Seminar Talk

"Personalized Prediction and Sparsity Pursuit in Latent Factor Models" by Yunzhang Zhu, Xiaotong Shen, and Changqing Ye, (2016)

Chao Cheng

School of Statistics and Management

May 9, 2018

Outline

Outline

1. Recommender System
2. Partial Latent Models and Sparse Factorizations
   - Motivation and Model Settings
   - Methods
   - Theory
   - Benchmark: MovieLens Data
3. An Extension

Introduction

How to deliver information precisely and efficiently

- Classified directory
- Search engine
- Recommender system

Generic form

- User-Item preference matrix: $\boldsymbol{R} = [r_{ui}]_{U \times I}$.
  r_{ui} is ordered or unordered categorical or continuous, rating, score, etc.
  Each row represents a user. Each column represents an item.
  Observe r_{ui} only over subset Ω.
- Goal:
  - Complete $\boldsymbol{R}$: Rating prediction.
  - Make recommendations: Top-N recommendation.

Evaluation

- Root Mean Squared Error(RMSE):

$$rmse = \sqrt{\frac{1}{n_{test}} \sum_{test} (r_{ui} - \hat{r}_{ui})^2}$$

- Mean Absolute Error:

$$rmse = \frac{1}{n_{test}} \sum_{test} |r_{ui} - \hat{r}_{ui}|$$

- Many other criteria:
  - Recall: $recall = \frac{\sum_{u \in U} |R(u) \bigcap T(u)|}{\sum_{u \in U} |T(u)|}$
  - Precision: $precision = \frac{\sum_{u \in U} |R(u) \bigcap T(u)|}{\sum_{u \in U} |R(u)|}$
  - Coverage: $coverage = \frac{\bigcup_{u \in U} |R(u)|}{|I|}$
  - Gini Index, Diversity, Serendipity, Trustworthy, Transparency, Robustness, Real-Time, etc.

General Methods

Filtering: Measure similarity

- Collaborative Filtering(User-CF):
  - Find similar users.
- Content-based Filtering(Item-CF):
  - Find similar items.
- Hybrid

					
Tom	?	4	5	?	?
Jerry	?	?	?	3	1
Denny	2	5	?	?	?
Sarah	?	?	5	?	?
Edwin	?	?	?	?	4

- Latent Factor Model(LFM):

$$\boldsymbol{R}_{U\times I} = \boldsymbol{P}_{U\times K}\boldsymbol{Q}^T_{I\times K}$$

The idea comes from text mining when researchers tring to find latent topic from text. Here, LFM is used to find comon latent class of users and items.
The user-item interactions are modeled as inner product.

LF 1		LF 2		LF 3	
Title	Genre	Title	Genre	Title	Genre
Catwalk (1995)	Documentary	In the Line of Duty 2 (1987)	Action	The Gay Deceivers (1969)	Comedy
See the Sea (1997)	Thriller	Simon Sez (1999)	Action	The Acid House (1998)	Comedy
The Secret Agent (1996)	Thriller	Taffin (1988)	Action	Mad Dog Time (1996)	Comedy

