Unified YouTube Video Recommendation via Cross-network Collaboration

Ming Yan[1,2], Jitao Sang[1,2], Changsheng Xu[1,2]
[1]National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing 100190, China
[2]China-Singapore Institute of Digital Media, Singapore, 117867, Singapore
{ming.yan, jtsang, csxu}@nlpr.ia.ac.cn

ABSTRACT

The ever growing number of videos on YouTube makes recommendation an important way to help users explore interesting videos. Similar to general recommender systems, YouTube video recommendation suffers from typical problems like new user, cold-start, data sparsity, etc. In this paper, we propose a unified YouTube video recommendation solution via cross-network collaboration: users' auxiliary information on Twitter are exploited to address the typical problems in single network-based recommendation solutions. The proposed two-stage solution first transfers user preferences from auxiliary network by learning cross-network behavior correlations, and then integrates the transferred preferences with the observed behaviors on target network in an adaptive fashion. Experimental results show that the proposed cross-network collaborative solution achieves superior performance not only in term of accuracy, but also in improving the diversity and novelty of the recommended videos.

Categories and Subject Descriptors

H.3.5 [**Online Information Services**]: Web-based services

General Terms

Theory

Keywords

YouTube video recommendation; cross-network collaboration; user modeling

1. INTRODUCTION

With the emergence and popularity of social media, people now usually engage in disparate Online Social Networks (OSNs) simultaneously for different purposes [1]. For example, the same individual may communicate with his/her friends on Facebook, follow real-time hot events on Twitter, subscribe and watch videos on YouTube, share and discuss favorite restaurants on Yelp, etc. These *cross-network* activities together record people's integral online footprints and reflect their demographics as well as interests from different perspectives.

Typical social media services, however, are usually conducted on one OSN. For example, the video recommendation service on YouTube, has been one of the most important ways to lead users to their interested videos from the huge repository [2]. The limitation of single network-based solutions is that, the available user data on one OSN are usually not sufficient to understand user interests and capture the ever-changing user preferences. The notorious cold-start and sparsity issues have significantly hindered accurate user modeling and practical personalized social media services [3]. Therefore, our work is motivated to exploit the scattered user data in multiple OSNs towards improved personalized services on the target OSN. In this paper, we aim to leverage users' rich cross-network activity data to help estimate their video preferences on YouTube, and design a unified video recommendation solution.

The unified YouTube video recommendation solution is expected to address the following three problems: (1) *New user*. When a user newly registers to YouTube and starts using the recommender, the system has no knowledge of the user's interactions on the videos. In literatures, new users are either modeled using the limited registration information [4] or treated as the average users with recommendation of the most popular items [5]. (2) *Cold-start*. This is related to situations in which the recommender is unable to provide accurate recommendations due to an initial lack of user preferences. We refer to the users with few historical behavior records as *light user* [1]. Current solutions to facilitate the light users include relying on their content information (e.g., demographics and tagging) and resorting to a content-based recommendation solution [6, 7], or exploiting the available social relations as regularization to predict user preferences and warm up the recommender [8]. (3) *Sparsity*. In typical recommender systems, most users have no chance to browse or rate most items and hence the user-item interaction matrix is very sparse. This is especially the case in systems with a very high item-to-user ratio, e.g., YouTube, which has hosted over 2 billion videos. Efforts have been taken to alleviate the sparsity problem by filling the missing user-item entries with default values [9], utilizing latent factor models and projecting the users and items to a low-dimension space to capture the salient structure [10],

[1] Conversely, the users with a lot of behavior records are referred to as *heavy user*.

ICMR'15, June 23–26, 2015, Shanghai, China.

http://dx.doi.org/10.1145/2671188.2749344.

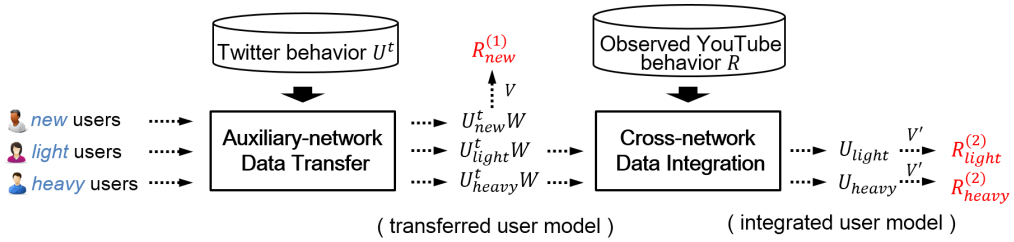

Figure 1: The proposed solution framework.

or discovering the high-order correlations between users by means of spread or iterative models [11].

New user, *cold-start*, and *sparsity* have been the most challenging problems in the field of recommender systems. Although received extensive attentions in the past decades, these problems still remain open. While the existing work propose to deal with one or two of the problems, a unified solution framework addressing all the three problems is unexplored. Moreover, most of the current efforts are devoted to designing advanced models to better exploit the limited and possibly unpromising data within the target OSN [2], but largely ignore the abundant user data available outside on the auxiliary OSNs. Recent practices on big data [12] have also suggested that "more data beats complex models". Therefore, in this work, we attempt to leverage more user data from Twitter, introduce a simple solution framework to simultaneously address all the mentioned problems, and benefit three types of typical YouTube users. Specifically, for *new users*, we estimate their video preferences on YouTube by analyzing the tweeting activities collected on Twitter, based on which an initial recommendation list is generated; for *light users*, we bootstrap the recommender by integrating the auxiliary information transferred from Twitter and the available information on YouTube; for *heavy users*, their recommendation is benefitted as a result of the reduced sparsity. The auxiliary information contributes to the correlation calculation between users.

