Social Poisson Factorization

By Zhao Yi

DataMining Lab

Overview

- A brief of PMF

Some issues about PMF model

- Poisson Factorization models

Hierarchical Poisson Factorization

& Poisson Trust Factorization

- Conclusion & Discussion

Probabilistic Matrix Factorization

- Basic PMF

Modeling & Derivation

- Automatic Complexity Control for PMF Models

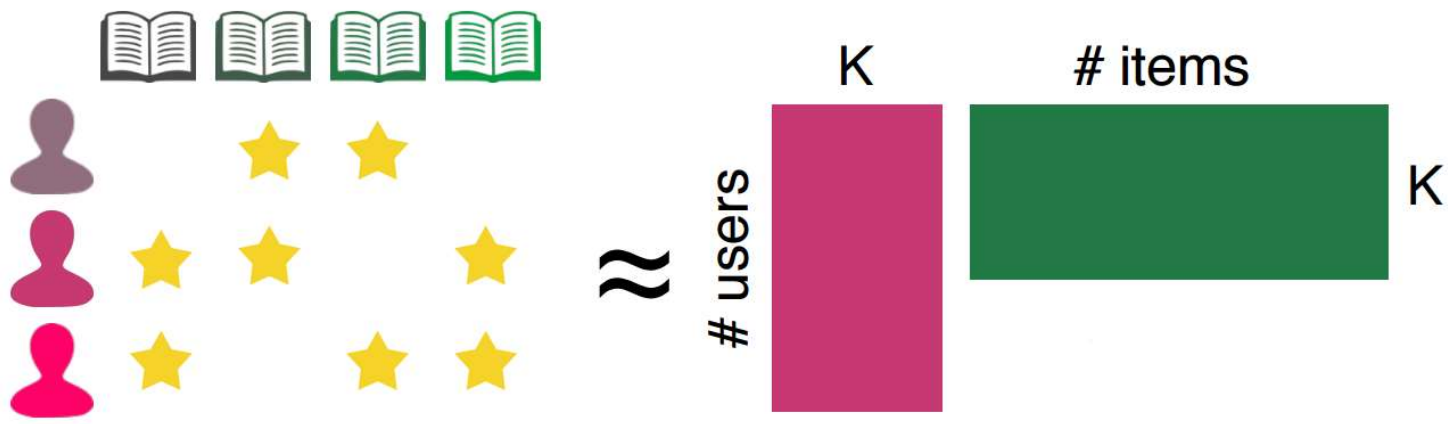
Complexity analysis

- Constrained PMF

Mnih, Andriy, and Ruslan Salakhutdinov. "Probabilistic matrix factorization." Advances in neural information processing systems. 2007.

Probabilistic Matrix Factorization

- Preliminary



Probabilistic Matrix Factorization

- Preliminary

Low-dimensional factor/ Low rank model

Probabilistic Matrix Factorization

- Preliminary

	国内	流行	烧脑
小明	-1.0	0.2	0.9

	国内	流行	烧脑
<Book>	1.0	0.8	-0.3

$Score_{小明,\ <Book}$ = -1.0 * 1.0 + 0.2*0.8 + 0.9*(-0.3) = -1.11

Probabilistic Matrix Factorization

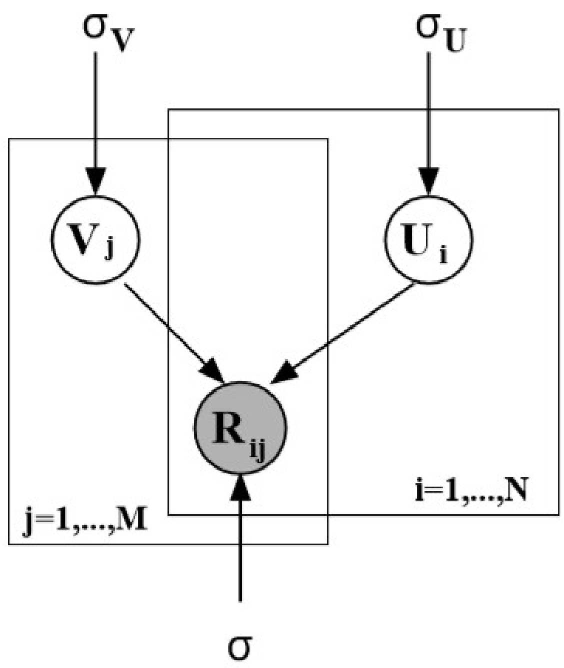
- Preliminary

$$\boldsymbol{R}_{\boldsymbol{N}\times\mathrm{M}} = \boldsymbol{U}_{\boldsymbol{N}\times\boldsymbol{D}}^{\boldsymbol{T}}\boldsymbol{V}_{\boldsymbol{D}\times\mathrm{M}}$$

