Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Little Is Much: Bridging Cross-Platform Behaviors through Overlapped Crowds

Meng Jiang, Peng Cui
Tsinghua University

Nicholas Jing Yuan, Xing Xie
Microsoft Research Asia

Shiqiang Yang
Tsinghua University

Abstract

People often use multiple platforms to fulfill their different information needs. With the ultimate goal of serving people intelligently, a fundamental way is to get comprehensive understanding about user needs. How to organically integrate and bridge cross-platform information in a human-centric way is important. Existing transfer learning assumes either fully-overlapped or non-overlapped among the users. However, the real case is the users of different platforms are partially overlapped. The number of overlapped users is often small and the explicitly known overlapped users is even less due to the lacking of unified ID for a user across different platforms. In this paper, we propose a novel *semi-supervised transfer learning* method to address the problem of *cross-platform behavior prediction*, called XPTRANS. To alleviate the sparsity issue, it fully exploits the small number of overlapped crowds to optimally bridge a user's behaviors in different platforms. Extensive experiments across two real social networks show that XPTRANS significantly outperforms the state-of-the-art. We demonstrate that by fully exploiting 26% overlapped users, XPTRANS can predict the behaviors of non-overlapped users with the same accuracy as overlapped users, which means the small overlapped crowds can successfully bridge the information across different platforms.

Introduction

An information platform acts as an information source to inform people about something or provide knowledge about it. The diversity of people's information needs intrinsically determine the multiplicity of platforms that people engage in. It is common for a user to watch videos in YouTube, browse images in Flickr and share social information in Facebook. The contents that the user interact with across different platforms often have explicit/implicit correlations, or correspond to different aspects of his/her needs. With the ultimate goal of serving people intelligently, a fundamental way is to get comprehensive understanding about user needs. However, the current information platforms are either isolated or their correlations are significantly undermined. How to organically integrate and bridge cross-platform information in a human-centric way is of paramount significance for better satisfying users' information seeking, or more broadly, maximizing the potential value of the big data isolated in different platforms. More specifically, we are interested in the problem of *cross-platform behavioral prediction*: how to predict user behaviors in one platform (e.g., e-commerce, health care) based on user behaviors in other platforms (e.g., social media, wearable sensor)?

A paucity of research works have investigated cross-domain behavior prediction problem based on transfer learning. CODEBOOK (Li, Yang, and Xue 2009a) assumes the auxiliary platform and target platform without overlapped users, Netflix and MovieLens, share the same user-item rating patterns. TPCF (Jing et al. 2014) integrates three types of data from auxiliary domains including (1) data with aligned users, (2) data with aligned items, and (3) homogeneous data that have the same rating scale as the target domain, but without the knowledge from the correspondences of users and items. The previous works assume either a fully-overlapped shape (Zhong, Fan, and Yang 2014) or a non-overlapped shape (Li, Yang, and Xue 2009a) among the users across different platforms or domains. However, the real case is in the middle, where there are some common users (partially overlapped users) across the different platforms. The number of overlapped users is often small, and the number of explicitly known overlapped users is even less, due to the lacking of unified ID for a user across different platforms. Thus the essential problem in cross-platform behavioral prediction is how to fully exploit the small number of overlapped crowds to optimally bridge a user's behaviors in different platforms. To this end, we entail several key challenges as follow:

- *Sparsity:* Users often adopt only a small portion of items in one platform. On average, a Douban user generates ratings for 60 books, 200 movies, and 100 songs from 50,000 items. A Weibo user has 4.5 tags from 10,000 and tweets about 5,000 entities from 100,000.
- *Heterogeneity:* Behavioral data sets are heterogeneous with different types of items and links. The weights of links can be integer values, non-negative integers, or floats: We have ratings from Douban, "Like" number from social media, review emoticons from e-commerce, running distance and blood pressure from wearable devices.
- *Different representations:* Rating, tweeting and purchasing behaviors should have different patterns. Users' behavioral patterns cannot be represented in the same la-

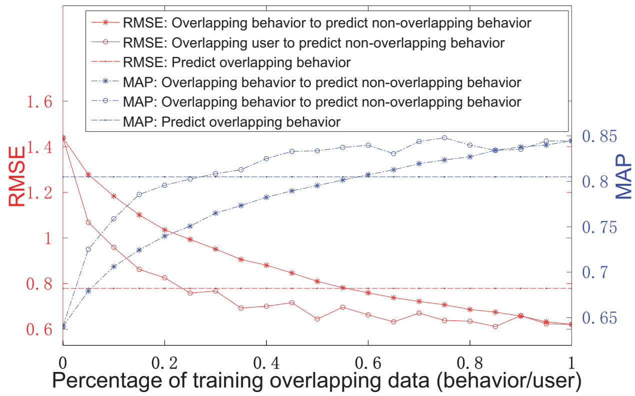


Figure 1: Exploiting 26% of active overlapped users between Weibo and Douban, XPTRANS can predict the movie rating behaviors of non-overlapped users with the same accuracy as overlapped users.

tent space, which is one of the points that make our work unique from traditional methods.

- *Partially overlapped crowds:* The natural bridge across platforms is overlapped users whose interests, tastes and personality are consistent. How to make full use of partially overlapped users' correspondences is still an open and challenging problem.

