Learning Geographical Preferences for Point-of-Interest Recommendation

Bin Liu

Yanjie Fu, Zijun Yao, Hui Xiong

Rutgers University

Outline

- Background and Motivation
- Geographical Probabilistic Factor Model
- Experimental Results
- Conclusion

Location based Services

- Location-based service becomes increasingly popular
- Develop rapidly, e.g., as 2011, Foursquare (**15 million** users) made over **3 million** check-ins per day; as Jan 2013, over 30 million people
- Users share check-in experiences, opinions, comments on a point-of-interest (a specific point location that someone may find useful or interesting, eg., restaurant, bar)

Point-of-Interest Recommendation

RUTGERS

- **Task**: to recommend POIs based users' check-in history and community opinions
- Existing solutions
  - Method: collaborative filtering based method to fuse information
  - Failed to consider the multiple factors in decision process of a user choose a POI ; lack of integrated analysis of the joint effect of the factors if considered part of them
- Various factors can influence POI check-in: user preferences, geographical influences, popularity and dynamic user mobility patterns

Check in histories

Recommender models

Similar user, similar items

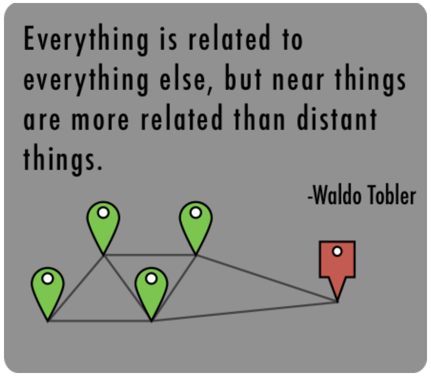
?!

Just arrive at a new place ???

What Make POI Recommendation Special

RUTGERS

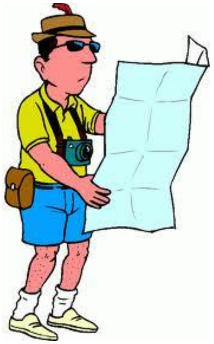
- **Law of geography**: everything is related to everything else, but near things are more related than distant thing
- **Regional popularity**: two POIs with similar semantic topics can have different popularity if located differently
- **Dynamic user mobility**: user may travel to different cities or even regions
- **Implicit user feedback**: need to infer user preferences from implicit user feedback in terms of user check-in count data



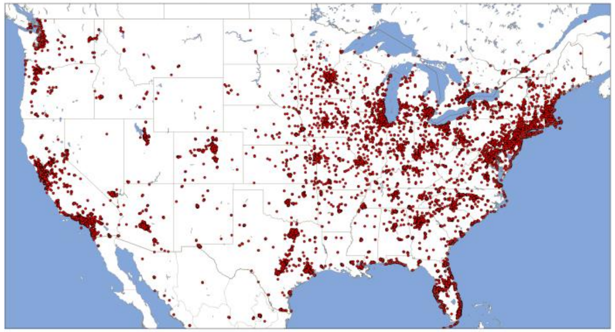
How Decision Process be Influenced

RUTGERS

- Geographical distance, the propensity of a user choose a POI is inversely proportional to the distance
- Utility matters, a user may prefer a far away POI than a nearby one for better satisfaction
- Popularity affects check-in behaviors, decision is largely affected by the word-of-mouth about the POI
- Dynamic mobility patterns: the check-in pattern may vary when people travel from one region to another

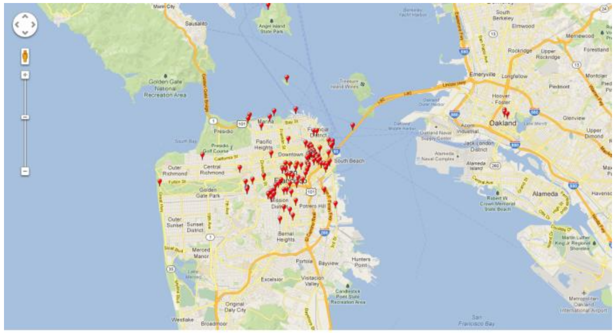


How Is a Typical User's Check-in Pattern



All POIs in different regions

A user's check-ins in different regions



User check-ins in San Francisco

- Need a model that jointly encodes the **personalized preferences**, **spatial influence**, **user mobility** and **popularity** into the user check-in decision process to learn geographical user preferences for effective POI recommendation

