LKT-FM: A Novel Rating Pattern Transfer Model for Improving Non-Overlapping Cross-Domain Collaborative Filtering

Yizhou Zang and Xiaohua Hu

Drexel University, Philadelphia PA 19104, USA,
yizhouzang@gmail.com

Abstract. Cross-Domain Collaborative Filtering (CDCF) has attracted various research works in recent years. However, an important problem setting, i.e., "users and items in source and target domains are totally different", has not received much attention yet. We coin this problem as Non-Overlapping Cross-Domain Collaborative Filtering (NOCDCF). In order to solve this challenging CDCF task, we propose a novel 3-step rating pattern transfer model, i.e. low-rank knowledge transfer via factorization machines (LKT-FM). Our solution is able to mine high quality knowledge from large and sparse source matrices, and to integrate the knowledge without losing much information contained in the target matrix via exploiting Factorization Machine (FM). Extensive experiments on real world datasets show that the proposed LKT-FM model outperforms the state-of-the-art CDCF solutions.

1 Introduction

Cross-Domain Collaborative Filtering(CDCF) is an emerging research topic in recommender systems. It aims to improve recommendations in an individual domain by drawing upon the knowledge acquired from related domains. Most CDCF models transfer knowledge based on explicit correspondence among entities of target and source domains [1][2][3]. However, few works have studied a more practical problem setting, in which "users and items in source and target domains are totally different". This setting is called Non-Overlapping CDCF (NOCDCF), which is the most challenging problem in cross-domain recommendation.

For the NOCDCF problem, the most well-known solutions may be the RPT-based (Rating Pattern Transfer) methods [5][9]. This set of methods shares across domains a group-level preference, which is referred to as a rating pattern. For example, suppose a newly opened book-selling website would like to build a recommender system. Due to the lack of visiting at the beginning, very few ratings are available for collaborative filtering. Fortunately, in the meanwhile, there is a popular movie review website shares its rating data to the public. Since movie domain is correlated to book domain in some aspects (First, they have correspondence in genres, e.g. comedy movies corresponds to humorous books.

Second, the user sets of two domains may be sampled from the same population and reflect similar social aspects [5][17] even though they don't overlap), similar user-item rating pattern is deemed to exist in both domains. RPT methods extract such rating pattern by co-clustering the rows (users) and columns (movies) of the source matrix. The knowledge of source domain (movie domain) is then transferred to the target domain (book domain) via sharing the co-clusters.

However, existing NOCDCF approaches have their limitations:

1. All these methods are based on the assumption that the source rating matrix is dense. Unfortunately, as we know, canonical CF datasets such as Amazon, Netflix and MovieLens rating sets are mostly sparse. The sparseness may considerably degrade the performance of existing methods. The reason for this is that current methods need to impute the missing values in the source matrix in order to apply co-clustering algorithms, which is undefined when the matrix is incomplete [4]. For sparse matrices, the imputation may easily distort the data and further affect the clustering quality. As a result, a lot of noise will be introduced into the co-clusters and thus lower the recommendation performance. Current methods try to avoid the impact of sparseness by exploiting a dense but small portion of the original source matrix [5][6]. However, this does not solve the whole problem: using only a dense subset might lose the useful knowledge contained in the remaining large though sparse portion of the original matrix. The decrease of recommendation accuracy of existing NOCDCF methods with the increase of data sparseness is also observed in our empirical studies in Section.4. Therefore, it raises a demand to devise a novel model that can extract knowledge from sparse matrices.
2. Current NOCDCF methods do not integrate the shared knowledge very well. An earlier work [19] observes that existing NOCDCF methods cannot transfer knowledge under some conditions. We argue that this is caused by the knowledge integration method they use. Most of these methods utilize Direct-Expansion (DE) for knowledge integration [6][9]. However, the DE approach relies too heavily on the shared knowledge and may miss useful knowledge contained in the target matrix itself. As a result, the transferred knowledge would hurt the recommendation performance. Such negative transfer is also observed in our empirical studies in Section.4. We find that current NOCDCF methods perform even worse than some single domain recommendation methods while using DE. Therefore, it brings a new challenge of how to integrate the shared knowledge while avoiding negative transfer.

To overcome the above limitations, we propose a novel 3-step RPT model, i.e. low-rank knowledge transfer via factorization machines (LKT-FM). In our first step, we factorize the source matrix into two low-rank matrices (i.e. user-factor and item-factor matrices), and then generate clusters from them. In the second step, we follow the idea in [5] to map users/items of target domain to corresponding clusters. In the third step, we expand the design matrix of Factorization Machine (FM), which incorporates the shared knowledge into the target data in a seamless manner. We then conduct extensive empirical studies on several real world datasets and achieve considerably better results than the state-of-the-art methods. Note that we also compare our LKT-FM with ordinary CDCF methods, which requires overlapping information as additional input, and achieve competitive results. Therefore, our model is a generic solution to CDCF problems, not limited to NOCDCF.

The contribution of this paper can be summarized as follows:

- We devise a novel rating pattern transfer solution LKT-FM for solving NOCDCF problem.
- We perform experiments on real world datasets and show that our LKT-FM method is considerably better than the state-of-the-art NOCDCF methods in terms of both knowledge extraction and knowledge integration, and is also competitive with ordinary CDCF methods even in an overlapping setting.

