PMF

Introduction

현재까지의 Collaborative Filtering 모델들은 1) very large datasets 2) sparse한 데이터셋을 잘 모델링하지 못했다. 이 논문에서는 이러한 문제점을 해결한 Probabilistic Matrix Factorization 모델을 제안하고, 이에 더불어 비슷한 영화에 평점을 매긴 사람들은 비슷한 선호를 가질 것이라는 가정 하에 Constrained PMF를 제안한다.

PMF

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

- $N(x|\mu,\sigma^2)$: probability density function of the Gaussian Distribution
- I_{ij} : indicator function

Priors

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

- zero-mean spherical Gaussian Priors

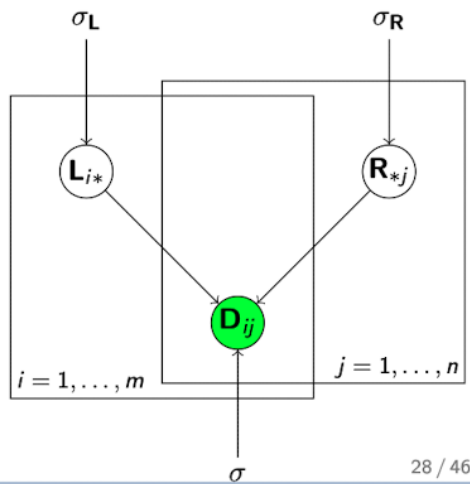
Posterior

$$\ln p(U,V|R,\sigma^2,\sigma_V^2,\sigma_U^2) = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}(R_{ij}-U_i^T V_j)^2 - \frac{1}{2\sigma_U^2}\sum_{i=1}^{N} U_i^T U_i - \frac{1}{2\sigma_V^2}\sum_{j=1}^{M} V_j^T V_j$$
$$-\frac{1}{2}\left(\left(\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}\right)\ln\sigma^2 + ND\ln\sigma_U^2 + MD\ln\sigma_V^2\right) + C, \qquad (3)$$

- Posterior를 Maximize하는 θ를 찾는 것이 MAP
- 이때, U, V 가 θ에 해당, Rating matrix가 $\mathbb{D}$에 해당

MAP

$$
\begin{aligned}
& p(\mathbf{L}, \mathbf{R} \mid \mathbf{D}, \sigma^2, \sigma_{\mathbf{L}}^2, \sigma_{\mathbf{R}}^2) \\
&= \frac{p(\mathbf{L}, \mathbf{R}, \mathbf{D} \mid \sigma^2, \sigma_{\mathbf{L}}^2, \sigma_{\mathbf{R}}^2)}{p(\mathbf{D} \mid \sigma^2, \sigma_{\mathbf{L}}^2, \sigma_{\mathbf{R}}^2)} \\
&\propto p(\mathbf{L}, \mathbf{R}, \mathbf{D} \mid \sigma^2, \sigma_{\mathbf{L}}^2, \sigma_{\mathbf{R}}^2) \\
&= p(\mathbf{D} \mid \mathbf{L}, \mathbf{R}, \sigma^2) p(\mathbf{L} \mid \sigma_{\mathbf{L}^2}) p(\mathbf{R} \mid \sigma_{\mathbf{R}}^2) \\
&\propto \left[\prod_{(i,j) \in \Omega} \exp\left(-\frac{(\mathbf{D}_{ij} - [\mathbf{L}\mathbf{R}]_{ij})^2}{2\sigma^2} \right) \right] \\
&\quad \cdot \left[\prod_{i,k} \exp\left(-\frac{\mathbf{L}_{ik}^2}{2\sigma_{\mathbf{L}}^2} \right) \right] \\
&\quad \cdot \left[\prod_{k,j} \exp\left(-\frac{\mathbf{R}_{ik}^2}{2\sigma_{\mathbf{R}}^2} \right) \right]
\end{aligned}
$$