General Idea

RUTGERS

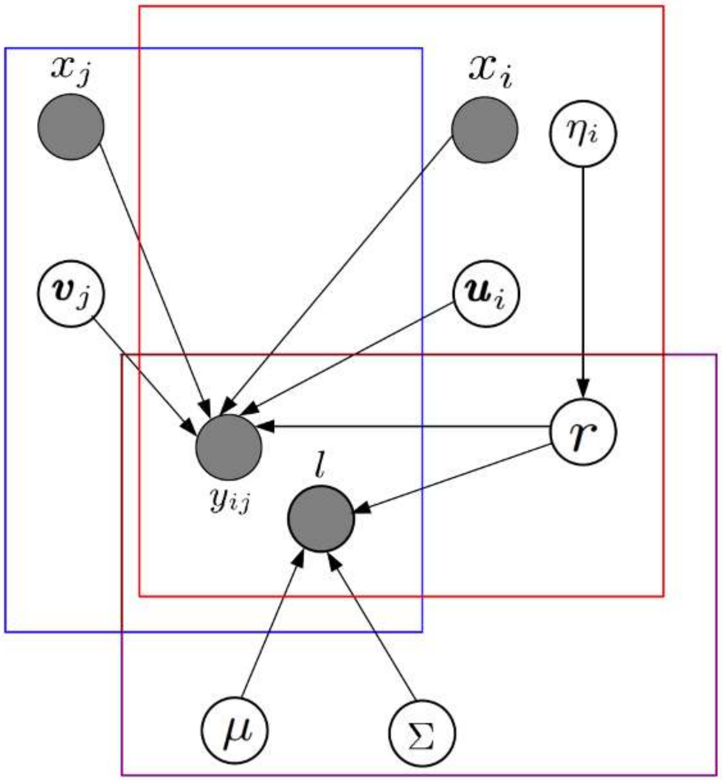
- By Tobler's first law of geography, POIs with similar services are likely to be clustered into the same geographical area $\boxed{l_j \sim \mathcal{N}(\mu_r, \Sigma_r)}$
- Users are most likely to check in a number of POIs and these POIs are usually limited to some geographical regions $\boxed{r \sim \text{Multinomial}(\eta_i)}$
- A user's propensity for a POI
  - Personalized interest
  - Regional popularity
  - Distance

$$p(i,j) \propto \alpha(i,j)\rho(j)(d_0 + d(i,j))^{-\tau}$$

Best personalization, maximum satisfaction, at lowest distance cost

Geographical Probabilistic Factor Model

RUTGERS



1. Draw a region $r \sim \text{Multinomial}(\eta_i)$. Mobility
2. Draw a location $l_j \sim \mathcal{N}(\mu_r, \Sigma_r)$. Law of geography
3. Draw a user preference Personalized Preference
   - a Generate user latent factor $\boldsymbol{u}_i \sim P(\boldsymbol{u}_i; \Psi_{\boldsymbol{u}_i})$.
   - b Generate item latent factor $\boldsymbol{v}_j \sim P(\boldsymbol{v}_j; \Psi_{\boldsymbol{v}_j})$.
   - c User-item preference $\alpha(i,j) = \boldsymbol{u}_i^\top \boldsymbol{v}_j + x_i^\top W x_j$.
4. $y_{ij} \sim P(f_{ij})$ where Check-in decision

$$f_{ij} = \left(\boldsymbol{u}_i^\top \boldsymbol{v}_j + x_i^\top W x_j\right) \cdot \rho(j) \cdot (d_0 + d(i,j))^{-\tau}$$

Table 1: Mathematical Notations

Symbol	Size	Description
R	$1 \times \mathbb{R}$	latent region set, r is a region in R
$\boldsymbol{\eta}$	$M \times \mathbb{R}$	user level region distribution
$\boldsymbol{\rho}$	$1 \times N$	item popularity
$\boldsymbol{\mu}$	$\mathbb{R}^2$	mean location of a latent region
$\boldsymbol{\Sigma}$	$\mathbb{R}^{2\times 2}$	covariance matrix of a latent region
$\boldsymbol{U}$	$M \times K$	user latent factor
$\boldsymbol{V}$	$N \times K$	item latent factor
$\boldsymbol{x}$	$(\cdot) \times \mathbb{K}$	user or item observable prosperities
$\boldsymbol{\theta}, \boldsymbol{\pi}$	$(\cdot) \times \mathbb{K}$	user or item topic distribution

