

Active Transfer Learning for Cross-System Recommendation

Lili Zhao[†], Sinno Jialin Pan[‡], Evan Wei Xiang^{*},
Erheng Zhong[†], Zhongqi Lu[†], Qiang Yang^{§†}

[†]Hong Kong University of Science and Technology, Hong Kong

[‡]Institute for Infocomm Research, Singapore

^{*}Baidu Inc., China

[§]Huawei Noah's Ark Lab, Science and Technology Park, Shatin, Hong Kong

^{†§}{lzhaoae, ezhong, zluab, qyang}@cse.ust.hk, [‡]jspace@i2r.a-star.edu.sg, ^{*}evan.xiang@gmail.com

Abstract

Recommender systems, especially the newly launched ones, have to deal with the data-sparsity issue, where little existing rating information is available. Recently, transfer learning has been proposed to address this problem by leveraging the knowledge from related recommender systems where rich collaborative data are available. However, most previous transfer learning models assume that entity-correspondences across different systems are given as input, which means that for any entity (e.g., a user or an item) in a target system, its corresponding entity in a source system is known. This assumption can hardly be satisfied in real-world scenarios where entity-correspondences across systems are usually unknown, and the cost of identifying them can be expensive. For example, it is extremely difficult to identify whether a user *A* from Facebook and a user *B* from Twitter are the same person. In this paper, we propose a framework to construct entity correspondence with limited budget by using active learning to facilitate knowledge transfer across recommender systems. Specifically, for the purpose of maximizing knowledge transfer, we first iteratively select entities in the target system based on our proposed criterion to query their correspondences in the source system. We then plug the actively constructed entity-correspondence mapping into a general transferred collaborative-filtering model to improve recommendation quality. We perform extensive experiments on real world datasets to verify the effectiveness of our proposed framework for this cross-system recommendation problem.

Introduction

Collaborative filtering (CF) technologies, especially matrix factorization methods, have achieved significant success in the field of recommender systems. CF aims to generate recommendations for a user by utilizing the observed preferences of other users whose historical behaviors are correlated with that of the target user. However, CF performs poorly when little collaborative information is available. This is referred to as the data sparsity problem, which is a common challenging problem in many newly launched recommender systems.

Recently, transfer learning (Pan and Yang 2010) has been proposed to address the data-sparsity problem in the target recommender system by using the data from some related recommender systems. A common motivation behind transfer learning is that many commercial Web sites often attract similar users (e.g., Twitter, Facebook, etc.), or provide similar product items (e.g., Amazon, eBay, etc.), thus, a source CF model built with rich collaborative data can be compressed as a prior to assist the training of a more precise CF model for the target recommender systems (Li, Yang, and Xue 2009a). This approach is also known as cross-system collaborative filtering.

Previous transfer-learning approaches to cross-system CF can be classified into two categories: (1) CF methods *with* cross-system entity correspondence, and (2) those *without* cross-system entity-correspondence. In the former category, Mehta and Hofmann (2006) and Pan *et al.* (2010) proposed to embed the cross-system entity-correspondences as constraints to jointly learn the CF models for the source and target recommender systems with an aim to improve the performance of the target CF system. Although these approaches have shown promising results, they require the existence of entity correspondence mappings, such as user correspondence or item correspondence, across different systems. This strong prerequisite is often difficult to satisfy in most real-world scenarios, as some specific users or items in one system may be missing in other systems. For example, user populations of Twitter and Facebook services are sometimes overlapping, but they are not identical, as is the case with Amazon and eBay. In addition, even though there may exist potential entity correspondences across different systems, they may be expensive or time-consuming to be recognized as users may use different names, or an item may be named differently in different online commercial systems.

In the second category where no assumption is made on pre-existing cross-system mappings, researchers have focused on capturing the group-level behaviors of users. For example, Li *et al.* (2009a) proposed a codebook-based-transfer (CBT) method for cross-domain CF, where entity-correspondences across systems are not required. The main assumption of CBT is that specific users or items may be different across systems, but the groups of them should behave similarly. Therefore, CBT aims to first generate a set of cluster-level user-item rating patterns from the source do-

main, which is referred to as a codebook. The codebook can be used as a prior for learning the CF model in the target system. Li *et al.* (2009b) further proposed a probabilistic model for cross-domain CF which shares a similar motivation with CBT. However, compared to the approaches in the former category, which make use of cross-system entity-correspondences as a bridge, these approaches are less effective for knowledge transfer across recommender systems.

In this paper, we assume that the cross-system entity-correspondences are unknown in general, but that these mappings can be identified with a cost. In particular, we propose a unified framework to actively construct entity-correspondence mappings across recommender systems, and then integrate them into a transfer learning approach with partial entity-correspondence mappings for the cross-system CF. The proposed framework consists of two major components:

- an active learning algorithm to construct entity-correspondences across systems with a fixed budget, and
- an extended transfer-learning based CF approach with partial entity-correspondence mappings for cross-domain CF.

