CoSoLoRec: Joint Factor Model with Content, Social, Location for Heterogeneous Point-of-Interest Recommendation

Hao Guo, Xin Li, Ming He,
Xiangyu Zhao, Guiquan Liu, Guandong Xu

University of Science and Technology of China

IFLYTEK Research

University of Technology Sydney

Outline

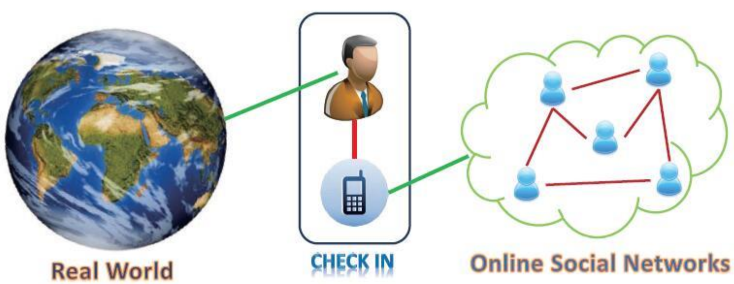
- Background and Motivation
- Model Construction
  - Geographical Influence
  - Social Correlation
  - Probabilistic Latent Factor Model
  - Textual Analysis
- Learning and Inference
- Experimental Results
- Conclusion

Outline

- Background and Motivation
- Model Construction
  - Geographical Influence
  - Social Correlation
  - Probabilistic Latent Factor Model
  - Textual Analysis
- Learning and Inference
- Experimental Results
- Conclusion

Location based Social Networks

- Location-based Social Networks(LBSNs) grow rapidly, such as Foursquare, Gowalla and so on.
- LBSNs break the boundary between the physical world and virtual networks.
- Point-of-Interest Recommendation not only benefit merchants but also benefit customers.



Heterogeneous Point-of-Interest Recommendation

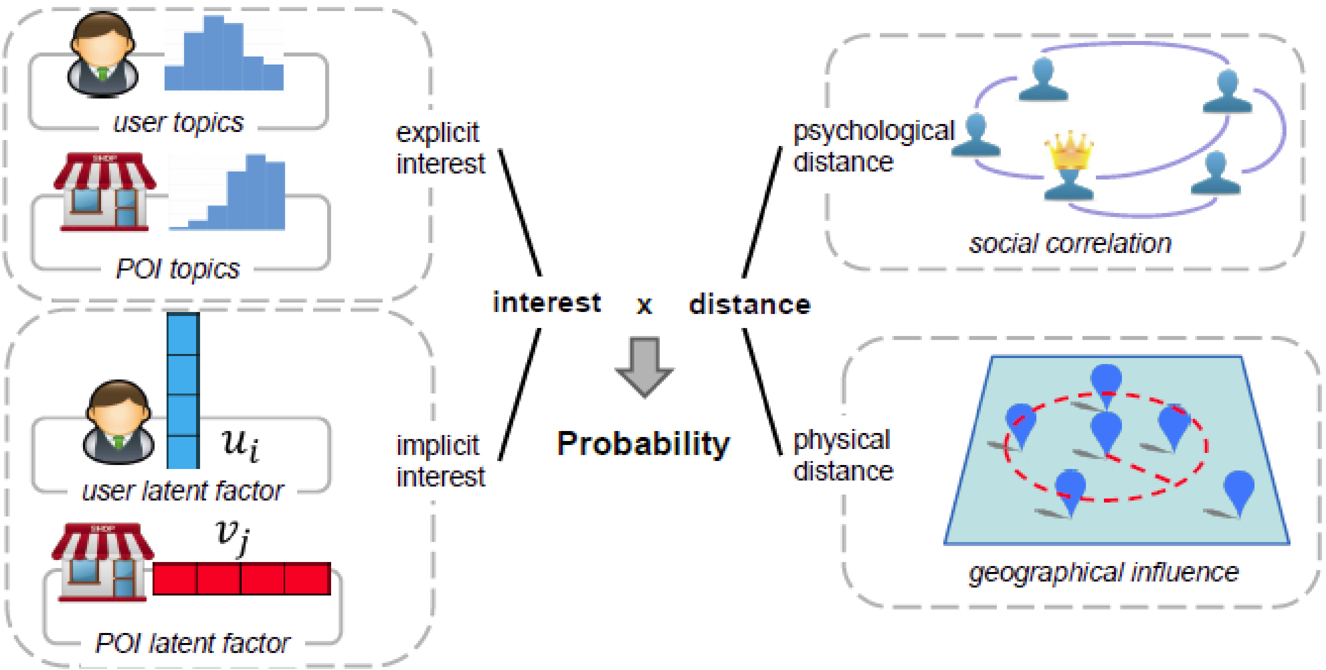
- Task: to recommend Point-of-Interest(POIs) based on different factors, such as geographical, social and textual information.
- Various factors can influence performance of recommendation:
  - Tobler's Law of Geographical Influence
  - Homophily of Social Correlation
  - Heterogeneous Information
- We propose a novel probabilistic latent factor model by considering the above three factors simultaneously.

