Browser-Oriented Universal Cross-Site Recommendation and Explanation based on User Browsing Logs

Yongfeng Zhang
Department of Computer Science and Technology
Tsinghua University, Beijing, 100084, China
zhangyf07@gmail.com

ABSTRACT

Our research aims to bridge the gap between different websites to provide cross-site recommendations based on browsers. Recent advances have made recommender systems essential to various online applications, such as e-commerce, social networks, and review service websites. However, practical systems mainly focus on recommending inner-site homogeneous items. For example, a movie review website usually recommends other movies within the site when a user has enjoyed a movie online. However, it would be exciting if the system recommends some attractive products related to this movie from some e-commerce websites like Amazon or eBay.

Such an ability to provide heterogeneous cross-site recommendations may shed light on brand new and promising business models, which could benefit both the online shops in expanding the marketing efforts, and the online users in discovering items of interest from a wider scope. In this research, we propose and formalize the problem of universal recommendation, record and analyze user browsing actions in web browsers, and provide browser-oriented cross-site recommendations when the users are surfing online.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Filtering; H.3.5 [**Online Information Services**]: Web-based services

Keywords

Recommender Systems; Collaborative Filtering; Browser Logs; Block Diagonal Form; Matrix Factorization

1. INTRODUCTION

With the development of Collaborative Filtering (CF) [15], the last few years have seen an increasingly important role that Recommender Systems (RS) [13] play in many online applications, such as product recommendation in Amazon, and friend recommendation in Facebook.

However, most practical recommender systems aim to provide inner-site homogeneous items according to the preferences of the target users, for example, to recommend similar videos when the user has watched a video online. Although intuitional and reasonable, the inner-site homogeneous recommendation paradigm is far from the tremendous benefits that recommender systems can bring into the Web.

The ability to recommend cross-site heterogeneous items benefits both the online businesses and the users. For example, after a user has finished watching the film *Titanic* on a video sharing website like YouTube, in which case it could be possible that she has been moved to tears, recommending an *Ocean Heart* necklace product from the online shopping website Amazon can be just no better.

The problem of cross-site recommendation is different from, but closely related to the problem of cross-domain recommendation, which attempts to provide recommendations of items in a target domain (e.g. books), by taking advantages of the user behaviours from other domains (e.g. movies or musics) [6, 19, 20]. Usually, recommendations are also provided in different domains simultaneously [14, 2].

However, an important problem troubling cross-domain recommendation in practice is the lack of inter-domain user behaviours, such as ratings, implicit feedbacks, or user-generated contents. This problem makes it even more difficult to conduct cross-site recommendation, because it is usual that the information of cross-site user behaviours may not even exist.

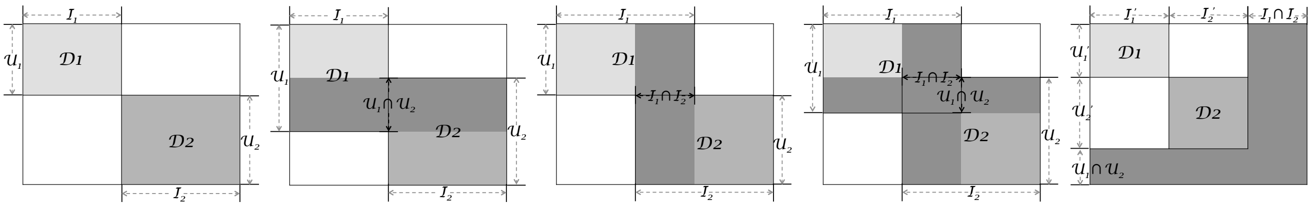
Fortunately, the vast amount of cross-site user browsing logs recorded by web browsers can serve as a kind of general inter-site relations in cross-site recommender systems. For example, a user who has finished watching the film *Titanic* on YouTube might continue to search for an *Ocean Heart* necklace from Amazon. By leveraging the wisdom of crowds and conducting CF according to the vast amount of similar browsing patterns, we can extract the hidden relationships between the heterogeneous items from different websites.

Web browser as the interface of browsing the internet can also serve as a platform of providing recommendations to the users. Due to the restrictions on security considerations, it is usually difficult for a web page to collect user browsing histories of other websites, which further makes it difficult for them to provide cross-site recommendations. However, such recommendations can be provided by the browser directly, which (if licenced by the user) is able to track the user browsing behaviors on different websites, and thus to provide recommended items from other websites according to the current browsing item and the preferences of the user.

