Collaborative Place Models

Berk Kapicioglu
Foursquare Labs
berk.kapicioglu@gmail.com

David S. Rosenberg
YP Mobile Labs
david.davidr@gmail.com

Robert E. Schapire
Princeton University
schapire@cs.princeton.edu

Tony Jebara
Columbia University
jebara@cs.columbia.edu

Abstract

A fundamental problem underlying location-based tasks is to construct a complete profile of users' spatiotemporal patterns. In many real-world settings, the sparsity of location data makes it difficult to construct such a profile. As a remedy, we describe a Bayesian probabilistic graphical model, called Collaborative Place Model (CPM), which infers similarities across users to construct complete and time-dependent profiles of users' whereabouts from unsupervised location data. We apply CPM to both sparse and dense datasets, and demonstrate how it both improves location prediction performance and provides new insights into users' spatiotemporal patterns.

1 Introduction[1]

During the last couple of years, positioning devices that measure and record our locations have become ubiquitous. The most common positioning device, the smartphone, is projected to be used by a billion people in the near future [Davie, 2012]. This surge in positioning devices has increased the availability of location data, and provided scientists with new research opportunities, such as building location-based recommendation systems [Hao *et al.*, 2010; Zheng *et al.*, 2010; 2009], analyzing human mobility patterns [Brockmann *et al.*, 2006; Gonzalez *et al.*, 2008; Song *et al.*, 2010], and modeling the spread of diseases [Eubank *et al.*, 2004].

A fundamental problem underlying many location-based tasks is modeling users' spatiotemporal patterns. For example, a navigation application that has access to traffic conditions can warn the user about when to depart, without requiring any input from the user, as long as the application can accurately model user's destination. Similarly, a restaurant application can provide a list of recommended venues, even reserve them while space is still available, by modeling the probable locations the user might visit for lunch.

One of the challenges in building such models is the sparsity of location datasets. Due to privacy considerations [Wernke *et al.*, 2012] and high energy consumption of positioning hardware [Oshin *et al.*, 2012], location datasets continue to grow larger while remaining relatively sparse per user. For example, most mobile phone operating systems only allow applications to log users' locations when the application is active, but not when it is running in the background. As a result, many location datasets have very little information per user. In Figure 1, we compare a sparse dataset that is retrieved from a mobile ad exchange and a dense dataset that is retrieved from a cellular carrier. The sparse dataset has 15 times less observations per user than the dense dataset, which makes it harder to infer its spatiotemporal patterns.

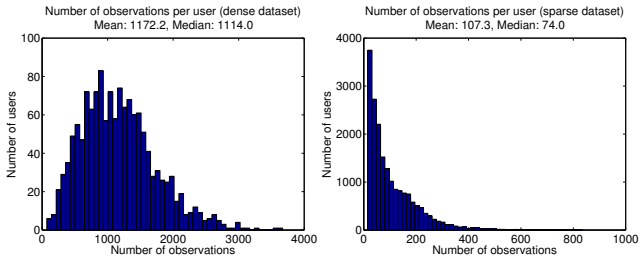

Figure 1: Number of observations per user. The left plot represents a dense mobile carrier dataset and the right plot represents a sparse mobile ad exchange dataset. The dense dataset has 15 times more observations per user than the sparse dataset.

In the past, there has been some work on spatiotemporal modeling as it pertains to location prediction. Cho et al. [Cho *et al.*, 2011] proposed a two-state mixture of Gaussians that leverages the social relationships between users, but they limited their model to only represent "home" and "work". Gao et al. [Gao *et al.*, 2012] designed a Markov model that takes temporal context into account in order to predict the user's location in the immediate future. De Domenico et al. [De Domenico *et al.*, 2012] presented a location prediction algorithm where the dataset consisted of tens of thousands of GPS observations, collected every few minutes and over the span of a year. Their algorithm exploited the high density of the location dataset, as well as the social relationships between users, by predicting the user's geographic coordinates in the immediate future based on the similarities between the user's most recent trajectory and all of his previous trajectories.

In most of these studies, the proposed models were designed for near-term forecasts, and they relied on making

[1] Supplements are available at http://www.berkkapicioglu.com.

predictions based on the most recent observations. However, there are many real-world applications where the test example and the most recent training example are temporally apart, and for all purposes, statistically independent. For such predictions, a model that relies on the most recent observations would not suffice; instead, the model would need to make predictions based on the user's global spatiotemporal patterns.

In addition to the work listed above, researchers have also studied spatiotemporal modeling as it pertains to the detection of significant places and routines. Eagle and Pentland [Eagle and Pentland, 2009] applied eigendecomposition to the Reality Mining dataset, where all locations were already labeled as "home" or "work", and extracted users' daily routines. Farrahi and Gatica-Perez [Farrahi and Perez, 2011] used the same dataset, but extracted the routines using Latent Dirichlet Allocation (LDA) instead of eigendecomposition. Liao et al. [Liao *et al.*, 2005; 2007] proposed a hierarchical conditional random field to identify activities and significant places from the users' GPS traces, and since their algorithm was supervised, it required locations to be manually labeled for training. In contrast to previous work, our model does not require labeled data; instead, it relies only on user IDs, latitudes, longitudes, and time stamps.

