Where Did You Go: Personalized Annotation of Mobility Records

Fei Wu Zhenhui Li
College of Information Sciences and Technology
Pennsylvania State University, University Park, PA, USA
{fxw133, jessieli}@ist.psu.edu

ABSTRACT

Recent advances in positioning technology have generated massive volume of human mobility data. At the same time, large amount of spatial context data are available and provide us with rich context information. Combining the mobility data with surrounding spatial context enables us to understand the semantics of the mobility records, e.g., *what is a user doing at a location* (e.g., dining at a restaurant or attending a football game). In this paper, we aim to answer this question by annotating the mobility records with surrounding venues that were actually visited by the user. The problem is non-trivial due to high ambiguity of surrounding contexts. Unlike existing methods that annotate each location record independently, we propose to use all historical mobility records to capture user preferences, which results in more accurate annotations. Our method does not assume the availability to any training data on user preference because of the difficulties to obtain such data in the real-world setting. Instead, we design a Markov random field model to find the best annotations that maximize the consistency of annotated venues. Through extensive experiments on real datasets, we demonstrate that our method significantly outperforms the baseline methods.

1. INTRODUCTION

Recently, driven by the advances in positioning technology and the prevalence of location-based services, there has been tremendous interest in mining human mobility data in the data mining community. Researchers have previously studied various mobility patterns such as frequent patterns [17], periodic behaviors [13], representative behaviors [8, 6], and activity recognition [15, 11, 22, 26, 39]. However, all these patterns and activities are extracted *purely from the mobility data*. While they are useful for understanding the inherent movement behaviors of a moving object, they do not provide any *contextual semantics* required for understanding the intended activities of the user.

CIKM'16 , October 24-28, 2016, Indianapolis, IN, USA
 ISBN 978-1-4503-4073-1/16/10...$15.00
DOI: http://dx.doi.org/10.1145/2983323.2983845

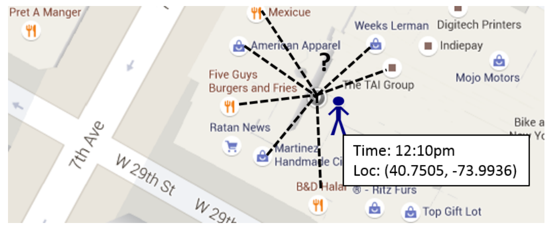


Figure 1: Ambiguity in annotating the location records with contexts.

In this paper, we study the problem of semantic annotation of mobility records, i.e., associating the mobility records in a person's trajectory with *relevant* surrounding contexts. In particular, we use rich venue datasets as the contexts and ask the question: *If a person is observed at location loc at time t, which venue is the true destination of this person?*

Knowing the actual venues a user visits allows us to profile individual users' interests, socioeconomic status, and health conditions. For example, if a person often visits a kindergarten and kids-friendly restaurants, he/she may have a young kid. Therefore we could recommend kids-related activities to this user over the weekend. If a person frequently visits fast food and convenience stores, but rarely visits recreational places (e.g., parks, fitness facilities), the person is likely to live a poor lifestyle and might be at a greater risk of chronic health conditions. Such semantic inferences rely on successfully associating the mobility records with the surrounding contexts.

However, associating the right contexts with the mobility records is a difficult problem. The key challenge lies in the ambiguity of surrounding contexts. Because of the data collection mechanism and GPS errors, we may not observe a person locating at the actual location where he was visiting. Take Figure 1 as an example. Given a GPS location of the user, how can we know whether he is visiting the clothing store American Apparel or eating at Five Guys restaurant? And even if we have accurate GPS locations, a geographical location could be associated with many venues at the same time. For example, the largest transition center of New York City, Penn Station, sits right beneath Madison Square Garden, a multi-purpose indoor arena. As another example, a multi-function building at Times Square could have restaurants, stores and offices, all of which sharing the same location. How can we know which venue is the true destination in these cases?

To tackle this problem, existing methods [1, 22, 28, 24] mainly consider the distance between the context location and the location of the user, i.e., the closer the context is, the more relevant it is to the mobility record. However, all the previous studies annotate each location record *independently*, without considering the history of this user's movement. As a result, different users at the same location and time will all be annotated with the same context regardless of their personal interests.

In this paper, we propose to explore the history of a user's mobility records to infer interest of the user for personalized annotation. Our intuition is that, user interests can be reflected by the features of venues associated with the user's mobility records. For example, a person preferring Asian food may visit restaurants of the same or similar categories. It is worth noting that our way of exploring the user preference is fundamentally different from that in location-based recommendation systems [30, 31, 34]. A recommendation system relies on the training data, where movements are already annotated with the venues visited (e.g., via check-in function). Such training data are often hard to obtain in the real world. Our problem does not require any training data as input.

Without any training data, our method is to take all the nearby venues as candidates and choose the ones that can maximize the inherent consistency among them (e.g., their categories). We propose a Markov random field (MRF) model to capture traditional factors such as the distance between venues and mobility records while enforcing the consistency on the selected venues. However, a naive implementation of MRF would yield poor performance because it tends to associate all records of a user with the same type of venues, whereas real users typically have multi-modal preferences (e.g., visiting places from different categories). To address this problem, we utilize two importance characteristics of human mobility: (i) spatial regularity, i.e., people tend to go to the same venue at similar locations; and (ii) temporal regularity, i.e., people tend to go to the same type of venues at similar times of the day. Experiment results demonstrate that, by incorporating these regularities, our improved MRF model significantly outperforms existing methods on annotating real mobility records.

In summary, the contributions of this paper are:

- We study an important and challenging problem of semantic annotation on mobility records. The annotations help bridge the gap between raw mobility records and surrounding contexts and greatly enrich our understanding of the mobility data.
- Different from existing studies which consider each mobility record independently, we propose a novel Markov random field model to capture personal preference using all the location records of each user.
- We conduct extensive experiments on large-scale datasets of three cities (i.e., more than $200,000$ records of $2,286$ users). Our proposed method significantly outperforms all existing methods on all datasets.

The rest of the paper is organized as follows. Related studies are first discussed in Section 2. Section 3 presents preliminaries. We describe our method in Section 4 and show experimental results in Section 5. Finally, the paper is concluded in Section 6.

2. RELATED WORK

Long lines of literature tackle the problem of mining patterns from mobility data, such as recurrence patterns, interaction patterns, and activities [42]. However, those methods do *not* explore external sources to discover the contextual semantics. Recently, many studies have begun to utilize external data to understand mobility data [38]. Xiao et al. [25] use a feature vector of POIs to represent each user for finding similar users. Zhong et al. [43] aggregate user check-ins to POIs for classifying demographic attributes. In these studies, user check-ins contain the information about which venues each user visited. In our problem, we use exclusively mobility records from the users and aim to infer the venues the user visited. Su et al. [23] use features of the roads to partition vehicle trajectories for summarizing vehicle status. In contrast to our ambiguity issue, the features of the road are unambiguously associated with trajectories.

