A Unified Video Recommendation by Cross-Network User Modeling

MING YAN, Institute of Automation, Chinese Academy of Sciences and University of Chinese Academy of Sciences, Beijing, China
JITAO SANG, Institute of Automation, Chinese Academy of Sciences and State Key Laboratory for Novel Software Technology, Nanjing University, People's Republic of China
CHANGSHENG XU, Institute of Automation, Chinese Academy of Sciences and University of Chinese Academy of Sciences, Beijing, China
M. SHAMIM HOSSAIN, Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Online video sharing sites are increasingly encouraging their users to connect to the social network venues such as Facebook and Twitter, with goals to boost user interaction and better disseminate the high-quality video content. This in turn provides huge possibilities to conduct cross-network collaboration for personalized video recommendation. However, very few efforts have been devoted to leveraging users' social media profiles in the auxiliary network to capture and personalize their video preferences, so as to recommend videos of interest. In this article, we propose a unified YouTube video recommendation solution by transferring and integrating users' rich social and content information in Twitter network. While general recommender systems often suffer from typical problems like cold-start and data sparsity, our proposed recommendation solution is able to effectively learn from users' abundant auxiliary information on Twitter for enhanced user modeling and well address the typical problems in a unified framework. In this framework, two stages are mainly involved: (1) auxiliary-network data transfer, where user preferences are transferred from an auxiliary network by learning cross-network knowledge associations; and (2) cross-network data integration, where transferred user preferences are integrated with the observed behaviors on a target network in an adaptive fashion. Experimental results show that the proposed cross-network collaborative solution achieves superior performance not only in terms of accuracy, but also in improving the diversity and novelty of the recommended videos.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; H.4.m [**Information Systems**]: Miscellaneous

General Terms: Algorithms, Experimentation, Performance

Additional Key Words and Phrases: Personalized video recommendation, cross-network collaboration, user modeling

53

This work is supported in part by National Basic Research Program of China (No. 2012CB316304), National Natural Science Foundation of China (No. 61432019, No. 61225009, No. 61303176, No. 61272256, No. 61373122, No. 61332016). The authors extend their appreciation to the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia for funding this work through the research group project no. RGP-228.
Authors' addresses: M. Yan, J. Sang, and C. Xu (corresponding author), National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing 100190, China; emails: {ming.yan, jtsang, csxu}@nlpr.ia.ac.cn; M. S. Hossain, Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia; email: mshossain@ksu.edu.sa.

DOI: http://dx.doi.org/10.1145/2957755

ACM Reference Format:
Ming Yan, Jitao Sang, Changsheng Xu, and M. Shamim Hossain. 2016. A unified video recommendation by cross-network user modeling. ACM Trans. Multimedia Comput. Commun. Appl. 12, 4, Article 53 (August 2016), 24 pages.
DOI: http://dx.doi.org/10.1145/2957755

1. INTRODUCTION

Being the world's largest video sharing platform, large amounts of videos are created and consumed in YouTube every day. The latest statistics show that every minute, 300 hours of video are uploaded to YouTube,[1] resulting in an estimate of more than 2 billion videos totally. Flooded with such a tremendous volume of videos, people usually find it hard to promptly and precisely discover related videos of their interest. Personalized recommendation service, which aims at providing relevant information tailored to user's interest or preference, stands out as an effective solution to deal with this information overload problem.

In general, traditional personalized video recommendation service mainly understands the user interests and preferences by analyzing various available information users leave in the online video sharing site, such as explicit and implicit user ratings [Koren 2008; Hu et al. 2008], social links [Wang et al. 2013], multimodal content information [Mei et al. 2011], etc. However, the user data available on one video sharing site are usually not sufficient to understand the typical user interests and preferences. The notorious cold-start and sparsity issues have significantly hindered accurate user modeling and practical personalized social media services [Ricci et al. 2011].

Fortunately, with the rise of social media, people now usually engage in disparate Online Social Networks (OSNs) simultaneously for different purposes [Chen et al. 2012]. 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, etc. According to the GWI social report,[2] each Internet user has 5.54 social media accounts and is actively using 2.82 social platforms on average. The *cross-network* activities together record people's integral online footprints and serve as a good addition to understand the comprehensive user interests from different perspectives. As a result, different OSNs are not isolated anymore and can exchange information with each other via the *overlapped* users.[3] Taking YouTube as an example, it has been allowed that users can directly log in in to YouTube via their Google+ accounts and YouTube has also released the share feature so that viewers can connect to their social media accounts and share videos with their networks. However, there have been few efforts on understanding the correlations between users' social media behaviors and their video preferences. In this article, we aim to leverage users' rich social media behaviors on the auxiliary OSN (i.e., Twitter) to help estimate their video preferences on the target OSN (i.e., YouTube), and design a unified video recommendation solution.

The unified video recommendation solution is expected to address the typical cold-start and data sparsity problems on YouTube and adaptively adjusts to different kinds of YouTube users. According to the sparsity level of available user data in YouTube, the users can be generally categorized into three kinds: (1) *New user*. This refers to the users who newly register to YouTube and there exist no historical video interactions for them. Traditional collaborative filtering methods that rely on the user-video interaction data cannot handle this kind of user. (2) *Light user*. This refers to the users with limited behavior records and the recommender may tend to give them a biased recommendation

[1] http://www.youtube.com/yt/press/statistics.html.
[2] http://www.globalwebindex.net/blog/internet-users-have-average-of-5-social-media-accounts.
[3] Overlapped users refer to the users who have user accounts in multiple OSNs.

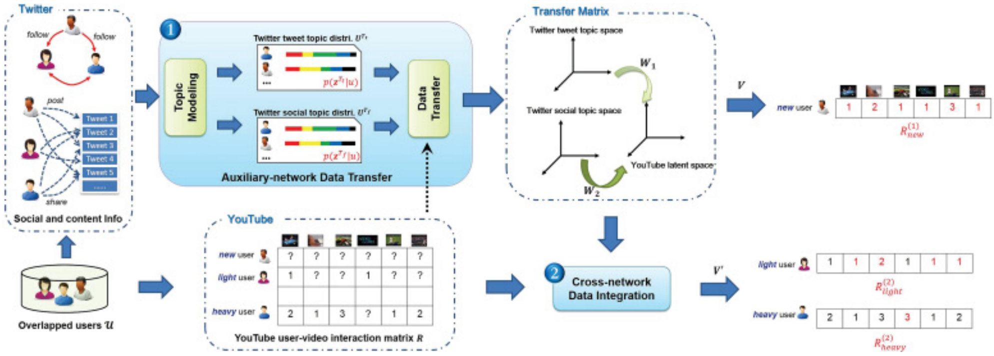


Fig. 1. Flow diagram of the proposed two-stage solution framework.

based on their sparse behaviors. (3) *Heavy user*. This acts as a contrast to the *Light user* and represents the users who frequently interact with the YouTube website. The proposed solution is thus designed to facilitate these three kinds of users in a unified recommendation framework.

Cold-start and *data sparsity* have been two of the most challenging problems in the field of recommender systems. Although they have received extensive attention in the past decade, these problems still remain open. Currently, most of the efforts are devoted to designing advanced models to better exploit the limited and possibly unpromising data within the target OSN, but largely ignore the abundant user data available outside on the auxiliary OSNs. Recent practices on big data [Mayer-Schönberger and Cukier 2013] 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 address the typical cold-start and data sparsity problems, and benefit three types of typical YouTube users. Specifically, for *new users*, we estimate their video preferences on YouTube by considering both their tweeting activities and social link information 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 the following three points. (1) The auxiliary user information on Twitter contains enormous tweeting activities as well as social link information, which indicate the typical user interests from different perspectives. It is not trivial to integrate them together toward a more comprehensive user interest understanding. (2) User data on different OSNs are heterogeneous. There exist no explicit connections between these cross-network users data. We cannot directly utilize the user tweeting data or social link information collected from Twitter to estimate the YouTube video preference. (3) 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 preceding challenges, we design a unified solution framework which consists of two stages, that is, auxiliary-network data transfer, and cross-network data integration (as depicted in Figure 1). At the first stage, to derive a more comprehensive user interest model on Twitter, we first represent each Twitter user on some generalized Twitter tweet topic space and Twitter social topic space, respectively, with different types of user information considered. Then we assume that the correlations between the auxiliary-network and target-network behaviors are embedded in a content transfer matrix and a social transfer matrix, by which the users' tweeting activities and social link information can be jointly mapped to a latent user space on YouTube from the respective Twitter topic spaces. With the derived transfer matrices, we can estimate a user's video preferences by jointly transferring his/her social and content information. 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 on different OSNs. 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 the Twitter tweeting activities, social link information, and historical interactions with YouTube videos. A straightforward recommender can then be designed to utilize the integrated user model for recommendation.