2 Background

2.1 Problem Definition

As mentioned in Section.1, in NOCDCF, there is no overlap between users and items across domains. More formally, assume we have a source domain D_S and a target domain D_T. Respectively for such domains, let U_S, U_T be their sets of users, I_S, I_T be their sets of items, and $\boldsymbol{X}_S$, $\boldsymbol{X}_T$ be their user-item rating matrices. Our goal is to predict the unobserved entries in $\boldsymbol{X}_T$ by taking advantage of the knowledge in $\boldsymbol{X}_S$ with the restriction that $U_S \cap U_T = \emptyset$ and $I_S \cap I_T = \emptyset$.

2.2 Non-Overlapping CDCF (NOCDCF)

The seminal paper [5] proposed one of the first NOCDCF methods that exploit rating pattern (a.k.a. codebook) to transfer knowledge. Its name is CodeBook Transfer (CBT). CBT is an adaptive knowledge transfer approach. It consists of 3 steps: 1) rating pattern construction, 2) cluster membership mapping and 3) rating pattern integration. In the first step, a rating pattern is constructed via applying co-clustering on the source matrix $\boldsymbol{X}_S$ so as to obtain a rating pattern $\boldsymbol{B}$. In the second step, each user/item is assigned to the cluster identified in the rating pattern $\boldsymbol{B}$. In the third step, the filled target matrix $\widehat{\boldsymbol{X}}_T$ is obtained by expanding the rating pattern $\boldsymbol{B}$.

In a later work, [9] extended the same idea and proposed a probabilistic approach that transfers knowledge in a collective manner. More recently, [6] believed that a rating pattern consists of two substructures, a domain-specific rating pattern and a common rating pattern. Each domain has its own domain-specific rating pattern, while all correlated domains share the common rating pattern. Furthermore, [10] learned the relatedness between different source domains and target domain, and then integrated appropriate amount of knowledge (i.e. rating pattern) from each source domain to the target domain. However, as mentioned in Section.1, all the existing NOCDCF methods depend on co-clustering the imputed matrix, thus do not work well in a sparse setting. In addition, all these methods do not leverage MF techniques for knowledge integration, which may lead to negative transfer due to the loss of knowledge contained in the target matrix.

2.3 Matrix Factorization (MF)

In this work, we apply MF as a pre-processing method to obtain a low-rank representation of users and items (i.e. $\boldsymbol{U}$ and $\boldsymbol{V}$), in order for further clustering. While there are several variants of MF [11][12][13], we only review the basic MF in this paper.

Matrix factorization (MF) may be the most common and successful technique for single-domain recommendation tasks [7]. The basic idea of MF is to approximate the rating matrix using the product of two low-rank latent factor matrices:

$$\boldsymbol{X}_{ij} \approx \widehat{\boldsymbol{X}}_{ij} = \boldsymbol{S}_{ij} + \boldsymbol{U}_i \boldsymbol{V}_j^T \ . \tag{1}$$

where $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ represents the rating matrix (m is number of users and n is number of items), $\boldsymbol{S}$ indicates the bias matrix [7], $\boldsymbol{U} \in \mathbb{R}^{m \times d}$ is the user-factor latent matrix and $\boldsymbol{V} \in \mathbb{R}^{n \times d}$ is the item-factor latent matrix. The system learns by minimizing the squared error function as follows, only considering the observed ratings:

$$\min \Sigma_{i=1}^{m} \Sigma_{j=1}^{n} \boldsymbol{I}_{ij} (\boldsymbol{X}_{ij} - \widehat{\boldsymbol{X}}_{ij})^2 + \lambda_1 ||\boldsymbol{U}||_F^2 + \lambda_2 ||\boldsymbol{V}||_F^2 \ . \tag{2}$$

where $\boldsymbol{I}_{ij}$is the indicator function that equals 1 if user rated item , and equals 0 otherwise, λ_1 and λ_2 are constants controlling the extent of regularization, and $|| \cdot ||_F^2$ denotes Frobenius norm.

2.4 Factorization Machine (FM)

In this work, we apply FM [8] for knowledge integration. FM is a generic predictive model that allows to mimic most collaborative filtering (CF) models by feature engineering. More specifically, in a rating prediction problem, let S denote the set of tuples $(\boldsymbol{x}, y)$ where $\boldsymbol{x} = (x_1, ..., x_k) \in \mathbb{R}^k$ is a k-dimensional feature vector and y is corresponding class label. FM models all possible interactions between variables in $\boldsymbol{x}$ using factorized interactions. The FM model considering pairwise interactions can be represented as follows:

$$\widehat{y}(\boldsymbol{x}) = w_0 + \Sigma_{i=1}^{k} w_i x_i + \Sigma_{i=1}^{k} \Sigma_{i'=i+1}^{k} \boldsymbol{v}_i \boldsymbol{v}_{i'} x_i x_{i'} \tag{3}$$

where $w_0 \in \mathbb{R}$ is the global bias, $w_i \in \mathbb{R}$ are the biases of feature i, vector $\boldsymbol{v}_i \in \mathbb{R}^{1 \times f}$ are interaction parameter vectors of feature i. In FM, the original interaction parameters $w_{ii'}$ are replaced by the product of $\boldsymbol{v}_i$ and $\boldsymbol{v}_{i'}$. By doing this, the number of parameters decreases significantly and thus the interactions can be estimated even under high data sparsity. In practice, FM performs prominently for various CF tasks [8] and thus is a very strong baseline for single-domain recommendation evaluation. Although FM works remarkably for single-domain recommendation problems, how to use it for CDCF problems still remains open. To the best of our knowledge, there is only one previous work using FM for CDCF. However, that work [15] solved a much easier CDCF task in which the users in different domains totally overlap. Thus, it is different from our goal of solving the NOCDCF problem.