Model Specification

- User mobility model
  - A user samples a region from all R regions following a multinomial distribution $r \sim \text{Multinomial}(\eta_i)$
  - An POI is assigned a normal distribution $l_j \sim \mathcal{N}(\mu_r, \Sigma_r)$
- Distance factor
  - Distance from region center to the POI $d(i,j) = ||\mu_r - l_j||_2$
  - the prob. a user choose a POI decays as the power-law of the distance between them $(d_0 + d(i,j))^{-\tau}$
- Regional popularity
  - Given a region

$$\rho_j = \frac{1}{2}\left\{\frac{\text{totalPeo}_j - 1}{\max_{j\in r}\{\text{totalPeo}_j\} - 1} + \frac{\text{totalCk}_j - 1}{\max_{j\in r}\{\text{totalCk}_j\} - 1}\right\}$$

Geographical-Topical Bayesian Non-negative Matrix Factorization

- Latent factor model: clod-start problem; normal assumption
- Poisson factor model for count data, and closely related to NMF
- GT-BNMF model: (1) encode the personalized preferences, spatial influence and popularity; (2) count data; (3) cold-start

$$y_{ij} \sim \mathcal{N}^R(y_{ij}|f_{ij}, \sigma^2), f_{ij} = \left(\boldsymbol{u}_i^\top \boldsymbol{v}_j + \theta_i^\top W \pi_j\right) \cdot \rho(j) \cdot (d_0 + d(i,j))^{-\tau}$$

i. Generate user latent factor $u_{ik} \sim \mathrm{Exp}(\alpha_k)$.

ii. Generate item latent factor $v_{jk} \sim \mathrm{Exp}(\beta_k)$.

iii. User-item preference $\alpha(i,j) = \boldsymbol{u}_i^\top \boldsymbol{v}_j + \theta_i^\top W \pi_j$.

iv. $\sigma^2 \sim \mathrm{Inv} - \mathrm{Gamma}(a, b)$.

Parameter Estimation

RUTGERS

- Given $\mathcal{D} = \{y_{ij}, l_j\}^{I_{ij}}$ where y_{ij} is the user check-in count and l_j is the location
- To maximize posterior $P(\Psi, \alpha, \beta; \mathcal{D}) = \prod_{\mathcal{D}} P(y_{ij}, l_j | \Psi, \Omega)$

parameter $\Psi = \{\boldsymbol{U}, \boldsymbol{V}, \sigma^2, W, \boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$, priors $\Omega = \{\alpha, \beta, a, b\}$

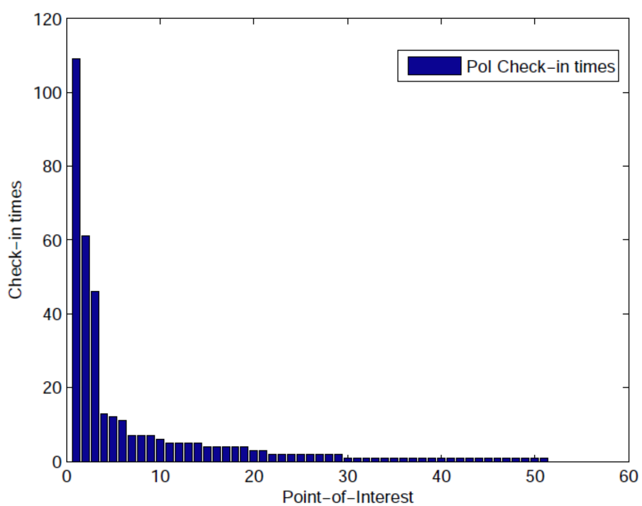
- Mixing Expectation Maximization (EM) and sampling algorithm to learn all the parameters by treating latent region r as a latent variable and introduce the hidden variable $P(r|l_j, \Psi)$
  - Geo-clustering updates the latent regions base on both location and check-in behaviors
  - GT-BNMF learns the graphical preference factors

Experimental Data

RUTGERS

- 12,422 users for 46, 194 POIs with 738,445 check-in observations from Foursquare with sparsity of 99.87%
- Wide range user check-in count data

Name: Otto Enoteca Pizzeria
Address: 1 5th Ave, New York, NY 10003
Tags: pizza wine bar italian olive oil cheese mario batali meat wine pasta gelato gluten free menu zagat rated pizza
Total people: 3,127, Total check-ins: 4,770.



