Information Sciences 332 (2016) 84–93

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Mixed factorization for collaborative recommendation with heterogeneous explicit feedbacks

Weike Pan, Shanchuan Xia, Zhuode Liu, Xiaogang Peng, Zhong Ming*

College of Computer Science and Software Engineering, Shenzhen University, China

ARTICLE INFO

Article history:
Received 28 November 2014
Revised 23 September 2015
Accepted 27 October 2015
Available online 10 November 2015

Keywords:
Preference learning
Collaborative recommendation
Heterogeneous explicit feedbacks
Mixed factorization
Transfer learning

ABSTRACT

Collaborative recommendation (CR) is a fundamental enabling technology for providing high-quality personalization services in various online and offline applications. Collaborative recommendation with heterogeneous explicit feedbacks (CR-HEF) such as 5-star grade scores and like/dislike binary ratings is a new and important problem, because it provides a rich and accurate source for learning users' preferences. However, most previous works on collaborative recommendation only focus on exploiting *homogeneous* explicit feedbacks such as grade scores or *homogeneous* implicit feedbacks such as clicks or purchases. In this paper, we study the CR-HEF problem, and design a novel and generic mixed factorization based transfer learning framework to fully exploit those two different types of explicit feedbacks. Experimental results on two CR-HEF tasks with real-world data sets show that our TMF is able to perform significantly better than the state-of-the-art methods.

1. Introduction

Collaborative recommendation [1,2] serves as an enabling technology for providing personalization services in various systems and applications, such as academic resource recommendation [3,4], entertainment video recommendation [5,6], telecom/mobile service recommendation [7,8], and people/community recommendation [9,10], etc. The main idea of collaborative recommendation is to learn users' hidden preferences via exploiting users' feedbacks in a collective way instead of studying each user separately [11,12]. However, we may not learn users' preferences well when users' feedbacks are few [13,14], in particular of the 5-star grade scores that most recommendation algorithms [15,16] rely on. This kind of data scarcity challenge will usually cause the overfitting problem when building a recommendation model.

Heterogeneous explicit feedbacks (HEF) such as grade scores from best to worst and binary ratings of like and dislike provide a rich and accurate source for learning users' preferences and constructing users' online profiles, which gives us an opportunity to address the data scarcity challenge of the grade scores [17]. However, most mathematical models in collaborative recommendation (CR) are designed for learning users' preferences from *homogeneous* explicit feedbacks such as 5-star grade scores [15,16], or from *homogeneous* implicit feedbacks such as clicks or purchases [18,19]. In this paper, we study this new and important problem, i.e., collaborative recommendation with heterogeneous explicit feedbacks (CR-HEF), which contains two different types of explicit feedbacks, i.e., 5-star grade scores and like/dislike binary ratings. We then focus on designing a novel and generic algorithm to fully exploit such heterogeneous explicit feedbacks in a principled way.

* Corresponding author. Tel.: +8675526534480.
E-mail addresses: panweike@szu.edu.cn (W. Pan), shaun.xia@outlook.com (S. Xia), zero.lzd@gmail.com (Z. Liu), patrickpeng@126.com (X. Peng), mingz@szu.edu.cn (Z. Ming).

http://dx.doi.org/10.1016/j.ins.2015.10.044
0020-0255/

Previous works on exploiting heterogeneous explicit feedbacks are very few, among which the major approach is matrix *collective* factorization [17,20,21]. Collective factorization based methods are usually designed to simultaneously factorize two preference matrices, i.e., a grade score matrix and a binary rating matrix in CR-HEF, where some latent features from the same users or the same items are shared so as to enable joint preference modeling and sharing. However, such two jointly conducted factorization tasks are still loosely coupled, which may not fully leverage the binary ratings to the grade scores.

Besides the collective factorization methods, there is a different technique called matrix *integrative* factorization for a related problem to CR-HEF, i.e., collaborative recommendation with grade scores and implicit feedbacks [15]. Integrative factorization based methods take the implicit feedbacks as instances (or feedback instances), and incorporate them seamlessly into the factorization task of grade scores as an additional term in the prediction rule. However, leveraging the implicit feedbacks to the grade scores in such an integrative manner may not well capture the implicit-feedback-dependent effect.

In this paper, we aim to overcome the aforementioned limitations of the two state-of-the-art factorization models, i.e., collective factorization and integrative factorization, and adapt them to our studied CR-HEF problem. Specifically, we first take the CR-HEF problem from a transfer learning view [22], in which grade scores are taken as target data and binary ratings are taken as auxiliary data. We then propose a novel and generic mixed factorization based transfer learning framework, i.e., *transfer by mixed factorization* (TMF), which consists of collective factorization and integrative factorization as different assembly components. The novelty of our TMF is that it unifies collective factorization and integrative factorization in one single transfer learning framework, which enables both feature-based and instance-based preference learning and transfer in a principled way. Furthermore, we can also derive some new algorithm variants from our TMF for leveraging different parts or combinations of auxiliary binary ratings.

TMF is expected to transfer more knowledge from binary ratings to grade scores than collective factorization, and to model binary-rating-dependent and -independent effect more accurately than integrative factorization. Experimental results on two real-world data sets show that our TMF performs significantly better than collective factorization or integrative factorization alone.

We organize the rest of the paper as follows. We first discuss some closely related works in Section 2. We then describe the proposed solution in detail in Section 3, and conduct extensive empirical studies of our TMF and the state-of-the-art methods in Section 4. Finally, we give some concluding remarks and future directions in Section 5.

2. Related work

Our TMF is designed from a transfer learning perspective with factorization techniques, which aims to learn users' preferences from heterogeneous explicit feedbacks (HEF) in collaborative recommendation. Hence, in this section, we discuss some related work of TMF in the context of transfer learning for heterogeneous feedbacks, and factorization for collaborative recommendation.

2.1. Transfer learning for heterogeneous feedbacks

Transfer learning [22] aims to conduct parameter learning and knowledge transfer for more than one domains, tasks or data, which has been a state-of-the-art solution in many applications, including text and image classification [23,24], personalization and social behavior learning [17,25], etc. Transfer learning algorithms have also been designed to learn users' preferences from heterogeneous feedbacks in collaborative recommendation, for which we will discuss about their problem settings and techniques.