In this paper, we propose a new Bayesian probabilistic graphical model, called Collaborative Place Model (CPM), which recovers the latent spatiotemporal structure underlying unsupervised location data by analyzing patterns shared across all users. CPM is a generalization of the Bayesian Gaussian mixture model (GMM), and assumes that each user is characterized by a varying number of place clusters, whose spatial characteristics, such as their means and covariances, are determined probabilistically from the data. However, unlike GMM, CPM also assigns users *weakly similar* temporal patterns; ones which do not force different users to have the same place distribution during the same weekhour.

The spatiotemporal patterns extracted by CPM are helpful in leveraging both sparse and dense datasets. In case of sparse data, the model infers a user's place distribution at a particular weekhour, even if the user has not been observed during that weekhour before. In case of dense data, sampling bias usually yields fewer observations for certain hours (e.g. users make more phone calls during day time than after midnight), and the model successfully infers the user's behavior during these undersampled hours. In both cases, the model combines the globally shared temporal patterns with user's own spatiotemporal patterns and constructs a customized, complete, and time-dependent profile of the user's locations.

Aside from its quantitative benefits, CPM also provides qualitative insights about the universal temporal patterns of human populations. Even though the model is given no prior information about the relationship between weekhours, it successfully extracts temporal clusters such as the hours spent during morning commute, work, evening commute, leisure time after work, and sleeping at night.

The paper proceeds as follows. In Section 2, we provide a formal description of CPM. In Section 3, we derive the inference algorithms. In Section 4, we demonstrate our model on two real location datasets. We conclude in Section 5.

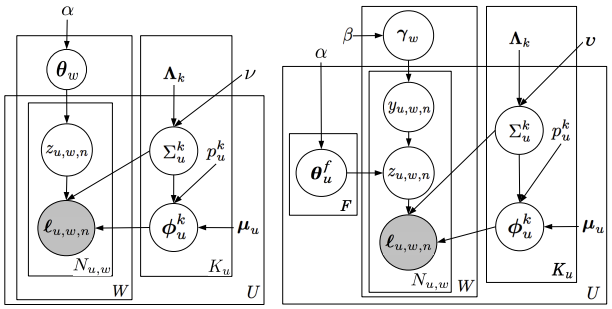

Figure 2: Graphical models. An earlier prototype we built is represented on the left and CPM is represented on the right. The geographic coordinates, denoted by $\boldsymbol{\ell}$, are the only observed variables.

2 Collaborative Place Model

In this section, we provide a formal description of CPM and describe its generative process. CPM comprises a spatial part, which represents the inferred place clusters, and a temporal part, which represents the inferred place distributions for each weekhour. The model is depicted on the right side of Figure 2.

The temporal part of CPM assumes that, for each user u and weekhour w, the corresponding place distribution is a convex combination of F factorized place distributions, each of which is represented by $\boldsymbol{\theta}_u^f$. These factorized place distributions are dependent on the user, but they are independent of the weekhour. In contrast, the coefficients of the convex combination, represented by $\boldsymbol{\gamma}_w$, are shared across all users. Using this factorization, CPM combines global temporal patterns with user-specific spatiotemporal patterns and infers each user's place distribution even for weekhours where the user has not been observed before. Furthermore, by not restricting each user to have the same number of places or each weekhour to have the same distribution over places, CPM allows a realistic and flexible representation of the users' location profiles.

The flexible temporal representation provided by CPM turns out to be essential for our problem. Early in the project, we have built a simpler prototype, depicted on the left side of Figure 2, which constrained all users to have the same weekhour distribution. However, it performed even worse than the baseline model with respect to the usual metrics (i.e. held-out log-likelihood), so we abandoned it. In contrast to this prototype, CPM allows us both to share temporal information across users and to provide a more flexible temporal representation for each user.

Let u represent a user, ranging from 1 to U, and let f represent a factor index, ranging from 1 to F. For each user u and factor index f, let $\boldsymbol{\theta}_u^f \in \mathbb{R}^{K_u}$ denote the Dirichlet random variable that represents the corresponding factorized place distribution, where K_u denotes the number of places for each user u. Intuitively, $\boldsymbol{\theta}_u^f \in \mathbb{R}^{K_u}$ describes the building blocks of user's place preferences. Furthermore, let w denote a weekhour, ranging from 1 to W. The Dirichlet random variable $\boldsymbol{\gamma}_w \in \mathbb{R}^F$ denotes the factor weights, which when combined with the user's factorized place distributions,

yields the user's place distribution for weekhour w.

At a high level, the generative process proceeds as follows. The variable $\boldsymbol{\gamma}_w$ generates $y_{u,w,n}$, the factor assignment for user u and observation n at weekhour w, and the factor assignment generates the place assignment $z_{u,w,n}$ from the factorized place distribution $\boldsymbol{\theta}_u^{y_{u,w,n}}$. The place assignment is in turn used to sample the observed coordinates $\boldsymbol{\ell}_{u,w,n}$ from the corresponding place cluster.

