Cross-OSN User Modeling by Homogeneous Behavior Quantification and Local Social Regularization

Jitao Sang, Zhengyu Deng, Dongyuan Lu, and Changsheng Xu, *Fellow, IEEE*

***Abstract*—In the context of social media services, data shortage has severally hindered accurate user modeling and practical personalized applications. This paper is motivated to leverage the user data distributed in disparate online social networks (OSN) to make up for the data shortage in user modeling, which we refer to as "cross-OSN user modeling." Generally, the data that the same user distributes in different OSNs consist of both behavior data (i.e., interaction with multimedia items) and social data (i.e., interaction between users). This paper focuses on the following two challenges: 1) how to aggregate the users' cross-OSN interactions with multimedia items of the same modality, which we call cross-OSN homogeneous behaviors, and 2) how to integrate users' cross-OSN social data with behavior data. Our proposed solution to address the challenges consist of two corresponding components as follows. 1) Homogeneous behavior quantification, where homogeneous user behaviors are quantified based on their importance in reflecting user preferences. After quantification, the examined cross-OSN user behaviors are aggregated to construct a unified user-item interaction matrix. 2) Local social regularization, where the cross-OSN social data is integrated as regularization in matrix factorization-based user modeling at local topic level. The proposed cross-OSN user modeling solution is evaluated in the application of personalized video recommendation. Carefully designed experiments on self-collected Google+ and YouTube datasets have validated its effectiveness and the advantage over single-OSN-based methods.**

***Index Terms*—Behavior fusion, cross-OSN user modeling, local social regularization, personalization, video recommendation.**

I. Introduction

SOCIAL MEDIA has exploded beyond anyone's wildest imagination, and User Generated Content (UGC) is propagated online at an unparalleled level. Taking YouTube for example, there were in total three billion videos on this website by the end of 2014 with more than 300 hours of new videos being uploaded every minute. In the face of the critical information overload and the increasing customized user needs, personalized recommendation services stand out as requiring solutions and play a more and more important role in the exploration and discovery of interesting resources [1].

In contrast to the huge volume of data to be served, the number of accessible personal user data is very limited. We have observed in our previous study that a typical YouTube user has less than 50 video-related behaviors (e.g., upload, favorite, add-to-playlist) [2]. In the past decade, many researchers have devoted to the problem of user modeling and proposed advanced models/algorithms to complement the user data shortage for personalized services. For example, factorization models are proposed to project the users and items onto a low-dimension space to capture the underlying structure [3], user behavior and registered information is exploited with regularization of the available social interactions [4]. However, the task of user modeling still remains open. The notorious *data sparsity* issue has severely hindered accurate user modeling and practical personalized services [5], [6].

Nowadays, many users are using a multitude of Online Social Network (OSN) services, such as Google+, Twitter, YouTube, etc. Global Web Index 2015 has reported that within the investigated 50 OSNs, each individual holds user accounts on an average of 5.54 OSNs, and actively participate in 2.82 OSNs.[1] In this context, users' data distribute among different OSNs which together record people's integral online footprint and reflect their demographics, as well as interests from a variety of perspectives. This work falls into the topic of *cross-OSN user modeling* [7], i.e., leveraging the user data distributed in different OSNs for user modeling to address the *data sparsity* issue under single-OSN situation.

Generally, in cross-OSN user modeling, there consists of two types of user data for fusion from different OSNs, i.e., the behavioral data indicating the interaction between user and the multimedia items, and the social data indicating the interaction between users. Regarding fusing the behavioral data, two types of data are further involved, the cross-OSN heterogeneous behaviors where the interacted items are from different modalities (e.g., YouTube video favorite and Twitter tweeting), and the cross-OSN homogeneous behaviors where the interacted items are from the same modality (e.g., YouTube video favorite and

Manuscript received June 22, 2015; revised September 25, 2015; accepted September 27, 2015. Date of publication October 12, 2015; date of current version November 13, 2015. This work was supported in part by National Basic Research Program of China under Grant 2012CB316304, in part by the National Natural Science Foundation of China under Grant 61225009, Grant 61332016, and Grant 61303176, and in part by the Beijing Natural Science Foundation under Grant 4131004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Cees Snoek.

J. Sang, Z. Deng, and C. Xu are with the National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: jtsang@nlpr.ia.ac.cn; zydeng@nlpr.ia.ac.cn; csxu@nlpr.ia.ac.cn).

D. Lu is with the School of Information Technology and Management, University of International Business and Economics, Beijing 100029, China (e-mail: ludy@uibe.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2015.2486524

[1][Online]. Available: http://www.globalwebindex.net/blog/internet-users-have-average-of-5-social-media-accounts.