The problem settings of heterogeneous feedbacks mainly include two categories, i.e., (i) different recommendation tasks with the same feedback type, and (ii) the same recommendation task with different feedback types. The first category includes recommendation tasks for books and movies with grade scores [13,26], recommendation tasks for books, movies and music with grade scores [27], and recommendation tasks for different Amazon products with grade scores [28,29], etc. This category is usually called cross-domain collaborative recommendation [30,31] because the items or product domains are different. The second category includes recommendation for movies with different types of feedbacks such as grade scores and implicit, binary and/or uncertain feedbacks [14,17,32]. Our studied CR-HEF problem in this paper also belongs to the second category. Transfer learning techniques for heterogeneous feedbacks in collaborative recommendation or other applications mainly contain model-based, feature-based and instance-based transfer [22,32], and the corresponding algorithms include adaptive, collective and integrative styles. We follow and expand the categorizations of transfer learning techniques in collaborative recommendation in [32], in particular of "how to transfer" in transfer learning [22], and show the relationship between our TMF and some typical works in Table 1. From Table 1, we can see that our TMF is different from all existing works w.r.t. either of the two dimensions, i.e., transfer learning approaches or transfer learning algorithm styles. Specifically, TMF is a *mixed* transfer learning approach of feature-based transfer and instance-based transfer, and a *mixed* transfer learning algorithm of collective factorization and integrative factorization.

2.2. Factorization for collaborative recommendation

Factorization techniques such as second order matrix factorization, higher order tensor factorization and their extensions have been well studied and successfully applied to many machine learning and data mining problems [33], among which collaborative recommendation is an important application [15,16,19,34–37].

Table 1
Summary of our TMF and some typical transfer learning works in collaborative recommendation from the perspective of "how to transfer" in transfer learning [22].

Transfer learning approaches	Transfer learning algorithm styles			
	Adaptive	Collective	Integrative	Mixed
Model-based transfer	CBT[13]	RMGM[26]		
Feature-based transfer	CST[14]	CMF[21], iTCF[20]		***TMF*** (proposed in this paper)
Instance-based transfer			SVD++[15]	

One of the state-of-the-art methods in collaborative recommendation is to factorize an explicit feedback or an implicit feedback into some latent feature factors. So far, many different factorization techniques have been proposed for different problem settings, including, (i) explicit feedbacks, (ii) implicit feedbacks, (iii) explicit feedbacks and implicit feedbacks, and (iv) heterogeneous explicit feedbacks such as CR-HEF studied in this paper.

For the first problem of explicit feedbacks, the most famous method is to approximate the grade score r_{ui} of user u and item i by $U_{u\cdot}V_{i\cdot}^T$, which is known as a pointwise regression method [15,34]. There are also some pairwise and listwise methods that directly optimize some ranking-oriented evaluation metrics [36,37]. For the second problem of implicit feedbacks, one of the state-of-the-art methods are based on a pairwise assumption, that is, the preference score of an observed (user, item) pair, denoted as (u, i), is larger than that of an unobserved one (u, j) [19], i.e., $U_{u\cdot}V_{i\cdot}^T > U_{u\cdot}V_{j\cdot}^T$. There are also some extensions of [19], such as that with users' social networks [38] and items' taxonomy [39]. For the third problem of explicit feedbacks and implicit feedbacks, SVD++ [15] and constrained PMF [34] are integrative factorization methods that incorporate the implicit feedbacks in a principled way, i.e., $U_{u\cdot}V_{i\cdot}^T + \sum_{j\in\mathcal{O}_u} O_{j\cdot}V_{i\cdot}^T/\sqrt{|\mathcal{O}_u|}$, where $\mathcal{O}_u$ is a set of implicitly examined items by user u. A recent generic feature engineering based factorization framework can also mimic such integration [16]. For the fourth problem of heterogeneous explicit feedbacks, a recent algorithm [20] approximates a grade score and a binary rating in a collective way via sharing the item-specific latent features, i.e., $U_{u\cdot}V_{i\cdot}^T$ and $W_{u\cdot}V_{i\cdot}^T$, and also introducing interactions between user-specific latent features, i.e., $U_{u\cdot}\leftrightarrow W_{u\cdot}$. An early work [17] on CR-HEF is based on a batch gradient descent algorithm, which may not be efficient as compared with the stochastic one [20]. Our TMF is also proposed for this problem, which unifies the collective factorization method [20] and the integrative factorization method [15] in a principled way.

In a summary, our TMF is different from previous works [13–15,20,21,26] due to its "mixed" characteristics w.r.t. knowledge transfer algorithm styles and transfer learning approaches, which can also be identified in Table 1.

3. Transfer by mixed factorization

3.1. Problem definition of CR-HEF

Collaborative recommendation with heterogeneous explicit feedbacks (CR-HEF) is a recently studied problem with few works [17,20]. Without loss of generality, we assume that we have n users and m items in a typical deployed recommendation system, where the user identities, item identities and preference scores are in $\{1, \ldots, n\}$, $\{1, \ldots, m\}$ and $\mathbb{G}$, respectively. We also have a set of target grade score records $\mathcal{R} = \{(u, i, r_{ui})\}$, where each record means that a user u has assigned a grade score r_{ui} to an item i. Note that such records are very few as compared with all possible grade score assignments, i.e., $|\mathcal{R}| \ll n \times m$, which is usually called the data scarcity challenge making it difficult to learn a reliable recommendation model.

Besides the target data of grade scores $\mathcal{R}$ as well studied already in previous works [15,16], we have an auxiliary data with binary ratings $\tilde{\mathcal{R}} = \{(u, i, \tilde{r}_{ui})\}$, where $\tilde{r}_{ui} \in \mathbb{B} = \{\text{like}, \text{dislike}\}$. We may leverage such auxiliary data to help alleviate the data scarcity problem.

The goal of CR-HEF is then to jointly model both the target grade scores $\mathcal{R}$ and the auxiliary binary ratings $\tilde{\mathcal{R}}$ so as to improve the preference learning accuracy. We illustrate the studied problem in Fig. 1, where the signs of *collective* and *integrative* denote two assembly components in the proposed mixed factorization framework that exploit heterogeneous explicit feedbacks in different styles.

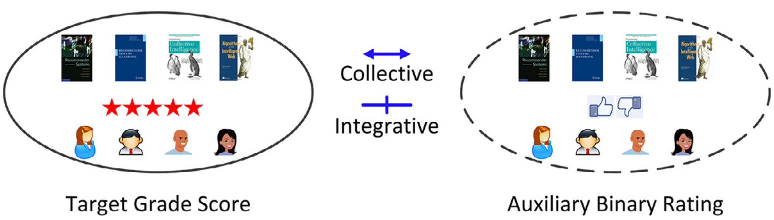

Fig. 1. Illustration of collaborative recommendation with heterogeneous explicit feedbacks (CR-HEF).

3.2. Preference representation

In factorization based methods, for the target grade score records $\mathcal{R}$, we usually represent a user's interest profile and an item's attribute description by latent vectors, i.e., $U_{u\cdot} \in \mathbb{R}^{1\times d}$ for user u and $V_{i\cdot} \in \mathbb{R}^{1\times d}$ for item i [15]. With such representation, a user u's quantitative preference to item i can be estimated as follows,

$$\hat{r}_{ui} = U_{u\cdot}V_{i\cdot}^T. \tag{1}$$

Note that we may introduce some scalar variables to the above prediction rule, i.e., $\hat{r}_{ui} = U_{u\cdot}V_{i\cdot}^T + b_u + b_i + \mu$, where b_u, b_i and μ are the user preference bias, the item preference bias and the global average preference score, respectively.

