Independent Study: Cross-Domain Recommendation: The Feasibility and the Value for the Cold Start Users

By: Shaghayegh Sahebi (Sherry)
Professor: Dr. Peter Brusilovsky

ABSTRACT

Most of the research studies on recommender systems are focused on single-domain recommendations. In these recommender systems, the domain of items used for training and the target item sets are within the same domain, like movies, books, etc. Cross-domain recommendations, or item recommendations in a multi-domain environment, are available in commercial systems like Amazon. But there are very few research studies on cross-domain recommender systems. Cross-domain recommendations can provide the relationship between two sets of items from various domains. They can provide extra information about the new users of a target domain (targeting the cold-start problem) and serendipity in recommendations. Hypothetically, they can also improve the quality of recommendations. In this paper, we first study the feasibility of cross-domain recommendations on movies and books domain using canonical correlation analysis (CCA), and then study both the cold-start problem and the hypothesis that cross-domain recommendations provide more accuracy. It is the first time that CCA is used for feasibility study of cross-domain recommendations.

1. INTRODUCTION

Cold-start [14] is a well-known issue in the area of recommender systems. Most seriously, cold-start affects recommendation approaches based on collaborative filtering that are fueled by the rating behavior of a community of users. For a new user who just entered the system, or who has not yet rated a sufficient number of items, collaborative filtering simply has too little information to reliably match him/her to the accumulated community wisdom and generate good recommendations. Similarly, collaborative filtering cannot effectively recommend new items that received no or too few ratings.

Both new-user and new-item problems were extensively discussed in recommender systems literature along with a number of possible solutions. This paper offers an extensive exploration of a specific approach to address a new-user problem: leveraging user rating data from a different domain. This approach is also known as cross-domain recommendation. We believe that the potential value of cross-domain recommendation is rapidly increasing in the modern social Web context. While just 10 years ago internet users had little chance to experience more than one recommender or similar rating-based system, nowadays the number and the diversity of these systems is increasing so rapidly that a typical internet user has ratings, votes, or bookmarks left in a number of different systems such as restaurants, hotels, movies, books, and other consumable products. In this context, cross-domain recommendation approaches could be very useful to offer a fast-start in a new domain by using the knowledge about user tastes and preferences accumulated in the other domains.

Despite this recognized potential, the number of reliable results obtained in the area of cross-domain recommendation is very small. While the problem of cross-domain recommendation was discussed in a good number of papers, very few papers reported empirical evaluation of reliable size. Worse, a good fraction of papers with empirical results was done in artificial context by subdividing a singe-domain dataset and considering that as separate domains e.g. separating the movie domain based on their genres [1,2,3]. As a result, there is still no extensive study of cross-domain recommendation in a cold-start context, there is even no reliable evidence that the cross-domain recommendation could, indeed, fulfill its promise as a cold-start helper.

This paper attempts to bridge this gap. Using a large dataset that has user ratings for two domains, books and movies, we perform an extensive exploration of the cross-domain recommendation in a cold start context. We start with exploring the very feasibility of cross-domain recommendation between this pair of domains. Using canonical correlation analysis, we investigate whether user ratings in these two domains have reasonable correlation. Without this correlation the information about user tastes from another domain will introduce more noise than help. After presenting the correlation data, we report the results of two experimental studies that explore the value of cross-domain recommendation as a function of the user profile size within and across the domains. The first study compares the accuracy of cross-domain and traditional recommendation for the cold start situation of difference severity (i.e. the number of rated items in the cold-start profile). The second study explores whether the size of the user profile in auxiliary (old) domain can also impact the quality of cross-domain recommendation in the target (new) domain.

2. BACKGROUND

2.1. Cross-Domain Recommendation

In recent years, cross-domain recommendation research is grown in popularity because of the broader availability of multi-domain user data. Social networks that include user connections and user interests in multiple domains are getting more and more users and data. In this context, cross-domain recommendation approaches can obtain more complete information about a user and improve both quality and serendipity of recommendations. For example, if a user only watches action movies but reads both action and horror books, his/her interest in horror movies can be captured by fusing the auxiliary book rating profiles into his/her movie rating profile. In addition, cross-domain recommendation can provide extra information in cold-start context mentioned in section 1. i.e., when a recommender system has insufficient information about new users in a target domain.

Although interest in cross-domain recommendations has been growing in recent years, there are very few research studies on this field. Berkovsky et al. [1, 2] and later Cremonesi et al. [3] offered an extensive discussion of cross-domain recommendation problems and suggested interesting generic approaches but were not able to explore these approaches in a true cross-domain context using instead artificial datasets produced by separation of single-domain user movie ratings into subdomains. Winoto and Tang analyzed the relatedness of music, movies, books, and computer games by categorizing them and studying the correlations between each two categories using an aggregating of all user ratings on each category [19]. The study, however, did not include an analysis using each user rating. Instead, the authors used different combinations of categories as candidate and target domain and studied the Mean Squared Error (MSE) resulting from performing collaborative filtering (CF) on these combinations. Jessenitschnig and Zanker [8] provided an architectural framework for a domain independent recommendation system without implementation or experimental results on real data.

Another group of projects explored the use of additional data sources to produce better cross-domain recommendations. Fernandez-Tobias et al. [5] suggested the use of semantic networks to integrate knowledge of different domains for cross-domain item recommendations. In [15], Shi et al. applied user cross-domain tags to perform cross-domain CF. In [13], Sahebi and Cohen used social links and cross-domain community detection approach to perform cross-domain recommendations.

Researchers in Machine Learning community explored transfer learning to perform cross-domain recommendations. Li et al. [9] attempted to solve the sparsity problem by compressing the ratings in one domain independent of users and then transferring the knowledge and reconstructing the target rating matrix. Wike et al. [12] also used transfer learning to solve the sparsity problem. They used two sets of data from the movie domain (MovieLens and Netflix) to run their experiments. In both of these studies, the authors do not assume a common user set for different domains.

2.2. Canonical Correlation Analysis

Canonical correlation analysis (CCA) is a multivariate statistical model that studies the interrelationships among sets of multiple dependent variables and multiple independent variables. It is the most generalized member of the family of multivariate statistical techniques [7]. Canonical correlation's goal is to find the strength of the relationship, between the sets of variables (independent and dependent). It is related to factor analysis in the sense that it creates composites of variables and is related to discriminant analysis in finding independent dimensions for each variable set. The goal of this analysis is to produce the maximum correlation between the dimensions. As a result, canonical correlation finds the optimum structure or dimensionality of each variable set that maximizes the relationship between independent and dependent variable sets.