In this research, we propose the problem of cross-site recommendation, and make it clear the similarities and differences between cross-site and cross-domain recommendation. We formalize the problem in a general and unified framework based on the Bordered Block Diagonal Form (BBDF) [25, 26, 24] structure of user-item relation matrices, and conduct cross-site recommendation with user browsing logs collected from web browsers.

RecSys'14, October 6–10, 2014, Foster City, Silicon Valley, CA, USA.

ACM 978-1-4503-2668-1/14/10 ...$15.00.
http://dx.doi.org/10.1145/2645710.2653367.



(a) No overlap (b) User overlap (c) Item overlap (d) User&Item overlap (e) BBDF structure

Figure 1: The user set and item set of two different domains may exhibit four relationships: (a) No common users or items; (b) Exist common users but no common items; (c) Exist common items but no common users; and (d) Exist both common users and items. The four relationships can be generalized into a unified bordered block diagonal form structured matrix (e) by permuting the common users/items to borders.

In the following part, we review some related work in Section 2, and introduce some of our research progress to date on cross-site recommendation in Section 3. We make further discussions, conclude the paper, and point out some of the future research directions in Section 4.

2. RELATED WORK

The most closely related research field is cross-domain personalized recommendation [2], which has been attracting the attention from research community over the past few years. One of the first studies was that presented by Winoto and Tang in [19], attempting to uncover the association between user preferences on related items across domains. They also discussed about the appropriate evaluation methods for cross-domain recommender systems.

The research of cross-domain recommendation has since then been driven into two main directions. One is to discover domain relations by Collaborative Filtering (CF) [6, 14, 7, 9, 27, 4, 5], and the other is to leverage various inter-domain content information [20, 18, 10, 1, 3, 16, 22].

One of the key techniques for CF approaches is Transfer Learning [8, 7, 9] and various of its extensions or variations. Singh [14] proposed Collective Matrix Factorization (CMF) to estimate the relations of multiple entities by factorizing several matrices simultaneously while sharing parameters in the latent space. Li [6] attempted to transfer user-item rating patterns from an auxiliary matrix in another domain to the target domain through Codebooks. Hu [5] and Zhao [27] extended transfer learning to triadic factorization and active learning for cross-domain recommendation, respectively. Gao [4] investigated cluster-level latent factor models to learn the common rating patterns shared across domains.

An important problem CF faces is data sparsity [4, 16, 9]. To alleviate the problem in a more direct way, researchers have been seeking the possibility of leveraging inter-domain relations and contents for cross-domain recommendations.

Shi [20] and Yang [22] made use of user generated tags as the bridge between domains to improve the performance. Chen [1] and Roy [10] used social relationships to recommend between domains in social media. More recently, Tobias [3, 18] attempted to leverage semantic network to discover the semantic-level relationships between objectives from different domains to make more informed recommendations.

The approach of leveraging inter-domain contents could be effective in specific applications. However, a key problem of such approaches is that the inter-domain information may be absent in many cases, and a substantial amount of manual work may have to be performed in order to align the content from different domains.

3. RESEARCH PROGRESS TO DATE

In this section, we introduce some of the current research progress on the topic of cross-site recommendation. We first review the four kinds of user-item relations in cross-domain recommendation, and further generalize them into a unified Bordered Block Diagonal Form (BBDF) structure [25, 24] to formalize the cross-site recommendation problem. User browsing logs are then integrated into this framework to serve as cross-site user and item relations. Besides, we further provide feature-level recommendation explanations [23] based on phrase-level sentiment analysis [21, 17] to improve the persuasiveness of cross-site recommendations.

3.1 Bordered Block Diagonal Forms

In cross-domain recommendation, the relationship between users and items from different domains can be generally classified according to their overlaps [2]. Without loss of generality, we consider the relationship between two domains. According to whether there exist common users (or items) between the user sets (or items sets) of the domains, there can be four kinds of relationships between the two domains, namely, no overlap users or items (Fig.1(a)), overlap users only (Fig.1(b)), overlap items only (Fig.1(c)), and both overlap users and items (Fig.1(d)).

