Cross-Domain Recommendation with Multiple Sources

Qian Zhang
Center for Artificial Intelligence
University of Technology Sydney
NSW, Australia
qian.zhang-1@uts.edu.au

Jie Lu
Center for Artificial Intelligence
University of Technology Sydney
NSW, Australia
jie.lu@uts.edu.au

Guangquan Zhang
Center for Artificial Intelligence
University of Technology Sydney
NSW, Australia
guangquan.zhang@uts.edu.au

***Abstract*—Data sparsity remains a challenging and common problem in real-world recommender systems, which impairs the accuracy of recommendation thus damages user experience. Cross-domain recommender systems are developed to deal with data sparsity problem through transferring knowledge from a source domain with relatively abundant data to the target domain with insufficient data. However, two challenging issues exist in cross-domain recommender systems: 1) domain shift which makes the knowledge from source domain inconsistent with that in the target domain; 2) knowledge extracted from only one source domain is insufficient, while knowledge is potentially available in many other source domains. To handle the above issues, we develop a cross-domain recommendation method in this paper to extract group-level knowledge from multiple source domains to improve recommendation in a sparse target domain. Domain adaptation techniques are applied to eliminate the domain shift and align user and item groups to maintain knowledge consistency during the transfer learning process. Knowledge is extracted not from one but multiple source domains through an intermediate subspace and adapted through flexible constraints of matrix factorization in the target domain. Experiments conducted on five datasets in three categories show that the proposed method outperforms six benchmarks and increases the accuracy of recommendations in the target domain.**

***Index Terms*—recommender system, cross-domain recommender system, knowledge transfer, collaborative filtering**

I. Introduction

Recommender systems have been in existence for more than twenty years and are prompted by both research and wide application in industry [1]. Recommender systems are now an indispensable part in personalization of products/services and are adopted by many websites, such as Amazon.com, Google, YouTube, Netflix, Yahoo and Facebook. Most of the recommender systems built are under the assumption of collaborative filtering, i.e. similar users share similar interests in the past will have similar interests in the future. However, the interactions between users and items are limited compared with the huge number of items existed. This is the data sparsity problem, a challenging and common issue in many existing recommender systems . Cross-domain recommender systems are developed to deal with the data sparsity problem. The abundance of data in another domain can assist the recommendation in a specific target domain with insufficient data. Further, user demands of diverse recommendation also prompted recommender systems to expand from single-domain to multi-domain [2].

Cross-domain recommender systems extract knowledge from domains that contain relatively rich data and adapt it to the target domain where data are insufficient. Two different types of cross-domain recommender systems have been developed [2]. Some methods connect multiple domains through auxiliary information such as user generated information [3], social information [4] or item attributes. On the other hand, some methods focus on preference data which are the most commonly collected data on e-commerce or online rating websites. Since the entity correspondence is not always fully available, some strategies are developed to match users or items in two domains [5]. In this paper, we focus on cross-domain recommender systems where users and items have no intersections, which is commonly existed between different websites or platforms. This is of great help for newly-launched recommender system, where data sparsity problem is particularly severe and challenging, to benefit from a mature recommender system.

Since no intersections exists in users/item between the source domain and the target domain, the knowledge shared is group-level rating pattern. How to extract the group-level knowledge and adapt it to the target domain is very crucial for the cross-domain recommendation. Two challenging issues exist in group-level knowledge extraction and adaptation: 1) Domain shift exists between the source and the target domain, so the group-level knowledge from source and target domains is not consistent. 2) The group-level knowledge extracted from one source domain is insufficient, while multiple sources potentially can assist recommendation in the target domain. The first issue is commonly recognized as domain shift in transfer learning [6] and raised as an issue in cross-domain recommender systems in [7]. To get rid of the inconsistent knowledge, inevitable loss of information aggravate the second issue. Most existing methods on cross-domain recommendation ignore the domain shifts and fail to extract consistent knowledge shared by two domains, let alone extracting consistent knowledge from multiple sources. Without handling the two issues, the effectiveness of knowledge transfer is not guaranteed, thus the accuracy of cross-domain recommendation is impaired.

In this paper, we develop a cross-domain recommendation method with multiple source (CDR-MS), which extracts group-level knowledge from multiple source domains to improve recommendation performance in a sparse target domain. We first cluster users and items into groups on multiple source domains and conduct group alignment with domain adaptation techniques, which ensures the knowledge consistency. Then, an intermediate subspace is learned through a Karcher mean to preserve knowledge from multiple source domains. We use a less restricted constrains on matrix factorization for the adaptation of the group-level knowledge from multiple sources to the target domain to increase the flexibility and ensure the consistency at the same time. The main contributions of this paper are as follows:

1) A domain adaptation method that aligns the subspaces of multiple source domains with the target domain. This method is able to conduct user and item latent group alignment in multiple domains to eliminate domain shift.
2) A matrix factorization method with constraints that are less restrictive in group-level knowledge adaptation. The constraints are more flexible than previous methods and ensure that more useful knowledge is transferred to the target domain.
3) A cross-domain recommendation method CDR-MS which enables transferring group-level knowledge from multiple sources to improve the recommendation accuracy in the target domain with insufficient data. The comparison results of the proposed method CDR-MS and six single-domain or cross-domain recommendation methods on five real-world datasets show that our proposed method outperforms other recommendation methods in sparse data.