The challenge lies in two-fold. (1) User data on different OSNs are heterogenous. There exist no explicit correlations between these cross-network user data. We cannot directly utilize the user tweeting data collected from Twitter to estimate the YouTube video preference. (2) For light and heavy users, the estimated preferences from Twitter and the observed behaviors on YouTube may contradict each other. It is critical to align the two potentially contradictory user models, and balance the contribution of either model in the final recommendation. To address the above challenges, as illustrated in Fig. 1, the proposed solution framework consists of two stages, i.e., auxiliary-network data transfer, and cross-network data integration. At the first stage, the correlations between the auxiliary-network and target-network behaviors are embedded in a transfer matrix, by which the users' tweeting activities on Twitter can be mapped to a latent user space on YouTube. With the derived transfer matrix, we can estimate a user's video preferences given his/her tweeting history. For new users, the recommender is ready to exploit the transferred video preferences to generate recommendations. At the second stage, viewing the transferred preference as a priori, we introduce a regularization-based formulation to integrate the two sources of user data. Moreover, a weighting matrix is added to adapt their contributions according to the amount of available YouTube data. In this way, the obtained user models for the light and heavy users consider both the Twitter tweeting activities and historical interactions with YouTube videos. A straightforward recommender can be designed to utilize the user model for recommendation.

We summarize the contributions of this work as follows:

1. We introduce a novel personalized recommendation solution by leveraging cross-network data. This is consistent with users' multi-OSN engagement phenomenon and entails user modeling from versatile aspects.

2. A unified video recommendation framework is presented, with goals to simultaneously address three longstanding problems in recommender system, i.e., new user, cold-start and sparsity.

3. Experimental results on the collected YouTube-Twitter dataset validate the effectiveness of the proposed solution on three kinds of typical users, in term of not only accuracy, but also diversity and novelty.

2. RELATED WORK

Cross-network collaborative applications have recently attracted attentions. One line is on cross-network user modeling, which focuses on integrating various social media activities. In [13], the authors introduced a cold-start recommendation solution by aggregating user profiles in Flickr, Twitter and Delicious. Deng et al. proposed a personalized YouTube video recommendation solution by incorporating user information from Google+ [14]. Another line is devoted to taking advantage of different OSNs' characteristics. Suman et al. exploited the real-time and socialized characteristics of the Twitter tweets to facilitate video applications in YouTube [15]. In [16], Twitter event detection is conducted by employing Wikipedia pages as the authoritative references. Qi et al. proposed a cross-network link prediction approach by borrowing links from another more densely linked network [17]. Our work belongs to the first line, where users' auxiliary behaviors on Twitter are transferred to facilitate user modeling on YouTube, and a unified solution is designed to address three problems of new user, cold-start and data sparsity.

To conduct cross-network collaborative applications, one important issue is the acquisition of cross-network user accounts corresponding to the same individual. Currently

[2] Due to the privacy concerns, the user demographic information obtained from registration is very sparse and sometimes not accurate, which may negatively affect the quality of a content-based recommender. Also, in many cases, the tagging behaviors are not adequate to reflect the user's preferences or customized needs. Using complex algorithms to complement these limitations in data doesnot always pay off.

Table 1: Statistics of users who share accounts in other OSNs within the 137,317 Google+ users.

	YouTube	Twitter	Facebook	Flickr
#account	52,390	43,772	31,020	12,242
proportion	0.3815	0.3188	0.2259	0.0892

Table 2: % user overlap between four OSNs.

	YouTube	Twitter	Facebook	Flickr
YouTube	1	0.4253	0.3109	0.1294
Twitter	0.5090	1	0.5376	0.2223
Facebook	0.5251	0.7586	1	0.2207
Flickr	0.5537	0.7948	0.5591	1

there are three ways. (1) Increasing people are voluntary to disclose their user accounts online, by filling in SNS registration information (such as Facebook, Google+) or maintaining an aggregated profiles on services like About.me and Friendfeed. (2) Many IT giants share identical account among their different OSNs, or allow third-party services to access their user base, such as Google account for YouTube and Google+, and Facebook's open platform. (3) With the trend that netizens are using a multitude of OSNs, many researchers are devoted to the field of user account linkage identification, which have achieved satisfied accuracies [18, 19]. In this work, we adopt the first way to construct our cross-network dataset, which will be detailed in the next section.

3. PROBLEM JUSTIFICATION

3.1 Data Set

Google+ encourages users to share their user accounts on other OSNs in the Google profile. We collected Google profiles of 137,317 Google+ users, and obtained 22,279 users who provide their user accounts on both YouTube and Twitter. We further examined these users on YouTube and Twitter via the respective APIs, and crawled data of 17,617 users who are publicly accessible and have behaviors on the both OSNs. These 17,617 users are recorded as the *overlapped users* in the rest of this paper. Specifically, on YouTube, for each of the overlapped users, we downloaded his/her uploaded videos, favorite videos and video playlists. For each video, the video tags, titles and descriptions are also collected. On Twitter, for each user we downloaded his/her recent 1,000 tweets and the user profile. As a result, the collected YouTube-Twitter dataset consists of 1,097,982 video-related behaviors and 9,253,729 tweet-related behaviors.

3.2 Data Analysis

To justify our motivation and the practical feasibility of cross-network collaborative solution, we conducted preliminary data analysis to answer two questions: (1) Is it easy to obtain the user accounts across different OSNs? (2) Are user's Twitter tweeting behaviors adequate to make up the video-related data shortage on YouTube?

