Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence

Recommendation by Mining Multiple User Behaviors with Group Sparsity

Ting Yuan, Jian Cheng, Xi Zhang, Shuang Qiu, Hanqing Lu
National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Science
{tyuan, jcheng, xi.zhang, shuang.qiu, luhq}@nlpr.ia.ac.cn

Abstract

Recently, some recommendation methods try to improve the prediction results by integrating information from user's multiple types of behaviors. How to model the dependence and independence between different behaviors is critical for them. In this paper, we propose a novel recommendation model, the Group-Sparse Matrix Factorization (GSMF), which factorizes the rating matrices for multiple behaviors into the user and item latent factor space with group sparsity regularization. It can (1) select out the different subsets of latent factors for different behaviors, addressing that users' decisions on different behaviors are determined by different sets of factors; (2) model the dependence and independence between behaviors by learning the shared and private factors for multiple behaviors automatically ; (3) allow the shared factors between different behaviors to be different, instead of all the behaviors sharing the same set of factors. Experiments on the real-world dataset demonstrate that our model can integrate users' multiple types of behaviors into recommendation better, compared with other state-of-the-arts.

1 Introduction

To deal with the information overload, recommender systems have emerged by suggesting users the potential enjoyed items. As the most widely studied methods, Collaborative Filtering (CF) techniques make predictions by mining users' historical behaviors on items (Sarwar et al. 2001; Koren 2008; Salakhutdinov and Mnih 2008). The historical data are usually represented by a user-item rating matrix, which is typically extremely sparse.

Matrix Factorization (MF) models (Salakhutdinov and Mnih 2008; Koren 2008) are the state-of-the-art CF methods. Based on the premise that users' tastes can be represented by a small number of factors, they factorize the rating matrix into two low-rank matrices that represent the latent factors for users and items. In other words, they map both users and items to a common latent factor space $\mathbb{R}^k$, where each dimension encodes a latent factor that determines users' decisions on items. Characterizing users and items by these latent factors, prediction can be made based on them. Traditionally, MF methods deal with single type of user behavior for each recommendation task.

However, with the prevalence of massive web applications, users often have various types of behaviors on the web, varying from providing scores for movies, joining communities, to favoring music or establishing friendships with others. Considering simultaneously multiple behaviors of an user may be helpful to model the user's taste better. Recently, some works have been proposed to address this issue (Singh and Gordon 2008; Zhang, Cao, and Yeung 2010; Li, Yang, and Xue 2009; Hu et al. 2013), where they attempt to model the dependencies between different types of behaviors, and transfer information between behaviors to improve the recommendation results. Among them, the most widely used method is Collective Matrix Factorization (CMF) (Singh and Gordon 2008), which decomposes the rating matrices for different types of user behaviors together, and transfers information by sharing the same user latent factor matrix across different behaviors. Inspired by the idea of CMF, some following works (Ma et al. 2008; Yang et al. 2011; Krohn-Grimberghe et al. 2012) have demonstrated that by sharing the same user latent factors across multiple types of behaviors, better predictions can be achieved than utilizing single type of user behavior. Nevertheless, a main issue of them is that all dimensions of the latent factors are shared when user make decisions on different behaviors.

As we know, when people conduct different behaviors, their decisions are determined by different sets of factors. Not all the factors can be shared across different behaviors. To explain it clearly we give a toy example in Figure 1. Supposing each user is characterized by four factors, such as profession, gender, whether he/she is an adventure enthusiast and whether she/he is a romantic. When watching movie, whether she/he likes an action movie or a love story may be affected by factors 2, 3 and 4; when listening to music, whether she/he likes a light music or heavy music is affected by factors 2 and 4; when establishing friendship with others, the decisions may be determined by factors 1 and 2. Among these factors, some are commonly shared by multiple types of behaviors while some are specifically private for certain behavior. Additionally, the shared factors between each two types of behaviors may also be different. In the literature, however, the shared and private factors are seldom distin-

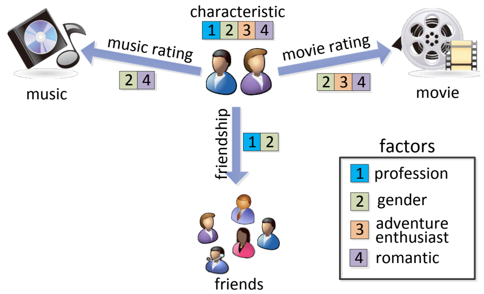

Figure 1: A toy example where users' decisions on different behaviors are affected by different sets of factors.

guished to model the independent parts of different behaviors, which may make improper information transferred and harm the recommendation results.

