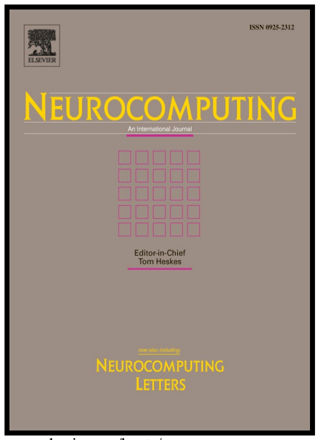
Author's Accepted Manuscript



www.elsevier.com/locate/neucom

A Survey of Transfer Learning for Collaborative Recommendation with Auxiliary Data

Weike Pan

PII: S0925-2312(15)01864-0
DOI: http://dx.doi.org/10.1016/j.neucom.2015.11.059
Reference: NEUCOM16455

To appear in: *Neurocomputing*

Received date: 2 March 2015
Revised date: 9 November 2015
Accepted date: 23 November 2015

Cite this article as: Weike Pan, A Survey of Transfer Learning for Collaborative Recommendation with Auxiliary Data, *Neurocomputing* http://dx.doi.org/10.1016/j.neucom.2015.11.059

A Survey of Transfer Learning for Collaborative Recommendation with Auxiliary Data

Weike Pan
College of Computer Science and Software Engineering, Shenzhen University
panweike@szu.edu.cn

Abstract

Intelligent recommendation technology has been playing an increasingly important role in various industry applications such as e-commerce product promotion and Internet advertisement display. Besides user feedbacks (e.g., numerical ratings) on items as usually exploited by some typical recommendation algorithms, there are often some additional data such as users' social circles and other behaviors. Such auxiliary data are usually related to user preferences on items behind numerical ratings. *Collaborative recommendation with auxiliary data* (CRAD) aims to leverage such additional information so as to improve personalized services. It has received much attention from both researchers and practitioners.

Transfer learning (TL) is proposed to extract and transfer knowledge from some auxiliary data in order to assist the learning task on the target data. In this survey, we consider the CRAD problem from a transfer learning view, especially on how to enable knowledge transfer from some auxiliary data, and discuss the representative transfer learning techniques. Firstly, we give a formal definition of transfer learning for CRAD (TL-CRAD). Secondly, we extend the existing categorization of TL techniques with three knowledge transfer strategies. Thirdly, we propose a novel and generic knowledge transfer framework for TL-CRAD.

Fourthly, we describe some representative works of each specific knowledge transfer strategy in detail, which are expected to inspire further works. Finally, we conclude the survey with some summarized discussions and several future directions.

Keywords: Collaborative Recommendation, Auxiliary Data, Transfer Learning

1. Introduction

Intelligent recommendation technology [1, 4, 18, 31, 45, 48] has been a standard component embedded in many Internet systems such as e-commerce and advertisement systems to provide personalized services. There are two main approaches widely used in personalized recommendation for an active user, i.e., content-based recommendation [3] and collaborative recommendation [14]. Content-based methods promote an item based on the relevance between a candidate item and the active user's consumed items, while collaborative recommendation techniques focus on collective intelligence and exploit the community's data so as to recommend preferred items from users with similar tastes. However, both methods are limited to users' feedbacks of explicit scores or implicit examinations, which may result in a challenging problem, data sparsity, due to the lack of users' behaviors.

Fortunately, there are often some additionally available data besides the users' feedbacks (e.g., numerical ratings) in a recommender system. There are at least four types of auxiliary data as shown in Table 1, such as content information [52, 56], time contextual information [23, 36], social or information networks [21, 49, 54] and additional feedbacks [19, 29, 39]. These auxiliary data have the potential to help relieve the aforementioned sparsity problem and thus improve the recommendation performance. In this survey, we study on how to exploit different types of auxiliary data in collaborative recommendation, which is coined as *collaborative recommendation with auxiliary data* (CRAD).

Specifically, we study the CRAD problem from an *inductive transfer learning* [37] view (instead of unsupervised or transductive transfer learning views [2]), in which we consider the users' feedback data as our *target data* or supervised information, and all the other additional information as our *auxiliary data*. In particular, we focus on how to enable knowledge transfer from some auxiliary data to the target data in order to address the aforementioned sparsity challenge. We discuss some representative transfer learning techniques, aiming to answer the fundamental question of transfer learning [37], i.e., "how to transfer". With this focus in our survey, we extend previous categorization of transfer learning techniques in collaborative filtering [38, 43], and answer the above question from two dimensions, including *knowledge transfer algorithm styles* (i.e., adaptive, collective and integrative knowledge transfer) and *knowledge transfer strategies* (i.e., prediction rule, regularization and constraint). Then, we propose a novel and generic knowledge transfer framework and describe some representative works in each category to answer the "how to transfer" question in detail, in particular the main idea that may be generalized to other applications. Finally, we conclude the survey with some summarized discussions and several exciting future directions.

2. Transfer Learning for Collaborative Recommendation with Auxiliary Data

2.1. Problem Definition

We have a target data set and an auxiliary data set. In the target data set, we have some feedbacks from n users and m items, which is usually represented as a rating matrix $\mathbf{R} = [r_{ui}]^{n\times m}$ and an indicator matrix $\mathbf{Y} \in \{0,1\}^{n\times m}$, where

Table 1: List of auxiliary data.

Content	
	user's static profile of demographics, affiliations, etc.
	item's static description of price, brand, location, etc.
	user-item pair's user generated content (UGC), etc.
Context	
	user's dynamic context of mood, health, etc.
	item's dynamic context of remaining quantities, etc.
	user-item pair's dynamic context of time, etc.
Network	
	user-user social network of friendship, etc.
	item-item relevance network of taxonomy, etc.
	user-item-user network of sharing items with friends, etc.
Feedback	
	user's feedback of rating on other items, etc.
	item's feedback of browsing by other users, etc.
	user-item pair's feedback of collection, etc.

$y_{ui} = 1$ means that the feedback r_{ui} is observed. In the auxiliary data set, we have some additional data such as content, context, network and feedback information as shown in Table 1. Our goal is to predict the unobserved feedbacks in $\mathbf{R}$ by transferring knowledge from the available auxiliary data. We illustrate the studied problem in Figure 1, where the left part is the target data of user feedbacks and the right part denotes different types of auxiliary data.