The rest of the paper is organized as follows. Section III gives some preliminary and a formal description of the problem. Section IV describes our method CDR-MS to enable cross-domain recommendation in multiple domains. In Section V, we present our experiments on five real-world datasets containing three categories. Finally, in Section VI, conclusions are provided with future study.

II. Related Work

In this section, related work of recommender systems and cross-domain recommender systems is presented.

Recommender systems are primarily devised to assist individuals who are short on experience or knowledge deal with the vast number of choices in relation to items. Recommender systems take advantage of various sources of information to predict preferences of users in relation to different items. Although recommender systems achieved great success, data sparsity is still one of the most challenging and common issue in developing real-world recommender systems, which impairs the accuracy of recommendation. If a system fails to provide practical support, new users will quickly lose interest and stop using it. The data sparsity problem can be alleviated from introducing side informations such as user generated information and item attributes [8]. However, sometimes it is more difficult to acquire side information since users need to spend time and efforts. Another solution is to develop cross-domain recommender systems. There may be insufficient data in one domain, but relatively rich data in another domain. The abundance of data in another domain can assist the recommendation in a specific target domain. Its ability to alleviate the data sparsity problem drives the development of the cross-domain recommender systems.

Cross-domain recommender systems based on preference data can be designed in various ways according to the overlap of users and item, the form the data takes, or the tasks the system needs to handle [2]. Non-overlapping methods tend to extract shared knowledge based on collective group-level user behavior. In overlapping methods, the original source and target rating matrixes are collectively factorized, then the entities' features are extracted. Constraints on each entity ensure these features are exactly the same in the source and target domains so they can act as a bridge for knowledge transfer.

Methods that handle two domains with non-overlapping entities transfer knowledge from a group-level. Users and items are clustered into groups and knowledge is shared through group-level rating patterns. For example, code book transfer (CBT) clusters users and items into groups and extracts group-level knowledge as a "codebook" [9]. Later, a probabilistic model named rating matrix generated model (RMGM) is extended from CBT, relaxing the hard group membership to soft membership [10]. These two methods cannot ensure that the information on the two groups from two different domains is consistent, and the effectiveness of knowledge transfer is not guaranteed. [11] extend CBT with common and domain-specific rating patterns and improves the accuracy through adjustment. A cross-domain recommendation method with consistent information transfer (CIT) [7] uses a domain adaptation technique and extract consistent knowledge from the source domain. This method is superior especially when the source domain data and the target domain data are divergent. Also, some cross-domain recommendation methods deal with data with overlapping entities. Transfer by collective factorization (TCF) [12] was developed to use implicit data in the source domain to help predict explicit feedback in the target domain, such as ratings. A kernel induced recommendation method is proposed to transfer knowledge even with a small number of overlapping entities [13].

The above methods are applied to only one source domain. Also, we review related cross-domain recommendation methods dealing with multiple sources. Cross-domain triadic factorization (CDTF) [14] is a user-item-domain tensor that integrates both explicit and implicit feedback. It assumes that users are fully overlapping and that the user factor matrix is the same, thus bridging the domains. Cluster-based matrix factorization (CBMF) [15] tries to extend CDTF to partially overlapping entities, but the core of the CBMF method is the same as with non-overlapping entities, which transfer knowledge based on groups rather than using the overlapping entities as a bridge. A cross-domain recommendation method

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

is developed [16] to extract group knowledge separately with non-linear constraints in a tensor form. The methods above can take advantages of knowledge from multiple source domains but they cannot handle the problem of the domain shift at the same time.

III. Preliminary and Problem Formation

In this section, cross-domain recommendation by tri-factorization is briefly introduced. The problem targeted in this paper is also formally formulated.

A. Cross-domain Recommendation by Tri-factorization

Matrix factorization projects both users and items onto the same latent space so that they are comparable, and through their inner products reconstructs the rating matrix [17]. Similarly, the rating matrix $\boldsymbol{R} \in \mathbb{R}^{M\times N}$ (bold letters represent matrixes) can be factorized into three matrixes (suppose there are M users and N items). Users and items are clustered into several latent groups and in the middle is the group-level rating pattern: $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$, where $\boldsymbol{U} \in \mathbb{R}^{M\times K}$ is user group membership matrix, representing users clustered into K groups, $\boldsymbol{V} \in \mathbb{R}^{N\times L}$ is item group membership matrix, representing items clustered into L groups and $\boldsymbol{S} \in \mathbb{R}^{K\times L}$ is the group rating pattern matrix, i.e. the group-level knowledge.