Outline

- Background and Motivation
- Model Construction
  - Geographical Influence
  - Social Correlation
  - Probabilistic Latent Factor Model
  - Textual Analysis
- Learning and Inference
- Experimental Results
- Conclusion

Model Construction

- Our model consists of four-fold: physical distance, psychological distance, explicit interest and implicit interest.



Model Construction

- Fused Probabilistic Latent Factor Model
  - each user i is associated with his or her interest $\eta(i,j)$ with respect to POI j
    - Implicit interest(topic distribution of users and POIs): $\eta_1(i,j) = \boldsymbol{\theta}_i^T \boldsymbol{\pi}_j$
    - Explicit interest(latent factor combination): $\eta_2(i,j) = \boldsymbol{u}_i^T \boldsymbol{v}_j$
  - each user has an intended visiting probability $p_f(i,j)$ with respect to POI j on the basis of friend-based Collaborative Filtering.
  - geographical influence impels user i to estimate the probability he or she will visits POI j denoted as $p_l(i,j)$ based on Kernel Density Estimation.
- Whole Model

1. Draw a user interest
   (a) Generator user latent factor $u_{iw} \sim Gamma(\alpha_U, \beta_U)$
   (b) Generator item latent factor $v_{jw} \sim Gamma(\alpha_V, \beta_V)$
   (c) user's explicit interest $\eta_1(i,j) = \boldsymbol{\theta}_i^T \boldsymbol{\pi}_j$, implicit interest $\eta_2(i,j) = \mathbf{u}_\mathbf{i}^\mathbf{T} \mathbf{v}_\mathbf{j}$
   (d) user's interest $\eta(i,j) = \eta_1(i,j) + \eta_2(i,j)$
2. $y(i,j) \sim P(p(i,j))$ where
   $p(i,j) = (\eta_1(i,j) + \eta_2(i,j))((1-\lambda)p_l(i,j) + \lambda p_f(i,j))$

Model Construction-Geographical Influence

- we employ Kernel Density Estimation (KDE) to model the geographical influence of POIs on users' visiting behaviors.

$$p_l(i,j) = P\left(\bigcup_{t=1}^{|L_i|}(c_t \to c_0)\right) = 1 - P\left(\bigcap_{t=1}^{|L_i|}\overline{c_t \to c_0}\right) = 1 - \prod_{t=1}^{|L_i|}\left(1 - P(c_t \to c_0)\right)$$

$$P(c_t \to c_0) = \frac{1}{|X_i|}\sum_{x \in X_i} K\left(\frac{z_t - x}{\delta}\right) = \frac{1}{\sqrt{2\pi}|X_i|}\sum_{x \in X_i} e^{-\frac{(z_t - x)^2}{2\delta^2}} \qquad \delta \approx 1.06\hat{\delta}|X_i|^{-1/5}$$

- The computational complexity grows rapidly with the increment of L_i . We use approximation algorithm.

Model Construction-Social Correlation

- we adopt the user-based collaborative filtering(CF) by regarding all of i's friends as neighbors named as Friend-based Collaborative Filtering.

$$p_f(i,j) = \frac{\sum_{i' \in \mathcal{F}_i} sim(i,i') r_{i'j}}{\sum_{i' \in \mathcal{F}_i} sim(i,i')} \cdot \frac{1}{r_{max}}$$

- We choose cosine similarity to validate $sim(i,j)$.

Model Construction-Probabilistic Latent Factor Model

- the process of Probabilistic Latent Factor Model is as follows:

1. for all w, generate $u_{iw} \sim p\left(u_{iw} | \Phi_{u_{iw}}\right)$
2. for all w, generate $v_{jw} \sim p\left(v_{jw} | \Phi_{v_{jw}}\right)$
3. generate $\hat{f_{ij}}$ from user i to location j with equation $\hat{f_{ij}} = \sum_{w=1}^{d} u_{iw} v_{jw} = \boldsymbol{u_i v_j}$
4. generate $\hat{y_{ij}} \sim P\left(\hat{f_{ij}}\right)$