Table: LFM on MovieLen 1M

General Methods

Matrix factorization and low-rank approximation

表2-13 MovieLens数据集中根据LFM计算出的不同隐类中权重最高的物品

1（科幻、惊悚）	3（犯罪）	4（家庭）	5（恐怖、惊悚）
《隐形人》（The Invisible Man，1933）	《大白鲨》（Jaws，1975）	《101真狗》(101 Dalmatians，1996)	《女巫布莱尔》（The Blair Witch Project，1999）
《科学怪人大战狼人》（Frankenstein Meets the Wolf Man，1943）	《致命武器》（Lethal Weapon，1987）	《回到未来》（Back to the Future，1985）	《地狱来的房客》（Pacific Heights，1990）
《哥斯拉》（Godzilla，1954）	《全面回忆》（Total Recall，1990）	《土拨鼠之日》（Groundhog Day，1993）	《异灵骇客2之恶灵归来》（Stir of Echoes: The Homecoming，2007）
《星球大战3：武士复仇》（Star Wars: Episode VI - Return of the Jedi，1983）	《落水狗》（Reservoir Dogs，1992）	《泰山》（Tarzan，2003）	《航越地平线》（Dead Calm，1989）
《终结者》（The Terminator，1984）	《忠奸人》（Donnie Brasco，1997）	《猫儿历险记》（The Aristocats，1970）	《幻象》（Phantasm，1979）
《魔童村》（Village of the Damned，1995）	《亡命天涯》（The Fugitive，1993）	《森林王子2》（The Jungle Book 2，2003）	《断头谷》（Sleepy Hollow，1999）
《异形》（Alien，1979）	《夺宝奇兵3》（Indiana Jones and the Last Crusade，1989）	《当哈利遇到莎莉》（When Harry Met Sally…，1989）	《老师不是人》（The Faculty，1998）
《异形2》（Aliens，1986）	《威胁2 :社会》（Menace II Society，1993）	《蚁哥正传》（Antz，1998）	《苍蝇》（The Fly，1958）
《天魔续集》（Damien: Omen II，1978）	《辣手神探》（Lashou shentan，1992）	《小姐与流浪汉》（Lady and the Tramp，1955）	《鬼哭神嚎》（The Amityville Horror，1979）
《魔鬼怪婴》（Rosemary's Baby，1968）	《真实罗曼史》（True Romance，1993）	《飞天法宝》（Flubber，1997）	《深渊》（The Abyss，1989）

Figure: Another result of the method, but with different algorithms and parameter settings.

Challenges

- Big Data: For Movielen dataset, number of rating observations ranges from 1m to 20m. In practical situation, this could easily be above 1B.
- Yet sparsity: For Movielen 1M, there are 6000 users and 4000 movies. Observational ratio is 4%. For Movielen 20M, there are 138000 users and 27000 movies, the ratio is 0.5%.
  Also these missing are non-ignorable.
- Cold start problem: How to make recommendations for a new user or item?
- How to define similarities? How to determine number of latent factors?

Outline

Outline

Partial Latent Models

We link $r_{ij} = G(\theta_{ij})$ with preference probability or mean parameter in a generalized linear model, where $G(\cdot)$ is a link function.

$$\theta_{ij} = \boldsymbol{x}_i^T \alpha + \boldsymbol{y}_j^T \beta + \boldsymbol{a}_i^T \boldsymbol{b}_j \quad (1)$$

where

$$\boldsymbol{x}_i = \left(x_{i1}, \cdots, x_{iU_0}\right)^T \quad , \quad \text{User-specific vector}$$
$$\boldsymbol{y}_j = \left(y_{j1}, \cdots, y_{jM_0}\right)^T \quad , \quad \text{Content-specific vector}$$

and

$$\boldsymbol{a}_i = (a_{i1}, \cdots, a_{iK})^T \quad , \quad \text{Unobserved user latent vector}$$
$$\boldsymbol{b}_j = (b_{j1}, \cdots, b_{jK})^T \quad , \quad \text{Unobserved item latent vector}$$

Partial Latent Models

Equation(1) can be expressed in a matrix form

$$\Theta = \boldsymbol{A}\boldsymbol{B}^T, \qquad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{x}_1^T & \beta^T & \boldsymbol{a}_1^T \\ \boldsymbol{x}_1^T & \beta^T & \boldsymbol{a}_1^T \\ \vdots & \vdots & \vdots \\ \boldsymbol{x}_U^T & \beta^T & \boldsymbol{a}_U^T \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} \boldsymbol{y}_1^T & \alpha^T & \boldsymbol{b}_1^T \\ \boldsymbol{y}_1^T & \alpha^T & \boldsymbol{b}_1^T \\ \vdots & \vdots & \vdots \\ \boldsymbol{y}_I^T & \alpha^T & \boldsymbol{b}_I^T \end{pmatrix} \qquad (2)$$

