang, X. Hu, and H. Liu. Modeling temporal effects of human mobile behavior on location-based social networks. In *Proceedings of CIKM'13*, 2013.

[9] H. Gao, J. Tang, and H. Liu. Exploring social-historical ties on location-based social networks. In *Proceedings of ICWSM'12*, 2012.

[10] H. Gao, J. Tang, and H. Liu. gscorr: modeling geo-social correlations for new check-ins on location-based social networks. In *Proceedings of CIKM'12*, pages 1582–1586. ACM, 2012.

[11] M. Gonzalez, C. Hidalgo, and A. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[12] M. Horner and M. O'Kelly. Embedding economies of scale concepts for hub network design. *Journal of Transport Geography*, 9(4):255–265, 2001.

[13] R. Kitamura, C. Chen, R. Pendyala, and R. Narayanan. Micro-simulation of daily activity-travel patterns for travel demand forecasting. *Transportation*, 27(1):25–51, 2000.

[14] I. Kontoyiannis, P. Algoet, Y. Suhov, and A. Wyner. Nonparametric entropy estimation for stationary processes and random fields, with applications to english text. *IEEE Transactions on Information Theory*, 44(3):1319–1327, 1998.

[15] D. Lian and X. Xie. Mining check-in history for personalized location naming. *ACM Trans. Intell. Syst. Technol.*, 5(2):32:1–32:25, Apr. 2014.

[16] D. Lian, X. Xie, V. W. Zheng, N. J. Yuan, F. Zhang, and E. Chen. Cepr: A collaborative exploration and periodically returning model for location prediction. *ACM Trans. Intell. Syst. Technol.*, 6(1):8:1–8:27, Apr. 2015.

[17] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui. Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Proceedings of KDD'14*, pages 831–840. ACM, 2014.

[18] D. Lian, Y. Zhu, X. Xie, and E. Chen. Analyzing location predictability on location-based social networks. In *Proceedings of PAKDD'14*, 2014.

[19] D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. *Journal of the American society for information science and technology*, 58(7):1019–1031, 2007.

[20] B. Liu, Y. Fu, Z. Yao, and H. Xiong. Learning geographical preferences for point-of-interest recommendation. In *Proceedings of KDD'13*, pages 1043–1051. ACM, 2013.

[21] B. Liu and H. Xiong. Point-of-interest recommendation in location based social networks with topic and location awareness. In *Proceedings of SDM'13*, pages 396–404. SIAM, 2013.

[22] J. McInerney, S. Stein, A. Rogers, and N. R. Jennings. Breaking the habit: Measuring and predicting departures from routine in individual human mobility. *Pervasive and Mobile Computing*, 2013.

[23] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti. Wherenext: a location predictor on trajectory pattern mining. In *Proceedings of KDD'09*, pages 637–646. ACM, 2009.

[24] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo. Mining user mobility features for next place prediction in location-based services. In *Proceedings of ICDM'12*, pages 1038–1043. IEEE, 2012.

[25] I. Rhee, M. Shin, S. Hong, K. Lee, S. Kim, and S. Chong. On the levy-walk nature of human mobility. *IEEE/ACM Trans. Netw. (TON)*, 19(3):630–643, 2011.

[26] A. Sadilek, H. Kautz, and J. Bigham. Finding your friends and following them to where you are. In *Proceedings of WSDM'12*, pages 723–732. ACM, 2012.

[27] C. Song, T. Koren, P. Wang, and A. Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010.

[28] L. Song, D. Kotz, R. Jain, and X. He. Evaluating location predictors with extensive wi-fi mobility data. In *Proceedings of INFOCOM'04*, volume 2, pages 1414–1424. IEEE, 2004.

[29] M. Szell, R. Sinatra, G. Petri, S. Thurner, and V. Latora. Understanding mobility in a social petri dish. *Scientific reports*, 2, 2012.

[30] C. Wang and B. A. Huberman. How random are online social interactions? *Scientific reports*, 2, 2012.

[31] X.-Y. Yan, X.-P. Han, B.-H. Wang, and T. Zhou. Diversity of individual mobility patterns and emergence of aggregated scaling laws. *Scientific reports*, 3, 2013.

[32] D. Yang, D. Zhang, Z. Yu, and Z. Wang. A sentiment-enhanced personalized location recommendation system. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media(HT'13)*, pages 119–128. ACM, 2013.

[33] Z. Yang, N. J. Yuan, X. Xie, D. Lian, Y. Rui, and T. Zhou. Indigenization of urban mobility. *arXiv preprint arXiv:1405.7769*, 2014.

[34] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of SIGIR'11*, pages 325–334. ACM, 2011.

[35] F. Zhang, N. J. Yuan, D. Lian, and X. Xie. Mining novelty-seeking trait across heterogeneous domains. In *Proceedings of the 23rd international conference on World wide web*, pages 373–384. International World Wide Web Conferences Steering Committee, 2014.

[36] V. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang. Collaborative filtering meets mobile recommendation: A user-centered approach. In *Proceedings of AAAI'10*. AAAl Press, 2010.