Say rating matrixes in D domains are available, denoted as $\boldsymbol{R}^D = \{\boldsymbol{R}^1, ..., \boldsymbol{R}^d, ..., \boldsymbol{R}^D\}$. The assumption of the cross-domain recommender systems is that the group-level knowledge can be shared if these domains are similar. Thus for the dth rating matrixes are reconstructed as:

$$\hat{\boldsymbol{R}}^d = \boldsymbol{U}^d\boldsymbol{S}(\boldsymbol{V}^d)^T \tag{1}$$

B. Problem Formulation

In our problem setting, there is no correspondence on the users/items across the domains and users/items are treated as completely different. We assume that on both the source and target domains the data are explicit ratings. The problem is formally defined as:

Given D rating matrixes $\boldsymbol{R}^D = \{\boldsymbol{R}^1, ..., \boldsymbol{R}^d, ..., \boldsymbol{R}^D\}$, $\boldsymbol{R}^d \in \mathbb{R}^{M^d\times N^d}$, our goal is to develop a multi-domain recommendation method to assist the recommendation task of predicting the ratings using knowledge in one target domain $\boldsymbol{R}^t$ from all the other rating matrixes in $\boldsymbol{R}^D$, where for each source domain data $\mathcal{U}^d \cap \mathcal{U}^t = \emptyset$ and $\mathcal{I}^d \cap \mathcal{I}^t = \emptyset$.

IV. Cross-Domain Recommendation from Multiple Sources

In this section, our proposed CDR-MS method is presented beginning with an overview of the method procedure containing five steps. Each of the five steps is then explained in detail.

A. The Method Overview

The proposed method CDR-MS ensures the knowledge extracted from multiple source domains is consistent with that in the target domain. The procedure consists five steps. 1) Users and items are clustered separately and user feature matrixes and item feature matrixes are obtained; 2) Use and item group alignment is conducted to extract a shared subspace among the multiple source domains, the group-level knowledge is extracted; 3) Group-level knowledge from multiple sources is transfered to the target domain with a less restricted constraint; 4) Feature representation is regulated in the target domain to retain domain specific characteristics; 5) Recommendation in the target domain.

B. The CDR-MS method

Our proposed method consists five steps.

1) Step 1: Clustering users and items in each domain: First, users and items in each domain are clustered separately for each source domain and the target domain. We choose the Flexible Mixture Model (FMM) to cluster the users and items separately [18], since this method allows both users and items to fall into multiple groups with different memberships. This fits to the situation that users may have various preferences and items may have diverse content.

Suppose users are clustered into K user groups $\{Z_u^{(1)}, \dots, Z_u^{(K)}\}$, while items are clustered into L item groups $\{Z_v^{(1)}, \dots, Z_v^{(L)}\}$. Z_u and Z_v are two latent variables that denote the user and item groups respectively. $P(Z_u|u)$ is the conditional probability of a user belonging to a user group, denoting the group membership of the user; $P(Z_v|v)$ is the conditional probability of an item belonging to an item group, denoting its group membership. Each user group has a rating preference for each item group. r is the variable representing the preference of user groups to item groups. $P(r|Z_u, Z_v)$ is the conditional probability of r given user group Z_u and item group Z_v. The rating for a coupled user-item pair is:

$$R(u, v) = \sum_r r \sum_{Z_u, Z_v} P(r|Z_u, Z_v)P(Z_u|u)P(Z_v|v) \tag{2}$$

Equation (2) can be rewritten into matrix form:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T \tag{3}$$

where $\boldsymbol{U} \in \mathbb{R}^{M\times L}$ and $\boldsymbol{V} \in \mathbb{R}^{N\times L}$ are the user and item feature matrix. U_{ij} represents the membership of user u_i for user group $Z_u^{(j)}$. U_{i*} is the ith row of matrix U representing membership of user u_i to each group. U_{*j} is the jth column of matrix $\boldsymbol{U}$ representing the membership of each user to user group $Z_u^{(j)}$. The same goes for items. $\boldsymbol{S} \in \mathbb{R}^{K\times L}$ is the group-level knowledge matrix. S_{ij} represents the preference of user group $Z_u^{(i)}$ for item group $Z_v^{(j)}$.

2) Step 2: User and item group alignment: After clustering, the user group and item group membership matrixes $\{\boldsymbol{U}_{s_1}^{(0)}, \dots, \boldsymbol{U}_{s_d}^{(0)}\}$, $\{\boldsymbol{V}_{s_1}^{(0)}, \dots, \boldsymbol{V}_{s_d}^{(0)}\}$ are acquired for the source domains and $\boldsymbol{U}_t^{(0)}$, $\boldsymbol{V}_t^{(0)}$ for the target domain.

To extract consistent group-level knowledge from two source domains, the user groups and item groups need to be aligned of the two source domains. There are many domain adaptation methods for two domains available but seldom for multiple domains. Domain adaptation is to find an intermediate space between subspaces from the source domain and the

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

target domain. When there are d source domains available, one way to deal with the d subspaces is to compute the mean of the d subspaces. For example, Karcher has defined the mean of points on a manifold [19]. With the Karcher mean calculated, the user aligned groups and item aligned groups are obtained for both source domains and the target domain as shown in Algorithm 1, denoted as $\bar{\boldsymbol{U}}_s$, $\bar{\boldsymbol{V}}_s$, $\boldsymbol{U}_t^{(1)}$, $\boldsymbol{V}_t^{(1)}$.