  - $\widehat{\boldsymbol{R}}_{\boldsymbol{N}\times\mathrm{M}} = \boldsymbol{U}_{\boldsymbol{N}\times\boldsymbol{D}}^{\boldsymbol{T}}\boldsymbol{V}_{\boldsymbol{D}\times\mathrm{M}}$
  - $\widehat{\boldsymbol{R}}_{\boldsymbol{N}\times\mathrm{M}} \approx \boldsymbol{R}_{\boldsymbol{N}\times\mathrm{M}}$ in observed parts
  - Constraining U and V in case of overfitting

Probabilistic Matrix Factorization

- Basic PMF



Probabilistic Matrix Factorization

- Basic PMF
  - Gauss Assumption

$$p(R|U,V,\sigma^2) = \prod_{i=1}^{N} \prod_{j=1}^{M} \left[\mathcal{N}(R_{ij}|U_i^T V_j, \sigma^2) \right]^{I_{ij}}$$

$$p(U|\sigma_U^2) = \prod_{i=1}^{N} \mathcal{N}(U_i|0, \sigma_U^2 \mathbf{I}), \qquad p(V|\sigma_V^2) = \prod_{j=1}^{M} \mathcal{N}(V_j|0, \sigma_V^2 \mathbf{I})$$

  - The log of posterior distribution:

$$\ln p(U,V|R,\sigma^2,\sigma_V^2,\sigma_U^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^{N} \sum_{j=1}^{M} I_{ij}(R_{ij} - U_i^T V_j)^2 - \frac{1}{2\sigma_U^2} \sum_{i=1}^{N} U_i^T U_i - \frac{1}{2\sigma_V^2} \sum_{j=1}^{M} V_j^T V_j$$
$$- \frac{1}{2} \left(\left(\sum_{i=1}^{N} \sum_{j=1}^{M} I_{ij} \right) \ln \sigma^2 + ND \ln \sigma_U^2 + MD \ln \sigma_V^2 \right) + C,$$

Probabilistic Matrix Factorization

- Basic PMF
  - Derivation

    Energy function:

$$E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}\left(R_{ij} - U_i^T V_j\right)^2 + \frac{\lambda_U}{2}\sum_{i=1}^{N} \| U_i \|_{Fro}^2 + \frac{\lambda_V}{2}\sum_{j=1}^{M} \| V_j \|_{Fro}^2$$

$$\lambda_U = \sigma^2/\sigma_U^2,\ \lambda_V = \sigma^2/\sigma_V^2$$

Probabilistic Matrix Factorization

- Basic PMF
  - Derivation

    Energy function:

$$E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}\left(R_{ij} - U_i^T V_j\right)^2 + \frac{\lambda_U}{2}\sum_{i=1}^{N} \| U_i \|_{Fro}^2 + \frac{\lambda_V}{2}\sum_{j=1}^{M} \| V_j \|_{Fro}^2$$

$$\lambda_U = \sigma^2/\sigma_U^2,\ \lambda_V = \sigma^2/\sigma_V^2$$

    Gradient descent…

Probabilistic Matrix Factorization

- Basic PMF
  - Derivation

    Energy function:

$$E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}\left(R_{ij} - U_i^T V_j\right)^2 + \frac{\lambda_U}{2}\sum_{i=1}^{N} \| U_i \|_{Fro}^2 + \frac{\lambda_V}{2}\sum_{j=1}^{M} \| V_j \|_{Fro}^2$$

$$\lambda_U = \sigma^2/\sigma_U^2,\ \lambda_V = \sigma^2/\sigma_V^2$$

    Gradient descent…

    …?

Probabilistic Matrix Factorization

- Basic PMF
  - Complexity

    D the numbers of the users' preference/ items' features

    - The larger D is, the more precise the model is
    - Large D → over fitting
    - The real data sets are unbalanced → Sparsity

Probabilistic Matrix Factorization

- Basic PMF
  - Complexity

    D the numbers of the users' preference/ items' features

    - The larger D is, the more precise the model is
    - Large D → over fitting
    - The real data sets are unbalanced → Sparsity

    λ the regularization parameters $\lambda_U = \sigma^2/\sigma_U^2, \lambda_V = \sigma^2/\sigma_V^2$

    - The larger σ is, the less reliable the model is
    - Large λ_U → small absolute values of U, V

$$E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}\left(R_{ij} - U_i^T V_j\right)^2 + \frac{\lambda_U}{2}\sum_{i=1}^{N} \| U_i \|_{Fro}^2 + \frac{\lambda_V}{2}\sum_{j=1}^{M} \| V_j \|_{Fro}^2$$