In other words, if we have two multivariate random variable sets x and y including random variables x_i ($1 \leq i \leq m$) and y_j ($1 \leq j \leq n$), we want to analyze the correlation between sets x and y. For each of these variable sets, we have n corresponding samples. We can represent these samples as vectors: $X = (x_1, \ldots, x_m)$ and $Y = (y_1, \ldots, y_n)$. CCA considers two linear directions w_x and w_y with the same column rank such that inner product of w^T_x (transpose of w_x) with x and inner product of w^T_y with y be projections in the same dimension vector space. If we consider $S_{x,wx} = w^T_x.X$ and $S_{y,wy} = w^T_y.Y$, CCA tries to maximize the correlation between $S_{x,wx}$ and $S_{y,wy}$ by finding the proper w_x and w_y. The objective function (ρ) is presented in the following equation.

$$\rho = \max_{w_x, w_y} \frac{S^T_{\{x,w_x\}}.S_{y,w_y}}{\left|S_{x,w_x}\right|\left|S_{y,w_y}\right|} \tag{1}$$

Canonical correlation analysis has not been used in cross-domain recommendation systems but it has been used in different sources of the same domain. For example, in the area of recommender systems, Faridani used CCA to predict hotel ratings from textual comments of the hotels and their sentiment analysis [4]. Also, in [11], Ohkushi used Kernel CCA to find the relationship between music pieces and human motion to recommend music to users. In social media analysis, CCA has been used for multi-mode networks integration by Tang and Liu in [17]. Takagi et al. have introduced a semi-supervised version of CCA for automatic audio tag classification [16] and Udupa and Khapra used CCA to address the problem of transliteration equivalence [18].

3. THE DATASET

In this study we used an anonymized dataset obtained from an online Russian social system called imhonet. Imhonet is relatively unique in several aspects including its diverse nature. It allows users to rate and review a range of items from books and movies to mobile phones and architectural monuments. This system also contains many aspects of a social network, including friendships, blogs and comments. We have access to a dataset that includes two richest sets of ratings - on books and movies. Each rating record in the dataset includes a user ID, an item ID, and a rating value between 0 (not rated) and 10. The same user ID indicates the same user across the sets of ratings. In total, the dataset includes about 16 million movie rating records on about 50,000 movies and more than 11.5 million book ratings records on about 195,000 available books in the dataset. Figure. 1 (a) shows the scale of the number of book ratings per user in log-log coordinates and Figure 1 (b) shows the number of ratings for each book. As the figure shows, the plot of number of user raters per book follows the usual power law distribution, but the plot of the number of book ratings per user doesn't follow a usual pattern, but looks like a combination of two distributions. This peculiar shape is produced by two interfaces for new users that imhonet offered at different times. One interface asked each new user to rate at least 20 books or movies to receive recommendations. Another interface allowed exploring the system right away adding ratings one by one.

Ratings on Movies (Showing the Number of Users Having K Number of Movies Rated).

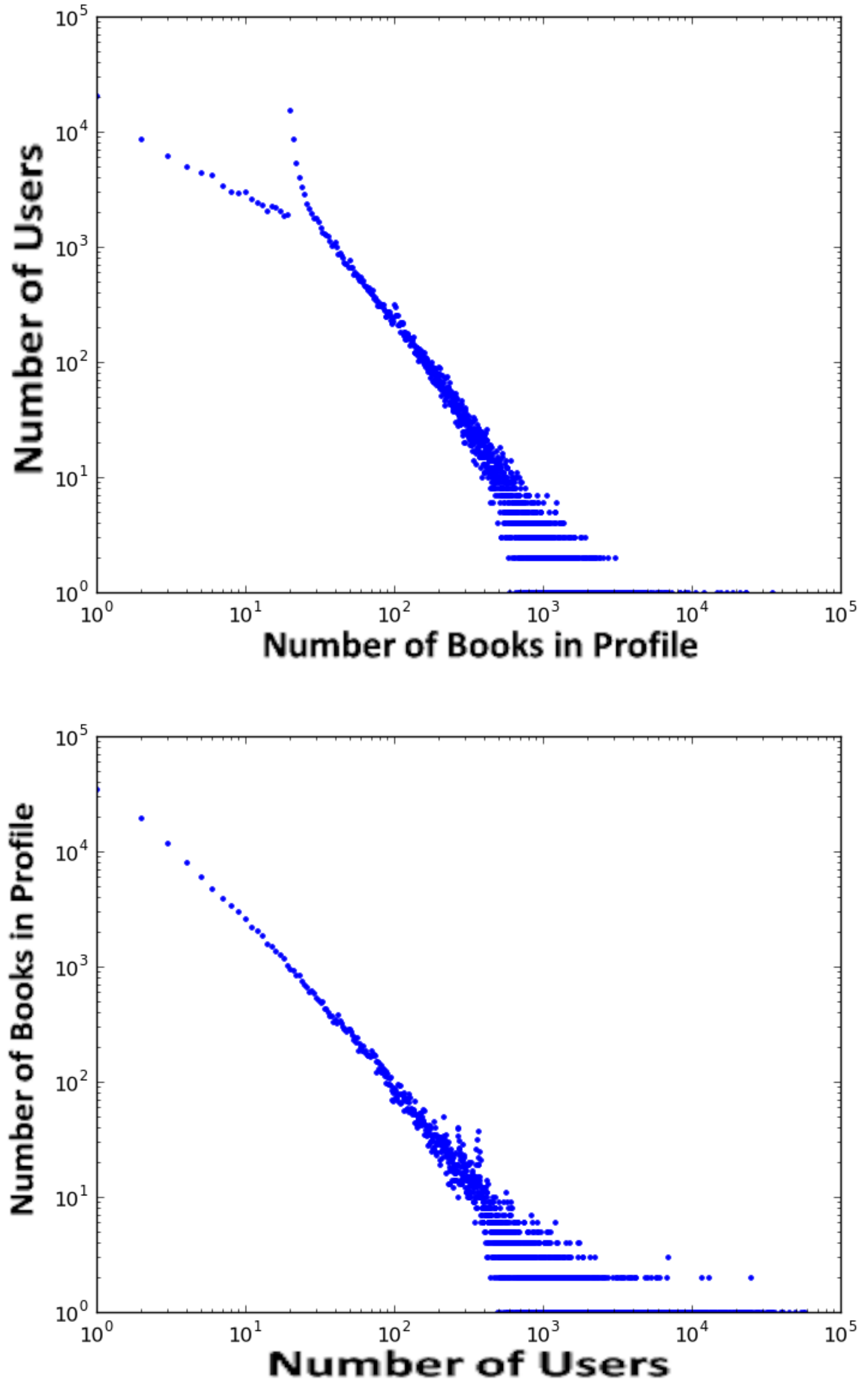
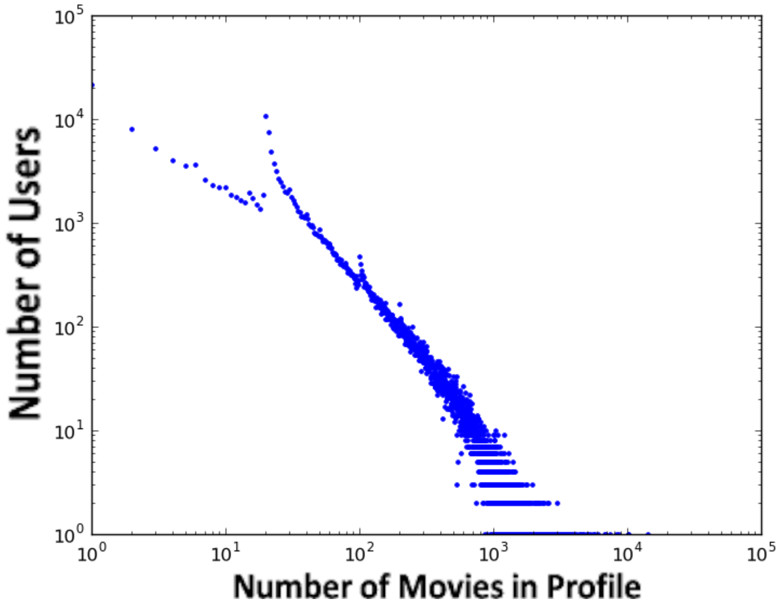