To address the above challenges, we propose a novel *semi-supervised transfer learning* method, called XPTRANS, based on matrix factorization that has been well adopted to represent heterogeneous data. First, XPTRANS jointly optimizes users' latent features on different platforms to alleviate the sparsity issue. Second, the latent space for user representation in one platform should be different from another. For different platforms, XPTRANS allows different settings for latent dimensions. From data study, we find that the similarities between overlapped users are consistent across different platforms. Thus, our method uses constraints of pairwise similarity to bring better flexibility for representations.

Figure 1 showcases the insights from the experiments on real data: XPTRANS predicts Douban users' movie rating behaviors by transferring tweet data from Sina Weibo. Exploiting more active overlapped user or more percentages of overlapping behavioral data, the prediction error consistently decreases and accuracy increases. We demonstrate that by fully exploiting 26% overlapped users, XPTRANS can predict the behaviors of non-overlapped users (who generated few records on one platform) with the same accuracy as the predictions on overlapped users (who generated more records on two platforms), which means the small overlapped crowds can successfully bridge the information across different platforms. Therefore, the *little* percentage of overlapped crowds is of *much* significance for cross-platform behavioral prediction.

Our contributions in this paper are as follows:

- We propose Cross-Platform Behavior Prediction as an open, challenging and promising problem for both the industry and research communities.
- We analyze users' behaviors across two real social networking platforms, which provides insights on exploiting overlapped crowds to alleviate sparsity issue.
- We develop a *semi-supervised transfer learning* method XPTRANS to predict user behaviors across platforms. Extensive experiments show that XPTRANS significantly outperforms state-of-the-art transfer learning methods. We demonstrate that (1) with the overlapping population increasing, prediction performances on non-overlapped users' behavior consistently improve, and (2) a little percentage of overlapped crowd is significant for cross-platform behavioral prediction.

Table 1: Data statistics: The overlapping population between Douban and Sina Weibo is 32,868.

	#User	#Item	#Link
Book	30,536	212,835	1,877,069
Movie	40,246	64,090	8,087,364
Music	33,938	286,464	4,141,708
Social tag	2,721,365	10,176	12,328,272
Tweet entity	25,586	113,591	141,908,323

Data and Preliminary Study

Data sets

We study two social networking data sets of *overlapped users*, Douban and Sina Weibo, to demonstrate the challenge of sparsity in behavior prediction problem. We have three types of items in Douban: *book*, *movie* and *music*; and we have two types in Weibo: *social tag* and *tweet entity*. We identified the overlapped users with their log-in accounts. Table 1 lists the data statistics. We observe that the sparsity of the five user-item matrices is serious but of different degrees: 95% in entity domain, 99.7% in movie domain, and 99.96-99.97% in book, music and tag domains.

Preliminary Study

Our idea of cross-platform behavioral modeling is to connect behavioral data from different platforms with overlapped crowds. We study the data sets to demonstrate the idea by answering the following two questions.

Q1: Do overlapped users have impacts on connecting latent spaces of different platforms?

A1: Yes! Given two platforms, if we use a user-based vector to represent an item on a platform, and on the other platform, we find one item that has a big value of user-based similarity with it, we can connect them. However, when there is NO overlapped user, the similarity is zero; with the percentage of overlapped users increasing, we may find a better similar item (of the maximum value of user-based similarity). For every two domains, Figure 2 plots curves of the percentage vs the similarity. We observe that

- When two platforms have more overlapping users, the value of cross-platform user-based similarity consistently increases (solid lines).

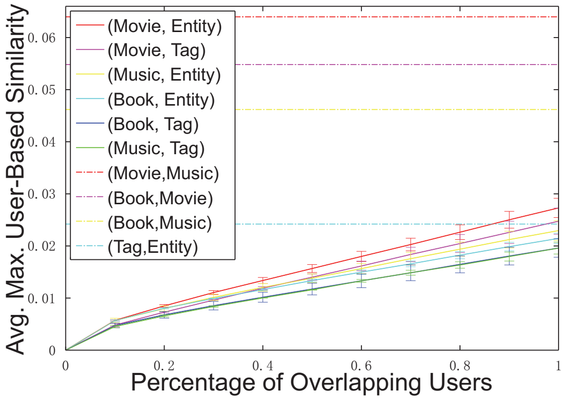


Figure 2: Overlapped users have significant impacts on connecting two platforms: (1) Cross-platform user-based similarity is smaller than within-platform similarity. (2) With more overlapped users, the similarity consistently increases.

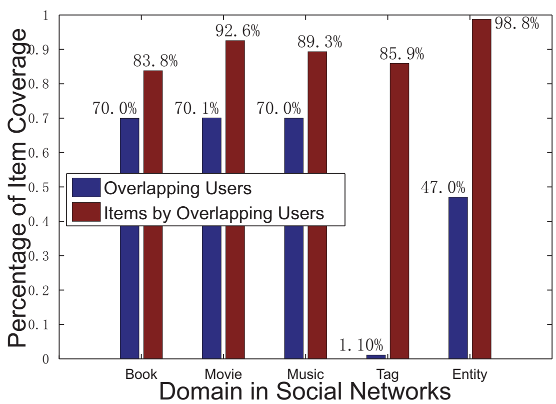


Figure 3: High coverage by overlapping users: they adopt >80% items in every domain, though only 1.1% users have accounts on both social networks.