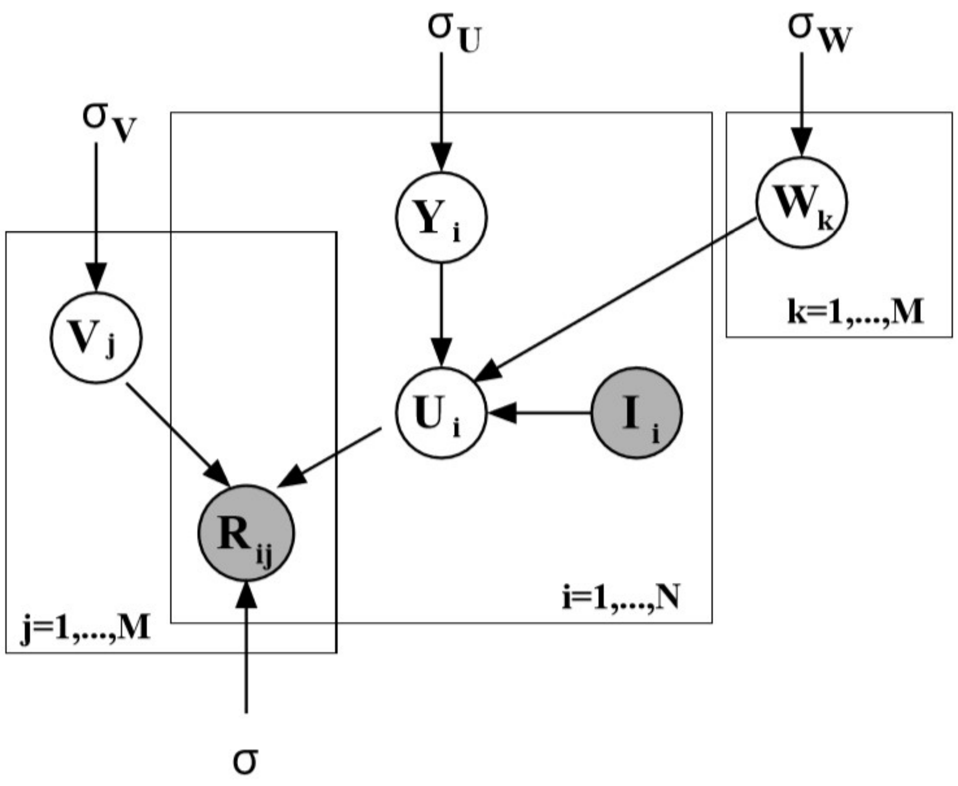
Probabilistic Matrix Factorization

- Automatic Complexity Control for PMF models
  - Hyperparameters for prior

$$\ln p(U, V, \sigma^2, \Theta_U, \Theta_V | R) = \ln p(R|U, V, \sigma^2) + \ln p(U|\Theta_U) + \ln p(V|\Theta_V) + \ln p(\Theta_U) + \ln p(\Theta_V) + C,$$

Probabilistic Matrix Factorization

- Constrained PMF



Probabilistic Matrix Factorization

- Constrained PMF
  - A latent similarity constrain matrix

    $W \in \ R_{D \times M}$

  - For the user i, the feature vector is:

$$U_i = Y_i + \frac{\sum_{k=1}^{M} I_{ik} W_k}{\sum_{k=1}^{M} I_{ik}}$$

Probabilistic Matrix Factorization

- Constrained PMF
  - A latent similarity constrain matrix

$$U_i = Y_i + \frac{\sum_{k=1}^{M} I_{ik} W_k}{\sum_{k=1}^{M} I_{ik}}$$

	M1	M2	…
D1	X_1	X_2	
D2	X_1’	X_2’	
D3	X_1’’	X_2’’	
…	…	…	

$$U_i = \left(\frac{x_1 + x_2 + \cdots + x_D}{2}, \frac{x_1' + x_2' + \cdots + x_D'}{2}, \ldots\right)$$

Probabilistic Matrix Factorization

- Constrained PMF
  - The conditional distribution

$$p(R|Y,V,W,\sigma^2) = \prod_{i=1}^{N}\prod_{j=1}^{M}\left[\mathcal{N}(R_{ij}|g([Y_i + \frac{\sum_{k=1}^{M} I_{ik}W_k}{\sum_{k=1}^{M} I_{ik}}]^T V_j), \sigma^2)\right]^{I_{ij}}$$

  - The Gauss prior on W

$$p(W|\sigma_W) = \prod_{k=1}^{M}\mathcal{N}(W_k|0, \sigma_W^2\mathbf{I})$$

  - The Energy function:

$$E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}\left(R_{ij} - g([Y_i + \frac{\sum_{k=1}^{M} I_{ik}W_k}{\sum_{k=1}^{M} I_{ik}}]^T V_j)\right)^2 + \frac{\lambda_Y}{2}\sum_{i=1}^{N} \| Y_i \|_{Fro}^2 + \frac{\lambda_V}{2}\sum_{j=1}^{M} \| V_j \|_{Fro}^2 + \frac{\lambda_W}{2}\sum_{k=1}^{M} \| W_k \|_{Fro}^2$$

Social Poisson Factorization

- Poisson Factorization

Modeling

Gopalan P, Hofman J M, Blei D M. "Scalable recommendation with hierarchical Poisson factorization."
Conference on Uncertainty in Artificial Intelligence. AUAI Press, 2015.