Notations & Preliminaries

Denote by \mathcal{D} a target CF task, which is associated with an extremely sparse preference matrix $\mathbf{X}^{(d)} \in \mathbb{R}^{m_d \times n_d}$, where m_d is the number of users and n_d is the number of items. Each entry $x_{uv}^{(d)}$ of $\mathbf{X}^{(d)}$ corresponds to user u 's preference on item v . If $x_{uv}^{(d)} \neq 0$, it means for user u , the preference on item v is observed, otherwise unobserved. Let \mathcal{I}^d be the set of all observed (u, v) pairs of $\mathbf{X}^{(d)}$. The goal is to predict users' unobserved preferences based on a few observed preferences. For rating recommender systems, preferences are represented by numerical values (e.g., $[1, 2, \dots, 5]$, one star through five stars). In cross-system CF, besides \mathcal{D} , suppose we have a source CF task \mathcal{S} which is associated with a relatively dense preference matrix $\mathbf{X}^{(s)} \in \mathbb{R}^{m_s \times n_s}$, where m_s is the number of users and n_s is the number of items. Similarly, let \mathcal{I}^s be the set of all observed (u, v) pairs of $\mathbf{X}^{(s)}$. Furthermore, we assume that the cross-system entity-correspondences are unknown, but can be identified with cost. Our goal is to 1) actively construct entity-correspondences across the source and target systems with budget, and 2) make use of them for knowledge transfer from the source task \mathcal{S} to the target task \mathcal{D} . In the sequel, we denote by $\mathbf{X}_{*,i}$ the i^{th} column of the matrix \mathbf{X} , and superscript \top the transpose of vector or matrix, and use the words "domain" and "system" interchangeably.

Maximum-Margin Matrix Factorization

Our active transfer learning framework for CF is based on Maximum-Margin Matrix Factorization (MMMF) (Srebro, Rennie, and Jaakkola 2005), which aims to learn a fully observed matrix $\mathbf{Y} \in \mathbb{R}^{m \times n}$ to approximate a target preference matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ by maximizing the predictive margin and minimizing the trace norm of \mathbf{Y} . Specifically, the

objective of MMMF for binary preference predictions is to minimize

$$\mathcal{J} = \sum_{(u,v) \in \mathcal{I}} h(y_{uv} \cdot x_{uv}) + \lambda \|\mathbf{Y}\|_{\Sigma}, \quad (1)$$

where \mathcal{I} is the set of observed (u, v) pairs of \mathbf{X} , $h(z) = (1 - z)_+ = \max(0, 1 - z)$ is the Hinge loss, $\|\cdot\|_{\Sigma}$ denotes the trace norm, and $\lambda \geq 0$ is a trade-off parameter. In binary preference predictions, $y_{uv} = +1$ denotes that user u likes the item v , while $y_{uv} = -1$ denotes dislike. The objective (1) can be extended to ordinal rating predictions, and solved efficiently (Rennie and Srebro 2005). Suppose $x_{uv} \in \{1, 2, \dots, R\}$, one can use $R - 1$ thresholds $\theta_1, \theta_2, \dots, \theta_{R-1}$ to relate the real-valued y_{uv} to the discrete-valued x_{uv} by requiring

$$\theta_{x_{uv}-1} + 1 \leq y_{uv} \leq \theta_{x_{uv}} - 1,$$

where $\theta_0 = -\infty$ and $\theta_R = \infty$. Furthermore, suppose \mathbf{Y} can be decomposed as $\mathbf{Y} = \mathbf{U}^{\top} \mathbf{V}$, where $\mathbf{U} \in \mathbb{R}^{k \times m}$ and $\mathbf{V} \in \mathbb{R}^{k \times n}$. The objective function of MMMF for ordinal rating predictions can be written as follows,

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}, \Theta} \mathcal{J} = & \sum_{(u,v) \in \mathcal{I}} \sum_{r=1}^{R-1} h(T_{uv}^r (\theta_{ur} - \mathbf{U}_{*u}^{\top} \mathbf{V}_{*v})) \\ & + \lambda (\|\mathbf{U}\|_F + \|\mathbf{V}\|_F), \end{aligned} \quad (2)$$

where $T_{uv}^r = +1$ for $r \geq x_{uv}$, while $T_{uv}^r = -1$ for $r < x_{uv}$, and $\|\cdot\|_F$ denotes the Frobenius norm. The thresholds $\Theta = \{\theta_{ur}\}$'s can be learned together with \mathbf{U}^d and \mathbf{V}^d from the data. Note that the thresholds $\{\theta_{ur}\}$'s here are user-specific. Alternative gradient descent methods can be applied to solve the optimization problem (2).

Active Transfer Learning for Cross-System CF Overall Framework

In this section, we introduce the overall framework on active transfer learning for cross-system CF as described in Algorithm 1. To begin with, we apply MMMF on the target collaborative data to learn a CF model. After that, we iteratively select K entities based on our proposed entity selection strategies to query their correspondences in the source system. We then apply the extended MMMF method in the transfer learning manner on the source and target collaborative data to learn an updated CF model.

In the following sections, we describe the entity selection strategies and the extended MMMF for transfer learning in detail, respectively.