- When few overlapped crowds exist, one more will make a big difference: The similarity increases faster when the percentage of overlapped users is between 0 to 10%.
- Cross-platform transfer is more difficult than with-platform cross-domain transfer. Cross-platform user-based similarity is smaller than within-platform similarity (dotted lines).

Q2: Can a small percentage of overlapped users cover most items on every platform?

A2: Yes! Figure 3 is a bar plot of the coverage of users and items in each domain on the two platforms. Though the percentage of overlapped users over the total population is very small, they adopt more than 80% items in every domain. Surprisingly, only 1.1% of the users who adopt at least one tag on Weibo have Douban accounts, they use as many as 85.9% of the tag set.

Cross-Platform Behavior Prediction

We first give the definition of cross-platform behavior prediction problem and then propose a novel semi-supervised transfer learning method and a scalable algorithm.

Notations and Problem Definition

First we introduce some symbols and notations that will be used through out the paper. Suppose we have two platforms P and Q. Platform P (Q) has m_P (m_Q) users and n_P (n_Q) items. We denote by r_P and r_Q the number of user/item clusters, i.e., the scale of latent spaces. We have user-item (rating) matrix $\mathbf{R}^{(P)} \in \mathbb{R}^{m_P \times n_P}$ and $\mathbf{R}^{(Q)} \in \mathbb{R}^{m_Q \times n_Q}$. The corresponding observation indicator matrices are $\mathbf{W}^{(P)}$ and $\mathbf{W}^{(Q)}$. Then we have user and item clustering matrices $\mathbf{U}^{(P)} \in \mathbb{R}^{m_P \times r_P}$ and $\mathbf{U}^{(Q)} \in \mathbb{R}^{m_Q \times r_Q}$ and $\mathbf{V}^{(P)} \in \mathbb{R}^{r_P \times n_P}$, $\mathbf{V}^{(Q)} \in \mathbb{R}^{r_Q \times n_Q}$. We denote by $\mathbf{W}^{(P,Q)} \in \mathbb{R}^{m_P \times m_Q}$ the user indicating matrix of the overlapped users where the entries are filled with either 1 or 0, depending on the availability. Now we can provide the problem definition as follows.

Problem 1 (Cross-platform behavior prediction (XPBP)) ***Given*** *a target platform P and an auxiliary platform Q, the user-item matrices $\mathbf{R}^{(P)}$ and $\mathbf{R}^{(Q)}$, the observation binary weights $\mathbf{W}^{(P)}$ and $\mathbf{W}^{(Q)}$, the overlapping user indicator matrix $\mathbf{W}^{(P,Q)}$,* ***find*** *the user clustering matrices $\mathbf{U}^{(P)}$ and $\mathbf{U}^{(Q)}$, the item clustering matrices $\mathbf{V}^{(P)}$ and $\mathbf{V}^{(Q)}$,* ***predict*** *missing values in $\mathbf{R}^{(P)}$.*

XPTrans: Semi-supervised Transfer Learning

To solve the above problem, we consider joint nonnegative matrix factorization (NMF) of (1) the term of behavioral data on the target platform P, (2) the term of behaviors on the auxiliary platform Q, and (3) the term that uses overlapping user indicator $\mathbf{W}^{(P,Q)}$ to supervise user-based similarity across platforms. This leads to the following optimization problem, which is to minimize

$$\begin{aligned}\mathcal{J} \quad = \quad & \textstyle\sum_{i,j} W_{i,j}^{(P)}\left(R_{i,j}^{(P)} - \sum_r U_{i,r}^{(P)} V_{r,j}^{(P)}\right)^2 \\ & + \lambda \textstyle\sum_{i,j} W_{i,j}^{(Q)}\left(R_{i,j}^{(Q)} - \sum_r U_{i,r}^{(Q)} V_{r,j}^{(Q)}\right)^2 \\ & + \mu \textstyle\sum_{i_1,j_1,i_2,j_2} W_{i_1,j_1}^{(P,Q)} W_{i_2,j_2}^{(P,Q)} \left(A_{i_1,i_2}^{(P)} - A_{j_1,j_2}^{(Q)}\right)^2, \quad (1)\end{aligned}$$

where we denote by $A_{i_1,i_2}^{(P)}$ the user-based similarity between u_{i_1} and u_{i_2} on platform P, and denote by $A_{j_1,j_2}^{(Q)}$ the user similarity between u_{j_1} and u_{j_2} on platform Q:

$$A_{i_1,i_2}^{(P)} = \sum_{r=1}^{r_P} U_{i_1,r}^{(P)} U_{i_2,r}^{(P)}, \; A_{j_1,j_2}^{(Q)} = \sum_{r=1}^{r_Q} U_{j_1,r}^{(Q)} U_{j_2,r}^{(Q)}.$$

