2011 23rd IEEE International Conference on Tools with Artificial Intelligence

Cross-Domain Collaborative Filtering: A Brief Survey

Bin Li
Centre for Quantum Computation and Intelligent Systems
University of Technology, Sydney
NSW 2007, Australia
Email: bin.li-1@uts.edu.au

***Abstract*—Cross-domain collaborative filtering (CF) is an emerging research topic in recommender systems. It aims to alleviate the sparsity problem in individual CF domains by transferring knowledge among related domains. In this paper, we will give a brief survey of the pilot studies in this research line in two dimensions: Collaborative Filtering Domains and Knowledge Transfer Styles. Some possible extensions for cross-domain CF will be discussed in the end.**

***Keywords*-survey; collaborative filtering; transfer learning; cross-domain; recommender systems**

I. INTRODUCTION

Collaborative filtering (CF) in recommender systems aims to predict missing ratings for a user or an item based on the observed ratings collected from like-minded users or similar items. Due to its high-efficiency and content-free advantages, CF has been widely adopted in today's recommender systems. However, since CF techniques purely rely on the observed rating data, the sparsity problem has become a major bottleneck for CF methods [1].

In real-world scenarios, we can easily find related CF domains that recommend similar items with the target one. A question was then asked in [2]: Can we establish a *bridge* between related CF domains and transfer useful knowledge from one another to improve the performance? This question has been answered "Yes" later on in a number of cross-domain CF methods [2], [3], [4], [5], [6], [7], [8], [9].

In the following, we will give a brief survey of the pilot studies on cross-domain CF in two dimensions, as shown in Table I: 1) Collaborative Filtering Domains (Section II) and 2) Knowledge Transfer Styles (Section III). Some possible extensions for cross-domain CF will be discussed in the conclusion (Section IV).

II. COLLABORATIVE FILTERING DOMAINS

A. System Domain

A common cross-domain CF scenario is that the data in the target recommender system (e.g., a new book website) are very sparse while the data in some related recommender systems are abundant (e.g., a popular movie website). In such cases, each recommender system is viewed as a "domain" and knowledge can be transferred over related system domains. A system domain is further decomposed into two sub-domains: user domain and item domain. Thus, knowledge can be transferred at either the user side or the item side or the both sides. Normally, the user/item sets in related system domains are not exactly the same but, in some cases, may have partial correspondence. The rationale for knowledge transfer over multiple user/item domains is that, in related system domains, users can be related in interests and items can be related in attributes. Example works in this problem setting include [2], [3], [5], [6], [9].

B. Data Domain

The second scenario is that multiple data sources have heterogeneous data types and each data source is viewed as a "domain". This problem setting is based on the assumption that some data types in the auxiliary domains (e.g., clickthrough data $0/1$ and polarity data $+/-$) can be obtained more easily than the data type in the target domain (e.g., five-star rating data $1 \sim 5$). Normally, this problem setting requires the user/item sets in different data domains to be the same. Then the knowledge can be discovered and transferred by finding relationships between data domains. Example works in this problem setting include [4], [8].

C. Temporal Domain

Another scenario can be presented when the rating data have time-stamps. By splitting the entire time-span of ratings into a number of time-slices, each time-slice can be viewed as a "domain". This problem setting is useful for capturing user-interest dynamics. In doing so, a user in different temporal domains can be treated as a set of user-counterparts with similar but different interests. Then the knowledge can be transferred along successive temporal domains to benefit the user-counterparts in each domain, which have very few data. An example in this problem setting is [7].

III. KNOWLEDGE TRANSFER STYLES

A. Rating-Pattern Sharing

Rating-pattern sharing was first proposed in [2], where it is also called CodeBook Transfer (CBT), for solving adaptive transfer learning (domain adaptation) problems in CF. Then the idea was incorporated into a probabilistic model, Rating-Matrix Generative Model (RMGM) [3], for solving collective transfer learning (multi-task learning) problems in CF. Both CBT and RMGM are cross-domain CF methods

1082-3409/11 $26.00 © 2011 IEEE
DOI 10.1109/ICTAI.2011.184

Table I
A CATEGORIZATION FOR CROSS-DOMAIN COLLABORATIVE FILTERING.

	System Domain	**Data Domain**	**Temporal Domain**
Rating-Pattern Sharing (Share $\mathbf{B}$, $\mathbf{X}^{(d)} = \mathbf{P}^{(d)}\mathbf{B}[\mathbf{Q}^{(d)}]^\top$)	CBT [2], RMGM [3]	-	RMGM-OT [7]
Latent-Feature Sharing (Share **U**/**V**, $\mathbf{X}^{(d)} = \mathbf{U}\mathbf{S}^{(d)}\mathbf{V}^\top$)	-	CST [4], TCF [8]	-
Domain Correlating (Correlate $\mathbf{F}^{(d)}$/$\mathbf{G}^{(d)}$, $\mathbf{X}^{(d)} = \mathbf{F}^{(d)}[\mathbf{G}^{(d)}]^\top$)	CLP [5], MCF [6], TagCDCF [9]	-	-

over system domains. More recently, this knowledge transfer style was adapted to temporal domains in [7].