In this paper, we propose to integrate multiple types of user behaviors into recommendation effectively by modeling the shared and private factors among them. In particular, we present a novel recommendation model, named Group-Sparse Matrix Factorization (GSMF), which factorizes the rating matrices for multiple behaviors into the user and item latent factor matrices with group sparsity regularization. Addressing that users' decisions on different behaviors are determined by different sets of factors, we encourage each type of behaviors to use a subset of the latent factors. Inspired by the idea of group sparsity (Bengio et al. 2009; Jia, Salzmann, and Darrell 2010; Yuan and Lin 2006), we exploit mixed-norm regularization to induce a sparse representation at the level of groups and thus select out the different subsets of the latent factors for multiple types of behaviors. As a consequence, for each type of behaviors, the latent factors are factorized into shared parts which transfer the related information with other types of behaviors, and private parts which model the independent information of it. Notice that, our method can learn these shared and private factors automatically. Furthermore, it allows the shared factors between each two types of behaviors to be different. To our knowledge, this is the first work to investigate group sparsity for analyzing multiple behaviors in recommendation. Experiments show that our method can achieve better recommendation results than other state-of-the-arts.

2 Related Work

(Li, Yang, and Xue 2009) establishes a codebook which represents the cluster-level user-item rating patterns between two types of behaviors to transfer information. However, it is always difficult to seek common patterns between different behaviors with highly sparse data. (Pan et al. 2010) assumes that the same entity has similar latent factors when participating in different activities. Thus, they learn user latent factor matrix $\mathbf{U}_A$ from an auxiliary rating matrix first and then generate the user latent factor matrix $\mathbf{U}_T$ for the target behavior with the regularization of penalizing the divergence between $\mathbf{U}_T$ and $\mathbf{U}_A$. However, above methods cannot be applied to the scenarios of multiple (more than two) behaviors as studied in this paper.

In (Berkovsky, Kuflik, and Ricci 2007) an early neighborhood based method integrating users' ratings on multiple domains is introduced, but it can easily fail to find similar users or items when data is very sparse. Collective Matrix Factorization (CMF) (Singh and Gordon 2008) can address the problem over multiple types of user behaviors by sharing the same user latent factors when decomposing the rating matrices for different behaviors. Recently, (Ma et al. 2008; Yang et al. 2011; Krohn-Grimberghe et al. 2012) have employed CMF for recommendation in social rating networks, where they share the same user latent factors when factoring the rating matrix and social network matrix together to transfer the dependencies between ratings and friendships. However, when sharing the same user latent factors, they do not distinguish the shared and private factors across different behaviors, and therefore may make some improper information transferred. (Zhang, Cao, and Yeung 2010) propose the Multi-domain Collaborative Filtering (MCF) method, where they use different user latent factor matrices for different domains, and utilize the covariance matrix between user latent factor matrices to model the relationships between domains. In this paper, instead of sharing all the latent factors as in CMF or sharing non of the latent factors as in MCF, we try to transfer information among multiple types of behaviors more effectively by modeling the shared and private latent factors automatically with group sparsity regularization.

Incorporating group information using mixed-norm regularization has been discussed in statistics and machine learning (Yuan and Lin 2006; Jenatton, Obozinski, and Bach 2009; Liu, Palatucci, and Zhang 2009; Li et al. 2012). Compared to the l_1-norm regularization which is well-known to promote a sparse representation (Tibshirani 1996), mixed-norm regularization, such as $l_{1,2}$-norm, induces a sparse representation at the level of groups. Recently, group sparsity are widely studied in computer vision area (Bengio et al. 2009; Jia, Salzmann, and Darrell 2010). (Bengio et al. 2009) extended sparse coding with $l_{1,2}$-norm regularization to promote each category(group) to use a subset of codewords for construction. (Nie et al. 2010) applied $l_{1,2}$-norm regularization to select out discriminative features across all data points with joint sparsity. With mixed-norm regularization, the group structure can be shared across data points and correlated variables can be selected jointly rather than individual variables. However, to our knowledge, algorithms of applying group sparsity regularization to recommendation have not been investigated before our work in this paper.

