CSE 8803RS: Recommendation Systems

Multi-Domain Collaborative Filtering

Arun Kumar Chithanar

School of Computational Science & Engineering
College of Computing
Georgia Institute of Technology

Introduction

- One of the biggest challenges in recommender systems: Data Sparsity — especially with memory-based techniques
- Matrix factorization techniques have become popular in recent times — Nevertheless, sparsity limits performance
- One approach is to use data from multiple sources — Multi-Domain Collaborative Filtering (MCF)

Applications

- This technique is particularly suitable for large e-commerce and social networking services
  —Multiple domains
  —Same set of users
- E.g. Amazon
  —Domains: Books, Electorinic Goods, ...

Multi-Domain Collaborative Filtering

- Each Domain constiutes a prediction problem
- For the same set of users, we have an incomplete rating matrix in different domains
- Exploit the correlation between the domains to alleviate data sparsity
- This paper proposes a probabilistic framework to model each prediction problem and correlation among different domains

Multi-Domain Collaborative Filtering

- Consider K domains
  — $X^i \in R^{m_i x n_i}$- rating matrix for the ith domain. i = 1,...,K
  — total no of users: m
- $U^i \in R^{dxm}$ and $V^i \in R^{dxn_i}$

- Conditional distribution over the observed ratings on the ith domain:

$$p(\mathbf{X}^i|\mathbf{U}^i, \mathbf{V}^i, \sigma_i) = \prod_{j=1}^{m} \prod_{k=1}^{n_i} \left[\mathcal{N}(X^i_{jk}|(\mathbf{U}^i_j)^T \mathbf{V}^i_k, \sigma_i^2)\right]^{I^i_{jk}}$$

$$p(\mathbf{U}^i|\lambda_i) = \prod_{j=1}^{m} \mathcal{N}(\mathbf{U}^i_j|\mathbf{0}_d, \lambda_i^2 \mathbf{I}_d)$$

$$p(\mathbf{V}^i|\eta_i) = \prod_{k=1}^{n_i} \mathcal{N}(\mathbf{V}^i_k|\mathbf{0}_d, \eta_i^2 \mathbf{I}_d),$$

Correlation between different domains

- Use a matrix-variate normal distribution $U = vec(U^1), ..., vec(U^k)$

$$p(\mathbf{U}|\mathbf{\Omega}) = \mathcal{MN}_{md \times K}(\mathbf{U}|\mathbf{0}_{md \times K}, \mathbf{I}_{md} \otimes \mathbf{\Omega}),$$

- matrix variate normal distribution is defined as

$$p(\mathbf{X}|\mathbf{M}, \mathbf{A}, \mathbf{B}) = \frac{\exp\left(-\frac{1}{2}\mathrm{tr}\left(\mathbf{A}^{-1}(\mathbf{X}-\mathbf{M})\mathbf{B}^{-1}(\mathbf{X}-\mathbf{M})^T\right)\right)}{(2\pi)^{ab/2}|\mathbf{A}|^{b/2}|\mathbf{B}|^{a/2}},$$

Parameter Learning

- The log posterior over $U^i andV^i$ is given by

$$
\begin{aligned}
&\ln p(\{\mathbf{U}^i\}, \{\mathbf{V}^i\} | \{\mathbf{X}^i\}, \boldsymbol{\sigma}, \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Omega}) \\
&= -\sum_{i=1}^{K} \frac{1}{2\sigma_i^2} \sum_{j=1}^{m} \sum_{k=1}^{n_i} I_{jk}^i \left(X_{jk}^i - (\mathbf{U}_j^i)^T \mathbf{V}_k^i \right)^2 \\
&\quad - \sum_{i=1}^{K} \frac{1}{2\lambda_i^2} \sum_{j=1}^{m} (\mathbf{U}_j^i)^T \mathbf{U}_j^i - \sum_{i=1}^{K} \frac{1}{2\eta_i^2} \sum_{k=1}^{n_i} (\mathbf{V}_k^i)^T \mathbf{V}_k^i \\
&\quad - \frac{1}{2} \sum_{i=1}^{K} (\ln \sigma_i^2 \sum_{j=1}^{m} \sum_{k=1}^{n_i} I_{jk}^i) - \frac{md}{2} \sum_{i=1}^{K} \ln \lambda_i^2 \\
&\quad - \sum^{K} \frac{dn_i}{2} \ln \eta_i^2 - \frac{1}{2} \mathrm{tr}(\mathbf{U}\boldsymbol{\Omega}^{-1}\mathbf{U}^T) - \frac{md}{2} \ln |\boldsymbol{\Omega}| + \mathrm{Const}
\end{aligned}
$$

$$
\Omega_{ij} = \frac{1}{md} \left(\mathrm{vec}(\mathbf{U}^i) \right)^T \mathrm{vec}(\mathbf{U}^j)
$$