When some auxiliary binary rating records $\tilde{\mathcal{R}}$ are available as shown in Fig. 1, we may have a similar representation for user u's preference to item i [20,21],

$$\hat{\tilde{r}}_{ui} = W_{u\cdot}V_{i\cdot}^T,\ W_{u\cdot} \leftrightarrow U_{u\cdot}, \tag{2}$$

where the term $W_{u\cdot}V_{i\cdot}^T$ in Eq. (2) and the term $U_{u\cdot}V_{i\cdot}^T$ in Eq. (1) mean that the item i's latent features $V_{i\cdot}$ are shared between $\mathcal{R}$ and $\tilde{\mathcal{R}}$ [21], and the term $W_{u\cdot} \leftrightarrow U_{u\cdot}$ in Eq. (2) denotes introduced interactions between user u's two different profiles $U_{u\cdot} \in \mathbb{R}^{1\times d}$ and $W_{u\cdot} \in \mathbb{R}^{1\times d}$ [20]. Specifically, we will encourage knowledge sharing between $U_{u\cdot}$ and $W_{u\cdot}$ in the learning process, where both $W_{u\cdot}$ and $U_{u\cdot}$ will be leveraged when updating $V_{i\cdot}$, regardless of the type of the sampled explicit feedback. The effect of the introduced interactions will affect the update rules in a smooth manner in the learning algorithm [20].

However, the knowledge transfer strategy via sharing the latent features (i.e., $V_{i\cdot}$) and/or encouraging interactions between $U_{u\cdot}$ and $W_{u\cdot}$ may not be able to fully exploit the auxiliary binary rating records. The reason is that such a *collective* factorization approach still only consists of two loosely-coupled regression or factorization tasks. In this paper, we propose to further introduce an *integrative* factorization method into the *collective* factorization approach, so as to reduce the independency between two factorization tasks and thus enable more effective knowledge transfer between the target data $\mathcal{R}$ and the auxiliary data $\tilde{\mathcal{R}}$.

In our daily life, a shopping guide may elicitate a customer's interests or preferences via asking what s/he likes and dislikes, based on which a recommendation may be made rather accurately. Hence, besides the user's interest profile $U_{u\cdot}$ for user u in the target data, we propose two additional interest profiles, including one from the user's *liked* items and one from the user's *disliked* items. Mathematically, as inspired by the integrative factorization work on modeling grade scores and implicit feedbacks [15], we propose two new interest profiles for user u,

$$\bar{P}_{u\cdot} = \frac{1}{\sqrt{|\mathcal{P}_u|}}\sum_{j\in\mathcal{P}_u} P_{j\cdot}, \quad \bar{N}_{u\cdot} = \frac{1}{\sqrt{|\mathcal{N}_u|}}\sum_{j\in\mathcal{N}_u} N_{j\cdot}, \tag{3}$$

where $\mathcal{P}_u$ and $\mathcal{N}_u$ are sets of liked and disliked items by user u, respectively. $\bar{P}_{u\cdot}$ are normalized user profiles constructed from user u's liked items' latent features $P_{j\cdot}$, $j \in \mathcal{P}_u$, and $\bar{N}_{u\cdot}$ are constructed from the corresponding disliked items' latent features $N_{j\cdot}$, $j \in \mathcal{N}_u$. We further combine these two profiles,

$$A_{u\cdot} = \delta_{\mathcal{P}} w_p \bar{P}_{u\cdot} + \delta_{\mathcal{N}} w_n \bar{N}_{u\cdot}, \tag{4}$$

where $A_{u\cdot}$ is a fused virtual user profile for user u with boolean variables $\delta_{\mathcal{P}}, \delta_{\mathcal{N}} \in \{0, 1\}$ and weights $w_p, w_n \geq 1$. The effect of the virtual user profile $A_{u\cdot}$ with different configurations of $\delta_{\mathcal{P}}$, $\delta_{\mathcal{N}}$, w_p and w_n will be studied in our experiments. Then, we obtain an improved preference representation as compared with that in Eq. (2),

$$\hat{r}_{ui} = U_{u\cdot}V_{i\cdot}^T + A_{u\cdot}V_{i\cdot}^T, \quad W_{u\cdot} \leftrightarrow U_{u\cdot}, \tag{5}$$

where the first term $U_{u\cdot}V_{i\cdot}^T$ is from *collective* factorization [21], the second term $A_{u\cdot}V_{i\cdot}^T$ is new and generalizes *integrative* factorization [15], and the third term $W_{u\cdot} \leftrightarrow U_{u\cdot}$ is for encouraging interactions between user profiles $W_{u}\cdot$ and $U_{u\cdot}$ [20]. We can see that the factorizations shown in Eqs. (1) and (5) are a *mix* of *collective* factorization and *integrative* factorization for the target data and the auxiliary data, and for this reason we call our approach as *transfer by mixed factorization* (TMF).

With the aforementioned user profiles, item descriptions and preference representation, we may describe them in a single graphical model, which is shown in Fig. 2(d). In Fig. 2(d), we can see that the target grade score r_{ui} is generated by $U_{u\cdot}$, $V_{i\cdot}$, $P_{j\cdot}$, $j \in \mathcal{P}_u$, and $N_{j\cdot}$, $j \in \mathcal{N}_u$, the auxiliary binary rating $\tilde{r}_{ui}$ is generated by $W_{u\cdot}$ and $V_{i\cdot}$, and the edge e_5 denotes the encouraged interactions between $U_{u\cdot}$ and $W_{u\cdot}$. Note that we follow [34], and use Gaussian distributions and priors for the observed ratings and the latent variables, respectively. We can also derive some reduced models via keeping some subsets of edges, e.g., RSVD for edges $\{e_1, e_2\}$ shown in Fig. 2(a), CMF for edges $\{e_1, e_2, e_3, e_4\}$ shown in Fig. 2(b), and iTCF for edges $\{e_1, e_2, e_3, e_4, e_5\}$ shown in Fig. 2(c). Note that our TMF contains all those seven edges in Fig. 2(d). From the graphical models in Fig. 2, we can also see a clear path of extensions from RSVD to CMF, from CMF to iTCF, and from iTCF to our TMF, which shows that our TMF is novel and generic.