3 Proposed Algorithm

Problem Statement

Suppose that we have a set of n users $\mathbf{u} = \{u_1, \cdots, u_n\}$ and their multiple types of behavior records. Each type of user behavior demonstrating her/his opinions on a kind of items can be regarded as ratings (binary or real values), thus we have an integrated matrix consisting of multiple rating matrices for different behaviors, denoted as $\mathbf{R} = \{\mathbf{R}^1, \cdots, \mathbf{R}^B\}$, where $\mathbf{R}^b = [R^b_{ij}]_{n\times m_b}$ denotes the rating matrix for the bth type of behavior. R^b_{ij} denotes the rating of u_i on item v^b_j, v^b_j denotes the jth item associated with the bth type of behav-

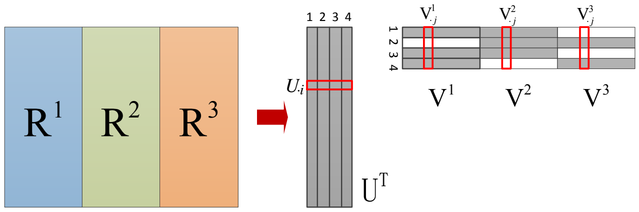

Figure 2: The idea of GSMF. It factorizes the rating matrices for different behaviors into user and item latent factor matrices $\mathbf{U}$ and $\mathbf{V}$, with group sparsity on $\mathbf{V}$. The gray and white rows of $\mathbf{V}^b$ ($b = 1, 2, 3$) represent nonzero and zero values, respectively.

ior, m_b is the number of items belonging to the bth type. Our goal is to predict the missing values in each behavior matrix $\mathbf{R}^b$ ($b = 1, \cdots, B$) by effectively exploiting the information from users' multiple types of behaviors.

Group-Sparse Matrix Factorization

We formulate our problems on the basis of Matrix Factorization (MF) models, which map both users and items to a common latent factor space $\mathbb{R}^k$ (each dimension encodes a latent factor). Let $\mathbf{U} \in \mathbb{R}^{k \times n}$ and $\mathbf{V} \in \mathbb{R}^{k \times m}$ be the user and item latent factor matrices respectively, with column vectors $U_{\cdot i}$ and $V_{\cdot j}$ representing the k-dimensional user-specific and item-specific latent factor vectors. Supposing users' decisions are determined by these latent factors, ratings can be approximated by the inner product, i.e., $R_{ij} = U_{\cdot i}^T V_{\cdot j}$. To learn the $\mathbf{U}$ and $\mathbf{V}$, many of the MF methods suggest minimizing the squared error between the predicted ratings and the observed ratings, as below

$$f(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{n} \sum_{j=1}^{m} I_{ij} (R_{ij} - U_{\cdot i}^T V_{\cdot j})^2 \quad (1)$$

where I_{ij} is the indicator function which is equal to 1 if the user u_i rated the item v_j and is 0 otherwise.

In our cases, for each user we have her/his multiple types of behavior records, leading to multiple rating matrices with the same user dimension. Based on MF, we characterize each user by k latent factors. Our goal is to learn $\mathbf{U}$ and $\mathbf{V}$ from these multiple rating matrices, and make recommendation based on them. In our case, the columns of $\mathbf{V}$ are divided into B groups as $\mathbf{V} = \{\mathbf{V}^1, \ldots, \mathbf{V}^B\}$, where each group is the items that belong to the same type of behavior, and $\mathbf{V}^b \in \mathbb{R}^{k \times m_b}$ denotes the latent factor matrix for the bth type of item with each column $V_{\cdot j}^b$ representing the item-specific latent factor vector for item v_j^b. Thus, the user u_i's rating on item v_j^b is predicted as $\hat{R}_{ij}^b = U_{\cdot i}^T V_{\cdot j}^b$.

Considering that users' decisions on different types of behaviors are determined by different sets of factors, we can not share all the latent factors across different behaviors. To correctly account for the dependencies and independencies between different behaviors, we cast the problem as selecting out the subset of latent factors that each type of behaviors based on. Our idea can be depicted by Figure 2. Supposing there are three types of user behavior records $\mathbf{R}^1$,$\mathbf{R}^2$,$\mathbf{R}^3$, and each user is characterized by four latent factors. We expect *group sparsity* on each $\mathbf{V}^b$ ($b = 1, 2, 3$), that is, the items belonging to the same type of behaviors share the same sparsity pattern in the latent factor space (the same zero-valued latent factor dimensions), resulting in zero-valued rows in each $\mathbf{V}^b$. Since R_{ij}^b is generated by $U_{\cdot i}^T V_{\cdot j}^b$, the zero-valued rows of $\mathbf{V}^b$ remove the influence of the corresponding latent factors, and therefore the factors which determine users' decisions on the bth type of behaviors are selected out. For example, in Figure 2, the decisions on $\mathbf{R}^1$ are affected by factors {1,3,4} with the second row of $\mathbf{V}^1$ to be zero. For different behaviors, different sets of factors may be selected out. As a result, the shared and private latent factors between different behaviors can be modeled automatically. For example, for behaviors 1 and 2, factors {1,3} are shared, while factor {4} and factor {2} are private for behavior 1 and 2, respectively. Thus, by enforcing group sparsity on each $\mathbf{V}^b$, we can achieve the goal of modeling the shared and private information between behaviors.