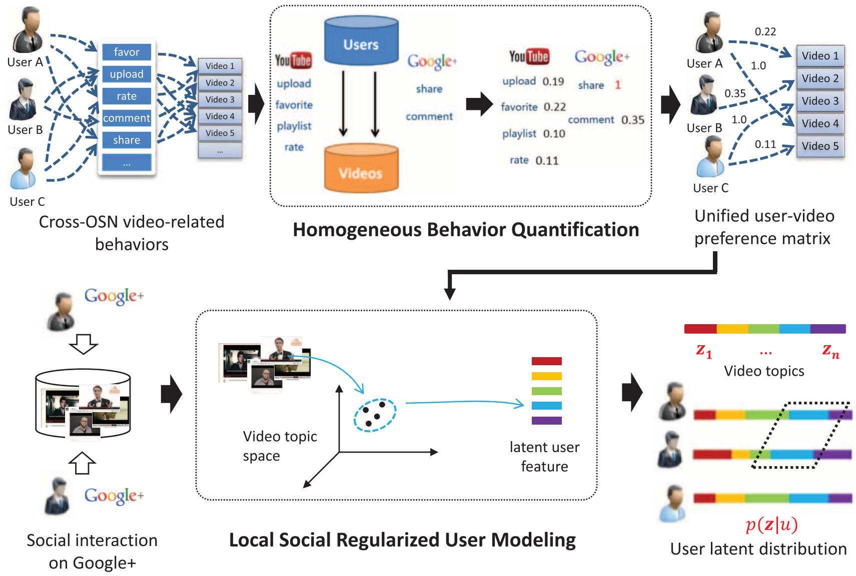


Fig. 1. Proposed solution framework.

Google+ video "+1"). In our previous work [8], [2], we have proposed solutions to address the fusion of cross-OSN heterogeneous behaviors by mining the association between users' interactions with different modalities of items. The first goal of this work is to fuse the other type of behavioral data, the cross-OSN homogeneous behaviors. Taking video-related cross-OSN behaviors as an example: the "upload", "favorite" behaviors on YouTube, and the "+1", "share" behaviors on Google+ all indicate users' interest in the videos. To fuse these behaviors for user modeling, we need to know how significant each behavior is in reflecting users' preferences. Till now, the discrepant significance of homogeneous behavior remains unexplored. Regarding fusing the social data, current solutions are devoted to adopting the social interactions as the global regularization, either by propagating interests among users with close social relations based on random walk [9], [10], or by modifying a latent factor model and regularizing the user factors [11], [12]. Both lines of solutions lead to the result that the socially interacting users will have similar feature vectors in all dimensions. However, in most cases, users connect with each other only because they share interests in a certain field. For example, colleagues in a securities company share interests in finance, daily friends share interests in classical music, family members share interests in TV shows, etc. Therefore, the second goal of this work is to fuse user-user social interaction with user-item behavioral data at a local level in the derived user model.

We achieve the above mentioned two goals by using the video-related behaviors and social interactions on YouTube and Google+ as example, and evaluating the cross-OSN user modeling performance in the context of personalized video recommendation application. Regarding fusing cross-OSN homogeneous behaviors, if each type of behavior constitutes a user-item interaction matrix, the most straightforward way is to construct a unified user-item matrix. Existing solutions either conduct the fusion at the decision stage by tuning the weight according to single behavior-based performance [13], or manually set the behavior significance according to priori knowledge and experiences. This work will conduct an exploratory study to quantify the different contributions for each behavior and how it indicates users' preferences. With each type of behavior defining a kernel, a multi-kernel learning strategy is designed to intimate the real preference and discover the behavioral importance as the kernel weight.

After behavior quantification, the cross-OSN homogeneous behaviors are aggregated to construct a unified user-video interaction matrix and we can conduct user modeling directly based on this matrix. To fuse the user social data into user model, we propose *local social regularization* to regularize the learned user models at fine-grained level. Two tasks need to be addressed to achieve this goal: 1) making the derived user models interpretable and able to be correlated with the videos; and 2) projecting the observed user social interactions onto certain dimensions of user models. For the first task, the work exploits Modified Fuzzy C-Means (MFCM), which combines the advantages of Matrix Factorization and Fuzzy C-Means and can obtain item clusters as well as the user distribution (membership) over these clusters. For the second task, the work investigates an iterative strategy to locally regularize the dynamically updated user models. We summarize the framework of the proposed solution in Fig. 1. The inputs include user video-related behaviors on Google+ and YouTube, user social interactions on Google+. The outputs are the generated video topics and user latent distributions. The main contributions of this paper are as follows.