MMMF with Partial Entity Correspondence

Denote by $\mathbf{U}_{\mathcal{C}}^{(s)}$ and $\mathbf{V}_{\mathcal{C}}^{(s)}$ the factor sub-matrices of $\mathbf{U}^{(s)}$ and $\mathbf{V}^{(s)}$ for the entities whose indices are in \mathcal{C} , respectively. Similarly, denote by $\mathbf{U}_{\mathcal{C}}^{(d)}$ and $\mathbf{V}_{\mathcal{C}}^{(d)}$ the factor sub-matrices for the entities whose indices are in \mathcal{C} respectively. Here \mathcal{C} denotes the indices of the corresponding entities (can be either users or items) between the source and target systems.

Algorithm 1 Active Transfer Learning for Cross-System CF

Input: $\mathbf{U}^{(s)}, \mathbf{V}^{(s)}, \mathbf{X}^{(d)}, T$, and K .

Output: $\mathbf{U}^{(d)}$, and $\mathbf{V}^{(d)}$.

Initialize:

Apply (2) on $\mathbf{X}^{(d)}$ to generate $\Theta_0^{(d)}, \mathbf{U}_0^{(d)}$ and $\mathbf{V}_0^{(d)}$.

for $t = 0$ to $T - 1$ **do**

Step 1: Set $\mathcal{C}^{(s)} = \text{ActiveLearn}(\Theta_t^{(d)}, \mathbf{U}_t^{(d)}, \mathbf{V}_t^{(d)}, K)$, where $\mathcal{C}^{(s)}$ is the set of the indices of the selected entities (either users or items), and $|\mathcal{C}^{(s)}| = K$.

Step 2: Query $\mathcal{C}^{(s)}$ in the source system to identify their corresponding indices $\mathcal{C}^{(d)}$.

Step 3: Apply $\text{MMMF}_{TL}(\mathbf{U}^{(s)}, \mathbf{V}^{(s)}, \mathbf{X}^{(d)}, \mathcal{C}^{(s)}, \mathcal{C}^{(d)})$ to update $\Theta_{t+1}^{(d)}, \mathbf{U}_{t+1}^{(d)}$, and $\mathbf{V}_{t+1}^{(d)}$.

end for

Return: $\mathbf{U}^{(d)} \leftarrow \mathbf{U}_T^{(d)}, \mathbf{V}^{(d)} \leftarrow \mathbf{V}_T^{(d)}$.

The proposed approach with partial entity-correspondences to cross-system CF can be written as,

$$\min_{\mathbf{U}, \mathbf{V}, \Theta} \mathcal{J} = \sum_{(u,v) \in \mathcal{I}} \sum_{r=1}^{R-1} h(T_{uv}^r (\theta_{ur} - \mathbf{U}_{*u}^T \mathbf{V}_{*v})) + \lambda \|\mathbf{U}\|_F + \lambda \|\mathbf{V}\|_F + \lambda_{\mathcal{C}} \mathcal{R}(\mathbf{U}_{\mathcal{C}}^{(d)}, \mathbf{V}_{\mathcal{C}}^{(d)}, \mathbf{U}_{\mathcal{C}}^{(s)}, \mathbf{V}_{\mathcal{C}}^{(s)}), \quad (3)$$

where the last term is a regularization term that aims to use $\mathbf{U}_{\mathcal{C}}^{(s)}$ and $\mathbf{V}_{\mathcal{C}}^{(s)}$ as priors to learn more precise $\mathbf{U}_{\mathcal{C}}^{(d)}$ and $\mathbf{V}_{\mathcal{C}}^{(d)}$, which can be expanded to obtain more precise $\mathbf{U}^{(d)}$ and $\mathbf{V}^{(d)}$ respectively. The associated $\lambda_{\mathcal{C}} \geq 0$ is a trade-off parameter to control the impact of the regularization term.

Intuitively, a simple way to define the regularization term is to enforce the target factor sub-matrices $\mathbf{U}_{\mathcal{C}}^{(d)}$ and $\mathbf{V}_{\mathcal{C}}^{(d)}$ to be the same as the source factor sub-matrices $\mathbf{U}_{\mathcal{C}}^{(s)}$ and $\mathbf{V}_{\mathcal{C}}^{(s)}$ respectively as follows,

$$\mathcal{R}(\mathbf{U}_{\mathcal{C}}^{(d)}, \mathbf{V}_{\mathcal{C}}^{(d)}, \mathbf{U}_{\mathcal{C}}^{(s)}, \mathbf{V}_{\mathcal{C}}^{(s)}) = \|\mathbf{W}_{\mathcal{C}}^{(s)} - \mathbf{W}_{\mathcal{C}}^{(d)}\|_F, \quad (4)$$

where $\mathbf{W}_{\mathcal{C}}^{(d)} = [\mathbf{U}_{\mathcal{C}}^{(d)} \mathbf{V}_{\mathcal{C}}^{(d)}]$ and $\mathbf{W}_{\mathcal{C}}^{(s)} = [\mathbf{U}_{\mathcal{C}}^{(s)} \mathbf{V}_{\mathcal{C}}^{(s)}]$. This ‘‘identical’’ assumption is similar to that of Collective Matrix Factorization (CMF) (Singh and Gordon 2008), and may not hold in practice. In the sequel, as a baseline, we denote by MMMF_{CMF} the extended MMMF method by plugging (4) into (3).