Inspired by prior researches on group sparsity (discussed in section 2), we can achieve group sparsity on each $\mathbf{V}^b$ by $l_{1,2}$-norm regularization, which is defined as:

$$\|\mathbf{Y}\|_{l_{1,2}} = \sum_{t=1}^{k} \|Y_{t\cdot}\|_2$$

That is, the $l_{1,2}$-norm of a matrix is the sum of vector l_2-norms of its rows. Penalization with $l_{1,2}$-norm promotes zero rows for the matrix. Therefore, with the $l_{1,2}$-norm on each $\mathbf{V}^b$ ($b = 1, \ldots, B$), it promotes that the items belonging to the same type share the same sparsity pattern in the latent factor space, and therefore the decisive latent factors for each type of behaviors can be selected out.

To address above problems, we have the following optimization function:

$$\min_{\mathbf{U},\mathbf{V}} \mathcal{L}(\mathbf{U}, \mathbf{V}) = \sum_{b=1}^{B} \alpha_b (f_b(\mathbf{U}, \mathbf{V}^b) + \beta \|\mathbf{V}^b\|_{l_{1,2}}) + \lambda(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) \quad (2)$$

$$f_b(\mathbf{U}, \mathbf{V}^b) = \sum_{i=1}^{n} \sum_{j=1}^{m_b} I_{ij}^b (R_{ij}^b - U_{\cdot i}^T V_{\cdot j}^b)^2$$

In Eq.(2), $f_b(\mathbf{U}, \mathbf{V}^b)$ ensures that our learnt $\mathbf{U}$ and $\mathbf{V}$ can well approximate the observed ratings in multiple types of user behaviors, where I_{ij}^b is the indicator function which is equal to 1 if the user u_i rated the the item v_j^b and is 0 otherwise. $\| \cdot \|_{l_{1,2}}$ ensure our learnt $\mathbf{V}^b$ are group sparse in rows, and thus select out the decisive factors for each behaviors. $\| \cdot \|_F$ is the Frobenius regularization norm which is used to avoid over-fitting. The parameter α_b is employed to control the contribution from each type of behavior. The parameters β and λ control the strength of each regularization term.

With group sparsity regularization, our model has the following characteristics:

- By selecting out the different sets of decisive latent factors for different behaviors, our method can model the shared and private factors automatically. Furthermore, it allows the shared factors between each two types of behaviors to be different, instead of all the behaviors sharing the same set of factors.

- For user latent factor vector, the shared factors are learnt based on the observed data from multiple types of behaviors which shared the factors, while the private factors are learnt from specific type of behavior separately. By this way, dependence information between behaviors are transferred by the shared factors while independence information are served by the private factors.
- CMF can be viewed as a special case of our model by restricting $\beta = 0$.

Optimization Algorithm

We apply an alternating optimization to solve the proposed formulation, which update $\mathbf{U}$ and $\mathbf{V} = \{\mathbf{V}^1, \dots, \mathbf{V}^B\}$ iteratively and alternatingly.

Optimizing $\mathbf{V}$, given $\mathbf{U}$: When $\mathbf{U}$ is fixed, the constraints are independent on each $\mathbf{V}^b$ ($b = 1, \dots, B$), suggesting that we can optimize each $\mathbf{V}^b$ separately. $\mathbf{V}^b$ can be obtained by solving the following problem,

$$\min_{\mathbf{V}^b} \mathcal{L}(\mathbf{V}^b) = \alpha_b (f_b(\mathbf{U}, \mathbf{V}^b) + \beta \|\mathbf{V}^b\|_{l_{1,2}}) + \lambda \|\mathbf{V}^b\|_F^2 \tag{3}$$

In order to learn the latent factor vector for each item, we take derivation of $\mathcal{L}(\mathbf{V}^b)$ with respect to $V_{\cdot j}^b$:

$$\frac{\partial \mathcal{L}(\mathbf{V}^b)}{\partial V_{\cdot j}^b} = \lambda V_{\cdot j}^b + \alpha_b (\beta \mathbf{D}^b V_{\cdot j}^b + \sum_{i=1}^{n} I_{ij}^b (U_{\cdot i} U_{\cdot i}^T V_{\cdot j}^b - R_{ij}^b U_{\cdot i})) \tag{4}$$

