Exploring Heterogeneity for Multi-Domain Recommendation with Decisive Factors Selection

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

ABSTRACT

To address the recommendation problems in the scenarios of multiple domains, in this paper, we propose a novel method, HMRec, which models both consistency and heterogeneity of users' multiple behaviors in a unified framework. Moreover, the decisive factors of each domain can also be captured by our approach successfully. Experiments on the real multi-domain dataset demonstrate the effectiveness of our model.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information filtering

Keywords

Recommendation; Heterogeneity; Multiple Domains

1. INTRODUCTION

Recently, recommender systems have been playing an increasingly critical role in coping with information overload. In online systems, there exists massive user rating data that is categorized into different domains such as book, movie or music. Instead of separately mining users' rating behaviors within a single domain [1], recent models (e.g., CMF [3]) resort to leveraging more collaborative information shared across multiple domains for better recommendation.

Existing methods [3] often assume that the users' underlying tastes remain the same across different domains. However, the above assumption ignores heterogeneity of users' multiple behaviors, but only concerns their consistency. Let's consider a real case of heterogeneity, where educational and occupational demands may count most for choosing books while users' preference on romantic or science fiction genre can significantly affect their choices of movies. Therefore, users' collective tastes on multiple domains should also be decided by different decisive factors.

With above concerns, a novel multi-domain approach, HMRec (**H**eterogeneous **M**ulti-domain **Rec**ommendation), is developed to integrate both the consistency and heterogeneity of users' multiple behaviors in a unified framework. More specifically, as shown in Fig.1, the user latent factor matrix of each domain is factorized to a combination of two components: the domain-shared factor matrix for consistency and the domain-specific one for heterogeneity, and behavior-decisive latent factors of each domain are successfully selected in domain-specific matrices. Experimental results on the multi-domain dataset show the superiority of our model.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742711.

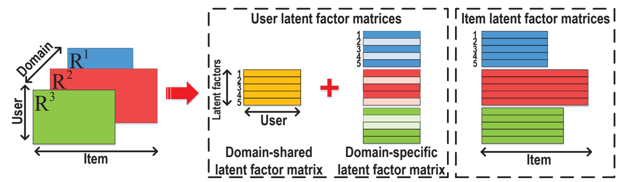

Figure 1: Framework of HMRec.

2. OUR APPROACH

2.1 Problem Statement

Suppose that we have user rating matrices for distinct B domains denoted as $\mathbf{R}^b \in \mathbb{R}^{n\times m_b}$ $(1 \le b \le B)$, where n and m_b are respectively the size of overall user set and the size of item set in the b^{th} domain. Then our goal is to predict the missing values in all rating matrices $\mathbf{R}^b$ by effectively mining the observed rating records across multiple domains.

2.2 HMRec

Our model is built on the basis of the matrix factorization (MF) technique. In the context of multi-domain behaviors, HMRec jointly factorizes the rating matrices to learn the domain-shared and the domain-specific user latent factor matrices as well as the item latent factor matrices of multiple domains.

In our model, let $\mathbf{U}^0$, $\mathbf{U}^b \in \mathbb{R}^{k\times n}$ denote the domain-shared user latent factor matrix and the domain-specific user latent factor matrix for the b^{th} domain, where k is the number of latent factors. Hence, user latent factor matrix $\tilde{\mathbf{U}}^b$ is eventually the function of $\mathbf{U}^0$ and $\mathbf{U}^b$, which is defined as $\tilde{\mathbf{U}}^b = g(\mathbf{U}^0, \mathbf{U}^b)$. Here we simply adopt the linear function to $g(\cdot,\cdot)$ as follows

$$\tilde{\mathbf{U}}^b = g(\mathbf{U}^0, \mathbf{U}^b) = \beta\mathbf{U}^0 + (1-\beta)\mathbf{U}^b \quad (1)$$

where $\beta(0 \le \beta \le 1)$ is the tradeoff parameter tending to balance the contribution of the two components. For simplicity, β is kept the same for each domain. Here $\mathbf{U}^0$ represents the common user features of all domains, and $\mathbf{U}^b$ embodies the domain-determined user features. Thereby, collaborative information will be transferred among domains by $\mathbf{U}^0$ while differences among domains are reflected by $\mathbf{U}^b$.

Since users' behaviors in different domains are determined by different set of factors, it is necessary to perform the factor selection on the domain-specific user latent factor matrix. In order to capture the most decisive factors in each domain, we introduce the $\ell_{2,1}$-norm to $\mathbf{U}^b$ in our model as follows

$$\|\mathbf{U}^b\|_{2,1} = \sum_{t=1}^{k} \|\mathbf{U}^b_{t\cdot}\|_2 \qquad (2)$$

where $\|\mathbf{U}^b_{t\cdot}\|_2$ is the ℓ_2-norm for each row of $\mathbf{U}^b$. $\ell_{2,1}$-norm can lead to row sparsity, which forces some rows of $\mathbf{U}^b$ to be close to 0. Thus the insignificant factors of each domain are eliminated and the most decisive factors are selected.

Then, with defining $\mathbf{V}^b \in \mathbb{R}^{k \times m_b}$ as the item latent factor matrix for domain b, we have rating prediction squared error on training data within domain b as the following form

$$f_b(\mathbf{U}^0, \mathbf{U}^b, \mathbf{V}^b) = \sum_{i=1}^{n} \sum_{j=1}^{m_b} \mathbf{I}^b_{ij} (\mathbf{R}^b_{ij} - (\tilde{\mathbf{U}}^b_{\cdot i})^T \mathbf{V}^b_{\cdot j})^2 \qquad (3)$$

Particularly, $\mathbf{I}^b$ is an indicator matrix whose entry $\mathbf{I}^b_{ij}$ is 1 if user u_i has rated item v_j and 0 otherwise. And $\tilde{\mathbf{U}}^b_{\cdot i}$, $\mathbf{V}^b_{\cdot j}$ is the i^{th} and j^{th} column of user and item latent factor matrix.