Algorithm 1 User and Item Group Alignment

Input:
$\{\boldsymbol{U}_{s_1}^{(0)}, \ldots, \boldsymbol{U}_{s_d}^{(0)}\}$, the source user group membership matrixes;
$\{\boldsymbol{V}_{s_1}^{(0)}, \ldots, \boldsymbol{V}_{s_d}^{(0)}\}$, the source item group membership matrixes;
$\boldsymbol{U}_t^{(0)}$, the target user membership matrix;
$\boldsymbol{V}_t^{(0)}$, the target item membership matrix;
Output:
$\bar{\boldsymbol{U}}_s$, the aligned source user matrix;
$\bar{\boldsymbol{V}}_s$, the aligned source item matrix;
$\boldsymbol{U}_t^{(1)}$, the aligned target user matrix;
$\boldsymbol{V}_t^{(1)}$, the aligned target item matrix;
1: Calculate Z-score of $\{\boldsymbol{U}_{s_1}^{(0)}, \ldots, \boldsymbol{U}_{s_d}^{(0)}\}$
Calculate Z-score of $\{\boldsymbol{V}_{s_1}^{(0)}, \ldots, \boldsymbol{V}_{s_d}^{(0)}\}$
2: Calculate the basis of user subspaces: $\boldsymbol{P}_u = \text{PCA}\Big(f_{zs}(\boldsymbol{U}_{s_1}^{(0)}), \ldots, f_{zs}(\boldsymbol{U}_{s_d}^{(0)})\Big)$
Calculate the basis of item subspaces: $\boldsymbol{P}_v = \text{PCA}\Big(f_{zs}(\boldsymbol{V}_{s_1}^{(0)}), \ldots, f_{zs}(\boldsymbol{V}_{s_d}^{(0)})\Big)$
3: Calculate Karcher mean of $\boldsymbol{P}_u$ and $\boldsymbol{P}_v$, denoted as $\text{KM}(\boldsymbol{P}_u)$ and $\text{KM}(\boldsymbol{P}_v)$
4: $\bar{\boldsymbol{U}}_s = \boldsymbol{\Psi}_s(\text{KM}(\boldsymbol{P}_u), \boldsymbol{U}_t^{(0)})$ similar to (16-17) in [7]
$\bar{\boldsymbol{V}}_s = \boldsymbol{\Phi}_s(\text{KM}(\boldsymbol{P}_v), \boldsymbol{V}_t^{(0)})$
$\boldsymbol{U}_t^{(1)} = \boldsymbol{\Psi}_t(\text{KM}(\boldsymbol{P}_u), \boldsymbol{U}_t^{(0)})$
$\boldsymbol{V}_t^{(1)} = \boldsymbol{\Phi}_t(\text{KM}(\boldsymbol{P}_v), \boldsymbol{V}_t^{(0)})$
5: **return** $\bar{\boldsymbol{U}}_s$, $\bar{\boldsymbol{V}}_s$, $\boldsymbol{U}_t^{(1)}$ and $\boldsymbol{V}_t^{(1)}$

3) Step 3: Group-level knowledge transfer in the target domain: With the aligned user and item group representations from the source domains, the shared knowledge of the multiple sources is extracted:

$$J_s(\boldsymbol{S}_s) = \sum_{d=1}^{D} \|\boldsymbol{I}_{s_d} \circ (\boldsymbol{X}_{s_d} - \bar{\boldsymbol{U}}_s \boldsymbol{S}_s (\bar{\boldsymbol{V}}_s)^T)\|_F + \lambda_s \|\boldsymbol{S}_s\|_F \quad (4)$$

where $\boldsymbol{I}_s$ is an indicator matrix for $\boldsymbol{X}_s$, if $(I_s)_{ij} = 1$, then $(X_s)_{ij} \neq 0$ and $(I_s)_{ij} = 0$, otherwise. The same applies to $\boldsymbol{I}_t$ for $\boldsymbol{X}_t$. $\circ$ is an entry-wise product, λ_s is the parameter for regularization.

Since $\bar{\boldsymbol{U}}_s$ and $\bar{\boldsymbol{V}}_s$ are treated as mean representation for multiple source domains, a restrictive constraint is not suitable for the knowledge share between the multiple source domains and the target domain. Different from [7] that extract consistent knowledge from both the source and the target domains, a less restrictive constraint between the group-level knowledge from the source domains and that from the target domain is applied. The constrains result in user groups who are similar tend to have similar preferences and item groups tend to have similar latent factors. Specifically, the regularization form is [20]:

$$\mathcal{R}_o(\boldsymbol{S}_t) = \text{tr}(\boldsymbol{S}_t^T \boldsymbol{L}_s^u \boldsymbol{S}_t) \quad (5)$$

where tr is the trace of the matrix, $\boldsymbol{L}$ denotes a Laplacian matrix, and $\boldsymbol{L}_s^u = \boldsymbol{D}_s^u - \boldsymbol{W}_s^u$. $\boldsymbol{W}_s^u$ is the user group similarity matrix, and $\boldsymbol{D}_s^u$ is a diagonal matrix defined as $[D_s^u]_{ii} = \sum_j [W_s^u]_{ij}$. User group similarities are measured from the user group-level knowledge by RBF measurement : $[W_s^u]_{ij} = e^{-\frac{\|S_{si*} - S_{sj*}\|^2}{\sigma^2}}$, where σ^2 is set to be the median of all the non-zero values calculated by $\|S_{si*} - S_{sj*}\|^2$. Similarly, the item constrains is calculated as $\boldsymbol{L}_s^v$ from the column view of $\boldsymbol{S}_s$.

