Cross-Domain Collaborative Recommendation in a Cold-Start Context: The Impact of User Profile Size on the Quality of Recommendation

Shaghayegh Sahebi and Peter Brusilovsky

Intelligent Systems Program
University of Pittsburgh
{shs106,peterb}@pitt.edu

Abstract. Most of the research studies on recommender systems are focused on single-domain recommendations. With the growth of multi-domain internet stores such as iTunes, Google Play, and Amazon.com, an opportunity to offer recommendations across different domains become more and more attractive. But there are few research studies on cross-domain recommender systems. In this paper, we study both the cold-start problem and the hypothesis that cross-domain recommendations provide more accuracy using a large volume of user data from a true multi-domain recommender service. Our results indicate that cross-domain collaborative filtering could significantly improve the quality of recommendation in cold start context and the auxiliary profile size plays an important role in it.

1 Introduction

For a new user 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 (Cold-start problem [6]). This paper offers an extensive exploration of a specific approach to address a new-user problem: leveraging user rating data from a different domain (or cross-domain recommendation). We believe that the potential value of cross-domain recommendation is rapidly increasing in the modern social Web context. Nowadays the number and the diversity of web systems are increasing so rapidly that a typical internet user has ratings, votes, or bookmarks left in a number of different systems. 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.

While the problem of cross-domain recommendation was discussed in a 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 single-domain dataset and considering that as separate domains e.g. separating the movie domain based on their genres [1, 2, ?] or by using different user and items sets [4]. Consequently, there is still no extensive study of cross-domain recommendation in a cold-start context [3]; there is even no reliable evidence that the cross-domain recommendation could fulfill its promise as a cold-start helper in a real multi-domain context. The study presented in this paper differs from the past studies in several ways: we use a

large dataset of user ratings in two different domains obtained from a real-world recommendation service; we do not consider any extra information such as content, tags, or product features, which differ from case to case and might not be available; and we explore the cross-domain recommendations in cold-start setting with two different point of views in two studies.

In this paper 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. 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   The Cold Start Studies

As explained in the introduction, the focus of this paper is an extensive study of cross-domain recommendation in the cold-start context. The idea 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. Our quality study pursues two goals: 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.

Previous studies of recommendation approaches in 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 (cold start boundary), a regular approach surpasses it. Following that, our first goal is to find the cold-start boundary for cross-domain recommendations. In other words, expecting that at some point the quality of single-domain recommendations may surpass the quality of the cross-domain recommendations. 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 goal, as a baseline, we use a traditional collaborative filtering approach based on target-domain profiles. As a comparable 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. We predict target user ratings on the target domain ratings by traditional collaborative filtering approach.

While the study introduced above represent a commonly accepted approach to study recommendation quality in a cold-start approach, it explores only one side of the phenomenon, i.e., the impact of the user profile size in the target domain on the quality of cross-domain recommendations. We believe, however, that 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. Thus, the second goal of our study is to discover the impact of the auxiliary-domain profile sized on the comparative recommendation quality of the cross-domain approach. To achieve this goal we run another study, which explores the quality of cross-domain recommendation as a function of cross-domain (auxiliary) profile size.

3   The Dataset and Experimental Setup

In this study we used an anonymized dataset obtained from an online Russian recommender service Imhonet (www.imhonet.com). Imhonet allows users to rate and review a range of items from books and movies to mobile phones and architectural monuments. For our study, we used an anonymized Imhonet 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. 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.

If we look at the distribution of movie and book ratings in this dataset we can see that based on imhonet's request, many users rated at least 20 books and 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. For more detailed information about the dataset please refer to [5].
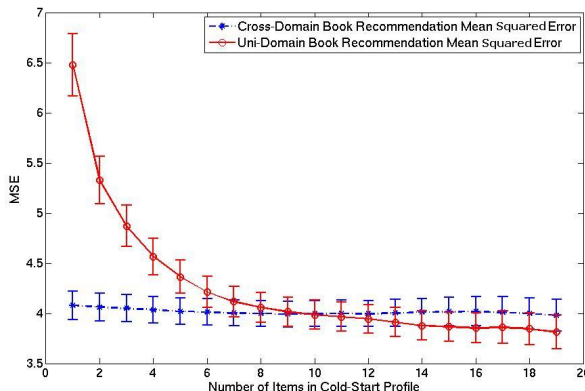

Fig. 1: MSE of Cross vs. Single -Domain Book Recommendations with 95% CI on Varying Cold-Start Profiles.