1) We present a novel framework of cross-OSN user modeling to investigate two previously unexplored challenges: homogeneous behavior discrepancy and social interaction locality.
2) The challenge of homogeneous behavior discrepancy is addressed by exploiting the correlation between different behaviors to quantify their respective significance to express user preferences. The challenge of social interaction locality is addressed by a novel iterative updating user modeling which integrate the user social interaction at fine-grained topic level.

The rest of this paper is organized as follows. The related work on cross-OSN user modeling is briefly reviewed in Section II. In Section III, we present the proposed homogeneous behavior quantification solution. Following that Section IV elaborates the local social regularized user modeling solution. Experimental results, analysis and discussion are reported in Section V. Finally, Section VI concludes this work with future directions.

II. Related Work

Nowadays, many users are not only using multiple Online Social Network (OSN) services but willing to disclose their cross-OSN user accounts. This opens up possibilities for the researchers to analyze people's online complete footprint and exploit the cross-OSN ata to solve many challenging problems which cannot be well explored under single-OSN situation. Szomszor and Alali [14] conducted a very early cross-OSN user modeling work by proposing to integrate user behaviors on Flickr and Delicious by matching user-contributed tags on the two OSNs to Wikipedia categories. Abel *et al.* analyzed the characteristics and overlap of tag-based user profiles on Flickr, Twitter and Delicious, and developed several heuristic cross-OSN user modeling solutions [15], [16]. Yuan *et al.* [17] proposed to model the lifestyle spectrum of a group of individuals based on their online behaviors in different types of OSNs. A hierarchical topic model is presented, where each topic corresponds to a specific life pattern in certain OSN.

Most of the above mentioned work are devoted to fusing the cross-OSN user data without considering the correlations between them. In recent years, we have conducted a series of work on cross-OSN analysis and addressing the task of user modeling by explicitly exploring the cross-OSN user data correlations. In [18], we observed that users have similar social and behavioral correlations on Twitter and Flickr, based on what we proposed a "coldest-start" Twitter friend recommendation problem by only utilizing the user behavior and social data from Flickr. In [19], based on the data observation that the same user responses faster on Twitter than on YouTube to the same emerging event/topics, we propose a temporal cross-OSN user modeling solution by exploiting users' Twitter and YouTube data for short-term and long-term interest estimation respectively. Recently, we proposed to mine the correlation between users' heterogeneous cross-OSN behaviors, e.g., the Twitter tweeting and YouTube watching behaviors. The discovered correlations are then utilized in the application of Twitter-assistant YouTube video promotion [8] and YouTube video recommendation [2]. In this paper, we aim at addressing two major challenges in cross-OSN user modeling that are not considered by previous work: 1) examining the significance of the cross-OSN interactions with the same type of items, and 2) integrating user social interactions on Social Networking Site (SNS) with user behaviors at fine-grained topic level.

III. Homogeneous Behavior Quantification

This work will use the homogeneous video-related cross-OSN behaviors as an example, i.e., favorite, rate, upload, add-to-playlist on YouTube, and share, comment, +1 on Google+. Both the motivation and the solution are expected to be readily generalized to other behaviors. To quantify and fuse the different video-related behaviors, the basic assumption is that: the different homogeneous behaviors are correlated and deserve similar significance weights if they are consistent in terms of user interactions with unique videos. For example, when a user watched an interesting video, he/she may comment, favorite it and add it to playlist. This indicates that some behaviors are consistent in expressing user preferences to some degree. If a group of users simultaneously conduct different behaviors on the same video, we are confident to claim that these behaviors are correlated and have similar significance in reflecting user preferences. Inspired by this, two solutions are proposed to examine the correlation between homogeneous behaviors for quantification, which will be detailed in the following two subsections respectively.

A. Heuristic Quantification

The first quantification solution is to directly calculate the behavior correlation based on their co-occurrence in terms of users' interaction with the same videos. Given a type of behavior, one user's interaction with the videos can be encoded into a binary vector where the element is set to 1 if this user has interacted with the video and 0 otherwise. Since the number of user interaction is very small compared to the video collection, the raw user-video vectors are extremely sparse and cannot be directly used to calculate the correlation. Therefore, we first cluster the videos based on their co-interactions with users. Specially, we view each user as a document and the videos user interacted with as words. Latent Dirichlet Allocation (LDA) [20] is adopted to learn the latent video topics which are represented by the occurrence probability of each video. Each video is represented over the discovered latent video topics, where K-means is utilized to obtain the video clusters.