The decomposition is latent if

$$\Theta_0 = \bar{\boldsymbol{A}}\bar{\boldsymbol{B}}^T, \quad r\left(\Theta_0\right) = r\left(\bar{\boldsymbol{A}}\right) = r\left(\bar{\boldsymbol{B}}\right) = r_0 \leq K$$

where $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{B}}$ are $U \times K$ and $I \times K$ matrices having the same locations of zero-columns simultaneously.

Sparse Latent Factorizations

$$\Theta_0 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 0.5 & 0 \\ 0.25 & 0.25 \end{pmatrix}$$
$$= \begin{pmatrix} -1.376 & -0.325 \\ -0.851 & 0.526 \end{pmatrix} \begin{pmatrix} -1.376 & -0.851 \\ -0.325 & 0.526 \end{pmatrix}$$

- The latent factorization in 1st row is sparser than that in the 2nd row. And we seek a sparsest latent factorization

$$(\boldsymbol{A}_0, \boldsymbol{B}_0) = \underset{\Theta_0 = \bar{\boldsymbol{A}}_0 \bar{\boldsymbol{B}}_0^T}{\operatorname{argmin}} \left(\|\bar{\boldsymbol{A}}_0\|_0 + \|\bar{\boldsymbol{B}}_0\|_0 \right)$$

  The L_0 norm could be replaced by other penalty, and the pursuit of sparsest factorization is achieved by the means of identifying zero entries of $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$.
- Identifiability issue due to scaling when developing the algorithm.

Outline

The negative log-likelihood(loss function) is

$$\sum_{(i,j)\in\Omega} l\left(r_{ij}, \boldsymbol{x}_i^T\alpha + \boldsymbol{y}_j^T\beta + \boldsymbol{a}_i^T\boldsymbol{b}_j\right)$$

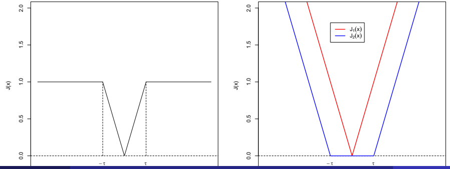
- No interactions of main effects.
- Ignorable missing.
- (Conditional) independent observations.

The choice of $l(\cdot,\cdot)$ depends on models underlying the observed data. This paper's strategy is to work with the likelihood of incomplete data.

$$S_q\left(\alpha, \beta, \tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}\right) = \sum_{(i,j)\in\Omega} l\left(r_{ij}, \boldsymbol{x}_i^T\alpha + \boldsymbol{y}_j^T\beta + \boldsymbol{a}_i^T\boldsymbol{b}_j\right) + \lambda\left(\sum_{i=1}^{U}\left\|\boldsymbol{a}_i\right\|_q + \sum_{j=1}^{I}\left\|\boldsymbol{b}_j\right\|_q\right) \quad (3)$$

where $q = 0, 1$. When $q = 2$, the real penalty is $\|\cdot\|_2^2$. For efficient computation, the L_0 norm is replaced by its continuous surrogate, the truncated L_1 function $J(u) = \frac{1}{\tau}\min(|u|, \tau)$ Shen, Pan, and Zhu(2012).