With above preliminary formulation, our model is eventually written as minimizing the following objective function

$$\mathcal{L}(\mathbf{U}^0, \{\mathbf{U}^b\}_{b=1}^B, \{\mathbf{V}^b\}_{b=1}^B) = \sum_{b=1}^{B} \alpha_b (f_b(\mathbf{U}^0, \mathbf{U}^b, \mathbf{V}^b) + \gamma \|\mathbf{U}^b\|_{2,1}) + \lambda \mathcal{R}(\mathbf{U}^0, \{\mathbf{U}^b\}_{b=1}^B, \{\mathbf{V}^b\}_{b=1}^B) \qquad (4)$$

where the regularization term is $\mathcal{R}(\mathbf{U}^0, \{\mathbf{U}^b\}_{b=1}^B, \{\mathbf{V}^b\}_{b=1}^B) = \|\mathbf{U}^0\|_F^2 + \sum_{b=1}^B \|\mathbf{U}^b\|_F^2 + \sum_{b=1}^B \|\mathbf{V}^b\|_F^2$ to avoid overfitting. And α_b balances the contribution of each domain, γ controls the effects of the row sparsity and λ controls the strength of the regularization term.

2.3 Optimization

Because Eq.(4) is convex w.r.t. one of the variables $\mathbf{U}^0$, $\{\mathbf{U}^b\}_{b=1}^B$, $\{\mathbf{V}^b\}_{b=1}^B$ when the others are fixed, we apply coordinate descent optimization for our model. Then the missing values are predicted as $\hat{\mathbf{R}}^b_{ij} = (\tilde{\mathbf{U}}^b_{\cdot i})^T \mathbf{V}^b_{\cdot j}$.

3. EXPERIMENTS

We perform experiments on a multi-domain dataset crawled from the website Douban. Douban is a famous Web2.0 website for users to provide their ratings, scaled from 1 to 5, on books, movies and music. We filtered out users with less than 10 ratings on the 3 domains and obtained a dataset of 5,916 users. The detailed description is presented in Table 1.

Table 1: Description of Douban Dataset

Domain	# Items	% Sparsity	# Ratings per User
Book	14,155	99.85	22
Music	15,492	99.75	38
Movie	7,845	98.87	88

Our method is compared with the following baselines: (1)PMF [2]: the basic MF method, making prediction in each domain separately. (2)NCDCF_U and NCDCF_I [1]: respectively user-based and item-based neighborhood methods for multi-domain scenarios. (3)CMF [3]: a multi-domain MF model which tries to share the same user latent factors across different domains. Note that our model is reduced to PMF if $\beta = 0$, $\gamma = 0$, and CMF when $\beta = 1$, $\gamma = 0$.

Table 2: Performance Comparisons (mean ± std.)

Methods	Domains		
	Book	Music	Movie
PMF	0.8604±0.0028	0.7433±0.0025	0.7438±0.0015
NCDCF_U	0.8305±0.0035	0.7710±0.0011	0.8599±0.0024
NCDCF_I	0.7701±0.0039	0.7230±0.0013	0.7668±0.0018
CMF	0.7836±0.0013	0.7063±0.0003	0.7379±0.0021
HMRec	**0.7622**±0.0021	**0.6884**±0.0010	**0.7292**±0.0014

In our experiments, we randomly select 80% observed ratings as training data and the rest are used as testing data. The random selection is preformed 5 times independently, and the parameters are determined. The best average results with standard deviations are then reported. Root Mean Square Error (RMSE) is employed as our evaluation metric, which is the most popular metric utilized for rating prediction tasks. Lower values of RMSE correspond to better recommendation performance.

Table 2 summarizes the performance comparisons. In the experiments, we set $\lambda = 0.05$, $\alpha_b = 1$ for our model and the dimension of the latent factors is fixed as $k = 10$. Then, the resulting optimal parameters of HMRec are $\{\beta = 0.6, \gamma = 80\}$ in each domain. We can observe that our approach always outperforms all the other baselines on each domain, including the single-domain approach (PMF) and the other multi-domain ones (NCDCF_U, NCDCF_I and CMF). The results prove the necessity of introducing domain-specific latent factors with factor selection into our proposed model.

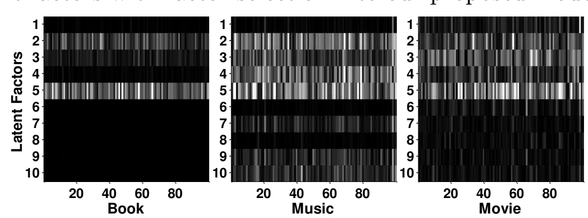

Figure 2: Decisive factor selection of $\mathbf{U}^b$ in each domain. Lighter blocks indicate larger values.

Then we randomly sample 100 items from each domain and show their domain-specific latent factor space. Fig.2 presents the learnt row-sparsity patterns of $\mathbf{U}^b$. And it shows that our method is able to mine the heterogeneity of users' collective tastes for different domains, and the decisive factors can be successfully selected.

4. CONCLUSIONS

Our novel recommendation approach manages to model users' multi-domain behaviors with domain-shared latent and domain-specific factors. And the most decisive factors for different domains are also effectively selected. Experimental results show the effectiveness of our method.

5. REFERENCES

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

[2] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *NIPS*, 2007.

[3] A. P. Singh and G. J. Gordon. Relational learning via collective matrix factorization. In *SIGKDD*, 2008.