After clustering the videos, we can transfer the user-video interaction vectors to user-cluster vectors. Person Correlation Coefficient (PCC) [21] is adopted to calculate the correlations between these homogeneous behaviors. Given user u, the correlation between YouTube "favorite" and Google+ "share" behavior is calculated as follows:

$$\mathrm{corr}(\mathbf{a}_{uf}, \mathbf{a}_{us}) = \frac{\sum_{i=1}^{N'}(a_{uf}^i - \bar{a}_{uf})(a_{us}^i - \bar{a}_{us})}{\sqrt{\sum_{i=1}^{N'}\left(a_{uf}^i - \bar{a}_{uf}\right)^2}\sqrt{\sum_{i=1}^{N'}\left(a_{us}^i - \bar{a}_{us}\right)^2}} \quad (1)$$

where $\mathbf{a}_{uf}$, $\mathbf{a}_{us}$ denote the user-cluster vector for YouTube "favorite" and Google+ "share" behaviors; a_{uf}^i, a_{us}^i denote the interaction value of user u on the ith cluster for YouTube "favorite" and Google+ "share" behaviors; $\bar{a}_{uf}$, $\bar{a}_{us}$ denote the corresponding average interaction value; N' denotes the number of clusters in K-means. The final relative significance of different behaviors is obtained by aggregating the correlation scores of all the examined users.

B. Learning-Based Quantification

The above introduced heuristic quantification is straightforward and easy to realize. In this subsection, we introduce another learning-based quantification solution to deal with the inevitable noise in the observed raw user-video interactions.

We utilized Multiple Kernel Learning (MKL) [22] to examine the consistency between users' different types of interactions with videos. In this case, each type of behavior is considered as a kernel. Selecting one user-video cluster matrix as the reference, the task is to use the combination of the other behavior kernel vectors to fit the selected reference kernel vector. Google+ "share" behavior is selected as the reference behavior due to the following reason: Google+ is an extension of YouTube video consuming to socializing. We observed that 76.5% of the shared videos in Google+ are from YouTube. Google+ "share" behavior connects the Google+ with YouTube where these cross-OSN shared videos strongly indicate user preferences.[2] The formulation of the linear combination for user u is as follows:

$$\mathbf{a}_{ur} = \sum_{i=1}^{N_k} \varphi_i * \mathbf{a}_{ui} \tag{2}$$

where $\mathbf{a}_{ur}$ is the reference user-video vector, φ_i is the linear parameter for the corresponding behavior, and N_k is the number of homogeneous behavior types except the reference behavior. For all users, the formulation is as follows:

$$\mathbf{A}_r = \sum_{i=1}^{N_k} \varphi_i * \mathbf{A}_i \tag{3}$$

where $\mathbf{A}_r \in \mathbb{R}^{M \times N'}$, $\mathbf{A}_i \in \mathbb{R}^{M \times N'}$, M indicates the number of users, and N' indicates the number of video clusters.

For model inference, the kernel matrices need first to be transferred to be square. Equation (3) is reformulated by multiplying the corresponding matrix transpose, i.e., $\mathbf{K}_r = \mathbf{A}_r^T * \mathbf{A}_r$, $\mathbf{K}_i = \mathbf{A}_i^T * \mathbf{A}_i$. Equation (3) is rewritten as

$$\mathbf{K}_r = \sum_{i=1}^{N_k} \varphi_i * \mathbf{K}_i \tag{4}$$

where the matrixes $\mathbf{K}_i \in \mathbb{R}^{N' \times N'}$, $\mathbf{K}_r \in \mathbb{R}^{N' \times N'}$.

A kernel-based learning technique is leveraged to find the optimal combination of multiple kernels by following the principles of KTA [23]. Specifically, we first centralize the kernel matrices [24] as follows:

$$[\mathbf{K}]_{ij} = K_{ij} - \frac{1}{N'} \sum_{i=1}^{N'} K_{ij} - \frac{1}{N'} \sum_{j=1}^{N'} K_{ij} + \frac{1}{N'^2} \sum_{i,j=1}^{N'} K_{ij}. \tag{5}$$

Kernel alignment is then adopted to measure the quality of the kernel $\mathbf{K}_i$ with respect to the target reference matrix $\mathbf{K}_r$

$$\rho(\mathbf{K}_r, \mathbf{K}_i) = \frac{\mathbf{E}[tr(\mathbf{K}_r \mathbf{K}_i)]}{\sqrt{\mathbf{E}[tr(\mathbf{K}_r \mathbf{K}_r)] \mathbf{E}[tr(\mathbf{K}_i \mathbf{K}_i)]}}. \tag{6}$$

[2]We emphasize that since the derived weights indicate the relative significance of different behaviors. The proposed solution has no requirements on the selection of certain behavior as the reference behavior.

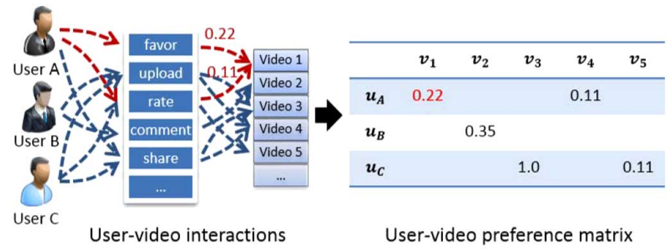


Fig. 2. User-video preference matrix.