Eq. (1) can be represented into NMF. We are minimizing

$$\begin{aligned}\mathcal{J} = & ||\mathbf{W}^{(P)} \odot (\mathbf{R}^{(P)} - \mathbf{U}^{(P)}\mathbf{V}^{(P)})||_F^2 \\ & + \lambda ||\mathbf{W}^{(Q)} \odot (\mathbf{R}^{(Q)} - \mathbf{U}^{(Q)}\mathbf{V}^{(Q)})||_F^2 \\ & + \mu (||\mathbf{W}^{(P,Q)}\mathbf{1}^{(Q)}\mathbf{W}^{(P,Q)\top} \odot \mathbf{U}^{(P)}\mathbf{U}^{(P)\top} \odot \mathbf{U}^{(P)}\mathbf{U}^{(P)\top}|| \\ & + ||\mathbf{W}^{(P,Q)\top}\mathbf{1}^{(P)}\mathbf{W}^{(P,Q)} \odot \mathbf{U}^{(Q)}\mathbf{U}^{(Q)\top} \odot \mathbf{U}^{(Q)}\mathbf{U}^{(Q)\top}|| \\ & - 2||\mathbf{U}^{(P)}\mathbf{U}^{(P)\top}\mathbf{W}^{(P,Q)}\mathbf{U}^{(Q)}\mathbf{U}^{(Q)\top}\mathbf{W}^{(P,Q)\top}||) \quad (2) \\ & s.t. \;\; \mathbf{U}^{(P)} > 0, \mathbf{V}^{(P)} > 0, \mathbf{U}^{(Q)} > 0, \mathbf{V}^{(Q)} > 0,\end{aligned}$$

where $\mathbf{1}^{(P)} \in \mathbb{R}^{m_P \times m_P}$ and $\mathbf{1}^{(Q)} \in \mathbb{R}^{m_Q \times m_Q}$ are filled with all 1s. Moreover, λ is a trade-off parameter determining the importance of knowledge transfer from auxiliary platform Q to target platform P, μ is a parameter determining the importance of the supervised term, $\odot$ is the Hadamard product, $||\cdot||$ is the 1-norm as a vector norm and $||\cdot||_F$ is the Frobenius norm. We use $\mathcal{L}_2$ norms of $\mathbf{U}^{(P)}$, $\mathbf{V}^{(P)}$, $\mathbf{U}^{(Q)}$, and $\mathbf{V}^{(Q)}$ as regularization terms but omit them for space.

As in the standard NMF (Lee, Yoo, and Choi 2010), the gradients of $\mathbf{U}^{(P)}$, $\mathbf{U}^{(Q)}$, $\mathbf{V}^{(P)}$ and $\mathbf{V}^{(Q)}$ to minimize Eq. (2) can be derived easily:

$$\begin{aligned}
\frac{\partial J}{\partial \mathbf{U}^{(P)}} &= -2[\mathbf{W}^{(P)} \odot (\mathbf{R}^{(P)} - \mathbf{U}^{(P)}\mathbf{V}^{(P)})]\mathbf{V}^{(P)\top} \\
&\quad +4\mu[\mathbf{W}^{(P,Q)}\mathbf{1}^{(Q)}\mathbf{W}^{(P,Q)\top} \odot \mathbf{U}^{(P)}\mathbf{U}^{(P)\top}]\mathbf{U}^{(P)} \\
&\quad -4\mu[\mathbf{W}^{(P,Q)}\mathbf{U}^{(Q)}\mathbf{U}^{(Q)\top}\mathbf{W}^{(P,Q)\top}]\mathbf{U}^{(P)}, \qquad (3)\\
\frac{\partial J}{\partial \mathbf{U}^{(Q)}} &= -2\lambda[\mathbf{W}^{(Q)} \odot (\mathbf{R}^{(Q)} - \mathbf{U}^{(Q)}\mathbf{V}^{(Q)})]\mathbf{V}^{(Q)\top} \\
&\quad +4\mu[\mathbf{W}^{(P,Q)\top}\mathbf{1}^{(P)}\mathbf{W}^{(P,Q)} \odot \mathbf{U}^{(Q)}\mathbf{U}^{(Q)\top}]\mathbf{U}^{(Q)} \\
&\quad -4\mu[\mathbf{W}^{(P,Q)\top}\mathbf{U}^{(P)}\mathbf{U}^{(P)\top}\mathbf{W}^{(P,Q)}]\mathbf{U}^{(Q)}, \qquad (4)\\
\frac{\partial \mathcal{J}}{\partial \mathbf{V}^{(P)}} &= -2\mathbf{U}^{(P)\top}[\mathbf{W}^{(P)} \odot (\mathbf{R}^{(P)} - \mathbf{U}^{(P)}\mathbf{V}^{(P)})], \qquad (5)\\
\frac{\partial \mathcal{J}}{\partial \mathbf{V}^{(Q)}} &= -2\mathbf{U}^{(Q)\top}[\mathbf{W}^{(Q)} \odot (\mathbf{R}^{(Q)} - \mathbf{U}^{(Q)}\mathbf{V}^{(Q)})]. \qquad (6)
\end{aligned}$$