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 has rated and the size-19 cold-start profile of this user has the first 19 movies he/she has rated. 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, 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. 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.

4   Experimental Results

4.1   Study I: Impact of User Profile Size in Target Domain

Our first study was conducted to achieve the first goal described in section 2. The results for recommending books based on cold-start book ratings of users and cross-domain profiles of users are shown in Figure 1. This Figure shows Mean
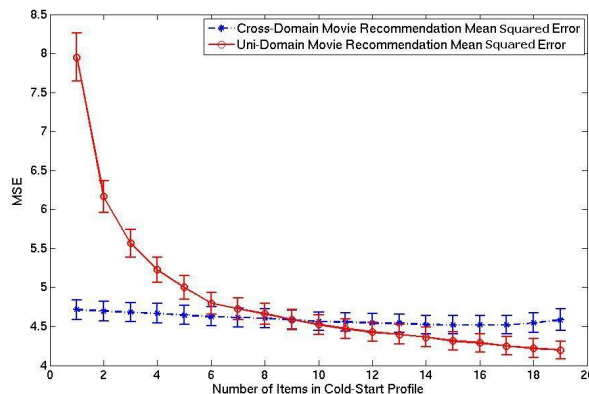

Fig. 2: MSE of Cross vs. Single -Domain Movie Recommendations with 95% CI on Varying Cold-Start Profiles.

Squared Error (MSE) of recommendations including a 95% confidence interval. MSE is a measure that shows how much our predicted ratings for users are different from users' actual ratings. As we can see in this picture, while the user has five or fewer book ratings, the cross-domain recommendations achieve significantly smaller MSE than the single-domain recommendations. After adding the 6th book, the single-domain MSE decreases to the point when it becomes statistically comparable with cross-domain MSEs. This trend continues once more book ratings are being added to the cold-start book profile: the single-domain MSEs is gradually surpassing cross-domain MSEs, although not becoming better significantly within the explored window size. In other words, once the target profile becomes reasonably large (12 ratings or more) the cross-domain ratings introduces more noise than value from the prospect of rating prediction. Note it doesn't mean that the cross-domain approach becomes useless at that point. It has long been argued that even with the lower rating prediction quality, the

cross-domain approach can offer other values such as novelty and diversity. Further studies are required to evaluate the comparative quality of two approaches from these prospects.
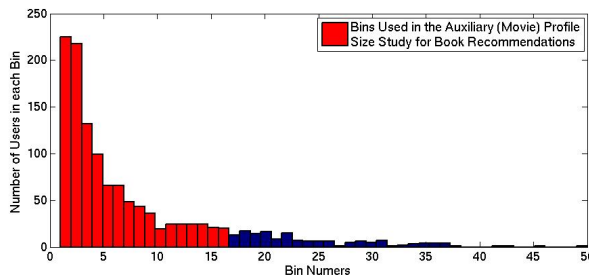

Fig. 3: Histogram of Number of Users in each Bin (Number of Movies in Auxiliary Profile) for Book Cross-Domain Recommendations.

The result of a reverse study with movies as target (cold-start) domain and books as auxiliary domain are shown in Figure. 2. As the data shows, the overall trend is similar to book recommendations: for the users who have one to five books in the cold-start profile, cross-domain recommendations work better than cold-start recommendations. In this case, however, single-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.