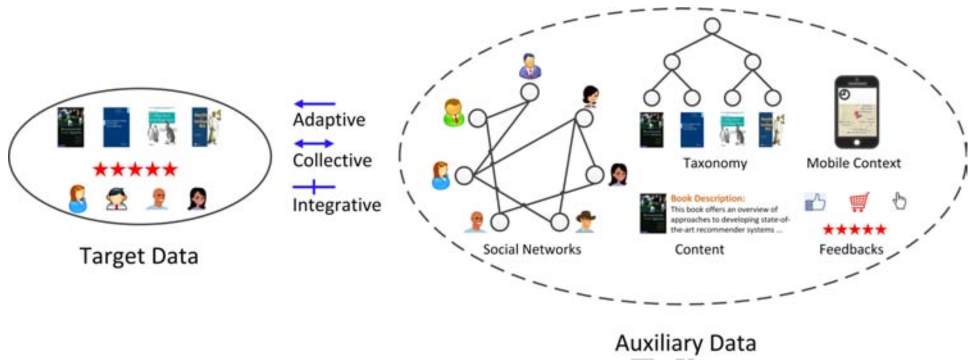


Figure 1: Illustration of Transfer Learning for Collaborative Recommendation with Auxiliary Data (TL-CRAD).

2.2. *Categorization of Transfer Learning Techniques*

Following the fundamental question of "how to transfer" in transfer learning [37, 43], we first categorize various transfer learning algorithms into (i) adaptive knowledge transfer, (ii) collective knowledge transfer and (iii) integrative knowledge transfer w.r.t. *knowledge transfer algorithm styles*. For each type of algorithm styles, we then study the related works in three specific *knowledge transfer strategies*, including (i) transfer via prediction rule, (ii) transfer via regularization and (iii) transfer via constraint, which are closely related to the three parts of a typical optimization problem [5], i.e., loss function, regularization and constraint.

Note that the binary categorization of adaptive knowledge transfer and collective knowledge transfer was first briefly described in [38], and was later expanded with one more category of integrative knowledge transfer in [43]. And in this survey, we further expand it with three specific knowledge transfer strategies in each algorithm style.

2.3. A Generic Knowledge Transfer Framework

We mainly survey some recent works of low-rank transfer learning methods for collaborative recommendation with auxiliary data (CRAD), in particular of matrix factorization based methods. The prosperity of matrix factorization based methods is mostly due to many successful stories in various public competitions and reported industry applications. Matrix factorization based methods are also the state-of-the-art in TL-CRAD because they are able to digest the sparse rating data well via learning latent variables and are also flexible to incorporate different types of auxiliary data.

Mathematically, matrix factorization based methods can be formulated with a loss function and a regularization term, i.e., $\min_{\Theta} \mathcal{E}(\Theta|\mathbf{R}) + \mathcal{R}(\Theta)$, where Θ is the model parameter. We extend such basic formulation and propose a novel and generic framework for TL-CRAD,

$$\begin{aligned} &\min_{\Theta,\mathbb{K}} \ \mathcal{E}(\Theta, \mathbb{K}|\mathbf{R}, \mathbb{A}) + \mathcal{R}(\Theta|\mathbb{K}, \mathbb{A}) + \mathcal{R}(\mathbb{K}), \\ &\text{s.t. } \Theta \in \mathcal{C}(\mathbb{K}, \mathbb{A}), \end{aligned} \tag{1}$$

which contains a loss function $\mathcal{E}(\Theta, \mathbb{K}|\mathbf{R}, \mathbb{A})$, two regularization terms $\mathcal{R}(\Theta|\mathbb{K}, \mathbb{A})$ and $\mathcal{R}(\mathbb{K})$, and a constraint $\Theta \in \mathcal{C}(\mathbb{K}, \mathbb{A})$. Specifically, $\mathbf{R}$ is the target user-item rating matrix, $\mathbb{A}$ is the auxiliary data, $\mathbb{K}$ is the extracted knowledge from $\mathbb{A}$, and Θ is the model parameter. Note that the prediction rule is not explicitly shown but embedded in the loss function $\mathcal{E}(\Theta, \mathbb{K}|\mathbf{R}, \mathbb{A})$. In the following sections, we will describe some representative works of TL-CRAD, which are instantiations of the generic framework in Eq.(1).

3. Adaptive Knowledge Transfer

Adaptive knowledge transfer aims to *adapt* the knowledge extracted from an auxiliary data domain to a target data domain. This is a *directed* knowledge transfer approach similar to traditional domain adaptation methods. In this section, we describe two adaptive knowledge transfer strategies as instantiated from Eq.(1), including (i) transfer via regularization, $\min_{\Theta} \mathcal{E}(\Theta|\mathbf{R}) + \mathcal{R}(\Theta|\mathbb{K})$, and (ii) transfer via constraint, $\min_{\Theta} \mathcal{E}(\Theta|\mathbf{R})$, s.t. $\Theta \in \mathcal{C}(\mathbb{K})$.

3.1. *Transfer via Regularization*

CST (Coordinate System Transfer) CST [42] studies knowledge transfer from auxiliary implicit feedbacks of browsing records to target explicit feedbacks of ratings. Specifically, it incorporates the coordinate systems (or latent features) extracted from auxiliary data into the target factorization system (i.e., $\mathbf{Y} \odot \mathbf{R} \sim \mathbf{U}\mathbf{B}\mathbf{V}^T$) via two biased regularization terms [42],

$$||\mathbf{U} - \dot{\mathbf{U}}||_F^2 + ||\mathbf{V} - \ddot{\mathbf{V}}||_F^2, \tag{2}$$

where $\dot{\mathbf{U}} \in \mathbb{R}^{n \times d}$ and $\ddot{\mathbf{V}} \in \mathbb{R}^{m \times d}$ are the user-specific feature matrix and the item-specific feature matrix, respectively. The two biased regularization terms in Eq.(2) are used to constrain the latent feature matrices $\mathbf{U} \in \mathbb{R}^{n \times d}$ and $\mathbf{V} \in \mathbb{R}^{m \times d}$ to be similar to $\dot{\mathbf{U}}$ and $\ddot{\mathbf{V}}$, respectively. Note that the factorization system

$\mathbf{Y} \odot \mathbf{R} \sim \mathbf{U}\mathbf{B}\mathbf{V}^T$ denotes the approximation of $\mathbf{Y} \odot \mathbf{R}$ via matrix tri-factorization $\mathbf{U}\mathbf{B}\mathbf{V}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices (i.e., $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$). Empirically, CST [42] works well when the auxiliary data are dense while the target ratings are few.