$$\|\boldsymbol{z}\|_0 = \sum_{k=1}^{K} J(z_k)$$



- Update $\hat{\boldsymbol{a}}_i$ given the rest; $i = 1, 2, \cdots, U$ by minimizing

$$\sum_{j \in R_{i.}} l\left(r_{ij}, \boldsymbol{x}_i^T \hat{\alpha} + \boldsymbol{y}_j^T \hat{\beta} + \boldsymbol{a}_i^T \hat{\boldsymbol{b}}_j\right) + \lambda \left\|\boldsymbol{a}_i\right\|_q \tag{4}$$

  where $R_{i.}$ is observations user i has rated.
- Update $\hat{\boldsymbol{b}}_j$ given the rest; $j = 1, 2, \cdots, I$ by minimizing

$$\sum_{i \in R_{.j}} l\left(r_{ij}, \boldsymbol{x}_i^T \hat{\alpha} + \boldsymbol{y}_j^T \hat{\beta} + \hat{\boldsymbol{a}}_i^T \boldsymbol{b}_j\right) + \lambda \left\|\boldsymbol{b}_j\right\|_q \tag{5}$$

- Update main effects $\left(\hat{\alpha}, \hat{\beta}\right)$ given rest by minimizing

$$\sum_{i \in R_{.j}} l\left(r_{ij}, \boldsymbol{x}_i^T \alpha + \boldsymbol{y}_j^T \beta + \hat{\boldsymbol{a}}_i^T \hat{\boldsymbol{b}}_j\right) \tag{6}$$

Let $\boldsymbol{s}_k$ and $\boldsymbol{t}_k$ be the k-th column of $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{B}}$ in equation(3), $k = 1, 2, \cdots, K$. And $c_k > 0$ is a scalar. Then replacing them with $c_k \boldsymbol{s}_k$ and $c_k^{-1} \boldsymbol{t}_k$ will not change the value of the loss function part. Also it won't change the predicited scores. But the penalty function part changes.

- For L_1 norm, note that

$$\sum_{i=1}^{U} \|\boldsymbol{a}_i\|_1 + \sum_{j=1}^{I} \|\boldsymbol{b}_j\|_1 = \sum_{k=1}^{K} (\|\boldsymbol{s}_k\|_1 + \|\boldsymbol{t}_k\|_1)$$

  Then

$$\|c_k \boldsymbol{s}_k\|_1 + \left\|c_k^{-1} \boldsymbol{t}_k\right\|_1 = c_k \|\boldsymbol{s}_k\|_1 + c_k^{-1} \|\boldsymbol{t}_k\|_1 \geq 2\sqrt{\|\boldsymbol{s}_k\|_1 \|\boldsymbol{t}_k\|_k}$$

  Equality holds $\iff c_k \|\boldsymbol{s}_k\|_1 = c_k^{-1} \|\boldsymbol{t}_k\|_1 \iff c_k = \sqrt{\frac{\|\boldsymbol{t}_k\|_1}{\|\boldsymbol{s}_k\|_1}}$

- For L_2 norm, the result is similar:

$$\sum_{i=1}^{U} \|\boldsymbol{a}_i\|_2^2 + \sum_{j=1}^{I} \|\boldsymbol{b}_j\|_2^2 = \sum_{k=1}^{K} \left(\|\boldsymbol{s}_k\|_2^2 + \|\boldsymbol{t}_k\|_2^2 \right)$$

$$\|c_k \boldsymbol{s}_k\|_2^2 + \left\| c_k^{-1} \boldsymbol{t}_k \right\|_2^2 = c_k^2 \|\boldsymbol{s}_k\|_2^2 + c_k^{-2} \|\boldsymbol{t}_k\|_2^2 \geq 2\sqrt{\|\boldsymbol{s}_k\|_2^2 \|\boldsymbol{t}_k\|_2^2}$$

The equality holds $\iff c_k^2 \|\boldsymbol{s}_k\|_2^2 = c_k^{-2} \|\boldsymbol{t}_k\|_2^2 \iff c_k = \frac{\|\boldsymbol{s}_k\|_2}{\|\boldsymbol{t}_k\|_2}$

Algorithm 1

L_1 method with missing values

1. Initialization. Input ratings r_{ij}, the upper bound K, tuning parameter λ, initial values for $(\alpha, \beta, \boldsymbol{A}, \boldsymbol{B})$
2. For each user, solve (4) to update $\hat{\boldsymbol{a}}_i$, $i = 1, 2, \cdots, U$.
3. For each item, solve (5) to update $\hat{\boldsymbol{b}}_j$, $j = 1, 2, \cdots, I$.
4. Solve (6) to update main effects $\left(\hat{\alpha}, \hat{\beta}\right)$.
5. Check and update the maximum improvement.
6. Apply equal scaling strategy if $\hat{\boldsymbol{A}}$ or $\hat{\boldsymbol{B}}$ is updated.
7. Iterate until stopping conditions met.