3.3. The optimization problem

Finally, with the preference representation in Eq. (2), Eq. (5) and Fig. 2, we reach the objective function to be minimized,

$$\min_{\Theta} \sum_{u=1}^{n}\sum_{i=1}^{m} y_{ui}\ell_{ui} + \lambda \sum_{u=1}^{n}\sum_{i=1}^{m} \tilde{y}_{ui}\tilde{\ell}_{ui}, \tag{6}$$

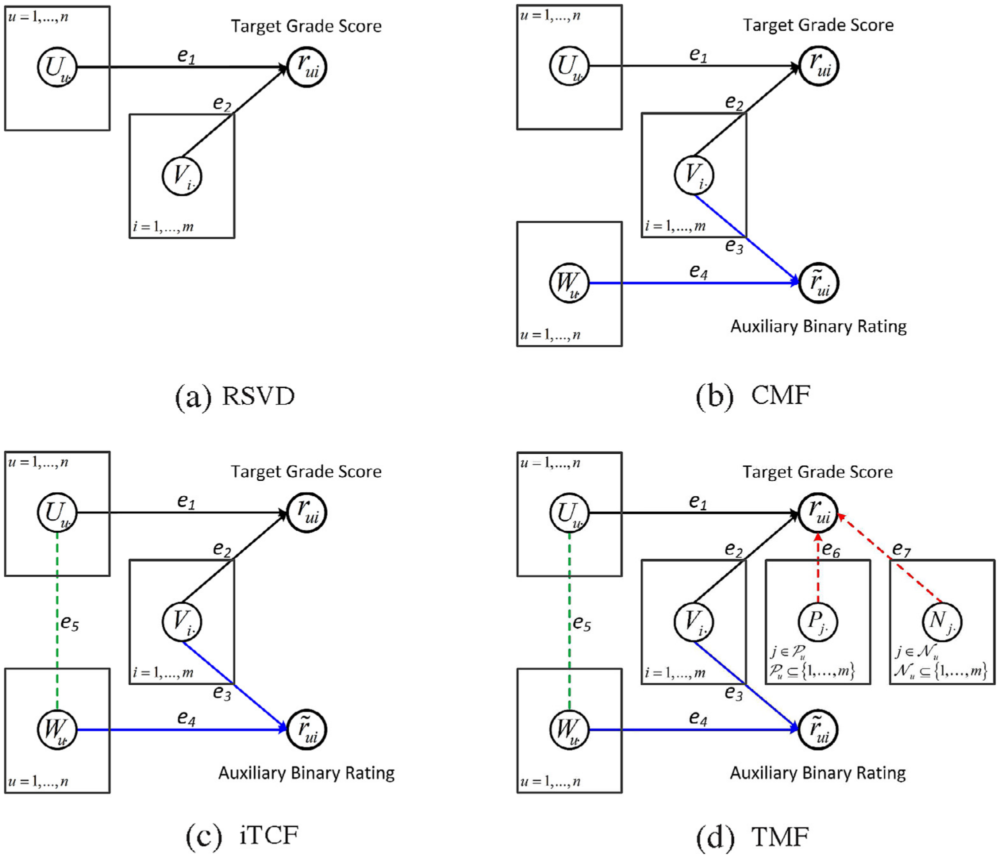

Fig. 2. Graphical models of *transfer by mixed factorization* (TMF) and other methods. Note that regularized Singular value decomposition (RSVD) [15], collective matrix factorization (CMF) [21] and interaction-rich transfer by collective factorization (iTCF) [20] contain subsets of edges, i.e., $\{e_1, e_2\}$, $\{e_1, \ldots, e_4\}$ and $\{e_1, \ldots, e_5\}$, respectively.

where $\ell_{ui} = \frac{1}{2}(r_{ui} - \hat{r}_{ui})^2 + \frac{\alpha_u}{2}\|U_{u\cdot}\|^2 + \frac{\alpha_v}{2}\|V_{i\cdot}\|^2 + \frac{\beta_u}{2}\|b_u\|^2 + \frac{\beta_v}{2}\|b_i\|^2 + \delta_{\mathcal{P}}\frac{\alpha_p}{2}\sum_{j\in\mathcal{P}_u}\|P_{j\cdot}\|^2 + \delta_{\mathcal{N}}\frac{\alpha_n}{2}\sum_{j\in\mathcal{N}_u}\|N_{j\cdot}\|^2$ and $\tilde{\ell}_{ui} = \frac{1}{2}(r_{ui} - \hat{\tilde{r}}_{ui})^2 + \frac{\alpha_w}{2}\|W_{u\cdot}\|^2 + \frac{\alpha_v}{2}\|V_{i\cdot}\|^2$ are regularized loss functions for a target grade score record (u, i, r_{ui}) and an auxiliary binary rating record $(u, i, \tilde{r}_{ui})$, respectively. Note that $\Theta = \{U_{u\cdot}, V_{i\cdot}, W_{u\cdot}, P_{j\cdot}, N_{j\cdot}, b_u, b_i, \mu\}$ with $u = 1, \ldots, n$ and $i, j = 1, \ldots, m$ is a set of model parameters to be learned, and $y_{ui}, \tilde{y}_{ui} \in \{1, 0\}$ are indicator variables for the corresponding records. Note that the interactions between $U_{u\cdot}$ and $W_{u\cdot}$ as shown in Eqs. (2) and (5) will be reflected in the gradients [20] in the sequent subsection.

3.4. The learning algorithm

We have the gradients of model parameters for a sampled target grade score record (u, i, r_{ui}) from ℓ_{ui},

$$\nabla \mu = -e_{ui}, \tag{7}$$

$$\nabla b_u = -e_{ui} + \beta_u b_u, \tag{8}$$

$$\nabla b_i = -e_{ui} + \beta_v b_i, \tag{9}$$

$$\nabla U_{u\cdot} = -e_{ui} V_{i\cdot} + \alpha_u U_{u\cdot}, \tag{10}$$

$$\nabla V_{i\cdot} = -e_{ui}(\rho U_{u\cdot} + (1-\rho) W_{u\cdot} + A_{u\cdot}) + \alpha_v V_{i\cdot}, \tag{11}$$

$$\nabla P_{j\cdot} = \delta_{\mathcal{P}}\left(-e_{ui} w_p \frac{1}{\sqrt{|\mathcal{P}_u|}} V_{i\cdot} + \alpha_p P_{j\cdot}\right), \quad j \in \mathcal{P}_u, \tag{12}$$

$$\nabla N_{j\cdot} = \delta_{\mathcal{N}}\left(-e_{ui} w_n \frac{1}{\sqrt{|\mathcal{N}_u|}} V_{i\cdot} + \alpha_n N_{j\cdot}\right), \quad j \in \mathcal{N}_u, \tag{13}$$

where $e_{ui} = (r_{ui} - \hat{r}_{ui})$ is the error on the target grade score. Note that the gradients $\nabla\mu$, ∇b_u, ∇b_i and $\nabla U_{u\cdot}$ are the same as that of RSVD [15], CMF [21] and iTCF [20]. The difference is from $\nabla V_{i\cdot}$ with a virtual user profile $A_{u\cdot}$, and two new gradients $\nabla P_{j\cdot}$ and $\nabla N_{j\cdot}$. Note that the interactions between $U_{u\cdot}$ and $W_{u\cdot}$ are reflected in the term $\rho U_{u\cdot} + (1-\rho) W_{u\cdot}$ in Eq. (11).