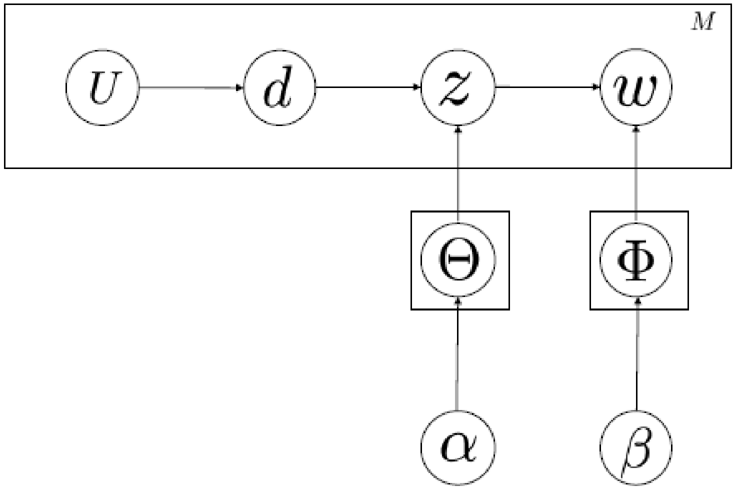
$$p\left(U | \alpha_U, \beta_U\right) = \prod_{i=1}^{m} \prod_{w=1}^{d} \frac{{u_{iw}}^{\alpha_U - 1} \exp\left(-u_{iw} / \beta_U\right)}{{\beta_U}^{\alpha_U} \Gamma\left(\alpha_U\right)}$$

$$p\left(V | \alpha_V, \beta_V\right) = \prod_{j=1}^{n} \prod_{w=1}^{d} \frac{{v_{jw}}^{\alpha_V - 1} \exp\left(-v_{jw} / \beta_V\right)}{{\beta_V}^{\alpha_V} \Gamma\left(\alpha_V\right)}$$

$$P\left(\hat{y_{ij}} | \hat{f_{ij}}\right) = \left(\boldsymbol{u}_i \boldsymbol{v}_j\right)^{\hat{y_{ij}}} \frac{\exp\left(-\boldsymbol{u}_i \boldsymbol{v}_j\right)}{\hat{y_{ij}}!}$$

Model Construction-Textual Analysis

- In order to extract users' explicit interest, we use an aggregated LDA model.
- In order to learn users' interests, we aggregate all the reviews written by each user into a document. Thus, user and document are interchangeable in reflecting user's interest.



$$\pi_{js} = \frac{n_j^{(s)} + \alpha}{\sum_{s=1}^{K} n_j^{(s)} + K\alpha}$$

$$\theta_{is} = \frac{n_i^{(s)} + \alpha}{\sum_{s=1}^{K} n_i^{(s)} + K\alpha}$$

Outline

- Background and Motivation
- Model Construction
  - Geographical Influence
  - Social Correlation
  - Probabilistic Latent Factor Model
  - Textual Analysis
- Learning and Inference
- Experimental Results
- Conclusion

Learning and Inference

- Given the observed data collection $\mathcal{D} = \{p(i,j)\}^{I_{ij}}$ where $p(i,j)$ is the user visiting probability, and $I_{ij} = 1$ when user i visited POI j , and $I_{ij} = 0$ otherwise.
- Maximum likelihood estimation (MLE) method to learn parameters $\Lambda = \{U, V\}$

$$\mathcal{L}(U,V;\mathcal{D}) = \sum_{i=1}^{M}\sum_{j=1}^{N} I_{ij}\left(y(i,j)\log p(i,j) - p(i,j)\right) + \sum_{i=1}^{M}\sum_{w=1}^{d}\left((\alpha_U - 1)\log u_{iw} - \frac{u_{iw}}{\beta_U}\right)$$
$$+ \sum_{j=1}^{N}\sum_{w=1}^{d}\left((\alpha_V - 1)\log v_{jw} - \frac{v_{jw}}{\beta_V}\right) \qquad ($$

$$p(i,j) = \left(\boldsymbol{u}_i^T \boldsymbol{v}_j + \boldsymbol{\theta}_i^T \boldsymbol{\pi}_j\right)\left((1-\lambda)\, p_l(i,j) + \lambda p_f(i,j)\right)$$

- Stochastic gradient descent(SGD) method to optimize them and update parameters iteratively using all training samples.

Outline

- Background and Motivation
- Model Construction
  - Geographical Influence
  - Social Correlation
  - Probabilistic Latent Factor Model
  - Textual Analysis
- Learning and Inference
- Experimental Results
- Conclusion

Experimental Results

- Datasets: Foursquare and Yelp

Table 2. Data Description

	Yelp	Foursquare
Number of users	366715	571700
Number of locations	61184	8318919
Review items	1569265	5550203
User-location matrix density	6.99×10^{-5}	1.17×10^{-6}
Number of Cities	10	50

- Foursquare:6895 users for 13208 POIs with 166989 ratings
- Yelp: 3059 users, 26446 business with 180755 review records.