We achieve group-level knowledge transfer by minimizing the following objective function:

$$\begin{aligned} J_t(\boldsymbol{S}_t) = &\|\boldsymbol{I}_t \circ (\boldsymbol{X}_t - \boldsymbol{U}_t^{(1)} \boldsymbol{S}_t (\boldsymbol{V}_t^{(1)})^T)\|_F \\ &+ \lambda_u \text{tr}(\boldsymbol{S}_t^T \boldsymbol{L}_s^u \boldsymbol{S}_t) + \lambda_v \text{tr}(\boldsymbol{S}_t \boldsymbol{L}_s^v \boldsymbol{S}_t^T) + \lambda_t \|\boldsymbol{S}_t\|_F \end{aligned} \quad (6)$$

where λ_u ,λ_v and λ_t are trade-off parameters of the regularization.

4) Step 4: Group representation regulation: In our problem setting, some domain-specific characteristics are embedded in the small amount of available data in the target rating matrix. To reveal these idiosyncrasies of the target domain, we amend feature representations of the target rating matrix to make the model fit better to the task in target rating matrix. The representation regulation is achieved through an optimization problem. The cost function is:

$$J_r(u, v) = \|\boldsymbol{I}_t \circ (\boldsymbol{X}_t - \boldsymbol{U}_t^{(1)} \boldsymbol{u} \boldsymbol{S}_t (\boldsymbol{V}_t^{(1)} \boldsymbol{v})^T)\|_F \quad (7)$$

The tuning factors can be learned through optimizing

$$\min_{\boldsymbol{u}, \boldsymbol{v}} J_r(\boldsymbol{u}, \boldsymbol{v})$$
$$\text{s.t. } u \geq 0, v \geq 0$$

The optimization problem is solved by alternatively estimating tuning factors $\boldsymbol{u}$ and $\boldsymbol{v}$. For more details, see [7].

5) Step 5: Recommendation in the target domain: The recommendation in target domain is given by Equation (8).

$$\hat{\boldsymbol{X}}_t = (\boldsymbol{U}_t^{(1)} \boldsymbol{u}) \boldsymbol{S} (\boldsymbol{V}_t^{(1)} \boldsymbol{v})^T \quad (8)$$

where $\hat{\boldsymbol{X}}_t$ is the reconstructed user-item rating matrix for prediction, $\boldsymbol{u}$, $\boldsymbol{v}$ are user and item tuning factors for target domain, $\boldsymbol{S}$ is the consistent knowledge, $\boldsymbol{U}_s^{(1)}$, $\boldsymbol{U}_t^{(1)}$ are user and item feature matrixes for the target domain after subspace alignment obtained from the above steps.

V. Experiments

In this section, the proposed method CDR-MS is evaluated. First, the datasets and evaluation metrics used is introduced, followed by experimental settings and the baseline methods. The results of the experiments are presented. The parameter analysis is in the end.

A. Datasets and Evaluation Metrics

To test our proposed method, we need to choose data from similar data so that transfer learning is meaningful, but divergence still exists between the source and target domains. Our experiments comprise nine cross-domain recommendation tasks on three categories. Five real-world datasets were used: EachMovie[1], Movielens1M[2], LibraryThing[3], Amazon Book[4] and YahooMusic[5]. Each is publicly available and has been used to test recommender systems in a variety of scenarios for recommender systems in single domain. But tests on these datasets in cross-domain setting, particularly on multi-domain setting are lacking. For AmazonBooks, we removed all users who had given exactly the same rating for every book, as these data are not effective for constructing a recommender system [7]. EachMovie and LibraryThing were normalized to the range of $\{1, 2, 3, 4, 5\}$ before conducting experiments.

The statistical information for original datasets is provided in Table I.

Across all the datasets, 1000 items that had been rated more than 10 times were randomly chosen. We then filtered out the users who had given less than a total of 20 ratings. For the source domain data, we randomly selected 500 users to be regular customers of the site. The source domain data were controlled to be more dense than the target domain data. For the target domain data, we randomly selected 300 users to be regular customers of the site, and another 200 users to be new customers. For new users, five observed ratings were given, and the rest of the ratings were used for evaluation. In the end, the rating matrixes for both the source and target domains were all 500×1000 matrixes. The details of the final datasets are summarized in Table II.

