Scalable Recommendation with Hierarchical Poisson Factorization

汇报人：贺语盈

2016.8.12

Poisson Recommendation

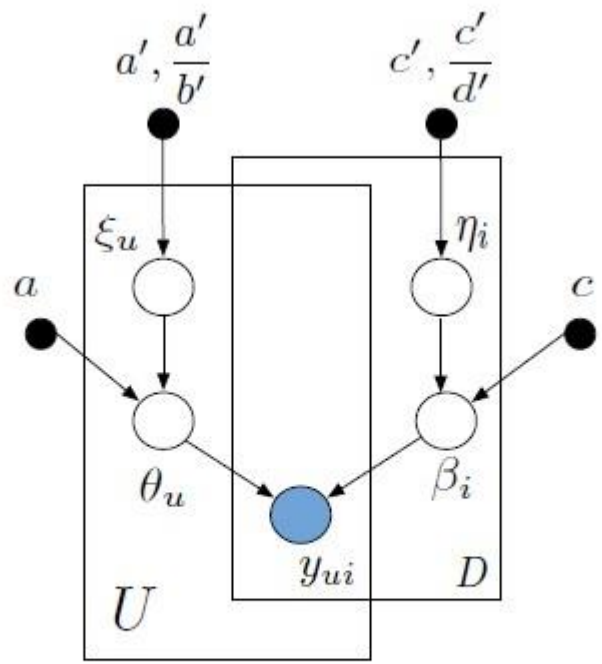


Figure 1: The hierarchical Poisson factorization model.

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

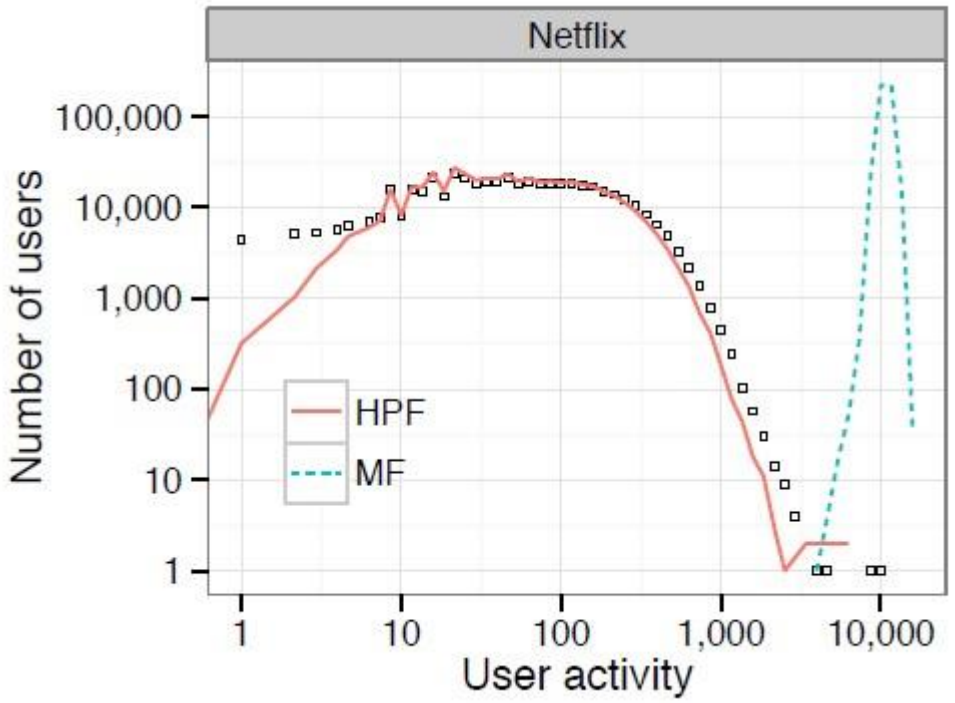
$$\beta_{ik} \sim \text{Gamma}(c, \eta_i).$$