There are several studies that address similar annotation problems as ours. Those studies annotate mobility records with landmarks [1], landscapes [22], or land-use categories [28]. All these methods measure the relevance of context based on the distance to the mobility records and annotate mobility records independently without considering the history of a user's movements. As a result, different users at the same place and time will always be annotated with the same venue. In [24], Wu et al. propose to annotate dynamic events to mobility records, where sources of annotations are collections of word occurrences in social media (i.e., Twitter) over the space. However, since our sources of annotations are venues, the Kernel Density Estimation (KDE) method proposed in [24] to model the spatial distribution of words cannot be applied to our problem.

For personalized records annotation, Yan et al. [28, 27] propose a hidden Markov model to consider the transitions of human movements. For example, people tend to go to restaurant after work and then go home. This model works well when the trajectories are densely sampled. However, human mobility data are often temporally sparse (e.g., two samples collected from social media may be separated by a few days), therefore a transition relationship may not exist at all. In contrast, our model does not make such assumption and can deal with sparse mobility records. Moreover, a transition matrix is needed as input in [28, 27]. Instead, we incorporate general characteristics of human mobility (i.e., temporal and spatial regularities) and infer the preference without any prior knowledge about the user. We will further compare the method proposed in [28, 27] in Section 5.

The problem of annotating a point with a venue is similar to the problem of recommending a venue to a user at one location. Following the line of POI recommendation on LBSN, existing methodologies can be divided into two categories: (i) content-based, and (ii) collaborative filtering [2]. In content-based approaches, profiles of users, such as age, gender, and preferred cuisines, are used to model preference [19, 29, 2]. For our annotation problem, we do not assume that the users' profiles are available. In the collaborative filtering framework, previous check-ins of a user are used to learn the preference of the user and the similarity between users. Studies incorporate various factors, e.g., geographical distance [34, 31, 16, 14], popularity [33, 32], correlation of check-ins [37], and temporal dependency [5, 35, 36]. In our annotation problem, we use only raw mobility records without any check-in information.

3. PRELIMINARIES

We consider the mobility records of a mobile user as a series of spatiotemporal points, $R = \{r_1, r_2, ..., r_n\}$, where each record consists of the raw geographic coordinates and the time, i.e., $r_i = (l, t)$. In practice, location data can be collected from a variety of services such as mobile phones, cell towers, and location-based social network. For continuous trajectory data, stopping point detection methods can be first applied to extract meaningful records [40].

Generally, context data can be defined as a set of objects associated with locations. In this paper, we use venue database as our context, $C = \{v_1, v_2, ..., v_m\}$, where for each venue v_j we have information about its geographic coordinates l and a set of features $\mathbf{f}$ describing the venue, i.e., $v_j = (l, \mathbf{f})$. Various forms of information can be included in the feature set, such as venue types, reviews, ratings, and check-in time series.

Given data from two separate channels, we now formally define our annotation problem as follows:

PROBLEM 1 (SEMANTIC ANNOTATION PROBLEM). *With mobility records of a user, $R = \{r_1, r_2, ..., r_n\}$, and the context dataset, $C = \{v_1, v_2, ..., v_m\}$, our goal is to find the most relevant venue $v^* \in C$ for each mobility record $r_i \in R$.*

To measure the cost of matching each mobility record r_i with a potential venue $v \in C$, one baseline is to rank the list of venues by their distances to the record r_i:

$$v^* = \arg\min_v dist(r_i.l, v.l), \tag{1}$$

where $dist(\cdot)$ is a distance measure of two geographic locations.

Moreover, the features associated with the venues, can also be considered when deciding the most relevant context. In such cases, the problem can be generally written as follows:

$$v^* = \arg\min_v E_i(r_i, v), \tag{2}$$

where E measures the cost of matching the context v to a location record r_i.

However, a fundamental problem with the above formulation is that each record r_i is annotated *independently*, without considering the mobility history of the user. The different users appearing at the same location at same time will always be assigned to the same venue, despite the fact that they are often visiting different venues based on their own preferences.

Therefore, it is necessary to consider user preference in order to find the most relevant venue for each location record. Unfortunately, such information is typically unavailable. In the following section, we discuss how one can learn the preference of a user and annotate all the mobility records simultaneously.

4. SEMANTIC ANNOTATION VIA PREFERENCE MODELING

Generally speaking, in our problem, a user's preference can be regarded as his/her consistent interests in some venues over others. Intuitively, such interests can be reflected by the features of venues associated with the user's mobility records (if known). For example, a person preferring Asian food may visit restaurants of the same or similar categories (e.g., Chinese and Japanese), and a sports fan is more likely to visit sports-related venues than shopping malls.

In this paper, we translate our intuition into a formal Markov random field framework and demonstrate its effectiveness in accurate annotation of location records.

4.1 A Markov Random Field Framework

Markov random fields are general models that have been successfully applied to obtain an optimal labeling under different application contexts [18]. A pairwise MRF consists of an undirected graph $G = (V, \mathcal{N})$, where each node is associated with a label $y_i \in C$ that can be in a finite number of states. The edge set $\mathcal{N}$ specifies nodes which are dependent. The Markov property in the graph states that a node only depends on its neighbors and is independent of all other nodes. Specifically, the joint probability of one set of assignments is defined as:

$$\begin{aligned} P(Y) &= P(y_1, y_2, ..., y_n) \\ &= \frac{1}{Z} \prod_i \phi_i(y_i) \prod_{<i,j>\in\mathcal{N}} \psi_{ij}(y_i, y_j), \end{aligned} \tag{3}$$

where Y denotes the label assignments of nodes, and Z is the normalizing constant. The node potential $\phi_i : C \to \mathbb{R}^+$ represents the initial assignment probability. The pairwise potential $\psi_{ij} : C \times C \to \mathbb{R}^+$ captures the likelihood of a node with label y_i to be connected to a node with label y_j through an edge.

For our semantic annotation problem, we let each node represent a mobility record r_i, and C be a set of venues: $C = \{v_1, v_2, \dots, v_m\}$. We further let the prior potential $\phi_i(y_i)$ represent the cost of assigning venue v_j to location record r_i, and the pairwise potential $\psi_{ij}(y_i, y_j)$ measures the similarity between the features of two venues. Intuitively, the pairwise potential $\psi_{ij}(y_i, y_j)$ is designed to improve the consistency of the features associated with the assigned venues.

