# Cross-Domain Collaborative Filtering with Factorization Machines

Babak Loni, Yue Shi, Martha Larson, Alan Hanjalic

Delft Multimedia Information Retrieval (D-MIR) Lab
Delft University of Technology, Netherlands
{b.loni,y.shi,m.a.larson,a.hanjalic}@tudelft.nl

**Abstract.** Factorization machines offer an advantage over other existing collaborative filtering approaches to recommendation. They make it possible to work with any auxiliary information that can be encoded as a real-valued feature vector as a supplement to the information in the user-item matrix. We build on the assumption that different patterns characterize the way that users interact with (i.e., rate or download) items of a certain type (e.g., movies or books). We view interactions with a specific type of item as constituting a particular domain and allow interaction information from an auxiliary domain to inform recommendation in a target domain. Our proposed approach is tested on a data set from Amazon and compared with a state-of-the-art approach that has been proposed for Cross-Domain Collaborative Filtering. Experimental results demonstrate that our approach, which has a lower computational complexity, is able to achieve performance improvements.

## 1 Introduction

Cross-domain Collaborative Filtering (CDCF) methods exploit knowledge from *auxiliary* domains (e.g., movies) containing additional user preference data to improve recommendation on a target domain (e.g. books). While relying on a broad scope of existing data in many cases is a key to relieving the problems of sparse user-item data in the target domain, CDCF can also simultaneously benefits different data owners by improving quality of service in different domains.

In most CDCF approaches (e.g., [4], [5]) it is assumed that user behavior in all domains is the same. This assumption is not always true since each user might have different domains of interest, for example, rating items consistently more frequently or higher in one domain than in another. In a recent work, Hu et al. [2] argue that CDCF should consider the full *triadic* relation *user-item-domain* to effectively exploit user preferences on items within different domains. They represent the user-item-domain interaction with a tensor of order three and adopt a tensor factorization model to factorize users, items and domains into latent feature vectors. The rating of a user for an item in a domain is calculated by element-wise product of user, item and domain latent factors. A major problem of tensor factorization however, is that the time complexity of this approach is exponential as it is $O(k^m)$ where $k$ is the number of factors and $m$ is the number of domains.

In this paper we exploit the insight that user preferences across domains could be deployed more effectively if they are modeled separately on separate domains, and then integrated to generate a recommendation on the target domain. We therefore address the problem with *factorization machines* (FM) [6], which make such modeling possible. In addition, the FMs are more flexible than the tensor representation regarding the ways of capturing the domain-specific user preferences and could lead to more reliable recommendations. Finally, FMs are polynomial in terms of $k$ and $m$, making them computationally less expensive than tensor factorization models [6].

FMs have already been applied to carry out CF in a single domain, [6, 7], but have yet to be exploited to address the CDCF problem. Here we apply FMs to cross-domain recommendation in a way that allows them to incorporate user interaction patterns that are specific to particular types of items. Note that in this work, we define a domain as a type of item. The set of users is not mutually exclusive between domains, but we assume that their interaction patterns differ sufficiently to make it advantageous to model domains separately. The novel contribution of our work is to propose an extension of FMs that incorporates domains in this pattern and to demonstrate its superiority to single domain approaches and to a state-of-the-art CDCF algorithm.

## 2 Related Work

**Cross-Domain Collaborative Filtering**: An overview of CDCF approaches is available in Li [9]. Here, we restrict our discussion of related CDCF approaches to mentioning the advantages of our approach compared to the major classes of existing algorithms. *Rating pattern sharing* algorithms, exemplified by [4], groups users and items into clusters and matches cluster-level rating patterns across domains. The success of the approach depends, however, on the level of sparseness of user-item information per domain. *Latent feature sharing* approaches, exemplified by [5], transfer knowledge between domains via a common latent space and are difficult to apply when more than two domains are involved [2]. *Domain correlation* approaches, exemplified by [8], use common information (e.g., user tags) to link domains and fail when such information is lacking.