Alternatively, we propose to use the similarities between entities estimated in the source system as priors to constrain the similarities between entities in the target system. The motivation is that if two entities in the source system are similar to each other, then their correspondences tend to be similar to each other in the target system as well. Therefore, we propose the following form of the regularization term,

$$\mathcal{R}(\mathbf{U}_{\mathcal{C}}^{(d)}, \mathbf{V}_{\mathcal{C}}^{(d)}, \mathbf{U}_{\mathcal{C}}^{(s)}, \mathbf{V}_{\mathcal{C}}^{(s)}) = \text{tr}(\mathbf{W}_{\mathcal{C}}^{(d)} \mathbf{L}_{\mathcal{C}}^{(s)} \mathbf{W}_{\mathcal{C}}^{(d)T}), \quad (5)$$

where $\text{tr}(\cdot)$ denotes the trace of a matrix, $\mathbf{L}_{\mathcal{C}}^{(s)} = \begin{bmatrix} \mathbf{L}_U^{(s)} & 0 \\ 0 & \mathbf{L}_V^{(s)} \end{bmatrix}$, and $\mathbf{L}_U^{(s)} = \mathbf{D}_U^{(s)} - \mathbf{A}_U^{(s)}$ is known as the

Laplacian matrix, where $\mathbf{A}_U^{(s)} = \mathbf{U}_C^{(s)\top} \mathbf{U}_C^{(s)}$ is the similarity matrix of the users in the source system, whose indices are in the set \mathcal{C} , and $\mathbf{D}_U^{(s)}$ is a diagonal matrix with diagonal entries $\mathbf{D}_{U_{ii}}^{(s)} = \sum_j \mathbf{A}_{U_{ij}}^{(s)}$. The definition of $\mathbf{L}_V^{(s)}$ on items is similar. Note that a similar regularization term has been proposed by (Li and Yeung 2009). However, their work is focused on utilizing relational information for single-domain CF, and the Laplacian Matrix is constructed by links between entities instead of entity-similarities in a source domain. In the sequel, we denote by MMMF_{TL} the proposed MMMF extension by plugging (5) into (3).

Actively Constructing Entity Correspondences

In this section, we describe a margin-based method for actively constructing entity correspondences. A common motivation behind margin-based active learning approaches is that given a margin-based model, the margin of an example denotes certainty to the prediction on the example. The smaller the margin is for an example, the lower the certainty is for its prediction.

Margins on User-Item Pairs Suppose that MMMF (2) or MMMF_{TL} (3) is performed on the collaborative data in the target system, then given a user u , for each threshold θ_k , where $k \in \{1, \dots, R-1\}$, the margin of a user-item pair (u, v) can be defined as

$$\begin{cases} \delta_k^{(d)}(u, v) = \mathbf{U}_{*,u}^{(d)T} \mathbf{V}_{*,v}^{(d)} - \theta_k, & \text{if } y_{u,v}^{(d)} > \theta_k, \\ \delta_k^{(d)}(u, v) = \theta_k - \mathbf{U}_{*,u}^{(d)T} \mathbf{V}_{*,v}^{(d)}, & \text{if } y_{u,v}^{(d)} \leq \theta_k, \end{cases} \quad (6)$$

where $y_{u,v}^{(d)} = x_{u,v}^{(d)}$ if $x_{u,v}^{(d)}$ is observed; otherwise, $y_{u,v}^{(d)} = \mathbf{U}_{*,u}^{(d)T} \mathbf{V}_{*,v}^{(d)}$. Based on the above definition, for each user-item pair (u, v) , we have $R-1$ margins. Among them, the margins to the left (lower) and right (upper) boundaries of the correct interval have the highest importance, which we denote by $\delta_L^{(d)}(u, v)$ and $\delta_R^{(d)}(u, v)$, respectively. Similar to other margin-based active learning methods, we assume that, for the unobserved user-item pairs $\{(u, v)\}$ ’s, the predictions of the CF model are correct, and thus we can obtain the ‘‘correct’’ intervals of the unobserved user-item pairs as well. Intuitively, for a pair (u, v) , when $\delta_L^{(d)}(u, v) = \delta_R^{(d)}(u, v)$, the confidence of the prediction is the highest. We define a normalized margin of a user-item pair (u, v) as follows,

$$\tilde{\delta}^{(d)}(u, v) = 1 - \frac{|\delta_L^{(d)}(u, v) - \delta_R^{(d)}(u, v)|}{\delta_L^{(d)}(u, v) + \delta_R^{(d)}(u, v)}. \quad (7)$$

Note that $\tilde{\delta}^{(d)}(u, v) \in [0, 1]$.