A general cross-platform representation: Now we extend XPTRANS into a general case. We denote by K the number of platforms and by $\mathbf{R}^{(k)} \in \mathbb{R}^{m_k \times n_k}$ the user-item matrix on the k^{th} platform, where m_k (n_k) is the number of users (items). Then we have user and item clustering matrices $\mathbf{U}^{(k)} \in \mathbb{R}^{m_k \times r_k}$, $\mathbf{V}^{(k)} \in \mathbb{R}^{r_k \times n_k}$, and we have observation binary matrices $\mathbf{W}^{(k)} \in \mathbb{R}^{m_k \times n_k}$. We denote by $\mathbf{W}^{(k,k')} \in \mathbb{R}^{m_k \times m_{k'}}$ the partially overlapping user indicator matrix between the k-th and k'-th platforms. Now the objective function is to minimize

$$\begin{aligned}
\mathcal{J} &= \textstyle\sum_k \lambda_k \sum_{i,j} W_{i,j}^{(k)} \left(R_{i,j}^{(k)} - \sum_r U_{i,r}^{(k)} V_{r,j}^{(k)}\right)^2 \qquad (7)\\
&\quad + \textstyle\sum_{(k,k')} \mu_{k,k'} \sum_{i_1,j_1,i_2,j_2} W_{i_1,j_1}^{(k,k')} W_{i_2,j_2}^{(k,k')} \left(A_{i_1,i_2}^{(k)} - A_{j_1,j_2}^{(k')}\right)^2,
\end{aligned}$$

where we have

$$A_{i_1,i_2}^{(k)} = \sum_{r=1}^{r_k} U_{i_1,r}^{(k)} U_{i_2,r}^{(k)};\ A_{j_1,j_2}^{(k')} = \sum_{r=1}^{r_{k'}} U_{j_1,r}^{(k')} U_{j_2,r}^{(k')},$$

and λ_k is the importance of user behavioral data from the k-th platform, and $\mu_{k,k'}$ is the similarity of overlapping users' behavioral patterns on the k-th and k'-th platforms.

Derivatives of the above error function with respect to $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k)}$ are given by

$$\begin{aligned}
\frac{\partial J}{\partial \mathbf{U}^{(k)}} &= -2\lambda_k[\mathbf{W}^{(k)} \odot (\mathbf{R}^{(k)} - \mathbf{U}^{(k)}\mathbf{V}^{(k)})]\mathbf{V}^{(k)\top} \\
&\quad +4\textstyle\sum_{k'} \mu_{k,k'}[\mathbf{W}^{(k,k')}\mathbf{1}^{(k')}\mathbf{W}^{(k,k')\top} \odot \mathbf{U}^{(k)}\mathbf{U}^{(k)\top}]\mathbf{U}^{(k)} \\
&\quad -4\textstyle\sum_{k'} \mu_{k,k'}[\mathbf{W}^{(k,k')}\mathbf{U}^{(k')}\mathbf{U}^{(k')\top}\mathbf{W}^{(k,k')\top}]\mathbf{U}^{(k)} \qquad (8)\\
\frac{\partial J}{\partial \mathbf{V}^{(k)}} &= -2\lambda_k\mathbf{U}^{(k)\top}[\mathbf{W}^{(k)} \odot (\mathbf{R}^{(k)} - \mathbf{U}^{(k)}\mathbf{V}^{(k)})], \qquad (9)
\end{aligned}$$

where $\mathbf{1}^{(k')} \in \mathbb{R}^{m_{k'} \times r_{k'}}$ is filled with all 1s.

XPTRANS **algorithm and complexity:** The basic algorithm procedure for solving Eq. (7) is shown in Algorithm 1. The computational complexity of XPTRANS is $\mathcal{O}(\sum_k m_k n_k r_k + \sum_{k,k'}(m_k m_{k'}(m_k + m_{k'} + r_{k'}) + m_k^2(r_k + r_{k'})))$. Since $n_k, m_k, m_{k'} \gg r_k, r'_k$ (constant), we know it reduces down to cubic time $\mathcal{O}(m(m^2+n))$.

Algorithm 1 XPTRANS: Semi-supervised transfer learning for cross-platform behavior prediction

Require: user-item matrix $\mathbf{R}^{(k)}$, observation binary matrix $\mathbf{W}^{(k)}$, overlapping user indicator $\mathbf{W}^{(k,k')}$

1: Initialize $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k)}$.
2: Repeat the following steps until convergence: (a) fixing $\mathbf{V}^{(k)}$, updating $\mathbf{U}^{(k)}$ by rule in Eq. (8); (b) fixing $\mathbf{U}^{(k)}$, updating $\mathbf{V}^{(k)}$ by rule in Eq. (9).
3: **return** $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k)}$, for $k = 1, ..., K$

Experiments

In this section we will present the empirical study on the XPTRANS method for cross-platform behavior prediction.

Experimental Settings

Data sets: User behaviors (e.g., ratings, adoptions) are often *sparse* on every platform (see Table 1). We use the Sina Weibo (tag, tweet entity) and Douban (book, movie, music) data sets in our experiments. Our extensive experiments have been conducted on 12 (target,auxiliary)-platform assignments to test the prediction performance. For example,

- From tweet entity to movie: Can we transfer user interests (e.g., political topics) from Weibo network to predict what movie (e.g., *House of Cards*) the user really likes?
- From book to tag: Can we predict users' social tags (e.g., *science freak*) with their tastes on Douban books (e.g., *A Brief History of Time*)?