Mean absolute error (MAE) and root mean square error (RMSE) were used as the evaluation metrics:

$$MAE = \sum_{u,v,X_{uv} \in Y} \frac{|\hat{X}_{uv} - X_{uv}|}{|Y|}$$

$$RMSE = \sqrt{\sum_{u,v,X_{uv} \in Y} \frac{(\hat{X}_{uv} - X_{uv})^2}{|Y|}}$$

where Y is the test set, and $|Y|$ is the number of test ratings.

B. Experimental Settings and Baselines

The rating average is a very important statistics of the data which we used in our experiments to represent whether data in two domains are of high similarity or not. According to Table II, we can see a big divergence in the rating average between the source domain data and the target domain data. This fits to our problem setting in Section III that data in the source domain and the target domain are similar but divergence still exists.

[1]http://www.cs.cmu.edu/∼lebanon/IR-lab/data.html#intro
[2]http://grouplens.org/datasets/movielens/1m/
[3]https://www.librarything.com
[4]http://jmcauley.ucsd.edu/data/amazon/
[5]https://webscope.sandbox.yahoo.com/catalog.php?datatype=r

Three non-transfer learning methods and two cross-domain methods were chosen as comparisons for the proposed method. The non-transfer learning methods were: Pearson's correlation coefficient (PCC) [21], FMM [18] and single value decomposition (SVD) [17]. The cross-domain methods were: CBT [9], RMGM [10] and CIT [7]. PCC uses user-based CF, and the number of neighborhoods was set at 50. For SVD, the latent feature number was fixed at 40, the regularization factor was set to 0.015, and the learning rate was set to 0.003. For FMM, CBT, RMGM and CIT, the user group number and item group number were both set to 40. For the proposed method, CDR-MS, the user and item group number were both set to 40. λ_s,λ_u ,λ_v and λ_t are trade-off parameters for the regularization in CDR-MS. They are adjusted with grid search and by a validation set within the range of 0.0001, 0.001,0.005, 0.01, 0.05. Further analysis of the parameters is provided later. All the methods (except for PCC) need to initialize the factorized matrix randomly, we ran 10 random initializations and report the averaged results and standard deviations.

C. Results

The experiment results of our proposed CDR-MS compared with the other six baselines on two accuracy metrics are presented in Table III, IV and V. Overall, CDR-MS has the best performance in recommendation tasks on the three categories. These results indicate that CDR-MS can extract knowledge from source sources that can help increase the recommendation accuracy in the target domain. Our analysis of the results revealed the following observations:

1) Comparison with recommendation methods in a single target domain. The performance of non-transfer learning methods was relatively poor on sparse data. As the basis of CBT and RMGM, FMM was designed to predict ratings for users with little available data. But in our experiment results, the performance of FMM in the target domain is greatly affected by data sparsity. In some severe setting, PCC fail to generate useful recommendations, e.g. in the book recommendation. In all the experiment results, KerKT significantly outperformed all the non-transfer learning recommendation techniques.
2) Comparison with cross-domain recommendation methods with one source domain. CBT, RMGM and CIT are three cross-domain recommendation methods that use knowledge from one source domain. CBT and RMGM showed improved precision in recommendations over its basis, FMM, but sometimes the improvement was not significant (see Table V) or sometimes suffered from negative transfer (see Table IV). CIT is the state-of-art method and it avoids negative transfer on the three categories, due to its advantage of effective knowledge transfer. Except for MAE in the music category, CDR-MS outperforms each of these methods, proving that the knowledge extracted by CDR-MS from multiple sources is more effective in assisting the recommendation in the target domain. Experiments are conducted on CBT,

TABLE I: Statistical Information on the Original Datasets

	EachMovie	Movielens 1M	LibraryThing	AmazonBook	YahooMusic_1	YahooMusic_2
#user	72916	6040	7279	8026324	200000	200000
#item	1628	3900	37232	2330066	136736	136736
#rating	2811983	1000209	749401	22507155	78344627	78742463
sparsity	97.63%	95.75%	99.72%	99.99%	99.71%	99.71%
range	0-1	1-5	0.5-5	1-5	1-5	1-5

TABLE II: Description of Data Subsets in Three Categories

Data type	Dataset	Domain	Sparsity	Average
Movie	EachMovie	source	96.00%	4.32
	Movielens1M	target	98.50%	2.91
Book	LibraryThing	source	87.43%	3.97
	AmazonBook	target	99.69%	3.12
Music	YahooMusic_1	source	95.70%	4.14
	YahooMusic_2	target	99.00%	2.70

RMGM and CIT on two source data respectively. The results of cross-domain recommendation are affected when the source data are different. It shows that CDR-MS is not relying on one source data but takes advantages of data from both sources, as CDR-MS outperforms each cross-domain recommendation with one source domain.