For the first question, within the total 137,317 Google+ users, we examined the number of available user accounts on four popular OSNs: YouTube, Twitter, Facebook, and Flickr. The results are shown in Table 1. We can see that a noticeable portion of Google+ users disclose their user accounts on other OSNs, especially on YouTube and Twitter (more than 30%). In Table 2, we also examined the user overlap among the four OSNs. The user overlap proportion between OSN A and B is calculated as $overlap(A, B) = \frac{|A \bigcap B|}{|A|}$, where $|A|$ indicates the number of available user accounts on OSN A. We can see that the user overlaps between different OSNs are significant, which is consistent with a Pew Internet study conducted over a global sample of online adults [20]. This validates the fact that users are voluntary to disclose their multiple-OSN accounts, and opens up opportunities for large-scale cross-network collaboration practices.

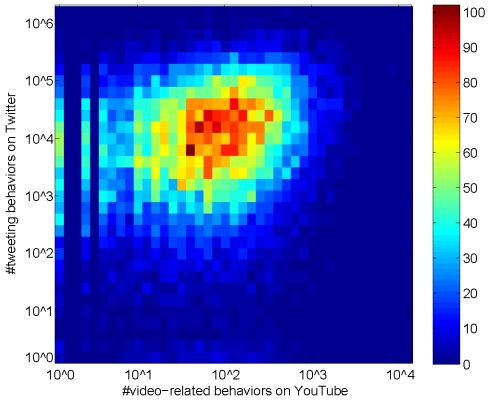

Figure 2: The heatmap of user behavior counts on YouTube and Twitter. (best viewed in color)

For the second question, for each user, we counted the number of his/her video-related behaviors on YouTube and tweeting behaviors on Twitter. In Fig. 2, x-axis and y-axis indicate the number of video-related and tweeting behaviors in log scale, respectively. We marked each user on the coordinate system and obtained a heatmap over all the 17,617 overlapped users. The depth of the red color is proportional to the number of users having the corresponding behavior count combination at this point. Two observations are made. (1) The red-color points locate largely in the upper left of the diagonal line. This indicates that most users have more tweeting behaviors on Twitter than video-related behaviors on YouTube. (2) For users who have sparse video-related behaviors, e.g., 10^0-10^1 along x-axis, the number of their available tweeting behaviors has a wide range from 10^2-10^5. This validates our motivation to address the cold-start and sparsity problems on YouTube by leveraging the auxiliary user data on Twitter.

4. A UNIFIED VIDEO RECOMMENDATION FRAMEWORK

This section introduces the proposed video recommendation solution via cross-network collaboration. We first provide some key notations to formulate the problem.

Given a set of overlapped users $\mathcal{U}$, each user $u \in \mathcal{U}$ corresponds to a two-dimensional tuple $[\mathcal{T}_u, \mathcal{V}_u]$, where $\mathcal{T}_u$ indicates his/her tweet collection on Twitter and $\mathcal{V}_u$ indicates the videos he/she has interacted with on YouTube. Each $v \in \mathcal{V}_u$ is represented by its contained textual words and visual keyframes $[\mathbf{w}_v, \mathbf{f}_v]$. We use R to denote the observed user-video interaction matrix on YouTube, where the row

$R_{i,\cdot}$ indicates the i^{th} user u_i's observed video interaction. The goal is to make use of the users' tweeting activity $\mathcal{T}_u$, and the video-related behaviors R, to design a unified video recommendation solution that facilitates three kinds of users whose $R_{i,\cdot}$ is empty, sparse or dense.

4.1 Preliminaries

Our proposed solution is based on regularized matrix factorization. In this subsection, we provide the necessary formulation of standard regularized matrix factorization (MF) model and introduce the preprocessing utilized in our work.

In recommender systems, MF model maps both users and videos to a latent factor space, where user-video interactions are modeled as the inner products. To avoid overfitting, state-of-the-art MF models suggest discovering the latent structure based only on the observed interactions [21, 10]. The standard formulation is as follows:

$$\min_{U,V} ||Y \odot (R - UV^{\mathrm{T}})||_F^2 + \lambda(||U||_F^2 + ||V||_F^2) \qquad (1)$$

where $U = \{\mathbf{u}_1; \cdots; \mathbf{u}_M\} \in \mathbb{R}^{M\times K}$, $V = \{\mathbf{v}_1; \cdots; \mathbf{v}_N\} \in \mathbb{R}^{N\times K}$ are the user and video representations in the K-dimension latent space, and $Y \in \mathbb{R}^{M\times N}$ is a binary mask matrix recording the observed user-video entries. Given the obtained latent factor representations U, V, we can directly estimate the user u_i's preference on video v_j as: $r_{ij} = \mathbf{u}_i \cdot \mathbf{v}_j^{\mathrm{T}}$.

In the context of our problem, we construct the user-video interaction matrix $R \in \mathbb{R}^{M\times N}$ by aggregating user's three kinds of video-related behaviors, i.e., upload, favorite and adding to playlists. Therefore, each entry $r_{ij} \in \{0, 1, 2, 3\}$.

To reduce the influence of sparsity, many existing recommendation solutions also incorporate the content information for regularization [22, 6]. The basic assumption is: the videos with similar content should have close representations in the derived latent factor space. Realizing this assumption, the objective function in Eq. (1) is further regularized with a Laplacian term and can be written as follows:

$$\min_{U,V} ||Y \odot (R - UV^{\mathrm{T}})||_F^2 + \theta Tr(V^{\mathrm{T}}LV) + \lambda(||U||_F^2 + ||V||_F^2) \qquad (2)$$

where $Tr(\cdot)$ is the matrix trace, L is a Laplacian matrix defined as $L = D - S$, θ is the weighting parameter controlling the importance of the Laplacian regularization. Here $S \in \mathbb{R}^{N\times N}$ is the similarity matrix between videos and $D \in \mathbb{R}^{N\times N}$ is a diagonal matrix with $D_{ii} = \sum_j s_{ij}$.