- Poisson Trust Factorization

Modeling

Chaney A J B, Gopalan P, Blei D M. "Poisson trust factorization for incorporating social networks into personalized item recommendation"
NIPS Workshop: What Difference Does Personalization Make. 2013.

Social Poisson Factorization

- Poisson Factorization

  Motivation: the classic PMF has two issues

  1. It does not prevent that the ratings become negative, which is a natural result of the Gauss distribution
  2. If no heuristics are applied, the model needs to model all zero ratings and therefore, dramatically impacting the predictive performance

$$\log P(X \mid \theta, \phi) = -\frac{1}{2\sigma^2} \sum_i \sum_j (X_{i,j} - \theta_i^T \phi_j)^2 - \frac{1}{2} N \times M \log \sigma^2 + C$$

Social Poisson Factorization

- Poisson Factorization

  Basic idea:

  $$X_{i,j} \sim Poisson\left(\theta_i^T \phi_j\right)$$

  $$p(X_{i,j} \mid \theta_i^T, \phi_j) = (\theta_i^T \phi_j)^{X_{i,j}} \frac{\exp\{-\theta_i^T \phi_j\}}{X_{i,j}!}$$

Social Poisson Factorization

- Poisson Factorization

  Motivation: the classic PMF has two issues

  1. It does not prevent that the ratings become negative, which is a natural result of the Gauss distribution
  2. If no heuristics are applied, the model needs to model all zero ratings and therefore, dramatically impacting the predictive performance

$$\log P(X \mid \theta, \phi) = -\frac{1}{2\sigma^2} \sum_i \sum_j (X_{i,j} - \theta_i^T \phi_j)^2 - \frac{1}{2} N \times M \log \sigma^2 + C$$

$$\log P(X \mid \theta, \phi) = \sum_{i,j} \left\{ X_{i,j} \log\left(\theta_i^T \phi_j\right) - \log X_{i,j}! - \theta_i^T \phi_j \right\}$$

Social Poisson Factorization

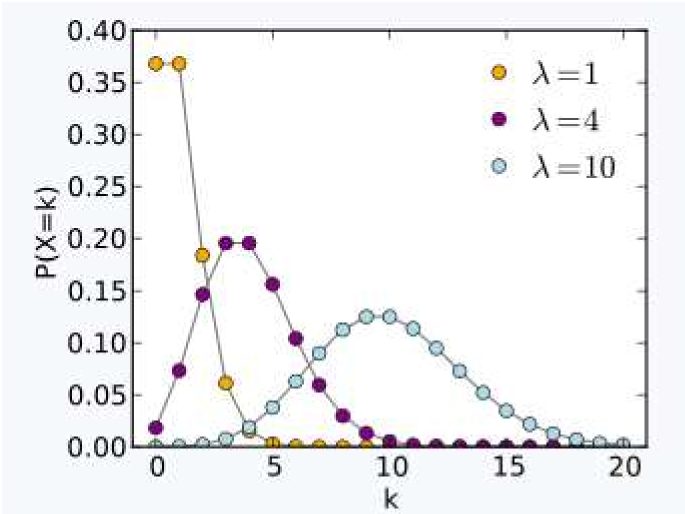
- Poisson Factorization

Basic idea:

$$X_{i,j} \sim Poisson\left(\theta_i^T \phi_j\right)$$

$$p(X_{i,j} \mid \theta_i^T, \phi_j) = (\theta_i^T \phi_j)^{X_{i,j}} \frac{\exp\{-\theta_i^T \phi_j\}}{X_{i,j}!}$$

$$\text{score}_{ui} = \mathrm{E}[\theta_u^\top \beta_i \mid y]$$



Social Poisson Factorization

- Poisson Factorization

Model:

1. For each user u:
   (a) Sample activity $\xi_u \sim \text{Gamma}(a', a'/b')$.
   (b) For each component k, sample preference

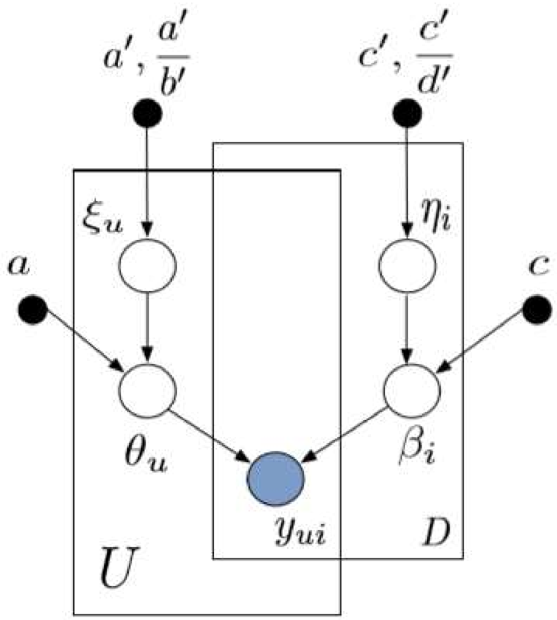
$$\theta_{uk} \sim \text{Gamma}(a, \xi_u).$$