Taking the negative log of Eq. 3, we can see that finding its MAP estimate is equivalent to minimizing the following objective function:

$$\begin{aligned} E(Y) &= \sum_i -\log\phi_i(y_i) + \sum_{<i,j>\in\mathcal{N}} -\log\psi_{ij}(y_i, y_j), \\ &= \sum_i E_i(y_i) + \sum_{<i,j>\in\mathcal{N}} E_{ij}(y_i, y_j) \end{aligned} \tag{4}$$

where the constant Z is omitted as it remains the same across all assignments. Note that, unlike Eq. (2) which finds the optimal assignment for each record independently, our goal now is to find the best set of assignments $\{y_1^*, y_2^*, ..., y_n^*\}$ for all records that minimizes Eq. (4). Overall, our formulation aims to achieve a balance between the consistency over venues visited and the relevance of the venues locally (which still allows for deviations of a user's behavior).

Our question now is how to design the pairwise MRF – that is, to determine the edge set $\mathcal{N}$, the node energy $E_i(\cdot)$, and the pairwise energy $E_{ij}(\cdot)$ – so that it properly reflects user preference. A simple strategy to determine $\mathcal{N}$ is to connecting all node pairs: $\mathcal{N} = \{< i, j >: \forall r_i, r_j, r_i \neq r_j\}$. It is easy to see that such a strategy favors a single-modal preference, i.e., a user visiting one type of venues consistently.

4.1.1 *Diversity of User Preference*

In practice, a user may express different preferences under different circumstances. We now make some observations on

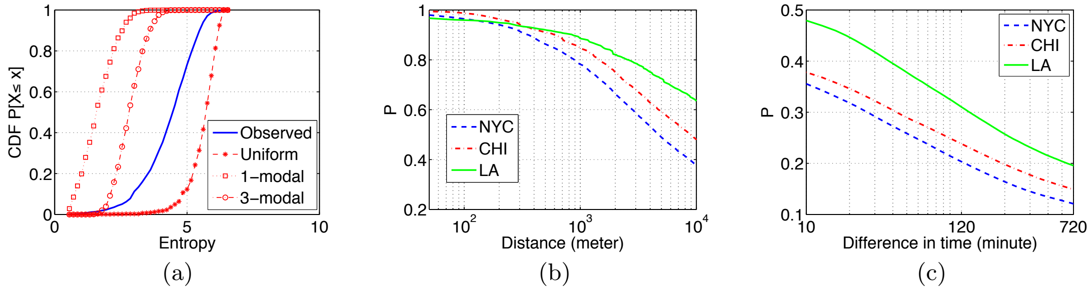


Figure 2: (a) The cumulative distribution function (CDF) of the entropy of user preference. (b) Probability of a check-in pair being in the same category of venues with respect to the distance threshold between two venues. (c) Probability of a check-in pair being in the same category of venues with respect to the threshold of (relative) time difference of the check-ins.

three check-in datasets from Foursquare (refer to Section 5 for details regarding the data) using the sub-categories (678 sub-categories, e.g., Sushi, Chinese food, and Sports bar) of check-ins to study the preference of each user.

For each user, the number of check-ins to each category forms a discrete distribution over categories. We use *Shannon entropy* to measure the diversity of the distributions. Figure 2(a) shows the cumulative distribution function (CDF) of the entropy of all users. We also generate distributions with 1-modal preference, 3-modal preference (prefers 3 categories of venues), and uniform preference (no preference)[1]. We can see from Figure 2(a) that the empirical CDF significantly deviates from the CDF of 1-model distributions. This suggests that users tend to have a multi-modal preference and therefore *it is inappropriate to connect all node pairs in our MRF formulation.*

4.1.2 *Regularity of Human Mobility*

As overall consistency across all location records of a user is non-existent, only connecting records in which a user would express the same preference becomes very important. Meanwhile, various studies have suggested that human movements exhibit (i) strong spatial regularity, i.e., a user visits a few places frequently [10, 21], and (ii) strong temporal regularity, i.e., a user appears at nearby locations at the same time of the day [9, 13, 7]. Based on previous studies, it is intuitive to assume that mobility records are spatially close or temporally close should be connected. We make observations on the Foursquare check-in datasets to verify the intuitions.

OBSERVATION 1 (SPATIAL REGULARITY). *A user tends to visit similar venues (e.g., defined by venue category) or the same venue when observed at close locations.*

We compute the probability of a check-in pair being in the same category of venues with respect to the distance between the two venues. Here, the probability is defined as the the number of “same category” pairs divided by the total number of pairs within certain distance. Figure 2(b) shows that 97%, 99%, and 96% of the venue pairs (checked-in by each user) within 100 meters are in the same category for NYC, CHI, and LA datasets, respectively. In addition, there are clear increasing trends in the probability as the distance between two checked-in venues decreases. Upon further investigation, we found that 90% of the check-in pairs within 100 meters that are in the same category actually correspond to the same venues. These results indicate that the regularity exist at both venue-level and category-level. We therefore incorporate both venue-level and category-level regularity into our model specification (Section 4.2).

OBSERVATION 2 (TEMPORAL REGULARITY). *A user tends to visit similar venues or the same venue at similar time of the day.*

We follow the same methodology above to study whether check-in pairs of each user are more likely to be in the same category as they are temporally closer (in a daily relative time frame). Figure 2(c) shows that the probability (of check-in pairs being in the same category) has an increase of 0.27, 0.22, and 0.28 as the time difference decreases from 12 hours to 10 minutes on three datasets respectively. The increasing trends indicate that check-ins at closer times of the a day are more likely to be in the same category of venues. In this paper, we limit our discussion to daily regularity. One may also further consider temporal regularity in a weekly, monthly, or yearly scale as extensions.

Our observations verify that human mobility exhibits strong spatial regularities and temporal regularities in our problem. In the next section, we define our MRF model according to these characteristics.

4.2 Specification of the Proposed Model

Edge list $\mathcal{N}$: The edge list $\mathcal{N}$ specifies the neighbors of mobility records among which the consistency should be enforced. Based on the observed spatial and temporal regularities, we define $\mathcal{N}$ as the set of nodes that are spatially or temporally close:

$$\mathcal{N} = \mathcal{N}_T \cup \mathcal{N}_D,$$
$$\mathcal{N}_D = \{< i, j >: D(r_i, r_j) < \xi_D\},$$
$$\mathcal{N}_T = \{< i, j >: T(r_i, r_j) < \xi_T\},$$

where $D(r_i, r_j)$ is a distance proximity measure, $T(r_i, r_j)$ is a proximity measure on the relative time, and ξ_T and ξ_D are thresholds. We use the Haversine function as our dis-

[1]We generate the reference distributions according to the number of check-ins and categories of each user. All categories that a user visited are assigned with one check-in. The rest of the check-ins are assigned to k random categories (evenly) for a k-modal reference distribution.

tance function for location records represented by GPS coordinates. Other distance functions may also be considered for other formats of location, e.g., l_p-norm when the coordinates are in a multi-dimensional feature space, and a binary indicator function $\mathbb{I} \in \{0, 1\}$ when location is discrete, e.g., cell-tower ID.