In our work, a topic-based method is utilized to calculate the video similarity matrix S. Specifically, for YouTube video $v : [\mathbf{w}_v, \mathbf{f}_v]$, $\mathbf{f}_v = \{f_1, \cdots, f_N\}$ is a collection of N visual feature vectors associated with v's keyframes, and $\mathbf{w}_v = \{w_1, \cdots, w_M\}$ is the collection of v's M caption and tag words. Viewing each video as one document, we modify a multi-modal topic model [23] to discover the YouTube video topics. After topic modeling, each video v can be represented as a topical distribution $\hat{\mathbf{v}} \in \mathbb{R}^{1\times K^v}$, where K^v is the dimension of the derived topic space. The similarity between the i^{th} and j^{th} video is then calculated as the histogram intersection of their topic distributions:

$$s_{ij} = \sum_{k=1}^{K^v} \min(\hat{v}_i^k, \hat{v}_j^k)$$

where $\hat{v}_i^k$ is the i^{th} video's distribution on the k^{th} topic.

4.2 Auxiliary-network Data Transfer

The goal at the first stage of our solution is to estimate the user's preference on YouTube videos given his/her tweeting activities on Twitter. Although user data on different OSNs are heterogeneous which prevents from direct aggregation, for the same overlapped user, the behaviors on different OSNs can be viewed as the reflections of his/her unique attributes. For example, the preference on advertising videos on YouTube and the interest to tweet news about the release of new products on Twitter are both related to the occupation as a market strategist. Therefore, it is reasonable to assume that the overlapped users' cross-network behaviors have some general association patterns.

We discover the association patterns by examining how the overlapped users' auxiliary data on Twitter can be transferred to their preferences on YouTube videos. Users' preferences on YouTube videos are readily embedded in the user-video interaction matrix R as mentioned above. To represent users' tweeting activities on Twitter, viewing each user's tweeting history as one document, we apply the standard LDA to the corpus constituted by all the Twitter users. As a result, each user can be represented as a topical distribution $\mathbf{u}^t \in \mathbb{R}^{1\times K^t}$, where K^t is the number of topics in the derived Twitter topic space.

For each overlapped user $u \in \mathcal{U}$, we assume that there exists a transfer matrix $W \in \mathbb{R}^{K^t\times K}$ entailing the map from his/her Twitter topical distribution $\mathbf{u}^t$ to his/her user representation $\mathbf{u}$ in the YouTube latent space extracted from R. This assumption is formulated as: $\mathbf{u} = \mathbf{u}^t \cdot W$. Therefore, the task of transferring auxiliary-network data changes to learn the transfer matrix W with observations of the overlapped users' Twitter and YouTube behaviors. We replace the user latent factor matrix U in Eq. (2) to incorporate W, and obtain the following objective function [3]:

$$\min_{W,V} ||Y\odot(R-U^tWV^{\mathrm{T}})||_F^2+\theta Tr(V^{\mathrm{T}}LV^)+\lambda(||W||_F^2+||V||_F^2) \qquad (3)$$

In the new formulation, instead of directly finding the optimal user representations, we change to optimize for the transfer matrix, which captures the association between users' behaviors on the auxiliary and target networks.

Since we are only interested to reconstruct the observed user-video entries in R, we define Ω as the collection of all the observed user-video pairs, i.e., $\forall(i,j) \in \Omega, Y_{ij} = 1$. Eq. (3) can be rewritten as:

$$\begin{aligned} \min_{\mathbf{v}_j, W} \sum_{(i,j)\in\Omega} (r_{ij} - \mathbf{u}_i^t W \mathbf{v}_j^{\mathrm{T}})^2 + \theta \sum_j F(\mathbf{v}_j) \\ + \lambda(||W||_F^2 + \sum_j ||\mathbf{v}_j||_F^2) \\ F(\mathbf{v}_j) \triangleq \mathbf{v}_j(V^{\mathrm{T}}L_j) + (V^{\mathrm{T}}L_j)^{\mathrm{T}}\mathbf{v}_j^{\mathrm{T}} - \mathbf{v}_j L_{jj} \mathbf{v}_j^{\mathrm{T}}. \end{aligned} \qquad (4)$$

where $\mathbf{v}_j$ is the j^{th} row of V, L_j is the j^{th} column of L, L_{jj} is the entry located in the j^{th} column and j^{th} row of L.

We can see that Eq. (4) is convex to $\mathbf{v}_j$ and W respectively with the other variables fixed. Therefore, we adopt stochastic gradient descent for solution and alternatively loop

[3] Since the goal is to learn the transfer matrix, at this stage, we only keep the users who have sufficient behaviors on both YouTube and Twitter as the training samples. A relative dense R and accurate $\mathbf{u}^t$ contribute to an improved inference of W.

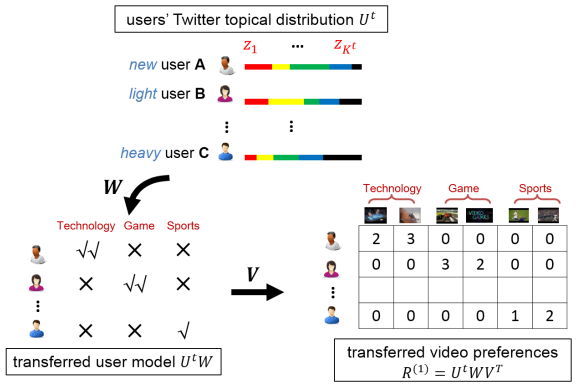

Figure 3: Toy example illustrating the first stage.

through all the observed user-video pairs in Ω [4]. Specifically, for $(i, j) \in \Omega$, let the prediction error $e_{ij} = r_{ij} - \mathbf{u}_i^t W \mathbf{v}_j^{\mathrm{T}}$, then the partial derivatives of the objective function can be derived as:

$$\frac{d}{d\mathbf{v}_j} = -2e_{ij}\mathbf{u}_i^t W + 2\theta(L_j^{\mathrm{T}} V - L_{jj}\mathbf{v}_j) + 2\lambda\mathbf{v}_j$$
$$\frac{d}{dW} = -2e_{ij}\mathbf{u}_i^{t\mathrm{T}}\mathbf{v}_j + 2\lambda W$$

Based on this, we update $\mathbf{v}_j$ and W iteratively until convergence or maximum iteration. The update rules are:

$$\mathbf{v}_j \leftarrow \mathbf{v}_j - \gamma\frac{d}{d\mathbf{v}_j}$$
$$W \leftarrow W - \gamma\frac{d}{dW}$$

where γ denotes the learning rate.