The place clusters, whose means and covariances are unique to each user, are modeled by the spatial part of CPM. Intuitively, we expect each place cluster to correspond to locations such as "home", "work", and "gym". Given a user u and a place index k, each place cluster is characterized by a bivariate normal distribution, with mean $\boldsymbol{\phi}_u^k$ and covariance $\boldsymbol{\Sigma}_u^k$. The observed coordinates $\boldsymbol{\ell}$ are considered to be noisy observations sampled from these place clusters.

Our model uses conjugate priors because of the computational advantages they provide in Bayesian inference. For the random variables $\boldsymbol{\theta}_w^f$ and $\boldsymbol{\gamma}_w$, the prior is a symmetric Dirichlet with concentration parameter $\alpha > 0$, where $\text{Dirichlet}_K(\alpha)$ denotes the K-dimensional distribution. For the random variables associated with the place clusters, $\boldsymbol{\phi}_u^k$ and $\boldsymbol{\Sigma}_u^k$, the prior is a normal-inverse-Wishart (NIW) distribution. $\boldsymbol{\Lambda} \in \mathbb{R}^{2\times 2}$ is a positive definite scale matrix, ν>1 indicates the degrees of freedom, and together they define the distribution over the covariance matrix. The parameter $\boldsymbol{\mu}_u$ is customized for each user u and is computed as the mean of the user's historical locations. The parameter p_k^u is set such that the prior covariance of the place mean (i.e. prior covariance of $\boldsymbol{\phi}_u^k$) is very large.

In location datasets, a user sometimes logs locations that are one-offs and are not representative of the user's regular location profile. To ensure that such outliers do not affect how the model infers the user's regular place clusters, we designate the last place, place K_u, as a special outlier place and set its covariance's prior mean (i.e. prior mean of $\boldsymbol{\Sigma}_u^{K_u}$), as determined by $\boldsymbol{\Lambda}_{K_u}$, to be very large. As for the covariance of regular clusters, we set the remaining scale matrices, $\boldsymbol{\Lambda}_{-K_u}$, such that each coordinate of the covariance's prior mean has a standard deviation of 150 meters. This is a reasonable size for a place cluster, as it is large enough to encapsulate both the potential inaccuracies of location hardware and the inherent noise in the users' locations, but small enough to ensure that each place cluster corresponds to a single intuitive place, such as "home" or "work".

Let $\mathcal{N}$ denote the normal distribution and IW denote the inverse-Wishart distribution. Let $\text{Dirichlet}_K(\cdot)$ denote a symmetric Dirichlet if its parameter is a scalar and a general Dirichlet if its parameter is a vector. The generative process of CPM is described in more formal terms below. Further technical details about the distributions we use in our model can be found in the supplement.

1. For each weekhour w, draw a distribution over factors $\boldsymbol{\gamma}_w \sim \text{Dirichlet}_F(\boldsymbol{\beta})$.
2. For each user u and factor f, draw a factorized place distribution $\boldsymbol{\theta}_u^f \sim \text{Dirichlet}_{K_u}(\alpha)$.
3. For each user u and place k,
   (a) Draw a place covariance $\boldsymbol{\Sigma}_u^k \sim IW(\boldsymbol{\Lambda}_k, \nu)$.
   (b) Draw a place mean $\boldsymbol{\phi}_u^k \sim \mathcal{N}\left(\boldsymbol{\mu}_u, \frac{\boldsymbol{\Sigma}_u^k}{p_k^u}\right)$.
4. For each user u, weekhour w, and observation index n,
   (a) Draw a factor assignment $y_{u,w,n} \sim \text{Categorical}(\boldsymbol{\gamma}_w)$.
   (b) Draw a place assignment $z_{u,w,n} \sim \text{Categorical}(\boldsymbol{\theta}_u^{y_{u,w,n}})$.
   (c) Draw a location $\boldsymbol{\ell}_{u,w,n} \sim \mathcal{N}(\boldsymbol{\phi}_u^{z_{u,w,n}}, \boldsymbol{\Sigma}_u^{z_{u,w,n}})$.

3 Inference

In this section, we derive our inference algorithm, and we present our derivation in multiple steps. First, we derive a collapsed Gibbs sampler to sample from the posterior distribution of the categorical random variables conditioned on the observed geographic coordinates. Second, we derive the conditional likelihood of the posterior samples, which we use to determine the sampler's convergence. Third, we derive formulas for approximating the posterior expectations of the non-categorical random variables conditioned on the posterior samples. Finally, in the last step, we combine all the previous derivations to construct a simple algorithm for efficient posterior inference. We state the main results below. The proofs are provided in the supplement.

In Lemmas 1 and 2, we describe the collapsed Gibbs sampler for variables $\boldsymbol{z}$ and $\boldsymbol{y}$, respectively. Given a vector $\boldsymbol{x}$ and an index k, let $\boldsymbol{x}_{-k}$ indicate all the entries of the vector excluding the one at index k. We assume that $i = (u, w, n)$ denotes the index of the variable that will be sampled. First, we state the posterior probability of $\boldsymbol{z}$.