The difference in relative timestamp is as follows:

$$T(r_i, r_j) = |r_i.t - r_j.t| \pmod{P},$$

where P is the period of choice. We found that daily temporal regularity is the strongest on a sub-category level, thus use one day as the period P.

Pairwise potential ψ_{ij}: Various similarity measures on venues [29, 37, 41] can be directly used as the pairwise potentiality. For our purpose, we measure the similarity of venues by sub-category for simplicity and generality (as the category information is easily accessible).

A venue should be most similar with itself, followed by the venues of the same category. Venues of different categories should not be considered as similar. Therefore, we define the pairwise potential ψ_{ij} as:

$$\psi_{ij}(y_i, y_j) = \begin{cases} 1 & \text{if } y_i = y_j \\ e^{-\alpha} & \text{if } y_i.c = y_j.c \wedge y_i \neq y_j \\ e^{-\beta} & \text{if } y_i.c \neq y_j.c \end{cases}, \tag{5}$$

where $y.c$ stands for the category of the venue y, α and β are parameters, and $\alpha < \beta$. Note that $y_i = y_j$ implies $y_i.c = y_j.c$; $y_i.c \neq y_j.c$ implies $y_i \neq y_j$. We define the potential in exponential form to avoid numerical underflow when computing the product (we can compute the sum of the exponents). Similar definitions are often seen from the definitions of Potts and Ising models [18].

Node potential ϕ_i: We use distance based measure to compute the node potential for each record r_i. Intuitively, venues closer to user locations are more likely to be visited by the user. The likelihood of a user visiting venue v_k at location r_i is modeled by a Gaussian distribution with center at the venue location, i.e.,

$$\phi_i(y_i = v_k) \leftarrow P(v_k.l | r_i.l) \propto \exp\left\{-\frac{D(r_i.l, v_k.l)}{2\sigma^2}\right\}, \tag{6}$$

where D is a function measuring the geographical distance, and σ^2 is a parameter. Finally, we normalize $\phi_i(y_i = v_k)$ by $\sum_{v_k} \phi_i(y_i = v_k)$.

Note that, as our focus is on leveraging user preference for semantic annotations, we simply use distance as our node potential for generality. Additional information (e.g., venue popularity, user ratings) can be easily incorporated into our framework through the design of the node potential.

4.3 Semi-Supervised Annotation

Although ground truth labels are hard to obtain, there may be occasions where a small amount of labeled data are available. Our proposed model can seamlessly incorporate labeled data. The integration of available labels does not require any learning on the data. Specifically, given a set of labeled data $R^* = \{(r_i, y_i^*)\}$ (y_i^* denotes the ground truth label), we can initialize the node potential $\phi_i(y_i) = 1$ if $y_i = y_i^*$, and 0 otherwise. For unlabeled data, the node potentials for each record is still estimated using Eq. (6). In the inference step, the given labels will be preserved, as any assignment violating those labels will have positive infinity as the energy. Aside from the initialization step, the inference procedure remains intact, which we discuss next.

4.4 Inference Method

Given the model specifications, the task is to infer the assignment of states (venues) that minimizes Eq. (4). While minimization on a MRF is NP-hard in general, a class of methods known as graph-cuts can solve the problem efficiently when the energy functions satisfy certain conditions [12]. This class of methods have been successfully applied to solve various computer vision problems [18].

Specifically, graph-cuts construct a graph such that the minimum cut on the graph also minimizes the energy function. The solution is exact for binary labels with semimetric energy. For multi-label cases, the α-expansion version of graph-cuts gives a local optimum that lies within a multiplicative factor of the optimal energy (the factor depends only on V) [4]. Furthermore, the time complexity is polynomial in the number of nodes, i.e., $O(|V|^3)$ or $O(|\mathcal{N}||V|\log|V|)$, where $|\mathcal{N}|$ is number of edges and $|V|$ is number of nodes [20]. We choose the α-expansion version of graph-cuts as our inference method for its theoretical guarantees and efficiency. We omit the details of the algorithm here for brevity, and refer interested readers to [4, 3, 12].

To use the α-expansion algorithm for multi-label problems, the pairwise energy needs to form a metric [4]. Next, we show that our definition of the pairwise energy indeed satisfies the requirement. Following the terminology in [4], we write the pairwise energy in our problem as:

$$E_{ij}(y_i, y_j) = -\log \psi_{ij}(y_i, y_j).$$

In order for E_{ij} to be a metric, we need to show that the following relation holds for all $y_i, y_j, \gamma \in C$:

$$\begin{aligned} & E_{ij}(y_i, y_j) + E_{ij}(\gamma, \gamma) \leq E_{ij}(y_i, \gamma) + E_{ij}(\gamma, y_j) \\ \Longleftrightarrow & E_{ij}(y_i, y_j) \leq E_{ij}(y_i, \gamma) + E_{ij}(\gamma, y_j), \end{aligned} \tag{7}$$

as $E_{ij}(\gamma, \gamma) = 0$ by our definition.

We now prove that Eq. (7) holds for our pairwise potential by exhaustion. When $y_i = y_j$, Eq. (7) always holds as $E_{ij}(y_i, y_j) = 0$. When $y_i \neq y_j$, we summarize the values of $E_{ij}(y_i, y_j)$, $E_{ij}(y_i, \gamma)$, and $E_{ij}(\gamma, y_j)$ for all possible cases in the following table (note that $\alpha < \beta$ by our definition):

$y_i.c = y_j.c$	$E_{ij}(y_i, y_j)$	$E_{ij}(y_i, \gamma)$	$E_{ij}(\gamma, y_j)$
$\gamma = y_i$ (or y_j)	α	0 (or α)	α (or 0)
$\gamma.c = y_i.c$	α	α	α
$\gamma.c \neq y_i.c$	α	β	β

$y_i.c \neq y_j.c$	$E_{ij}(y_i, y_j)$	$E_{ij}(y_i, \gamma)$	$E_{ij}(\gamma, y_j)$
$\gamma = y_i$ (or y_j)	β	0 (or β)	β (or 0)
$\gamma.c = y_i.c$ (or $y_j.c$)	β	α (or β)	β(or α)
$\gamma.c \neq y_i.c$ $\gamma.c \neq y_j.c$	β	β	β

From the above table, we can see that Eq. (7) holds for all possible choices of y_i, y_j, and γ.