Biased regularization is a classical approach commonly used in machine learning, in particular of domain adaptation methods [22]. Note that it may also be considered as a soft constraint as compared with the hard constraint in collective knowledge transfer methods [38, 52] in Section 4.1.

3.2. Transfer via Constraint

CBT (CodeBook Transfer) CBT [26] is an early transfer learning algorithm, which studies knowledge transferability between two distinct data domains, which may be regarded as *far transfer of learning* in psychology [15]. Specifically, it transfers knowledge of cluster-level rating behavior from auxiliary data of movies to target data of books. Firstly, a cluster-level rating pattern (a.k.a., codebook), $\breve{\mathbf{B}} \in \mathbb{R}^{d\times d}$, is constructed from the auxiliary data $\breve{\mathbf{R}} \in \mathbb{R}^{\breve{n}\times\breve{m}}$ via some co-clustering algorithm, where each entry of $\breve{\mathbf{B}}$ denotes the average rating of the corresponding co-cluster. Secondly, the codebook is transferred to the target data via codebook expansion $\mathbf{U}\mathbf{B}\mathbf{V}^T$ with the following constraint [26],

$$\mathbf{B} = \breve{\mathbf{B}}, \tag{3}$$

which means that the rating pattern is shared between target data and auxiliary data. Note that $\mathbf{U} \in \{0,1\}^{n\times d}$ and $\mathbf{V} \in \{0,1\}^{m\times d}$ are membership indicator matrices.

A later extension called RMGM (Rating-Matrix Generative Model) [27] combines codebook construction and codebook expansion in CBT [26] into one single step with soft membership indicator matrices. Considering the existence of more than one auxiliary data, the codebook in CBT [26] may also be extended to multiple codebooks with different correlation weights [35]. Furthermore, a recent work generalizes the codebook by including a data-independent rating pattern and a data-dependent rating pattern, which is shown to be more accurate than sharing the data-independent common knowledge only [13].

The idea of transferring compact group-level knowledge may also be applied to other applications such as text mining and bio-informatics. For example, a matrix tri-factorization-based classification framework (MTrick) [61] extends CBT [26] and RMGM [27] with supervised label information and studies its effectiveness in cross-domain document categorization. Furthermore, the 2-D cluster-level rating pattern may also be generalized to high dimensions, such as transferring a 3-D cluster of knowledge from an auxiliary tagging data in the form of (user, item, tag) to a target one [11].

Cluster-level rating pattern in the above works is a kind of group behavior, which is more stable than individual behavior and has higher transferability. It is thus particularly useful when the explicit correspondences or overlaps are not available between entities of the target data and the auxiliary data.

4. Collective Knowledge Transfer

Collective knowledge transfer usually *jointly* learns the shared knowledge and unshared effect of the target data and the auxiliary data simultaneously, which is a *bi-directed* knowledge transfer approach with richer interactions similar to multi-task learning algorithms. We describe some representative works of collective knowledge transfer via constraint on model parameters, $\min_{\Theta,\mathbb{K}} \mathcal{E}(\Theta|\mathbf{R}) + \mathcal{R}(\Theta) +$

$\mathcal{E}(\mathbb{K}|\mathbb{A})+\mathcal{R}(\mathbb{K})$, s.t. $\Theta \in \mathcal{C}(\mathbb{K})$, which is also an instantiation of Eq.(1). Note that the model parameter Θ and the shared knowledge $\mathbb{K}$ are learned simultaneously, instead of the two-step style adopted in adaptive knowledge transfer.

4.1. Transfer via Constraint

CMF (Collective Matrix Factorization) CMF [52] is proposed to collectively factorize one user-item rating matrix $\mathbf{R} \in \mathbb{R}^{n\times m}$, $\mathbf{Y}\odot\mathbf{R}\sim\mathbf{U}\mathbf{V}^T$, and one item-content matrix $\ddot{\mathbf{R}} \in \mathbb{R}^{\ddot{n}\times m}$, $\ddot{\mathbf{R}}\sim\ddot{\mathbf{U}}\ddot{\mathbf{V}}^T$, by sharing the same item-specific latent features $\mathbf{V}$ [52],

$$\mathbf{V} = \ddot{\mathbf{V}}, \tag{4}$$

which implies that the item-specific latent feature matrix $\ddot{\mathbf{V}}$ is shared as a bridge to enable knowledge transfer between two data. We may also use different link functions on the factorized variables $f(U_{u\cdot}V_{i\cdot}^T)$ [52]. A similar model is proposed independently in the context of social recommendation [34], which generalizes the basic matrix factorization model by jointly factorizing a user-item rating matrix $\mathbf{R} \in \mathbb{R}^{n\times m}$, $\mathbf{Y}\odot\mathbf{R}\sim\mathbf{U}\mathbf{V}^T$, and a user-user social network matrix $\dot{\mathbf{R}} \in \mathbb{R}^{n\times n}$, $\dot{\mathbf{R}}\sim\dot{\mathbf{U}}\dot{\mathbf{V}}$, with the constraint of $\mathbf{U} = \dot{\mathbf{U}}$.

The underlying assumption that same users (or items) in the target data and the auxiliary data are associated with the same latent factors is quite universal. Various models with a similar spirit have been proposed to fuse user-side and/or item-side auxiliary data via sharing latent features or topic distributions. WNMCTF (Weighted Non-negative Matrix Co-Tri-Factorization) [55] follows non-negative matrix tri-factorization [24] and collectively factorizes one user-item rating matrix, one user-demographics matrix and one item-content matrix with the constraint of sharing both the same user-specific latent feature matrix and the item-specific latent feature matrix. MRMF (Multi-Relational Matrix Factorization) [28] and HYRES (HYbrid REcommendation System) [16] collectively factorize more than two matrices from both user-side and item-side content information with the same idea of sharing latent factors. JMF (Joint Matrix Factorization) [50] collectively factorizes one user-item rating matrix and one item-item similarity matrix mined from item-side auxiliary data of movies' mood descriptions. MCF-LF (Multi-domain Collaborative Filtering with Link Function) [58], CLP-GP (Collective Link Prediction via Gaussian Process) [6] and NB-MCF (Nonparametric Bayesian Multitask Collaborative Filtering) [7] apply the same idea of sharing common latent features for problems with multiple user-side auxiliary rating matrices, and learn users' preferences and similarities between the target and auxiliary data simultaneously, which are shown to be more effective as compared with sharing the latent features alone. LOCABAL (LOCAl and gloBAL) [53] collectively factorizes one user-item rating matrix weighted by users' global reputations and one user-user social matrix weighted by cosine similarities, with the constraint of sharing the same user-specific latent feature matrix. STLCF (Selective Transfer Learning for Collaborative Filtering) [32] learns the user preferences in a joint and selective manner from multiple user-aligned data via selectively transferring high quality knowledge of consistent data, which is empirically more accurate than collective knowledge transfer without selection. The effectiveness of selective knowledge transfer has also been verified in collaborative recommendation with heterogeneous implicit feedbacks [44].