We have the gradients of model parameters for a sampled auxiliary binary rating record $(u, i, \tilde{r}_{ui})$ from $\tilde{\ell}_{ui}$,

$$\nabla W_{u\cdot} = \lambda(-\tilde{e}_{ui}V_{i\cdot} + \alpha_u W_{u\cdot}), \tag{14}$$

$$\nabla V_{i\cdot} = \lambda(-\tilde{e}_{ui}(\rho W_{u\cdot} + (1-\rho)U_{u\cdot}) + \alpha_v V_{i\cdot}), \tag{15}$$

where $\tilde{e}_{ui} = (\tilde{r}_{ui} - \hat{\tilde{r}}_{ui})$ is the error on the auxiliary binary rating. Note that the gradients $\nabla W_{u\cdot}$ and $\nabla V_{i\cdot}$ on the auxiliary data are the same as that of iTCF [20], because the *mixed* factorization is mainly reflected in the factorization of the target data as shown in Eq. (5).

With the above gradients for model parameters, we can learn the model parameters via the following update rule,

$$\theta = \theta - \gamma \nabla \theta, \tag{16}$$

where θ can be any member of Θ.

We put the gradients in Eqs. (7–15) and the update rule in Eq. (16) in a stochastic gradient descent (SGD) based algorithmic framework shown in Algorithm 1. In Algorithm 1, the most expensive steps are calculating the gradients $\nabla P_{j\cdot}$, $\nabla N_{j\cdot}$ and updating the parameters $P_{j\cdot}$, $N_{j\cdot}$ for each positive or negative feedback. Specifically, the time cost of TMF is $O(Td|\mathcal{P}||\mathcal{N}|)$, where $|\mathcal{P}|$ and $|\mathcal{N}|$ are the average numbers of positive feedbacks and negative feedbacks by a certain user, respectively, which are usually small.

Algorithm 1 The algorithm of *transfer by mixed factorization* (TMF).

Initialization of parameters Θ.
for $t = 1, 2, \ldots, T$ **do**
  for $iter = 1, 2, \ldots, |\mathcal{R}| + |\tilde{\mathcal{R}}|$ **do**
    Sample a grade score record (u, i, r_{ui}) with $y_{ui} = 1$ or a binary rating record $(u, i, \tilde{r}_{ui})$ with $\tilde{y}_{ui} = 1$ from $\mathcal{R} \cup \tilde{\mathcal{R}}$.
    if $y_{ui} = 1$ **then**
      Get the set of liked items by user u, i.e., $\mathcal{P}_u$.
      Get the set of disliked items by user u, i.e., $\mathcal{N}_u$.
      Calculate the virtual user profile $A_{u\cdot}$ in Eq. (4).
      Calculate the gradients $\nabla\mu, \nabla b_u, \nabla b_i, \nabla U_{u\cdot}, \nabla V_{i\cdot}, \nabla P_{j\cdot}, j \in \mathcal{P}_u$, and $\nabla N_{j\cdot}, j \in \mathcal{N}_u$ in Eq. (7–13).
      Update the parameters $\mu, b_u, b_i, U_{u\cdot}, V_{i\cdot}, P_{j\cdot}, j \in \mathcal{P}_u$, and $N_{j\cdot}, j \in \mathcal{N}_u$ via Eq. (16).
    if $\tilde{y}_{ui} = 1$ **then**
      Calculate the gradients $\nabla W_{u\cdot}$ and $\nabla V_{i\cdot}$ in Eqs. (14–15).
      Update the parameters $W_{u\cdot}$ and $V_{i\cdot}$ via Eq. (16).
  Decrease the learning rate via $\gamma = \gamma \times 0.9$.

3.5. Analysis

When $\delta_{\mathcal{P}} = \delta_{\mathcal{N}} = 0$, TMF reduces to iTCF [20], which does not involve the virtual user profile $A_{u\cdot}$. When $\delta_{\mathcal{P}} = \delta_{\mathcal{N}} = 0$ and $\lambda = 0$, TMF further reduces to RSVD [15], which focus on user preference modeling with the target grade score records only. Hence, we may represent the relationship among TMF, iTCF and RSVD as follows,

$$\text{TMF} \xrightarrow{\delta_{\mathcal{P}}=\delta_{\mathcal{N}}=0} \text{iTCF} \xrightarrow{\lambda=0} \text{RSVD}, \tag{17}$$

from which we can see that our TMF is generic and absorbs iTCF and RSVD as special cases. Furthermore, we may fix $\lambda = 0$ and assign different values to $\delta_{\mathcal{P}}$ and $\delta_{\mathcal{N}}$, and obtain different algorithm variants of SVD, i.e., SVD+ for $(\delta_{\mathcal{P}}, \delta_{\mathcal{N}}) = (1, 0)$, SVD- for $(\delta_{\mathcal{P}}, \delta_{\mathcal{N}}) = (0, 1)$, and SVD+- for $(\delta_{\mathcal{P}}, \delta_{\mathcal{N}}) = (1, 1)$. Similarly, we can obtain TMF+, TMF- and TMF+- when $\lambda > 0$. We will study the effect of different configurations of $(\delta_{\mathcal{P}}, \delta_{\mathcal{N}})$ in our experiments.

4. Experiments

4.1. Data sets and evaluation metrics

We use two data sets from [20], including MovieLens10M (denoted as ML10M)[1] and Flixter[2] [40], each of which contains five copies of (i) target grade score records in the form of (u, i, r_{ui}) with $r_{ui} \in \mathbb{G} = \{0.5, 1, 1.5, \ldots, 5\}$, (ii) auxiliary binary rating records in the form of $(u, i, \tilde{r}_{ui})$ with $\tilde{r}_{ui} \in \mathbb{B} = \{\text{like}, \text{dislike}\}$, and test grade score records in the form of (u, i, r_{ui}) with $r_{ui} \in \mathbb{G}$. Note that the auxiliary binary rating records are constructed via converting ratings larger than or equal to 4 as likes and the others as dislikes [20].

[1] http://grouplens.org/datasets/movielens/
[2] http://www.cs.sfu.ca/~sja25/personal/datasets/

Table 2
Prediction performance of TMF and other methods on ML10M and Flixter. Note that the results of RSVD, CMF and iTCF are copied from [20].