Here $\mathbf{D}^b$ is a $k \times k$ diagonal matrix with the tth diagonal element as $D_{tt}^b = \frac{1}{\|V_{t\cdot}^b\|_2}$, where $V_{t\cdot}^b$ is the tth row of $\mathbf{V}^b$. Setting $\frac{\partial \mathcal{L}(\mathbf{V}^b)}{\partial V_{\cdot j}^b} = 0$, we have:

$$V_{\cdot j}^b = (\frac{\lambda}{\alpha_b} \mathbf{E} + \beta \mathbf{D}^b + \sum_{i=1}^{n} I_{ij}^b U_{\cdot i} U_{\cdot i}^T)^{-1} (\sum_{i=1}^{n} I_{ij}^b R_{ij}^b U_{\cdot i}) \tag{5}$$

Where $\mathbf{E}$ is a $k \times k$ identity matrix. Note that $\mathbf{D}^b$ is dependent to $\mathbf{V}^b$, we propose to update $\mathbf{D}^b$ iteratively too. That is, in each iteration, $\mathbf{D}^b$ is calculated with the current $\mathbf{V}^b$, and then $\mathbf{V}^b$ is updated based on the current calculated $\mathbf{D}^b$. By this way, the update rule in Eq. (5) can monotonically decrease the value of $\mathcal{L}(\mathbf{V}^b)$. The proof process is similar to that in (Nie et al. 2010), and we omit it here due to limited space.

Optimizing $\mathbf{U}$, given $\mathbf{V}$: When $\mathbf{V}$ is fixed, $\mathbf{U}$ can be obtained by the following optimization problem,

$$\min_{\mathbf{U}} \mathcal{L}(\mathbf{U}) = \sum_{b=1}^{B} \alpha_b f_b(\mathbf{U}, \mathbf{V}^b) + \lambda \|\mathbf{U}\|_F^2 \tag{6}$$

Similarly, solving $\frac{\partial \mathcal{L}(\mathbf{U})}{\partial U_{\cdot i}} = 0$, we have:

$$U_{\cdot i} = (\lambda \mathbf{E} + \sum_{b=1}^{B} \alpha_b \sum_{j=1}^{m_b} I_{ij}^b V_{\cdot j}^b {V_{\cdot j}^b}^T)^{-1} (\sum_{b=1}^{B} \alpha_b \sum_{j=1}^{m_b} I_{ij}^b R_{ij}^b V_{\cdot j}^b) \tag{7}$$

Algorithm 1 Optimization Algorithm for GSMF

Require: rating matrices for multiple type of user behavior $\mathbf{R}^b (b = 1, \dots, B)$, Parameters α_b, β, λ
Ensure: latent factor matrices $\mathbf{U}$ and $\mathbf{V}^b (b = 1, \dots, B)$

```
Initialize V^b (b = 1,...,B) and U;
Repeat
  for b = 1 to B do
    Compute the diagonal matrix D^b, where D^b_tt = 1/||V^b_t.||_2;
  end for
  for b = 1 to B do
    Update V^b_.j, ∀1 ≤ j ≤ m_b with Eq. (5);
  end for
  Update U_.i, ∀1 ≤ i ≤ n with Eq. (7);
Until convergence
Return U and V^b (b = 1,...,B)
```

Based on the above analysis, we summarize the detailed optimization algorithm in Algorithm 1. After achieving $\mathbf{U}$ and $\mathbf{V}$, we can predict the missing values by $\hat{R}_{ij}^b = U_{\cdot i}^T V_{\cdot j}^b$.

4 Experiments

Datasets

To evaluate our model's recommendation quality, we crawled the dataset from the publicly available website Douban[1], where users can provide their ratings for movie, books and music, as well as establish social relations with others. Thus, we have four types of user behaviors here. To have sufficient observations to be split in various proportions of training and testing data for our evaluation, we filtered out users who have rated less than 10 books, or 10 movie, or 10 music, and then removed users without social relationships with others. Retrieving all items rated by the selected users, we have a dataset containing 5,916 users with their ratings on 14,155 books, 15,492 music and 7,845 movie, as well as their social relations between each other. The ratings are real values in the range [1,5], while the social relations are binary, indicating whether or not a social relation exists. The detailed statistics are showed in Table 1.