TCF (Transfer by Collective Factorization) [38] collectively factorizes a 5-star numerical target data $\mathbf{R}$ and a binary like/dislike auxiliary data, and assumes that both the user-specific and item-specific latent feature matrices are the same. Besides the shared latent features, TCF uses two inner matrices to capture the data-dependent information, which is different from the inner matrix used in CBT [26] and RMGM [27] as described in Section 3.2. The strategy to share some common knowledge and to not share some specific effect is a sophisticated knowledge transfer approach, which is expected to be more applicable to other applications. A recent extension called iTCF (interaction-rich TCF) [40] achieves a good balance between the efficiency of stochastic update rules or stochastic learning in CMF [52] and the effectiveness of knowledge transfer for heterogeneous data in TCF [38]. iTCF is an efficient transfer learning algorithm, which shares user-specific latent features in a smooth manner in addition to requiring the same item-specific latent features.

5. Integrative Knowledge Transfer

Integrative knowledge transfer *incorporates* the raw auxiliary data as known knowledge into the learning task on the target data. It can be considered as an *embedded* knowledge transfer approach similar to feature engineering, information fusion and data integration methods. Mathematically speaking, we can instantiate the generic framework in Eq.(1), and have (i) transfer via prediction rule, $\min_\Theta \mathcal{E}(\Theta|\mathbf{R}, \mathbb{A})+\mathcal{R}(\Theta)$, (ii) transfer via regularization, $\min_\Theta \mathcal{E}(\Theta|\mathbf{R})+\mathcal{R}(\Theta|\mathbb{A})$, and (iii) transfer via constraint, $\min_\Theta \mathcal{E}(\Theta|\mathbf{R}) + \mathcal{R}(\Theta)$, s.t. $\Theta \in \mathcal{C}(\mathbb{A})$. It is interesting to see that we include the raw auxiliary data $\mathbb{A}$ instead of the extracted knowledge $\mathbb{K}$, which is thus different from the way of adaptive knowledge transfer as shown in Section 3.

5.1. Transfer via Prediction Rule

Typically, once a recommendation model has been built using some training data, we can use a prediction rule such as [23] $\hat{r}_{ui} = \mu + b_u + b_i + U_{u\cdot} V_{i\cdot}^T$ or $\hat{r}_{ui} = U_{u\cdot} V_{i\cdot}^T$ in order to predict user u's preference on item i. Note that $U_{u\cdot}, V_{i\cdot} \in \mathbb{R}^{1\times d}$ are user u's and item i's latent feature vectors, respectively, and μ is the global mean, b_u is user u's bias, and b_i is item i's bias.

FM (Factorization Machines) FM [47] represents the user-item feedback matrix $\mathbf{R}$ in a novel way, i.e., a design matrix $\mathbf{X} \in \{1, 0\}^{q\times(n+m)}$ and a rating vector $\boldsymbol{r} \in \{1, 2, 3, 4, 5\}^{q\times 1}$, where q is the number of ratings in $\mathbf{R}$. For a rating triple (u, i, r_{ui}) of the user-item feedback matrix $\mathbf{R}$, the corresponding row of the design matrix is $\boldsymbol{x} = \{(u, 1), (i, 1)\} \in \mathbb{R}^{1\times(n+m)}$, where the uth and $(n+i)$th entries are 1's and all other entries are 0's, and the value of the corresponding entry of the rating vector $\boldsymbol{r}$ is r_{ui}. Then, we have a revised prediction rule with pairwise interactions between latent factors [47],

$$\hat{r}_{ui} = w_0 + \sum_{j=1}^{n+m} w_j x_j + \sum_{j=1}^{n+m} \sum_{j'=j+1}^{n+m} x_j x_{j'} w_{jj'}, \tag{5}$$

where $w_{jj'}$ denotes the inner product of two latent feature vectors. With the new representation via the design matrix, we can augment it with some auxiliary data in a simple but effective pre-processing step of feature engineering, such as user-side auxiliary ratings [30]. Note that, when no auxiliary data is fused, the prediction rule is the same as that of basic matrix factorization, i.e., $\hat{r}_{ui} = w_0 + w_u + w_{n+i} + w_{u,n+i}$, where w_0 is the global mean (i.e., μ), w_u is user u's bias (i.e., b_u), w_{n+i} is item i's bias (i.e., b_i), and $w_{u,n+i} = U_{u\cdot} V_{i\cdot}^T$ is the inner product of user u's and item i's latent feature vectors.

Besides the design matrix used in FM [47], there are some other approaches to incorporate auxiliary data via a revised prediction rule. RSTE (Recommendation with Social Trust Ensemble) [33] designs a mixed prediction rule with two terms, $\hat{r}_{ui} = \lambda U_{u\cdot} V_{i\cdot}^T + (1-\lambda) \sum_{u' \in T_u^+} \tilde{e}_{u'i} U_{u'\cdot} V_{i\cdot}^T$, where T_u^+ is the set of trusted friends of user u and $\sum_{u' \in T_u^+} \tilde{e}_{u'i} U_{u'\cdot} V_{i\cdot}^T$ represents the friends' overall taste on item i. Note that $\tilde{e}_{u'i}$ is estimated from the user-side social network or the target user-item rating matrix [33]. BMFSI (Bayesian Matrix Factorization with Side Information) [46] designs an integrated prediction rule for both the target feedback data and the auxiliary data, $\hat{r}_{ui} = U_{u\cdot} V_{i\cdot}^T + \boldsymbol{w}_u \dot{\boldsymbol{x}}_{ui}^T + \boldsymbol{w}_i \ddot{\boldsymbol{x}}_{ui}^T$, where $\dot{\boldsymbol{x}}_{ui} \in \mathbb{R}^{1 \times \dot{d}_x}$ and $\ddot{\boldsymbol{x}}_{ui} \in \mathbb{R}^{1 \times \ddot{d}_x}$ are user-side and item-side raw features related to the rating at (u, i) of $\mathbf{R}$, including the rating's time information, user u's latest two ratings, the ratings provided by user u on those 5 most similar movies measured in Pearson correlation coefficient, the features of movie directors and actors extracted from Wikipedia[1], etc. Another recent feature engineering based model called SVDfeature [10] is an efficient but restricted case of FM, which is also able to incorporate users' demographics, items' descriptions and contextual information. The rich pairwise interactions in FM [47] as shown in Eq.(5) are able to capture more complex correlations among the variables, which are thus expected to generate better recommendations.