As we discussed earlier, our model can also incorporate more sophisticated similarity measures on venues. For similarity measures that are not metric, one may choose the $\alpha - \beta$ swap algorithm of graph-cuts for inference. The $\alpha - \beta$ swap algorithm does not require the Eq. (7) to hold at the cost of losing the theoretical guarantees [18].

5. EXPERIMENT

5.1 Datasets

We collect data from three major cities in United States, namely, New York City (NYC), Chicago (CHI), and Los Angeles (LA). The statistics of our datasets are summarized in Table 1. We describe the details of our data collection below.

Dataset	#users	#records	#venues
New York City (NYC)	1,125	149,208	380,380
Chicago (CHI)	681	79,554	120,000
Los Angeles (LA)	480	58,397	157,411

Table 1: Statistics of datasets.

Venue Context Data: To obtain context datasets, we query Foursquare API to get information about the venues (i.e., venue name, venue category, and venue coordinates). Foursquare has a hierarchy of categories. We use the the lowest level of category (e.g., Sushi bar and Chinese restaurant) as it provides a fine-grained semantics revealing user interests. There are 687 categories in total. Venues that are checked in by at least one user from Foursquare are kept. Note that the API at most returns 50 venues for each query. In order to obtain an exhaustive set of venues, we query venues by partitioning the map into grids. For grids that the API return 50 or more venues, we further partition them into smaller sub-grids and query the sub-grids. We perform these steps recursively until less than 50 venues are returned from the API for each sub-grid.

It is worth noting that the number of users visiting each venue follows a long-tail distribution, i.e., most venues have low likelihoods to be visited. As we mentioned before, prior information about venue popularity can be easily incorporated into our MRF model. And we have observed consistent improvement in performance given such prior. In this paper, we report experiment results in a more general setting where we do not assume to have such additional information.

Human Mobility Data: To obtain mobility datasets, we crawl geo-tagged tweets (i.e., with GPS coordinates) by user over a period of two years, i.e., from Oct, 2011 to July, 2013. We are interested in check-in tweets posted from Foursquare and only keep those tweets, as the check-in information provides us ground truth for evaluation. In particular, we only select users who have at least 40 check-in tweets. Geo-tags of the check-in tweets are mapped to coordinates of the corresponding venues. We add noise to the coordinates of the check-ins to obtain the final mobility datasets. The noise is drawn from a zero-mean Gaussian with variance σ^2. By default, we set $\sigma = 0.0002$, i.e., the probability of observing a noise larger then 40 meters being drawn is less than 0.05.

Ground Truth: To obtain ground truth for evaluation, we match those check-in tweets with the venues in our venue context dataset. The check-in posts start with *"I'm at"* followed by a venue name. A check-in tweet is matched with a venue if: (i) the geo-tag of the tweet is within 100 meters of the venue location, and (ii) similarity between venue names is large than 0.8, where the similarity is defined as the length of the longest common subsequence divided by the length of the longer sequence.

	Accuracy			Relative improvement		
	NYC	CHI	LA	NYC	CHI	LA
DIST	0.40	0.43	0.48	0	0	0
GDM	0.34	0.40	0.51	-0.14	-0.06	0.05
HMM	0.03	0.04	0.06	-0.91	-0.89	-0.87
NR	0.17	0.20	0.23	-0.58	-0.52	-0.52
TR	0.47	0.50	0.55	0.16	0.16	0.14
SR	0.59	0.62	**0.68**	0.47	0.44	**0.39**
STR	**0.60**	**0.63**	**0.68**	**0.49**	**0.46**	**0.39**
STR$^+$	NYC	CHI	LA	NYC	CHI	LA
5%	0.61	0.64	0.69	0.51	0.49	0.42
10%	**0.64**	**0.68**	**0.73**	**0.60**	**0.57**	**0.50**

Table 2: Performance of all methods on three datasets. STR$^+$ is STR under a semi-supervised setting where partial labels are given.

5.2 Evaluation Metrics

Annotation accuracy on all records. We evaluate all the method using accuracy and relative improvement. Accuracy is defined by $\frac{N_m}{N}$, where N_m is the number of cases correctly annotated by the method, and N is the total number of records. The relative improvement is computed by $\frac{N_m - N_{\mathsf{DIST}}}{N_{\mathsf{DIST}}}$, where N_{DIST} is the number of records correctly annotated by using distance only. Both our method and HMM output the most probable context (by MAP inference). Since the output does not include scores for unmatched context, we do not use precision and recall as the metric.

Improvement by number of users. The accuracy measures evaluate the performance over all the records, but they cannot reflect whether methods are generalizable to different users. Therefore, we also look at the percentage of users on whom the annotation accuracy shows improvement (over a baseline method). A better method should also demonstrate improvement on larger number of users.

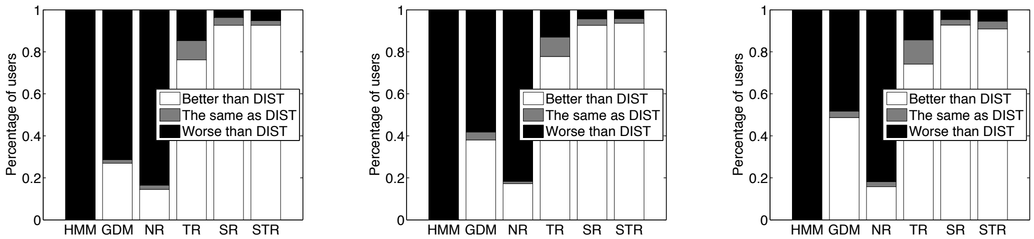
5.3 Methods for Comparison

In this paper, we compare our method to the following baselines:

Distance-based method (DIST). This method simply matches the closest venue to each mobility record. This is equivalent to using the node potential $\phi(y_i)$ only in our model for annotation.

Gaussian de-noising model (GDM). The spatial regularity suggests that spatially close records are likely to be repeated visits to the same venue. Those repeated visits are assumed to follow a Gaussian distribution. Following this intuition, we fit the location records by grids with a Gaussian distribution, where the grid size is 500×500 meters (consistent with the grid size used in the pre-processing step of our models). Then, the venue closest to the distribution center of each grid is used to annotate records in that grid. Note that this strategy of specifying spatial constraint is similar to that of [36].

Hidden Markov model (HMM). Yan et al. [28, 27] proposed a hidden Markov model to address the trajectory annotation problem that is similar to ours. The proposed method is designed for annotation on densely sampled trajectory data, and assumes that each point depends on the point observed before. For our problem, we consider the



(a) NYC (b) CHI (c) LA

Figure 3: Evaluation of all methods by user.

states to be a set of venues and the transition to be similarity between venues. We use the same pairwise potentials (ψ_{ij}) between venues and node potentials as in our models for HMM.