Margins on Entities With the margin definition of a user-item pair in (7), we are now ready to define the margin of an entity (either a user or an item). For simplicity, in the rest of the section, we only describe the definition on the margin of a user, and propose the user selection strategies based on the definition. The margin of an item can be defined similarly, and item selection strategies can be designed accordingly as well. By observing that given a preference matrix $\mathbf{X}^{(d)}$

with m_d users and n_d items, a user u can be represented by the pairs between the user and each item (i.e., n_d pairs in total). It is reasonable to assume that the margin of the user u can be decomposed to the margins of the user-item pairs $\{(u, v_i)\}$'s. Furthermore, for each user u , ratings on some items are observed, whose item indices are denoted by \mathcal{I}_u^d , while the others are unobserved, whose item indices are denoted by $\widehat{\mathcal{I}}_u^d$. Therefore we propose the margin of a user as follows,

$$\delta_u^{(d)} = \eta \frac{1}{|\mathcal{I}_u^d|} \sum_{v \in \mathcal{I}_u^d} \tilde{\delta}^{(d)}(u, v) + (1-\eta) \frac{1}{|\widehat{\mathcal{I}}_u^d|} \sum_{v \in \widehat{\mathcal{I}}_u^d} \tilde{\delta}^{(d)}(u, v), \quad (8)$$

where the first term can be referred to as the average of the ‘‘true’’ margins of the user-item pairs with observed ratings, and the second term can be referred to as the average of the ‘‘predictive’’ margins of the user-item pairs with unobserved ratings. The tradeoff parameter $\eta \in [0, 1]$ is to balance the impact of the two terms to the overall margin of the user. In this paper, we simply set $\eta = 0.5$.

Based on the margin of a user as defined in (8), we propose two user-selection strategies as follows.

- **MG_{min}**: in each iteration, we rank users in ascending order in terms of their corresponding margin $\{\delta_u^{(d)}\}$'s, and select the top K users to construct \mathcal{C} for query.

This strategy can return the most uncertain users in the current CF model. However, due to the *long-tail* problem in CF (Park and Tuzhilin 2008), many items or users in the long tail have only few ratings. Thus, the most uncertain users in the target recommender system tend to be in the long tail with high probabilities. Furthermore, since we assume the source and target recommender systems be similar, if the users are in the long tail in the target system, then their counterparts tend to be long-tail users in the source system as well. This implies that the factor sub-matrices $\mathbf{U}_C^{(s)}$ and $\mathbf{V}_C^{(s)}$ to be transferred from the source system may not be precise, resulting in limited knowledge transfer through (5). Thus, we propose another user-selection strategy as follows.

- **MG_{hybrid}**: in each iteration, we first apply **MG_{min}** to select K_1 users, denoted by \mathcal{C}_1 , where $K_1 < K$. After that for the rest users $\{u_i\}$'s, we apply the scoring function defined in (9) to rank them in descending order, and select $K - K_1$ users to construct \mathcal{C}_2 . Finally, we set $\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2$.

$$\Delta^{(d)}(u_i, \mathcal{C}_u) = \frac{\sum_{u_j \in \mathcal{C}_u} \text{sim}(u_i, u_j) \delta_{u_i}^{(d)}}{\sum_{u_j \in \mathcal{C}_u} \text{sim}(u_i, u_j)}, \quad (9)$$

where $\text{sim}(u_i, u_j) = \frac{|\mathcal{I}_{u_i}^d \cap \mathcal{I}_{u_j}^d|}{\max(|\mathcal{I}_{u_i}^d|, |\mathcal{I}_{u_j}^d|)}$ is the measure of correlation between the users u_i and u_j based on their rating behaviors. The motivation behind the scoring function (9) is that we aim to select users who are 1) informative (with large values of $\{\delta_{u_i}^{(d)}\}$'s) and thus supposed to be ‘‘active’’ instead of the long tail; 2) of strong correlation to the pre-selected most uncertain users in \mathcal{C}_1 (with large values of $\sum_{u_j \in \mathcal{C}_u} \text{sim}(u_i, u_j)$) and thus supposed to be helpful to recommend items for them based on the intrinsic assumption in CF.

Experiments

Datasets and Experimental Setting

We evaluate our proposed framework on two datasets: Netflix¹ and Douban². The Netflix dataset contains more than 100 million ratings given by more than 480,000 users on around 18,000 movies with ratings in $\{1, 2, 3, 4, 5\}$. Douban is a popular recommendation website in China. It contains three types of items including movies, books and music with rating scale $[1, 5]$.

For Netflix dataset, we filter out movies with less than 5 ratings for our experiments. The dataset is partitioned into two parts with disjoint sets of users, while sharing the whole set of movies. One part consists of ratings given by 50% users with 1.2% rating density, which serves as the source domain. The remaining users are considered as the target domain with 0.7% rating density. For Douban, we collect a dataset consisting of 12,000 users and 100,000 items with only movies and books. Users with less than 10 ratings are discarded. There remain 270,000 ratings on 3,500 books, and 1,400,000 ratings on 8,000 movies, given by 11,000 users. The density of the ratings on books and movies are 0.6% and 1.5% respectively. We consider movie ratings as the source domain and book ratings as the target domain. In this task, all users are shared but items are disjoint. Furthermore, since there are about 6,000 movies shared by Netflix and Douban, we extract ratings on the shared movies from Netflix and Douban respectively, and obtain 490,000 ratings given by 120,000 users from Douban with rating density 0.7%, and 1,600,000 ratings given by 10,000 users from Netflix with density 2.6%. We consider ratings on Netflix as the source domain and those on Douban as the target domain. In total, we construct three cross-system CF tasks, and denote by **Netflix**→**Netflix**, **DoubanMovie**→**DoubanBook** and **Netflix**→**DoubanMovie**, respectively.