Algorithms: We implement the following two variants of our proposed method and baseline methods:

- XPTRANS: It uses overlapping crowds as supervisory knowledge. The function is shown in Eq. (2).
- CMF (Collective Matrix Factorization with regularization) (Singh and Gordon 2008): It does not transfer knowledge from the auxiliary platform.
- CBT (Codebook Transfer) (Li, Yang, and Xue 2009a): It transfers more useful knowledge through cluster-level user-item rating patterns, called "codebook". However, it does not use the supervisory overlapping part.
- XPTRANS-ALIGN: It is a variant of our cross-platform method with too strong assumption. It assumes that the behavioral patterns have the same representation across platforms ($r = r_P = r_Q$). So, the latent features of overlapping users can be aligned in the same space. This object function replaces the 3^{rd} term in Eq. (2) with

$$\mu \textstyle\sum_{i,j} W_{i,j}^{(P,Q)} \sum_r \left(U_{i,r}^{(P)} - U_{j,r}^{(Q)}\right)^2. \qquad (10)$$

Training, testing and auxiliary parameters: We run hold-out experiments to test the performance of walking across platforms. We set the percentage of *training* behavioral entries in $\mathbf{R}^{(P)}$ by *non-overlapping* users as 70%, the percentage of *auxiliary* behavioral entries in $\mathbf{R}^{(Q)}$ by *non-overlapping* users as 70%, and the other two parameters:

- $\alpha_{\mathbf{R}}^{(P\cap Q)} \in [0, 100\%]$: the percentage of *overlapping behavioral entries* in $\mathbf{R}^{(P)}$ and $\mathbf{R}^{(Q)}$;

Table 2: XPTRANS outperforms the state-of-the-art methods in predicting cross-platform behaviors.

	Q: Weibo tweet entity $\rightarrow$ P: Douban movie				Q: Douban book $\rightarrow$ P: Weibo tag			
	RMSE		MAP		RMSE		MAP	
	$P \cap Q$	$P \setminus Q$	$P \cap Q$	$P \setminus Q$	$P \cap Q$	$P \setminus Q$	$P \cap Q$	$P \setminus Q$
CMF (Singh and Gordon 2008)	1.379	1.439	0.651	0.640	0.418	0.429	0.477	0.464
CBT (Li, Yang, and Xue 2009a)	0.767	1.290	0.808	0.676	0.261	0.419	0.675	0.477
XPTRANS-ALIGN	0.757	1.164	0.811	0.702	0.256	0.411	0.681	0.487
XPTRANS	**0.715**	**0.722**	**0.821**	**0.820**	**0.236**	**0.374**	**0.705**	**0.533**
XPTRANS vs. CBT	↓6.8%	↓44.0%	↑1.62%	↑21.3%	↓9.6%	↓10.8%	↑4.5%	↑11.7%
XPTRANS vs. XPTRANS-ALIGN	↓5.5%	↓38.0%	↑1.3%	↑16.8%	↓8.0%	↓9.0%	↑3.6%	↑9.4%

- $\alpha_{\mathbf{U}}^{(P \cap Q)} \in [0, 100\%]$: the percentage of *the most active overlapping users* in $\mathbf{R}^{(P)}$ and $\mathbf{R}^{(Q)}$.

We conduct experiments for 10 times and report the average performance.

Evaluation: The results will be evaluated with two criteria: Root Mean Square Error (*RMSE*) and Mean Average Precision (*MAP*). We test prediction performance on two tasks: hold-out behaviors by *overlapping* users ($P \cap Q$) and by *non-overlapping* users ($P \setminus Q$). The performance is better if we spot smaller RMSE and higher MAP.

Experimental Results

Cross-platform prediction performance: Table 2 shows the RMSE and MAP of our proposed XPTRANS and baseline methods. Typically, we show the performances of transferring Weibo tweet entity to Douban movie and Douban book to Weibo social tag. Our XPTRANS improves the performance of predicting non-overlapping users' behavior. In predicting Douban movie ratings, XPTRANS reduces the RMSE by 44% over the state-of-the-arts and 38% over XPTRANS-ALIGN; it increases the MAP by 21% over CBT and 16.8% over XPTRANS-ALIGN. In predicting Weibo social tag, XPTRANS has 9.4% improvement. This result demonstrates the positive effects of the auxiliary knowledge and supervisory information. The assumption on user representations by CBT and XPTRANS-ALIGN are too strong: for a single user, the representations are the same in different platforms. The term of pair-wise similarity in XPTRANS provides more *flexibility* on the user representations. For new comer recommendation, if we focus on users whose movie-rating behaviors have been completely hidden, RMSE scores of CBT and XPTRANS are 1.737 and 1.569 (smaller is better). XPTRANS can jointly optimize (1) the users latent representations in $U^{(Q)}$ using $R^{(Q)}$ and (2) representations in $U^{(P)}$ using overlapped user similarity constraints, for a new comer in P who has behaviors in Q.

Figure 4 reports the RMSE decreasing percentage of every pair of target-auxiliary platforms. We observe that

- Consistently as in Table 2, the improvement on predicting non-overlapping users' behavior is higher than the improvement on predicting overlapping users' behavior. The sparsity problem of the non-overlapping users is more challenging and XPTRANS can alleviate the problem.
- Transferring Weibo's behavioral data to predict interest-driven item adoptions on Douban can reduce the RMSE by 38.0%, indicating that it is effective to incorporate knowledge from social platforms for content sharing.
- Using auxiliary book/movie/music data to predict social tags can reduce the RMSE by 9.0%, while the RMSE decrease of predicting tweet entity is small. Predicting tweet entity adoptions is such a difficult task that the performance cannot benefit much from cross-platform transfer.