We also compare our proposed model with its own variants. Note that all models use the same pairwise potential as defined in Eq. (5), where $\alpha = 0.1$ and $\beta = 0.2$ (please see Section 5.7 for experiment results on parameter sensitivity). The differences lie in the definitions of the edge list.

Non-regularity model (NR). NR model does not incorporate any regularity assumption. We simply use a fully connected edge list as our $\mathcal{N}$.

Temporal regularity-based model (TR). TR model incorporates the temporal regularity we observed from the data. Mobility records that are temporally close are connected, i.e., we use $\mathcal{N}_T$ as the edge list. We set $\xi_T = 1$ (hour) for $\mathcal{N}_T$.

Spatial regularity-based model (SR). SR model incorporates the spatial regularity we observed from the data. Similar to TR model, SR model connects spatially close records. i.e., we use $\mathcal{N}_D$ as the edge list. We set $\xi_D = 100$ (meters) for $\mathcal{N}_D$.

Full model (STR). We name our full model STR and follow definitions in Section 4.2 to construct the model. Same parameters are used for all models.

In our problem, venues that are far away from a mobility record will have negligible probabilities to be assigned. Therefore, we only compute node potentials for the venues that are within 500 meters of each record, and keep the top-5 closest venues as candidates for annotation. Accordingly, we filtered records that cannot be annotated correctly and evaluate on the rest for all methods.

5.4 Performance Comparison

Annotation accuracy on all records

We start by comparing the accuracy and the relative improvement of all methods. Table 2 summarizes the result on three datasets. STR outperforms all the other methods on NYC and CHI datasets, achieving an improvement in accuracy of 0.49 and 0.46 compared with DIST, respectively. On LA dataset, STR and SR both achieve an relative improvement of 0.39. Meanwhile, TR consistently improves over DIST by at least 0.14 on all three datasets. These results suggest that one can improve the annotation over DIST by considering either the spatial or temporal regularity, and achieve the best performance overall when considering both. Furthermore, SR achieves a better performance than TR. This suggests that considering spatial regularity alone is more effective than considering temporal regularity alone. NR shows no improvement over DIST on three datasets. The poor performance of NR is a result of its over-strict single-modal preference assumption. As shown in Figure 2(a), people tend to have multi-modal preferences, i.e., visiting several categories of venues frequently.

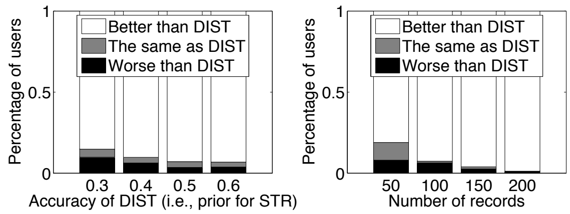
As for other baselines, GDM performs similarly compared with DIST on all three datasets. The center estimations are often erroneous when there are only a few records within a grid. As a result, GDM performs poorly on grids that have less than 10 records. Meanwhile, HMM performs significantly worse than DIST. This is because the original HMM proposed in [28] is designed for annotation task on continuously sampled trajectory data. Therefore, the model focuses on capturing the transition characteristics, such as people tend to go to restaurant after work and then go home. In our problem, the mobility records are temporally sparse, i.e., there may be a gap of several days between two consecutive records. Consequently, these records may not exhibit any significant transitional relationship.

Improvement by number of users

In this section, we study how the annotation accuracy of each method compares with that of DIST on each user. Figure 3(a-c) show the percentage of users on whom the annotation accuracy by STR is higher than DIST on all three datasets. As one can see, both STR and SR have improvements on more than 90% of the users. Moreover, STR shows improvement on the largest number of users. This suggests that STR is applicable on more users than any other method.

However, there are 5% of the users on whom STR cannot improve over DIST. We further examine STR to see how the node potential ϕ_i of STR affects the final annotation result. Recall that annotation using the node potential only is equivalent to DIST. In this experiment, we bin the users by annotation accuracy values of the node potential and study the performance of STR on the users falling into each bin. Figure 4(a) shows that STR improves over DIST on more users, as the node potential becomes more effective, i.e., increases from 0.3 to 0.6. The result suggests that STR is more likely to perform worse than DIST when the annotation accuracy by node potential is lower.

We also investigate how the number of records affects the annotation result. Figure 4(b) shows the performance of STR with respect to different numbers of records. As the number of records increases from 50 to 200, the percentage of users on whom STR shows improvement increases from



(a) (b)

Figure 4: Evaluation by user. (a) Performance of STR with respect to the accuracy of DIST. (b) Performance of STR with respect to the number of records for each user.

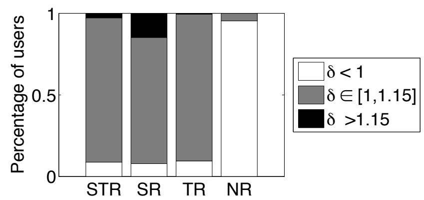


Figure 5: Evaluation on the energy functions of our method and its variants.

80% to 99%. This trend suggests that STR performs better for users with more records.

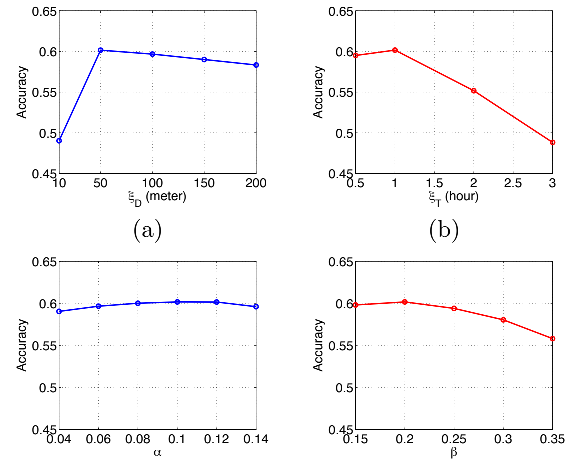
5.5 Annotation with Partially Labeled Data

As we discussed in Section 4.3, labeled data can be easily incorporated into our method. Next we investigate how much the annotation accuracy can be improved via semi-supervision, i.e., providing a small amount (i.e., 5% and 10% for each user) of labeled check-ins to STR. The labeled records are randomly selected and the average accuracy of 10 independent runs are reported. From Table 2, we can see that STR$^+$(10%) achieves better accuracy than other methods. Furthermore, the annotation accuracy increases as more labels are given. Specifically, when the percentage of labeled data increases from 5% to 10%, the accuracy increases by 0.03 to 0.06 on the three datasets.