3 LKT-FM

Our proposed LKT-FM solution follows the 3-step framework of CBT model. We illustrate our model in Fig.1. Specifically, in the first step, we aim to extract high quality rating pattern via low-rank clustering; in the second step, we assign each user/item in target domain to corresponding cluster; and in the third step, we propose to integrate the extracted rating pattern through feature expansion of factorization machines (FM). These three steps are described in detail in the following three subsections.

3.1 Low-rank Rating Pattern Construction

As discussed in Section.1, constructing rating patterns through user-item co-clustering has potential issues when the source matrix is sparse. Thus, we propose a new construction method to alleviate the sparseness problem.

We first preprocess the source matrix $\boldsymbol{X}_S \in \mathbb{R}^{m_S \times n_S}$ by applying basic MF [7]. As shown in the upper half part of Fig.1, the user-item rating matrix is factorized into two low-rank matrices $\boldsymbol{U}_S \in \mathbb{R}^{m_S \times d}$ and $\boldsymbol{V}_S \in \mathbb{R}^{n_S \times d}$, $d \ll m_S, n_S$. $\boldsymbol{U}_S$ is user latent factor matrix and $\boldsymbol{V}_S$ is item latent factor matrix. Secondly, we apply K-means clustering on the row vectors of $\boldsymbol{U}_S$ and $\boldsymbol{V}_S$ respectively to generate user/item clusters. Note that, since $\boldsymbol{U}_S$ and $\boldsymbol{V}_S$ are both complete matrices (i.e. no missing values), no imputation is needed before clustering. Thus, it effectively avoids the impact of imputation, which may introduce much noise when source matrix is sparse. Fig.1 shows the obtained user-cluster membership matrix $\boldsymbol{P}_S \in \{0,1\}^{m_S \times p}$ and item-cluster membership matrix $\boldsymbol{Q}_S \in \{0,1\}^{n_S \times q}$, where p, q are given cluster numbers. Then the clustering-level rating pattern $\boldsymbol{B}$ is constructed as follows:

$$\boldsymbol{B} = [\boldsymbol{P}_S^T \boldsymbol{X}_S \boldsymbol{Q}_S] \oslash [\boldsymbol{P}_S^T \mathbf{1}\mathbf{1}^T \boldsymbol{Q}_S] \tag{4}$$

where $\oslash$ denotes the entry-wise division. Eq.4 means averaging the ratings of each user-item co-cluster as an entry in $\boldsymbol{B}$.

3.2 Cluster Membership Mapping

In the second step, source and target domains are bridged by mapping users/items in the source domain to the clusters identified in $\boldsymbol{B}$. We adopt the mapping

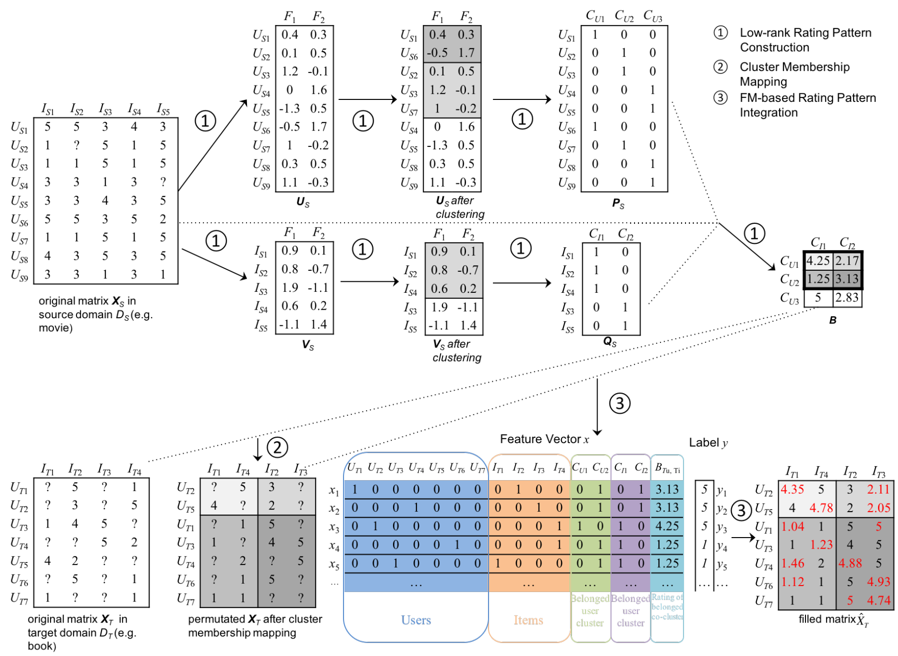

Fig. 1. Illustration of LKT-FM

method proposed in [5]. We learn such mapping by minimizing the following quadratic loss function:

$$\mathcal{L} = ||[\boldsymbol{X}_T - \boldsymbol{P}_T \boldsymbol{B} \boldsymbol{X}_T^T] \circ \boldsymbol{W}||_F^2, \quad \text{s.t.} \boldsymbol{P}_T \mathbf{1} = \mathbf{1}, \boldsymbol{Q}_T \mathbf{1} = \mathbf{1} \tag{5}$$

where $\boldsymbol{X}_T \in \mathbb{R}^{m_T \times n_T}$ is the target matrix, $\boldsymbol{P}_T \in \{0,1\}^{m_T \times p}$ and $\boldsymbol{Q}_T \in \{0,1\}^{n_T \times q}$ are cluster membership matrices, $\circ$ denotes the entry-wise product, and $\boldsymbol{W}$ denotes a binary weighting matrix where $\boldsymbol{W}_{ij} = 1$ if user i rated item j and $\boldsymbol{W}_{ij} = 0$ otherwise. We learn the parameters $\boldsymbol{P}_T$ and $\boldsymbol{Q}_T$ by applying algorithm.2 in [5].