Lemma 1. *The unnormalized probability of z_i conditioned on the observed location data and remaining categorical variables is*

$$p\left(z_i = k \mid y_i = f, \boldsymbol{z}_{-i}, \boldsymbol{y}_{-i}, \boldsymbol{\ell}\right) \propto t_{\tilde{v}_k^u - 1}\left(\boldsymbol{\ell}_i \mid \tilde{\boldsymbol{\mu}}_k^u, \frac{\tilde{\boldsymbol{\Lambda}}_u^k (\tilde{p}_k^u + 1)}{\tilde{p}_k^u (\tilde{v}_k^u - 1)}\right)\left(\alpha + \tilde{m}_{u,\cdot}^{k,f}\right).$$

The parameters $\tilde{v}_k^u$, $\tilde{\boldsymbol{\mu}}_k^u$, $\tilde{\boldsymbol{\Lambda}}_u^k$, and $\tilde{p}_k^u$ are defined in the proof. t denotes the bivariate t-distribution and $\tilde{m}_{u,\cdot}^{k,f}$ denotes counts, both of which are defined in the supplement.

Next, we state the posterior probability of $\boldsymbol{y}$.

Lemma 2. *The unnormalized probability of y_i conditioned on the observed location data and remaining categorical variables is*

$$p\left(y_i = f \mid z_i = k, \boldsymbol{y}_{-i}, \boldsymbol{z}_{-i}, \boldsymbol{\ell}\right) \propto \frac{\alpha + \tilde{m}_{u,\cdot}^{k,f}}{K_u \alpha + \tilde{m}_{u,\cdot}^{\cdot,f}}\left(\beta_{w,f} + \tilde{m}_{\cdot,w}^{\cdot,f}\right),$$

where the counts $\tilde{m}_{u,\cdot}^{k,f}$, $\tilde{m}_{u,\cdot}^{\cdot,f}$, and $\tilde{m}_{\cdot,w}^{\cdot,f}$ are defined in the supplement.

In Lemma 3, we describe the conditional log-likelihoods of the posterior samples conditioned on the observed geographical coordinates. We use these conditional log-likelihoods to determine the sampler's convergence. Later in the paper, when we present the algorithm for posterior inference, we will use these conditional likelihoods to determine the algorithm's convergence. Let Γ denote the gamma function, let Γ_2 denote the bivariate gamma function, and let $|\cdot|$ denote the determinant.

Lemma 3. *The log-likelihood of the samples $\boldsymbol{z}$ and $\boldsymbol{y}$ conditioned on the observations $\boldsymbol{\ell}$ is*

$$\begin{aligned}\log p\left(\boldsymbol{z}, \boldsymbol{y} \mid \boldsymbol{\ell}\right) = &\left(\sum_{w=1}^{W}\sum_{f=1}^{F}\log\Gamma\left(\beta_{w,f}+m_{\cdot,w}^{\cdot,f}\right)\right)\\ &+\left(\sum_{u=1}^{U}\sum_{f=1}^{F}-\log\Gamma\left(\alpha K_u+m_{u,\cdot}^{\cdot,f}\right)\right)\\ &+\left(\sum_{u=1}^{U}\sum_{f=1}^{F}\sum_{k=1}^{K_u}\log\Gamma\left(\alpha+m_{u,\cdot}^{k,f}\right)\right)\\ &+\left(\sum_{u=1}^{U}\sum_{k=1}^{K_u}\left(\log\Gamma_2\left(\frac{\hat{v}_k^u}{2}\right)-m_{u,\cdot}^{k,\cdot}\log\pi\right)\right)\\ &+\left(\sum_{u=1}^{U}\sum_{k=1}^{K_u}\left(-\frac{\hat{v}_k^u}{2}\log\left|\hat{\boldsymbol{\Lambda}}_u^k\right|-\log\hat{p}_k^u\right)\right)+C,\end{aligned}$$

where C denotes the constant terms. The counts $m_{\cdot,w}^{\cdot,f}$, $m_{u,\cdot}^{\cdot,f}$, $m_{u,\cdot}^{k,\cdot}$, and $m_{u,\cdot}^{k,f}$ are defined in the supplement. The parameters $\hat{v}_k^u$, $\hat{\boldsymbol{\Lambda}}_u^k$, and $\hat{p}_k^u$ are defined in the proof.

In Lemmas 1 and 2, we described a collapsed Gibbs sampler for sampling the posteriors of the categorical random variables. In Lemma 4, we show how these samples, denoted as $\boldsymbol{y}$ and $\boldsymbol{z}$, can be used to approximate the posterior expectations of $\boldsymbol{\gamma}$, $\boldsymbol{\theta}$, $\boldsymbol{\phi}$, and $\boldsymbol{\Sigma}$.