Integrating auxiliary data into the prediction rule is an effective approach for knowledge transfer, where the knowledge of the raw auxiliary data (or more specifically part of model parameters) is learned automatically. However, the revised prediction rule will also make the learning and prediction procedures more expensive regarding the time and space complexity.

[1] http://www.wikipedia.org/

5.2. *Transfer via Regularization*

The main idea of integrative transfer via regularization is to constrain the latent feature matrices or vectors to be similar between related users or items, e.g., similar users according to tagging information [59] or socially connected users [17].

TagiCoFi (Tag Informed Collaborative Filtering) TagiCoFi [59] is proposed for incorporating social tagging data into the target numerical rating data. Specifically, it first constructs a user-user similarity matrix from social tagging data and then introduces an additional regularization term to the basic matrix factorization [59],

$$\sum_{u=1}^{n} \sum_{u'=1}^{n} \dot{s}_{uu'} ||U_{u\cdot} - U_{u'\cdot}||_F^2, \tag{6}$$

where $\dot{s}_{uu'}$ is the similarity between users u and u', and thus transfers knowledge of the nearest neighbors' taste by constraining the user-specific features to be similar in the latent space.

SocialMF (Social Matrix Factorization) [17] studies the effect of trust propagation and generalizes the basic matrix factorization model by introducing a different additional regularization term from trusted friends [17],

$$\sum_{u=1}^{n} ||U_{u\cdot} - \sum_{u' \in T_u^+} \dot{s}_{uu'} U_{u'\cdot}||_F^2, \tag{7}$$

where T_u^+ is the set of trusted friends of user u (excluding user u himself/herself) and $\dot{s}_{uu'}$ is the similarity between users u and u' obtained from social networks. Similarly, it transfers knowledge of the friends' taste via constraining the user-specific features to be similar in the latent space.

We can see that the regularization term in Eq.(6) focuses on the distance between a user' feature vector and each of his/her friends' feature vectors, while the regularization term in Eq.(7) defines the distance between one user's feature vector and a weighted sum of his/her friends' feature vectors. A recent work on TV channel recommendation studies the effect of combining those two regularization terms for both users and items and obtains better recommendation performance [57], which shows the complementary effect of those two types of regularization in knowledge transfer. Another interesting point is that the regularization terms in Eq.(6), Eq.(7) and Eq.(2) can all be considered as soft constraints imposed on latent features.

Knowledge transfer via regularization usually increases the computational complexity in the training step, but remains the same in the prediction step since the prediction rule is not changed, which is thus more efficient than the aforementioned strategy of transfer via prediction rule in Section 5.1.

5.3. Transfer via Constraint

TIF (Transfer by Integrative Factorization) TIF [43] aims to incorporate knowledge from social impressions or anchoring effects, which are represented as score intervals called uncertain ratings. Specifically, it studies on how to leverage auxiliary uncertain ratings, denoted by $[a_{ui}, b_{ui}]$, to the target data of numerical ratings. Different from most previous works, it incorporates auxiliary data through some constraints defined on the score intervals in addition to the basic matrix factorization [43],

$$\hat{r}_{ui} \in \mathcal{C}(a_{ui}, b_{ui}), \tag{8}$$

where the constraint, $\hat{r}_{ui} \in \mathcal{C}(a_{ui}, b_{ui})$, requires that the estimated preference by the learned model should be in the range of the corresponding auxiliary uncertain rating.

Integrative knowledge transfer via constraint is related to knowledge-based recommendation [18], where a user's additional constraints or requirements need to be satisfied during recommendation. Incorporating auxiliary data via constraints are also flexible since auxiliary data can usually be represented as some constraints.

6. Discussions and Future Directions

6.1. Discussions

Table 2: Some representative works of transfer learning for collaborative recommendation with auxiliary data (TL-CRAD) in the perspective of "how to transfer" in transfer learning, including different knowledge transfer algorithm styles and different knowledge transfer strategies. We also include the corresponding mathematical formulations as instantiations of the generic framework in Eq.(1), i.e., $\min_{\Theta,\mathbb{K}} \mathcal{E}(\Theta,\mathbb{K}|\mathbf{R},\mathbb{A}) + \mathcal{R}(\Theta|\mathbb{K},\mathbb{A}) + \mathcal{R}(\mathbb{K}), \text{s.t.} \Theta \in \mathcal{C}(\mathbb{K},\mathbb{A})$.

Style \ Strategy	Prediction rule	Regularization	Constraint
Adaptive		$\min_{\Theta} \mathcal{E}(\Theta\|\mathbf{R}) + \underline{\mathcal{R}(\Theta\|\mathbb{K})}$ e.g., CST [42], etc.	$\min_{\Theta} \mathcal{E}(\Theta\|\mathbf{R}),\ \underline{\text{s.t. } \Theta \in \mathcal{C}(\mathbb{K})}$ e.g., CBT [26], etc.
Collective			$\min_{\Theta,\mathbb{K}} \mathcal{E}(\Theta\|\mathbf{R}) + \mathcal{R}(\Theta)$ $+\mathcal{E}(\mathbb{K}\|\mathbb{A}) + \mathcal{R}(\mathbb{K}),\ \underline{\text{s.t. } \Theta \in \mathcal{C}(\mathbb{K})}$ e.g., CMF [52], etc.
Integrative	$\min_{\Theta} \underline{\mathcal{E}(\Theta\|\mathbf{R},\mathbb{A})} + \mathcal{R}(\Theta)$ e.g., FM [47], etc.	$\min_{\Theta} \mathcal{E}(\Theta\|\mathbf{R}) + \underline{\mathcal{R}(\Theta\|\mathbb{A})}$ e.g., tagiCoFi [59], etc.	$\min_{\Theta} \mathcal{E}(\Theta\|\mathbf{R}) + \mathcal{R}(\Theta),\ \underline{\text{s.t. } \Theta \in \mathcal{C}(\mathbb{A})}$ e.g., TIF [43], etc.