It is important to note that, the four kinds of relationships can be generalized into a unified BBDF matrix structure by permuting the common users and items (if they exist) of the two domains to matrix borders [25, 24], as shown in Fig.1(e). The row and/or column borders in the BBDF structure may be absent if there is no common user and/or item between the two domains, respectively.

In the case of cross-site recommendation, the existence of common users and/or items may depend on the type of websites as well as the user and item information collected by the browsers. For example, two e-commerce websites like Amazon and eBay may contain a well amount of similar or even the same online products. By aligning the products from the two websites, we may extract the common items to serve as cross-site relations in personalized recommendation. However, an e-commerce website may hardly contain any common item with a video sharing website like YouTube, which leads to the absence of such cross-site item relations. Similarly, a substantial amount of users own accounts on many online service websites like Facebook, Twitter and Amazon. By user account alignment, we can extract the underlying common users among the websites, which also serve as the cross-site relationships.

In many common cases however, there could be no common users or items between different websites, which leaves us with two independent domains as shown in Fig.1(a), and thus there is no row or column border in Fig.1(e). Such a case also makes it difficult or even impossible to make cross-site recommendation if no external information is provided.

3.2 Browser-Oriented Recommendation

Cross-site recommendations can be best provided by web browsers. Traditional recommender systems are usually embedded in various online applications, like the product recommendation in e-commerce websites, and the friend recommendation in social networks. However, it is usually difficult for these web applications to collect user and item information from other sites, and such technical difficulties prevent them from providing cross-site recommendations. Besides, from a business perspective, some web service providers may be unwilling to present cross-site recommendations due to the business competitions, although a product from another website may better meet the current user need.

However, the cross-site recommendations can be provided beyond the websites and by the web browser directly [23], as shown in Fig.2. Besides, by tracking the user browsing actions across various websites during a period of time, the browser can better understand user needs than each single website, thus to make more informed recommendations. Additionally, the web browser logs can be easily integrated into the unified BBDF structure for cross-site recommendation. In the following sections, we discuss about the ways to incorporate browser logs, as well as the approaches to conduct CF on the BBDF structure to generate the recommendations.

3.3 Incorporating Browser Logs

We collected a vast amount of real user browsing logs from a commercial web browser, including 2.1 million users and 24 million web items (i.e. distinct urls). The browsing actions include visit time, duration, click position, and page title, etc, which makes it possible for us to analyze the user behaviours and preferences in real-world scenarios.

The browser logs provide a general way to construct inter-site relations by recovering the row borders (user relations) and column borders (item relations). Given an user-item relation matrix of two websites without any overlap users or items (as in Fig.1(a)), we adopt the following four approaches to recover the cross-site relations (as in Fig.1(e)):

1. Common user identification: The browser can identify that two user IDs from different websites actually correspond to the same user when she signs in on the websites. In such a case, the row vectors from different domains in Fig.1(a) that correspond to the same user are merged and permuted to serve as a row border.

2. Common item identification: We also merge two column vectors and permute them to border when the corresponding items are identified as the same one. For example, when a user follows an account on Twitter, we can also recommend her to follow the same account on Facebook.

3. Auxiliary users: Some browser users never sign in on the concerned websites, and thus it is difficult to map them to the existing user IDs of these sites. However, they can be appended to the BBDF structured relation matrix as row borders to help discover the item relations.

4. Auxiliary items: Symmetrically, we append the unmapped web items to the BBDF structure as column borders, and they help to estimate the collaboration relationship among the users.

(a) In browser recommendation (b) Explanations

Figure 2: (a) Cross-site recommendations can be provided by browser plugins directly, beyond the specific website that the user is currently browsing. (b) The browser can also provide intuitional recommendation explanations, e.g., a feature word cloud.

Once the user and item information mined from browser logs is integrated into the BBDF structure and properly normalized, the cross-site relations (i.e. matrix borders) make it possible to conduct CF for personalized recommendation.

3.4 Localized Matrix Factorization

Our preliminary work on Localized Matrix Factorization (LMF) [26] provides a theoretical paradigm for conducting CF with matrix factorization techniques on the BBDF structured matrices. In this research, we adopt the LMF framework to make cross-site predictions and recommendations.