The important role of overlapping users: When does the prediction on non-overlapping users' behaviors with cross-platform transfer achieve the same performance as the prediction on overlapping behaviors without the transfer? From Figure 1 we can spot that in transferring tweet entity data to predict movie, more overlapping users can better improve the performance than overlapping behaviors: using $\alpha_{\mathbf{U}}^{(P \cap Q)}$=26% of active overlapping users and using $\alpha_{\mathbf{R}}^{(P \cap Q)}$=60% of overlapping behaviors can make the performance of predicting non-overlapping users' behavior to reach the same RMSE and MAP as that of predicting overlapping users' behavior.

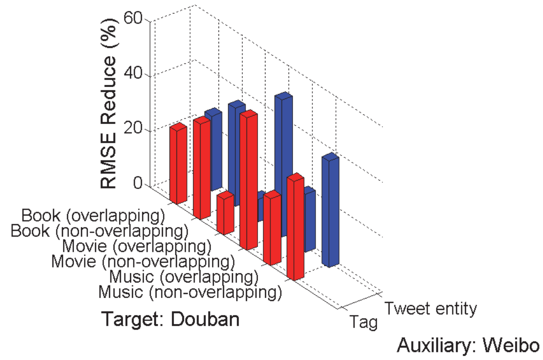
Related Works

Cross-Domain Collaborative Filtering

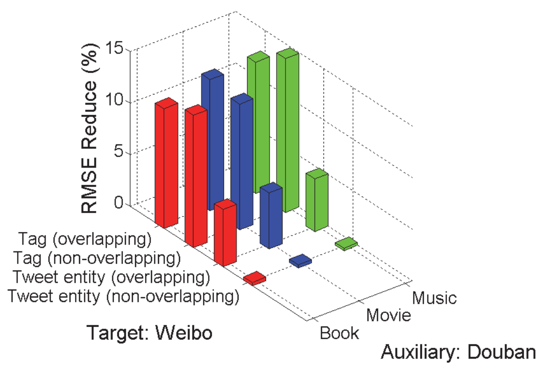
Adomavicius *et al.* surveyed the field of recommender systems, described limitations and discussed possible extensions (Adomavicius and Tuzhilin 2005). The next generation of recommender systems should learn across multiple data sources (Jiang et al. 2012)(Gao et al. 2013)(Montañez, White, and Huang 2014)(Jiang et al. 2015). Shi *et al.* proposed a generalized tag-induced CDCF approach to enjoy the benefit of using social tags as representatives of explicit links between domains (Shi, Larson, and Hanjalic 2011). They reviewed two categories of collaborative filtering beyond the user-item matrix: rich side information and interaction information (Shi, Larson, and Hanjalic 2014). However, for cross-platform scenarios, a user does not always perform on every platfrom. The number of overlapping users between the two million-user platforms are usually very small due to data access and user matching techniques. Cross-platform behavior modeling is still an open and challenging problem.

Transfer Learning for Behavior Prediction

Pan *et al.* surveyed the categories of transfer learning, with many works having attempted to further enrich the field (Pan



Weibo→Douban



Douban→Weibo

Figure 4: XPTRANS ($P \rightarrow Q$) significantly improves the performance in target platform Q via knowledge transfer from auxiliary platform P, especially for non-overlapping users.

and Yang 2010). Li *et al.* proposed a generative model to learn between multiple rating matrices. Li *et al.* proposed CODEBOOK (Li, Yang, and Xue 2009a) to transfer knowledge through a latent space, which assumed the auxiliary domain (Netflix) and target domain (MovieLens) shared the same user-item rating patterns(Li, Yang, and Xue 2009b). Yang *et al.* proposed *heterogeneous transfer learning* across user-tag and user-image networks (Yang et al. 2009). Chen *et al.* used tensor factorization method to fuse user, tag, and book/movie information in a single model (Chen, Hsu, and Lee 2013). Tan *et al.* proposed transfer learning with multiple views and multiple sources (Tan et al. 2014b). Lin *et al.* assumes that users from homogeneous data (two domains of the same rating scale) share the same prior parameters in their generative model (Jing et al. 2014). However, users have different types of behaviors and form different behavioral patterns on different platforms, which demands different representations (Pan et al. 2011)(Moreno et al. 2012)(Tan et al. 2014a). Our work aims at cross-platform data where the bridge is the overlapping population of users. The above transfer learning methods either align all users in the same space or assume that the two domains are independent. We have pointed out the significance of supervision of overlapping users for cross-platform study.

Semi-Supervised NMF

Recently, a number of approaches have studied how to conduct semi-supervised learning using non-negative matrix factorization (Koren 2008)(Singh and Gordon 2008)(Jiang et al. 2014)(Al-Shedivat et al. 2014). For multi-label learning, Liu *et al.* proposed constrained NMF to minimize the difference between input patterns and class memberships(Liu, Jin, and Yang 2006). For clustering problems, Li *et al.* used NMF to integrate various forms of background knowledge from distributed data sources into clustering (Li, Ding, and Jordan 2007). Wang *et al.* co-clustered data sets of different types with constraints in their matrix factorizations (Wang, Li, and Zhang 2008). Lee *et al.* conducted both document clustering and EEG classification tasks in the presented semi-supervised NMF method (Lee, Yoo, and Choi 2010). Inspired by the above methods and their practice in different applications, we consider using the overlapping population across platforms as supervisory information.