foursquare

Evaluation Method

RUTGERS

- Baselines: SVD, PMF, NMF, BNMF, F-BNMF
- Metrics:
  - Recall and precision

$$\text{Precision@}N = \frac{|S_{N,\text{rec}} \bigcap S_{\text{visited}}|}{N}$$

$$\text{Recall@}N = \frac{|S_{N,\text{rec}} \bigcap S_{\text{visited}}|}{|S_{\text{visited}}|}$$

  - Relative recall and relative precision, measure the improvement over a random recommendation

$$\text{rPrecision@}N = \frac{\text{Precision@}N}{|S_{\text{visited}}|/|C|} = \frac{|S_{N,\text{rec}} \bigcap S_{\text{visited}}| \cdot |C|}{|S_{\text{visited}}| \cdot N}$$

$$\text{rRecall@}N = \frac{\text{Recall@}N}{N/|C|} = \frac{|S_{N,\text{rec}} \bigcap S_{\text{visited}}| \cdot |C|}{|S_{\text{visited}}| \cdot N}$$

  - Initialize the algorithm with K-means

Precision and Recall

K	Pre	SVD	PMF	NMF	BNMF	F-BNMF	GT-BNMF
10	@1	0.0041	0.0034	0.0125	0.0181	0.0192	**0.0347**
	@5	0.0066	0.0062	0.0169	0.0197	0.0208	**0.0288**
	@10	0.0081	0.0080	0.0202	0.0224	0.0237	**0.0306**
20	@1	0.0052	0.0029	0.0126	0.0147	0.0166	**0.0326**
	@5	0.0067	0.0059	0.0163	0.0160	0.0177	**0.0278**
	@10	0.0088	0.0079	0.0202	0.0197	0.0210	**0.0304**

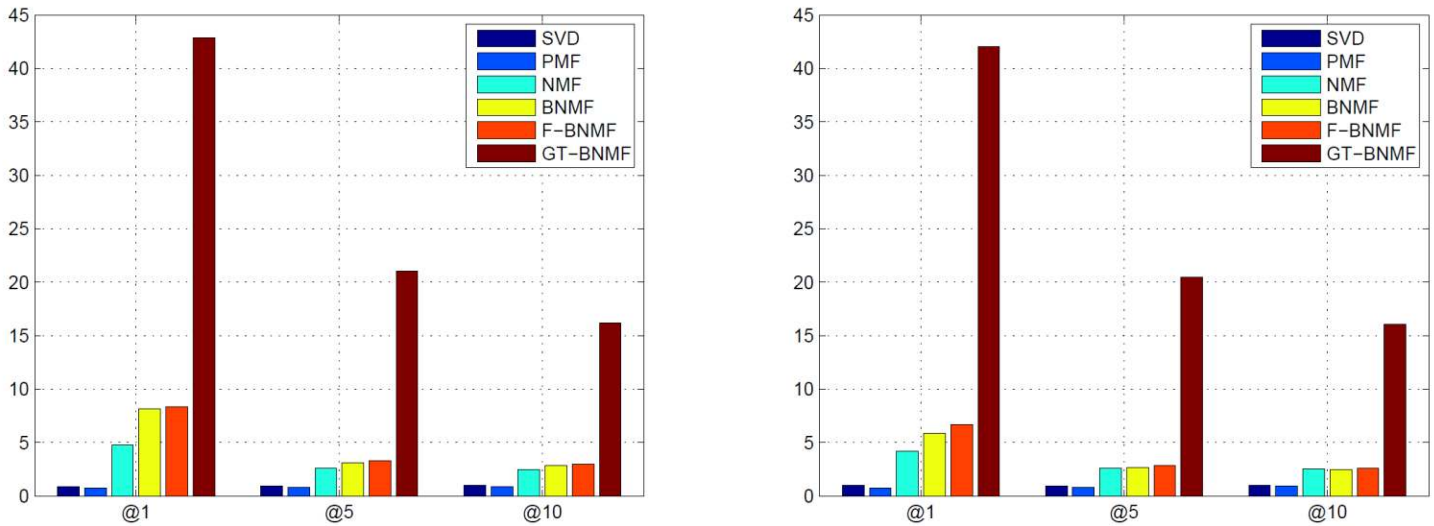
Table 3: Precision @N with different latent dimensions K.