5.6 Evaluation on the Energy Functions

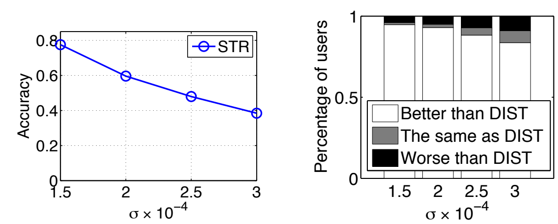
To further understand the behavior of our method, we also examine the energy functions for our full model (STR) and its variants (NR,SR,TR) on NYC dataset. The differences among the objectives lie in the definitions of the edge list. Ideally, if an energy function is well-designed, the ground truth assignments should yield the lowest energy values among all possible assignments. While this is difficult to verify directly, we can still gain insight about each model by comparing the energy value achieved by the ground truth assignment (E_{truth}) with the energy value obtained by graph-cuts (E_{model}). Specifically, the model must be ill-designed if $E_{model} < E_{truth}$. Further, if $E_{model} > E_{truth}$, one can conclude that graph-cuts was unable to find the true minimizer.

To this end, we compute the ratio $\delta = E_{model}/E_{truth}$ and show the percentage of users with δ being in different intervals in Figure 5. As one can see, E_{truth} is less than E_{model} for more than 90% of the users for STR, SR, and TR. However, for 95% of the users we have $E_{NR}/E_{truth} < 1$.



(a) (b) (c) (d)

Figure 6: Parameter sensitivity.



(a) (b)

Figure 7: Sensitivity to data noises. (a) Accuracy of DIST and STR as a function of the noise level. (b) Performance of STR with respect to noise level by user.

This further suggests that the energy function of NR is ill-designed as the result of the over-strict 1-modal preference assumption.

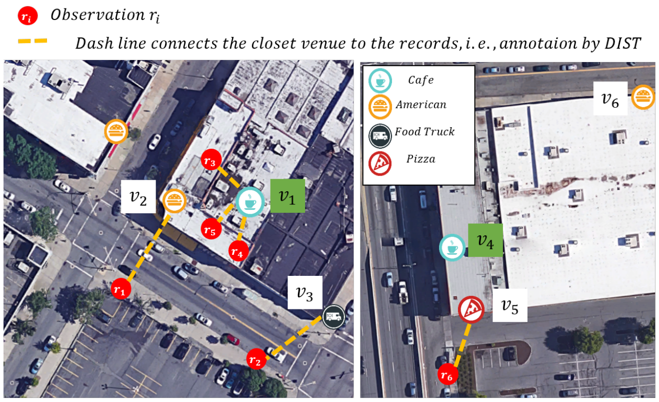
In addition, for STR, SR, and TR, we have $\delta \in [1, 1.15]$ for more than 75% of the users. The fact that E_{model} is close to E_{truth} implies that the inference method probably obtains a good approximation to the optimum for most of the users.

5.7 Parameter Sensitivity

In this section, we study the sensitivity of parameters on NYC dataset. We first study the effect of two thresholds in the definitions of spatial and temporal neighbors, i.e., ξ_D and ξ_T. We vary ξ_D from 10 meters to 200 meters and ξ_D from 0.5 hour to 3 hours. Figure 6(a) and Figure 6(b) show the accuracy as a function of ξ_D and ξ_T, respectively. The accuracy remains relatively stable for $\xi_D \in [50, 200]$ meters (varies by 0.02), and for $\xi_T \in [0.5, 2]$ hours (varies by 0.05).

Next, we fix $\xi_D = 100$ meters and $\xi_T = 1$ hour, and vary α from 0.01 to 0.15 and β from 0.1 to 0.35. Figure 6(c) shows that the accuracy varies less than 0.01 as α changes and achieves the best performance when $\alpha = 0.1$. Similarly, the accuracy varies by 0.04 as β changes. The best performance is attained with $\beta = 0.2$.

The result suggests that STR is generally insensitive to the parameters. Comparing with α and β, parameters ξ_D and ξ_T have more impacts on the annotation result.



(a) (b)

Ground truth
Assigments by STR
Assigments by DIST

r_i(timestamp)	v_1	v_2	v_3
$r_1(4:54\ p.m.)$	1.42	0.59	4.6
$r_2(5:53\ p.m.)$	1.31	1.77	0.58
$r_3(5:40\ p.m.)$	0.27	1.56	3.51
$r_4(12:54\ p.m.)$	0.54	1.17	2.21
$r_5(11:13\ p.m.)$	0.36	1.43	2.81

	v_4	v_5	v_6
$r_6(6:11\ p.m.)$	1.27	0.34	4.61

Node energy of each venue being visited, i.e., ϕ_i estimated by distance factor.

(c)

Figure 8: Case study on one user from NYC dataset. (a-b). Mobility records and venues are shown on the map. Each red node is one observation (i.e., mobility record). The yellow dash lines connect the records with annotations by DIST. The ground truth venues are colored green. (c). The table shows the node energy for each venue. The venues closest to the records are marked red. The annotations by STR are marked blue.

5.8 Sensitivity to Data Noises

In this experiment, we study how STR performs with respect to the noise levels. Noises drawn from a zero-mean Gaussian with varying σ are added to the locations of the mobility records. We vary the value of σ from 0.0001 to 0.0003. This is equivalent to vary the distance range in which 95% of the deviations fall from 20 meters to 60 meters. Figure 7(a) shows the accuracy of STR with respect to different noise levels. The accuracy of STR decreases from 0.79 to 0.4 as σ increases from 0.0001 to 0.0003.

We also study the performance of both methods for each user. Figure 7(b) shows that, as the noise level increases, the percentage of users on whom STR outperforms DIST decreases. However, the number of users on whom STR is worse than DIST remains similar. In all cases, our MRF model which combines the node and pairwise potentials rarely yields worse performance than using the node potential only.

5.9 Case Study

In this section, we conduct case study on the mobility records of a user in NYC dataset to further demonstrate the effectiveness of STR. We pick six mobility records from this user and visualize them in Figure 8(a-b). From the check-ins of the user, we know that the user visited two cafes, i.e., venues v_1 and v_4, for records r_1 to r_5 and record r_6, respectively (marked green in the figure). There are also other venues nearby, such as American restaurants. Figure 8(c) shows the node energy for each venue computed as in Eq. (6). The venues with the highest node potentials (i.e., closest to the records) are marked red, which are also the venues annotated by DIST. The annotations by STR are colored blue in the table.

From Figure 8(a), we can see that DIST incorrectly annotates venue v_2 to record r_1, and venue v_3 to record r_2, due to the noise in the observations. Since records r_1 to r_5 are spatially close, STR prefers annotating them with the same venue while also considering the distance, resulting in the correct annotations. Similarly, in Figure 8(b), DIST incorrectly annotates venue v_5 to record r_6. At the same time, STR prefers annotating record r_6 with a venue that has the category cafe, as we inferred from the records r_1 to r_5 that this user often goes to cafe in the afternoon.