Data	Algorithm	MAE	RMSE
ML10M	AF	0.6766 ± 0.0006	0.8735 ± 0.0007
	RSVD	0.6438 ± 0.0011	0.8364 ± 0.0012
	CMF	0.6334 ± 0.0012	0.8273 ± 0.0013
	iTCF	0.6197 ± 0.0006	0.8091 ± 0.0008
	TMF	**0.6124** ± 0.0007	**0.8005** ± 0.0008
Flixter	AF	0.6867 ± 0.0005	0.9128 ± 0.0007
	RSVD	0.6561 ± 0.0007	0.8814 ± 0.0010
	CMF	0.6423 ± 0.0009	0.8710 ± 0.0012
	iTCF	0.6373 ± 0.0005	0.8636 ± 0.0010
	TMF	**0.6348** ± 0.0007	**0.8615** ± 0.0012

The ML10M data set is associated with $n = 71,567$ users and $m = 10,681$ items. Each copy contains $|\mathcal{R}| = 4,000,022$ target records, $|\tilde{\mathcal{R}}| = 4,000,022$ auxiliary records and $|\mathcal{T}_E| = 2,000,010$ test records, where the ratio is $|\mathcal{R}| : |\tilde{\mathcal{R}}| : |\mathcal{T}_E| = 2 : 2 : 1$.

The Flixter data set is associated with $n = 147,612$ users and $m = 48,794$ items. Each copy contains $|\mathcal{R}| = 3,278,431$ target records, $|\tilde{\mathcal{R}}| = 3,278,431$ auxiliary records and $|\mathcal{T}_E| = 1,639,215$ test records, where the ratio is also $|\mathcal{R}| : |\tilde{\mathcal{R}}| : |\mathcal{T}_E| = 2 : 2 : 1$.

We conduct experiments on those five copies of each data set and report the average results on two popular evaluation metrics, including Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

4.2. Baselines and parameter settings

We study the effectiveness of our mixed factorization based transfer learning framework, i.e., TMF, with some state-of-the-art methods, including regularized Singular value decomposition (RSVD) [15], collective matrix factorization (CMF) [21] and interaction-rich transfer by collective factorization (iTCF) [20].

- RSVD approximates an observed target grade score via learning some latent variables of the corresponding user and item, which works well for data of grade scores;
- CMF extends RSVD via sharing items' latent variables for the target grade scores and the auxiliary binary ratings, which can be applied to our CR-HEF problem; and
- iTCF further extends CMF via introducing interactions between users' latent variables, which was reported to be more accurate than CMF.

We choose the above methods as our major baselines because they are closely related to our TMF as shown in Fig. 2. The comparative empirical studies are designed to verify the effectiveness of the extensions from RSVD, CMF and iTCF to our TMF.

We also include algorithm variants of Singular value decomposition (SVD) with different configurations, including SVD++, SVD-, SVD+ and SVD+- for leveraging different subsets or combinations of positive feedbacks and negative feedbacks.

Furthermore, we also include a simple but effective method for grade score prediction, i.e., average filling (AF) via $b_u + b_i + \mu$, where $b_u = \sum_i y_{ui} r_{ui} / \sum_i y_{ui} - \mu$, $b_i = \sum_u y_{ui} r_{ui} / \sum_u y_{ui} - \mu$ and $\mu = \sum_{u,i} y_{ui} r_{ui} / \sum_{u,i} y_{ui}$ are statistics calculated from the target grade score records $\mathcal{R}$.

The model parameters in factorization-based methods are initialized with small random values in the same with that of [20]. The weight on auxiliary data λ, the weight for interactions ρ, the latent dimension d, the tradeoff on regularization terms, and the iteration number T are also set the same, i.e., $\lambda = 1$, $\rho = 0.5$, $\alpha_u = \alpha_v = \alpha_w = \beta_u = \beta_v = 0.01$, $d = 20$ for ML10M and $d = 10$ for Flixter, and $T = 50$. For the weight on positive feedbacks and negative feedbacks, we first fix $w_p = 2$, $w_n = 1$ to emphasize the importance of positive feedbacks, and then change the value of $w_p \in \{1, 2, 3, 4, 5\}$ in order to investigate the effect of positive feedbacks and negative feedbacks. Note that for the auxiliary binary ratings, we replace "like" with grade score 5 and "dislike" with grade score 1 [20].

4.3. Experimental results

4.3.1. Main results

The prediction performance of AF, RSVD, CMF, iTCF and our TMF on ML10M and Flixter are shown in Table 2. We can see that the overall performance ordering of the studied methods is AF < RSVD < CMF < iTCF < TMF, which demonstrates the effectiveness of the factorization-based methods (as compared with the average filling baseline), in particular of the designed *mixed* factorization approach. We conduct significant test and find that our TMF is significantly better than all other methods on both data sets (the p-value[3] is smaller than 0.01).

[3] http://www.mathworks.com/help/stats/ttest2.html

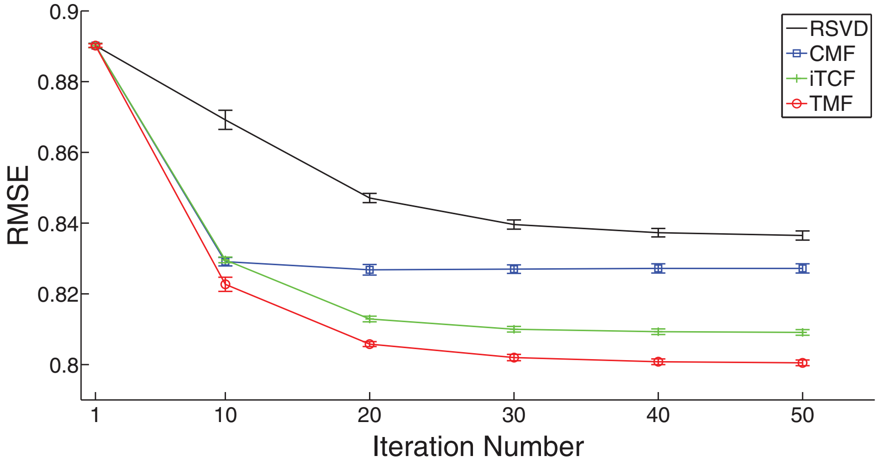

Fig. 3. Prediction performance of RSVD, CMF, iTCF and TMF with different iteration numbers on ML10M.

Table 3
Prediction performance of algorithm variants of the SVD family and TMF family on ML10M and Flixter, which are associated with different configurations of "++", "−", "+" and "+−".