Figure 1 a) Log-Log Scale of the Number of Book Ratings of Users (Showing the Number of Users Having K Number of Books Rated); b) Log-Log Scale of the Number of Ratings on Books (Showing the Number of Books Having K Number of Users Rating them).



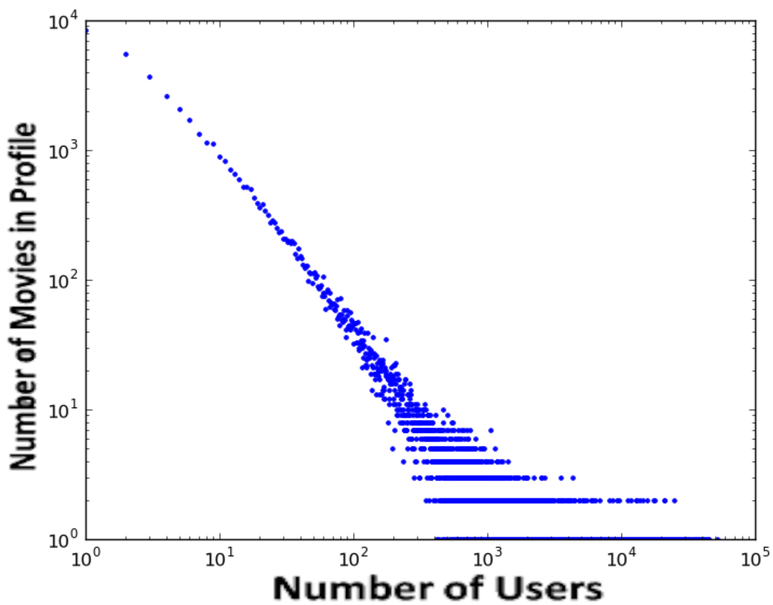


Figure 2 a) Log-Log Scale of the Number of Movie Ratings of Users (Showing the Number of Movies Having K Number of Users Rating them); b) Log-Log Scale of the Number of Ratings on Movies (Showing the Number of Users Having K Number of Movies Rated).

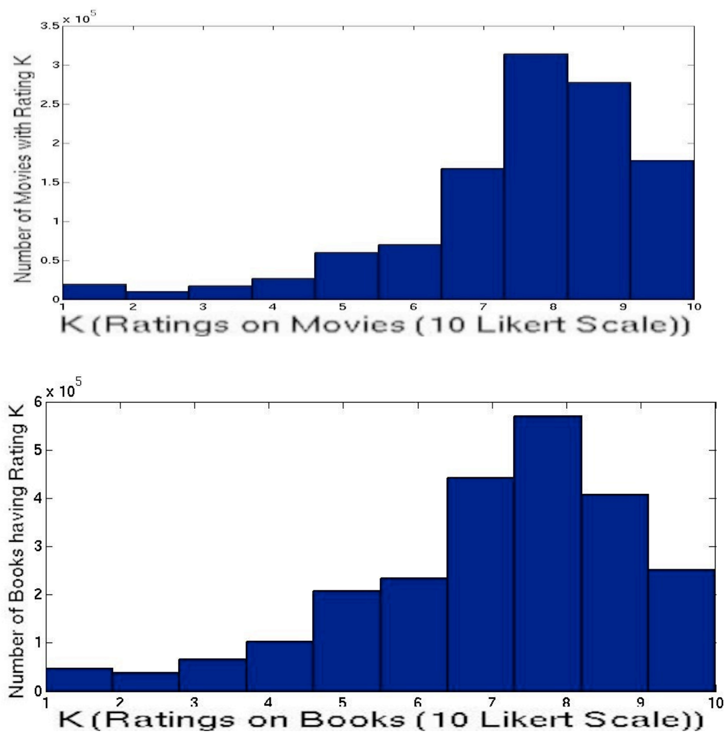


Figure 3 a) Histogram of Movie Ratings on all Movies, Showing how many Movies have Rating K; b) Histogram of Book Ratings on all Books, Showing how many Books have Rating K

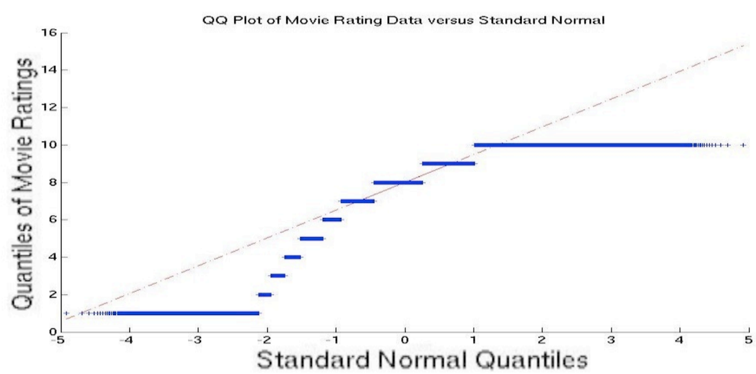


Figure 4 QQPlot of Movie Ratings (Showing how many Movies have Rating K) on all Movies.}

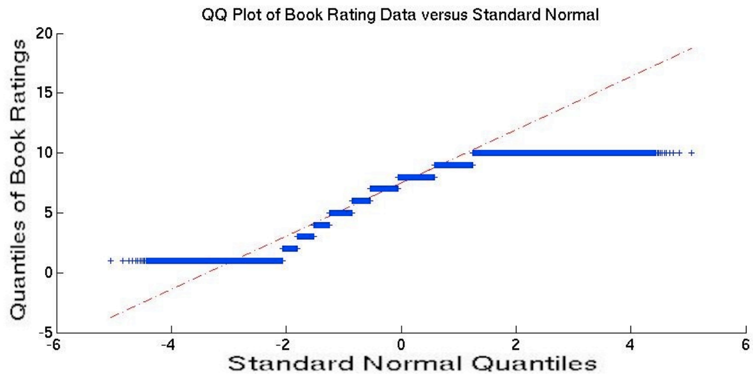


Figure 5 QQPlot of Book Ratings on all Books (Showing how many Books have Rating K).