We summarize some representative works of transfer learning for collaborative recommendation with auxiliary data (TL-CRAD) in Table 2. We can see that integrative knowledge transfer via prediction rule and collective knowledge transfer via constraint have recently received more attention, which are also the state-of-the-art TL-CRAD algorithms w.r.t. recommendation accuracy in corresponding problem settings. The interaction between auxiliary data and target data usually becomes richer from adaptive, collective to integrative algorithm styles, which are believed to enable more effective knowledge transfer. However, the time complexity may also increase from adaptive to integrative algorithm styles, especially of the cost caused by sophisticated prediction rules and regularization terms used in integrative knowledge transfer approaches. We can also see that there are some blank and few-work entries in Table 2, which provide opportunities for further studies.

Parallel to various *data modeling* methods in Table 2, there are also some recommendation approaches based on *user modeling*, which may be developed for TL-CRAD so as to further expand the two-dimensional categorization used in Table 2. A recent brief survey [25] studies cross-domain collaborative filtering in the perspective of *collaborative domain* (i.e., source domain and target domain in classic transfer learning [37]) and *knowledge transfer style*, which is different from our focus on "how to transfer" in transfer learning in a more general and practical recommendation problem. An extended survey of cross-domain recommendation [12] mainly focuses on *relations* between domains, including content-based relations and collaborative filtering based relations. The most recent comprehensive survey on collaborative filtering with additional information [51] focuses on different memory-based and model-based methods on exploiting rich side information. Those three surveys consistently emphasize the importance of problem settings or recommendation scenarios such as domains, relations and side information. Note that the types of auxiliary data may also be introduced as an additional dimension for TL-CRAD such as "where to transfer" and "what to transfer", or even which part of auxiliary data can be transferred [20], for different *TL settings*, which is an orthogonal dimension with our focus of *TL techniques* in this survey.

Besides our focus of TL techniques in this survey, we may also study the representative works in Table 2 from the perspective of TL settings, including four different types of auxiliary data in Table 1, and four different sides of auxiliary data, i.e., user side, item side, frontal side (or user-item interaction [51]) and that without overlap. Specifically, CST [42] is for two-side implicit feedbacks, CBT [26] and RMGM [27] are for auxiliary explicit feedbacks without overlap, CMF [52] is for item-side content, TCF [38] is for frontal-side binary feedbacks, TIF [43] is for frontal-side uncertain ratings, tagiCoFi [59] is for frontal-side tags, BMFSI [46] is for two-side features, and FM [47] is for frontal-side context or user-side content information.

In this survey, we do not include empirical studies of the surveyed representative works, because (i) different TL techniques are usually designed for different recommendation scenarios, (ii) a typical TL technique is usually developed to improve some specific non-TL techniques (e.g., techniques without leveraging auxiliary data or techniques exploiting auxiliary data without explicitly addressing the data difference), and (iii) some TL techniques are designed for different goals though for the same recommendation problem, e.g., TCF [38] and iTCF [40] are for accuracy and efficiency, respectively. Note that the aforementioned related surveys [12, 25, 51] do not include empirical evaluations either.

For ease of investigation of new transfer learning algorithms, we compile a list of popular and public data sets below:

- *MovieLens* (`http://grouplens.org/datasets/movielens/`) is a famous benchmark that contains target data of numerical ratings and auxiliary data of content (e.g., item description, tag) and context (e.g., time);

- *Amazon reviews* (`https://snap.stanford.edu/data/web-Amazon.html`) is a real industry data that contains target data of numerical ratings and auxiliary data of content (e.g., reviews, item description), context (e.g., time), network (e.g., item-item relations derived from users' transactions), and feedback (e.g., numerical ratings of a related product domain);

- *Yahoo! Music user ratings* (`https://webscope.sandbox.yahoo.com/catalog.php?datatype=c`) is used in KDD Cup 2011, which contains target data of numerical ratings and auxiliary data of context (e.g., time) and network (e.g., item-item relations reflected in the item taxonomy);

- *HetRec 2011* is used for a specific workshop held within RecSys 2011, which contains target data of numerical ratings and auxiliary data of content (e.g., tag), context (e.g., time), and network (e.g., social connections); and

- *Flixter* (`http://socialcomputing.asu.edu/datasets/Flixster`), *Epionions* and *Ciao* (`http://www.public.asu.edu/~jtang20/datasetcode/truststudy.htm`) are commonly used benchmarks with target data of numerical ratings and auxiliary data of social connections.

6.2. Future Directions

As a fertile interdisciplinary research area of recommendation and transfer learning, there are various exciting directions worth further exploration in TL-CRAD. In this section, we include several major directions w.r.t. techniques, data, objectives, explanation and security.

Heterogeneous Techniques Ensemble Different transfer learning techniques for CRAD as described in the survey have their own advantages and disadvantages regarding recommendation effectiveness, and learning and prediction efficiency. It is thus natural to design some *heterogeneous* knowledge transfer algorithm styles and *heterogeneous* knowledge transfer strategies [41] in order to achieve a better balance among flexibility, effectiveness and efficiency. Such heterogeneous TL techniques are expected to be superior than a simple combination of existing TL techniques.

Heterogeneous Data Integration An existing transfer learning technique is usually designed for a typical recommendation scenario, while a real recommendation application usually contains more than one type of auxiliary data such as social networks, mobile context and others. Hence, it is very useful to develop a unified framework for *heterogeneous* auxiliary data integration. Furthermore, with more and more available data, a scalable and distributed framework for heterogenous data is also needed.

Multi-Objective Recommendation Existing transfer learning techniques in CRAD are mainly for rating prediction and item recommendation, while a real recommender system requires a multi-objective evaluation such as accuracy, diversity and even serendipity, or even the effectiveness and quality of online network services [9] when items are of large sizes. Hence, it is well motivated to design a more sophisticated objective function with different evaluation metrics when exploiting the auxiliary data.

Explanation and Security Most of transfer learning techniques in CRAD are developed for sparsity reduction in the target data. For a real recommender system, auxiliary data may be taken as a source for explanation generation of the recommended items, and even for robustness against malicious attacks or fake views [8].

Practice of leveraging auxiliary data in collaborative recommendation via transfer learning also expands the traditional categorization of recommendation approaches with one more branch, i.e., *collaborative recommendation with auxiliary data*, in addition to the two main approaches of *content-based recommendation* and *collaborative recommendation*. Researches or practices of TL-CRAD are also quite interesting in the big data and AI era, especially of the data variety or heterogeneity as commonly known as one of the major properties of today's data [60].