Lemma 4. *The expectations of $\boldsymbol{\gamma}$, $\boldsymbol{\theta}$, $\boldsymbol{\phi}$, and $\boldsymbol{\Sigma}$ given the observed geographical coordinates and the posterior samples are*

$$\hat{\gamma}_{w,f} = \mathbb{E}\left[\gamma_{w,f} \mid \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\ell}\right] = \frac{\beta_{w,f}+m_{\cdot,w}^{\cdot,f}}{\sum_f\left(\beta_{w,f}+m_{\cdot,w}^{\cdot,f}\right)},$$

$$\hat{\theta}_{u,k}^f = \mathbb{E}\left[\theta_{u,k}^f \mid \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\ell}\right] = \frac{\alpha+m_{u,\cdot}^{k,f}}{K_u\alpha+m_{u,\cdot}^{\cdot,f}},$$

$$\hat{\boldsymbol{\phi}}_u^k = \mathbb{E}\left[\boldsymbol{\phi}_u^k \mid \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\ell}\right] = \hat{\boldsymbol{\mu}}_k^u,$$

$$\hat{\boldsymbol{\Sigma}}_u^k = \mathbb{E}\left[\boldsymbol{\Sigma}_u^k \mid \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\ell}\right] = \frac{\hat{\boldsymbol{\Lambda}}_u^k}{\hat{v}_k^u-3}.$$

Counts $m_{\cdot,w}^{\cdot,f}$, $m_{u,\cdot}^{k,f}$, and $m_{u,\cdot}^{\cdot,f}$ are defined in the supplement. Parameters $\hat{\boldsymbol{\mu}}_k^u$, $\hat{\boldsymbol{\Lambda}}_u^k$, and $\hat{v}_k^u$ are defined in the proof.

Finally, we combine our results to construct an algorithm for efficient posterior inference. The algorithm uses Lemmas 1 and 2 to implement the collapsed Gibbs sampler, Lemma 3 to determine convergence, and Lemma 4 to approximate the posterior expectations. The sufficient statistics are defined in the supplement.

Algorithm 1 Collapsed Gibbs sampler for CPM.

Input: Number of factors F, number of places K_u, hyperparameters α, β, ν, p_k, $\boldsymbol{\Lambda}_k$, and $\boldsymbol{\mu}_u$.

Initialize $\boldsymbol{y}$ and $\boldsymbol{z}$ uniformly at random.
Initialize $m_{u,\cdot}^{k,f}$, $m_{u,\cdot}^{\cdot,f}$, $m_{\cdot,w}^{\cdot,f}$, $m_{u,\cdot}^{k,\cdot}$, $\boldsymbol{S}_k^u$, and $\boldsymbol{P}_k^u$ based on $\boldsymbol{y}$ and $\boldsymbol{z}$.
while the log-likelihood in Lemma 3 has not converged **do**
  Choose index $i=(u,w,n)$ uniformly at random.
  $m_{u,\cdot}^{z_i,y_i} \leftarrow m_{u,\cdot}^{z_i,y_i}-1$, $m_{u,\cdot}^{\cdot,y_i} \leftarrow m_{u,\cdot}^{\cdot,y_i}-1$, $m_{\cdot,w}^{\cdot,y_i} \leftarrow m_{\cdot,w}^{\cdot,y_i}-1$.
  Sample y_i with respect to Lemma 2.
  $m_{u,\cdot}^{z_i,y_i} \leftarrow m_{u,\cdot}^{z_i,y_i}+1$, $m_{u,\cdot}^{\cdot,y_i} \leftarrow m_{u,\cdot}^{\cdot,y_i}+1$, $m_{\cdot,w}^{\cdot,y_i} \leftarrow m_{\cdot,w}^{\cdot,y_i}+1$.
  Choose index $i=(u,w,n)$ uniformly at random.
  $m_{u,\cdot}^{z_i,y_i} \leftarrow m_{u,\cdot}^{z_i,y_i}-1$, $m_{u,\cdot}^{z_i,\cdot} \leftarrow m_{u,\cdot}^{z_i,\cdot}-1$, $\boldsymbol{S}_{z_i}^u \leftarrow \boldsymbol{S}_{z_i}^u-\boldsymbol{\ell}_i$, $\boldsymbol{P}_{z_i}^u \leftarrow \boldsymbol{P}_{z_i}^u-\boldsymbol{\ell}_i\boldsymbol{\ell}_i^T$.
  Sample z_i with respect to Lemma 1.
  $m_{u,\cdot}^{z_i,y_i} \leftarrow m_{u,\cdot}^{z_i,y_i}+1$, $m_{u,\cdot}^{z_i,\cdot} \leftarrow m_{u,\cdot}^{z_i,\cdot}+1$, $\boldsymbol{S}_{z_i}^u \leftarrow \boldsymbol{S}_{z_i}^u+\boldsymbol{\ell}_i$, $\boldsymbol{P}_{z_i}^u \leftarrow \boldsymbol{P}_{z_i}^u+\boldsymbol{\ell}_i\boldsymbol{\ell}_i^T$.
Compute $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\phi}}$, and $\hat{\boldsymbol{\Sigma}}$ using Lemma 4.
return $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\phi}}$, and $\hat{\boldsymbol{\Sigma}}$.

4 Experiments

In this section, we demonstrate the performance of CPM on two real-world location datasets, a dense cellular carrier dataset and a sparse mobile ad exchange dataset, both of which are depicted in Figure 1.