If we look at log-log plot of movie rating distribution in Figure 2, we can see the same behavior: based on imhonet's request, many users rated at least 20 movies. The availability of many users who had at least 20 rated items in each of the used domain caused us to select this number as a basis of a smaller, but richer dataset that we used in our studies. To produce this dataset, we selected all ratings of users who had at least 20 ratings on books and 20 ratings on movies. The resulting dataset contains 2,363,394 movie ratings of 6,089 users on 12,621 movies and 1,138,401 book ratings of the same users on 17,907 books.
Figures 3 (a) and (b) show the histogram of movie and book ratings we have used in our experiments. They show how many movies and books had rating one to 10. As we can see, the histograms are negatively skewed, which is a common behavior for extended Likert scales. Figures 4 and 5 represents the quantile-quantile plot of books and movies ratings as a test of normality. We can see that none of the ratings that books or movies achieved are normal.

4. THE FEASIBILITY ASSESSMENT

Before running any experiments on cross-domain recommendations, we must make sure that there is a relationship between the utilized domains. In some cases, different domains might seem to be semantically closer to each other, like movies and songs domains; while in other cases, this relationship might not be that clear. To determine the relationship between books and movies domain in our dataset, we used canonical correlation analysis. As mentioned in section 2.2., CCA is a tool to evaluate latent linear correlations between two sets of variables. As a result, it fits our problem of finding relationships between two domains. By analyzing books and movies domains of our dataset using CCA, we can make sure that there is a correlation between these two domains and the results of our cross-domain recommendations are concluded from this relationship. If we cannot find any linear correlations, still there might be a non-linear relationship between the two domains but we will need more tests to capture that.

In this paper, we follow the analysis steps provided by chapter 8 of [6]. These steps are: defining objectives of canonical correlation, designing a canonical correlation analysis and testing the assumptions, deriving the canonical functions and assessing the overall fit, and interpreting the canonical variates. In the correlation between two sets of variables, some of those variables might be more important and effective than others. In interpreting the canonical variates, we should examine the canonical functions, to determine the relative importance of each of the original variables in deriving the canonical relationship. Since we have an anonymized dataset that only contains userIDs and IDs of rated items, we do not have any access to the content of rated items, such as title, genre, or description of them. We did this step and found the important variables with most canonical weights, but because we could not interpret the results more, due to the lack of data, we omitted this section from the paper. Other performed steps of analysis are listed in below sections. The terminology used in the following sections is described here:

- Canonical correlation: Measure of the strength of the overall relationships between the linear composites (canonical variates) for the independent and dependent variables.
- Canonical function: Relationship (correlational) be- tween two linear composites (canonical variates). Each canonical function has two canonical variates, one for the set of dependent variables and one for the set of independent variables.
- Canonical variates: Linear combinations that represent the weighted sum of two or more variables and can be defined for either dependent or independent variables.

4.1. Objectives of Canonical Correlation

We choose all the books and movies in the reduced dataset as input data. All of the 12621 movies are designated as the set of independent variables and all of the 17907 books are specified as the set of dependent variables. In practice, the canonical correlation is the same for both dependent and independent variates in the canonical function. As a result, we do not study the case in which books are independent variables and movies are dependent variables. The statistical problem is identifying any latent linear relationship between composites of variables: a customer's rating of books and his/her ratings on movies.

4.2. Designing a Canonical Correlation Analysis and Testing the Assumptions

The 12,621 movie variables result in a 0.482 to 1 ratio of observations to variables and the 17,907 book variables result in a 0.34 to 1 ratio of observations to variables. This ratio represents the number of observations we have for each variable in the dataset. The small value of this ratio happens because of the sparsity of our data, which is very common in recommendation systems. Sparsity of the data might result to over-fitting because there are not enough observations for each variable in the dataset. As a result, we admit that this small ratio might affect the statistical significance of the results because of over-fitting.

There are two statistical assumptions that are made for CCA. Since these assumptions are not very restrictive, they lead to generality of this approach. The first one is the assumption of linearity and the second one is desirable normality of the data. As for linearity, we assume that the correlation coefficient between any two variables is a linear relationship. If we couldn't find this linear relationship, we might transform one or both variables. Also, we assume that the relationship between variates is also linear.

The second assumption of CCA does not include any strict assumption of normality but normality of variables is recommended for significance of the results. In Figures 3 to 5 that show the histogram and quantile-quantile plots of movie and book ratings, we can see that the ratings do not follow a normal distribution which is recommended but is not necessary.

4.3. Deriving the Canonical Functions and Assessing Overall Fit

In this step, we run CCA and derive the relationships between the two sets of variables. The number of canonical functions derived (latent variants) is 5,953 variants. We look at the level of statistical significance and the practical significance of the canonical correlation to determine the number of canonical functions to include in the interpretation stage.

To examine the statistical significance, we test the canonical correlation for each of the 5,953 functions separately. The levels of canonical correlations vary between 0.1717 and 1 and 5,938 of these 5,953 functions show a significant correlation between the two sets of books and movies variables. In addition to this, we look at multivariate tests of all functions using Wilks' Lambda, Rao's approximate F-statistic and Bartlett's approximate chi-squared statistic. Based on Wilks' Lambda, 5,936 of the functions are having significant correlations at the 0.05 alpha level. Looking at Rao's approximate F-statistic and Bartlett's approximate chi-squared statistic, 5,898 of these functions are significantly related at the 0.05 alpha level. In addition to statistical significance, we look at the practical significance of the correlations. Usually, if the correlations are more than a specific value (e.g. 0.5) they are counted as practically significant. Our test shows that 5,938 of the canonical correlations are of sufficient size (> 0.5) to be considered practically significant.

As a result of these analyses, we can see that most of the derived canonical functions has the required level of significance. Thus, we can conclude that there is a correlation between the two domains of our study and based on that, we can trust our studies of cross-domain recommendations established on these domains. In other words, it shows for a user, his/her movie ratings is linearly correlated with the same user's book ratings and as a result, each of these ratings can be predicted based on the other one.

5. THE COLD START STUDIES

As we saw in the previous section, based on CCA results, the two domains of book rating and movie ratings are linearly correlated with each other. We think that this feasibility analysis, although not directly improving the recommendation results, is necessary before actually performing cross-domain recommendation. It convinces us that adding the source domain as an auxiliary profile is more than just adding some noise to the problem in hand. Being motivated by the results of feasibility studies, we can go ahead and perform cross-domain recommendations.

As explained in the introduction, one of the foci of this paper is an extensive study of cross-domain recommendation in the cold-start context. The idea of cross-domain recommendation in cold-start context is simple: if a user is new in a target domain (i.e., has few to none rated items), but has a rating history in another (auxiliary) domain, we can use his/her auxiliary domain profile to recommend relevant items, in the target domain.

The feasibility study demonstrated that cross-domain recommendations in the movies-books pair of domains is potentially possible, but didn't answer the question about the quality of such recommendations. This section presents the studies that we run to investigate the value and the quality of the cross-domain recommendation.