Table 1: Statistics of the Datasets

Behavior Type	#Items	Sparsity	#Ratings per User
Book Rating	14,155	99.85%	22
Music Rating	15,492	99.75%	38
Movie Rating	7,845	98.87%	88
Social Relation	5,916	99.72%	17

Experimental Setups

In the experimental study, we focus on the task of rating prediction in recommendation to evaluate our models' quality. The most popular metric, Root Mean Square Error (RMSE)

[1]http://www.douban.com

Table 2: Performance Comparison on different sparsity cases

Behavior	Training	PMF	NCDCF_U	NCDCF_I	CMF	MCF	GSMF
Book	80%	0.8150	0.8355	0.7976	0.7849	0.8065	**0.7487**
	60%	0.8329	0.8367	0.8026	0.8011	0.8147	**0.7631**
	40%	0.8500	0.8394	0.8143	0.8181	0.8308	**0.7707**
Music	80%	0.7309	0.7826	0.7367	0.7112	0.7027	**0.6815**
	60%	0.7326	0.7812	0.7376	0.7187	0.7067	**0.6881**
	40%	0.7639	0.7859	0.7465	0.7411	0.7307	**0.6998**
Movie	80%	0.7577	0.8941	1.0866	0.7452	0.7328	**0.7235**
	60%	0.7671	0.8954	1.0920	0.7581	0.7476	**0.7286**
	40%	0.7955	0.8971	1.1060	0.7790	0.7700	**0.7398**

is used to measure the prediction quality.

$$RMSE = \sqrt{\frac{1}{T}\sum_{i,j}(R_{ij} - \hat{R}_{ij})^2} \tag{8}$$

where R_{ij} and $\hat{R}_{ij}$ denotes the true and predicted ratings respectively, and T denotes the number of tested ratings. Smaller RSME value means a better performance.

Since the social relation prediction belongs to the task of link prediction, which is different from rating prediction task and unsuitable to be evaluated by RMSE, here we use social relations as a kind of auxiliary behavior and do not do the social relation prediction task. Thus, we will report the RMSE results for book, music and movie rating prediction.

To validate the effectiveness of our GSMF model, we implement the following baselines for comparison with it.

- **PMF** (Salakhutdinov and Mnih 2008): The well known MF method for single type of user behavior. It learns the user and item latent factors for each type of behavior separately and no information is transferred across them.
- **NCDCF_U** (Berkovsky, Kuflik, and Ricci 2007): The user-based neighborhood method integrating user's multiple types of behavior.

  NCDCF_I: The item-based neighborhood method integrating user's multiple types of behavior.
- **CMF** (Singh and Gordon 2008): The method decomposes the rating matrices for different behaviors and transfers information by sharing the same user latent factors.
- **MCF** (Zhang, Cao, and Yeung 2010): The method uses different user latent factor matrices when decomposing the rating matrices for different behaviors, and learns the covariance matrix between these user latent factor matrices to transfer the relationships between behaviors.

Among them, PMF is the single behavior based method, the remaining four methods are the multiple behavior based methods. In our experiments, all the multiple behavior based methods integrated the above mentioned four types of user behaviors on Douban when doing predictions.

To perform comprehensive comparison, we conducted experiments on different training sets (80%, 60% and 40%) to test the models' performance under different sparsity cases. For example, for training data 80%, we randomly select 80% of the data from each types of the behaviors for training and the rest for testing. The random selection was carried out 5 times independently, and we report the average results.

Results

Performance Comparison. We evaluate the rating prediction performance for book, music and movie using the above constructed training/testing sets. The experimental results using 10 dimensions to represent the latent factors are shown in Table 2. The parameter values of our GSMF are: $\alpha_b = 1$ $(b = 1, 2, 3, 4)$, $\lambda = 0.05$ for the three training sets. $\beta = 70$ for 80% and 60% training sets, and $\beta = 40$ for 40% training set. Notice that parameter α_b control the contribution of each behavior, we set all of them the same for simplicity. β controls the strength of group sparse regularization. Later, we will further analyze the impact of β on the performance.