It is easy to obtain the optimal kernel weights by maximizing the alignment ρ over $\mathbf{K}_i$. The solution φ^* of the optimization problem is given by

$$\boldsymbol{\varphi}^* = \arg\min \boldsymbol{\varphi}^T \mathbf{M} \boldsymbol{\varphi} - 2\boldsymbol{\varphi}^T \mathbf{b} \tag{7}$$

where $\mathbf{b} = [\mathbf{tr}(\mathbf{K_1 K_r}), \ldots, \mathbf{tr}(\mathbf{K_{N_k} K_r})]^{\mathbf{T}}$, and $[\mathbf{M}]_{\mathbf{ij}} := \mathbf{tr}(\mathbf{K_i K_j})$, $i,j \in N_k$.

Assuming the significance weight of the reference behavior as 1, the derived normalized weight parameters are the desired behavior significance to reflecting user preferences. Given the behavior significance weights, we can easily aggregate the different behaviors to construct a unified user-video interaction matrix denoted by $\mathbf{R}$. $\mathbf{R}$ is defined with each element setting as the maximum weight among the user-video interaction behaviors, i.e.

$$[\mathbf{R}]_{mn} = \begin{cases} 1, & \text{user } u_m \text{ shared video } v_n \\ \max I_{mn}^i \varphi_i, & \text{otherwise} \end{cases}$$

where I_{mn}^i is an indicator function that equals 1 if user u_m has the corresponding behavior on video v_n and equals 0 otherwise. One example is illustrated in Fig. 2.

IV. Local Social Regularization

OSNs have different functionality focuses, e.g., YouTube focuses on user-content behavioral interaction, and Google+ focuses on user-user social interaction. This section focuses on integrating the cross-OSN social data with behavioral data, i.e., users' social interaction on Google+ with the aggregated homogeneous behaviors from above section. For integrating the social interaction between users into the final user model, the basic premise is that: users usually interact with each other because they share interests in certain fields. For example, one user is interested in sport and finance. When he/she shared a video on "European Cup" on Google+, his/her sport friends who share an interest in sport are very likely to comment or reshare it, while those who share an interest in finance may not interact with this video. These social interactions only reflect the common interests for sports. Therefore, at this stage, the goal is to leverage the observed social interactions as local regularization that corresponds to certain dimensions (e.g., sports) in the derived user models. We have proposed a MFCM based solution to achieve this goal.

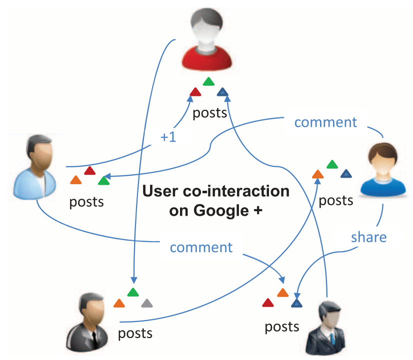


Fig. 3. Implicit social interactions on Google+.

The first step is to associate the user models with the videos. We briefly introduce how to utilize MFCM to factorize the unified user-video matrix to obtain video latent topics and user model (user distribution over these topics). Specifically, given the user-video interaction matrix $\mathbf{R} \in \mathbb{R}^{M \times N}$, where $M(N)$ is the number of users (videos), the objective function of MFCM is as follows:

$$\begin{aligned} H(\mathbf{Z}, \mathbf{C}) &= \|\mathbf{R} - \mathbf{Z}\mathbf{C}^T\|_F^2 + \lambda(\|\mathbf{Z}\|_F^2 + \|\mathbf{C}\|_F^2) \\ &= \frac{1}{2} \sum_{u,v \in \mathcal{P}} \left(r_{u,v} - \sum_{k=1}^{K} z_{uk} c_{vk} \right)^2 \\ &\quad + \frac{\lambda}{2} \left(\sum_{u \in \mathcal{P}} \sum_{k=1}^{K} z_{uk}^2 + \sum_{v \in \mathcal{P}} \sum_{k=1}^{K} c_{vk}^2 \right) \\ s.t. \quad & \sum_{k=1}^{K} z_{uk} = 1 \quad \text{and} \quad z_{uk} \geq 0 (k \in [1, K]) \end{aligned} \tag{8}$$

where $\mathbf{Z} \triangleq (\mathbf{z}_1, \ldots, \mathbf{z}_M)^T \in \mathbb{R}^{M \times K}, \mathbf{C} \triangleq (\mathbf{c}_1, \ldots, \mathbf{c}_K) \in \mathbb{R}^{N \times K}$, $\mathbf{z}_i(\mathbf{c}_i)$ is the distribution vector of the ith user on each video topic, K is the number of video topics, $r_{u,v}$ is the observed interaction value of user u on video v, z_u is the derived user model where z_{uk} indicates the preference significance of user u on video topic k, c_{vk} is the probability of video v belonging to topic k, λ is the penalty parameter to prevent overfitting.