With the derived transfer matrix W and video latent factor representations V, given a test user u_i with his/her tweeting activity and Twitter topical distribution $\mathbf{u}_i^t$, we can estimate u_i's preferences on YouTube videos as:

$$R_{i,\cdot}^{(1)} = \mathbf{u}_i^t W V^{\mathrm{T}} \tag{5}$$

In Fig. 3, we show a toy example by simulating three typical YouTube users. Transferred by W, we estimate the users' interests on latent YouTube video topics indicating *Technology*, *Game*, and *Sports* ($\times$, $\checkmark$, $\checkmark\checkmark$denote "not like", "like", and "very like", respectively). Further multiplied by V, we can discover their preferences on specific videos. Therefore, even no behavior records are available on the target network, we can still build an initial user model by transferring the data from auxiliary networks. This has actually addressed the first scenario in our unified recommendation solution: *new user*.

4.3 Cross-network Data Integration

For the light and heavy users, they already have some observed behaviors on the target network. It is not practical to directly aggregate the observed behaviors with the estimated preference $R_{i,\cdot}^{(1)}$ as they may contradict each other. Therefore, at the second stage of our solution, we introduce another formulation to update the user latent representation U for light and heavy users, by considering both the observed YouTube user-video matrix R and the transferred user model U^tW.

In the new formulation, we view the transferred user model U^tW as a prior to the integrated user model U. This can be interpreted in two-fold: (1) For users with few observed behaviors on the target network, the transferred user model serves as good indication of the integrated user model, i.e., U should resemble U^tW. This actually corresponds to the cold-start problem. (2) The obtained transfer matrix W defines a latent space where the users are located. This helps to measure the correlation between users, and can be employed to alleviate the sparsity problem in the target user-video matrix R. Based on these two interpretations, we introduce the formulation for cross-network data integration as follows [5]:

$$\begin{aligned}\min_{U,V'} \; & ||Y \odot (R - UV'^{\mathrm{T}})||_F^2 + \alpha||P(U - U^tW)||_F^2 \\ & + \beta||V' - V||_F^2 + \lambda(||U||_F^2 + ||V'||_F^2)\end{aligned} \tag{6}$$

Two notations for the above objective function: (1) V is known quantity as the output from the first stage. The reason we also update the learnt video latent representations V' is to better couple with the update of U in fitting the observed user-video matrix R. (2) $P = diag(p_1, \cdots, p_{|\mathcal{U}|})$ is a diagonal matrix to control the contribution of the transferred information. Different from the weighting parameters α, β, and λ which apply on all users, P works on micro-level and defines adaptive weights for different users. To define the user-specific weight p_i, we expect that the user having dense behaviors on YouTube deserves a small p_i, which indicates that more emphasis should be given on modeling of his/her observed video interactions on YouTube to estimate the integrated user model. Specifically, we employ the relative amount of behaviors each user has on YouTube to that on Twitter to define p_i, and the definition is:

$$p_i = \left(1 + e^{\frac{|\mathcal{V}_{u_i}|}{avg.(|\mathcal{V}_u|)} - \frac{|\mathcal{T}_{u_i}|}{avg.(|\mathcal{T}_u|)}}\right)^{-1} \tag{7}$$

where $|\mathcal{T}_{u_i}|$, $|\mathcal{V}_{u_i}|$ indicate the numbers of available tweets and interacted videos of user u_i, $avg.(|\mathcal{T}_u|)$, $avg.(|\mathcal{V}_u|)$ are the corresponding numbers averaged over all the examined users.

Similar to the manipulation of Eq. (3), we rewrite Eq. (6) as follows:

$$\begin{aligned}\min_{\mathbf{u}_i,\mathbf{v}_j'} \; & \sum_{(i,j)\in\Omega} (r_{ij} - \mathbf{u}_i\mathbf{v}_j'^{\mathrm{T}})^2 + \alpha\sum_i ||p_i(\mathbf{u}_i - \mathbf{u}_i^t W)||_F^2 + \\ & \beta\sum_j ||\mathbf{v}_j' - \mathbf{v}_j||_F^2 + \lambda(\sum_i ||\mathbf{u}_i||_F^2 + \sum_j ||\mathbf{v}_j'||_F^2)\end{aligned} \tag{8}$$

The stochastic gradient descent is again adopted to learn the model parameters. For each observed user-video pair $(i, j) \in \Omega$, we can update the parameters with learning rate γ as:

$$\begin{aligned}\mathbf{u}_i &\leftarrow \mathbf{u}_i + \gamma(e_{ij}\mathbf{v}_j' - \alpha p_i^2(\mathbf{u}_i - \mathbf{u}_i^t W) - \lambda\mathbf{u}_i) \\ \mathbf{v}_j' &\leftarrow \mathbf{v}_j' + \gamma(e_{ij}\mathbf{u}_i - \beta(\mathbf{v}_j' - \mathbf{v}_j) - \lambda\mathbf{v}_j')\end{aligned} \tag{9}$$

[4]The fact that each video latent representation vector $\mathbf{v}_j$ can be updated independently of other vectors provides opportunities for parallel implementation. Distributed storing also allows for processing large matrices with huge number of videos.