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

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

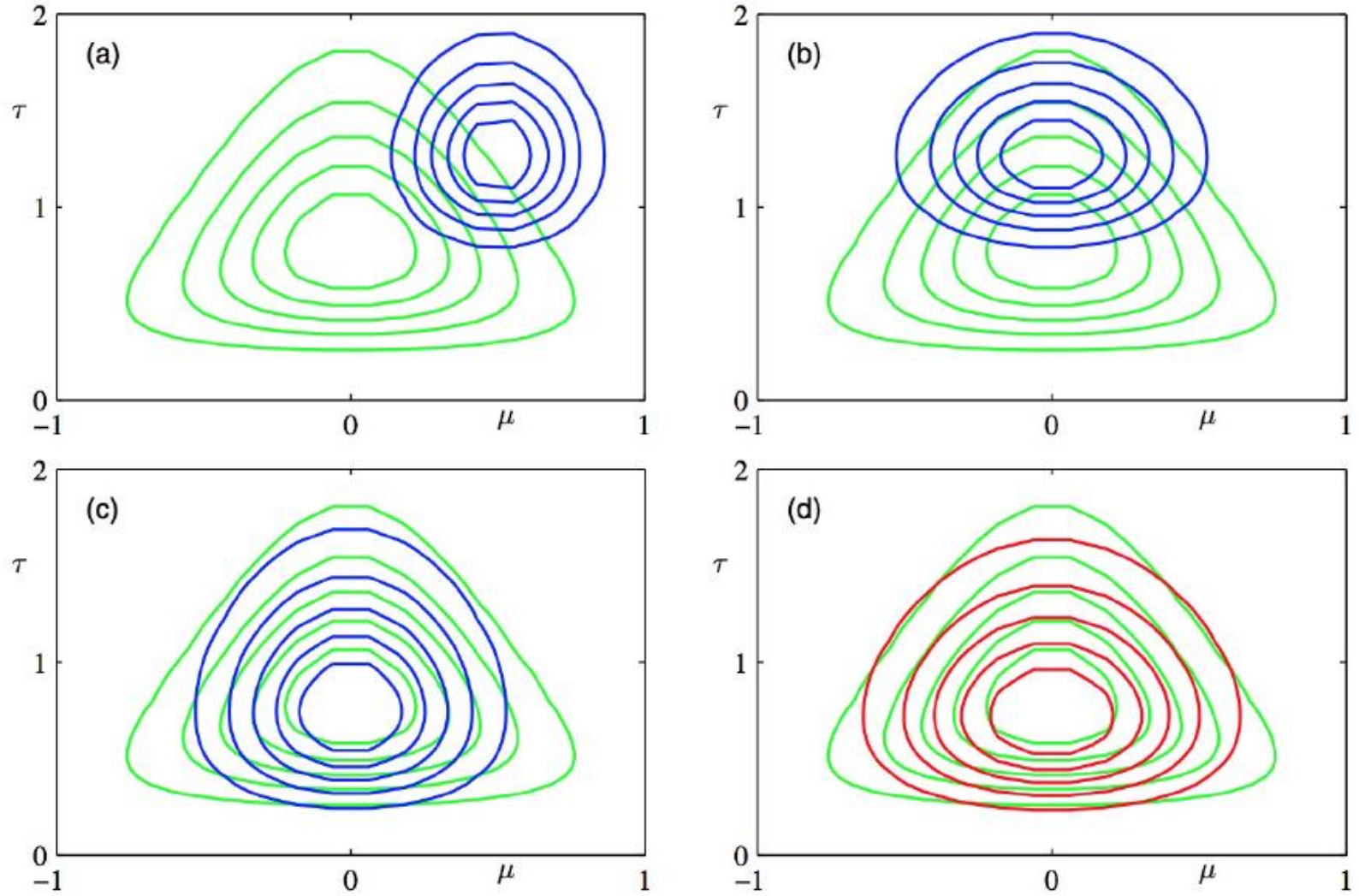
$$\text{score}_{ui} = \text{E}[\theta_u^\top \beta_i \mid y].$$

Properties of HPF



- HPF models the long-tail of users and items.
- HPF downweights the effect of zeros.
- HPF captures sparse factors.
- Fast inference with sparse matrices.