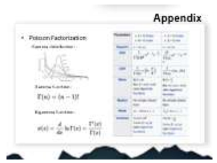
2. For each item i:
   (a) Sample popularity $\eta_i \sim \text{Gamma}(c', c'/d')$.
   (b) For each component k, sample attribute

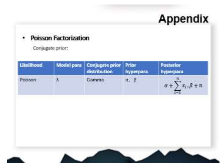
$$\beta_{ik} \sim \text{Gamma}(c, \eta_i).$$

3. For each user u and item i, sample rating

$$y_{ui} \sim \text{Poisson}(\theta_u^\top \beta_i).$$







Social Poisson Factorization

- Poisson Factorization

  Advantages over classical Gauss MF:

  - **HBF captures sparse factors**

    Small shape parameter → most of the weight will be close to 0

Social Poisson Factorization

- Poisson Factorization

  Advantages over classical Gauss MF:

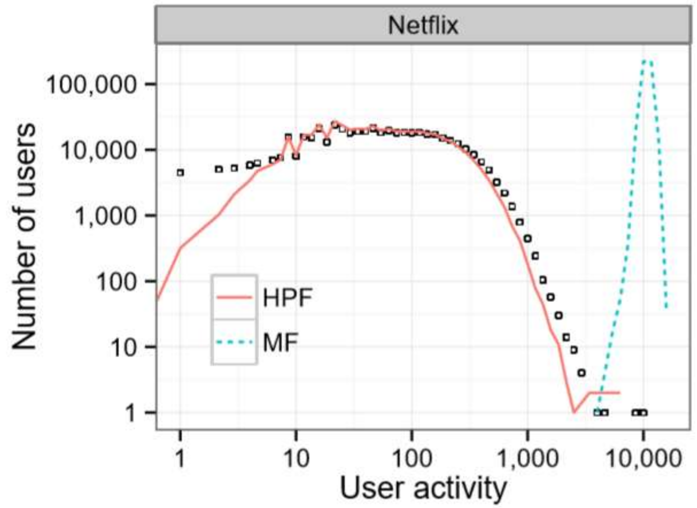
  - **HBF captures sparse factors**

    Small shape parameter → most of the weight will be close to 0

  - **HBF models the long-tail of users and items**

$$b_u \sim \mathrm{Poisson}(\theta_u^T \sum_i \beta_i)$$

$$[y_{u1}, \cdots, y_{uM}] \sim \mathrm{Mult}(b_u, \frac{\theta_u^T \beta_i}{\theta_u^T \sum_i \beta_i})$$



Social Poisson Factorization

- Poisson Factorization

  Inference:

  - Pick a family of distribution q over the latent variables with its own **variation parameters**
  - Optimizing q to approximate the posterior p
  - To choose q, we use the **mean field assumption**: each variable is independent , allowing q to factorize
  - Use coordinate ascent: iteratively optimize each variable, holding the others fixed

Social Poisson Factorization

- Poisson Factorization

Inference:

For all users and items, initialize the user parameters γ_u, κ_u^{rte} and item parameters λ_i, τ_i^{rte} to the prior with a small random offset. Set the user activity and item popularity shape parameters:

$$\kappa_u^{\text{shp}} = a' + Ka; \quad \tau_i^{\text{shp}} = c' + Kc$$

Repeat until convergence:

1. For each user/item such that $y_{ui} > 0$, update the multinomial:

$$\phi_{ui} \propto \exp\{\Psi(\gamma_{uk}^{\text{shp}}) - \log \gamma_{uk}^{\text{rte}} + \Psi(\lambda_{ik}^{\text{shp}}) - \log \lambda_{ik}^{\text{rte}}\}.$$

2. For each user, update the user weight and activity parameters:

$$\gamma_{uk}^{\text{shp}} = a + \textstyle\sum_i y_{ui}\phi_{uik}$$

$$\gamma_{uk}^{\text{rte}} = \frac{\kappa_u^{\text{shp}}}{\kappa_u^{\text{rte}}} + \textstyle\sum_i \lambda_{ik}^{\text{shp}}/\lambda_{ik}^{\text{rte}}$$

$$\kappa_u^{\text{rte}} = \frac{a'}{b'} + \sum_k \frac{\gamma_{uk}^{\text{shp}}}{\gamma_{uk}^{\text{rte}}}$$