Evaluation Method

- Baselines: PMF, NMF, BNMF, GT-BNMF, Geo-PFM
- Our model: CoSoLoRec, CoSoLo-PMF, CoSoLo-NMF, CoSoLo-BNMF.
- Evaluation Metrics:
  - Relative precision:

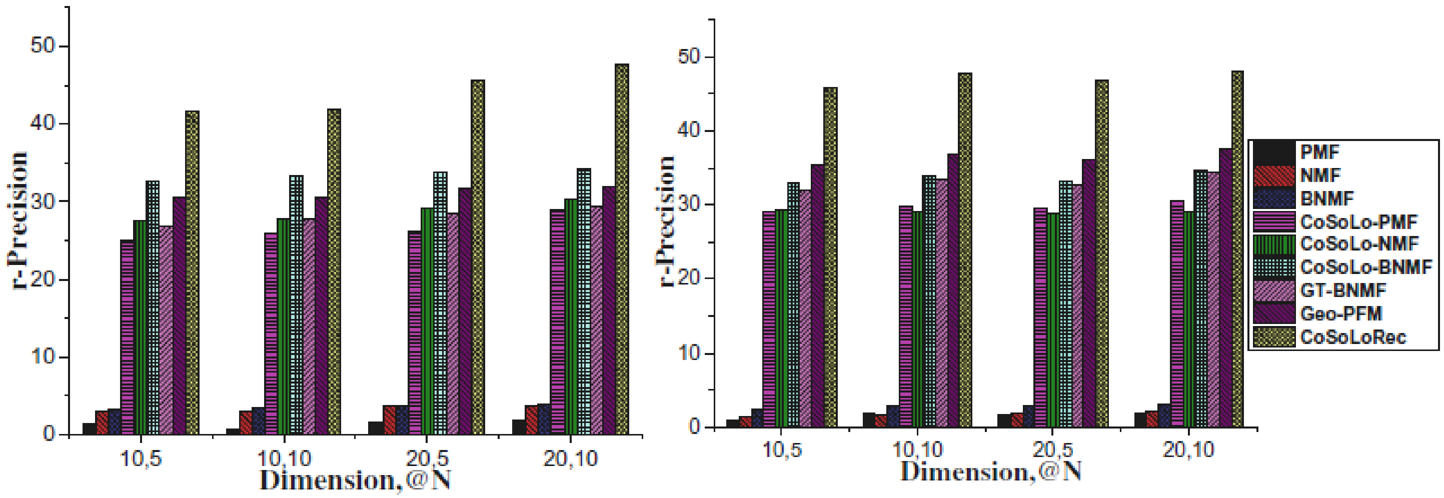
$$rPrecision@N = \frac{|S_{N,rec} \cap S_{visited}| \cdot |C|}{|S_{visited}| \cdot N}$$

  - RMSE:

$$RMSE = \sqrt{\frac{1}{N} \sum_{(u,i) \in E} (r_{ui} - \hat{r_{ui}})^2}.$$

Relative precision

- CoSoLoRec model outperforms all the baselines in these two datasets under our situations.
- CoSoLo-PMF, CoSoLo-NMF and CoSoLo-BNMF show almost equivalent performance in precision.
- Heterogeneous information can reflect user's interests accurately.