3.3 FM-based Rating Pattern Integration

Once the rating pattern $\boldsymbol{B}$ and the membership matrices $\boldsymbol{P}_T$ and $\boldsymbol{Q}_T$ are obtained, existing NOCDCF approaches [5][6] usually incorporate the transferred knowledge in $\boldsymbol{B}$ by expanding $\boldsymbol{B}$ directly. More specifically, target matrix $\boldsymbol{X}_T$ is reconstructed by duplicating the rows and columns of $\boldsymbol{B}$ using $\boldsymbol{P}_T \boldsymbol{B} \boldsymbol{Q}_T^T$. Fig.2 illustrates this incorporation process. More formally, the target matrix $\boldsymbol{X}_T$ is approximated by $\widehat{\boldsymbol{X}}_T$, which is defined as follows:

$$\widehat{\boldsymbol{X}}_T = \boldsymbol{W} \circ \boldsymbol{X}_T + [\mathbf{1} - \boldsymbol{W}] \circ [\boldsymbol{P}_T \boldsymbol{B} \boldsymbol{Q}_T^T] \tag{6}$$

We call this method Direct Expansion (DE), which may miss useful knowledge in the target matrix itself according to the previous analysis. Thus, in this work, we propose a new knowledge integration approach to solve this problem. We treat rating pattern as the side information of collaborative filtering tasks and incorporate it by using Factorization Machines (FM). The new approach is superior to the conventional solution as it takes advantage of both RPT and matrix factorization techniques.

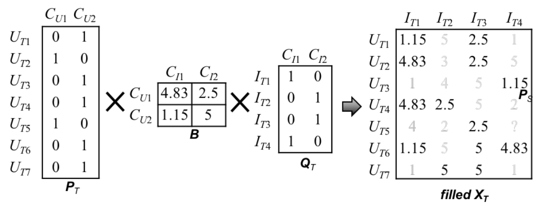

Fig. 2. Rating Pattern Incorporation by Direct Expansion (DE)

First, assume U_T and I_T to be the sets of users and items in the target domain D_T. The rating prediction problem in D_T can be modeled by a target function $f : U_T \times I_T \to \mathbb{R}$. According to FM, each user-item interaction $(u, i) \in U_T \times I_T$ in $\boldsymbol{X}_T$ is represented by a feature vector $\boldsymbol{x} \in \mathbb{R}^k$, $k = |U_T| + |I_T|$. The feature vector $\boldsymbol{x}$ consists of binary variables indicating which user rated which item. In other words, for each non-zero entry x_{ui} in $\boldsymbol{X}_T$, its corresponding feature vector $\boldsymbol{x}$ can be represented as:

$$\boldsymbol{x} = (\underbrace{0, ..., 0, 1, 0, ..., 0}_{|U_T|}, \underbrace{0, ..., 0, 1, 0, ..., 0}_{|I_T|}) \tag{7}$$

where the first $|U_T|$ binary indicator variables represent user , and the following $|I_T|$ binary indicator variables represent item .

Given the rating pattern $\boldsymbol{B}$, cluster membership matrices $\boldsymbol{P}_T$ and $\boldsymbol{Q}_T$, we can then incorporate the rating pattern by adding more features into $\boldsymbol{x}$. There are various possible ways of extending the vector $\boldsymbol{x}$. We first provide a straightforward solution, which adds three types of information, as follows:

$$\boldsymbol{x} = (\underbrace{0, ..., 0, 1, 0, ..., 0}_{|U_T|}, \underbrace{0, ..., 0, 1, 0, ..., 0}_{|I_T|}, \underbrace{0, ..., 0, 1, 0, ..., 0}_{p}, \underbrace{0, ..., 0, 1, 0, ..., 0}_{q}, \boldsymbol{B}_{C_u, C_i}) \tag{8}$$

where p, q denote the numbers of user clusters and item clusters. In the added part, the first p binary indicator variables represent which cluster user u belongs to, the next q binary indicator variables refer to which cluster item i belongs to, and $\boldsymbol{B}_{C_u, C_i}$ denotes the cluster-level rating of user cluster C_u to item cluster C_i. The extended feature vector $\boldsymbol{x}$ then serves as the input for Eq.3, and the output y is the predicted rating of user u gives to item i.

In our empirical studies, we also tried other six solutions for expanding design matrix. Unlike Eq.8 that integrates all the three new features, those six solutions make use of partial information extracted. The details of these six expansions are given in Appendix. We compare and discuss the differences of these expansions in Section.4.

3.4 Algorithm

Input: Source rating matrix $\boldsymbol{X}_S$ and target rating matrix $\boldsymbol{X}_T$.

Output: The predicted value r for each missing entry in $\boldsymbol{X}_T$

Step 1. Low-rank rating pattern construction

Step 1.1. Reduce dimension of $\boldsymbol{X}_S$ by applying basic matrix factorization algorithm as shown in Eq.1 so as to obtain two low-rank latent factor matrices $\boldsymbol{U}_S$ and $\boldsymbol{V}_S$

Step 1.2. Cluster $\boldsymbol{U}_S$ and $\boldsymbol{V}_S$ to obtain user-cluster membership matrix $\boldsymbol{P}_S$ and item-cluster membership matrix $\boldsymbol{Q}_S$, based on which the rating pattern $\boldsymbol{B}$ is constructed according to Eq.4

Step 2. Cluster membership mapping

Step 2.1. Map each user and item in $\boldsymbol{X}_T$ to the cluster identified in $\boldsymbol{B}$ using Eq.5