In the following, we first introduce the traditional global social regularization in MFCM model. Thereafter, we elaborate how to measure user similarity at topic level and how to integrate the topic-level similarity into MFCM model to realize local social regularization.

A. Global Social Regularization

In factor models, existing methods assume that the similar users will have similar feature vectors and introduce users' similarities to constrain all dimensions of their feature vectors. The similarities can be measured by either the explicit (e.g., social relation) or the implicit (e.g., co-interaction with the same multimedia item) user social interactions. On Google+, since the explicit social relation (i.e., Google circle) information is unaccessible via API, we use the co-interaction with Google+ posts to calculate the social similarity between users. The co-interaction between users on Google+ is illustrated in Fig. 3. Typically the user similarity is integrated into factor models as global regularization term.

The objective function of MFCM with global social regularization is as follows:

$$\begin{aligned} H(\mathbf{Z}, \mathbf{C}) &= \|\mathbf{R} - \mathbf{Z}\mathbf{C}^T\|_F^2 + \lambda(\|\mathbf{Z}\|_F^2 + \|\mathbf{C}\|_F^2) \\ &\quad + \lambda_s \sum_{u,u' \in \mathcal{P}} \text{sim}(u, u') \|\mathbf{z}_u - \mathbf{z}'_u\|_F^2 \\ &= \frac{1}{2} \sum_{u,v \in \mathcal{P}} \left(r_{u,v} - \sum_{k=1}^{K} z_{uk} c_{vk} \right)^2 \\ &\quad + \frac{\lambda}{2} \left(\sum_{u \in \mathcal{P}} \sum_{k=1}^{K} z_{uk}^2 + \sum_{v \in \mathcal{P}} \sum_{k=1}^{K} c_{vk}^2 \right) \\ &\quad + \frac{\lambda_s}{2} \sum_{u,u' \in \mathcal{P}} \sum_{k=1}^{K} \text{sim}(u, u')(z_{uk} - z_{u'k})^2 \\ s.t. \quad & \sum_{k=1}^{K} z_{uk} = 1 \quad \text{and} \quad z_k \geq 0 (k \in [1, K]) \end{aligned} \tag{9}$$

where $sim(u, u')$ denotes the similarity between the users u and u', λ_s is the weight parameter to control the influence of social interactions.

B. Local Social Regularization

As mentioned in the introduction, users often share similar interests on certain topic. The social interactions between users can reflect their shared interests at topic level, which should be considered in user modeling as the local regularization. By the standard MFCM, we can associate the derived user models with video topics. This provides a way to project the observed user-user social interactions with videos onto the derived user models, i.e., the topic-level user similarity. However, the video latent topics obtained by MFCM dynamically change during the iteration process. Therefore, the topic-level user similarity must keep pace with the update of the video latent topics. Since each video topic is represented by the occurrence probability (membership) of videos, our basic premise to address this problem is to calculate the user similarity based on the probability distribution over the videos in advance. The topic-level user similarity can then be dynamically computed by projecting the probability distribution to the derived video topics. Enlightened by this, we design a scheme to dynamically integrate topic-level user similarity.

We use $\mathbf{v}_i$ to denote the TF-IDF vector of the ith video calculated from its metadata including video title, tags and description. The social interactions between users u and u' can be represented by a similar TF-IDF vector $\mathbf{l}_{uu'}$ constituted by the textual metadata of the videos co-interacted by them. Therefore, we project the user similarity $\mathbf{l}_{uu'}$ to each video vector $\mathbf{v}_i$ and can obtain the probability distributions of user similarity over the videos, denoted by $\mathbf{s}_{uu'} \in \mathbb{R}^N$, where the element in $\mathbf{s}_{uu'}$ is defined as

$$s_{uu'i} = \mathbf{l}_{uu'}^T \mathbf{v}_i = \sum_{w=1}^{W} l_{uu'w} v_{iw}, i \in [1, N] \tag{10}$$

where $l_{uu'w}, v_{iw}$ is the weight of the wth word in vector $\mathbf{l}_{uu'}$ *and* $\mathbf{v}_n$; W is the size of the word vocabulary.