$$
\begin{aligned}
&\mathrm{MAP}(\mathbf{L}, \mathbf{R} \mid \mathbf{D}, \sigma^2, \sigma_{\mathbf{L}}^2, \sigma_{\mathbf{R}}^2) \\
&= \underset{\mathbf{L},\mathbf{R}}{\operatorname{argmax}}\, p(\mathbf{L}, \mathbf{R}, \mid \mathbf{D}, \sigma^2, \sigma_{\mathbf{L}}^2, \sigma_{\mathbf{R}}^2) \\
&= \underset{\mathbf{L},\mathbf{R}}{\operatorname{argmin}} - \ln p(\mathbf{L}, \mathbf{R} \mid \mathbf{D}, \sigma^2, \sigma_{\mathbf{L}}^2, \sigma_{\mathbf{R}}^2) \\
&= \underset{\mathbf{L},\mathbf{R}}{\operatorname{argmin}} \frac{1}{2\sigma^2} \sum_{(i,j)\in\Omega} (\mathbf{D}_{ij} - [\mathbf{LR}]_{ij})^2 + \frac{1}{2\sigma_{\mathbf{L}}^2} \sum_{i,k} \mathbf{L}_{ik}^2 + \frac{1}{2\sigma_{\mathbf{R}}^2} \sum_{k,j} \mathbf{R}_{kj}^2 \\
&= \underset{\mathbf{L},\mathbf{R}}{\operatorname{argmin}} \sum_{(i,j)\in\Omega} (\mathbf{D}_{ij} - [\mathbf{LR}]_{ij})^2 + \frac{\sigma^2}{\sigma_{\mathbf{L}}^2} \sum_{i,k} \mathbf{L}_{ik}^2 + \frac{\sigma^2}{\sigma_{\mathbf{R}}^2} \sum_{k,j} \mathbf{R}_{kj}^2 \\
&= \underset{\mathbf{L},\mathbf{R}}{\operatorname{argmin}} \sum_{(i,j)\in\Omega} (\mathbf{D}_{ij} - [\mathbf{LR}]_{ij})^2 + \lambda_{\mathbf{L}} \|\mathbf{L}\|_F^2 + \lambda_{\mathbf{R}} \|\mathbf{R}\|_F^2
\end{aligned}
$$

- PMF + MAP = latent factor model with L2 regularization
- Precision $\lambda_{\mathbf{L}} = \sigma^2/\sigma_{\mathbf{L}}^2$ relates variation of noise and factors
- Similarly $\lambda_{\mathbf{R}} = \sigma^2/\sigma_{\mathbf{R}}^2$

PMF with sigmoid

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

- simple linear-Gaussian model을 사용하는 것보다 $g(x) = 1/(1 + exp(-x))$를 통과시키는 것이 더 성능이 좋다고 서술
- $1, 2, \cdots, K$ 까지의 rating을 $t(x) = (x-1)/(K-1)$ 에 통과시켜 같은 range로 맞춰줌

Constrained PMF

- users with very few ratings은 prior mean이나 average user에 가까워지는 경향 존재
- user-specific한 feature vector를 통해 infrequent users들에 대한 제약 가하는 방식

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

- $W \in R^{D \times M}$: latent similarity constraint matrix
- $i^{th}\ column\ of\ W$: 특정 영화를 평가한 것이 offset으로 더해짐 → 유사한 영화를 본 사용자들은 vector가 비슷해짐

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

Priors

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

MAP

$$\begin{aligned} E \quad = \quad & \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{M} I_{ij} \big(R_{ij} - g\big([Y_i + \frac{\sum_{k=1}^{M} I_{ik} W_k}{\sum_{k=1}^{M} I_{ik}}]^T V_j)\big)^2 \\ & + \frac{\lambda_Y}{2} \sum_{i=1}^{N} \parallel Y_i \parallel_{Fro}^2 + \frac{\lambda_V}{2} \sum_{j=1}^{M} \parallel V_j \parallel_{Fro}^2 + \frac{\lambda_W}{2} \sum_{k=1}^{M} \parallel W_k \parallel_{Fro}^2, \end{aligned}$$

- 일반적인 PMF에서 진행했던 동일한 MAP 과정을 진행