Our contributions in this work can be summarized as follows:

—We introduce a novel personalized video recommendation solution by leveraging users' cross-network social and content data. This is consistent with users' multi-OSN engagement phenomenon and entails user modeling from different aspects.
—A unified video recommendation framework is presented, under which different types of user information on Twitter are jointly transferred and integrated in an early-fusion way for enhanced user modeling on YouTube. With this framework, we also address the typical cold-start and data sparsity problems in recommender system and benefit three different types of users in an adaptive fashion.

A preliminary version of this work was published in Yan et al. [2015]. The extension in this article includes five aspects: (1) Different from the solution in Yan et al. [2015] where only users' tweeting activities are used, this article proposes to also incorporate users' social link information on Twitter for enhanced user modeling and cross-network data transfer. (2) We explicitly present an early-fusion method to better transfer and integrate both the users' social and content information on Twitter for enhanced user modeling on YouTube in Sections 4.3 and 4.4. (3) We add a discussion on the feasibility of the proposed cross-network solution in Section 6, including (a) a statistical justification on the assumption that there exists correlation between users' social media interests and video preferences, (b) a visualization on the derived correlations by the proposed approach, and (c) a discussion on the limitations of the proposed approach. (4) A more comprehensive survey of the related work has been presented in Section 2, including the following: (a) the literature of personalized video recommendation has been added as one subsection, and (b) the research topic of cross-domain recommendation has also been introduced in a cross-network collaboration mechanism. (5) More quantitative experiments are conducted to validate the effectiveness of the proposed solution framework, including the following: (a) a more focused investigation on the sensitivity of the proposed solution with respect to different key model parameters has been added in Section 5.2.2 and Figure 5, and (b) a further exploration on the influence of weighting parameter α to the *diversity* metrics has been made in Section 5.2.3 and Figure 7.

2. RELATED WORK

2.1. Personalized Video Recommendation

With the Internet delivery of video content surging to an unprecedented level, personalized video recommendation has become a very promising online service. Similar to the traditional recommender systems, content collaboration and collaborative filtering have been widely used in video recommendation technologies [Mei et al. 2011; Koren 2008; Balabanović and Shohom 1997; Hong and Shao 2012]. Content collaboration-based methods are devoted to recommending related videos that a given user has liked in the past by matching up the user profiles and video metadata with content analysis technologies. By considering the textual, visual, and aural content of the videos together, Mei et al. [2011] and Yang et al. [2007] designed a unified video recommendation framework based on multimodal fusion and relevance feedback. Park et al. [2010] proposed a framework for recommending online videos by constructing user profiles as an aggregate of tag clouds, while collaborative filtering methods make predictions of the unknown user preferences based on the past user-video interactions, among which latent factor models and neighborhood models prove to be more effective ones [Koren 2008]. A survey of collaborative filtering techniques is given in Su and Khoshgoftaar [2009]. However, there exist certain limitations for both of these kinds of methods: (1) content-based methods only match the user profiles with video metadata using content analysis, which may not reflect user's actual interest or preference; and (2) collaborative filtering techniques often suffer from great data sparsity and cold-start problems due to an initial lack of a users' preference database.

To overcome the preceding limitations, there have been some studies on combining content collaboration and collaborative filtering techniques for better recommendation performance. Burke [2002] gave a good survey on the landscape of actual and possible hybrid recommenders and introduced a novel hybrid system that combines knowledge-based recommendation and collaborative filtering. Basilico and Hofmann [2004] proposed a unified recommendation approach that systematically integrates the past user-item ratings with attributes of users or items by learning a suitable kernel function. Recently, with the rise of social media, social recommendation has attracted attention, which often leverages existing social relations to boost the recommendations. Wang et al. [2013] have proposed a joint video recommendation framework by leveraging social propagation, content similarity, and users' social activities together. Shi et al. [2014] also gave a comprehensive survey of the recent collaborative filtering techniques, which exploits information beyond the user-item matrix.

In this article, we study how users' rich social and content information in another auxiliary network can be utilized to facilitate the personalized video recommendation in the target network.

2.2. Cross-Network Collaboration

With various social media networks growing in prominence, netizens are using a multitude of social media services for social connection and information sharing. Cross-network user modeling, which focuses on integrating various social media activities [Abel et al. 2011; Deng et al. 2013; Yan et al. 2013], has recently attracted more and more attention. In Abel et al. [2011], the authors introduced a cold-start recommendation problem by aggregating user profiles in Flickr, Twitter, and Delicious. Sang et al. [2015] proposed a method to aggregate video behaviors of the same modality between YouTube and Google+, toward a better user preference understanding. Roy et al. [2012] utilized the real-time social streams in Twitter to facilitate video recommendation in YouTube, by building an intermediate topic space. This work mainly focuses on directly aggregating more user data from different networks, which share some common semantics (e.g., common words or tags). However, there exist certain gaps between data in different networks and the common semantics may not always exist for the heterogeneous user data. Our work aims to bridge the gap by building cross-network correlations via the collective behaviors of a collection of *overlapped* users, of which the correlations are embedded in a transfer matrix. The idea of transfer matrix is similar to that in Gao et al. [2014]; the difference lies in that their work utilizes the transfer matrix to model relationships among sentiment, opinions, and actions in the same network, while our work uses it in the context of cross-network correlations. Moreover, we further apply the transferred information to an adaptive integration with the user behaviors in the target network.

In the literature of recommender systems, cross-domain recommendation has been studied for years, which is the most similar to our work. It aims to generate or enhance recommendations in a target domain by exploiting knowledge from source domains. According to Cantador and Cremonesi [2014], there are basically two types of cross-domain approaches, based on how knowledge from the source domain is exploited. The first one focuses on aggregating knowledge from both source and target domains, such as aggregating user ratings [Loni et al. 2014] or social tags [Abel et al. 2011] from multiple domains. The other one is devoted to linking and transferring knowledge from source domain to target domain [Singh and Gordon 2008; Li et al. 2009a, 2009b; Hong et al. 2014]. For example, Singh and Gordon [2008] proposed a collective matrix factorization method to jointly factorize multiple relational matrices by sharing common latent features. Li et al. [2009a, 2009b] introduced a novel codebook transfer method by transferring the user-item rating patterns across multiple domains. A brief survey of cross-domain recommendation can also be found in Fernández-Tobías et al. [2012]. However, to the best of our knowledge, most of the cross-domain recommendation approaches are devoted to finding correlations between related domains, such as movies and books, while our work aims to establish correlations between more complicated and totally uncorrelated social networks via *overlapped* users. Moreover, the user preferences are usually expressed in some well-defined rating matrix or binary matrix in cross-domain recommendation, while more intricate social and content data are involved in our cross-network scenario.

3. PROBLEM JUSTIFICATION

3.1. Dataset

Google+ encourages users to share their user accounts on other OSNs in the Google profile. Therefore, to construct a dataset with user account linkage across different networks, we started from the Google+ website and randomly selected some seed users, then the snowball sampling method is utilized to collect 137,317 Google+ users from their social connections. The external account links of these users are kept, among which 22,279 users provide their user accounts on both YouTube and Twitter. We further examined these users on YouTube and Twitter via the respective Application Programming Interfaces (APIs), and crawled data of 17,617 users who are publicly accessible and have activities on both OSNs.[4] These 17,617 users are recorded as the *overlapped users* in the rest of this article. 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 1000 tweets, total friend collection, and the user profile. As a result, the collected YouTube-Twitter dataset[5] consists of 1,097,982 video-related behaviors (including upload, favorite and adding to playlist) and 9,253,729 tweeting behaviors. On average, each user in Twitter has followed about 1042 friends.

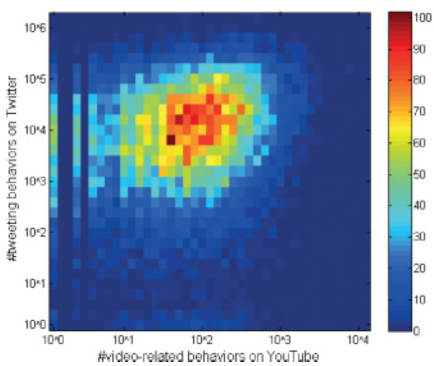
[4]The accounts of public figures and popular organizations are also filtered out to represent the general user collection.
[5]The dataset is available at http://www.nlpr.ia.ac.cn/mmc/homepage/jtsang/dataset.html.