[5]Note that different from the training samples used at the first stage (Eq. (3)), R and Y in this equation correspond to the test light and heavy users.

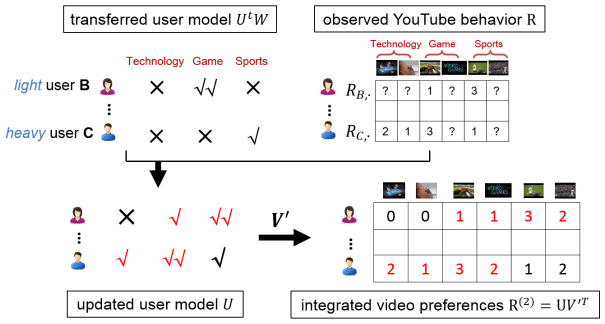

Figure 4: Toy example illustrating the second stage.

In this way, the user latent representation $\mathbf{u}_i$ and video latent representation $\mathbf{v}_j{}'$ are updated. For each test light or heavy user u_i, his/her preferences on YouTube videos can be calculated as:

$$R^{(2)}_{i,\cdot} = \mathbf{u}_i V'^{\mathrm{T}} \tag{10}$$

In Fig. 4, continuing the toy example in Fig. 3, we show how the user models of light and heavy users are updated at the second stage. It is shown that the transferred user model and video representations are revised according to the observed behaviors on YouTube (revised entries are highlighted with red color). The obtained final video preferences consider both the auxiliary information and well fit the observed target-network information.

5. EXPERIMENT

5.1 Experimental Settings

5.1.1 Dataset Partition

To construct a dataset with adequate user behaviors for model learning and performance evaluation, we filtered the raw overlapped user set by keeping the ones who interacted with over 10 YouTube videos and posted over 10 tweets on Twitter. The YouTube videos interacted by less than three users are also filtered out. This results in a dataset of 2,560 users and 4,414 YouTube videos. The resultant user-video matrix has a sparsity score of 99.45%.

Within the experimental dataset, we first randomly selected 1,060 active users who have more than 30 video-related interactions and posted more than 200 tweets, to construct the training dataset used at the first stage to learn the transfer matrix W. Since the proposed solution is expected to be evaluated on three kinds of users, we evenly separated the remaining 1,500 users into three subsets according to the number of their video-related interactions in ascending order, which are denoted as $\mathcal{U}^{new}$, $\mathcal{U}^{light}$, and $\mathcal{U}^{heavy}$. For user $u \in \mathcal{U}^{new}$, all the observed video-related interactions are hidden in the training stage and taken as ground truth for evaluation. For user $u \in \mathcal{U}^{light}$, 30% of the video-related interactions are used to update the user model $\mathbf{u}$, with the rest 70% for evaluation. For user $u \in \mathcal{U}^{heavy}$, 80% of the video-related interactions are used as training data, with the rest 20% for evaluation. We can see that the training partitions from $\mathcal{U}^{light}$ and $\mathcal{U}^{heavy}$ actually constitute the dataset used in in Eq. (6) at the second stage. The statistics of the three user sets is summarized in Table 3.

Table 3: Statistics (per user) of video-related interactions for three kinds of user sets.

Dataset	Statistics	$\mathcal{U}^{new}$	$\mathcal{U}^{light}$	$\mathcal{U}^{heavy}$
Train	*min.*	0	6	24
	avg.	0	7.3	43.3
	max.	0	9	172
Test	*min.*	10	10	11
	avg.	10.8	12.1	19.2
	max.	12	15	74

5.1.2 Parameter Settings

In preprocessing, the topical distributions of YouTube videos and Twitter users are derived from topic modeling. We resorted to the standard perplexity measure [24] and selected the topic number that leads to small perplexity and fast convergence. As a result, the topic number is set as: $K^v = 70$ and $K^t = 60$. The hyperparameters are fixed as $\alpha_{LDA} = 0.8$ and $\beta_{LDA} = 0.1$ according to the empirical expectation for the output distribution.

In the proposed video recommendation solution, six parameters are involved: the dimension of latent factor space K, regularization coefficient λ, learning rate γ, and weighting parameters θ, α, β. Considering the settings without video Laplacian regularization, i.e., set $\theta = 0$, we firstly jointly select K and λ in Eq. (3) by grid search and 2-fold cross validation. As a result, we set $K = 40$ and $\lambda = 0.1$. The learning rate γ is fixed as a small value 0.005 to ensure the convergence to the local minimum. With K, λ and γ fixed, we finally select θ, α, β by the same grid search strategy respectively, and set the parameters leading to the best results, i.e., $\theta = 0.45, \alpha = 20, \beta = 1$.

5.2 Experimental Results and Analysis

To evaluate the effectiveness of the proposed two-stage solution on addressing the mentioned three problems, we implemented four single-network baselines and two different settings of our solution. The six examined methods are listed as follows:

- *Popularity*: recommending popular videos with the most view count, which serves as a simple baseline to address the new user problem;
- *KNN*: the typical item-based collaborative filtering recommendation algorithm [25];
- *LFM*: state-of-the-art Latent Factor Model [10], which is mainly designed to address the sparsity problem;
- *rPMF*: probabilistic Matrix Factorization method incorporating video content Laplacian regularization [6], as shown in Eq. (2);
- *auxTransfer*: the proposed solution that only considers auxiliary-network data transfer, shown in Eq. (3);
- *crossIntegration*: the proposed solution considering both auxiliary-network data transfer and cross-network data integration, shown in Eq. (6).

We view personalized video recommendation as a top-k recommendation task and adopt *top-k precision*, *recall* and *F-score* as the evaluation metrics [26]. For each test user u_i, we recommend the top k YouTube videos with the highest entry score ($r^{(1)}_{ij}$ for new users, $r^{(2)}_{ij}$ for light and heavy users). The evaluation metrics are calculated by examining whether the recommended videos are included in u_i's interested video set $\mathcal{V}_{u_i}$. The final results are averaged over all the test users.