Our quality study pursues three goals: examining the value of using cross-domain (auxiliary) profiles in cold-start recommendations, examining the impact of the user profile size in the target domain on the comparative quality of ``cold-start'' recommendation, and examining the impact of the user profile size in the auxiliary domain on the quality of recommendations.

Our first goal is relatively simple: to find out whether cross-domain recommendations have any practical value. Since within-domain recommendation is considered to be a standard approach now, to achieve this goal we have to investigate whether cross-domain recommendations can beat this standard, i.e., to offer more accuracy than traditional within-domain recommendations in cold-start context.

A simple proof of value, however, is not sufficient to reveal the whole picture. Previous studies of recommendation approaches suggested for cold-start context demonstrated that the comparative quality of alternative approaches frequently depend on the number of items in the user profile. I.e., a specific cold-start approach could perform better when the number of items in the target domain user profile is very small, but after the profile grows over a threshold point, a regular approach surpasses it. This threshold point is sometimes considered as a cold start boundary, i.e., users with profiles that are smaller than this threshold are considered as cold start users. Following that, our second goal is to find the cold-start boundary for cross-domain recommendations. In other words, expecting that at some point the quality of within-domain recommendations may surpass the quality of the cross-domain recommendations, we want to see how many items the user should rate in the target domain to achieve best-quality recommendations without using ratings from auxiliary domain. To determine that we have to compare the quality of cross-domain and in-domain recommendations independently for each number of items rated in the target domain.

To achieve the first two goals, we run a comparative study of two recommendation approaches. As a baseline, we use a traditional collaborative filtering approach based on target-domain profiles. As a cross-domain approach, we use the most natural extension of the baseline approach to the cross-domain context: collaborative filtering based on concatenated profiles in auxiliary and target domains. First, we perform collaborative filtering on the target domain using cold-start profiles. Then, we do the same using both auxiliary and cold-start profiles by concatenating these two profiles together. We predict target user ratings on the target domain ratings using Equation. 2 in collaborative filtering. Here, n is the number of similar users we would like to take into account, α is a normalizer, $v_{i,j}$ is the vote of user i on item j, v_i is the average rating of user i and w(a,i) is the similarity weight of this n similar users. In cross-domain recommendations, this similarity weight is calculated based on a concatenation of auxiliary and target domain profiles of users; while in cold-start recommendations, it is only based on the cold-start profiles.

$$p_{a,j} = \overline{v_a} + \sum_{i=1}^{n} w(a,i)(v_{i,j} - \overline{v_i}) \qquad (2)$$

While the study introduced above would be sufficient to explore the value of a regular cold-start approach, it is not sufficient to explore the cold-start problem in the cross-domain context. Since in this context we have two domains, the cold-start issue should be explored in respect to each of the domains. Given that the number of items rated in the target domain can impact the quality of recommendation, it is natural to expect that the number of items rated in the auxiliary domain have a similar impact. Our third goal is to discover this potential effect. To achieve this goal we run another study, in some sense, symmetric to the first one. This study explores the quality of cross-domain recommendation as a function of cross-domain (auxiliary) profile size. The following sections present the details and the results of both studies.

5.1. The Experimental Setup

We pick 80% of the users randomly as training users and the rest as target users. We build 19 cold-start profiles for each target user in the movies domain: starting from having one movie in the profile to having 19 movies in the profile in a time-based order. In other words, the size-one cold-start profile of the target user has the first movie he/she rated and the size-19 cold-start profile of this user has the first 19 movies he/she has rated. Although we do not consider a time-dependent algorithm in providing the recommendations, we think that including the recommendations in the natural order of rating might has an effect on capturing user interest. For single-domain cold-start recommendations, we used collaborative filtering on the whole movie ratings of the training users and the cold-start profiles of the target users to predict the rating of target users on movies. We run the CF algorithm 19 times on this data: once for each of the cold-start profiles.

To compare the results of single-domain recommendations with cross-domain recommendations, we run 19 experiments using both cold-start profiles and book rating profiles of the users: for each target user, we combine all of his/her book ratings with the cold-start movie profiles and for each training user, we combine all of his/her books and movies as a whole profile. This setting is shown in Figure 6. To see the effects of having movies on predicting books, we set up the same experiment having 19 cold-start profiles for the book domain and the movie ratings as an auxiliary domain. We use k-NN algorithm with K = 20 for rating predictions in collaborative filtering.

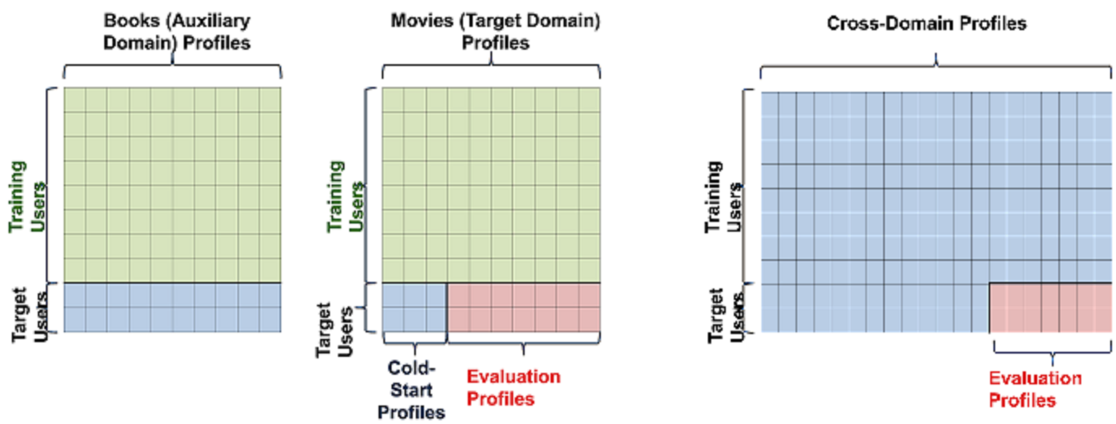


Figure 6 Setup for the Experiments.

5.2. Experimental Results

5.2.1. First Study: the Impact of Target Domain Ratings' Volume

We conducted our experiments to achieve the first and second goals described in section 5. We ran CF algorithm on both cross-domain and single-domain cold-start profiles to see if cross-domain recommendations perform better than single-domain recommendations. Also, we changed the number of items in the cold-start profiles from one to 19 to see which size of cold-start profile is the cold-start boundary size. In other words, how many ratings in a profile are enough with no need of cross-domain information to have good recommendations for users. The results for recommending books based on cold-start book ratings of users and cross-domain profiles of users are shown in Figure 7.