Conclusion

In this paper, we proposed cross-platform behavior prediction problem to alleviate sparsity with behavioral data from auxiliary platforms. We discovered that overlapped crowds had significant impacts on knowledge transfer across platforms. In response, we developed semi-supervised transfer learning method XPTRANS to exploit how the small number of the overlapped crowds can bridge user's behaviors in different platforms. Extensive experiments on two real social networks show that XPTRANS significantly outperforms the baseline transfer learning methods.

Acknowledgement

This work was supported by National Program on Key Basic Research Project, No. 2015CB352300; National Natural Science Foundation of China, No. 61370022, No. 61531006 and No. 61210008; International Science and Technology Cooperation Program of China, No. 2013DFG12870. Thanks for the support of NExT Research Center funded by MDA, Singapore, under the research grant, WBS:R-252-300-001-490 and the research fund of Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology.

References

Adomavicius, G., and Tuzhilin, A. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE TKDE* 17(6):734–749.

Al-Shedivat, M.; Wang, J. J.-Y.; Alzahrani, M.; Huang, J. Z.; and Gao, X. 2014. Supervised transfer sparse coding. In *AAAI*, 1665–1672.

Chen, W.; Hsu, W.; and Lee, M. L. 2013. Making recommendations from multiple domains. In *ACM SIGKDD*, 892–900.

Gao, S.; Luo, H.; Chen, D.; Li, S.; Gallinari, P.; and Guo, J. 2013. Cross-domain recommendation via cluster-level latent factor model. In *Machine Learning and Knowledge Discovery in Databases*. 161–176.

Jiang, M.; Cui, P.; Wang, F.; Yang, Q.; Zhu, W.; and Yang, S. 2012. Social recommendation across multiple relational domains. In *Proceedings of the 21st ACM international conference on Information and knowledge management (CIKM)*, 1422–1431. ACM.

Jiang, M.; Cui, P.; Wang, F.; Xu, X.; Zhu, W.; and Yang, S. 2014. Fema: flexible evolutionary multi-faceted analysis for dynamic behavioral pattern discovery. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (SIGKDD)*, 1186–1195. ACM.

Jiang, M.; Cui, P.; Chen, X.; Wang, F.; Zhu, W.; and Yang, S. 2015. Social recommendation with cross-domain transferable knowledge. *Knowledge and Data Engineering, IEEE Transactions on (TKDE)*.

Jing, H.; Liang, A.-C.; Lin, S.-D.; and Tsao, Y. 2014. A transfer probabilistic collective factorization model to handle sparse data in collaborative filtering. In *IEEE ICDM*.

Koren, Y. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *ACM SIGKDD*, 426–434.

Lee, H.; Yoo, J.; and Choi, S. 2010. Semi-supervised nonnegative matrix factorization. *IEEE Signal Processing Letters* 17(1):4–7.

Li, T.; Ding, C.; and Jordan, M. I. 2007. Solving consensus and semi-supervised clustering problems using nonnegative matrix factorization. In *IEEE ICDM*, 577–582.

Li, B.; Yang, Q.; and Xue, X. 2009a. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction. In *IJCAI*, volume 9, 2052–2057.

Li, B.; Yang, Q.; and Xue, X. 2009b. Transfer learning for collaborative filtering via a rating-matrix generative model. In *ICML*, 617–624.

Liu, Y.; Jin, R.; and Yang, L. 2006. Semi-supervised multi-label learning by constrained non-negative matrix factorization. In *NCAI*, volume 21, 421.

Montañez, G. D.; White, R. W.; and Huang, X. 2014. Cross-device search. In *ACM CIKM*, 1669–1678.

Moreno, O.; Shapira, B.; Rokach, L.; and Shani, G. 2012. Talmud: transfer learning for multiple domains. In *ACM CIKM*, 425–434.

Pan, S. J., and Yang, Q. 2010. A survey on transfer learning. *IEEE TKDE* 22(10):1345–1359.

Pan, W.; Liu, N. N.; Xiang, E. W.; and Yang, Q. 2011. Transfer learning to predict missing ratings via heterogeneous user feedbacks. In *IJCAI*, volume 22, 2318.

Shi, Y.; Larson, M.; and Hanjalic, A. 2011. Tags as bridges between domains: Improving recommendation with tag-induced cross-domain collaborative filtering. 305–316.

Shi, Y.; Larson, M.; and Hanjalic, A. 2014. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM CSUR* 47(1):3.

Singh, A. P., and Gordon, G. J. 2008. Relational learning via collective matrix factorization. In *ACM SIGKDD*, 650–658.

Tan, B.; Zhong, E.; Ng, M. K.; and Yang, Q. 2014a. Mixed-transfer: transfer learning over mixed graphs. In *SDM*.

Tan, B.; Zhong, E.; Xiang, E. W.; and Yang, Q. 2014b. Multi-transfer: Transfer learning with multiple views and multiple sources. *Statistical Analysis and Data Mining*.

Wang, F.; Li, T.; and Zhang, C. 2008. Semi-supervised clustering via matrix factorization. In *SDM*, 1–12.

Yang, Q.; Chen, Y.; Xue, G.-R.; Dai, W.; and Yu, Y. 2009. Heterogeneous transfer learning for image clustering via the social web. In *Joint Meeting of ACL, AFNLP*, 1–9.

Zhong, E.; Fan, W.; and Yang, Q. 2014. User behavior learning and transfer in composite social networks. *ACM TKDD* 8(1):6.