In the experiments, we split each target domain data into a training set of 80% preference entries and a test set of 20% preference entries, and report the average results of 10 random times. The parameters of the model, i.e., the number of latent factors k and the number of iterations T are tuned on some hand-out data of **Netflix**→**Netflix**, and fixed to all experiments.³ Here, $T = 10$, and $k = 20$. For evaluation criterion, we use Root Mean Square Error (RMSE) defined as,

$$\text{RMSE} = \sqrt{\frac{\sum_{(u,v) \in \mathcal{I}} (x_{uv} - \hat{x}_{uv})^2}{|\mathcal{I}|}},$$

where x_{uv} and \hat{x}_{uv} are the true and predicted ratings respectively, and $|\mathcal{I}|$ is the number of test ratings. The smaller is the value, the better is the performance.

Overall Comparison Results

In the first experiment, we qualitatively show the effectiveness of our proposed active transfer learning framework for

¹<http://www.netflix.com>

²<http://www.douban.com>

³Suppose total budget is ρ which is the total number of correspondences to be constructed, we set the number of correspondences actively constructed in each iteration as $K = \rho/T$.

Table 1: Overall comparison results on the three datasets in terms of RMSE.

Tasks	Methods						
	NoTransf (w/o corr.)		NoTransf (0.1% corr.)		CBT (w/o corr.)	MMMF _{TL}	
	MF	MMMF	MF	MMMF		(0.1% corr.)	(100% corr.)
Netflix→Netflix	0.8900 (± 0.0004)	0.880 (± 0.0001)	0.9112 (± 0.0002)	0.9103 (± 0.0004)	0.8846 (± 0.0002)	0.8692 (± 0.0003)	0.8527 (± 0.0002)
Movie→Book (Douban)	0.8804 (± 0.0017)	0.8784 (± 0.0002)	0.8876 (± 0.0003)	0.8837 (± 0.0001)	0.8656 (± 0.0002)	0.8292 (± 0.0003)	0.8126 (± 0.0002)
Netflix→DoubanMovie	0.8520 (± 0.0003)	0.8578 (± 0.0002)	0.8643 (± 0.0001)	0.8589 (± 0.0002)	0.8246 (± 0.0002)	0.7740 (± 0.0001)	0.7576 (± 0.0001)

cross-domain CF as compared with the following baselines:

- NoTransf without correspondences: to apply state-of-the-art CF models on the target domain collaborative data directly without either active learning or transfer learning. In this paper, for state-of-the-art CF models, we use low-rank Matrix Factorization (MF) (Koren, Bell, and Volinsky 2009) and MMMF.
- NoTransf with actively constructed correspondences: to first apply active learning strategy to construct cross-domain entity-correspondences, and then align the source and target domain data to generate a unified item-user matrix. Finally, we apply state-of-the-art CF models on the unified matrix for recommendations.
- CBT: to apply the codebook-based-transfer (CBT) method on the source and target domain data for recommendations. As introduced in the first section, CBT does not require any entity-correspondences to be constructed.
- MMMF_{TL} with full correspondences: to apply the proposed MMMF_{TL} on the source and target domain data with full entity-correspondences for recommendations. Note that this method which assumes all entity-correspondences be available can be considered as an upper bound of the active transfer learning method.

The overall comparison results on the three cross-domain tasks are shown in Table 1. For the active learning strategies, we use MG_{hybrid} as proposed in (9). As can be observed from the first group of columns in the table, applying state-of-the-art CF models on the extremely sparse target domain data directly is not able to obtain precise recommendation results in terms of RMSE. The results of the second group of columns in the table suggest that aligning all the source and target data to a unified item-user matrix and then performing state-of-the-art CF models on it cannot help to boost the recommendation performance, but may even hurt the performance compared to that of applying CF models on the target domain data only. This is because the alignment makes the matrix to be factorized larger but still very sparse, resulting in a more difficult learning task. From the table we can also observe that the transfer learning method CBT performs better than the NoTransf methods. However, our proposed active transfer learning method MMMF_{TL} with only 0.1% entity-correspondences achieves much better performance than CBT in terms of RMSE. This verifies the conclusion that making use of cross-system entity-correspondences as a bridge is useful for knowledge transfer across recommender systems. Finally, by considering the

performance of MMMF_{TL} with full entity-correspondences as the knowledge-transfer upper bound, and the performance of MMMF as the baseline, our proposed active transfer learning method can achieve around 70% knowledge transfer ratio on average on the three tasks while only requires 0.1% entity-correspondences to be labeled.

Experiments on Diff. Active Learning Strategies

In the second experiment, we aim to verify the performance of our proposed active transfer learning framework plugging with different entity selection strategies. Here, we use MMMF_{TL} as the base transfer learning approach to cross-domain CF. For entity selection strategies, besides the two strategies, MG_{min} and MG_{hybrid}, presented in the model introduction section, we also conduct comparison experiments on the following strategies.

- Rand: to select entities randomly in the target domain to query their correspondences in the source domain.
- Many: to select the entities with most historical ratings in the target domain to query their correspondences in the source domain.
- Few: to select the entities with fewest historical ratings in the target domain to query their correspondences in the source domain.
- MG_{max}: to select the entities with largest $\{\delta_u^{(d)}\}$'s as defined in (8) in the target domain to query their correspondences in the source domain.