Parameter Learning

- After plugging in the equations in the negative of log posterior

$$
\begin{aligned}
&J(\{\mathbf{U}^i\}, \{\mathbf{V}^i\}, \boldsymbol{\sigma}, \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Omega}) \\
=&\sum_{i=1}^{K} \frac{1}{2\sigma_i^2} \sum_{j=1}^{m} \sum_{k=1}^{n_i} I_{jk}^i \left(X_{jk}^i - (\mathbf{U}_j^i)^T \mathbf{V}_k^i \right)^2 \\
&+ \frac{1}{2} \sum_{i=1}^{K} (\ln \sigma_i^2 \sum_{j=1}^{m} \sum_{k=1}^{n_i} I_{jk}^i) + \frac{md}{2} \sum_{i=1}^{K} \ln \left(\sum_{j=1}^{m} (\mathbf{U}_j^i)^T \mathbf{U}_j^i \right) \\
&+ \sum_{i=1}^{K} \frac{dn_i}{2} \ln \left(\sum_{k=1}^{n_i} (\mathbf{V}_k^i)^T \mathbf{V}_k^i \right) + \frac{1}{2(md)^{K-1}} \ln |\mathbf{U}^T \mathbf{U}|.
\end{aligned}
$$

$$
\ln \left(\sum_{j=1}^{m} (\mathbf{U}_j^i)^T \mathbf{U}_j^i \right) = \ln \operatorname{tr} \left(\mathbf{U}^i (\mathbf{U}^i)^T \right)
$$

$$
\ln \left(\sum_{k=1}^{n_i} (\mathbf{V}_k^i)^T \mathbf{V}_k^i \right) = \ln \operatorname{tr} \left(\mathbf{V}^i (\mathbf{V}^i)^T \right)
$$

Link Function

- Gaussian likelihood is not suitable for integral ratings.
  —Therefore, use a link function $g(:\theta)$
  —$Z^i_{jk} = g_i(X^i_{jk})$

$$p(\mathbf{Z}^i|\mathbf{U}^i,\mathbf{V}^i,\sigma_i) = \prod_{j=1}^{m}\prod_{k=1}^{n_i}\left[\mathcal{N}(Z^i_{jk}|(\mathbf{U}^i_j)^T\mathbf{V}^i_k,\sigma_i^2)\right]^{I^i_{jk}}$$

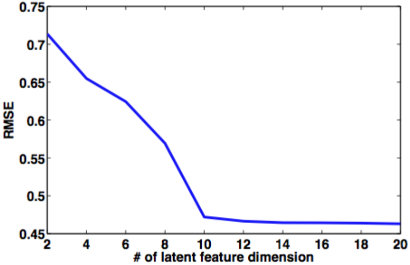
- Using Jacobian transformation

$$p(\mathbf{X}^i|\mathbf{U}^i,\mathbf{V}^i,\sigma_i) = \prod_{j=1}^{m}\prod_{k=1}^{n_i}\left[\mathcal{N}\left(g_i(X^i_{jk})|(\mathbf{U}^i_j)^T\mathbf{V}^i_k,\sigma_i^2\right)g_i'(X^i_{jk})\right]^{I^i_{jk}}$$

Experiments

- Datasets: MovieLens and Book-Crossing
- Evaluation Metric: Root Mean Squared Error
- Base Line Systems: Independent collaborative filtering using PMF and Collective Matrix Factorization (CMF)

Variation with Latent Feature Dimension



Comparison over MovieLens Dataset

Method	1st domain	2nd domain	3rd domain	4th domain	5th domain	Total
PMF	0.9642	1.2104	0.9377	1.0035	1.0352	1.0092
CMF	0.8272	0.7977	0.8120	0.7945	0.7987	0.8088
MCF	0.8061	0.7914	0.7907	0.7761	0.7859	0.7913
MCF-LF	**0.8017**	**0.7644**	**0.7806**	**0.7607**	**0.7504**	**0.7755**

Comparison over Book-Crossing Dataset

Method	1st domain	2nd domain	3rd domain	4th domain	5th domain	Total
PMF	0.9180	0.9795	0.8308	0.8699	0.8812	0.9269
CMF	0.9620	1.0207	0.9777	0.8465	1.0449	0.9960
MCF	0.7023	0.7046	0.7585	0.7555	0.7371	0.7158
MCF-LF	**0.5686**	**0.5791**	**0.6047**	**0.6001**	**0.5953**	**0.5811**

References

- Y. Zhang, B. Cao, and D. Y. Yeung. Multi-domain collaborative filtering. In Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI), Catalina Island, California, USA, 2010.
- R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, Advances in Neural Information Processing Systems 20, Vancouver, British Columbia, Canada, 2007