Table 4: Top-10 precision, recall and F-score for the examined methods on three test user sets.

Test set	Metrics	Popularity	KNN	LFM	rPMF	auxTransfer	crossIntegration
new users	*precision*	0.0108	-	-	-	**0.0246**	**0.0246**
	recall	0.0101	-	-	-	**0.0229**	**0.0229**
	F-score	0.0105	-	-	-	**0.0237**	**0.0237**
light users	*precision*	0.0126	0.0160	0.0060	0.0190	0.0242	**0.0274**
	recall	0.0105	0.0083	0.0050	0.0159	0.0201	**0.0227**
	F-score	0.0115	0.0109	0.0055	0.0173	0.0220	**0.0248**
heavy users	*precision*	0.0076	0.0286	0.0088	0.0300	0.0330	**0.0436**
	recall	0.0047	0.0181	0.0045	0.0157	0.0170	**0.0222**
	F-score	0.0058	0.0221	0.0060	0.0206	0.0224	**0.0294**

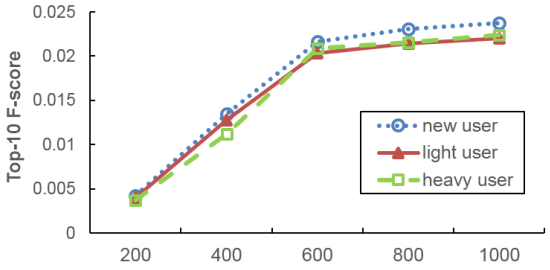

Figure 5: Top-10 F-score as the number of overlapped users at the first stage changes.

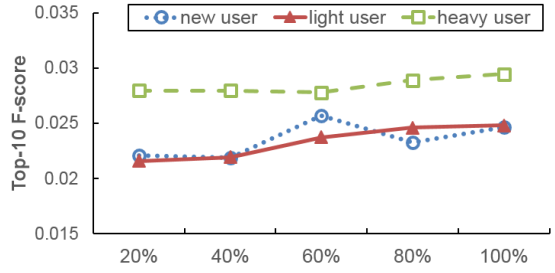

Figure 6: Top-10 F-score as the ratio of Twitter user activities changes.

The evaluation results of the examined methods are shown in Table 4. It is easy to find that the two settings of our solution well address different kinds of users and obtain the best performances. Other observations include: (1) Among the four compared baselines, only *Popularity* can address all three kinds of users. However, recommending the global popular videos fails to capture users' personalized needs and thus achieves inferior performances (heavy users with even lower F-score). (2) For light users, the fact that *auxTransfer* outperforms the four single-network baselines validates our motivation of exploiting auxiliary-network information. *crossIntegration* performs slightly better than *auxTransfer* by further incorporating the limited target-network information. (3) For heavy users, single-network baselines (*KNN*, *rPMF*) achieve comparable results. The improvement of *crossIntegration* ascribes to the prior from auxiliary user model, where user-user correlations are pre-defined to help alleviate the sparsity.

Influence of #overlapped users. We further examined how the number of overlapped users influence the inference of transfer matrix and thus contributes to the final recommendation performance. Different numbers of users are randomly sampled to construct the training set in Eq. (3). We utilize the obtained respective W and evaluate the performance on the *auxTransfer* setting. The top-10 F-score for the three test user sets is shown in Fig. 5, showing a gradual improvement as the number of overlapped users increases. This leads to a coarse conclusion that, more users for training leads to more accurate estimation of W and thus positively affect the recommendation performance.

Sensitivity to #auxiliary-network data. We also investigated whether the performance is sensitive to the number of available auxiliary activities at the cross-network data integration stage. In Fig. 6 we show the top-10 F-score on *crossIntegration* setting by varying the test users' ratio of Twitter activities (i.e., 20%, 40%, 60%, 80%, 100%). It is observed that the performance is not a monotonically increasing function of the Twitter activities. We can understand this result by looking into Eq. (6), that it is the Twitter topical distribution $\mathbf{u}^t$ that directly relates to the update of user model. As long as the scale of Twitter activities is adequate to obtain an accurate topical distribution, more activities will not much influence the final performance.

5.3 Discussion

The goal of cross-network collaboration is to complement the data shortage on target networks. A natural question arises: compared with the case when users have adequate target-network behaviors, will cross-network solution on limited target-network behaviors with auxiliary-network information beat single-network solution on adequate target-network behaviors? will cross-network collaborative recommendation have other advantages except for accuracy?

To investigate into this question, we considered the recommendation to the heavy users $\mathcal{U}^{heavy}$ with two different data settings: (1) using all their training interactions on YouTube videos, and (2) keeping only 20% of the training YouTube interactions with all the available Twitter activities. These data settings simulate the case of adequate target-network behaviors (*adequate*) and limited behaviors with auxiliary-network information (*limited*), respectively. We employ the single-network solutions of *KNN* and *rPMF* on the *adequate* data setting, and the proposed *crossIntegration* on the *limited* data setting. The results of top-10 F-score in Table 5 show that in term of accuracy, cross-network collaborative solution with limited target-network behaviors is not so effective compared with single-network solutions with adequate target-network behaviors.

In practical recommender systems, only considering the accuracy is not sufficient to provide useful recommendation-

Table 5: Performance comparison in term of different evaluation metrics.