Figure 1 shows the results of MMMF_{TL} with different entity selection strategies under varying proportions of entity-correspondences to be labeled. From the figure, we can observe that the margin-based approaches (i.e., MG_{min}, MG_{max}, and MG_{hybrid}) perform much better than other approaches based on different criteria. In addition, compared with MG_{min} and MG_{max}, our proposed MG_{hybrid} not only selects uncertain entities but also selects informative entities which have strong correlations to the most uncertain entities, thus performs slightly better.

Experiments on Diff. Cross-domain Regularizers

As mentioned in the model introduction section, the regularization term $\mathcal{R}(\mathbf{U}_c^{(d)}, \mathbf{V}_c^{(d)}, \mathbf{U}_c^{(s)}, \mathbf{V}_c^{(s)})$ in (3) for cross-system knowledge transfer can be substituted by different forms, e.g., (4) or (5), resulting in different transfer learning approaches, MMMF_{CMF} or MMMF_{TL} accordingly. Therefore, in the third experiment, we use MG_{hybrid}

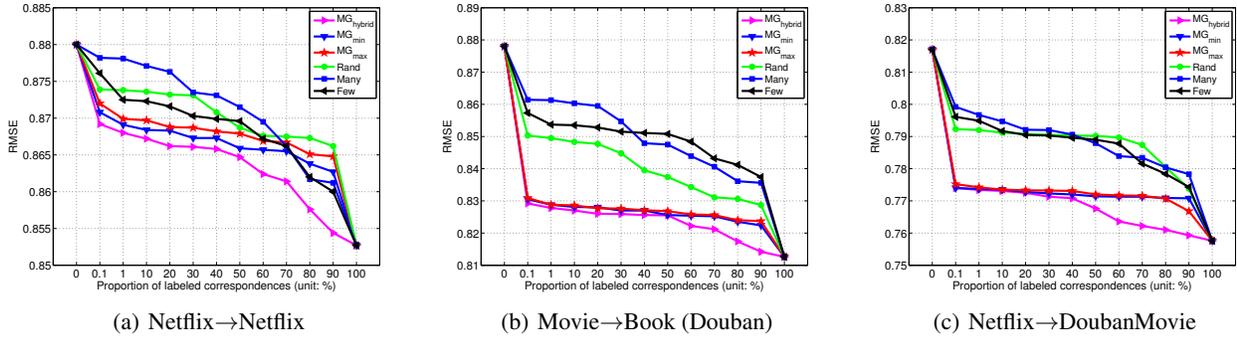


Figure 1: Results on different entity selection strategies under varying proportions of entity-responses to be labeled.

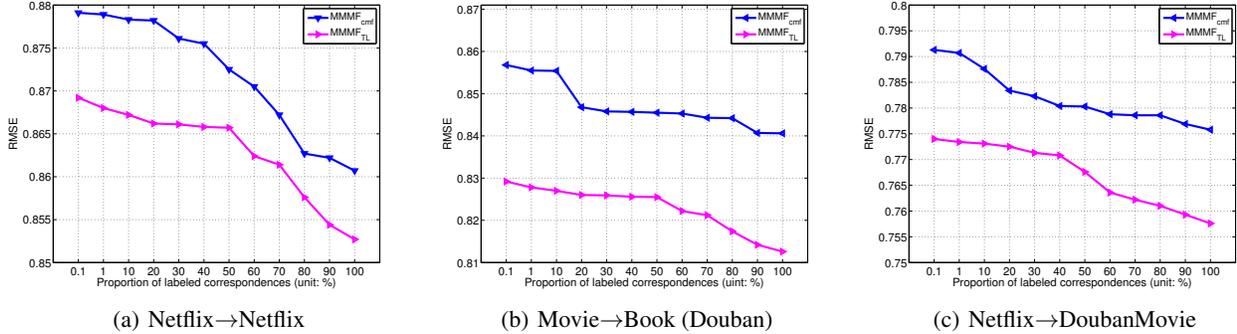


Figure 2: Results on different cross-domain regularizers under varying proportions of entity-responses to be labeled.

as the entity selection strategy, and compare the performance of $MMMF_{CMF}$ and $MMMF_{TL}$ in terms of RMSE. As can be seen from Figure 2, the proposed $MMMF_{TL}$ outperforms $MMMF_{CMF}$ consistently on the three cross-system tasks under varying proportions of the labeled entity-correspondences. This implies that using similarities between entities from the source domain data as priors is more safe and useful for knowledge transfer across recommender systems than using the factor matrices factorized from the source domain data directly.

Related Work

Besides the works introduced in the first section, there are several other related works on applying transfer learning for CF. Pan *et al.* (2012) developed an approach known as TIF (Transfer by Integrative Factorization) to integrate the auxiliary uncertain ratings as constraints into the target matrix factorization problem. Cao *et al.* (2010) and Zhang *et al.* (2010) extended the CMF method to solve multi-domain CF problems in a multi-task learning manner respectively.