Data	Metric	Conf.	SVD family	TMF family
ML10M	MAE	++	0.6285 ± 0.0006	0.6169 ± 0.0006
		−	0.6305 ± 0.0006	0.6176 ± 0.0003
		+	0.6233 ± 0.0005	0.6152 ± 0.0003
		+−	0.6206 ± 0.0006	**0.6124** ± 0.0007
	RMSE	++	0.8187 ± 0.0007	0.8058 ± 0.0008
		−	0.8211 ± 0.0009	0.8066 ± 0.0006
		+	0.8123 ± 0.0007	0.8036 ± 0.0007
		+−	0.8093 ± 0.0008	**0.8005** ± 0.0008
Flixter	MAE	++	0.6494 ± 0.0008	0.6373 ± 0.0008
		−	0.6479 ± 0.0008	0.6373 ± 0.0011
		+	0.6456 ± 0.0012	0.6366 ± 0.0008
		+−	0.6422 ± 0.0011	**0.6348** ± 0.0007
	RMSE	++	0.8747 ± 0.0009	0.8635 ± 0.0011
		−	0.8734 ± 0.0008	0.8633 ± 0.0011
		+	0.8709 ± 0.0013	0.8626 ± 0.0012
		+−	0.8680 ± 0.0011	**0.8615** ± 0.0012

We also show the recommendation performance of the factorization methods RSVD, CMF, iTCF and our TMF with different iteration numbers on ML10M in Fig. 3. We can again see the superior prediction ability of our TMF over other methods. Specifically, the factorization-based methods' performance at different iterations is consistent with that in Table 2 when the model parameters are sufficiently trained.

4.3.2. *Effect of positive and negative feedbacks*

In order to have a deep understanding of the effect of positive feedbacks (i.e., likes) and negative feedbacks (i.e., dislikes), we also study two families of algorithm variants of SVD and TMF with different configurations, including (i) "++" for a union set of both positive and negative feedbacks without distinction, i.e., $\mathcal{P}_u \cup \mathcal{N}_u$ for user u, (ii) "−" for a set of negative feedbacks only, (iii) "+" for a set of positive feedbacks only, and (iv) "+−" for one set of positive feedbacks and one set of negative feedbacks. The prediction performance are shown in Table 3. We can see that the overall performance ordering of the algorithms with different configurations from either the SVD family or the TMF family is "++" ≈ "−" < "+" < "+−", from which we can have the following observations;

- a simple combination of positive feedbacks and negative feedbacks without distinction is harmful since the configuration "++" does not perform well;
- positive feedbacks are more useful than negative feedbacks in modeling users' preferences, which can be explained by the fact that users usually prefer to assign grade scores to liked items than to disliked items, e.g., the global average preference scores of ML10M and Flixter are 3.51 and 3.61, respectively; and
- positive feedbacks and negative feedbacks are complementary for the prediction performance because the algorithm variant with configuration "+−" is better than that with either "+" or "−" on both ML10M and Flixter.

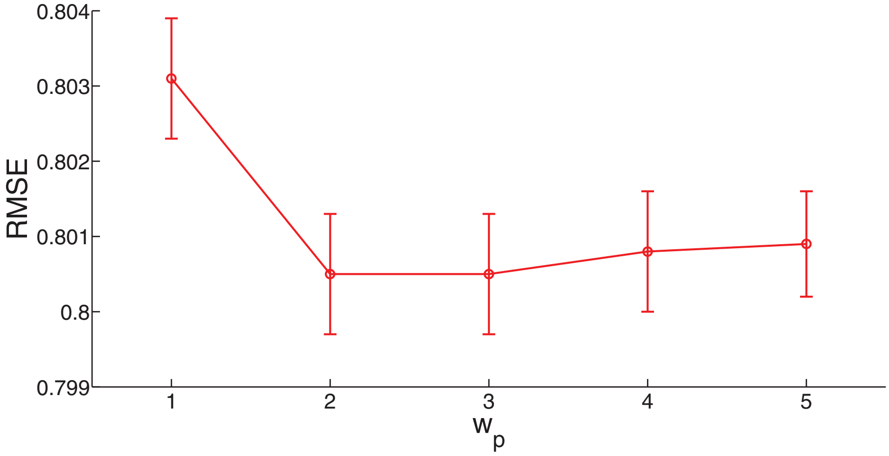

Fig. 4. Prediction performance of TMF with different values of w_p on ML10M.

Note that if no user avoids assigning a low score to an item when he or she disliked an item, the usefulness of negative feedbacks may be shown more clearly, because those information can be used to exclude some items in recommendation.

4.3.3. *Effect of different weight on positive feedbacks*

We further study the relative importance of positive feedbacks and negative feedbacks via adjusting the weight on positive feedbacks, i.e., $w_p \in \{1, 2, 3, 4, 5\}$. We show the prediction performance of TMF with different values of w_p in Fig. 4. We can see that putting more weight on positive feedbacks, e.g., $w_p = 2$ or $w_p = 3$, can usually generate better performance, which echoes the observations from Table 3, i.e., positive feedbacks are more useful in modeling users' preferences.

5. Conclusions and future work

In this paper, we propose a novel and generic mixed factorization based transfer learning framework, i.e., *transfer by mixed factorization* (TMF), for collaborative recommendation with heterogeneous explicit feedbacks (CR-HEF). Specifically, TMF unifies two state-of-the-art factorization models in transfer learning and collaborative recommendation, i.e., feature-based transfer via collective factorization [20,21] and instance-based transfer via integrative factorization [15,32], in one single framework in a principled way. TMF is able to model users' preferences more accurately via transferring more feedback-independent knowledge than either collective factorization or integrative factorization alone. Experimental results on two real-world data sets show that our TMF can achieve significantly better prediction performance than the state-of-the-art methods on two CR-HEF tasks. Furthermore, we study the effect of different parts or combinations of positive feedbacks and negative feedbacks of binary ratings, and find that our TMF can leverage each part of auxiliary feedbacks significantly better than the state-of-the-art method.

For future works, we are mainly interested in generalizing our TMF with temporal context in a multi-objective oriented optimization manner.

Acknowledgment

We thank the support of Natural Science Foundation of Guangdong Province No. 2014A030310268, National Natural Science Foundation of China (NSFC) No. 61502307,61170077,61272303, and Natural Science Foundation of SZU No. 201436. We also thank the editors and reviewers for their constructive and expert comments, and our colleague George Basker for his help on linguistic quality improvement.

References

[1] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, IEEE Trans. Knowl. Data Eng. 17 (6) (2005) 734–749.
[2] J. Bobadilla, F. Ortega, A. Hernando, A. GutiéRrez, Recommender systems survey, Knowl. Based Syst. 46 (2013) 109–132.
[3] O. Küçüktunç, E. Saule, K. Kaya, U.V. Çatalyürek, Diversifying citation recommendations, ACM Trans. Intell. Syst. Technol. 5 (4) (2014) 55:1–55:21, doi:10.1145/2668106.
[4] A. Tejeda-Lorente, C. Porcel, E. Peis, R. Sanz, E. Herrera-Viedma, A quality based recommender system to disseminate information in a university digital library, Inf. Sci. 261 (2014) 52–69, doi:10.1016/j.ins.2013.10.036.
[5] E. Amolochitis, I.T. Christou, Z.-H. Tan, Implementing a commercial-strength parallel hybrid movie recommendation engine, IEEE Intell. Syst. 29 (2) (2014) 92–96, doi:10.1109/MIS.2014.23.
[6] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, D. Sampath, The youtube video recommendation system, in: Proceedings of the 4th ACM Conference on Recommender Systems, 2010, pp. 293–296.
[7] C. Biancalana, F. Gasparetti, A. Micarelli, G. Sansonetti, An approach to social recommendation for context-aware mobile services, ACM Trans. Intell. Syst. Technol. 4 (1) (2013) 10:1–10:31, doi:10.1145/2414425.2414435.