K	Recall	SVD	PMF	NMF	BNMF	F-BNMF	GT-BNMF
10	@1	0.0008	0.0008	0.0049	0.0077	0.0081	**0.0061**
	@5	0.0038	0.0036	0.0103	0.0121	0.0127	**0.0147**
	@10	0.0060	0.0060	0.0153	0.0167	0.0176	**0.0212**
20	@1	0.0011	0.0006	0.0046	0.0059	0.0068	**0.0060**
	@5	0.0037	0.0034	0.0098	0.0097	0.0107	**0.0144**
	@10	0.0065	0.0059	0.0151	0.0148	0.0158	**0.0210**

Table 4: Recall @N with two different latent dimensions K.

Relative Performances

The relative performance @N measures the improvement over a random recommendation

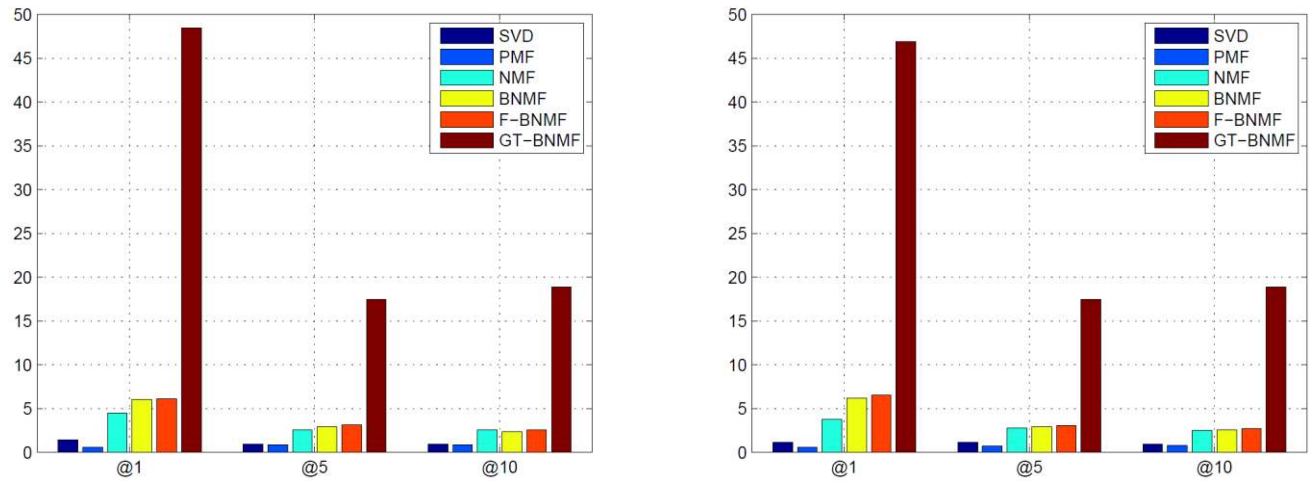


(a) $K = 10$.　(b) $K = 20$.

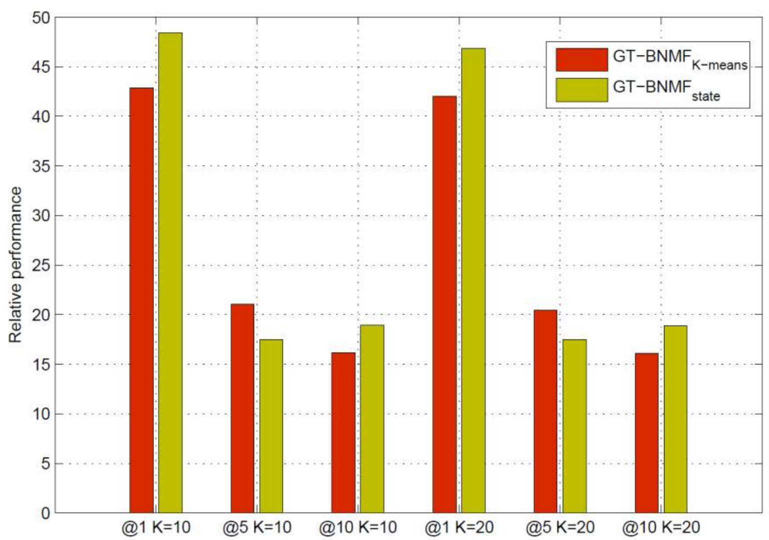
K	@N	SVD	PMF	NMF	BNMF	F-BNMF	GT-BNMF
10	@1	0.8729	0.7280	4.7736	8.1243	8.3408	**42.8835**
	@5	0.9345	0.8251	2.5871	3.0713	3.2956	**21.0357**
	@10	0.9829	0.8856	2.4480	2.8588	2.9585	**16.1661**
20	@1	1.0148	0.7124	4.1618	5.8585	6.6864	**42.0183**
	@5	0.9287	0.8271	2.6095	2.6219	2.8506	**20.4742**
	@10	1.0084	0.9067	2.5131	2.4717	2.6043	**16.0662**