Outline

Convergence properties of the algorithms

Lemma

The estimate $\left(\alpha^{L_1}, \beta^{L_1}, \tilde{\boldsymbol{A}}^{L_1}, \tilde{\boldsymbol{B}}^{L_1}\right)$ *and* $\left(\alpha^{L_0}, \beta^{L_0}, \tilde{\boldsymbol{A}}^{L_0}, \tilde{\boldsymbol{B}}^{L_0}\right)$*, computed from Algorithm 1 and a* L_0 *version of Algorithm 1, are stationary points of* $S_1\left(\alpha, \beta, \tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}\right)$ *and* $S_0\left(\alpha, \beta, \tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}\right)$*, respectively.*

Notations and assumptions

- Degree of sparsity s_q:

$$s_q = \min_{\Theta_0 = \boldsymbol{A}\boldsymbol{B}^T} \left(\|\boldsymbol{A}\|_q + \|\boldsymbol{B}\|_q \right), \qquad q = 0, 1$$
$$s_2 = \min_{\Theta_0 = \boldsymbol{A}\boldsymbol{B}^T} \left(\|\boldsymbol{A}\|_2^2 + \|\boldsymbol{B}\|_2^2 \right)$$

- Parameter space: $\mathcal{F} = \left\{ \theta = \boldsymbol{A}\boldsymbol{B}^T : \|\boldsymbol{A}\|_\infty \leq L, \|\boldsymbol{B}\|_\infty \leq L, \boldsymbol{A} \in \mathbb{M}(U, K), \boldsymbol{B} \in \mathbb{M}(I, K) \right\}$ where $\mathbb{M}(M, N)$ is a class of $M \times N$ matrices, $L > 0$ is a constant.

- Hellinger-distance $h(\Theta_1, \Theta_2) = (UI)^{-1} \sum_{i=1}^{U} \sum_{j=1}^{I} h\left(\theta_{ij}^1, \theta_{ij}^2\right)$, where

$$h\left(\theta_{ij}^1, \theta_{ij}^2\right) = \int \left(f^{1/2}\left(r_{ij}, z_{ij}, \theta_{ij}^1\right) - f^{1/2}\left(r_{ij}, z_{ij}, \theta_{ij}^2\right) \right)^2 \mathrm{d}\mu\left(r_{ij}, z_{ij}\right)$$

Assumption A

For some constant $d_0 > 0$, any θ_{ij}^1 and θ_{ij}^2,

$$\left| f^{1/2}\left(r_{ij}, z_{ij}, \theta_{ij}^1\right) - f^{1/2}\left(r_{ij}, z_{ij}, \theta_{ij}^2\right) \right| \leq G\left(r_{ij}, \delta_{ij}\right) \left|\theta_{ij}^1 - \theta_{ij}^2\right|$$

with $\sup_{1 \leq i \leq U, 1 \leq j \leq I} EG\left(r_{ij}, \delta_{ij}\right) \leq d_0$

Error bound for L_0 method

Theorem (1)