Afterwards, we project the video-level distribution vector $\mathbf{s}_{uu'}$ to each video topic to obtain the topic-level user similarity distribution vector, denoted by $\mathbf{t}_{uu'} \in \mathbb{R}^K$. The element in $\mathbf{t}_{uu'}$ is calculated as follows:

$$t_{uu'k} = \mathbf{s}_{uu'}^T \mathbf{c}_k = \sum_{i=1}^{N} s_{uu'i} c_{ik}, \quad k \in [1, K]. \tag{11}$$

The derived topic-level user similarity vector $\mathbf{t}_{uu'}$ can be then integrated into MFCM model as follows:

$$\begin{aligned} H(\mathbf{Z}, \mathbf{C}) &= \|\mathbf{R} - \mathbf{Z}\mathbf{C}^T\|_F^2 + \lambda\left(\|\mathbf{Z}\|_F^2 + \|\mathbf{C}\|_F^2\right) \\ &\quad + \lambda_s \sum_{u,u' \in \mathcal{P}} \|(\mathbf{z}_u - \mathbf{z}_{u'}) \odot \mathbf{t}_{uu'}\|_F^2 \\ &= \frac{1}{2} \sum_{u,v \in \mathcal{P}} \left(r_{u,v} - \sum_{k=1}^{K} z_{uk} c_{vk}\right)^2 \\ &\quad + \frac{\lambda}{2} \left(\sum_{u \in \mathcal{P}} \sum_{k=1}^{K} z_{uk}^2 + \sum_{v \in P} \sum_{k=1}^{K} c_{vk}^2\right) \\ &\quad + \frac{\lambda_s}{2} \sum_{u,u' \in \mathcal{P}} \sum_{k=1}^{K} (z_{uk} - z_{u'k})^2 \left(\sum_{n=1}^{N} s_{uu'n} c_{nk}\right)^2 \\ s.t. \quad & \sum_{k=1}^{K} z_{uk} = 1 \quad and \quad z_k \geq 0 (k \in [1, K]). \end{aligned} \tag{12}$$

It is easy to conceive that the users' similarities at topic-level are introduced into MFCM to constrain the users' latent representation in the user model, which are changing with the update of the video topic space.

C. Model Inference

We introduce two strategies to infer the proposed model: 1) one strictly subjects to the constraints, called MFCMS and 2) the other loosely subjects to the constraints, called MFCML.

1) MFCMS: The constraints in Eq. (12) is equivalent to the formulation as follows:

$$z_{uk} = \frac{e^{q_{uk}}}{\sum_{l=1}^{K} e^{q_{ul}}}. \tag{13}$$

By incorporating the constraints into the objective function, (12) can be rewritten as follows:

$$\begin{aligned} H(\mathbf{Q}, \mathbf{C}) = &\frac{1}{2} \sum_{u,v \in \mathcal{P}} \left(r_{u,v} - \frac{1}{\sum_{l=1}^{K} e^{q_{ul}}} \sum_{k=1}^{K} e^{q_{uk}} c_{vk}\right)^2 \\ &+ \frac{\lambda_s}{2} \sum_{u,u' \in \mathcal{P}} \sum_{k=1}^{K} \left(\frac{e^{q_{uk}}}{\sum_{l=1}^{K} e^{q_{ul}}} - \frac{e^{q_{u'k}}}{\sum_{l=1}^{K} e^{q_{u'l}}}\right)^2 \\ &* \left(\sum_{v=1}^{N} s_{uu'v} c_{vk}\right)^2 + \frac{\lambda}{2} \left(\sum_{u \in \mathcal{P}} \sum_{k=1}^{K} q_{uk}^2 + \sum_{v \in \mathcal{P}} \sum_{k=1}^{K} c_{vk}^2\right). \end{aligned} \tag{14}$$

Then, we adopt gradient descending strategy to infer the model. Let the partial derivatives of $H(\mathbf{Q}, \mathbf{C})$ on q_{uk} and c_{vk} be equal to 0

$$\frac{\partial H}{\partial q_{uk}} = 0, \qquad \frac{\partial H}{\partial c_{vk}} = 0.$$

For each user-video interaction pair $u, v \in \mathcal{P}$, we have

$$\begin{aligned} \delta q_{uk} &= e_{uv} z_{uk} (\hat{r}_{uv} - c_{vk}) \\ &\quad + \lambda q_{uk} + \lambda_s \sum_{u' \in \mathcal{P}} \left(\sum_{n=1}^{N} s_{uu'n} c_{nk}\right)^2 \\ &\quad \times (z_{uk} - z_{u'k}) z_{uk} (1 - z_{uk}) \\ \delta c_{vk} &= -e_{uv} z_{uk} + \lambda c_{vk} \\ &\quad + \lambda_s \sum_{u' \in \mathcal{P}} \sum_{n=1}^{N} s_{uu'n} c_{nk} s_{uu'v} (z_{uk} - z_{u'k})^2 \\ q_{uk} &= q_{uk} - \eta * \delta_{q_{uk}} \\ c_{vk} &= c_{vk} - \eta * \delta_{c_{vk}} \end{aligned} \tag{15}$$

where η denotes the learning rate, $\hat{r}_{uv}$ is the estimated interaction value, $e_{uv} = r_{u,v} - \hat{r}_{uv}$ is the error between real value and estimated value.