6.3. Conclusion

In this survey, we have discussed three knowledge transfer strategies for collaborative recommendation with different types of auxiliary data. Generally, each strategy can be applied to any type of auxiliary data, though different strategies may result in different effectiveness and efficiency. In order to fully exploit the complementary property of different strategies, we believe that designing hybrid knowledge transfer strategies will provide better performance in most cases.

Acknowledgement

I would like to thank Prof. Qiang Yang for advice and comments, the editors and reviewers for constructive suggestions, and the support of Natural Science Foundation of Guangdong Province No. 2014A030310268, National Natural Science Foundation of China No. 61502307,61170077,61472258 and Natural Science Foundation of SZU No. 201436.

References

[1] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17:734–749, June 2005.

[2] Andrew Arnold, Ramesh Nallapati, and William W. Cohen. A comparative study of methods for transductive transfer learning. In *Workshops Proceedings of the 7th IEEE International Conference on Data Mining (ICDM 2007)*, pages 77–82, 2007.

[3] Chumki Basu, Haym Hirsh, and William Cohen. Recommendation as classification: using social and content-based information in recommendation. In *Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence*, AAAI ’98/IAAI ’98, pages 714–720, Menlo Park, CA, USA, 1998. American Association for Artificial Intelligence.

[4] J. Bobadilla, F. Ortega, A. Hernando, and A. GutiéRrez. Recommender systems survey. *Knowledge-Based Systems*, 46:109–132, 2013.

[5] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[6] Bin Cao, Nathan Nan Liu, and Qiang Yang. Transfer learning for collective link prediction in multiple heterogenous domains. In *Proceedings of the 27th International Conference on Machine Learning*, pages 159–166, June 2010.

[7] Sotirios Chatzis. Nonparametric bayesian multitask collaborative filtering. In *Proceedings of the 22nd ACM International Conference on Conference on Information and Knowledge Management*, pages 2149–2158, October–November 2013.

[8] Liang Chen, Yipeng Zhou, and Dah Ming Chiu. Analysis and detection of fake views in online video services. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 11(2s):44:1–44:20, February 2015.

[9] Liang Chen, Yipeng Zhou, and Dah-Ming Chiu. Smart streaming for online video services. *IEEE Transactions on Multimedia*, 17(4):485–497, 2015.

[10] Tianqi Chen, Weinan Zhang, Qiuxia Lu, Kailong Chen, Zhao Zheng, and Yong Yu. Svdfeature: A toolkit for feature-based collaborative filtering. *Journal of Machine Learning Research*, 13(1):3619–3622, December 2012.

[11] Wei Chen, Wynne Hsu, and Mong Li Lee. Making recommendations from multiple domains. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 892–900, August 2013.

[12] Ignacio Fernandez-Tobias, Ivan Cantador, Marius Kaminskas, and Francesco Ricci. Cross-domain recommender systems: A survey of the state of the art. In *Proceedings of the 2nd Spanish Coference on Information Retrieval*, June 2012.

[13] Sheng Gao, Hao Luo, Da Chen, Shantao Li, Patrick Gallinari, and Jun Guo. Cross-domain recommendation via cluster-level latent factor model.

In *Proceedings of the 2013 European Conference on Machine Learning and Knowledge Discovery in Databases - Part II*, pages 161–176, September 2013.

[14] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. Using collaborative filtering to weave an information tapestry. *Communnication of ACM*, 35:61–70, December 1992.

[15] Robert E. Haskell. *Transfer of Learning: Cognition, Instruction, and Reasoning*. Educational psychology series. Academic Press, September 2000.

[16] Niklas Jakob, Stefan Hagen Weber, Mark Christoph Müller, and Iryna Gurevych. Beyond the stars: exploiting free-text user reviews to improve the accuracy of movie recommendations. In *Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion*, pages 57–64, November 2009.

[17] Mohsen Jamali and Martin Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the 4th ACM Conference on Recommender Systems*, pages 135–142, September 2010.

[18] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. *Recommender Systems: An Introduction*. Cambridge University Press, New York, NY, USA, September 2010.

[19] Gawesh Jawaheer, Peter Weller, and Patty Kostkova. Modeling user preferences in recommender systems: A classification framework for explicit and

implicit user feedback. *ACM Transactions on Interactive Intelligent Systems*, 4(2):8:1–8:26, June 2014.

[20] Meng Jiang, Peng Cui, Xumin Chen, Fei Wang, Wenwu Zhu, and Shiqiang Yang. Social recommendation with cross-domain transferable knowledge. *IEEE Transactions on Knowledge and Data Engineering*, 27(11):3084–3097, 2015.

[21] Meng Jiang, Peng Cui, Fei Wang, Qiang Yang, Wenwu Zhu, and Shiqiang Yang. Social recommendation across multiple relational domains. In *21st ACM International Conference on Information and Knowledge Management (CIKM'12)*, pages 1422–1431, 2012.

[22] Wolf Kienzle and Kumar Chellapilla. Personalized handwriting recognition via biased regularization. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 457–464, June 2006.

[23] Yehuda Koren. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 447–456, June 2009.

[24] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *Annual Conference on Neural Information Processing Systems 13*, pages 556–562, December 2001.

[25] Bin Li. Cross-domain collaborative filtering: A brief survey. In *Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence*, pages 1085–1086, November 2011.

[26] Bin Li, Qiang Yang, and Xiangyang Xue. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 2052–2057, July 2009.

[27] Bin Li, Qiang Yang, and Xiangyang Xue. Transfer learning for collaborative filtering via a rating-matrix generative model. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 617–624, June 2009.

[28] Christoph Lippert, Stefan Hagen Weber, Yi Huang, Volker Tresp, Matthias Schubert, and Hans-Peter Kriegel. Relation prediction in multi-relational domains using matrix factorization. In *Workshop on Structured Input-Structured Output: Annual Conference on Neural Information Processing Systems 20*, December 2008.

[29] Nathan N. Liu, Evan W. Xiang, Min Zhao, and Qiang Yang. Unifying explicit and implicit feedback for collaborative filtering. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pages 1445–1448, October 2010.

[30] Babak Loni, Yue Shi, Martha A Larson, and Alan Hanjalic. Cross-domain collaborative filtering with factorization machines. In *Proceedings of the 36th European Conference on Information Retrieval*, April 2014.

[31] Linyuan Lü, Matus Medo, Chi Ho Yeung, Yi-Cheng Zhang, Zi-Ke Zhang, and Tao Zhou. Recommender systems. *Physics Reports*, 519(1):1–50, 2012.