This Figure shows Mean Squared Error (MSE) of recommendations including a 95% confidence interval. MSE is a measure that shows how much our predicted ratings for users is different from users' actual ratings. As we can see in this picture, at first, with the cold-start user profiles having less than five book ratings, the cross-domain recommendations has significantly less MSE than the cold-start recommendations. After adding the 6th book, cold-start MSE results gets almost similar to the cross-domain MSEs. We can see that after adding more book ratings to the cold-start profile and completing it, the trend of cold-start profile MSEs are going to get better than cross-domain MSEs. This effect might be due to the extra noise cross-domain profiles add to the recommendations.

The result of a reverse study having movies as cold-start domain and books as auxiliary domain are shown in Figure. 8. The overall trend is similar to book recommendations: having one to five books in the cold-start profile, cross-domain recommendations work significantly better than cold-start recommendations. In this case, however, within-domain recommendations gets more advantage as the cold-start profile size increases. Starting from 16 movies in the cold-start profiles, these recommendations achieve reliably better MSE than cross-domain recommendations.

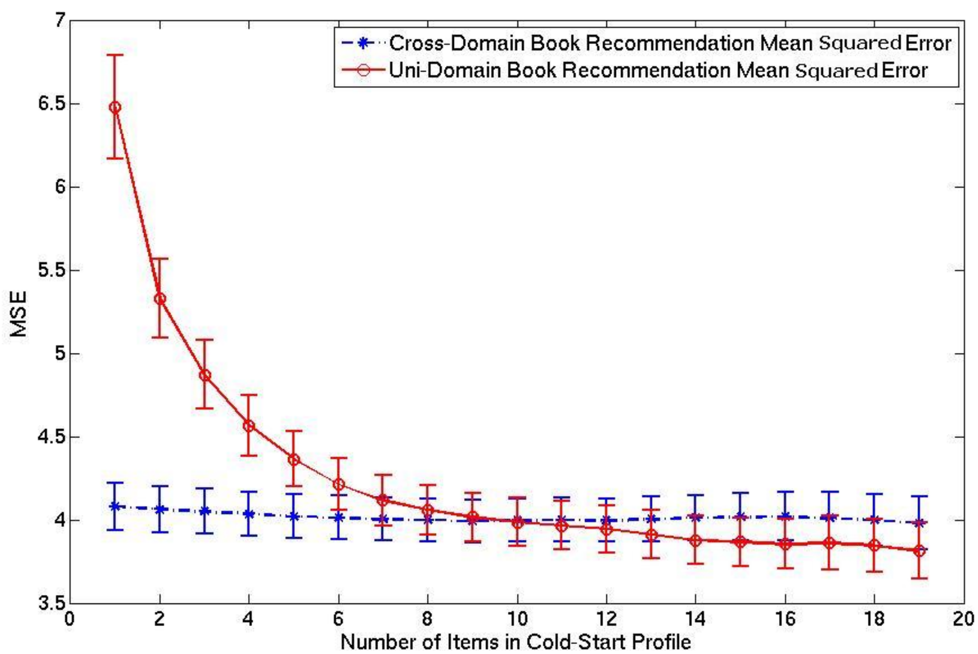


Figure 7 Mean Squared Error of Cross-Domain vs. single-domain Book Recommendations with 95% Confidence Interval on Varying Cold-Start Profiles.

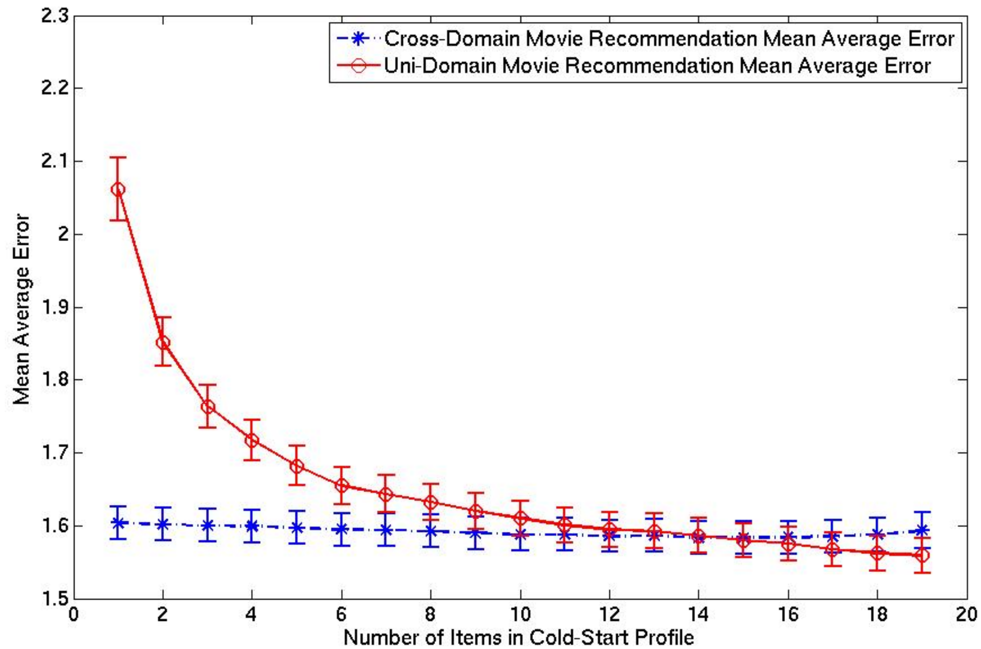


Figure 8 Mean Squared Error of Cross-Domain vs. single-domain Movie Recommendations with 95% Confidence Interval on Varying Cold-Start Profiles.

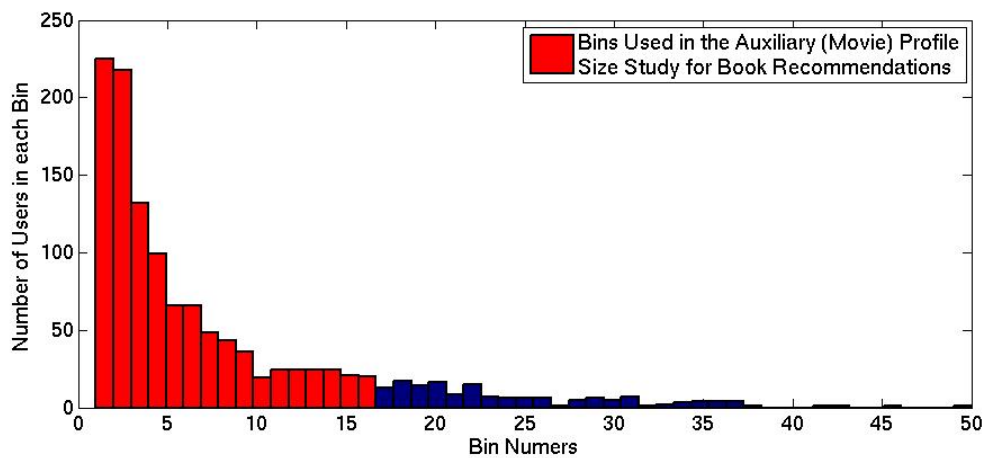


Figure 9 Histogram of Number of Users in each Bin (Number of Movies in Auxiliary Profile) for Book Cross-Domain Recommendations.