Variational inference



Variational inference

It is universally true that:

$$\ln\left(p(X)\right) = \ln\left(p(X,Z)\right) - \ln\left(p(Z|X)\right)$$

It's also true (a bit silly) that:

$$\ln\left(p(X)\right) = [\ln(p(X,Z)) - \ln(q(Z))] - [\ln(p(Z|X)) - \ln(q(Z))]$$

The above is so that we can insert an arbitrary pdf $q(Z)$ into, now we get:

$$\ln\left(p(X)\right) = \ln\left(\frac{p(X,Z)}{q(Z)}\right) - \ln\left(\frac{p(Z|X)}{q(Z)}\right)$$

Taking the expectation on both sides, given $q(Z)$:

$$\begin{aligned}
\ln\left(p(X)\right) &= \int q(Z)\ln\left(\frac{p(X,Z)}{q(Z)}\right)\mathrm{d}Z - \int q(Z)\ln\left(\frac{p(Z|X)}{q(Z)}\right)\mathrm{d}Z \\
&= \underbrace{\int q(Z)\ln(p(X,Z))\mathrm{d}Z - \int q(Z)\ln(q(Z))\mathrm{d}Z}_{\mathcal{L}(q)} + \underbrace{\left(-\int q(Z)\ln\left(\frac{p(Z|X)}{q(Z)}\right)\mathrm{d}Z\right)}_{\mathbb{KL}(q\|p)} \\
&= \mathcal{L}(q) + \mathbb{KL}(q\|p)
\end{aligned}$$

we can find approximate and optimal $q_i^*(Z_i)$, such that:

$$\ln\left(q_i^*(Z_i)\right) = \mathbb{E}_{i\neq j}\left[\ln\left(p(X,Z)\right)\right]$$

Poisson Factorization

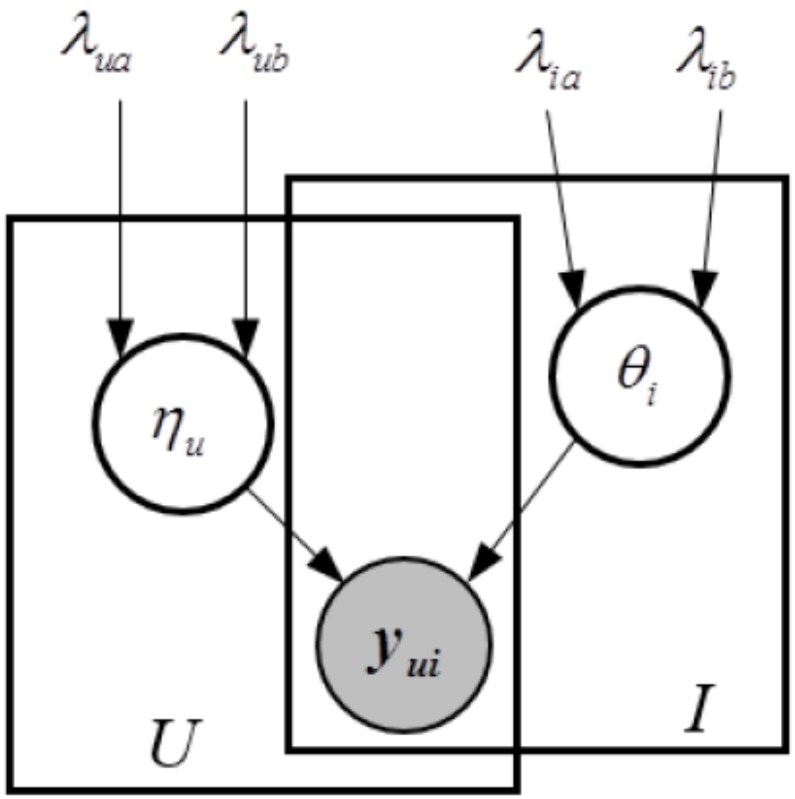


Fig.2 Poisson Factorization Model

$$\eta_u \sim Gamma(\lambda_{ua}, \lambda_{ub})$$
$$\theta_i \sim Gamma(\lambda_{ia}, \lambda_{ib})$$
$$y_{ui} \sim Poisson(\eta_u^T \theta_i)$$

引入隐变量
因子*k*

$$y_{ui} = \sum_k y_{ui,k}$$

$$\eta_{u,k} \sim Gamma(\lambda_{ua,k}, \lambda_{ub,k})$$
$$\theta_{i,k} \sim Gamma(\lambda_{ia,k}, \lambda_{ib,k})$$
$$y_{ui,k} \sim Poisson(\eta_{u,k}^T \theta_{i,k})$$