Table I. Statistics of Users Who Share Accounts in Other OSNs Within the 137,317 Google+ Users

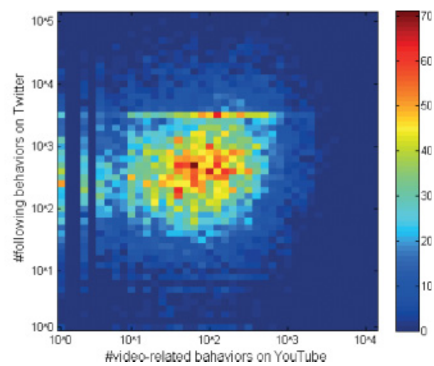
	YouTube	Twitter	Facebook	Flickr
#account	52,390	43,772	31,020	12,242
proportion	0.3815	0.3188	0.2259	0.0892

Table II. % 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



(a) tweeting behaviors considered



(b) following behaviors considered

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

3.2. Data Analysis

Since our motivation is to leverage users' rich social behaviors on some auxiliary network to help estimate their video preferences on the target network, two questions naturally arise: (1) Is it easy to obtain the cross-network user accounts? (2) Can user's social behaviors on Twitter (i.e., auxiliary network) make up the video-related data shortage on YouTube (i.e. target network) in the derived dataset? Therefore, we first conducted a preliminary data analysis on our dataset to answer these questions.

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 I. 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 II, 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 the mentioned OSNs are significant, which is consistent with a Pew Internet study conducted over a global sample of online adults [Duggan and Smith 2013]. This also lays the foundation for the cross-network collaboration practices on the derived dataset.

For the second question, for each user, we counted the number of his/her tweeting behaviors and following behaviors on Twitter with respect to the number of video-related behaviors on YouTube, respectively. In Figure 2, x-axis and y-axis indicate the number of different behaviors on YouTube and Twitter in log scale, respectively. With one type of Twitter social behavior considered each time, we marked each user on the coordinate system and obtained a heatmap over all 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 points locate largely in the upper left of the diagonal line, especially in Figure 2(a). This indicates that in our dataset most users have more social behaviors on Twitter than video-related behaviors on YouTube. (2) For users who have sparse video-related behaviors, for example, 10^0–10^1 along the x-axis, the number of their available social behaviors on Twitter is mostly more sufficient, with a wide range from 10^2 to 10^5 for tweeting behaviors and 10^1 to 10^4 for following behaviors. This provides a huge possibility to address the cold-start problem on YouTube by leveraging the auxiliary user data on Twitter.

4. A UNIFIED VIDEO RECOMMENDATION FRAMEWORK

4.1. Problem Formulation

Definition 1 (*Unified Video Recommendation Framework*). Given a set of overlapped users $\mathcal{U}$, each user $u \in \mathcal{U}$ corresponds to a three-dimensional tuple $\mathcal{T}_u, \mathcal{F}_u, \mathcal{V}_u]$, where $\mathcal{T}_u$ and $\mathcal{F}_u$ indicate his/her tweet collection and friend 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 $\mathcal{T}_u$, following relation $\mathcal{F}_u$, 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.

4.2. Preliminaries

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

In recommender systems, the 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 [Mnih and Salakhutdinov 2007; Koren 2008]. The standard formulation is as follows:

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

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, 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 equally aggregating user's three kinds of video-related behaviors, that is, 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 [Balabanović and Shoham 1997; Hong et al. 2015]. In this work, we adopt the spectral regularization method to preserve the local similarity of the video content[6] [Gao et al. 2010]. 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:

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

where $Tr(\cdot)$ is the matrix trace, L is a Laplacian matrix defined as $L = D - S$, and 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 $v : [\mathbf{w}_v, \mathbf{f}_v]$, $\mathbf{f}_v = \{f_1, \ldots, f_N\}$ is a collection of N visual feature vectors associated with v's keyframes, and $\mathbf{w}_v = \{w_1, \ldots, w_M\}$ is the collection of v's M caption and tag words. Viewing each video as one document, we modify a multimodal topic model [Blei and Jordan 2003] 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:

$$s_{ij} = \sum_{k=1}^{K^v} \min\big(\hat{v}_i^k, \hat{v}_j^k\big), \tag{3}$$

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

4.3. 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 and friend-following behaviors on Twitter. Although user data on different OSNs are heterogeneous, which prevents 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 or follow various social media influencers 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. In this stage, we aim to discover the association patterns by examining how the overlapped users auxiliary data on Twitter can be transferred to their preferences on YouTube videos.

Two challenges are mainly involved in this stage: (1) how to effectively leverage both users' tweeting activities and following behaviors on Twitter for comprehensive user modeling, and (2) how to jointly transfer users' preferences on Twitter to their preferences on YouTube videos. To address the preceding challenges, we first separately conduct a semantic-based and a network-based topic modeling on the overlapped users with different types of user behaviors considered each time. Then, with all the overlapped users represented in the derived Twitter topic spaces and their user-video interactions on YouTube observed, we further design a data transfer mechanism to allow the user preferences transfer between the Twitter topic spaces and YouTube latent space by mining some general cross-network association patterns.

[6]Note that other methods, such as the collective factorization method [Gao et al. 2015], can also be used to incorporate the content information. Here the purpose is to reduce the data sparsity problem and we do not expect to introduce more freedom by factorizing the sparse video-content matrix, so we adopt the spectral method instead.

4.3.1. Topic Modeling. In Twitter, to capture the user interest or preference from different perspectives, we conduct user topic modeling with the tweeting activities and the friend-following behaviors, respectively.

Semantic-Based Topic Modeling. Since our ultimate goal is to recommend videos to users and videos distribute more on a semantic level, it is natural to also represent Twitter users in some semantic topic space. Therefore, we aggregate each Twitter user's generated tweets and keep only the nouns and hashtags.[7] Then the standard Latent Dirichlet Allocation (LDA) model [Blei et al. 2003] is applied to each user for topic modeling, with user as *document*, and the nouns or hashtags of his/her tweets as *word*. In this way, the derived Twitter topics can capture some co-occurred semantic concepts frequently used by many users that may also be found in YouTube video semantic space.

Network-Based Topic Modeling. Except for sharing information and expressing opinions with tweets, users in Twitter also maintain their social links with others by following them. To choose whom to follow can also indicate much of the typical user interest and preference. For example, a person who follows many musical stars such as *Britney Spears* and *Taylor Swift* on Twitter may result from his/her interest in *music*. Therefore, we are also interested in capturing the user interest or preference from the persons he/she chooses to follow on Twitter. Some recent work finds it very effective to extract topics from the social graph data with a traditional topic modeling technique, especially when only linkage data is available [Cha and Cho 2012]. In this way, we represent each Twitter user (*document*) with all his/her friends (*words*) and apply the standard LDA for topic modeling. Since topic modeling exploits co-occurrence relationships, the Twitter topics derived in this way actually capture the typical user interest shared by a subset of Twitter friends. High document-topic distribution indicates user's significant interest in a class of Twitter friends.

After topic modeling, we can obtain (1) Twitter user tweet topic distribution matrix $U^{T_t} = \{\mathbf{u}_1^{T_t}; \cdots; \mathbf{u}_{|\mathcal{U}|}^{T_t}\}$; (2) Twitter user social topic distribution matrix $U^{T_f} = \{\mathbf{u}_1^{T_f}; \cdots; \mathbf{u}_{|\mathcal{U}|}^{T_f}\}$. Each user $u \in \mathcal{U}$ is represented as $\mathbf{u}^{T_t} = \{u_1^{T_t}, \ldots, u_{K^{T_t}}^{T_t}\}$ and $\mathbf{u}^{T_f} = \{u_1^{T_f}, \ldots, u_{K^{T_f}}^{T_f}\}$ in the corresponding topic spaces, where K^{T_t} and K^{T_f} are the number of topics in the derived Twitter topic spaces, and $u_k^{T_t} = p(z_k^{T_t}|u)$ and $u_k^{T_f} = p(z_k^{T_f}|u)$ are user u's topic distributions on the k^{th} topic. In this way, user's social and content information are both considered to model user interest from different perspectives, which will be integrated to help estimate the user's preference on YouTube videos in the data transfer stage.