Meanwhile, in this example, GDM would annotate records r_1 to r_5 with venue v_2 (not shown). Because there are only 5 records within the grid, the estimation of venue location by GDM is erroneous.

6. CONCLUSION

In this paper, we address the problem of annotating venue to mobility records of mobile users. Our proposed method incorporates characteristics of human mobility and models user preference as consistency of the annotations.

There are several extensions that worth further investigation. First, we only use category information of the venue as features. We can further consider more sophisticated features in order to measure the similarity between venues. Second, there are different ways to fine tune the model specification. For example, we can further consider different scales of temporal regularity to define edge list, such as weekly, monthly, and yearly. The edge potential can also be modified to include features of the venues (e.g., price, user rating, and hierarchy of categories). Finally, we focus on mobility records of individual users in this work. We plan to further investigate how to understand the semantics of patterns in collective mobility data.

7. ACKNOWLEDGEMENTS

The work was supported in part by NSF award #1618448 and #1544455. The views and conclusions contained in this paper are those of the authors and should not be interpreted as representing any funding agencies.

8. REFERENCES

[1] L. O. Alvares, V. Bogorny, B. Kuijpers, J. A. F. de Macedo, B. Moelans, and A. Vaisman. A model for enriching trajectories with semantic geographical information. In *Proc. ACM GIS*, 2007.

[2] J. Bao, Y. Zheng, and M. F. Mokbel. Location-based and preference-aware recommendation using sparse geo-social networking data. In *ACM GIS*, 2012.

[3] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *PAMI*, 2004.

[4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 2001.

[5] C. Cheng, H. Yang, M. R. Lyu, and I. King. Where you like to go next: Successive point-of-interest recommendation. In *IJCAI*, 2013.

[6] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proc. ACM KDD*, 2011.

[7] V. M. de Lira, S. Rinzivillo, C. Renso, V. C. Times, and P. C. Tedesco. Investigating semantic regularity of human mobility lifestyle. In *IDEAS'14*. ACM.

[8] N. Eagle, A. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. In *Proc. PNAS*, 2009.

[9] N. Eagle and A. S. Pentland. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066, 2009.

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

[11] B. Guc, M. May, Y. Saygin, and C. Körner. Semantic annotation of gps trajectories. In *Proc AGILE*, 2008.

[12] V. Kolmogorov and R. Zabin. What energy functions can be minimized via graph cuts? *PAMI*, 2004.

[13] Z. Li, B. Ding, J. Han, R. Kays, and P. Nye. Mining periodic behaviors for moving objects. In *Proc. ACM KDD'10*.

[14] 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 *Proc. ACM KDD*, 2014.

[15] L. Liao. *Location-based activity recognition*. PhD thesis, University of Washington, 2006.

[16] B. Liu, Y. Fu, Z. Yao, and H. Xiong. Learning geographical preferences for point-of-interest recommendation. In *Proc. ACM KDD*, 2013.

[17] N. Mamoulis, H. Cao, G. Kollios, M. Hadjieleftheriou, Y. Tao, and D. Cheung. Mining, indexing, and querying historical spatiotemporal data. In *Proc. ACM KDD*, 2004.

[18] K. P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[19] M.-H. Park, J.-H. Hong, and S.-B. Cho. Location-based recommendation system using bayesian users preference model in mobile devices. In *UIC*. 2007.

[20] R. Sedgewick and K. Wayne. Algorithms 4a. edition.

[21] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science'10*.

[22] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, and C. Vangenot. A conceptual view on trajectories. *Trans. IEEE TKDE*, 2008.

[23] H. Su, K. Zheng, K. Zeng, J. Huang, S. Sadiq, N. J. Yuan, and X. Zhou. Making sense of trajectory data: A partition-and-summarization approach. In *ICDE'15*.

[24] F. Wu, Z. Li, W.-C. Lee, H. Wang, and Z. Huang. Semantic annotaion of mobility data using social media. In *Proc. WWW*, 2015.

[25] X. Xiao, Y. Zheng, Q. Luo, and X. Xie. Finding similar users using category-based location history. In *GIS'10*. ACM.

[26] K. Xie, K. Deng, and X. Zhou. From trajectories to activities: a spatio-temporal join approach. In *Proc. LBSN*, 2009.

[27] Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer. Semantic trajectories: Mobility data computation and annotation. *ACM Trans. TIST'13*.

[28] Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer. Semitri: a framework for semantic annotation of heterogeneous trajectories. In *Proc. EDBT*, 2011.

[29] M. Ye, D. Shou, W.-C. Lee, P. Yin, and K. Janowicz. On the semantic annotation of places in location-based social networks. In *Proc. ACM SIGKDD*, 2011.

[30] M. Ye, P. Yin, and W.-C. Lee. Location recommendation for location-based social networks. In *Proc. SIGSPATIAL GIS*. ACM, 2010.

[31] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *ACM SIGIR'11*.

[32] H. Yin, Y. Sun, B. Cui, Z. Hu, and L. Chen. Lcars: a location-content-aware recommender system. In *Proc. ACM SIGKDD*, 2013.

[33] H. Yin, X. Zhou, Y. Shao, H. Wang, and S. Sadiq. Joint modeling of user check-in behaviors for point-of-interest recommendation. In *ACM CIKM'15*.

[34] Y. Yu and X. Chen. A survey of point-of-interest recommendation in location-based social networks. In *Workshops at AAAI*, 2015.

[35] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann. Time-aware point-of-interest recommendation. In *ACM SIGIR'13*.

[36] W. Zhang and J. Wang. Location and time aware social collaborative retrieval for new successive point-of-interest recommendation. In *CIKM'15*. ACM.

[37] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Collaborative location and activity recommendations with gps history data. In *WWW*, 2010.

[38] Y. Zheng. Methodologies for cross-domain data fusion: an overview. *Big Data, IEEE Transactions on*, 2015.

[39] Y. Zheng, Y. Chen, Q. Li, X. Xie, and W.-Y. Ma. Understanding transportation modes based on gps data for web applications. *Trans. ACM TWEB*, 2010.

[40] Y. Zheng, X. Xie, and W.-Y. Ma. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, 2010.

[41] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proc. WWW*, 2009.

[42] Y. Zheng and X. Zhou. *Computing with spatial trajectories*. Springer Science & Business Media, 2011.

[43] Y. Zhong, N. J. Yuan, W. Zhong, F. Zhang, and X. Xie. You are where you go: Inferring demographic attributes from location check-ins. In *Proc. ACM WSDM*, 2015.