Matrix Factorization (MF) techniques have proved to be effective and efficient in modeling the hidden relationships of users and items in a latent factor space. However, MF algorithms fail to make cross-site predictions if there is no border in the BBDF structure as in Fig.1(a) [26, 25]. The reason is similar to that of the neighbourhood-based approaches, for which it would be impossible to compute the user similarities if there is no overlap item between two row vectors. As a result, the essence of incorporating bowser logs to recover the borders lies in that, it bridges the users and items across different sites, which makes it possible to model the relations in a common latent factor space.

The LMF framework is capable of integrating any decomposable MF algorithm, including many of the commonly used approaches, such as Singular Value Decomposition, Nonnegative Matrix Factorization, and Probabilistic Matrix Factorization. Besides, the framework can be easily parallelized to tackle the vast amount of cross-site user feedbacks.

3.5 Recommendation Explanation

Providing appropriate recommendation explanations could be more important in cross-site recommendation than inner-site recommendations, because it might be difficult for users to understand why a heterogeneous item is recommended instead of a similar product. Persuasive and informative explanations also attract the attention of users and guide the users across various websites that we expect them to visit. The ability to guide user browsing actions during the browsing process is of key importance for web browsers to gain the initiative in web economy.

Our preliminary work on Explicit Factor Models (EFM) [23] leverages phrase-level sentiment analysis to provide feature-level explanations by telling the user which feature of the recommended item may attract her. Practical online experiments verified the persuasiveness of the explanations. In this research, we further develop the idea of feature-level explanation by extracting item features from different websites and domains. Except for conducting phrase-level sentiment analysis on user reviews, the features can also be extracted from various other information sources, such as tags, metadata, and wikipedia categorization pages, etc.

3.6 Experimentation and Evaluation

Many commonly used evaluation metrics mainly focus on the precision of rating prediction or top-k recommendation, including MAE, RMSE, MRR, Precision@k, Recall@k and F-Measure, etc [11]. However, it is widely considered that the advantage of cross-domain/site recommendation may not be the improvement in accuracies, but the added novelty and diversity of the recommendations, which may offer better satisfaction and utility to the users [12, 11]. In this context, the recently proposed novelty and diversity metrics could also be taken into consideration [11].

Besides, we can perform online evaluation with large-scale practical users in real-world scenarios, which is a more flexible way to evaluate the performance of cross-site recommendation and explanation, and to conduct comparative studies with traditional inner-site recommendations.

4. CONCLUSIONS AND RESEARCH PLAN

Cross-site recommendation is more than a site-wide view of cross-domain recommendation. The research of cross-site recommendation touches various important and cutting-edge problems of recommender systems, *e.g.*, the problems of code-start, novelty, explanations, transfer-learning, data integration, and recommender interfaces. This makes cross-site recommendation a promising task to integrate and extend the research achievements of personalized recommendation. Besides, the browser-oriented recommendation may also bring brand new business models to the Web.

In the previous studies, we have formalized cross-site recommendation in a unified BBDF framework, investigated the feature-level explanation of recommendation based on sentiment analysis, and designed browser-based recommendation interfaces. Both offline and online experimental results verified the effectiveness and efficiency of our proposed methods in [26, 25, 24, 23, 21]. However, these are still preliminary studies of cross-site recommendation, and there is much room for improvement. Some of the key problems to investigate in the future includes: Integrating (perhaps inconsistent) data from different websites into the unified BBDF framework; Extracting explicit features from different information sources for explainable recommendations; Understanding the user needs during the browsing process; and designing appropriate metrics as well as online experiments for the evaluation of cross-site recommendations.

Acknowledgment

I sincerely thank my supervisors Professor Shaoping Ma, Min Zhang and Yiqun Liu for the consistent support and mentoring during my research, and special thanks to Prof. Yi Zhang of UCSC and Prof. Min-Yen Kan of NUS for the fruitful discussions and valuable suggestions.

5. REFERENCES

[1] W. Chen, W. Hsu, and M. L. Lee. Making recommendations from multiple domains. *KDD*, 2013.

[2] P. Cremonesi, A. Tripodi, and R. Turrin. Cross-Domain Recommender Systems. *The 11th IEEE International Conference on Data Mining Workshops*, 2011.

[3] I. Fernandez-Tobias. Mining Semantic Data, User Generated Contents and Contextual Information for Cross-Domain Recommendation. *UMAP*, 2013.