4.3.2. Data Transfer. With the derived Twitter user topic distributions and observed user-video interactions on YouTube, we present the solution for user preference transfer in this stage. For each overlapped user $u \in \mathcal{U}$, we assume that there exists a content transfer matrix $W_1 \in \mathbb{R}^{K^{T_t} \times K}$ and a social transfer matrix $W_2 \in \mathbb{R}^{K^{T_f} \times K}$ entailing the maps from his/her Twitter tweet topic distribution $\mathbf{u}^{T_t}$ and Twitter social topic distribution $\mathbf{u}^{T_f}$ to his/her user representation in the YouTube latent space extracted from R, respectively. To integrate the user preferences transferred from different types of user behaviors, we further assume that the final user representation $\mathbf{u}$ in the YouTube latent space can be seen as a weighted sum of the transferred user representations from both types of user behaviors. This assumption is formulated as $\mathbf{u} = \eta \cdot \mathbf{u}^{T_t} W_1 + (1-\eta) \cdot \mathbf{u}^{T_f} W_2$, where η is a trade-off parameter to balance the contribution of different types of user behaviors on Twitter. Therefore, the task of transferring auxiliary-network data changes to learn the transfer matrices W_1 and W_2 with observations of the overlapped users'

[7]A hashtag is a word or an unspaced phrase prefixed with the number sign ("#") in Twitter that shows some meaningful message.

Twitter and YouTube behaviors. We replace the user latent factor matrix U in Eq. (2) to incorporate W_1 and W_2, and obtain the following objective function[8]:

$$\min_{W_1, W_2, V} ||Y \odot \big(R - \big(\eta \cdot U^{T_t} W_1 + (1-\eta) \cdot U^{T_f} W_2\big) V^{\mathrm{T}}\big)||_F^2 + \theta Tr(V^{\mathrm{T}} L V) + \lambda\big(||W_1||_F^2 + ||W_2||_F^2 + ||V||_F^2\big) \tag{4}$$

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

Since we are only interested in reconstructing the observed user-video entries in R, we define Ω as the collection of all the observed user-video pairs, that is, $\forall (i, j) \in \Omega, Y_{ij} = 1$. Eq. (4) can be rewritten as:

$$\begin{aligned} \min_{\mathbf{v}_j, W_1, W_2} & \sum_{(i,j)\in\Omega} \big(r_{ij} - \big(\eta \cdot \mathbf{u}_i^{T_t} W_1 + (1-\eta) \cdot \mathbf{u}_i^{T_f} W_2\big)\mathbf{v}_j^{\mathrm{T}}\big)^2 \\ & + \theta \sum_j F(\mathbf{v}_j) + \lambda \left(||W_1||_F^2 + ||W_2||_F^2 + \sum_j ||\mathbf{v}_j||_F^2 \right) \\ & F(\mathbf{v}_j) \triangleq \mathbf{v}_j\big(V^{\mathrm{T}} L_j\big) + \big(V^{\mathrm{T}} L_j\big)^{\mathrm{T}} \mathbf{v}_j^{\mathrm{T}} - \mathbf{v}_j L_{jj} \mathbf{v}_j^{\mathrm{T}} \end{aligned} \tag{5}$$

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. (5) is convex to $\mathbf{v}_j$, W_1 and W_2 respectively with the other variables fixed. Therefore, we adopt stochastic gradient descent for solution and alternatively loop through all the observed user-video pairs in Ω. Specifically, for $(i, j) \in \Omega$, let the prediction error $e_{ij} = r_{ij} - (\eta \cdot \mathbf{u}_i^{T_t} W_1 + (1-\eta) \cdot \mathbf{u}_i^{T_f} W_2)\mathbf{v}_j^{\mathrm{T}}$, then the partial derivatives of the objective function can be derived as:

$$\begin{aligned} \frac{d}{d\mathbf{v}_j} &= -2e_{ij}\big(\eta \cdot \mathbf{u}_i^{T_t} W_1 + (1-\eta) \cdot \mathbf{u}_i^{T_f} W_2\big) \\ &\quad + 2\theta\big(L_j^{\mathrm{T}} V - L_{jj}\mathbf{v}_j\big) + 2\lambda \mathbf{v}_j, \\ \frac{d}{dW_1} &= -2\eta \cdot e_{ij}\big(\mathbf{u}_i^{T_t}\big)^{\mathrm{T}} \mathbf{v}_j + 2\lambda W_1, \\ \frac{d}{dW_2} &= -2(1-\eta) \cdot e_{ij}\big(\mathbf{u}_i^{T_f}\big)^{\mathrm{T}} \mathbf{v}_j + 2\lambda W_2. \end{aligned}$$

Based on this, we update $\mathbf{v}_j$ and W_1, W_2 iteratively until convergence or maximum iteration. The update rules are

$$\begin{aligned} \mathbf{v}_j &\leftarrow \mathbf{v}_j - \gamma \frac{d}{d\mathbf{v}_j}, \\ W_1 &\leftarrow W_1 - \gamma \frac{d}{dW_1}, \\ W_2 &\leftarrow W_2 - \gamma \frac{d}{dW_2}, \end{aligned}$$

where γ denotes the learning rate.

[8]Since the goal is to learn the transfer matrices, 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 $\mathbf{u}^{T_t}$, $\mathbf{u}^{T_f}$ contribute to an improved inference of W_1 and W_2.

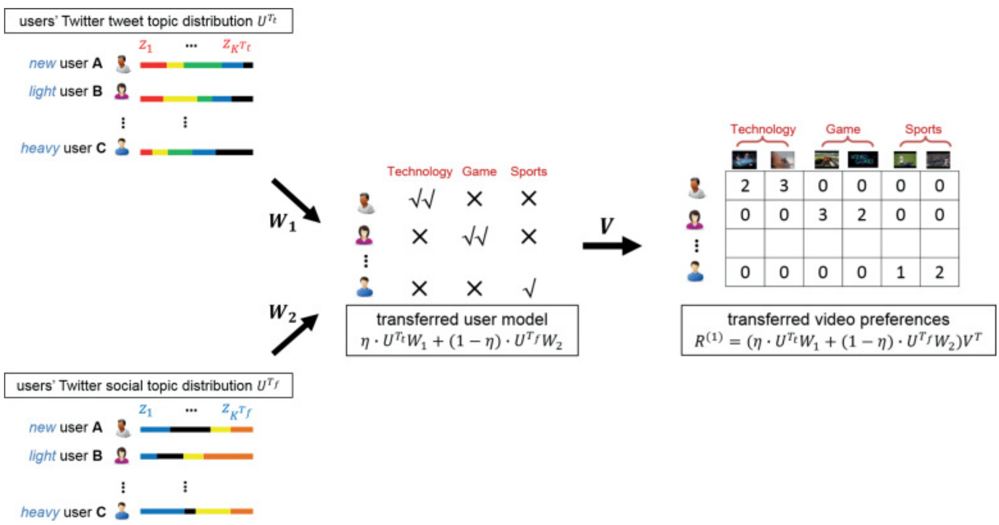


Fig. 3. A toy example illustrating the auxiliary-network data transfer stage.

With the derived transfer matrices W_1, W_2 and video latent factor representations V, given a test user u_i with his/her tweeting activity, friend collection, and the corresponding Twitter topical distributions $\mathbf{u}_i^{T_t}$, $\mathbf{u}_i^{T_f}$, we can estimate u_i's preferences on YouTube videos as

$$R_{i,\cdot}^{(1)} = \left(\eta \cdot \mathbf{u}_i^{T_t} W_1 + (1-\eta)\cdot \mathbf{u}_i^{T_f} W_2\right) V^{\mathrm{T}}. \tag{6}$$

In Figure 3, we show a toy example by simulating three typical YouTube users. Transferred by W_1 and W_2, 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 though 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*.

4.4. 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 $U^{trans} = \eta \cdot U^{T_t} W_1 + (1-\eta)\cdot U^{T_f} W_2$.