Under Assumption A, for $\hat{\Theta}^{L_0}$, if $K \geq r_0 \equiv r(\Theta_0)$, then there exists a constant $c_1 > 0$, such that for $(|\Omega|, U, I)$,

$$P\left(h\left(\hat{\Theta}^{L_0}, \hat{\Theta}_0\right) \geq \epsilon_{0,|\Omega|}\right) \leq 4\exp\left(-c_1 |\Omega| \epsilon^2_{0,|\Omega|}\right) \tag{7}$$

provided that $\lambda = s_0^{-1} c_3 \epsilon^2_{0,|\Omega|}$, where $\epsilon^2_{0,|\Omega|} = \log\left(\frac{(U+I)K}{s_0}\right)\frac{s_0}{|\Omega|}$, which is $\log\left(\frac{(U+I)r_0}{s_0}\right)\frac{s_0}{|\Omega|}$ when $K = r_0$ is tuned. As $|\Omega|$, U, $I \to \infty$.

Error bound for L_1 and L_2 method

This is a parallel result of Theorem 1

$$\epsilon_{1,|\Omega|} = \sqrt{\frac{s_1^2 \log\left(\left(U+I\right)K\right)}{|\Omega|}}$$

$$\epsilon_{2,|\Omega|} = \sqrt{\frac{\left(U+I\right)K \log s_2}{|\Omega|}}$$

In summary

$$\epsilon_{0,|\Omega|} \preceq \epsilon_{|\Omega|}^{\text{mat}} \preceq \epsilon_{1,|\Omega|} \preceq \epsilon_{2,|\Omega|}$$

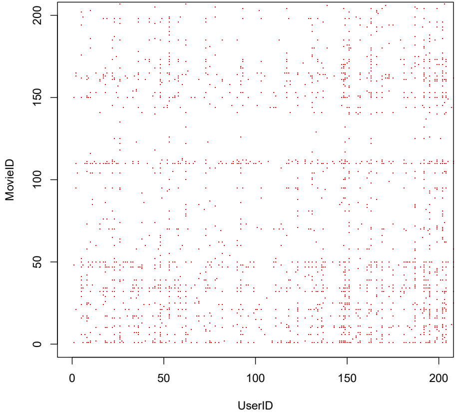
Here $a_n \preceq b_n$ means $a_n \leq cb_n$ for some $c > 0$, for all sufficiently large n.

Outline

Data description

- Data collected by GroupLens Research. 1M MovieLens: 100209 ratings(1-5) from 6040 users on 3900 movies.
- Predictors:
  - user-related covariates: gender, age, occupation and zip-code
  - content-related covariates: movie genres
- There are also 10M, 20M at their website.

Red points are observed ratings. UserID and MovieID both range from 1 to 200.



Data description
Nonignorable Missing

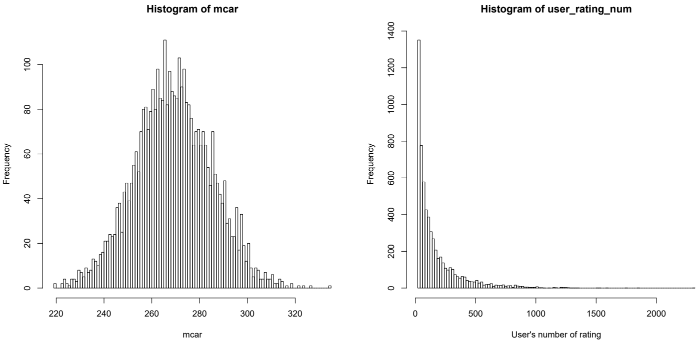


Figure: User's number of ratings compared with MCAR number of ratings

Data description
Nonignorable Missing

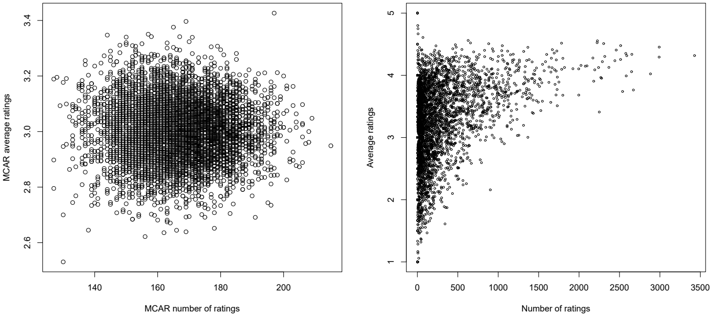


Figure: Movie's rating pattern compared with MCAR

Benchmark Result

My reproduction of L_1 method gives $rmse = 0.95$. But there are some problems in the computation.
The result of the paper is:

Method	1M MovieLens	10M MovieLens
Constant only	1.1186	1.0214
Predictors only	1.0960	1.0025
Soft-Imput	1.0656	1.0175
Agarwal & Chen	1.0520	1.0185
L_2 w/o predictors	1.0577	1.0152
L_1 w/o predictors	1.0523	1.0105
L_0 w/o predictors	1.0502	1.0104
L_2 w predictors	1.0516	1.0023
L_1 w predictors	1.0480	0.9998
L_0 w predictors	1.0478	0.9995

Figure: RMSE's of various methods

Outline

Smooth Neighbourhood Recommender Systems

Consider this cost function for a smooth neighbourhood recommender system

$$\mathcal{L}(\boldsymbol{P}, \boldsymbol{Q}) = \frac{1}{UI} \sum_{i=1}^{U} \sum_{j=1}^{I} \left(\sum_{(i',j') \in \Omega} \omega_{ij,i'j'} \left(r_{i'j'} - \boldsymbol{p}_i^T \boldsymbol{q}_j \right)^2 \right) + \lambda_1 \sum_{i=1}^{U} P(\boldsymbol{p}_i) + \lambda_2 \sum_{j=1}^{I} P(\boldsymbol{q}_j) \tag{8}$$

and the weight function $\omega_{ij,i'j'}$ is defined as

$$\omega_{ij,i'j'} = \frac{\mathcal{K}_h(\boldsymbol{x}_{ij}, \boldsymbol{x}_{i'j'}) S_{ij}^{i'j'}}{\sum_{(i'j') \in \Omega} \mathcal{K}_h(\boldsymbol{x}_{ij}, \boldsymbol{x}_{i'j'}) S_{ij}^{i'j'}} \tag{9}$$

where $\mathcal{K}_h(\cdot)$ is a kernel function and $S_{ij}^{i'j'}$ is another measure of similarity between (i, j) and (i', j'). Source of these information could be other than the rating-related data.

Some thoughts

- Can deal with cold start problem.
- Complete the rating matrix to make estimation and prediction.
- Each subproblem needs more computational rescourse.
- The performance releys on not only the method itself, but also the information of the outside network.

We want to make recommendations for students about which courses are suitable for he/she.

- Goal: improve students' score, hoping they will be more competitive when appling for jobs.
- Rating: students' score(or taecher evaluation).
- User: students, gender, major, book renting history, etc.
- Item: courses(just?), school and department information, teacher's information
- Cold start: the cold start ratio of students is about 25%.
- Practical issues: data storage and transmission, data cleanning, monitoring task table, APIs