General Framework: Given a number of rating matrices $\{\mathbf{X}^{(d)}\}$ for cross-domain CF, we factorize each rating matrix by $\mathbf{X}^{(d)} = \mathbf{P}^{(d)}\mathbf{B}[\mathbf{Q}^{(d)}]^\top$, where $\mathbf{P}^{(d)}$ ($\mathbf{P}^{(d)} \geq 0, \mathbf{P}^{(d)}\mathbf{1} = \mathbf{1}$) is a user-group membership matrix, $\mathbf{Q}^{(d)}$ ($\mathbf{Q}^{(d)} \geq 0, \mathbf{Q}^{(d)}\mathbf{1} = \mathbf{1}$) is an item-group membership matrix, and $\mathbf{B}$ is a group-level rating matrix. This approach matches user/item groups from multiple CF domains and encodes group-level rating patterns in a shared codebook $\mathbf{B}$. Then the knowledge can be transferred across domains through the shared $\mathbf{B}$ (i.e., the *bridge*).

B. Latent-Feature Sharing

The idea of latent-feature sharing has been widely adopted in relational learning (e.g., collective matrix factorization). This knowledge transfer style was introduced into cross-domain CF via a method named Coordinate System Transfer (CST) [4] over multiple data domains. Later, another setting of multiple data domains was studied in [8].

General Framework: Given a number of rating matrices $\{\mathbf{X}^{(d)}\}$ for cross-domain CF, we factorize each rating matrix by $\mathbf{X}^{(d)} = \mathbf{U}\mathbf{S}^{(d)}\mathbf{V}^\top$, where $\mathbf{U}$ ($\mathbf{U}^\top\mathbf{U} = \mathbf{I}$) and $\mathbf{V}$ ($\mathbf{V}^\top\mathbf{V} = \mathbf{I}$) are the bases for the user and the item latent feature spaces, respectively, and $\mathbf{S}$ is a scaling and rotation matrix. This approach shares user/item latent feature spaces across CF domains and the knowledge can be transferred through the shared $\mathbf{U}$ and $\mathbf{V}$ (i.e., the *bridge*).

C. Domain Correlating

The idea of exploiting correlations among CF domains was first mentioned in [10], but no solution was provided therein. Some methods based on this idea were proposed later, including Collective Link Prediction (CLP) [5] and Multi-domain Collaborative Filtering (MCF) [6]. Both CLP and MCF explore domain correlations via learning. More recently, a method resorts to estimating domain correlations based on explicit cues (tag similarities) [9].

General Framework: Given a number of rating matrices $\{\mathbf{X}^{(d)}\}$ for cross-domain CF, we factorize each rating matrix by $\mathbf{X}^{(d)} = \mathbf{F}^{(d)}[\mathbf{G}^{(d)}]^\top$, where $\mathbf{F}^{(d)}$ and $\mathbf{G}^{(d)}$ are the user and the item latent features, respectively. This approach tries to explore the correlations between user domains $\{\mathbf{F}^{(d)}\}$ and/or item domains $\{\mathbf{G}^{(d)}\}$ and the knowledge can be transferred across domains through the correlation matrices $Corr(\{\mathbf{F}^{(d)}\})$ and $Corr(\{\mathbf{G}^{(d)}\})$ (i.e., the *bridge*).

IV. CONCLUSION

We have given a brief survey of the pilot studies on cross-domain CF and summarize the related works in Table I. There are some blank cells in Table I that suggest possible extensions. One possible extension is to apply "Rating-Pattern Sharing" to "Data Domain", for example, using the rating patterns in binary data domains as prior knowledge to help learn five-star rating patterns. Another possible extension is to apply "Domain Correlating" to "Temporal Domain" by exploring correlations between user/item-counterparts in successive temporal domains.

ACKNOWLEDGMENT

This work was supported by Australian Research Council's Discovery Project No. DP1093762.

REFERENCES

[1] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in Artificial Intelligence*, 2009.

[2] B. Li, Q. Yang, and X. Xue, "Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction," in *IJCAI*, 2009, pp. 2052–2057.

[3] ——, "Transfer learning for collaborative filtering via a rating-matrix generative model," in *ICML*, 2009, pp. 617–624.

[4] W. Pan, E. W. Xiang, N. N. Liu, and Q. Yang, "Transfer learning in collaborative filtering for sparsity reduction," in *AAAI*, 2010, pp. 230–235.

[5] B. Cao, N. N. Liu, and Q. Yang, "Transfer learning for collective link prediction in multiple heterogenous domains," in *ICML*, 2010, pp. 159–166.

[6] Y. Zhang, B. Cao, and D.-Y. Yeung, "Multi-domain collaborative filtering," in *UAI*, 2010, pp. 725–732.

[7] B. Li, X. Zhu, R. Li, C. Zhang, X. Xue, and X. Wu, "Cross-domain collaborative filtering over time," in *IJCAI*, 2011, pp. 2293–2298.

[8] W. Pan, N. N. Liu, E. W. Xiang, and Q. Yang, "Transfer learning to predict missing ratings via heterogeneous user feedbacks," in *IJCAI*, 2011, pp. 2318–2323.

[9] Y. Shi, M. Larson, and A. Hanjalic, "Tags as bridges between domains: Improving recommendation with tag-induced cross-domain collaborative filtering," in *UMAP*, 2011, pp. 305–316.

[10] S. Berkovsky, T. Kuflik, and F. Ricci, "Cross-domain mediation in collaborative filtering," in *UM*, 2007, pp. 355–359.