In the new formulation, we view the transferred user model U^{trans} as a prior to the integrated user model U. This has a twofold interpretation: (1) For users with few observed behaviors on the target network, the transferred user model serves as a good indication of the integrated user model, that is, U should resemble U^{trans}. This actually corresponds to the cold-start problem. (2) The obtained transfer matrices W_1 and W_2 define 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[9]:

$$\min_{U,V'} ||Y \odot (R - UV'^{\mathrm{T}})||_F^2 + \alpha ||P(U - U^{trans})||_F^2 + \beta ||V' - V||_F^2 + \lambda\big(||U||_F^2 + ||V'||_F^2\big). \tag{7}$$

Two remarks for the preceding objective function are as follows: (1) V is a 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) $P = diag(p_1, \ldots, p_{|\mathcal{U}|})$ is a diagonal matrix to control the contribution of the transferred information. Different from the weighting parameters α, β, and λ that apply on all users, P works on microlevel 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 the 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

$$p_i = \left(1 + e^{\frac{|\mathcal{V}_{u_i}|}{avg.(|\mathcal{V}_u|)} - (\eta \cdot \frac{|\mathcal{T}_{u_i}|}{avg.(|\mathcal{T}_u|)} + (1-\eta)\cdot \frac{|\mathcal{F}_{u_i}|}{avg.(|\mathcal{F}_u|)})}\right)^{-1}, \tag{8}$$

where $|\mathcal{T}_{u_i}|$, $|\mathcal{F}_{u_i}|$, $|\mathcal{V}_{u_i}|$ indicate the numbers of available tweets, friends, and interacted videos of user u_i, and $avg.(|\mathcal{T}_u|)$, $avg.(|\mathcal{F}_u|)$, $avg.(|\mathcal{V}_u|)$ are the corresponding numbers averaged over all the examined users. In this formulation, we first empirically normalize each user's behavior amount with respect to the average amount of all the examined users to measure the relative activeness of each user, so that the activeness of each user across different networks can be directly compared. Then the sigmoid function is adopted to project user's activeness difference between Twitter and YouTube to a probability score to define the adaptive weight p_i.

Similar to the manipulation of Equation (4), we rewrite Equation (7) as follows:

$$\min_{\mathbf{u}_i, \mathbf{v}_j'} \sum_{(i,j)\in\Omega} \left(r_{ij} - \mathbf{u}_i \mathbf{v}_j'^{\mathrm{T}}\right)^2 + \alpha \sum_i ||p_i\left(\mathbf{u}_i - \mathbf{u}_i^{trans}\right)||_F^2 + \beta \sum_j ||\mathbf{v}_j' - \mathbf{v}_j||_F^2 + \lambda \left(\sum_i ||\mathbf{u}_i||_F^2 + \sum_j ||\mathbf{v}_j'||_F^2\right). \tag{9}$$

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

$$\begin{aligned} \mathbf{u}_i &\leftarrow \mathbf{u}_i + \gamma\left(e_{ij}\mathbf{v}_j' - \alpha p_i^2\left(\mathbf{u}_i - \mathbf{u}_i^{trans}\right) - \lambda \mathbf{u}_i\right), \\ \mathbf{v}_j' &\leftarrow \mathbf{v}_j' + \gamma\left(e_{ij}\mathbf{u}_i - \beta(\mathbf{v}_j' - \mathbf{v}_j) - \lambda \mathbf{v}_j'\right). \end{aligned} \tag{10}$$

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

$$R_{i,\cdot}^{(2)} = \mathbf{u}_i V'^{\mathrm{T}}. \tag{11}$$

In Figure 4, continuing the toy example in Figure 3, we show how the user models of light and heavy users are updated at the second stage. It is shown that the transferred

[9]Note that different from the training samples used at the first stage (Equation (4)), R, Y, and U^{trans} in this equation correspond to the test light and heavy users.

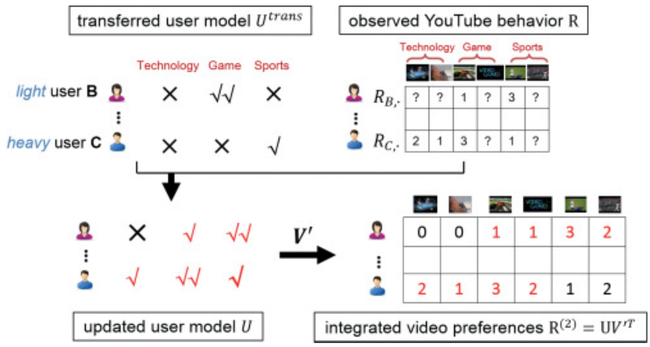


Fig. 4. A toy example illustrating the cross-network data integration stage.

user model and video representations are revised according to the observed behaviors on YouTube (revised entries are highlighted with red). The obtained final video preferences consider both the auxiliary information and well fit the observed target-network information.

5. EXPERIMENTS

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, followed more than 10 other users on Twitter. The YouTube videos interacted by less than three users are also filtered out. This results in a dataset of 2560 users and 4414 YouTube videos, with 75,156 user-video interaction records totally. The resultant user-video matrix has a sparsity score of 99.45%.

Within the dataset, we first randomly selected 1060 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 matrices W_1 and W_2. Since the proposed solution is expected to be evaluated on three kinds of users, we evenly separated the remaining 1500 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 remaining 20% for evaluation. We can see that the training partitions from $\mathcal{U}^{light}$ and $\mathcal{U}^{heavy}$ actually constitute the dataset used in Equation (7) at the second stage. The statistics of the three user sets is summarized in Table III.

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 [Blei et al. 2003] and selected the topic number that leads to small perplexity and fast convergence. As a result, the topic number is set as $K^v = 70$, $K^{T_t} = 60$, and $K^{T_f} = 80$.[10]

[10]The hyperparameters are fixed as $\alpha_{LDA} = 0.8$ and $\beta_{LDA} = 0.1$ according to the empirical expectation for the output distribution.

Table III. 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

In the proposed video recommendation solution, seven parameters are mainly involved: the dimension of latent factor space K, regularization coefficient λ, learning rate γ, and weighting parameters $\eta, \theta, \alpha, \beta$. Considering the settings without video Laplacian regularization and giving equal weight to the social transfer and content transfer, that is, set $\theta = 0,\ \eta = 0.5$, we first jointly select K and λ in Equation (4) by grid search and twofold cross validation. As a result, we set $K = 100$ and $\lambda = 0.1$. The learning rate γ is fixed as a small value 0.01 to ensure the convergence to the local minimum. With K, λ, and γ fixed, we finally select $\eta, \theta, \alpha, \beta$ by the same grid search strategy, respectively, and set the parameters leading to the best results, that is, $\eta = 0.4,\ \theta = 0.2,\ \alpha = 2,\ \beta = 0.6$.

5.1.3. Evaluation Metrics and Comparison Methods. To evaluate the effectiveness of the proposed two-stage solution on addressing the cold-start and data sparsity 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 deal with *new user*;
—*KNN*: the typical item-based collaborative filtering recommendation algorithm [Karypis 2001];
—*LFM*: state-of-the-art Latent Factor Model [Koren 2008], which is mainly designed to address the sparsity problem;
—*rPMF*: probabilistic Matrix Factorization method incorporating video content Laplacian regularization [Pazzani and Billsus 2007], as shown in Equation (2);
—*auxTransfer*: the proposed solution that only considers auxiliary-network data transfer, shown in Equation (4);
—*crossIntegration*: the proposed solution considering both auxiliary-network data transfer and cross-network data integration, shown in Equation (7).

We view personalized video recommendation as a top-k recommendation task and adopt *top-k precision*, *recall*, and *F-score* as the evaluation metrics [Herlocker et al. 2004]. For each test user u_i, we recommend the top k YouTube videos with the highest entry score ($r_{ij}^{(1)}$ for new users, $r_{ij}^{(2)}$ 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.

5.2. Experimental Results and Analysis

5.2.1. Comparison of Different Methods. The evaluation results of the examined methods are shown in Table IV. 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 the following: (1) Among the four compared baselines, only *Popularity* can address all three kinds of users. However, recommending the global popular videos fails to

Table IV. Top-10 Precision, Recall, and F-Score for the Examined Methods on Three Test User Sets