We start by describing how we set up the experiments. Each data point consists of a user ID, a local time, and geographic coordinates represented as latitudes and longitudes. First, we check if a user has logged multiple observations during the same hour, and if so, we replace these observations with their geometric median, computed using Weiszfeld's algorithm. Since geometric median is a robust estimator, this step removes both noisy and redundant observations. Then, we sort the datasets chronologically, and for each user, split the user's data points into 3 partitions: earliest 60% is added to the training data, middle 20% to the validation data, and final 20% to the test data. After preprocessing, both datasets contain approximately 2 million data points, the dense dataset contains 1394 users, and the sparse dataset contains 19247 users. Furthermore, since the temporal gap between the training data and the test data ends up being at least a week apart, our datasets become inappropriate for models that make near-term forecasts based on the most recent observations.

We train our model in two stages. In the first stage, we determine both the optimal number of places for each user (i.e. K_u) and the optimal setting for each spatial hyperparameter (i.e. ν, p_k, $\boldsymbol{\Lambda}_k$, and $\boldsymbol{\mu}_u$). To do so, we extract the spatial part of CPM into a GMM and train a different GMM for all

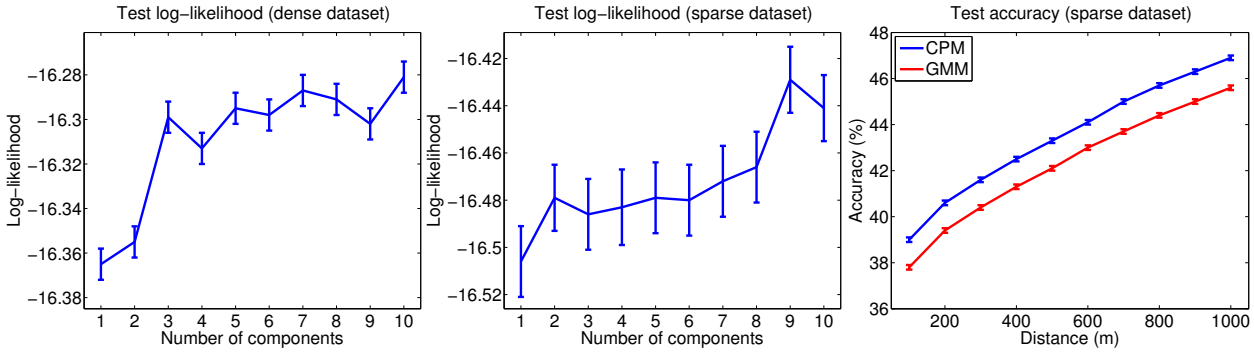

Figure 3: Comparison of CPM and GMM on held-out data. The left and middle plots represent the held-out log-likelihoods on dense and sparse datasets, respectively. In these plots, GMM is represented as a CPM with a single factor. The right plot shows the percentage of predictions that have distance errors less than a given threshold. In general, CPM performs better than GMM.

combinations of users, number of places, and spatial hyperparameter settings. Then, for each hyperparameter setting, we iterate over all users, and assign each user to a number of place that minimizes the user's validation log-likelihood. The final spatial hyperparameter setting we output is the one that minimizes the total validation log-likelihood summed across all users.

In the second stage, we use the assignments obtained in the previous stage, train a different CPM for each non-spatial hyperparameter setting (i.e. F, α, β), and again use the validation log-likelihood to select the best model. We set the hyperparameters by choosing the number of factors from $F \in \{1, \ldots, 10\}$ and the Dirichlet parameters from $\alpha, \beta \in \{0.01, 0.1, 1, 10\}$.

We evaluate how well our model fits the data by measuring the held-out test log-likelihood. The results for GMM and CPM are displayed in Figure 3. Since GMM is equivalent to a CPM with a single factor, the log-likelihood plots compare the held-out performance of the two models and demonstrate the advantages of temporal factorization. As the number of factors increases, the held-out log-likelihood also increases, illustrating how CPM can represent increasingly complex temporal patterns without overfitting. Once we convert log-likelihood to likelihood, we observe that CPM improves performance by 8% for both sparse and dense datasets.

An alternative way to evaluate model performance is to measure the distance between predicted coordinates and observed coordinates. Let $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\phi}}$, and $\hat{\boldsymbol{\Sigma}}$ be the posterior expectations computed in Section 3. Both GMM and CPM can use these posteriors to estimate a maximum a posteriori (MAP) location for user u at weekhour w. In Figure 3, we use these location estimates to plot the percentage of predictions that have distance errors less than a given threshold. Again, CPM outperforms GMM, and the difference in accuracy grows as the distance threshold increases. Note that compared to evaluations based on log-likelihoods, evaluations based on point estimates do not capture the full extent of information encoded by our models, since they ignore the probabilities assigned to places other than the highest weighted ones.

The spatiotemporal patterns extracted by CPM not only fit the data well, but also provide qualitative insights about location data. In Figure 4, we plot $\hat{\boldsymbol{\gamma}}$, which represents the time-dependent posterior weights associated with the factorized place distributions. Intuitively, we would expect these weights to capture the latent temporal patterns that are shared across all users. CPM has no prior notion of how weekhours are related to one another, nor does it have any explicit sequential constraints; nevertheless, it is able to extract these temporal patterns from raw location data alone. Furthermore, even though it looks like we haven't encoded any temporal dependencies in our model, an application of Bayes ball algorithm reveals that $\boldsymbol{\gamma}_w$ are actually dependent on one another conditioned on the observed data. In fact, we observe this temporal dependency in Figure 4.