Methods/*data setting*	Evaluation metrics		
	F-score	*Similarity*	*Novelty*
KNN/*adequate*	**0.0221**	0.4312	0.0142
rPMF/*adequate*	0.0206	0.3909	0.0139
crossIntegration/*limited*	0.0211	**0.3430**	**0.0159**

s. Therefore, we also investigated into other advantages of cross-network collaborative recommendation, by examining evaluation metrics of *diversity* and *novelty* [6]. From Table 5 we can see that, *crossIntegration* achieves improved diversity and novelty over single-network solutions, even with much fewer target-network behaviors. This shows the advantage of cross-network collaborative recommendation in exploiting users' versatile interests in different domains and the potentials in serendipity recommendation.

6. CONCLUSION AND FUTURE WORK

We have introduced a unified YouTube video recommendation solution to address three typical problems in recommender systems, i.e., new user, cold-start, data sparsity. The evaluation results on different metrics of accuracy, diversity and novelty suggest that, by incorporating auxiliary-network information and employing a cross-network collaborative solution, novel recommenders may lead to a higher satisfaction and utility for users.

To interpret the mechanism of proposed solution, we are conducting some case studies to examine into the obtained transfer matrix. An updated formulation is expected to allow for non-linear cross-network behavior correlation. Moreover, the selection of auxiliary network and auxiliary information is critical to the performance of cross-network collaborative recommendation. Except for the user behavior information, we are also very interested in exploiting the social relation information, e.g., to transfer user preferences from Twitter friend network.

7. ACKNOWLEDGMENT

This work is supported in part by National Basic Research Program of China (No. 2012CB316304), National Natural Science Foundation of China (No. 61225009, 61332016, 61303176, 61432019, U1435211), and Beijing Natural Science Foundation (No. 4131004).

8. REFERENCES

[1] Terence Chen, Mohamed Ali Kaafar, Arik Friedman, and Roksana Boreli. Is more always merrier?: a deep dive into online social footprints. In *Proceedings of the 2012 ACM workshop on Workshop on online social networks*, pages 67–72. ACM, 2012.

[2] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, et al. The youtube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296. ACM, 2010.

[3] Francesco Ricci, Lior Rokach, and Bracha Shapira. *Recommender Systems Handbook: A Complete Guide for Scientists and Practioners*. Springer, 2011.

[6]While *diversity* measures the average pairwise similarity between all the top recommended videos, *novelty* encourages the recommendation of less popular videos. Due to space limitation, we skip the detailed definition. Please refer to [27] for details.

[4] Marco Degemmis, Pasquale Lops, and Giovanni Semeraro. A content-collaborative recommender that exploits wordnet-based user profiles for neighborhood formation. *User Modeling and User-Adapted Interaction*, 17(3):217–255, 2007.

[5] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. *Recommender systems: an introduction*. Cambridge University Press, 2010.

[6] Michael J Pazzani and Daniel Billsus. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007.

[7] Zi-Ke Zhang, Chuang Liu, Yi-Cheng Zhang, and Tao Zhou. Solving the cold-start problem in recommender systems with social tags. *EPL (Europhysics Letters)*, 92(2):28002, 2010.

[8] Hao Ma, Dengyong Zhou, Chao Liu, Michael R Lyu, and Irwin King. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 287–296. ACM, 2011.

[9] Mukund Deshpande and George Karypis. Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems (TOIS)*, 22(1):143–177, 2004.

[10] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *ACM SIGKDD 2008*, pages 426–434. ACM.

[11] Zan Huang, Hsinchun Chen, and Daniel Zeng. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. *ACM Transactions on Information Systems (TOIS)*, 22(1):116–142, 2004.

[12] Viktor Mayer-Schönberger and Kenneth Cukier. *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, 2013.

[13] Fabian Abel, Samur Araújo, Qi Gao, and Geert-Jan Houben. Analyzing cross-system user modeling on the social web. In *Web Engineering*, pages 28–43. Springer, 2011.

[14] Zhengyu Deng, Jitao Sang, and Changsheng Xu. Personalized video recommendation based on cross-platform user modeling. In *ICME 2013*, pages 1–6. IEEE.

[15] Suman Deb Roy, Tao Mei, Wenjun Zeng, and Shipeng Li. Socialtransfer: cross-domain transfer learning from social streams for media applications. In *ACM Multimedia 2012*, pages 649–658. ACM.

[16] Miles Osborne, Saša Petrovic, Richard McCreadie, Craig Macdonald, and Iadh Ounis. Bieber no more: First story detection using twitter and wikipedia. In *TAIA 2012*, volume 12.

[17] Guo-Jun Qi, Charu C Aggarwal, and Thomas Huang. Link prediction across networks by biased cross-network sampling. In *Data Engineering (ICDE), 2013 IEEE 29th International Conference on*, pages 793–804. IEEE, 2013.

[18] Jing Liu, Fan Zhang, Xinying Song, Young-In Song, Chin-Yew Lin, and Hsiao-Wuen Hon. What's in a name?: an unsupervised approach to link users across communities. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 495–504. ACM, 2013.

[19] Reza Zafarani and Huan Liu. Connecting users across social media sites: a behavioral-modeling approach. In *ACM SIGKDD 2013*, pages 41–49. ACM.

[20] Maeve Duggan and Aaron Smith. Social media update 2013. *Pew Internet and American Life Project*, 2013.

[21] Andriy Mnih and Ruslan Salakhutdinov. Probabilistic matrix factorization. In *Advances in neural information processing systems*, pages 1257–1264, 2007.

[22] Marko Balabanović and Yoav Shoham. Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66–72, 1997.

[23] David M Blei and Michael I Jordan. Modeling annotated data. In *SIGIR 2003*, pages 127–134.

[24] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[25] George Karypis. Evaluation of item-based top-n recommendation algorithms. In *Proceedings of the tenth international conference on Information and knowledge management*, pages 247–254. ACM, 2001.

[26] Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1):5–53, 2004.

[27] Guy Shani and Asela Gunawardana. Evaluating recommendation systems. In *Recommender systems handbook*, pages 257–297. Springer, 2011.