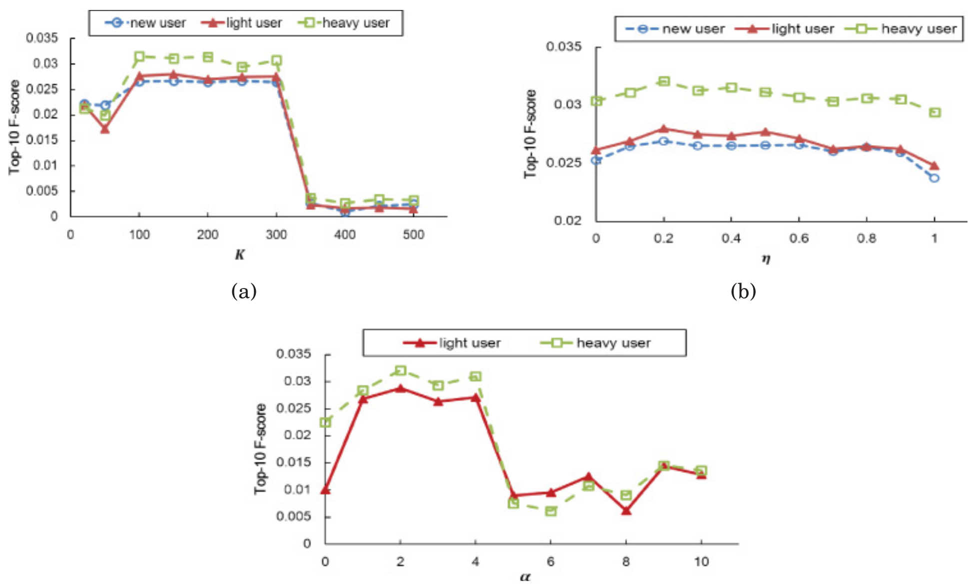
Test set	**Metrics**	Popularity	KNN	LFM	rPMF	auxTransfer	crossIntegration
new users	*precision*	0.0108	-	-	-	**0.0276**	**0.0276**
	recall	0.0101	-	-	-	**0.0256**	**0.0256**
	F-score	0.0105	-	-	-	**0.0266**	**0.0266**
light users	*precision*	0.0126	0.0160	0.0060	0.0190	0.0282	**0.0308**
	recall	0.0105	0.0083	0.0050	0.0159	0.0234	**0.0252**
	F-score	0.0115	0.0109	0.0055	0.0173	0.0256	**0.0277**
heavy users	*precision*	0.0076	0.0286	0.0088	0.0300	0.0388	**0.0456**
	recall	0.0047	0.0181	0.0045	0.0157	0.0206	**0.0241**
	F-score	0.0058	0.0221	0.0060	0.0206	0.0269	**0.0315**

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 predefined to help alleviate the sparsity. (4) For both light and heavy users, *crossIntegration* can always outperform *auxTransfer* by also considering observed video interaction data on YouTube. This also validates the effectiveness of the cross-network data integration stage. Moreover, the improvement is more significant for the heavy users, indicating that more observed behaviors on YouTube contribute to better recommendation performance when conducting data integration.

5.2.2. Influence of Different Parameters. In order to better explore the internal mechanism of the proposed method, we conducted more experiments to examine how different parameters or settings contribute to the final experiment performance.

Sensitivity to model parameters K, η, and α. Three parameters are very important in the proposed method, that is, the dimension of latent factor space K and two weighting parameters η and α, where K defines the size of shared latent factor space where users and videos are projected, η decides on how much the proposed model depends on the social transfer and content transfer, respectively, and α balances the utilization of the transferred user model from Twitter and the observed user-video interactions on YouTube at the cross-network data integration stage. With the other parameters in our model fixed as stated in Section 5.1.2, we tuned K, η, and α at a proper scale in turn and tested the final performance at the cross-network data integration stage, respectively. The top-10 F-score results with respect to different parameter settings are shown in Figure 5.[11] From the results, we can see that (1) the proposed model obtains relatively stable and superior performance when K ranges from 100 to 300, while larger and smaller K will both decrease the performance a lot. This may be due to the reason that, when K is less than 50, the number of latent topics in YouTube is not big enough to capture the actual topical interests of the users and videos, and when K is more than 300, much redundancy will be brought in to the model, which leads to the overfitting of this model and finally results in a bad local minimum. (2) The performance is relatively stable when η varies in a range from 0 to 1, but a little smaller when it is exactly 0 or 1 where only one type of data transfer is considered. The best performance can be

[11]In the proposed model, the parameter α only works on the light and heavy users, so that only the performances of light and heavy users are given in Figure 5(c).



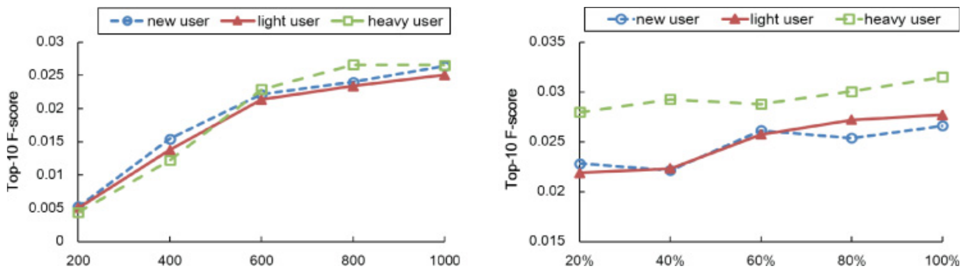
(a) (b)

(c)

Fig. 5. Top-10 F-score with respect to different model parameters: K, η, and α.

obtained when $\eta = 0.2$, and the improvement over the settings when $\eta = 1$ (i.e., when only tweet data is used to transfer) is more than 10% for the new user and light user. This also validates the assumption that more data may lead to better results where in our case social and content data in Twitter are both utilized. (3) The performance changes a lot with different α, where relatively stable and superior performances can be obtained when α varies between 1 and 4. It is easy to see that excessive reliance only on the sparse target network data (i.e., $\alpha < 1$) or auxiliary network data (i.e., $\alpha > 4$) can both lead to poor performances and a proper balance should be learnt to integrate the cross-network user data, which is very significant in our method.

Influence of #overlapped users and #auxiliary-network data. We further examined how the number of overlapped users influence the inference of transfer matrices and thus contributes to the final recommendation performance. Different numbers of users are randomly sampled to construct the training set in Equation (4). We utilize the obtained respective W_1, W_2 and evaluate the performance on the *auxTransfer* setting. The top-10 F-score for the three test user sets is shown in Figure 6(a), 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_1, W_2 and thus positively affect the recommendation performance. We also investigated whether the performance is sensitive to the number of available auxiliary data at the cross-network data integration stage. In Figure 6(b) we show the top-10 F-scores on *crossIntegration* setting by varying the test users' ratio of Twitter behaviors (i.e., 20%, 40%, 60%, 80%, 100%, both for users' social and content behaviors). It is observed that the performance is not a monotonically increasing function of the Twitter activities. We can understand this result by looking into Equation (7), which is the Twitter topical distributions $\mathbf{u}^{T_t}$ and $\mathbf{u}^{T_f}$ that directly relate to the update of the user model. As long as the scale of Twitter user behaviors is adequate to obtain accurate user topical distributions, more user data will not much influence the final performance.



(a) influence of #overlapped users (b) influence of #auxiliary-network data

Fig. 6. Top-10 F-score with respect to different #overlapped users and #auxiliary-network data.

Table V. Performance Comparison in Terms of Different Evaluation Metrics

Methods/*data setting*	Evaluation metrics		
	F-score	*Diversity*	*Novelty*
KNN/*adequate*	**0.0228**	0.4418	0.0148
rPMF/*adequate*	0.0210	0.3983	0.0142
crossIntegration/*limited*	0.0219	**0.3360**	**0.0166**

5.2.3. Other Advantages Beyond Accuracy. 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 these questions, we considered the recommendation to 300 randomly selected heavy users 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 user behaviors. 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 V show that in terms 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 recommendations. Therefore, we also investigated other advantages of cross-network collaborative recommendation, by examining evaluation metrics of *diversity* and *novelty*. For *diversity*, we adopt the intralist similarity used in Ziegler et al. [2005] and calculate the average pairwise content-based similarity between all the top recommended videos. The diversity of a test user $u \in \mathcal{U}^{test}$ to a list of videos $\mathcal{V}_u^{rec}$ recommended to him/her is calculated as

$$diversity(u) = \frac{\sum_{v_i \in \mathcal{V}_u^{rec}} \sum_{v_j \in \mathcal{V}_u^{rec}, v_i \neq v_j} sim(v_i, v_j)}{N_u^{pair}},$$

where $sim(v_i, v_j)$ is the content similarity between video v_i and v_j derived from Equation (3), and N_u^{pair} is the total number of video-video pairs in $\mathcal{V}_u^{rec}$. Then the final *diversity* score is averaged over all the test users $\mathcal{U}^{test}$. For *novelty*, the modified exponential decay score introduced in Shani et al. [2008] is used, which calculates the prediction

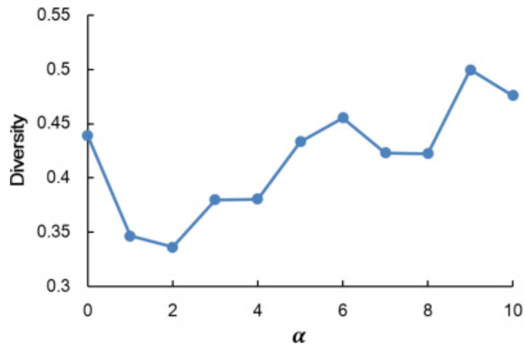


Fig. 7. Diversity for *crossIntegration* method on the *limited* data setting as the model parameter α changes.