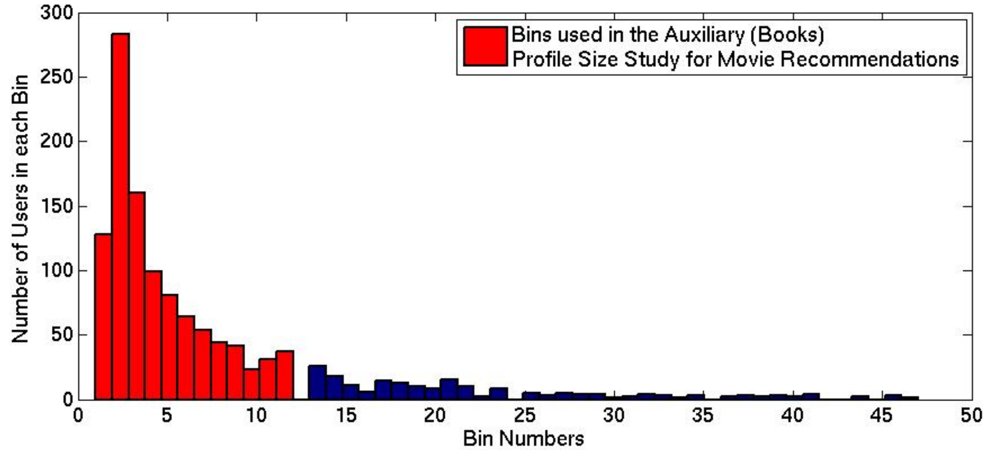


Figure 10 Histogram of Number of Users in each Bin (Number of Books in Auxiliary Profile) for Movie Cross-Domain Recommendations.

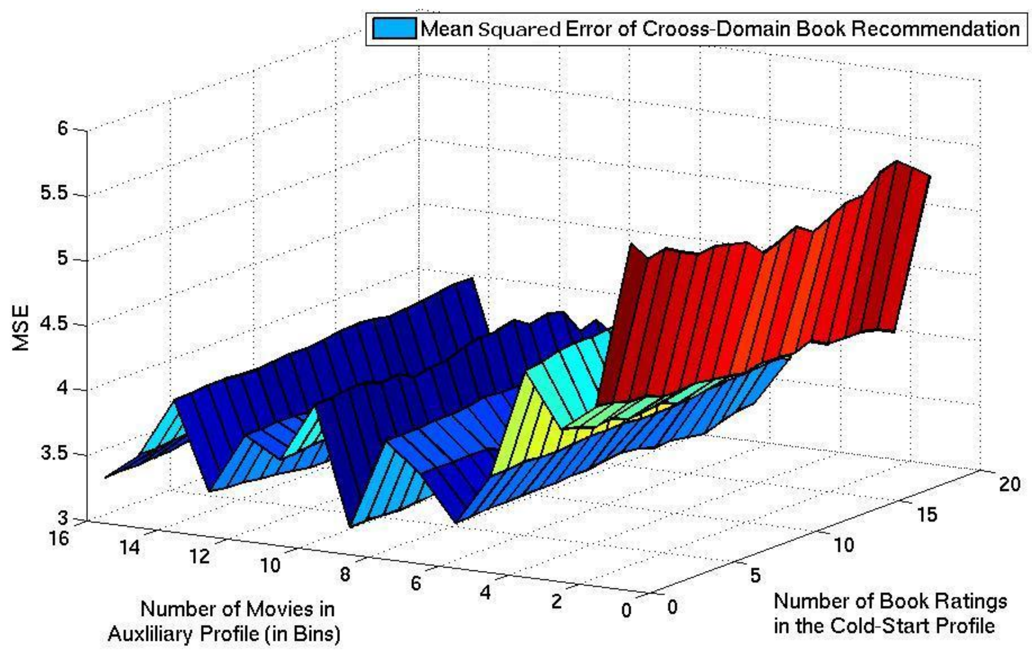


Figure 11 Mean Squared Error of Cross-Domain Book Recommendations with Varying Cold-Start and Auxiliary Profiles.

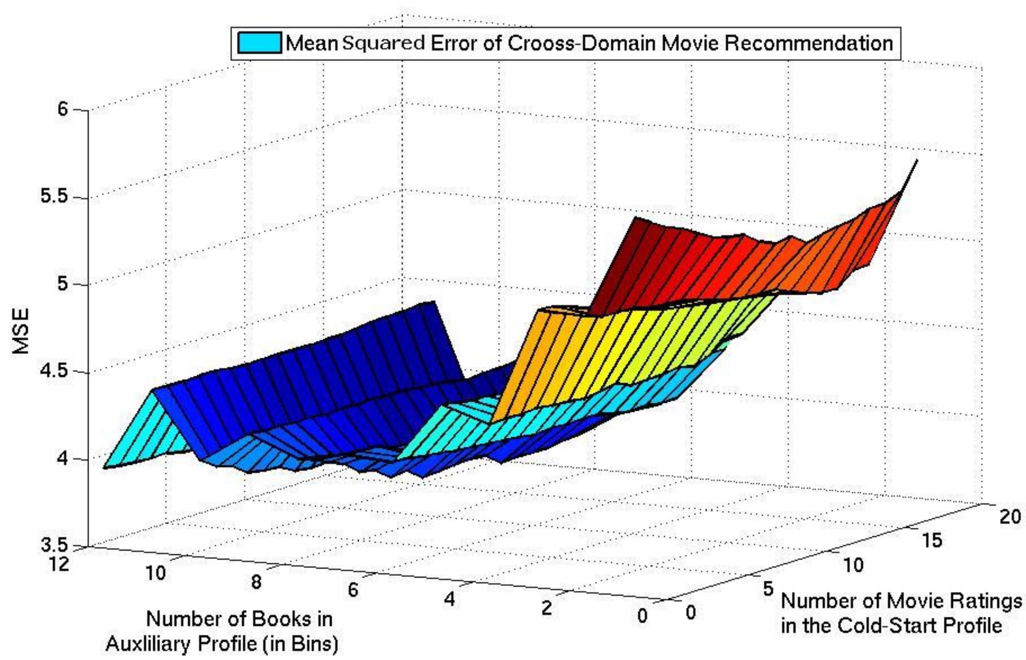


Figure 12 Mean Squared Error of Cross-Domain Movie Recommendations with Varying Cold-Start and Auxiliary Profiles.