Our work is also related to previous works on active learning to CF (Shi, Zhao, and Tang 2012; Mello, Aufaure, and Zimbrão 2010; Rish and Tesauro 2008; Jin and Si 2004; Boutilier, Zemel, and Marlin 2003), which assumed that the users are able to provide ratings to every item of the system. However, this assumption may not hold in many real-world scenarios because users may not be familiar with all items of the system, thus they may fail to provide ratings on them. Alternatively, we propose to actively construct entity-correspondence mappings across systems.

Another related research topic is developing a unified framework for active learning and transfer learning. Most previous works on this topic are focused on standard classification tasks (Saha *et al.* 2011; Rai *et al.* 2010; Shi, Fan, and Ren 2008; Chan and Ng 2007). In this paper, our study on active transfer learning is focused on addressing the data-sparsity problem in CF, which is different from the previous tasks on classification or regression. The existing frameworks on combining active learning and transfer learning cannot be directly applied to our problem.

Conclusions and Future Work

In this paper, we present a novel framework on active transfer learning for cross-system recommendations. In the proposed framework, we 1) extend previous transfer learning approaches to CF in a partial entity-corresponding manner, and 2) propose several entity selection strategies to actively construct entity-correspondences across different recommender systems. Our experimental results show that compared with the transfer learning method which requires full entity-correspondences, our proposed framework can achieve 70% knowledge-transfer ratio, while only requires 0.1% of the entities to have correspondence. For future work, we are planning to apply the proposed framework to other applications, such as cross-system link prediction in social networks.

Acknowledgement

We thank the support of Hong Kong RGC grants 621812 and 621211.

References

- Boutillier, C.; Zemel, R. S.; and Marlin, B. 2003. Active collaborative filtering. In *UAI*, 98–106.
- Cao, B.; Liu, N. N.; and Yang, Q. 2010. Transfer learning for collective link prediction in multiple heterogeneous domains. In *ICML*, 159–166.
- Chan, Y. S., and Ng, H. T. 2007. Domain adaptation with active learning for word sense disambiguation. In *ACL*.
- Jin, R., and Si, L. 2004. A bayesian approach toward active learning for collaborative filtering. In *UAI*, 278–285.
- Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer* 42(8):30–37.
- Li, W.-J., and Yeung, D.-Y. 2009. Relation regularized matrix factorization. In *IJCAI*, 1126–1131.
- Li, B.; Yang, Q.; and Xue, X. 2009a. Can movies and books collaborate?: cross-domain collaborative filtering for sparsity reduction. In *IJCAI*, 2052–2057.
- Li, B.; Yang, Q.; and Xue, X. 2009b. Transfer learning for collaborative filtering via a rating-matrix generative model. In *ICML*, 617–624.
- Mehta, B., and Hofmann, T. 2006. Cross system personalization and collaborative filtering by learning manifold alignments. In *KI*. 244–259.
- Mello, C. E.; Aufaure, M.-A.; and Zimbrão, G. 2010. Active learning driven by rating impact analysis. In *RecSys*, 341–344.
- Pan, S. J., and Yang, Q. 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22(10):1345–1359.
- Pan, W.; Xiang, E. W.; Liu, N. N.; and Yang, Q. 2010. Transfer learning in collaborative filtering for sparsity reduction. In *AAAI*.
- Pan, W.; Xiang, E.; and Yang, Q. 2012. Transfer learning in collaborative filtering with uncertain ratings. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.
- Park, Y.-J., and Tuzhilin, A. 2008. The long tail of recommender systems and how to leverage it. In *RecSys*, 11–18.
- Rai, P.; Saha, A.; Daumé, III, H.; and Venkatasubramanian, S. 2010. Domain adaptation meets active learning. In *NAACL HLT 2010 Workshop on Active Learning for Natural Language Processing*, 27–32.
- Rennie, J. D. M., and Srebro, N. 2005. Fast maximum margin matrix factorization for collaborative prediction. In *ICML*, 713–719.
- Rish, I., and Tesauro, G. 2008. Active collaborative prediction with maximum margin matrix factorization. In *ISAIM*.
- Saha, A.; Rai, P.; III, H. D.; Venkatasubramanian, S.; and DuVall, S. L. 2011. Active supervised domain adaptation. In *ECML/PKDD (3)*, 97–112.
- Shi, X.; Fan, W.; and Ren, J. 2008. Actively transfer domain knowledge. In *ECML/PKDD (2)*, 342–357.
- Shi, L.; Zhao, Y.; and Tang, J. 2012. Batch mode active learning for networked data. *ACM Transaction on Intelligent Systems and Technology* 3(2):33:1–33:25.
- Singh, A. P., and Gordon, G. J. 2008. Relational learning via collective matrix factorization. In *KDD*, 650–658.
- Srebro, N.; Rennie, J. D. M.; and Jaakkola, T. S. 2005. Maximum-margin matrix factorization. In *NIPS 17*, 1329–1336.
- Tang, J.; Yan, J.; Ji, L.; Zhang, M.; Guo, S.; Liu, N.; Wang, X.; and Chen, Z. 2011. Collaborative users’ brand preference mining across multiple domains from implicit feedbacks. In *AAAI*.
- Zhang, Y.; Cao, B.; and Yeung, D.-Y. 2010. Multi-domain collaborative filtering. In *UAI*, 725–732.