TABLE III: Recommendation Results on the Movie Target Domain

Methods		Source Data	MAE	RMSE
Single-Domain	PCC	–	1.2123	1.5722
	FMM	–	1.3543±0.0171	1.6767±0.0201
	SVD	–	1.0935±0.0036	1.3542±0.0057
Cross-Domain	CBT	book	1.1785±0.0198	1.4659±0.0317
		music	1.2234±0.0096	1.5452±0.0181
	RMGM	book	1.2966±0.0193	1.6284±0.0226
		music	1.2394±0.0145	1.5453±0.0162
	CIT	book	1.0185±0.0033	1.2291±0.0020
		music	1.0176±0.0021	1.2279±0.0027
Multiple	CDR-MS	book&music	**1.0014±0.0023**	**1.2206±0.0013**

TABLE IV: Recommendation Results on the Book Target Domain

Methods		Source Data	MAE	RMSE
Single-Domain	PCC	–	-	-
	FMM	–	1.5426±0.0219	1.8730±0.0214
	SVD	–	1.0417±0.0057	1.2839±0.0078
Cross-Domain	CBT	movie	1.2259±0.0311	1.5487±0.0272
		music	1.1110±0.0089	1.4147±0.0216
	RMGM	movie	1.2911±0.0198	1.6167±0.0223
		music	1.2426±0.0154	1.5585±0.0207
	CIT	movie	0.9969±0.0085	1.2086±0.0085
		music	0.9931±0.0038	1.2061±0.0050
Multiple	CDR-MS	movie&music	**0.9810±0.0014**	**1.2034±0.0014**

D. Parameter Analysis and Complexity Analysis

We analyzed how the parameters K and L affect the performance of CDR-MS. Due to the space limitation, only the result on the movie category is presented. To analyze K and L, grid search is used with evaluation metrics of both MAE and RMSE as shown in Fig. 1. Since the performance of other baselines is not comparable with the ones in Fig. 1, they are omitted. The result of analysis shows that the performance of SVD, CIT and CDR-MS is not greatly affected by these parameters. However, the complexity of the method will significantly increase with the increase of K and L. The time consumed by different K is shown in Table VI. For simplicity, the setting of L is set to be the same as K. The time consumption shown in Table VI is for 10 iterations of the proposed method. For fair comparison with other baselines, we choose 40 for both K and L for CDR-MS in our comparison experiments. The total complexity of CDR-MS is $O(n)$.

TABLE V: Recommendation Results on the Music Target Domain

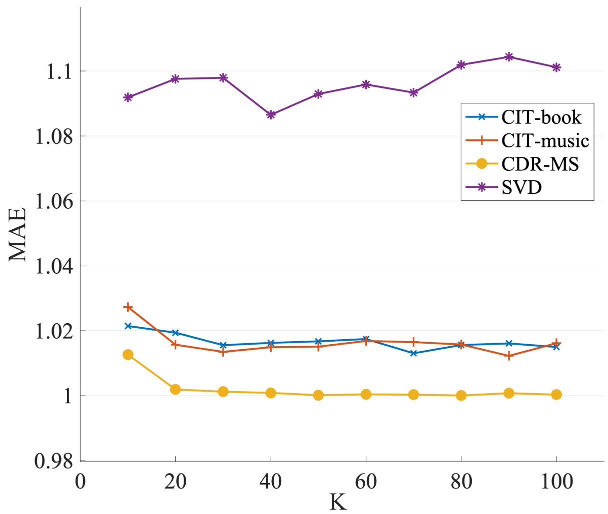
Methods		Source Data	MAE	RMSE
Single-Domain	PCC	–	1.7936	2.2436
	FMM	–	1.4717±0.0211	1.8185±0.0219
	SVD	–	1.3426±0.0053	1.5663±0.0062
Cross-Domain	CBT	movie	1.7635±0.0211	2.1091±0.0220
		book	1.5963±0.0153	1.8764±0.0227
	RMGM	movie	1.4972±0.0194	1.8216±0.0263
		book	1.4648±0.0305	1.7623±0.0352
	CIT	movie	**1.3243±0.0092**	1.5104±0.0096
		book	1.3309±0.0101	1.5168±0.0090
Multiple	CDR-MS	movie&book	1.3379±0.0020	**1.5004±0.0016**

TABLE VI: Time consumption with different settings of K

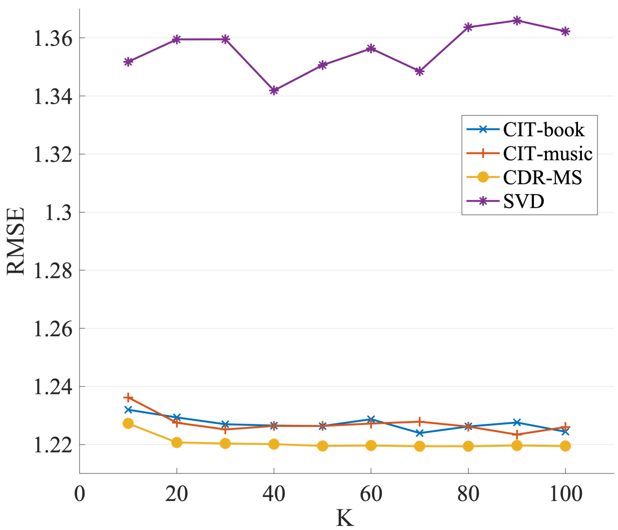
K	MAE	RMSE	Time(s)
$K = 10$	1.0139	1.2297	200.00
$K = 20$	1.0022	1.2208	321.34
$K = 30$	1.0008	1.2206	515.72
$K = 40$	1.0014	1.2207	759.87
$K = 50$	1.0014	1.2203	1051.55
$K = 60$	1.0002	1.2194	1751.07
$K = 70$	1.0006	1.2195	2385.67
$K = 80$	1.0002	1.2194	2977.87
$K = 90$	1.0002	1.2194	3527.90
$K = 100$	1.0001	1.2192	4687.98