Step 3. FM-based rating pattern integration

Step 3.1. Incorporate rating pattern $\boldsymbol{B}$ into the target matrix $\boldsymbol{X}_T$ via expanding design matrix as shown in Eq.8

Step 3.2. Factorize design matrix using factorization machines as shown in Eq.3

Fig. 3. The algorithm of LKT-FM (low-rank knowledge transfer via factorization machines)

We depict the above three steps of *low-rank rating pattern construction*, *cluster membership mapping* and *FM-based rating pattern integration* in Fig.3, which contains five components of *dimension reduction*, *clustering*, *mapping*, *incorporation* and *factorization*. In specific, we first employ dimension reduction technique to obtain the low-rank user-factor and item-factor matrices, which are then used by clustering algorithms to extract a group-level rating matrix. After that, we map each user and item into its corresponding cluster. In the end, we transfer the rating pattern via incorporating it into the design matrix and factorize the expanded design matrix using factorization machines.

Note that our algorithm is generic and flexible because we may find alternatives to each of these five components to derive a new solution according to the requirements of real-world applications. For example, we may employ another matrix factorization algorithm rather than basic MF, or another clustering algorithm rather than the simple K-means.

4 Experiments

4.1 Experiment Goal

The goal of our experiments is to answer the following research questions:

- **RQ1 Evaluation of Knowledge Extraction**: Can the proposed low-rank knowledge transfer solution (LKT) extract higher quality rating pattern than the traditional co-clustering method?
- **RQ2 Evaluation of Knowledge Integration**: Can factorization machines (FM) better integrate the rating pattern than traditional Direct-Expansion (DE) methods?
- **RQ3 Overall Performance**: Overall, how is the performance of our LKT-FM model compared with state-of-the-art NOCDCF techniques?
- **RQ4 Performance in Overlapping Setting**: How is the performance of our LKT-FM model compared with ordinary CDCF techniques, when users/items are overlapped across domains?

Note that RQ1, RQ2, RQ3 aim at evaluating our model in a NOCDCF setting from different aspects, while RQ4 aims at exploring its generality. We all know that NOCDCF has much wider range of applications in real world. And this is in fact the major motivation of this paper to solve this problem. But we are also interested in evaluating our model on overlapping CDCF datasets. If our model performs better than current NOCDCF approaches and in the meanwhile is competitive with existing Overlapping-CDCF methods, we can safely say our model is a generic solution to all CDCF tasks. And that is the reason why we evaluate RQ4. In the following sub-sections, we will describe datasets, baselines and experiment setups for these two settings respectively.

4.2 Datasets

For RQ1, RQ2, RQ3, we use three benchmark real-world datasets for evaluation:

- MovieLens 1M dataset[1]: A movie rating dataset contains 1,000,209 ratings of 3,900 movies made by 6,040 users (rating ratio 4.2%). Since we want to explore the effect of the sparseness of source matrix on the performance of NOCDCF algorithms, besides using the original dataset, we also use three sub-matrices with different sparseness as source matrix: 224,745 ratings by 500 users on 1000 movies (rating ratio 44.9%); 476,409 ratings by 1000 users on 2000 movies (rating ratio 23.8%); 634,680 ratings by1500 users on 3000 movies (rating ratio 14.1%). We use the same sub-matrix extraction method in [5][6].

[1] http://www.grouplens.org/node/73

– Book-Crossing dataset[2]: A book rating data set contains more than 1.1 million ratings (scales 0-9) by 278,858 users on 271,379 books. Following [5][6], we obtain target matrix by randomly choosing 500 users with at least 20 ratings, and 1000 movies (rating ratio 3.03%).
– EachMovie dataset[3]: A movie rating dataset contains 2.8 million ratings (scales 1-6) by 72,916 users on 1682 movies. We still randomly choose 500 users with at least 20 ratings, and 1000 movies for experiment (rating ratio 12.4%).

Note that, in order for rating scale consistency, we normalize the rating scales from 1 to 5 for Book-Crossing and EachMovie dataset.

For RQ4, we use Amazon dataset for evaluation:

– Amazon dataset[20]: A diverse product rating data set contains 7,593,243 ratings provided by 1,555,170 users over 548,552 different products including 393,558 books, 103,144 music CDs, 19,828 DVDs and 26,132 VHS video tapes.

4.3 Baseline

For RQ1, RQ2, RQ3, we consider three state-of-the-art techniques as baselines: Codebook Transfer (CBT) [5], Cluster-Level Latent Factor Model (CLFM) [6] and Factorization Machines (FM) [8]. Note that, this work focuses on solving the single-domain knowledge-transfer problem [1], in which only a target domain and a source domain are considered. There are also some models designed for multi-domain knowledge-transfer problem, such as TALMUD [10], which however are identical to CBT when only two domains are involved. Therefore, we didn't adopt these methods for evaluation.

– CBT is a widely used RPT method in previous works [6][10]. It adopts orthogonal nonnegative matrix tri-factorization (ONMTF) algorithm [4] for codebook (i.e. rating pattern) extraction and is also one of the first approaches that can handle NOCDCF problem.
– CLFM is a more recent work that extends the idea of CBT. It not only learns a common rating pattern shared by all domains, but also learns domain-specific rating pattern for each individual domain.
– FM is a generic predictive model for single domain recommendations. To the best of our knowledge, no previous work adopted FM as a baseline algorithm for NOCDCF problem. This may be because cross-domain approaches were deemed to perform better than single domain approaches. However, cross-domain recommendation may easily have negative transfer issue. Thus, including a prominent single-domain CF approach is a necessity for the evaluation of CDCF algorithms. Thus in this work, we adopt FM as a baseline to explore whether NOCDCF methods can really improve the recommendation.