Poisson Factorization

According to mean-field:

$$p(\eta, \theta, y) = \prod_{u,k} p(\eta_{u,k}) \prod_{i,k} p(\theta_{i,k}) \prod_{ui,k} p(y_{ui,k})$$

In terms of per $\eta_{u,k}$:

$$p(\eta_{u,k} \mid y_{ui,k}, \lambda_{ua,k}, \lambda_{ub,k})$$

$$\propto \eta_{u,k}^{\lambda_{ua,k}-1} e^{-\lambda_{ub,k}\eta_{u,k}} \prod_i \left(\eta_{u,k}^T \theta_{i,k}\right)^{y_{ui,k}} e^{-\eta_{u,k}^T \theta_{i,k}}$$

$$\propto \eta_{u,k}^{\lambda_{ua,k}+\sum_i y_{ui,k}-1} e^{-\left(\lambda_{ub,k}+\sum_i \theta_{i,k}\right)\lambda_{u,k}}$$

$$= Gamma\left(\lambda_{ua,k} + \sum_i y_{ui,k}, \lambda_{ub,k} + \sum_i \theta_{i,k}\right)$$

Update the user weight parameters:

$$(\lambda_{ua,k}, \lambda_{ub,k}) \leftarrow \left(\lambda_{ua,k} + \sum_i y_{ui,k}, \lambda_{ub,k} + \sum_i \theta_{i,k}\right)$$

Poisson Factorization

对于$\theta_{i,k}(\forall k \in K)$的推导方法同上。对于$y_{ui,k}$，因为存在$y_{ui} = \sum_k y_{ui,k}$，因此可以增加变量$\phi_{ui,k}$，使得$y_{ui,k} = y_{ui}\phi_{ui,k}$。且$\sum_k \phi_{ui,k} = 1$

$$\phi_{ui,k} = \frac{\eta_{u,k}\theta_{i,k}}{y_{ui}}$$
$$\propto \exp\left(\log\left(E_q\left[\eta_{u,k}\theta_{i,k}\right]\right)\right)$$
$$= \exp\left(E_q\left[\log\left(\eta_{u,k}\theta_{i,k}\right)\right]\right)$$
$$\propto \log\left(\eta_{u,k}\right) + \log\left(\theta_{i,k}\right)$$

因为存在$\log\left(\eta_{u,k}\right) = \Psi\left(\lambda_{ua,k}\right) - \log\left(\lambda_{ub,k}\right)$，最终得到$\phi_{ui,k}$的迭代公式

$$\phi_{ui,k} \leftarrow \exp\left\{\Psi\left(\lambda_{ua,k}\right) - \log\left(\lambda_{ub,k}\right) + \Psi\left(\lambda_{ia,k}\right) - \log\left(\lambda_{ib,k}\right)\right\}$$

Poisson Factorization

$y_{ui,k} = y_{ui}\phi_{ui,k}$，且 Gamma 分布 $\theta_{i,k}$ 的期望为 $\frac{\lambda_{ia,k}}{\lambda_{ib,k}}$，因此可将 $(\lambda_{ua,k}, \lambda_{ub,k})$ 的迭代公式改写为

$$(\lambda_{ua,k}, \lambda_{ub,k}) \leftarrow \left(\lambda_{ua,k} + \sum_i y_{ui}\phi_{ui,k}, \lambda_{ub,k} + \sum_i \frac{\lambda_{ia,k}}{\lambda_{ib,k}}\right)$$

$$q(\beta, \theta, \xi, \eta, z) = \prod_{i,k} q(\beta_{ik} \mid \lambda_{ik}) \prod_{u,k} q(\theta_{uk} \mid \gamma_{uk}) \prod_u q(\xi_u \mid \kappa_u) \prod_i q(\eta_i \mid \tau_i) \prod_{u,i} q(z_{ui} \mid \phi_{ui}).$$

Poisson Factorization

Inference for HPF

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:

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

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

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

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

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

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

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

Empirical Study

Data sets:

- The Mendeley data set of scientific articles is a binary matrix of 80,000 users and 260,000 articles, with 5 million observations.
- The Echo Nest music data set is a matrix of 1 million users and 385,000 songs, with 48 million observations.
- The New York Times data set is a matrix of 1,615,675 users and 103,390 articles, with 80 million observations.
- The Netflix data set contains 480,000 users and 17,770 movies, with 100 million observations.