An example

Representive courses of latent factors

yyz1	yyz2	yyz3	yyz4	yyz5
二维动画设计与制作&李小	面向对象程序设计&阳万安	声乐4&未知教师	专业实习&陈世海	经济法&曾品红
Android课程设计&唐远翔	多媒体数据库系统&阳万安	土木工程材料&殷涛	旅游产品开发与设计&王莎	广告学&曾品红
网站设计与开发实训&曹莉	程序设计基础&阳万安	视唱练耳2&未知教师	写生与摄影&王莎	经济法&陈云岗
数字图像处理&李朝荣	数据库原理及应用&阳万安	基础乐理2&未知教师	专业外语&海萍	经济法&洪叶
计算机网络安全&赵灵锴	面向对象程序设计—(面向对	曲式与作品分析1&未知教师	数学教学设计&郭家庆	广告学&徐向峰
面向对象课程设计&唐远翔	刑法学总论&雷安军	中国古典舞基本功训练4&未	PLC实验与设计&张雪平	管理学原理&曾品红
动画技术与应用&李小美	微观经济学&杨波	财务管理&郭蕾	声乐小组课2&夏毅	组织行为学&徐向峰
速写&祝正锋	光学&段志春	成本管理会计&郭蕾	家具设计&王松	经济法&吴新华
Java课程设计&唐自力	法理学初阶&雷安军	剧目2&未知教师	器乐演奏1&侯凌燕	市场营销学&未知教师
软件产品设计&未知教师	数字电子技术&高曾辉	合唱与指挥常识2&未知教师	动物分类学&李操	财务管理学&冯丽丽
艺术考察&李波	环境微生物实验&未知教	形势与政策2&周跃军	公共关系学&未知教师	高级日语听说2&万玲玲
法律协会活动&寇杪	试验设计与优化&赵成	形势与政策1&李玉芳	网页设计&唐远翔	学前教育科研方法&张国平
旅游产品开发与设计&王海	食品科学专业综合实验&张	毛泽东思想与中国特色社会	数学史&未知教师	大学物理&黄兴勇
商业展示及设施设计&黎沛	毕业论文&曾安平	毛泽东思想与中国特色社会	人力资源管理&曾品红	实变函数&张正亮
植物学教学实习&王宇	毕业论文&王勇	形势与政策1&黄诗玉	网页设计&赵灵锴	大学物理&曾志强
毕业创作&未知教师	毕业论文&张春	形势与政策1&高琴	网页设计&未知教师	环境景观设计&尤川宝
社会工作活动设计&杨东	毕业论文&陈支那	形势与政策1&周跃军	线性代数与概率统计&刘金	装置与装饰艺术&郭莉
行政诉讼模拟法庭&王夏玮	毕业论文&邹良洁	毛泽东思想与中国特色社会	线性代数与概率统计&龙希	证据法学理论与实务&古强
知识产权审判实践演练&王	毕业论文&吴宦熙	中国近现代史纲要&李玉芳	中小学教育研究方法&未知	国际法学&刘霓
知识产权案例教学&王海燕	基础生物学实验A4&邓鸳远	教育心理学&曲燕	中小学教育研究方法&张国	自动控制原理&周桂宇

Figure: Representvie courses of latent factors

An example

Top-n recommendations

id	xh	tjkc	js	yccj
0	141207001	毕业实习	陈世海	89.6426999072417
1	141207001	毕业实习	未知教师	89.0735635677487
2	141207001	毕业实习	赵跃	87.9233436907185
3	141207001	毕业实习	寇彬	87.7157554621421
4	141207001	毕业舞蹈专场汇报	未知教师	84.5361109123564
5	141207001	毕业论文	陈世海	83.6103365585082
6	141207001	毕业论文	段志高	83.427790464891
7	141207001	环境教育	魏贵玉	83.111115905422
8	141207001	毕业论文	吴宦熙	[illegible]261656114
9	141207001	毕业论文	李彦	83.0393028371424

Figure: An example of top-10 recommendations

id	xh	tjkc	js	yccj
3592	161209016	合唱2	未知教师	91.0835536121502
3593	161209016	三维动画设计与...	李小美	90.4380339011638
3594	161209016	田野采风	未知教师	90.2814582027063
3595	161209016	网站设计与开发...	曹莉兰	90.0129582874451
3596	161209016	艺术实践4	未知教师	89.8415930195798
3597	161209016	N2专训	薛春	89.7671985783883
3598	161209016	专业见习	未知教师	89.5273677595598
3599	161209016	英汉语言对比	李丽娟	89.4795195580549
3600	161209016	化工工艺学	唐红梅	89.1415924713858
3601	161209016	智能技术实践2	未知教师	89.1158577194814

Figure: Another example of top-10 recommendations. This student's major is product design.