[4] S. Gao, H. Luo, D. Chen, S. Li, P. Gallinari, and J. Guo. Cross-Domain Recommendation via Cluster-Level Latent Factor Model. *ECML*, 2013.

[5] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and C. Zhu. Personalized Recommendation via Cross-Domain Triadic Factorization. *WWW*, pages 595–605, 2013.

[6] B. Li, Q. Yang, and X. Xue. Can Movies and Books Collaborate? Cross-Domain Collaborative Filtering for Sparsity Reduction. *IJCAI*, pages 2052–2057, 2009.

[7] B. Li, Q. Yang, and X. Xue. Transfer Learning for Collaborative Filtering via a Rating-Matrix Generative Model. *ICML*, pages 617–624, 2009.

[8] S. J. Pan and Q. Yang. A Survey on Transfer Learning. *TKDE*, 22(10):1345–1359, 2010.

[9] W. Pan, N. N. L. Evan W. Xiang, and Q. Yang. Transfer Learning in Collaborative Filtering for Sparsity Reduction. *AAAI*, pages 230–235, 2011.

[10] S. D. Roy, T. Mei, W. Zeng, and S. Li. SocialTransfer: Cross-Domain Transfer Learning from Social Streams for Media Applications. *MM*, 2012.

[11] S. Vargas, P. Castells. Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems. *RecSys*, 2011.

[12] G. Shani and A. Gunawardana. Evaluating Recommender Systems. *Recommender Systems Handbook*, 2011.

[13] 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 Comp. Surv.*, 47(1), 2014.

[14] A. P. Singh and G. J. Gordon. Relational Learning via Collective Matrix Factorization. *KDD*, 2008.

[15] X. Su and T. M. Khoshgoftaar. A Survey of Collaborative Filtering Techniques. *Advances in AI*, 2009(4), 2009.

[16] S. Tan, J. Bu, X. Qin, C. Chen, and D. Cai. Cross domain recommendation based on multi-type media fusion. *Neurocomputing*, pages 124–134, 2014.

[17] Y. Tan, Y. Zhang, M. Zhang, Y. Liu, and S. Ma. A Unified Framework for Emotional Elements Extraction based on Finite State Matching Machine. *NLPCC*, 400:60–71, 2013.

[18] I. F. Tobias, M. Kaminskas, I. Cantador, and F. Ricci. A Generic Semantic-based Framework for Cross- Domain Recommendation. *HetRec*, pages 25–32, 2011.

[19] P. Winoto and T. Tang. If You Like the Devil Wears Prada the Book, Will You also Enjoy the Devil Wears Prada the Movie? A Study of Cross-Domain Recommendations. *New Generation Computing*, 26(3):209–225, 2008.

[20] Y. Shi, M. Larson, A. Hanjalic. Tags as Bridges between Domains: Improving Recommendation with Tag-Induced Cross-Domain Collaborative Filtering. *UMAP*, 2011.

[21] Y. Zhang, H. Zhang, M. Zhang, Y. Liu, S. Ma. Do Users Rate or Review? Boost Phrase-level Sentiment Labeling with Review-level Sentiment Classification. *SIGIR*, 2014.

[22] D. Yang, Y. Xiao, Y. Song, J. Zhang, K. Zhang, and W. Wang. Tag Propagation Based Recommendation across Diverse Social Media. *WWW*, 2014.

[23] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma. Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis. *SIGIR*, 2014.

[24] Y. Zhang, M. Zhang, Y. Liu, and S. Ma. A General Collaborative Filtering Framework based on Matrix Bordered Block Diagonal Forms. *Hypertext*, 2013.

[25] Y. Zhang, M. Zhang, Y. Liu, and S. Ma. Improve Collaborative Filtering Through Bordered Block Diagonal Form Matrices. *SIGIR*, 2013.

[26] Y. Zhang, M. Zhang, Y. Liu, S. Ma, and S. Feng. Localized Matrix Factorization for Recommendation based on Matrix Block Diagonal Forms. *WWW*, 2013.

[27] L. Zhao, S. J. Pan, E. W. Xiang, E. Zhong, Z. Lu, and Q. Yang. Active Transfer Learning for Cross-System Recommendation. *AAAI*, 2013.