[2] http://www.informatik.uni-freiburg.de/?cziegler/BX/
[3] http://www.cs.cmu.edu/?lebanon/IR-lab.htm

For RQ4, we consider two state-of-the-art CDCF methods as baselines: PF2-CDTF [18] and CDFM [15].

- PF2-CDTF is a tensor-based factorization model that can capture triadic relation between users, items and domains to improve CDCF recommendation.
- CDFM is a Factorization-Machine-based CDCF method that incorporates different domains by expanding the design matrix of FM. This method is relevant to our model but requires user correspondence information as additional input, which makes it only applicable in the Overlapping-CDCF setting.

4.4 Experimental Setup

For RQ1, RQ2, RQ3, we choose MovieLens as source domain, Book-Crossing and EachMovie as target domains. Following the work in [6], we evaluate our method under different configurations. For each target dataset, 300 users are randomly selected as the training set, and the remaining users for testing. For each test user, three different sizes of observed rating (Given5, Given10, Given15) are provided to avoid cold-start and the remaining ratings are for evaluation. Note that, as mentioned in Section 4.1, we obtained three sub-matrices of MovieLens 1M dataset with different levels of sparseness. In our experiments, all the three sub-matrices were used as source matrix, besides the original one, in order to explore the impact of data sparseness.

For RQ4, we choose Amazon-Music and Amazon-Book as target domain, the rests as source domains. We build the training and test set in two different ways similar to [15][18] to allow comparison with them. In the first setup, TR75, 75% of data is considered as training set and the rest as test set, and in the second setup, TR20, only 20% of data is considered as training set and the rest as test set.

We adopt Mean Absolute Error (MAE) as evaluation metric. MAE is computed as $\mathrm{MAE} = (\varSigma_{i \in T} |r_i - \widehat{r}_i|)/|T|$, where T means the test set, r_i is true value and $\widehat{r}_i$ is the predicted rating.

4.5 Experimental Results

RQ1 Evaluation of Knowledge Extraction
Since the performance of recommendation depends on both knowledge construction and integration methods, to evaluate the quality of knowledge extraction method, we need to make sure all the model adopt the same integration method. Thus, we use Direct-Expansion (DE) instead of Factorization Machines (FM) for our model in this part. We re-denoted our model as LKT-DE and compare it with two baselines CBT and CLFM.

Note that, in Fig.4, "ML (224,745, 44.9%)" denotes that the source is a subset of MovieLens, containing 224,745 ratings and the rating ratio is 44.9%. "BX-5" denotes the target is Book-Crossing dataset and 5 ratings are given for each test user. Similarly, "EM-10" denotes the target is EachMovie dataset and 10 ratings are given for each test user. We use these denotations in the rest of the paper.

From Fig.4 we can see that our method LKT-DE outperforms CBT and CLFM all the time, which means our knowledge extraction method behaves better than the co-clustering methods used by baselines. It is interesting to see that, in general, as the source matrix ML becomes sparser, the performances of two baselines degrade (i.e. MAE becomes larger). For example, in the first chart of Fig.4, the blue line, representing baseline method CBT, goes up from 0.56 to 0.63 while the rating radio of source matrix decrease from 44.9% to 4.3%. This supports the analysis in Section.1 that current NOCDCF methods do not work well in sparse settings.

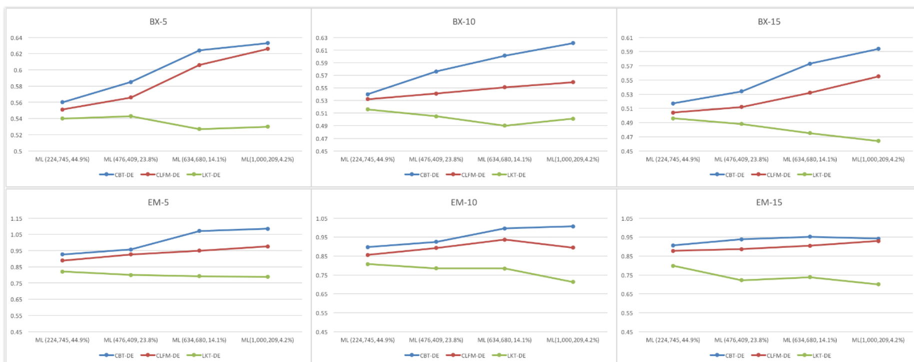

Fig. 4. MAE comparison of different knowledge integration methods

On the contrary, our method is not affected by the sparseness. As the source matrix becomes sparser, the performance becomes even better. This might be because as the source matrix ML becomes sparser, number of ratings in it also increases (from 224,745 to 1,000,209). Thus, although the sparseness makes it harder to extract knowledge for baseline methods, our method takes advantage of the increased ratings, and thus extract even more knowledge from them.

RQ2 Evaluation of Knowledge Integration

Before comparing our FM-based integration method with the traditional Direct-Expansion (DE) approach, we first compare the seven different expansions of design matrix discussed in Section3.3. Similar to RQ1, we make sure all the integration methods, including our seven solutions and the baseline DE, use the same rating pattern constructed by LKT.

Fig.5 shows the result on EM (EachMovie) dataset using ML (476,409, 23.8%) as source matrix. All the other seven lines are below the black line representing FM, which only uses target matrix for recommendation. This indicates that all the three types of information, item cluster index, user cluster index and co-cluster rating in source domain are useful for enhancing recommendation tasks in the target. Among the seven, FM(U+R+I) which includes all the three types of information U, R, I, performs the best.