4.2   Study II: Impact of Auxiliary Domain Ratings' Volume

To explore the behavior of cross-domain recommendation in cold-start context more extensively and to achieve our second 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. 3. 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 achieve more reliable results, we ignored the auxiliary profile sizes/bins with less than 30 users. The red bars in Figure. 3 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. The results are shown in Figure 4. The results reveal that the number of ratings in the auxiliary profile has a strong impact on the accuracy of cross-domain recommendations. The accuracy is the lowest (MSE highest) for the users in the first bin with the number of ratings between 20 and 55. The accuracy improves considerably for the users in the second bin and then in the third bin. This tendency continues further although once the number of users in a bin falls below 100, the accuracy increase become less monotonic due to the small bin size. Overall, the data shows that the impact of the user profile size in the auxiliary domain is qualitatively similar to the impact of the profile size

in the target domain: the more ratings are available in the auxiliary profile (i.e., the more information we have about user preferences), the better is the accuracy of recommendations. At the same time, the scale of this impact is very different. While within the target domain, adding each next rating visibly improves the accuracy, within the auxiliary domain this improvement is much more slow.
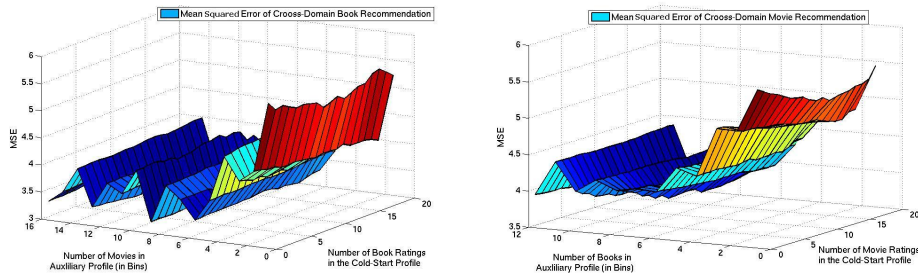

Fig. 4: MSE of Cross-Domain (left) Book and Movie (right) Recommendations with Varying Cold-Start and Auxiliary Profiles.

We conducted the same experiments for recommending movies based on auxiliary book profiles. In a similar way, we divided the target user profiles into 100 bins starting from 20 books to 2840 books. The bin size is about 23 books. 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 results are shown in Figure 4. We can see the same pattern here: as the number of items increases in the auxiliary profile, accuracy of recommendation increases (MSE decreases).

5 Conclusion

In this paper we reported an extensive exploration of cross-domain recommendation in a cold start context with a relatively large number of movie and book ratings. Our studies demonstrated that even a simple cross-domain approach can increase the quality of recommendation when the number of in-domain rated items is small. At the same time the added value of the cross-domain recommendation decreases quite rapidly. Finally, we also explored the impact of the user profile size in auxiliary domain on the quality of the cross-domain recommendation. As expected, with the increase of the auxiliary profile size, cross-domain approach was able to produce better recommendations. This result indicates that the recommendation power of a cross-domain approach is not free: it is fueled by a considerable number of user ratings in a an auxiliary domain. For users with too few auxiliary domain ratings, cross-domain recommendation performs much worse than for the users with size-able profiles.

In future work we plan to explore cross-domain recommendation for other domain pairs. In addition, the study part of this paper explored a relatively simple cross-domain approach and a traditional collaborative filtering approach. A smarter fusion of information from auxiliary and target domains could result in better cross-domain recommendations.

References

1. S. Berkovsky, T. Kuflik, and F. Ricci. Cross-domain mediation in collaborative filtering. User Modeling, UM'07, 2007.
2. S. Berkovsky, T. Kuflik, and F. Ricci. Mediation of user models for enhanced personalization in recommender systems. *User Modeling and User-Adapted Interaction, UMUAI'08*, 18(3), Aug. 2008.
3. P. Cremonesi, A. Tripodi, and R. Turrin. Cross-domain recommender systems. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on Data Mining*, pages 496 –503, dec. 2011.
4. W. Pan, E. W. Xiang, N. N. Liu, and Q. Yang. Transfer learning in collaborative filtering for sparsity reduction. In *Artificial Intelligence, AAAI'10*, 2010.
5. S. Sahebi and W. Cohen. Community-based recommendations: a solution to the cold start problem. In *RSWEB, in conjunction with ACM RecSys'11*, 2011.
6. A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock. Methods and metrics for cold-start recommendations. Special Interest Group in Information Retrieval, SIGIR '02, New York, NY, USA, 2002.