5.2.2. Second Study: the Impact of Auxiliary Domain Ratings' Volume

To explore the behavior of cross-domain recommendation in cold-start context more extensively and to achieve our third goal, we examined the effect of the auxiliary profile size on mean squared error of cold-start recommendations. To do that in book recommendations, we divided the target user profiles into 100 bins starting from having 20 movies in the auxiliary profile to about 3400 movies in the profile. Each bin has a gap of about 35 movies. Histogram of number of users in each bin is shown in Figure. 9. We do not display the bins with numbers over 50 to show a clearer picture. As we can see, the number of users decreases as the number of movies rated in the profile increases. To have more accurate results, we ignored the auxiliary profile sizes/bins with less than 30 users in them. The red bars in Figure. 9 indicate utilized bins. As a result, we used only the first 16 bins in study of auxiliary profile size on error rate of cross-domain recommendations. It means that, we used 207 target users among all 6089 of them. The results are shown in Figure 11. Based on these results, in the first bins, having less number of ratings in the auxiliary profile, we have less accurate results. By increasing the size of auxiliary profile, the MSE decreases and we get better results. In higher number of bins, MSE gets a little bit higher which might be due to the noise auxiliary profile incorporates into the recommendations. Since the number of users decreases by increasing the bin number, we expect to have more trustable results in the first bins versus the last ones.

We conducted the same experiments for recommending movies based on auxiliary movie profiles. Again, we divided the target user profiles into 100 bins starting from

20 books to 2840 books. The bin size is about 23 books. Histogram of number of users in each bin is shown in Figure. 10. We cut the bin numbers larger than 50 to show a clearer picture. Again, we see that the number of users in the bins decrease as the bins increase. Thus, we ignore the auxiliary profile sizes/bins with less than 30 users in them. As a result, we used only the first 12 bins in this study of auxiliary profile size on error rate of cross-domain recommendations. The red bars in Figure. 10 indicate utilized bins. The results are shown in Figure 12. We can see the same pattern here: as the number of items increases in the auxiliary profile, the error rate decreases.

6. CONCLUSION

In this paper we reported an extensive exploration of cross-domain recommendation in a cold start context using imhonet dataset with a relatively large number of movie and book ratings produced by the same users. We started with a feasibility study. Using canonical correlation analysis, we demonstrated that user ratings in these two domains have reasonable correlation. This data provide a good ground for the exploration of different cross-domain recommendation approaches. We followed with an empirical study of a relatively simple cross-domain collaborative filtering approach based on concatenation of user profiles in two domains. A comparative study of this approach and a matching within-domain approach demonstrated that even a simple cross-domain approach can remarkably increase the quality of recommendation when the number of in-domain rated items is small. At the same time, as our data shows, the added value of the cross-domain recommendation decreases quite rapidly. The quality of within-domain recommendation surpasses the quality of cross-domain recommendation before the target domain profile reaches 20 ratings. While in the book domains the quality of cross- and within-domain recommendation still can be reliably distinguished at this point, the tendency looks clear. Finally, we also explored the impact of the user profile size in auxiliary domain on the quality of the cross-domain recommendation. As expected, additional information matters: with the increase of the auxiliary profile size, cross-domain approach was able to produced better recommendations.

Our study had several limitations. Most importantly, we studied a relatively similar pair of domains - movies and books. This pair has been already recognized as a promising ground for cross-domain recommendations and was a focus of several studies [19, 9, 10]. While our study demonstrated both the feasibility and the value of cross-domain recommendation between books and movies, the results for a less similar pair of domains (such as books and hotels) can be completely different. In our future work we plan to explore cross-domain recommendation for other domain pairs. On the other hand, the study part of this paper explored a relatively simple cross-domain approach based on profile concatenation. Due to the inertia of a large concatenated profile, this approach was not able to benefit from in-domain ratings and was surpassed by the within-domain approach relatively early. A smarter fusion of information from auxiliary and target domains could result in better cross-domain recommendation approaches. Further improvement of cross-domain recommendation can be obtained by using social linking and tagging information that is frequently present in modern social systems. We already started exploring this source of information in [13] and plan to continue this work.

REFERENCES

[1] S. Berkovsky, T. Kuflik, and F. Ricci. Cross-domain mediation in collaborative filtering. In Proceedings of the 11th international conference on User Modeling, UM '07, Berlin, Heidelberg, 2007. Springer-Verlag.

[2] S. Berkovsky, T. Kuflik, and F. Ricci. Mediation of user models for enhanced personalization in recommender systems. User Modeling and User-Adapted Interaction, 18(3), Aug. 2008.

[3] P. Cremonesi, A. Tripodi, and R. Turrin. Cross-domain recommender systems. In 2011 IEEE 11th International Conference on Data Mining Workshops. IEEE. Discussed many approaches, but evaluated using articicially separated netflix dataset. Mostly alg comparizon.

[4] S. Faridani. Using canonical correlation analysis for generalized sentiment analysis, product recommendation and search. In Proceedings of the fifth ACM conference on Recommender systems, RecSys '11, New York, NY, USA, 2011. ACM.

[5] I.Fern´andez-Tob´ıas,I.Cantador,M.Kaminskas,and F. Ricci. A generic semantic-based framework for cross-domain recommendation. In Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems, HetRec '11, New York, NY, USA, 2011. ACM.

[6] J. F. Hair, R. L. Tatham, R. E. Anderson, and W. Black. Multivariate Data Analysis (5th Edition). Prentice Hall, 5th edition, Mar. 1998.

[7] H. Hotelling. Relations Between Two Sets of Variates. Biometrika, 28(3/4), 1936.

[8] M. Jessenitschnig and M. Zanker. Iseller: A flexible personalization infrastructure for e-commerce applications. In Proceedings of the 10th International Conference on E-Commerce and Web Technologies, EC-Web 2009, Berlin, Heidelberg, 2009. Springer-Verlag.

[9] B. Li, Q. Yang, and X. Xue. Can movies and books collaborate?: cross-domain collaborative filtering for sparsity reduction. In Proceedings of the 21st international jont conference on Artifical intelligence, IJCAI'09, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.

[10] A. Loizou. How to recommend music to film buffs: enabling the provision of recommendations from multiple domains. PhD thesis, 2009.

[11] H. Ohkushi, T. Ogawa, and M. Haseyama. Kernel cca-based music recommendation according to human motion robust to temporal expansion. In Communications and Information Technologies (ISCIT), 2010 International Symposium on, oct. 2010.

[12] W. Pan, E. W. Xiang, N. N. Liu, and Q. Yang. Transfer learning in collaborative

filtering for sparsity reduction. In AAAI, 2010.
[13] S. Sahebi and W. Cohen. Community-based recommendations: a solution to the cold start problem. In Workshop on Recommender Systems and the Social Web, in conjunction with ACM RecSys'11. ACM, 2011.
[14] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock. Methods and metrics for cold-start recommendations. In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '02, New York, NY, USA, 2002. ACM.
[15] Y. Shi, M. Larson, and A. Hanjalic. Tags as bridges between domains: improving recommendation with tag-induced cross-domain collaborative filtering. In Proceedings of the 19th international conference on User modeling, adaption, and personalization, UMAP'11, Berlin, Heidelberg, 2011. Springer-Verlag.
[16] J. Takagi, Y. Ohishi, A. Kimura, M. Sugiyama, M. Yamada, and H. Kameoka. Automatic audio tag classification via semi-supervised canonical density estimation. In Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, may 2011.