[8] Z. Zhang, H. Lin, K. Liu, D. Wu, G. Zhang, J. Lu, A hybrid fuzzy-based personalized recommender system for telecom products/services, Inf. Sci. 235 (2013) 117–129, doi:10.1016/j.ins.2013.01.025.
[9] L.A.S. Pizzato, T. Rej, J. Akehurst, I. Koprinska, K. Yacef, J. Kay, Recommending people to people: the nature of reciprocal recommenders with a case study in online dating, User Model. User-Adapt. Interact. 23 (5) (2013) 447–488, doi:10.1007/s11257-012-9125-0.
[10] A. Sharma, B. Yan, Pairwise learning in recommendation: experiments with community recommendation on linkedin, in: Proceedings of the 7th ACM Conference on Recommender Systems, 2013, pp. 193–200.
[11] D. Goldberg, D. Nichols, B.M. Oki, D. Terry, Using collaborative filtering to weave an information tapestry, Communications of the ACM (CACM) 35 (12) (1992) 61–70.
[12] T. Segaran, Programming Collective Intelligence, first ed., O'Reilly, 2007.
[13] B. Li, Q. Yang, X. Xue, Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction, in: Proceedings of the 21st International Joint Conference on Artificial Intelligence, 2009, pp. 2052–2057.
[14] W. Pan, E.W. Xiang, N.N. Liu, Q. Yang, Transfer learning in collaborative filtering for sparsity reduction, in: Proceedings of the 24th AAAI Conference on Artificial Intelligence, 2010, pp. 230–235.
[15] Y. Koren, Factorization meets the neighborhood: a multifaceted collaborative filtering model, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008, pp. 426–434.
[16] S. Rendle, Factorization machines with libfm, ACM Trans. Intell. Syst. Technol. 3 (3) (2012) 57:1–57:22, doi:10.1145/2168752.2168771.
[17] W. Pan, Q. Yang, Transfer learning in heterogeneous collaborative filtering domains, Artif. Intell. 197 (2013) 39–55.
[18] S. Kabbur, X. Ning, G. Karypis, FISM: Factored item similarity models for top-n recommender systems, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013, pp. 659–667.
[19] S. Rendle, C. Freudenthaler, Z. Gantner, S.-T. Lars, BPR: Bayesian personalized ranking from implicit feedback, in: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, 2009, pp. 452–461.
[20] W. Pan, Z. Ming, Interaction-rich transfer learning for collaborative filtering with heterogeneous user feedbacks, IEEE Intell. Syst. 29 (6) (2014) 48–54.
[21] A.P. Singh, G.J. Gordon, Relational learning via collective matrix factorization, in: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008, pp. 650–658.
[22] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359.
[23] W. Dai, Q. Yang, G.-R. Xue, Y. Yu, Boosting for transfer learning, in: Proceedings of the 24th International Conference on Machine Learning, 2007, pp. 193–200.
[24] Y. Zhu, Y. Chen, Z. Lu, S.J. Pan, G.-R. Xue, Y. Yu, Q. Yang, Heterogeneous transfer learning for image classification, in: Proceedings of the 25th AAAI Conference on Artificial Intelligence, 2011, pp. 1304–1309.
[25] E. Zhong, W. Fan, Q. Yang, User behavior learning and transfer in composite social networks, ACM Trans. Knowl. Discov. Data 8 (1) (2014) 6:1–6:32, doi:10.1145/2556613.
[26] B. Li, Q. Yang, X. Xue, Transfer learning for collaborative filtering via a rating-matrix generative model, in: Proceedings of the 26th Annual International Conference on Machine Learning, 2009, pp. 617–624.
[27] B. Cao, N.N. Liu, Q. Yang, Transfer learning for collective link prediction in multiple heterogenous domains, in: Proceedings of the 27th International Conference on Machine Learning, 2010, pp. 159–166.
[28] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, C. Zhu, Personalized recommendation via cross-domain triadic factorization, in: Proceedings of the 22nd International Conference on World Wide Web, 2013, pp. 595–606.
[29] B. Loni, Y. Shi, M.A. Larson, A. Hanjalic, Cross-domain collaborative filtering with factorization machines, in: Proceedings of the 36th European Conference on Information Retrieval, 2014.
[30] I. Fernandez-Tobias, I. Cantador, M. Kaminskas, F. Ricci, Cross-domain recommender systems: a survey of the state of the art, in: Proceedings of the 2nd Spanish Coference on Information Retrieval, 2012.
[31] B. Li, Cross-domain collaborative filtering: a brief survey, in: Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence, 2011, pp. 1085–1086.
[32] W. Pan, E.W. Xiang, Q. Yang, Transfer learning in collaborative filtering via uncertain ratings, in: Proceedings of the 26th AAAI Conference on Artificial Intelligence, 2012, pp. 662–668.
[33] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, SIAM Review 51 (3) (2009) 455–500.
[34] R. Salakhutdinov, A. Mnih, Probabilistic matrix factorization, in: Proceedings of Annual Conference on Neural Information Processing Systems, 2008, pp. 1257–1264.
[35] Y. Shi, A. Karatzoglou, L. Baltrunas, M. Larson, N. Oliver, A. Hanjalic, CLiMF: Learning to maximize reciprocal rank with collaborative less-is-more filtering, in: Proceedings of the 6th ACM Conference on Recommender Systems, 2012, pp. 139–146.
[36] Y. Shi, M. Larson, A. Hanjalic, List-wise learning to rank with matrix factorization for collaborative filtering, in: Proceedings of the 4th ACM Conference on Recommender Systems, 2010, pp. 269–272.
[37] M. Weimer, A. Karatzoglou, A. Smola, Improving maximum margin matrix factorization, Mach. Learn. 72 (3) (2008) 263–276.
[38] L. Du, X. Li, Y.-D. Shen, User graph regularized pairwise matrix factorization for item recommendation, in: Proceedings of the 7th International Conference on Advanced Data Mining and Applications, 2011, pp. 372–385.
[39] B. Kanagal, A. Ahmed, S. Pandey, V. Josifovski, J. Yuan, L. Garcia-Pueyo, Supercharging recommender systems using taxonomies for learning user purchase behavior, Proc. VLDB Endow. 5 (10) (2012) 956–967.
[40] M. Jamali, M. Ester, A matrix factorization technique with trust propagation for recommendation in social networks, in: Proceedings of the 4th ACM Conference on Recommender Systems, 2010, pp. 135–142.