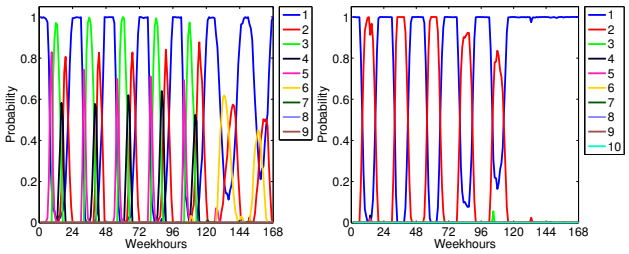

Figure 4: Posterior $\hat{\boldsymbol{\gamma}}$ inferred by CPM. These are distributions over factors and they capture the latent temporal patterns that are shared across all users. The weights learned from the dense dataset are on the left and from the sparse dataset are on the right.

The left plot in Figure 4 corresponds to the dense dataset, and it roughly clusters days into four groups: monday through thursday, friday, saturday, and sunday. Factor 1 approximates time spent at home, which is the dominant factor except between 8am and 8pm monday through thursday, between 8am and 11pm friday, between noon and 10pm saturday, and between 1pm and 8pm sunday. The weekday hours between 8am and 6pm, which typically correspond to work hours, are partitioned into three factors: factor 5 roughly corresponds to a morning commute, factor 4 to an evening commute, and factor 3 to the work hours in between. Leisure time after work is approximately represented by factor 2, and other than home, it is the only factor which dominates a time segment during each day: between 5pm and 8pm monday through thursday,

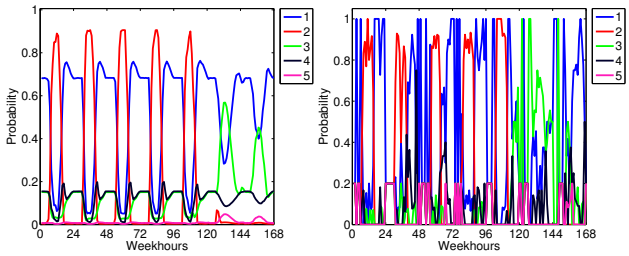

Figure 5: Posterior and empirical place distributions for the dense dataset. The left plot is the place distribution inferred by CPM and the right plot is the empirical place distribution computed using a GMM. The CPM distribution is much more smooth than the empirical distribution computed using GMM, making it easier to visually inspect the user's temporal patterns. Note that, in contrast to Figure 4, these are distributions over a user's place clusters, not factors.

between 6pm and 11pm on friday, between 5pm and 10pm on saturday, and between 4pm and 8pm on sunday. Factor 6 only appears during weekends and dominates couple hours in the early afternoon. All of these temporal patterns correspond to intuitive temporal clusters and they are extracted from all users in an unsupervised fashion.

The right plot in Figure 4 corresponds to the sparse dataset, and due to scarcity of information per user, CPM chooses a coarser temporal pattern: a split between work and non-work hours with smooth transitions between them. As we will demonstrate later, this temporal factorization allows us to make predictions for weekhours where we have not observed any data.

These temporal patterns do not necessarily mean that all users are constrained to the same routine. In fact, one of the key benefits of our model compared to the earlier prototype in Figure 2 is that it allows each user a certain level of flexibility, by allowing everyone to have their own distinct factorized place distributions. In Figure 5, we show the final place distribution of an arbitrarily chosen user from the dense dataset, computed by *combining* the global factor weights in the left plot of Figure 4 *with* user's own factorized place distributions. In Figure 5, the left plot is the place distribution inferred by CPM and the right plot is the empirical place distribution computed using a GMM. The difference in the temporal patterns between left plots of Figures 4 and 5 demonstrates how CPM successfully customizes global patterns to each user. Furthermore, the distribution inferred by CPM is much smoother than the empirical distribution computed using GMM, making it easier to visually inspect the user's temporal patterns.

In Figure 6, we continue analyzing the same user. The left plot shows some of the places inferred by CPM. The circles represent the training data, the colors represent the place assignments, and the pins represent the place cluster means. A visual inspection of the temporal patterns in Figure 5 reveals that Place 1 is user's home. The right plot shows the empirical and inferred probability distributions over the weekhours, assuming that the user is already at Place 1. The empirical distribution is very noisy, and due to sampling bias, it shows that user is not at home between midnight and 8am! The distribution inferred by CPM corrects that bias.

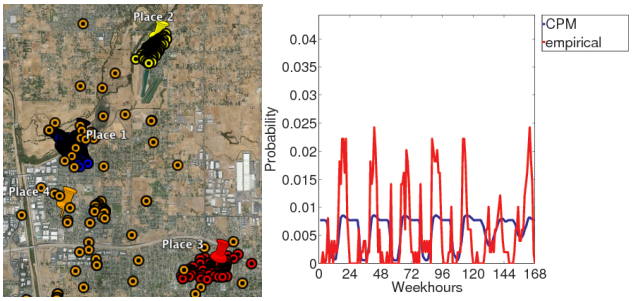

Figure 6: Left plot shows some of the places inferred by CPM. The circles represent the training data, the colors represent the place assignments, and the pins represent the place cluster means. Right plot shows a comparison of the empirical and inferred probability distributions over the weekhours given that the user is already at Place 1 (home). Note that, in contrast to Figure 5, this is a distribution over weekhours.

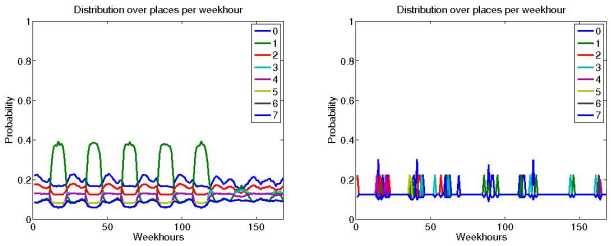

Figure 7: Posterior and empirical place distributions for the sparse dataset. The left plot is the place distribution inferred by CPM and the right plot is the empirical place distribution computed using a GMM. The GMM distribution is sparse whereas the CPM distribution is both complete and clearly delineates home and work.

Lastly, in Figure 7, we show the final place distribution of an arbitrarily chosen user from the sparse dataset, computed by *combining* the global factor weights in the right plot of Figure 4 *with* user's own factorized place distributions. In Figure 7, the left plot is the place distribution inferred by CPM and the right plot is the empirical place distribution computed using a GMM. In contrast to the sparse distribution computed using GMM, CPM clearly delineates the temporal cycles associated with home and work and is able to make predictions for all weekhours.

5 Conclusion

In this paper, we present the Collaborative Place Model, which recovers the latent spatiotemporal structure underlying unsupervised location data by analyzing patterns shared across all users. CPM combines population-wide inferences with user-specific ones, allows users to have different number of places or weekhours to have different distributions over places, and constructs a realistic and flexible representation of users' location profiles. We apply CPM to two real datasets, one sparse and one dense, and demonstrate how it both improves location prediction performance and provides new insights into users' spatiotemporal patterns.

For future work, we are interested in generalizing our

model using the hierarchical Dirichlet process, such that training is simplified and the number of places for each user is learned more efficiently.

References

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

[Cho *et al.*, 2011] Eunjoon Cho, Seth A. Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pages 1082–1090, New York, NY, USA, 2011. ACM.

[Davie, 2012] William R. Davie. *Telephony*, pages 251–264. Focal Press, 2012.

[De Domenico *et al.*, 2012] Manlio De Domenico, Antonio Lima, and Mirco Musolesi. Interdependence and predictability of human mobility and social interactions. In *Proceedings of the Nokia Mobile Data Challenge Workshop in conjunction with International Conference on Pervasive Computing*, 2012.

[Eagle and Pentland, 2009] Nathan Eagle and Alex S. Pentland. Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066, May 2009.

[Eubank *et al.*, 2004] Stephen Eubank, Hasan Guclu, V. S. Anil Kumar, Madhav V. Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, May 2004.

[Farrahi and Perez, 2011] Katayoun Farrahi and Daniel G. Perez. Discovering routines from large-scale human locations using probabilistic topic models. *ACM Trans. Intell. Syst. Technol.*, 2(1), January 2011.

[Gao *et al.*, 2012] Huiji Gao, Jiliang Tang, and Huan Liu. Mobile location prediction in spatio-temporal context. In *Proceedings of the Nokia Mobile Data Challenge Workshop in conjunction with International Conference on Pervasive Computing*, 2012.

[Gonzalez *et al.*, 2008] Marta C. Gonzalez, Cesar A. Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, June 2008.

[Hao *et al.*, 2010] Qiang Hao, Rui Cai, Changhu Wang, Rong Xiao, Jiang M. Yang, Yanwei Pang, and Lei Zhang. Equip tourists with knowledge mined from travelogues. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 401–410, New York, NY, USA, 2010. ACM.

[Liao *et al.*, 2005] Lin Liao, Dieter Fox, and Henry Kautz. Location-based activity recognition. In *In Advances in Neural Information Processing Systems (NIPS*, pages 787–794, 2005.

[Liao *et al.*, 2007] Lin Liao, Dieter Fox, and Henry Kautz. Extracting places and activities from GPS traces using hierarchical conditional random fields. *Int. J. Rob. Res.*, 26(1):119–134, January 2007.

[Oshin *et al.*, 2012] T. O. Oshin, S. Poslad, and A. Ma. Improving the Energy-Efficiency of GPS based location sensing smartphone applications. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on*, pages 1698–1705. IEEE, June 2012.

[Song *et al.*, 2010] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, February 2010.

[Wernke *et al.*, 2012] Marius Wernke, Pavel Skvortsov, Frank Dürr, and Kurt Rothermel. A classification of location privacy attacks and approaches. pages 1–13, 2012.

[Zheng *et al.*, 2009] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei Y. Ma. Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 791–800, New York, NY, USA, 2009. ACM.

[Zheng *et al.*, 2010] Vincent W. Zheng, Yu Zheng, Xing Xie, and Qiang Yang. Collaborative location and activity recommendations with GPS history data. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 1029–1038, New York, NY, USA, 2010. ACM.