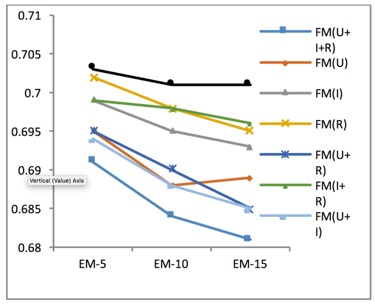

Fig. 5. MAE comparison of different design matrix expansions when source is ML (476,409, 23.8%) and target is EM

Note that, in our future work, we are planning to improve the current approach by allowing automatic selection of feature expansion for FM, but in this work, we choose the simple but effective FM(U+R+I) expansion for further comparisons.

To save space, we don't show the comparison results when using other two source matrices, ML (224,745, 44.9%) and ML (634,680, 14.1%). We report the final results in the following.

Table 1 shows the comparison of our FM-based method with the baseline DE, which is now re-denoted as LKT-DE.

Table 1. MAE comparison of different rating pattern incorporation methods

	ML(224,745 44.9%)		ML (476,409 23.8%)		ML (634,680 14.1%)		ML (1,000,209 4.2%)	
	LKT-DE	LKT-FM	LKT-DE	LKT-FM	LKT-DE	LKT-FM	LKT-DE	LKT-FM
BX-5	0.540	**0.527**	0.543	**0.526**	0.527	**0.503**	0.530	**0.510**
BX-10	0.516	**0.494**	0.505	**0.480**	0.490	**0.468**	0.501	**0.475**
BX-15	0.496	**0.470**	0.488	**0.461**	0.475	**0.449**	0.464	**0.431**
EM-5	0.821	**0.705**	0.800	**0.701**	0.792	**0.695**	0.788	**0.683**
EM-10	0.807	**0.696**	0.785	**0.684**	0.774	**0.669**	0.763	**0.662**
EM-15	0.798	**0.694**	0.775	**0.681**	0.768	**0.672**	0.744	**0.660**

It is clear that our method always performs better than the baseline DE. We also find that the improvement on EM dataset is prominent. It may be because EM and ML are both movie-rating dataset while BX is book-rating dataset. The relatedness between ML and EM is higher than that between ML and BX. Thus, more useful knowledge is transferred from ML to EM and thus improves recommendation accuracy more.

RQ3 Overall Performance
To answer this question, three baselines CBT, CLFM and FM are compared with the proposed LKT-FM method. Note that, FM generates the prediction only with target-domain data. Thus, the MAE values for FM do not change when the source matrix changes.

Table 2. Table 2. MAE comparison of different NOCDCF methods

	ML (224,745, 44.9%)				ML (476,409, 23.8%)			
	CBT	CLFM	FM	LKT-FM	CBT	CLFM	FM	LKT-FM
BX-5	0.560	0.551	0.541	**0.527**	0.585	0.566	0.541	**0.526**
BX-10	0.540	0.532	0.508	**0.494**	0.576	0.541	0.508	**0.480**
BX-15	0.517	0.504	0.491	**0.470**	0.534	0.512	0.491	**0.461**
EM-5	0.927	0.889	**0.703**	**0.705**	0.957	0.927	0.703	**0.691**
EM-10	0.897	0.856	0.701	**0.696**	0.925	0.893	0.701	**0.684**
EM-15	0.906	0.877	0.701	**0.694**	0.938	0.886	0.701	**0.681**
	ML (634,680, 14.1%)				ML(1,000,209, 4.2%)			
	CBT	CLFM	FM	LKT-FM	CBT	CLFM	FM	LKT-FM
BX-5	0.624	0.606	0.541	**0.503**	0.633	0.626	0.541	**0.510**
BX-10	0.601	0.551	0.508	**0.468**	0.621	0.559	0.508	**0.475**
BX-15	0.573	0.532	0.491	**0.449**	0.534	0.525	0.491	**0.431**
EM-5	1.071	0.950	0.703	**0.695**	0.998	0.901	0.703	**0.683**
EM-10	0.996	0.937	0.701	**0.669**	0.957	0.894	0.701	**0.662**
EM-15	0.952	0.904	0.701	**0.672**	0.942	0.929	0.701	**0.660**

The experimental results are shown in Table 2. We can see that the proposed method outperforms all the other models. It is very surprising to find that FM behaves better than the two state-of-the-art NOCDCF methods (CBT and CLFM), especially on the EachMovie dataset. This indicates that existing NOCDCF methods have negative transfer issue in this case. As analyzed in Section.1, existing NOCDCF methods incorporate rating pattern via direct expansion, which may rely too excessively on the knowledge from source domain and miss knowledge in the target domain itself. When the loss of knowledge from target domain exceeds the gain from source domain, current NOCDCF methods become even less effective than single-domain algorithms. On the contrary, our model overcomes this by striking a good balance between getting knowledge from the source and the target, and thus outperforms both current single-domain (FM) and cross-domain algorithms(CBT, CLFM) .

RQ4 Performance in Overlapping Setting
Fig.6 and Fig.7 show the comparison of different CDCF methods on Amazon dataset. In these two figures we can see that CDFM, which is a FM-based CDCF method, outperforms other methods, including our LKT-FM. This is not sur-

prising because both LKT-FM and CDFM adopt FM for knowledge integration, while CDFM also utilizes user correspondence information as addition input. But except for CDFM, we can see that our LKT-FM outperforms other CDCF methods including the prominent PRE2-CDTF algorithm. Note that PRE2-CDTF performs even worse than FM when 20% of data is used as training set (TR20), which means negative transfer happens. On the contrary, our LKT-FM outperforms FM consistently. These results indicate that our LKT-FM is competitive with other methods even in the overlapping setting. Thus, we can argue that our LTK-FM model is an appropriate approach for solving various CDCF tasks, not limited to NOCDCF problems.