3. For each item, update the item weight and popularity parameters:

$$\lambda_{ik}^{\text{shp}} = c + \textstyle\sum_u y_{ui}\phi_{uik}$$

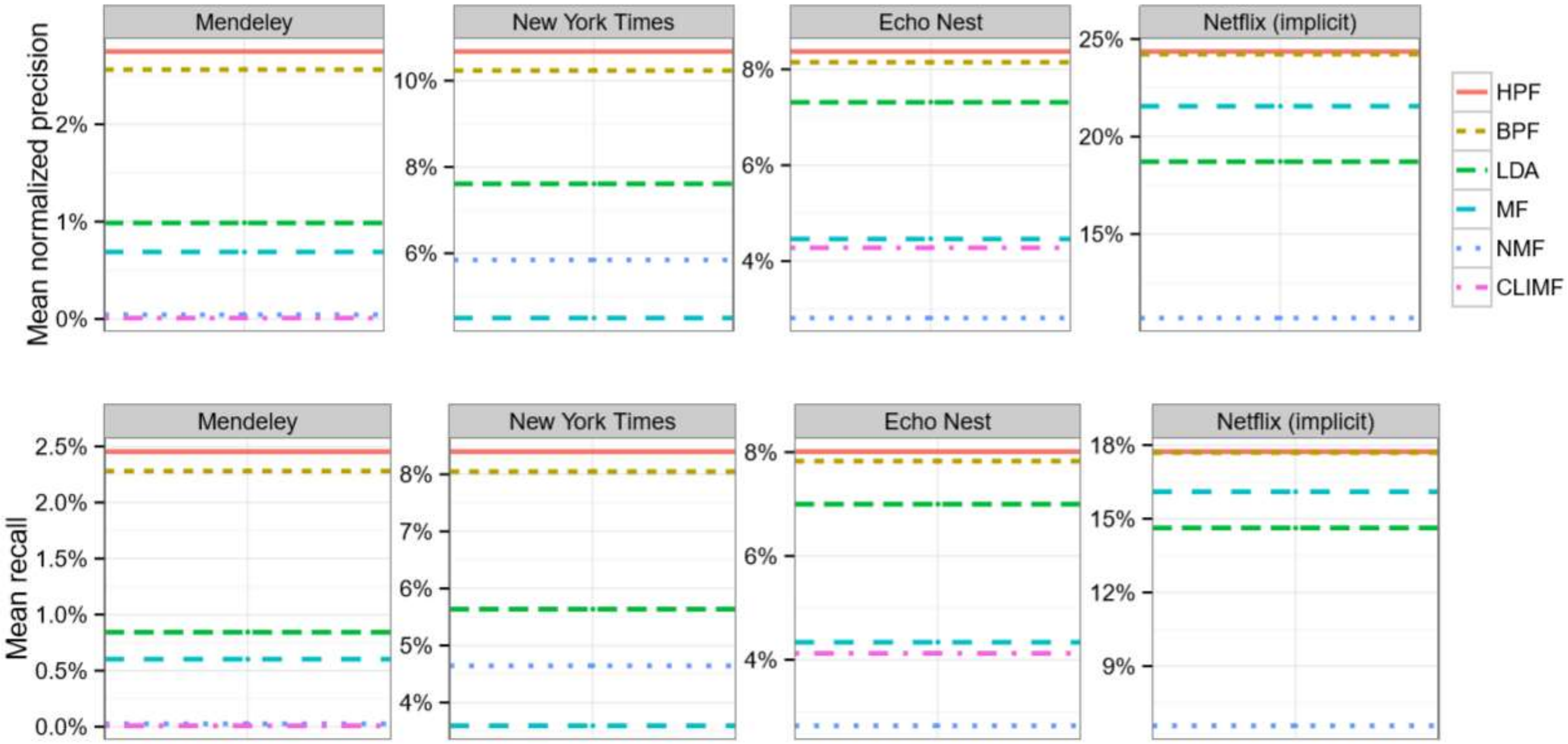
$$\lambda_{ik}^{\text{rte}} = \frac{\tau_i^{\text{shp}}}{\tau_i^{\text{rte}}} + \textstyle\sum_u \gamma_{uk}^{\text{shp}}/\gamma_{uk}^{\text{rte}}$$

$$\tau_i^{\text{rte}} = \frac{c'}{d'} + \sum_k \frac{\lambda_{ik}^{\text{shp}}}{\lambda_{ik}^{\text{rte}}}$$

Social Poisson Factorization

- Poisson Factorization

  performance:



Social Poisson Factorization

- Poisson Trust Factorization

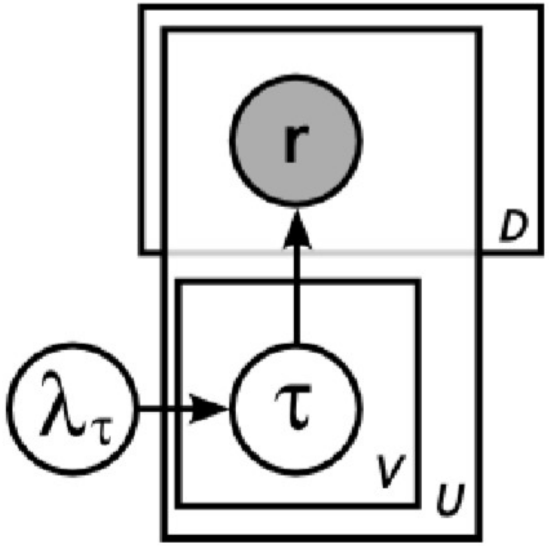
Basic trust model:

User trust:

$$\tau_{uv} \sim \text{Gamma}(f, g)$$

Ratings

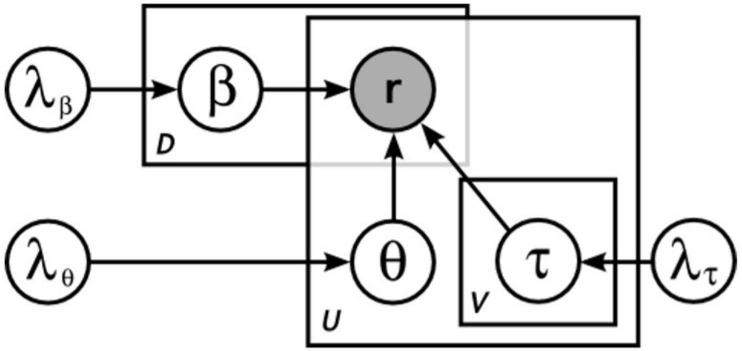
$$r_{ui} \sim \text{Poisson}\left(\sum_{v \in N(u,i)} \tau_{uv} r_{vi}\right)$$



Social Poisson Factorization

- Poisson Trust Factorization

  Combining Matrix Factorization and Friend Trust:



Rating:

$$r_{ui} \sim \text{Poisson}\left(\theta_u^\top \beta_i + \sum_{v \in N(u,i)} \tau_{uv} r_{vi}\right)$$

Social Poisson Factorization

- Poisson Trust Factorization

Generative model

User preference:

$$\theta_u \sim Gamma(a, b)$$

Item features:

$$\beta_i \sim Gamma(c, d)$$

User trust:

$$\tau_{uv} \sim Gamma(f, g)$$

Rating:

- Poisson Trust Factorization

Generative model

User preference:

$$\theta_u \sim Gamma(a, b)$$

Item features:

$$\beta_i \sim Gamma(c, d)$$

User trust:

$$\tau_{uv} \sim Gamma(f, g)$$

Social Poisson Factorization

- Poisson Trust Factorization

Generative model

Auxiliary latent variables $r_{ui} = \sum_n z_{uin}$ where

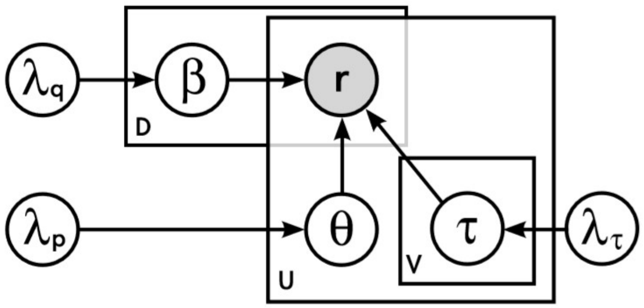
$$n = \{n_1^{MF}, \dots, n_K^{MF}, n_1^{trust}, \dots, n_V^{trust}\}$$

$$z_{uik}^{MF} \sim Poisson(\theta_{uk}\beta_{ik})$$

$$z_{ui}^{trust} \sim Poisson(\tau_{uv} r_{vi})$$

Social Poisson Factorization

- Poisson Trust Factorization

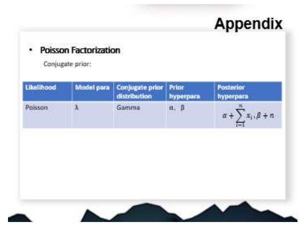


Inference

$$\theta_{uk} \mid \beta, \tau, z, m, r \sim \text{Gamma}\left(a + \sum_i z_{uik},\; b + \sum_i \beta_{ik}\right)$$