(a) rPrecision@N on Yelp (b) rPrecision@N on Foursquare

RMSE

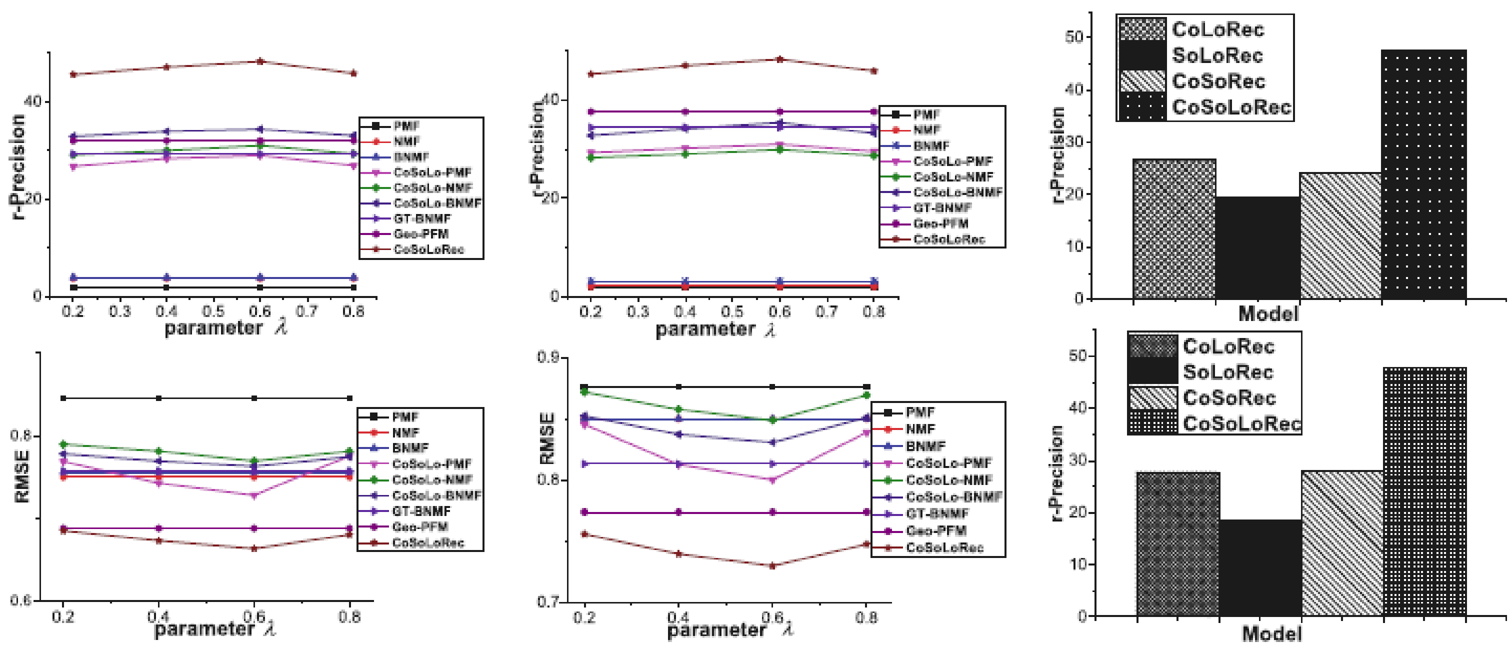
- Our model(CoSoLoRec) achieves less RMSE than baselines with different dimensions of latent factors
- Heterogeneous information can ensure more accuracy in recommending POIs

Table 3. Performance comparison in different dimensions

	D	Metrics	PMF	NMF	BNMF	C-PMF	C-NMF	C-BNMF	GT-BNMF	Geo-PFM	CoSoLoRec
Yelp	10	RMSE	0.8225	0.7644	0.766	0.7639	0.7824	0.7769	0.7241	0.7076	**0.6692**
		Improve	18.64 %	12.45 %	12.64 %	12.40 %	14.47 %	13.86 %	7.58 %	5.43 %	
	20	RMSE	0.8455	0.7502	0.7564	0.7365	0.7716	0.7672	0.7573	0.6881	**0.6693**
		Improve	20.84 %	10.78 %	11.52 %	9.12 %	13.26 %	12.76 %	11.62 %	2.73 %	
4sq	10	RMSE	0.8792	0.8515	0.8624	0.8335	0.8612	0.8454	0.8282	0.7815	**0.7476**
		Improve	14.97 %	12.20 %	13.31 %	10.31 %	13.19 %	11.57 %	9.73 %	4.34 %	
	20	RMSE	0.8763	0.8498	0.85	0.8019	0.8509	0.8334	0.8132	0.7739	**0.7319**
		Improve	16.48 %	13.87 %	13.89 %	8.73 %	13.99 %	12.18 %	10.00 %	5.43 %	

Parameter Sensitivity and Model Robust

- Both geographical and social influence play comparative roles.
- User's text information contributes greater than geographical and social influence in precision of recommending.



(a) r-Pre and RMSE on Yelp (b) r-Pre and RMSE on 4sq (c) Robust for Yelp and 4sq

Outline

- Background and Motivation
- Model Construction
  - Geographical Influence
  - Social Correlation
  - Probabilistic Latent Factor Model
  - Textual Analysis
- Learning and Inference
- Experimental Results
- Conclusion

Conclusion

- we proposed CoSoLoRec model which fused geographical information, social information and text information.
- Experimental results show that our fused model is superior to all other approaches evaluated, such as PMF, NMF, GT-BNMF and Geo-PFM.
- Our model performs better in any different combinations between geographical and social influence.
- Text information is more important than above two factors.

Q&A

Thank You!