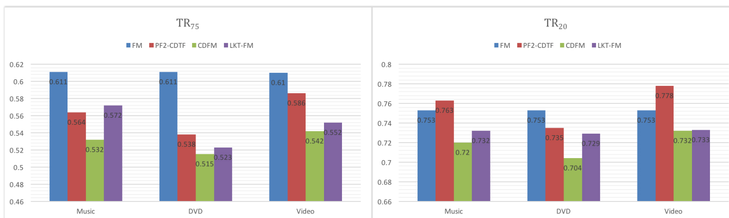

Fig. 6. MAE comparison of different CDCF methods on Amazon dataset (target: Book)

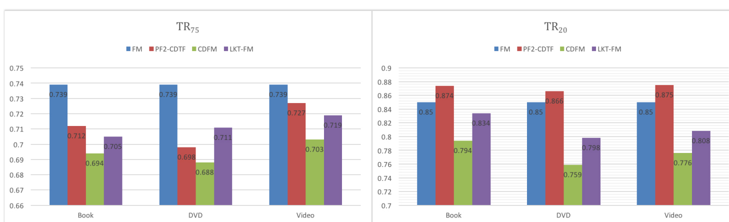

Fig. 7. MAE comparison of different CDCF methods on Amazon dataset (target: Music)

5 Conclusion and Future Work

In this paper we presented LKT-FM, a novel rating pattern transfer model, which aims at addressing the Non-Overlapping CDCF (NOCDCF) problem. The proposed model consists of 3 components: 1) a low-rank clustering method that enables knowledge extraction on large and sparse matrices; 2) a membership mapping algorithm which assigns users and items into clusters identified in the rating pattern; 3) a FM-based knowledge integration method that incorporates the shared knowledge into the target data in a seamless manner. Our experimental results showed LKT-FM outperforms state-of-the-art single-domain and NOCDCF approaches. In addition, LKT-FM is competitive with ordinary CDCF methods even in overlapping settings, which makes it a generic solution to all CDCF tasks.

References

1. Fernndez-Tobas, I., Cantador, I., Kaminskas, M., Ricci, F.: Cross-domain Recommender Systems: A Survey of the State of the Art. CERI, 187-198 (2012)
2. Berkovsky, S., Kuflik, T. & Ricci, F. : Cross-Domain Mediation in Collaborative Filtering. User Modeling, 355–359 (2007)
3. Tang, J., Wu, S., Sun, J. & Su, H. : Cross-domain collaboration recommendation. KDD, 1285-1293 (2012)
4. Ding, C. H. Q., Li, T., Peng, W. & Park, H. : Orthogonal nonnegative matrix t-factorizations for clustering. KDD, 126-135 (2006)
5. Li, B., Yang, Q. & Xue, X. : Can Movies and Books Collaborate? Cross-Domain Collaborative Filtering for Sparsity Reduction. IJCAI, 2052-2057 (2009)
6. Gao, S., Luo, H., Chen, D., Li, S., Gallinari, P. & Guo, J. : Cross-Domain Recommendation via Cluster-Level Latent Factor Model. ECML/PKDD, 161-176 (2013)
7. Koren, Y., Bell, R. & Volinsky, C. : Matrix Factorization Techniques for Recommender Systems. Computer, 42, 30-37 (2009)
8. Rendle, S. : Factorization Machines with libFM. ACM Trans. Intell. Syst. Technol., 3, 57:1–57:22. (2012)
9. Li, B., Yang, Q. & Xue, X. : Transfer learning for collaborative filtering via a rating-matrix generative model. ICML, 617-624 (2009)
10. Moreno, O., Shapira, B., Rokach, L. & Shani, G. : TALMUD: transfer learning for multiple domains. CIKM, 425-434 (2012)
11. Koren, Y. : Factorization meets the neighborhood: a multifaceted collaborative filtering model. KDD, 426-434 (2008)
12. Hu, Y., Koren, Y. & Volinsky, C. : Collaborative Filtering for Implicit Feedback Datasets. ICDM, 263-272 (2008)
13. Koren, Y. : Collaborative filtering with temporal dynamics. KDD, 426-434 (2009)
14. Ji, K., Sun, R., Li, X. & Shu, W. : Improving matrix approximation for recommendation via a clustering-based reconstructive method. Neurocomputing, 173, 912-920 (2016)
15. Loni, B., Shi, Y., Larson, M. & Hanjalic, A. : Cross-Domain Collaborative Filtering with Factorization Machines. ECIR, 656-661 (2014)
16. Shi, Y., Larson, M. & Hanjalic, A. : Tags as Bridges between Domains: Improving Recommendation with Tag-Induced Cross-Domain .Collaborative Filtering. UMAP, 305-316 (2011)
17. Coyle, M. & Smyth, B. : (Web Search)shared: Social Aspects of a Collaborative, Community-Based Search Network. AH, 103-112 (2008)
18. Hu, L., Cao, J., Xu, G., Cao, L., Gu, Z. & Zhu, C. : Personalized recommendation via cross-domain triadic factorization. WWW, 595-606 (2013)
19. Cremonesi, P. & Quadrana, M. : Cross-domain recommendations without overlapping data: myth or reality? RecSys , 297-230 (2014)
20. Leskovec, J., Adamic, L. A. & Huberman, B. A. : The dynamics of viral marketing. ACM Trans. Web, 1, 5 (2007)