From Table 2, we can observe that the two multiple behavior based MF methods, CMF and MCF, are consistently better than the PMF, which demonstrates that integrating information from multiple types of user behaviors is useful for recommendation. However, the two multiple behavior based neighborhood methods, NCDCF_U and NCDCF_I, do not get consistently better results, which may because that our dataset is very sparse and the neighborhood based methods usually fail to find similar neighbors under such sparse data. Note that CMF performs worse than MCF in music and movie rating prediction, which may because that by sharing all the user latent factors, CMF do not model the private information of different behaviors and make some improper information transferred. However, MCF performs worse than CMF in book rating prediction. It may because that without sharing any of the latent factors, MCF transfers information only based on the learned covariance matrix, when the covariance matrix cannot be learned well in some cases, the shared relationship between behaviors will be badly modeled and in turn harm the results. Thus, it is important to model both the independence and dependence between behaviors well when integrating multiple behaviors into recommendation. It is obvious that our GSMF model consistently outperforms other approaches in all sparsity cases, especially achieving significant improvement over CMF and MCF, which illustrates that the by modeling the shared and private factors among behaviors, GSMF can transfer the information more effectively between behaviors.

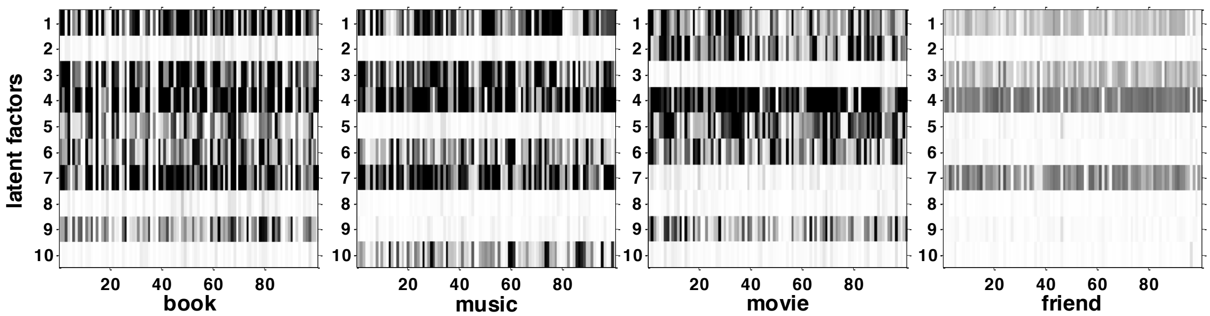

Figure 3: Latent Factor Vectors for each type of items. Each column of the matrix indicates the latent factor vector for one item and each row corresponds to a latent factor entry. The dark block means a large value on the corresponding factor while white block on the opposite.

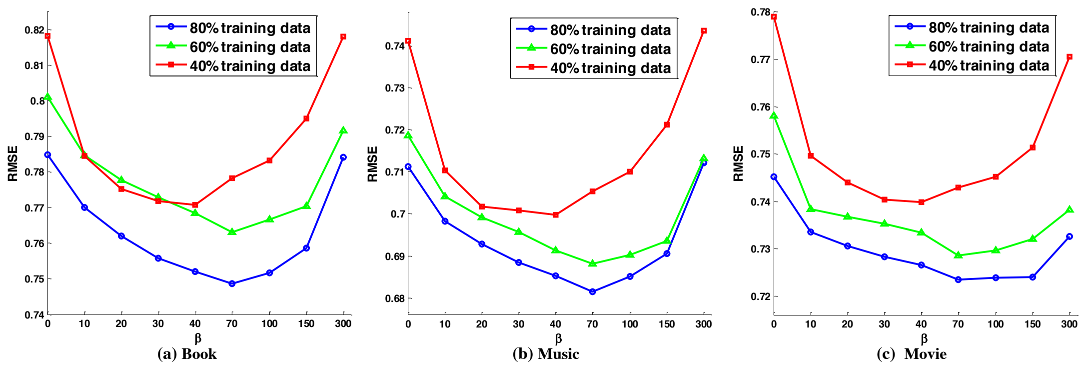

Figure 4: Impact of Parameter β

Analysis on Latent Factor Space. In Figure 3, we showed our learned latent factor vectors for the four types of items associated with the four types of behaviors on Douban. Since the number of items is very large, we can hardly display all of them for limited space. Here, we randomly sampled 100 items from each type of items and showed their latent factor space. In Figure 3, each column of the matrix indicates the 10-dimensional latent factor vector for one item. From the Figure, we can observe white rows[2] in each type of item latent factor matrix, which select out the decisive factors for each type of items. For example, factors $\{1,3,4,5,6,7,9\}$ are selected for books and factors $\{1,3,4,7\}$ are selected for friends. At the same time, our method has modeled the shared-private factors between behaviors automatically, and the shared factors between different behaviors are different. For example, factors $\{1,3,4,6,7\}$ are shared between book and music, while factors $\{1,4,6\}$ are shared between music and movie. Altogether, the results demonstrate that our model can really achieve the goal of modeling the shared and private information between behaviors, which will make information transferred more effectively.