[32] Zhongqi Lu, Weike Pan, Evan Wei Xiang, Qiang Yang, Lili Zhao, and Er-Heng Zhong. Selective transfer learning for cross domain recommendation. In *Proceedings of the 13th SIAM International Conference on Data Mining*, pages 641–649, May 2013.

[33] Hao Ma, Irwin King, and Michael R. Lyu. Learning to recommend with explicit and implicit social relations. *ACM Transactions on Intelligent Systems and Technology (ACM TIST)*, 2:29:1–29:19, May 2011.

[34] Hao Ma, Haixuan Yang, Michael R. Lyu, and Irwin King. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pages 931–940, October 2008.

[35] Orly Moreno, Bracha Shapira, Lior Rokach, and Guy Shani. Talmud: Transfer learning for multiple domains. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pages 425–434, October–November 2012.

[36] Jing Pan, Zhao Ma, Yanwei Pang, and Yuan Yuan. Robust probabilistic tensor analysis for time-variant collaborative filtering. *Neurocomputing*, 119:139–143, 2013.

[37] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, August 2010.

[38] Weike Pan, Nathan N. Liu, Evan W. Xiang, and Qiang Yang. Transfer learning to predict missing ratings via heterogeneous user feedbacks. In *Proceed-*

ings of the 22nd International Joint Conference on Artificial Intelligence, pages 2318–2323, July 2011.

[39] Weike Pan, Zhuode Liu, Zhong Ming, Hao Zhong, Xin Wang, and Congfu Xu. Compressed knowledge transfer via factorization machine for heterogeneous collaborative recommendation. *Knowledge-Based Systems*, 85:234–244, 2015.

[40] Weike Pan and Zhong Ming. Interaction-rich transfer learning for collaborative filtering with heterogeneous user feedback. *IEEE Intelligent Systems*, 29(6):48–54, 2014.

[41] Weike Pan, Shanchuan Xia, Zhuode Liu, Xiaogang Peng, and Zhong Ming. Mixed factorization for collaborative recommendation with heterogeneous explicit feedbacks. *Information Sciences*, 2016.

[42] Weike Pan, Evan W. Xiang, Nathan N. Liu, and Qiang Yang. Transfer learning in collaborative filtering for sparsity reduction. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, pages 230–235, July 2010.

[43] Weike Pan, Evan W. Xiang, and Qiang Yang. Transfer learning in collaborative filtering via uncertain ratings. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, pages 662–668, July 2012.

[44] Weike Pan, Hao Zhong, Congfu Xu, and Zhong Ming. Adaptive bayesian personalized ranking for heterogeneous implicit feedbacks. *Knowledge-Based Systems*, 73:173–180, 2015.

[45] Deuk Hee Park, Hyea Kyeong Kim, Il Young Choi, and Jae Kyeong Kim. A

literature review and classification of recommender systems research. *Expert Systems with Applications*, 39(11):10059–10072, 2012.

[46] Ian Porteous, Arthur Asuncion, and Max Welling. Bayesian matrix factorization with side information and dirichlet process mixtures. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, pages 563–568, July 2010.

[47] Steffen Rendle. Factorization machines with libfm. *ACM Transactions on Intelligent Systems and Technology (ACM TIST)*, 3(3):57:1–57:22, May 2012.

[48] Francesco Ricci, Francesco Ricci, Bracha Shapira, and Paul B. Kantor. *Recommender Systems Handbook*. Springer, Secaucus, NJ, USA, October 2010.

[49] Bracha Shapira, Lior Rokach, and Shirley Freilikhman. Facebook single and cross domain data for recommendation systems. *User Modeling and User-Adapted Interaction*, 23(2-3):211–247, April 2013.

[50] Yue Shi, Martha Larson, and Alan Hanjalic. Mining contextual movie similarity with matrix factorization for context-aware recommendation. *ACM Transactions on Intelligent Systems and Technology (ACM TIST)*, 4(1):16:1–16:19, February 2013.

[51] Yue Shi, Martha Larson, and Alan Hanjalic. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys*, 47(1):3:1–3:45, May 2014.

[52] Ajit P. Singh and Geoffrey J. Gordon. Relational learning via collective matrix factorization. In *Proceeding of the 14th ACM SIGKDD International*

Conference on Knowledge Discovery and Data Mining, pages 650–658, August 2008.

[53] Jiliang Tang, Xia Hu, Huiji Gao, and Huan Liu. Exploiting local and global social context for recommendation. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, pages 2712–2718, August 2013.

[54] Xiwang Yang, Yang Guo, Yong Liu, and Harald Steck. A survey of collaborative filtering based social recommender systems. *Computer Communications*, 41:1–10, 2014.

[55] Jiho Yoo and Seungjin Choi. Weighted nonnegative matrix co-tri-factorization for collaborative prediction. In *Proceedings of the 1st Asian Conference on Machine Learning: Advances in Machine Learning*, pages 396–411, November 2009.

[56] Liyan Zhang, Jie Xu, and Chunping Li. A random-walk based recommendation algorithm considering item categories. *Neurocomputing*, 120:391–396, 2013.

[57] Ya Zhang, Weiyuan Chen, and Zibin Yin. Collaborative filtering with social regularization for tv program recommendation. *Knowledge-Based Systems*, 54:310–317, December 2013.

[58] Yu Zhang, Bin Cao, and Dit-Yan Yeung. Multi-domain collaborative filtering. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, pages 725–732, July 2010.

[59] Yi Zhen, Wu-Jun Li, and Dit-Yan Yeung. Tagicofi: tag informed collaborative filtering. In *Proceedings of the 3rd ACM Conference on Recommender Systems*, pages 69–76, October 2009.

[60] Yu Zheng. Methodologies for cross-domain data fusion: An overview. *IEEE Transactions on Big Data*, 1(1):16–33, 2016.

[61] Fuzhen Zhuang, Ping Luo, Hui Xiong, Qing He, Yuhong Xiong, and Zhongzhi Shi. Exploiting associations between word clusters and document classes for cross-domain text categorization. *Statistical Analysis and Data Mining*, 4(1):100–114, February 2011.

Weike Pan received the Ph.D. degree in Computer Science and Engineering from the Hong Kong University of Science and Technology, Kowloon, Hong Kong, China, in 2012. He is currently a lecturer (research oriented) with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. His research interests include transfer learning, recommender systems, and statistical machine learning.