Implications of Latent Regions

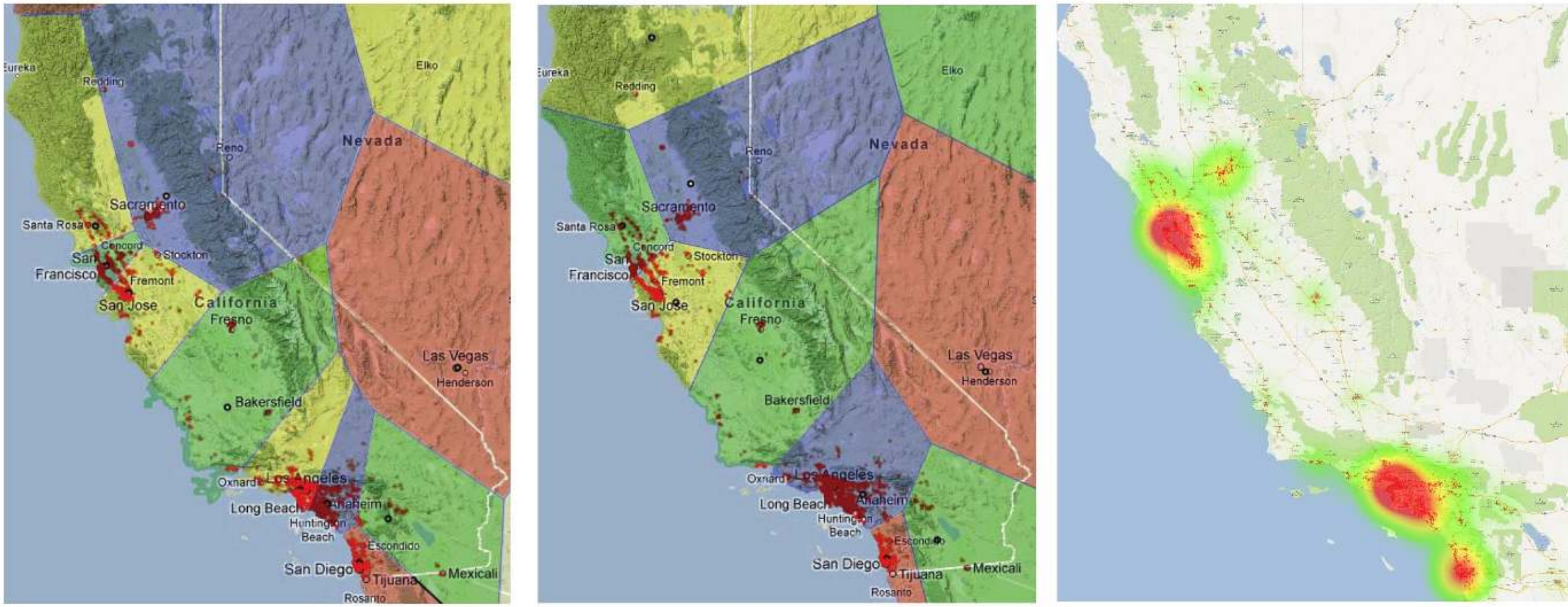
- How about initialize the algorithm by states



- Robust to region initiations



Latent Region Analysis



(a) K-means. (b) Latent region. (c) Ground truth.

- Voronoi visualization of POI segmentation in California area. (b) latent regions learned from our model and (a) initiation by K-means. (c) true user collaborative activity clusters. Deeper color (red) indicates more check-ins for a POI, as contrary to light color (green).
- Latent regions learned from our model is more coherent to real user activity

Conclusion

- Proposed a general framework to learn geographical preferences for POI recommendation
  - Captured the geographical influence on a user's check-in behavior by taking into consideration of geographical factors
  - Effectively modeled the user mobility patterns
  - Extended the latent factor in explicit rating recommendation to implicit feedback recommendation settings
  - Proposed model is flexible and could be extended to incorporate different later factor models
- The proposed model not only improves recommendation performances, but also provides an interesting perspective on POI segmentation

Questions?

Thank You !