Impact of Parameter β. We investigate the effects of the important parameter in GSMF: β, which controls the strength of group sparsity regularization. In the extreme case, if $\beta = 0$, it degenerates to CMF, which will not distinguish the shared and private factors between behaviors. If β with a very large value, the group sparsity regularization will dominate the learning processes. Figure 4 shows the performance of GSMF on different training sets with different values of β. We can see that, in all cases the RMSE results decrease at first, when β goes greater than a threshold the RMSE increase. This observation coincides with the intuitions: modeling the shared and private information between behaviors with group sparsity regularization may help us integrate the information from users' multiple types of behaviors better and is useful for recommendation; yet, if too much weight is given to the group sparsity regularization, it may pollute the other parts' affection on the model's learning and in turn harm the recommendation performance.

[2]Notice that these "white rows" are not absolutely zero-valued, but most of the elements in these rows are with very small values.

5 Conclusion

In this paper, we propose a novel recommendation model, GSMF, to integrate multiple types of user behaviors effectively by modeling the shared and private factors among them. Specifically, we exploit group sparsity regularization to select out the different subsets of latent factors which different types of behaviors are based on. As a result, the shared and private latent factors between different behaviors can be modeled automatically. Experiment on real-world dataset demonstrate that the proposed method can achieve better recommendation results than other competitors.

Acknowledgements

This work was supported in part by 973 Program (Grant No. 2010CB327905), National Natural Science Foundation of China (Grant No. 61170127, 61332016).

References

Bengio, S.; Pereira, F.; Singer, Y.; and Strelow, D. 2009. Group sparse coding. In *Advances in Neural Information Processing Systems*, 82–89.

Berkovsky, S.; Kuflik, T.; and Ricci, F. 2007. Cross-domain mediation in collaborative filtering. In *User Modeling 2007*. 355–359.

Hu, L.; Cao, J.; Xu, G.; Cao, L.; Gu, Z.; and Zhu, C. 2013. Personalized recommendation via cross-domain triadic factorization. WWW, 595–606.

Jenatton, R.; Obozinski, G.; and Bach, F. 2009. Structured sparse principal component analysis. *arXiv preprint arXiv:0909.1440*.

Jia, Y.; Salzmann, M.; and Darrell, T. 2010. Factorized latent spaces with structured sparsity. In *Advances in Neural Information Processing Systems*, 982–990.

Koren, Y. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 426–434.

Krohn-Grimberghe, A.; Drumond, L.; Freudenthaler, C.; and Schmidt-Thieme, L. 2012. Multi-relational matrix factorization using bayesian personalized ranking for social network data. In *Proceedings of the fifth ACM international conference on Web search and data mining*, 173–182.

Li, C.; Liu, Q.; Liu, J.; and Lu, H. 2012. Learning ordinal discriminative features for age estimation. In *CVPR*. IEEE.

Li, B.; Yang, Q.; and Xue, X. 2009. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction. In *IJCAI*, volume 9, 2052–2057.

Liu, H.; Palatucci, M.; and Zhang, J. 2009. Blockwise coordinate descent procedures for the multi-task lasso, with applications to neural semantic basis discovery. In *Proceedings of the 26th Annual International Conference on Machine Learning*, 649–656.

Ma, H.; Yang, H.; Lyu, M. R.; and King, I. 2008. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management*, 931–940.

Nie, F.; Huang, H.; Cai, X.; and Ding, C. H. 2010. Efficient and robust feature selection via joint l2, 1-norms minimization. In *Advances in Neural Information Processing Systems*, 1813–1821.

Pan, W.; Xiang, E. W.; Liu, N. N.; and Yang, Q. 2010. Transfer learning in collaborative filtering for sparsity reduction. In *AAAI*, volume 10, 230–235.

Salakhutdinov, R., and Mnih, A. 2008. Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems*.

Sarwar, B.; Karypis, G.; Konstan, J.; and Riedl, J. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, 285–295.

Singh, A. P., and Gordon, G. J. 2008. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 650–658.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 267–288.

Yang, S.-H.; Long, B.; Smola, A.; Sadagopan, N.; Zheng, Z.; and Zha, H. 2011. Like like alike: joint friendship and interest propagation in social networks. In *Proceedings of the 20th international conference on World wide web*, 537–546.

Yuan, M., and Lin, Y. 2006. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68(1):49–67.

Zhang, Y.; Cao, B.; and Yeung, D.-Y. 2010. Multi-domain collaborative filtering. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI)*.