accuracy by also considering the user and item popularity. By taking the item popularity first, the novelty of a test user $u \in \mathcal{U}^{test}$ to his/her video recommendation $\mathcal{V}_u^{rec}$ is defined as

$$novelty(u) = \sum_{i \in \mathcal{V}_u^{rec} \cap \mathcal{V}_u} \log\left(\frac{|\mathcal{U}^{test}|}{|\mathcal{U}_i^{test}|}\right),$$

where $\mathcal{V}_u$ is the groundtruth video set he/she has interacted with on YouTube, and $|\mathcal{U}_i^{test}|$ is the number of test users who have interacted with video i. Then the final *novelty* score is derived by also considering user popularity as

$$novelty = \frac{\sum_{u \in \mathcal{U}^{test}} \log\left(\frac{|\mathcal{V}^{test}|}{|\mathcal{V}_u^{test}|}\right) \cdot novelty(u)}{\sum_{u \in \mathcal{U}^{test}} \log(\frac{|\mathcal{V}^{test}|}{|\mathcal{V}_u^{test}|}) \cdot \max(novelty(u))},$$

where $|\mathcal{V}^{test}|$ is the total number of test videos, and $|\mathcal{V}_u^{test}|$ is the number of test videos user u has interacted with.

From Table V 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.

Since the model parameter α controls the balance between the contributions of transferred user model from Twitter and observed user behaviors on YouTube so that more variance from both networks can be brought in, we also conducted more experiments to investigate how the *diversity* changes with different α for the *limited* data setting. The result is shown in Figure 7. We can see that with some appropriate balance chosen, that is, α ranges from 1 to 3, much diversity can be brought in to the recommendation results, which well captures users' versatile interests in different networks.

6. DISCUSSION

To better interpret the feasibility of the proposed cross-network solution, we further conducted some statistical experiments and visualization in this section. Moreover, the limitation of the proposed approach is also discussed.

6.1. Correlation between Social Media Interests and Video Preferences

The key of the proposed solution lies in how to establish some proper correlation between Twitter and YouTube, of which one basic premise is that there should exist a correlation between them in practical situations. Therefore, we made some correlation

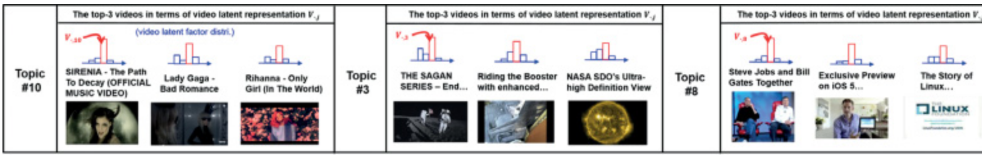


Fig. 8. Visualization of discovered YouTube video latent topics.

Table VI. Visualization of Discovered Twitter Tweet Semantic Topics

Topic	The top-5 probable tweet words in terms of $p(w\|z^{T_t})$				
#20	travel	space	nasa	moon	science
#6	job	looking	career	hr	engineer
#23	google	android	apple	phone	windows

analysis between the derived Twitter topics and the actual YouTube categories on the sampled overlapped users, and check if there are Twitter topics that are highly predictive of YouTube video categories. Specifically, we first define two categorical variables $\mathcal{F}$ and $\mathcal{E}$. $\mathcal{F}$ is defined on the sampled space of users, and associates each user to the top-1 derived Twitter topic that the user possesses the highest expertise. $\mathcal{E}$ associates each user to the favorite YouTube video category that the user is most interested in.

We compute the correlation between the derived Twitter topics and actual YouTube video categories by applying the Pearson's chi-square test [Camilli and Hopkins 1978] on $\mathcal{F}$ and $\mathcal{E}$. The chi-square test checks if the null-hypothesis that two random variables are independent is true or not. On Twitter side, the tweeting behaviors and following behaviors are considered separately, and the result of our test is a strong rejection of the null hypothesis with confidence $p > 0.99$ for both kinds of behaviors (i.e., there exists some correlation between the derived Twitter topics and YouTube video preferences).

This result encouragingly suggests that the interests on the derived Twitter topics can somehow help predict the YouTube video preferences. To better interpret the correlations between YouTube and Twitter, we also made some visualization to the obtained transfer matrices W_1 and W_2 in Equation (4), which are the key to embed the cross-network correlations in our approach. Before that, we first visualized some of the discovered topics in YouTube and Twitter, respectively. Figure 8 shows three sampled YouTube video latent topics derived in Section 4.3.2. For each topic, we provide the top three videos ranked based on the video latent factor distribution on this topic $V_{\cdot,j}$ and each video is represented by its title and thumbnail. It is very easy to interpret the domain knowledge associated with these latent topics. On Twitter side, we visualized some tweet semantic topics and social topics derived in Section 4.3.1, respectively. Table VI shows three sampled tweet semantic topics. For each topic, the top-5 probable tweet words are provided. It is also easy to interpret the semantics concerning these topics and some topics capture similar semantics as in YouTube video latent topics. Table VII shows three sampled Twitter social topics, with each visualized by the top three probable Twitter friends and the friends' profile information. We can see that the network-based topic modeling can discover more flexible topics in a crowdsourcing way. Both the semantic-related topics (e.g., #37 for photography) and geographic-related topics (e.g., #38 for Berlin) are identified.

For the transfer matrix, the entry value $W_{(ij)}$, located at the ith row and jth column of the transfer matrix, indicates the pairwise correlation strength between the Twitter topic i and YouTube topic j. With different user behaviors considered on Twitter, W_1 reflects the correlations between Twitter tweet semantic topics and YouTube latent topics, while W_2 between Twitter social topics and YouTube latent topics. For each

Table VII. Visualization of Discovered Twitter Social Topics

Topic	Username	Location	#Followers	Self-description
#37	PetaPixel	California	149,811	A photography blog for the Web 2.0 generation...
	Chase Jarvis	Seattle + NYC	238,640	Maniac Photographer Director
	Scott Kelby	Oldsmar, Florida USA	135,085	Photoshop and photography book author, trainer...
#77	Lady Gaga	real life gypsy	40,879,332	A pop star from the 70's trapped in 2013...
	Ellen DeGeneres	California	24,071,133	Comedian, talk show host and ice road trucker...
	Katy Perry	REALITY	48,440,525	LET THE LIGHT IN. PRISM. OUT NOW...
#38	Sascha Lobo	Berlin, Germany	161,099	Author, Internet.
	netzpolitik	Berlin, Germany	120,014	Entrepreneur, activist, organizer of @republica.
	Mario Sixtus	Berlin, Germany	60,542	Journalist, Photographer. Hier mehr oder weniger.

kind of social behavior, we visualized the two most significant correlated topic pairs indicated from W_1 and W_2, respectively. Among the $K^{T_t} \times K = 6000$ correlated topic pairs in W_1, two of the most significant are $\{z_3^Y, z_{20}^{T_t}\}$ and $\{z_8^Y, z_6^{T_t}\}$, which have been visualized in Figure 8 and Table VI. It is shown that some interpretable correlations are discovered by the proposed approach. Obviously, the first pair of topics are both related with the semantic topic of space travel. For the second pair, it may indicate that users who are concerned about looking for engineering jobs on Twitter may tend to pay attention to the newest technology videos on YouTube. On the other hand, among the $K^{T_f} \times K = 8000$ correlated topic pairs in W_2, the two most significant are $\{z_{10}^Y, z_{77}^{T_f}\}$ and $\{z_{14}^Y, z_{37}^{T_f}\}$, where the first pair is about the topic of popular music and the second one is related with photography and travel.[12]

Based on the preceding analysis and visualization, we can see that there exists a certain correlation between Twitter and YouTube. The proposed approach can provide a way to successfully discover these underlying correlations, which may further enable the user interest transfer between different networks.