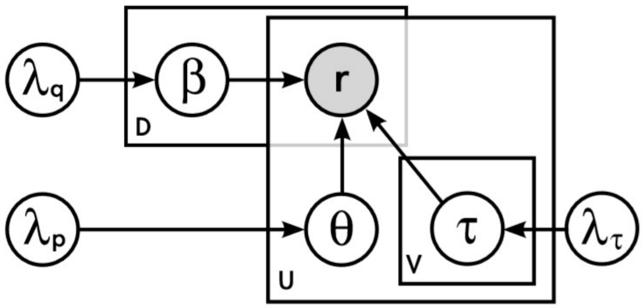
$$\beta_{ik} \mid \theta, \tau, z, m, r \sim \text{Gamma}\left(c + \sum_u z_{uik},\; d + \sum_u \theta_{uk}\right)$$

$$\tau_{uv} \mid \theta, \beta, z, m, r \sim \text{Gamma}\left(f + \sum_{i \in D(u,v)} z_{uiv},\; g + \sum_{i \in D(v)} r_{vi}\right)$$



Social Poisson Factorization

- Poisson Trust Factorization



Inference

$$\theta_{uk} \mid \beta, \tau, z, m, r \sim \text{Gamma}\left(a + \sum_i z_{uik},\ b + \sum_i \beta_{ik}\right)$$

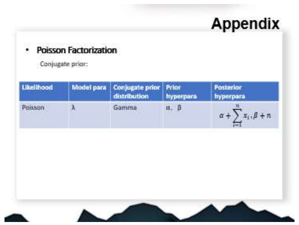
$$\beta_{ik} \mid \theta, \tau, z, m, r \sim \text{Gamma}\left(c + \sum_u z_{uik},\ d + \sum_u \theta_{uk}\right)$$

$$\tau_{uv} \mid \theta, \beta, z, m, r \sim \text{Gamma}\left(f + \sum_{i \in D(u,v)} z_{uiv},\ g + \sum_{i \in D(v)} r_{vi}\right)$$

Complete condition:

$$z_{ui} \mid \theta, \beta, \tau, m, r \sim \text{Mult}\left(r_{ui}, \phi\right)$$

$$\phi \propto \langle \theta_{u1}\beta_{i1}, \cdots, \theta_{uK}\beta_{iK}, \tau_{u1} r_{1i}, \cdots, \tau_{uV} r_{Vi} \rangle$$



Conclusion & Discussion

- Poisson model is more interpretable

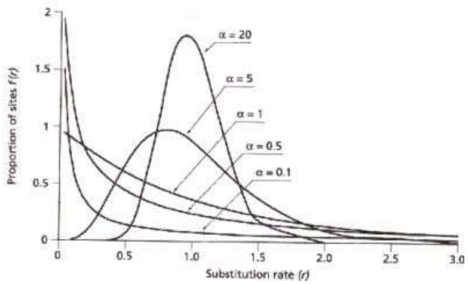
1. A Probabilistic Model for Using Social Networks in Personalized Item Recommendation
2. Bayesian Poisson Tucker Decomposition for Learning the Structure of International Relations
3. Dynamic Poisson Factorization
4. Poisson Trust Factorization for Incorporating Social Networks into Personalized Item Recommendation
5. Scalable Recommendation with Hierarchical Poisson Factorization

THANK YOU.

Appendix

- Poisson Factorization

Gamma distribution:



Gamma function:

$$\Gamma(n) = (n-1)!$$

Digamma function:

$$\psi(x) = \frac{d}{dx} \ln \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$$

Parameters	• $k > 0$ shape • $\theta > 0$ scale	• $\alpha > 0$ shape • $\beta > 0$ rate
Support	$x \in (0, \infty)$	$x \in (0, \infty)$
PDF	$\frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ [1]
CDF	$\frac{1}{\Gamma(k)} \gamma\left(k, \frac{x}{\theta}\right)$	$\frac{1}{\Gamma(\alpha)} \gamma(\alpha, \beta x)$
Mean	$\mathbf{E}[X] = k\theta$ $\mathbf{E}[\ln X] = \psi(k) + \ln(\theta)$ (see digamma function)	$\mathbf{E}[X] = \frac{\alpha}{\beta}$ $\mathbf{E}[\ln X] = \psi(\alpha) - \ln(\beta)$ (see digamma function)
Median	No simple closed form	No simple closed form
Mode	$(k-1)\theta$ for $k \geq 1$	$\frac{\alpha-1}{\beta}$ for $\alpha \geq 1$
Variance	$\operatorname{Var}[X] = k\theta^2$ $\operatorname{Var}[\ln X] = \psi_1(k)$ (see trigamma function)	$\operatorname{Var}[X] = \frac{\alpha}{\beta^2}$ $\operatorname{Var}[\ln X] = \psi_1(\alpha)$ (see trigamma function)

Appendix

- Poisson Factorization

  Conjugate prior:

Likelihood	Model para	Conjugate prior distribution	Prior hyperpara	Posterior hyperpara
Poisson	λ	Gamma	α, β	$\alpha + \sum_{i=1}^{n} x_i, \beta + n$