**Factorization Machines**: Factorization Machines (FM) [6] are general models that factorize user-item collaborative data into real valued feature vectors. Most factorization models such as Matrix Factorization can be modeled as a special case of FM [6]. Despite typical CF models where collaboration between users and items are represented by a rating matrix, in factorization machines the interaction between user and item is represented by a feature vector and the rating is considered as class label for this vector. More specifically lets assume that the data of a rating prediction problem is represented by a set $S$ of tuples $(\mathbf{x}, y)$ where $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is a $n$-dimensional feature vector and $y$ is its corresponding label. Factorization machines model all interactions between features using factorized interaction parameters. In this work we adopted a FM model with order $d = 2$ where only *pairwise* interaction between features are considered. This model can be represented as follows:

$$\hat{y}(\mathbf{x}) = w_0 + \sum_{j=1}^{n} w_j x_j + \sum_{j=1}^{n} \sum_{j'=j+1}^{n} w_{j,j'} x_j x_{j'} \tag{1}$$

where $w_j$ are model parameters and $w_{j,j'}$ are factorized interaction parameters and are defined as $w_{j,j'} = \mathbf{v}_j.\mathbf{v}_{j'}$ where $\mathbf{v}_j$ is $k$-dimensional factorized vector for feature $j$. For a FM with $n$ as the dimensionality of feature vectors and $k$ as the dimensionality of factorization, the model parameters that need to be learnt are $\Theta = \{w_0, w_1, \ldots, w_n, v_{1,1}, \ldots, v_{n,k}\}$. Three learning approaches have been proposed to learn FMs [6]: Stochastic Gradient Descent (SGD), Alternating Least-Squares (ALS) and Markov Chain Monte Carlo (MCMC) method. We exploit all 3 methods in this work.

## 3 Cross-Domain CF with Factorization Machines

Assume we are given collaborative data of users and items in $m$ different domains $\{\mathbb{D}_1, \ldots, \mathbb{D}_m\}$. The domains are different based on the type of items that exists in them. While rating information for a user might be very sparse in one domain (e.g. Books), he might have rich collaborative data in another domain (e.g. movies). The purpose of cross-domain CF is to transfer knowledge from different auxiliary domains to a target domains to improve rating predictions in the target domain.

To understand our approach, without loss of generality lets assume $\mathbb{D}_1$ is the target domain and $\{\mathbb{D}_2, \ldots, \mathbb{D}_m\}$ are the auxiliary domains. Also consider $U_j$ and $I_j$ as the set of users and items in domain $\mathbb{D}_j$. The standard rating prediction problem in the target domain $\mathbb{D}_1$ can be modeled by a target function $y : U_1 \times I_1 \rightarrow \mathbb{R}$. We represent each user-item interaction $(u, i) \in U_1 \times I_1$ with a feature vector $\mathbf{x} \in \mathbb{R}^{|U_1|+|I_1|}$ with binary variables indicating which user rated which item. In other words, if user $u$ rated item $i$ the feature vector $\mathbf{x}$ is represented as:

$$\mathbf{x} = (\underbrace{0, \ldots, 0, 1, 0, \ldots, 0}_{|U_1|}, \underbrace{0, \ldots, 0, 1, 0, \ldots, 0}_{|I_1|}) \tag{2}$$

where non-zero elements are corresponding to user $u$ and item $i$. The feature vector $\mathbf{x}$ can also be represented by its sparse representation $\mathbf{x}(u, i) = \{(u, 1), (i, 1)\}$.

Given the feature vector $\mathbf{x}(u, i)$ in the target domain, our cross-domain CF approach extend this vector by adding collaborative information of user $u$ from other domains. Now lets assume that $s_j(u)$ represents all items in domain $\mathbb{D}_j$ which are rated by user $u$. For each auxiliary domain $\mathbb{D}_j$, $j = 2, \ldots, m$, our method extend $\mathbf{x}(u, i)$ with a vector $\mathbf{z}_j(u)$ with the following sparse representation:

$$\mathbf{z}_j(u) = \{(l, \phi_j(u, l)) : l \in s_j(u)\} \tag{3}$$

where $\phi_j(u,l)$ is a domain-dependent real valued function. We define $\phi_j$ based on the rating of user $u$ to item $l$ and normalize it based on total number of items which is rated by user $u$ in domain $\mathbb{D}_j$:

$$\phi_j(u,l) = \frac{r_j(u,l)}{|s_j(u)|} \tag{4}$$

where $r_j(u,l)$ specifies the rating of user $u$ to item $l$ in domain $\mathbb{D}_j$. In the above definition $\phi_j$ is a function of $r_j(u,l)$ which reflects rating patterns of user $u$ in different domains. Furthermore it is normalized by considering the number of items which are rated by user in an auxiliary domain. This means that if a user is a frequent rater in an auxiliary domain, the contribution of each single rated item in this domain would be less compared to a rated item in an auxiliary domain with smaller number of user ratings. The above definition of $\phi_j$ prevents the model to be overwhelmed by too much information from auxiliary domains. This is one of the main advantages of factorization machines, namely to allow control of the amount of knowledge that is transferred from auxiliary domains. Note that the function $\phi_j$ can be also defined in other forms to reflect contribution of various domains in different ways. Based on our experiments we found the above definition of $\phi_j$ simple yet effective to transfer knowledge from auxiliary domains. Given the above definitions, we can now represent the extended vector $\mathbf{x}$ with the following sparse form:

$$\mathbf{x}(u,i,s_2(u),\ldots,s_m(u)) = \{\underbrace{(u,1),(i,1)}_{\text{target knowledge}},\underbrace{\mathbf{z}_2(u),\ldots,\mathbf{z}_m(u)}_{\text{auxiliary knowledge}}\} \tag{5}$$

The above feature vector serves as the input into the FM model in Equation (1), while the output variable $y$ is the rating of user $u$ to item $i$ in the target domain. Based on our proposed feature expansion method, the FM will only need to focus on the users in the target domain, resulting in an improvement in terms of computational cost.

## 4 Experiments

We conducted our experiments on Amazon dataset [3] which consists of rating information of users in 4 different domains: books, music CDs, DVDs and video tapes. The dataset contains 7,593,243 ratings on the scale 1-5 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.

We build the training and test set in two different ways similar to [2] to be able to compare our approach with them. In the first setup, $TR_{75}$, 75% of data is considered as training set and the rest as test set, and in the second setup, $TR_{20}$, only 20% of data is considered as training set and the rest as test set.

We implemented a recommendation framework with C#[1] on top of two open source libraries for recommender systems: MyMediaLite [1] which implements

[1] https://github.com/babakx/delft-recommendation-framework

**Table 1.** Comparison of different CF methods on the Amazon dataset

| Method \ Setup | $TR_{75}$ | | $TR_{20}$ | |
|---|---|---|---|---|
| **Target: Book** | MAE | RMSE | MAE | RMSE |
| MF-SGD (Book) | 0.62 | 0.86 | 0.89 | 1.14 |
| FM-SGD (Book) | 0.69 | 0.92 | 0.74 | 0.96 |
| FM-ALS (Book) | 0.72 | 0.99 | 0.75 | 1.07 |
| FM-MCMC (Book) | 0.60 | 0.79 | 0.72 | 0.94 |
| FM-All-MCMC (Book) | 0.60 | 0.79 | 0.76 | 0.99 |
| FM-MCMC (Book, {Music, DVD, Video}) | **0.46** | 0.64 | **0.69** | 0.92 |
| PF2-CDCF (Book, {Music, DVD, Video}) [2] | 0.50 | - | 0.76 | - |
| **Target: Music** | | | | |
| FM-MCMC (Music) | 0.71 | 0.95 | 0.77 | 1.00 |
| FM-MCMC (Music, {Book, DVD, Video}) | **0.67** | 0.91 | **0.74** | 0.98 |
| PF2-CDCF (Music, {Book, DVD, Video}) [2] | 0.70 | - | 0.82 | - |

most common CF approaches including Matrix Factorization, and LibFM [6] which implements FM learning algorithms. We first compared FMs with matrix factorization method on two different single domains and then we compare the results of our proposed method with the state-of-the-art CDCF work [2] on the same dataset. We also compare our method with a *blind* combination of all items from all domains to show that the improvement of our results is not only due to additional training data. We used mean absolute error (MAE) and root mean square error (RMSE) as evaluation metrics in our experiments. Table 1 lists the MAE and RMSE scores on the two different setups $TR_{75}$ and $TR_{20}$ and based on the following approaches:

- **MF-SGD ($D$)**: Matrix Factorization method using SGD learning algorithm on single domain $D$.
- **FM-X (D)**: Factorization Machine method on single domain D based on learning algorithm X (SGD, ALS or MCMC).
- **FM-All-X (D)**: Combining all rating data into single domain (blind combination) and testing target domain $D$ by using FM with algorithm X. This approach simply increases the size of training data by including the rating data of all domains. In other words, the feature vector $\mathbf{x}$ is represented as in equation (2) and all items in different domains are treated the same.
- **FM-X ($D_T, \{D_A\}$)**: Factorization Machine method on target domain $D_T$ and auxiliary domains $\{D_A\}$ based on algorithm X.
- **PF2-CDCF**: The Cross-Domain CF method which is proposed by Hu et al. [2] on the same dataset.

Comparison of results on single domains in table 1 shows that by using MCMC learning method, FM method performs better than matrix factorization. Comparison of FM-MCMC and FM-All-MCMC methods reveals that simply including the rating data of auxiliary domains into target domain does not cause any improvement on rating prediction and it can also hurt the result since the additional data can be noisy for the target domain. The best results, FM-MCMC

(Book, {Music, DVD,Video}) and FM-MCMC (Music, {Book, DVD,Video}), are obtained using our adopted cross-domain method with MCMC learning method and are better than PF2-CDCF on the same dataset.

## 5 Discussion and Future Directions

In this work we adopted a model using factorization machines to exploit additional knowledge from auxiliary domains to achieve performance improvement in cross-domain CF. The success of CDCF is highly dependent on effectively transferring knowledge from auxiliary domains, which can be well exploited with FMs. A key factor of success of our approach is the ability to encode domain-specific knowledge in terms of real-valued feature vector, which became possible with FMs and which enables better exploitation of the interaction patterns in auxiliary domains. The experimental results show that our adopted method can perform better than state-of-the-art CDCF methods while it benefits from low computational cost of FMs.

In the future, we want to apply our method to more complicated CDCF scenarios particularly when the source and target domains are more heterogeneous. Another extension to our approach is to also use *contextual* information from both target and auxiliary domains to investigate whether exploiting context can result in even better CDCF performance.

## Acknowledgments

This research is supported by funding from the European Commission's 7th Framework Program under grant agreement no. 610594 (CrowdRec).

## References

1. Z. Gantner, S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme. Mymedialite: a free recommender system library. RecSys '11, New York, NY, USA, 2011. ACM.
2. L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and C. Zhu. Personalized recommendation via cross-domain triadic factorization. WWW '13, 2013.
3. J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Trans. Web*, 1(1), May 2007.
4. B. Li, Q. Yang, and X. Xue. Can movies and books collaborate?: cross-domain collaborative filtering for sparsity reduction. IJCAI'09, 2009.
5. W. Pan, E. W. Xiang, N. N. Liu, and Q. Yang. Transfer learning in collaborative filtering for sparsity reduction. In *AAAI*. AAAI Press, 2010.
6. S. Rendle. Factorization machines with libfm. *ACM Trans. Intell. Syst. Technol.*, 3(3), May 2012.
7. S. Rendle, Z. Gantner, C. Freudenthaler, and L. Schmidt-Thieme. Fast context-aware recommendations with factorization machines. SIGIR '11. ACM, 2011.
8. Y. Shi, M. Larson, and A. Hanjalic. Tags as bridges between domains: improving recommendation with tag-induced cross-domain collaborative filtering. UMAP'11, 2011.
9. 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*, To appear.