Empirical Study



Figure 4: Predictive performance on data sets. The top and bottom plots show normalized mean precision and mean recall at 20 recommendations, respectively. While the relative performance of the competing methods varies across data sets, HPF consistently outperforms each of them.

Empirical Study

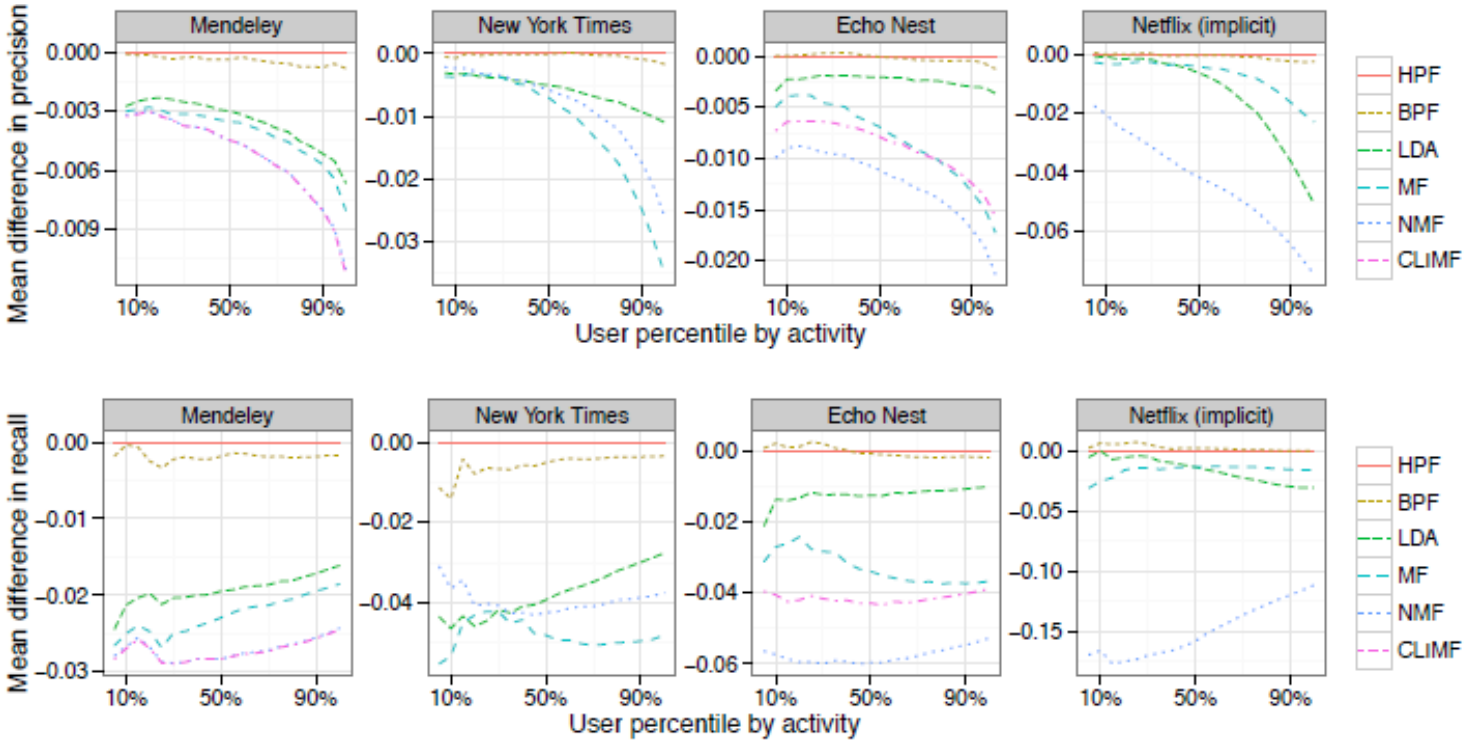


Figure 5: Predictive performance across users. The top and bottom plots show the mean difference in precision and recall to HPF at 20 recommendations, respectively, by user activity.

Thanks!