6.2. Limitations of the Proposed Approach

The key to the proposed algorithm lies in the learning of a content transfer matrix W_1 and a social transfer matrix W_2, so that user interests across different topic spaces can be directly transferred. The preceding statistical analysis and visualization have already justified the reasonability of this assumption and also demonstrated the effectiveness of this method in capturing the underlying cross-network correlations. However, this kind of design suffers from the restriction on the linear correlation of the transfer matrix. In practical situations, the correlation between the topics across different networks may be more complicated, which cannot be well represented by a linear fitting. To allow for more freedom of the cross-network correlation, some nonlinear formulation of the solution should also be explored. In that case, the correlation may not be simply measured in the pairwise level, but it can also involve more than two topics. For example, the interest on both topic A and topic B in the auxiliary network may lead to a higher confidence of interest on topic C in the target network, but neither one alone will lead to the conclusion. To deal with this, we are also considering using the association rule mining method in data mining to derive more confident topic correlations, which involve different numbers of topics. Moreover, the recent deep learning-based techniques and some kernel transformation can also be used to discover the latent correlation or representation between different topic spaces, which focus on nonlinear transformation on the original topic spaces.

[12]Here we do not visualize YouTube topic z_{14}^Y for space consideration.

In the proposed approach, the cross-network correlations are established via a collection of overlapped users with adequate user behaviors on both networks. Therefore, the proposed approach is more suitable for the situation when different OSNs share a large active user base between them and more overlapped users obtained can contribute to more reliable correlations established. The association patterns derived by our method are also biased on the typical user collection in our dataset. Nevertheless, the proposed approach provides an alternative method to mine the cross-network association patterns via the overlapped users and we believe that with more and more overlapped users obtained, the derived association patterns will be more generally applicable.

7. CONCLUSION

In this article, we have introduced a unified YouTube video recommendation solution via cross-network collaboration, to address the typical cold-start and data sparsity problems in recommender systems. First, users' tweeting activities and social link information are jointly utilized for enhanced user modeling on Twitter. Then, users' preferences are transferred from Twitter to YouTube by learning some cross-network associations. Finally, the transferred user preferences from Twitter and the observed user behaviors on YouTube are properly integrated for better video recommendation. 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. In the future, we will work toward designing an updated formulation of the solution by allowing for nonlinear cross-network behavior correlation. Moreover, we are also considering incorporating the time factor into the solution so as to capture the swift drift of user interest for dynamic user modeling and more accurate video recommendation.

REFERENCES

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

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

Marko Balabanović and Yoav Shohom. 1997. Content-based, callaborative recommendation. *Communications of the ACM* 40, 3 (1997).

Justin Basilico and Thomas Hofmann. 2004. Unifying collaborative and content-based filtering. In *Proceedings of the 21st International Conference on Machine Learning*. ACM, 9.

David M. Blei and Michael I. Jordan. 2003. Modeling annotated data. In *SIGIR 2003*. 127–134.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. *The Journal of Machine Learning Research* 3 (2003), 993–1022.

Robin Burke. 2002. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction* 12, 4 (2002), 331–370.

Gregory Camilli and Kenneth D. Hopkins. 1978. Applicability of chi-square to 2×2 contingency tables with small expected cell frequencies. *Psychological Bulletin* 85, 1 (1978), 163.

Iván Cantador and Paolo Cremonesi. 2014. Tutorial on cross-domain recommender systems. In *Proceedings of the 8th ACM Conference on Recommender Systems*. ACM, 401–402.

Youngchul Cha and Junghoo Cho. 2012. Social-network analysis using topic models. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 565–574.

Terence Chen, Mohamed Ali Kaafar, Arik Friedman, and Roksana Boreli. 2012. Is more always merrier? A deep dive into online social footprints. In *Proceedings of the 2012 ACM Workshop on Workshop on Online Social Networks*. ACM, 67–72.

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

Maeve Duggan and Aaron Smith. 2013. Social media update 2013. *Pew Internet and American Life Project* (2013).

Ignacio Fernández-Tobías, Iván Cantador, Marius Kaminskas, and Francesco Ricci. 2012. Cross-domain recommender systems: A survey of the state of the art. In *Spanish Conference on Information Retrieval*.

Huiji Gao, Jalal Mahmud, Jilin Chen, Jeffrey Nichols, and Michelle Zhou. 2014. Modeling user attitude toward controversial topics in online social media. In *8th International AAAI Conference on Weblogs and Social Media*.

Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. 2015. Content-aware point of interest recommendation on location-based social networks. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*.

Shenghua Gao, Ivor Wai-Hung Tsang, Liang-Tien Chia, and Peilin Zhao. 2010. Local features are not lonely—Laplacian sparse coding for image classification. In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 3555–3561.

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

Richang Hong and Ling Shao. 2012. Learning from social media network. *Neurocomputing* 95 (2012), 1–2.

Richang Hong, Meng Wang, Yue Gao, Dacheng Tao, Xuelong Li, and Xindong Wu. 2014. Image annotation by multiple-instance learning with discriminative feature mapping and selection. *IEEE Transactions on Cybernetics* 44, 5 (2014), 669–680.

Richang Hong, Yang Yang, Meng Wang, and Xian-Sheng Hua. 2015. Learning visual semantic relationships for efficient visual retrieval. *IEEE Transactions on Big Data* 1, 4 (2015), 152–161.

Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *8th IEEE International Conference on Data Mining (ICDM'08)*. IEEE, 263–272.

George Karypis. 2001. Evaluation of item-based top-n recommendation algorithms. In *Proceedings of the 10th International Conference on Information and Knowledge Management*. ACM, 247–254.

Yehuda Koren. 2008. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 426–434.

Bin Li, Qiang Yang, and Xiangyang Xue. 2009a. Can movies and books collaborate? Cross-domain collaborative filtering for sparsity reduction. In *IJCAI*, Vol. 9. 2052–2057.

Bin Li, Qiang Yang, and Xiangyang Xue. 2009b. Transfer learning for collaborative filtering via a rating-matrix generative model. In *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 617–624.

Babak Loni, Yue Shi, Martha Larson, and Alan Hanjalic. 2014. Cross-domain collaborative filtering with factorization machines. In *Advances in Information Retrieval*. Springer, 656–661.

Viktor Mayer-Schönberger and Kenneth Cukier. 2013. *Big Data: A Revolution that Will Transform How we Live, Work, and Think*. Houghton Mifflin Harcourt.

Tao Mei, Bo Yang, Xian-Sheng Hua, and Shipeng Li. 2011. Contextual video recommendation by multimodal relevance and user feedback. *ACM Transactions on Information Systems (TOIS)* 29, 2 (2011), 10.

Andriy Mnih and Ruslan Salakhutdinov. 2007. Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems*. 1257–1264.

Jonghun Park, Sang-Jin Lee, Sung-Jun Lee, Kwanho Kim, Beom-Suk Chung, and Yong-Ki Lee. 2010. Online video recommendation through tag-cloud aggregation. *IEEE MultiMedia* 1 (2010), 78–87.

Michael J. Pazzani and Daniel Billsus. 2007. Content-based recommendation systems. In *The Adaptive Web*. Springer, 325–341.

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

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

Jitao Sang, Zhengyu Deng, Dongyuan Lu, and Changsheng Xu. 2015. Cross-OSN user modeling by homogeneous behavior quantification and local social regularization. *IEEE Transactions on Multimedia* 17, 12 (2015), 2259–2270.

Guy Shani, Max Chickering, and Christopher Meek. 2008. Mining recommendations from the web. In *Proceedings of the 2008 ACM Conference on Recommender Systems*. ACM, 35–42.

Yue Shi, Martha Larson, and Alan Hanjalic. 2014. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys (CSUR)* 47, 1 (2014), 3.

Ajit P. Singh and Geoffrey J. Gordon. 2008. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 650–658.

Xiaoyuan Su and Taghi M. Khoshgoftaar. 2009. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence* 2009 (2009), 4.

Zhi Wang, Lifeng Sun, Wenwu Zhu, Shiqiang Yang, Hongzhi Li, and Dapeng Wu. 2013. Joint social and content recommendation for user-generated videos in online social network. *IEEE Transactions on Multimedia* 15, 3 (2013), 698–709.

Ming Yan, Jitao Sang, Tao Mei, and Changsheng Xu. 2013. Friend transfer: Cold-start friend recommendation with cross-platform transfer learning of social knowledge. In *2013 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6.

Ming Yan, Jitao Sang, and Changsheng Xu. 2015. Unified YouTube video recommendation via cross-network collaboration. In *Proceedings of the 5th ACM Conference on International Conference on Multimedia Retrieval*. ACM, 19–26.

Bo Yang, Tao Mei, Xian-Sheng Hua, Linjun Yang, Shi-Qiang Yang, and Mingjing Li. 2007. Online video recommendation based on multimodal fusion and relevance feedback. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*. ACM, 73–80.

Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web*. ACM, 22–32.

Received September 2015; revised January 2016; accepted May 2016