VI. Conclusion and Future Work

In this paper, we develop a cross-domain recommendation method with multiple sources named CDR-MS, to improve recommendation performance in a sparse target domain. To address two challenging issues: 1) group-level knowledge inconsistency and 2) group-level knowledge insufficiency, a domain adaptation method and a matrix factorization method with flexible constraints are developed. In this way, CDR-MS ensures the effectiveness of the group level knowledge extraction and adaptation from multiple sources with domain shift existed. With the virtue of knowledge extracted from multiple source domains, CDR-MS alleviates the data sparsity problem



(a) Parameter analysis on MAE



(b) Parameter analysis on RMSE

Fig. 1: Parameter analysis with different setting on K.

and increases the prediction accuracy in cross-domain recommendation. Experiments on five real-world datasets with three categories demonstrate that our method CDR-MS achieves the best performance compared with six baselines including both single-domain and cross-domain recommendation methods.

In the future, we will solve the problem of cross-domain recommendation with multiple target domains together with multiple source domains. We will also try to develop active learning strategies for cross-domain recommender system with multiple sources, which is necessary when the number of source domains increases.

Acknowledgment

This work was supported by the Australian Research Council (ARC) under the Australian Laureate Fellowship [FL190100149].

References

[1] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, "Recommender system application developments: a survey," *Decision Support Systems*, vol. 74, pp. 12–32, 2015.

[2] I. Cantador, I. Fernández-Tobías, S. Berkovsky, and P. Cremonesi, "Cross-domain recommender systems," in *Recommender Systems Handbook*. Springer, 2015, pp. 919–959.

[3] P. Hao, G. Zhang, L. Martinez, and J. Lu, "Regularizing knowledge transfer in recommendation with tag-inferred correlation," *IEEE Transactions on Cybernetics*, no. 99, pp. 1–14, 2017.

[4] M. Jiang, P. Cui, X. Chen, F. Wang, W. Zhu, and S. Yang, "Social recommendation with cross-domain transferable knowledge," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3084–3097, 2015.

[5] W. Chen, H. Yin, W. Wang, L. Zhao, W. Hua, and X. Zhou, "Exploiting spatio-temporal user behaviors for user linkage," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 2017, pp. 517–526.

[6] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: a survey," *Knowledge-Based Systems*, vol. 80, pp. 14–23, 2015.

[7] Q. Zhang, D. Wu, J. Lu, F. Liu, and G. Zhang, "A cross-domain recommender system with consistent information transfer," *Decision Support Systems*, vol. 104, pp. 49–63, 2017.

[8] Y. Shi, M. Larson, and A. Hanjalic, "Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges," *ACM Computing Surveys*, vol. 47, no. 1, p. 3, 2014.

[9] B. Li, Q. Yang, and X. Xue, "Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction," in *IJCAI*, vol. 9, 2009, pp. 2052–2057.

[10] ——, "Transfer learning for collaborative filtering via a rating-matrix generative model," in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 617–624.

[11] S. Gao, H. Luo, D. Chen, S. Li, P. Gallinari, and J. Guo, "Cross-domain recommendation via cluster-level latent factor model," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2013, pp. 161–176.

[12] W. Pan and Q. Yang, "Transfer learning in heterogeneous collaborative filtering domains," *Artificial Intelligence*, vol. 197, pp. 39–55, 2013.

[13] Q. Zhang, J. Lu, D. Wu, and G. Zhang, "A cross-domain recommender system with kernel-induced knowledge transfer for overlapping entities," *IEEE Transactions on Neural Networks and Learning Systems*, 2018.

[14] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and C. Zhu, "Personalized recommendation via cross-domain triadic factorization," in *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 2013, pp. 595–606.

[15] N. Mirbakhsh and C. X. Ling, "Improving top-n recommendation for cold-start users via cross-domain information," *ACM Transactions on Knowledge Discovery from Data*, vol. 9, no. 4, p. 33, 2015.

[16] Y.-F. Liu, C.-Y. Hsu, and S.-H. Wu, "Non-linear cross-domain collaborative filtering via hyper-structure transfer," in *International Conference on Machine Learning*, 2015, pp. 1190–1198.

[17] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, 2009.

[18] L. Si and R. Jin, "Flexible mixture model for collaborative filtering," in *Proceedings of the 20th International Conference on Machine Learning*, 2003, pp. 704–711.

[19] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on Pure and Applied Mathematics*, vol. 30, no. 5, pp. 509–541, 1977.

[20] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2010.

[21] M. Deshpande and G. Karypis, "Item-based top-n recommendation algorithms," *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 143–177, 2004.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE