Collaborative Place Models

Berk Kapicioglu — *Foursquare Labs*
David S. Rosenberg — *YP Mobile Labs*
Robert E. Schapire — *Microsoft Research*
Tony Jebara — *Columbia University*

General Problem

Goals:

- Given **sparse** location data from a user, can we still predict where the user is going to be at any arbitrary time in the future?

Applications:

- Automated traffic alerts.
- Contextual venue recommendations.
- Location-based advertisement.

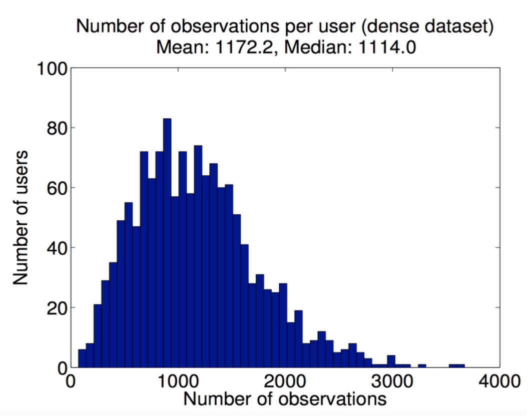
Sparsity in Location Data

Sparsity: Very few observations per user.

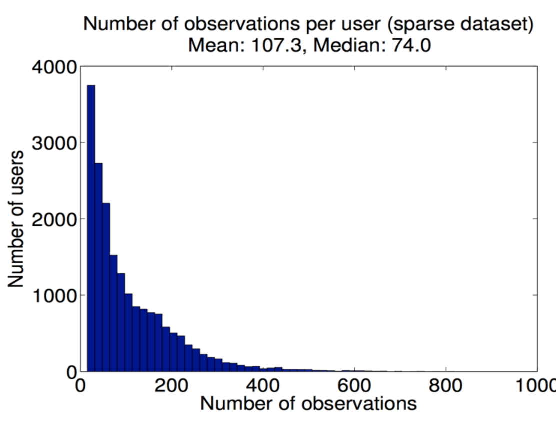
Causes:

- Battery constraints.
- User privacy.

Dense Dataset



Sparse Dataset



Remarks:

- Sparse dataset has 15x less observations per user than dense dataset!
- Harder to infer spatiotemporal patterns under sparsity.
- Sparsity is becoming increasingly common among location datasets.

I/O Specification

Input

User ID	Local Time	Latitude	Longitude
...	...	...	...
User ID	Local Time	Latitude	Longitude

Query: User ID, Local Time → Model → **Output**: Distribution over latitudes and longitudes

Collaborative Place Model

- A generalization of Gaussian Mixture Model (GMM).
- Like GMM, it learns the latent place clusters for each user (i.e. "home", "work", "gym", etc.).
- Unlike GMM, it also learns the latent temporal patterns shared across the users.

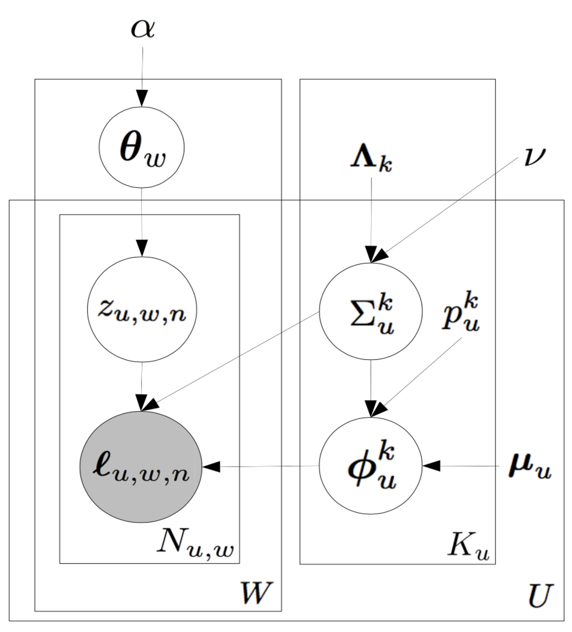
Continued…

- Thus, even if we observe a user very few times, we can complete the user's missing data by relying on population-wide temporal patterns.

Temporal Assumptions

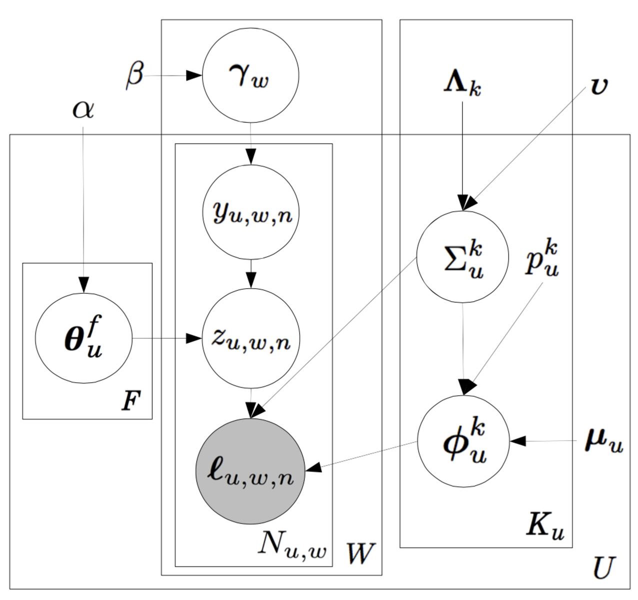
- We can encode the temporal patterns using either the "Strong Similarity" or "Weak Similarity" assumptions.
- *Strong similarity* assumes that, at a given weekhour, all users share the exact same place distributions.
- "Probability that all users are at home Sunday 9am is 98%."

$$
\begin{aligned}
\boldsymbol{\theta}_w &\sim \mathrm{Dirichlet}(\alpha)\\
z_{u,w,n} \mid \boldsymbol{\theta}_w &\sim \mathrm{Multinomial}(\boldsymbol{\theta}_w)\\
\Sigma_u^k \mid \boldsymbol{\Lambda}_k, \nu_0 &\sim IW_{v_0}(\boldsymbol{\Lambda}_k)\\
\boldsymbol{\phi}_u^k \mid \boldsymbol{\mu}_u, \Sigma_u^k, p_k &\sim \mathcal{N}\left(\boldsymbol{\mu}_u, \frac{\Sigma_u^k}{p_k}\right)\\
\vec{\ell}_{u,w,n} \mid z_{u,w,n} = k &\sim \mathcal{N}\left(\boldsymbol{\phi}_u^k, \Sigma_u^k\right)
\end{aligned}
$$



- Unfortunately, under this assumption, users are forced to have
  - The same number of places,
  - The same distribution over these places.
- *Weak similarity* assumes that
  - Each user has her own factorized place distribution,
  - Each weekhour has its own factor weights shared across all users.
  - A user's place distribution is a convex combination of user-specific factor distributions and weekhour-specific factor weights.

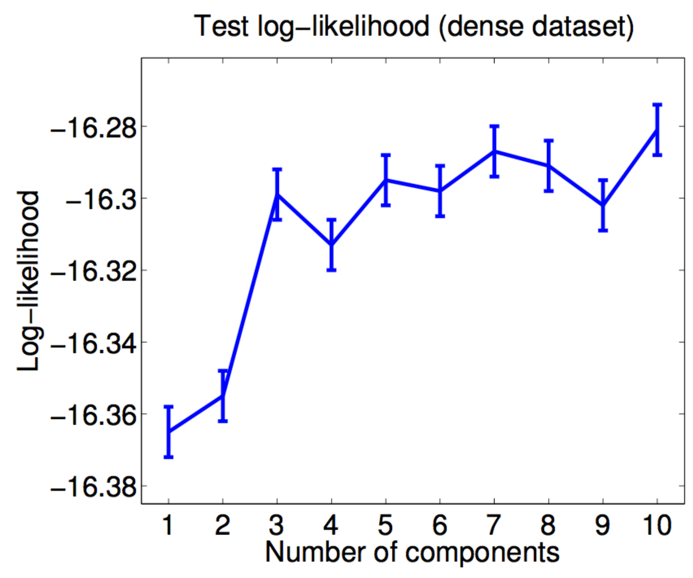
$$
\begin{aligned}
\boldsymbol{\theta}_{u,w} &\sim \mathrm{Dirichlet}(\alpha)\\
\boldsymbol{\gamma}_w &\sim \mathrm{Dirichlet}(\beta_w)\\
y_{u,w,n} \mid \boldsymbol{\gamma}_w &\sim \mathrm{Multinomial}(\boldsymbol{\gamma}_w)\\
z_{u,w,n} \mid y_{u,w,n} = t &\sim \mathrm{Multinomial}(\boldsymbol{\theta}_{u,t})\\
\Sigma_u^k \mid \boldsymbol{\Lambda}_k, \nu_0 &\sim IW_{v_0}(\boldsymbol{\Lambda}_k)\\
\boldsymbol{\phi}_u^k \mid \boldsymbol{\mu}_u, \Sigma_u^k, p_k &\sim \mathcal{N}\left(\boldsymbol{\mu}_u, \frac{\Sigma_u^k}{p_k}\right)\\
\vec{\ell}_{u,w,n} \mid z_{u,w,n} = k &\sim \mathcal{N}\left(\boldsymbol{\phi}_u^k, \Sigma_u^k\right)
\end{aligned}
$$

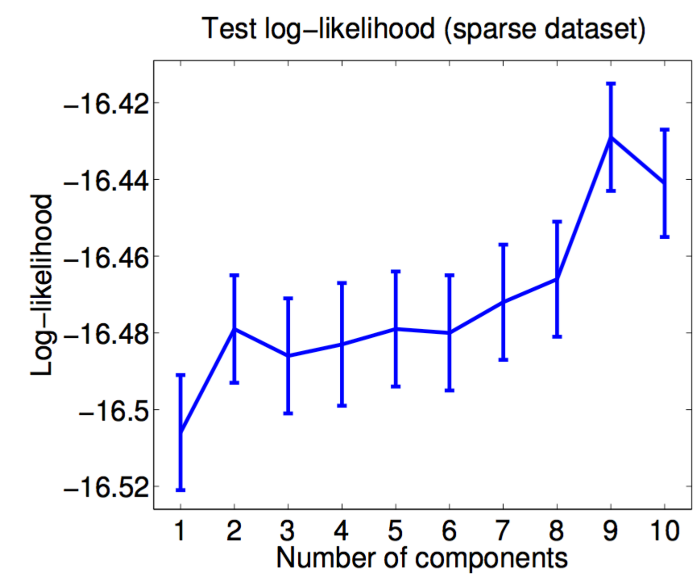


- *Weak similarity* assumption is **better** because
  - Unlike *strong similarity*, users can have different number of places and different distribution over these places,
  - Like *strong similarity*, we can make predictions for weekhours even if the user has not been observed during that hour before.

Experiments

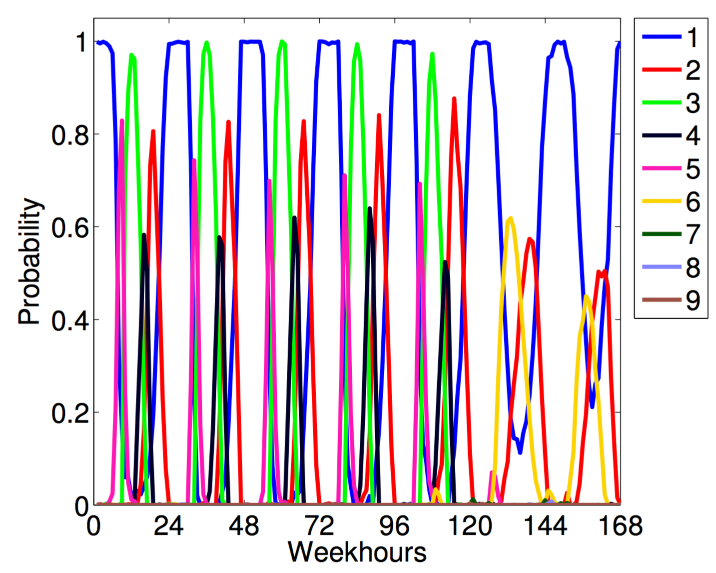
- For inference, we derive and implement a collapsed Gibbs sampler.
- We run experiments on both sparse and dense datasets.
- First, we compare the held-out log-likelihoods of CPM vs. GMM:

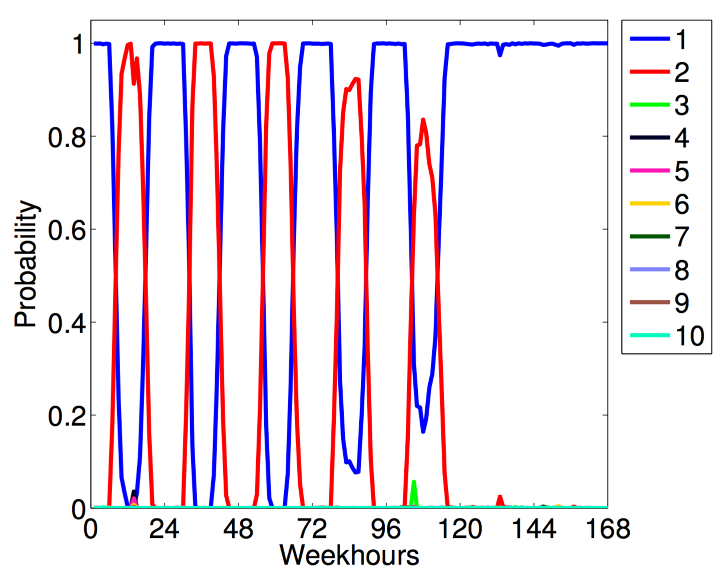




- The model with only one factor is GMM.
- As number of factors increase, CPM predicts better.
- With respect to likelihood, CPM outperforms GMM by 8%.

- Second, we show the weekhour-specific factor weights (i.e. $\hat{\gamma}$) that are shared across all users.
- Dense dataset is at top and sparse dataset is at bottom.





- For dense dataset, Factor 1 is time at home, Factor 5 a morning commute, Factor 4 an evening commute, Factor 3 workhours in between, and Factor 2 leisure time after work.

- For sparse dataset, factors are home and work.
- Lastly, we compare the inferred and empirical place distributions of a user from the sparse dataset.
- The inferred distribution (left) is much smoother than the empirical distribution (right).