2) MFCML: The constraints in (12) can also be directly incorporated into the objective function as follows:

$$\begin{aligned} H(\mathbf{Z}, \mathbf{C}) = &\|\mathbf{R} - \mathbf{Z}\mathbf{C}^T\|_F^2 + \lambda\left(\|\mathbf{Z}\|_F^2 + \|\mathbf{C}\|_F^2 + \|\mathbf{Z}^T\mathbf{1} - \mathbf{1}\|^2 \right. \\ &\left. + \|\mathbf{Z}_{-}\|_F^2\right) + \lambda_s \sum_{u,u'} \|(\mathbf{z}_u - \mathbf{z}_{u'}) \odot \mathbf{t}_{u,u'}\|_F^2 \\ = &\frac{1}{2} \sum_{u,v \in \mathcal{P}} \left(r_{u,v} - \sum_{k=1}^{K} z_{uk} c_{vk}\right)^2 + \frac{\lambda}{2} \left(\sum_{v \in \mathcal{P}} \sum_{k=1}^{K} c_{vk}^2 \right. \\ &\left. + \sum_{u \in \mathcal{P}} \left(\sum_{k=1}^{K} z_{uk} - 1\right)^2 + \sum_{u \in \mathcal{P}} \sum_{k=1}^{K} (z_{uk}^2 + z_{uk\_}^2)\right) \\ &+ \frac{\lambda_s}{2} \sum_{u,u' \in \mathcal{P}} \sum_{k=1}^{K} (z_{uk} - z_{u'k})^2 \left(\sum_{n=1}^{N} s_{uu'n} c_{nk}\right)^2 \end{aligned} \tag{16}$$

where $\mathbf{Z}_{-} = (z_{1\_}, \ldots, z_{K\_})$, $z_{i\_} \triangleq max(0, -z_i)$, $\|\mathbf{Z}_{-}\|_F^2$ forces $z_k \geq 0 (k \in [1, K])$. The term $\|\mathbf{Z}^T\mathbf{1} - \mathbf{1}\|^2$ is to force $\sum_{k=1}^{K} z_{uk} = 1$.

Gradient descending is further adopted for model inference. For each user-video rating pair $u, v \in \mathcal{P}$, we have

$$\begin{aligned} \delta z_{uk} &= -e_{uv} c_{vk} + \lambda \left(z_{uk} - z_{uk\_} + \sum_{l=1}^{K} z_{ul} - 1\right) \\ &\quad + \lambda_s \sum_{u' \in \mathcal{P}} (z_{uk} - z_{u'k}) * \left(\sum_{n=1}^{N} s_{uu'n} c_{nk}\right)^2 \\ \delta c_{vk} &= -e_{uv} z_{uk} + \lambda c_{vk} \\ &\quad + \lambda_s \sum_{u' \in \mathcal{P}} \sum_{n=1}^{N} s_{uu'n} c_{nk} s_{uu'v} (z_{uk} - z_{u'k})^2 \\ z_{uk} &= z_{uk} - \eta * \delta_{z_{uk}} \\ c_{vk} &= c_{vk} - \eta * \delta_{c_{vk}}. \end{aligned} \tag{17}$$

Since the topic-level user similarity must be calculated and updated at each iteration, the time complexity of the proposed local social regularization solution is relative high. Therefore, we adopt a strategy to accelerate the model convergence speed. A simple strategy can be designed to speed up the model convergence. Since topic modeling outputs a similar structure as

TABLE I
NUMBER OF DIFFERENT VIDEO-RELATED BEHAVIORS ON YOUTUBE AND GOOGLE+

YouTube					Google+	
#favor	#upload	#playlist	#comm	#rate	#share+	#comm+
534814	223106	456622	50851	469620	114970	53124

TABLE II
STATISTIC OF USER-VIDEO MATRIX IN OUR DATASET

Statistic	User	Video
Min. No. of interactions	0	5
Max. No. of interactions	215	129
Avg. No. of interactions	18.9	11.2

MFCM, the solution is inspired to leverage the output from topic modeling as the starting-point for the iteration process. Specifically, we run the standard topic modeling method (e.g., Latent Dirichlet Allocation, LDA) and utilize the derived user-topic and topic-word distribution to initialize the model. In experiments, we compare the performance of our model with random and LDA initialization.

V. EXPERIMENTS

A. Dataset and Evaluation Metrics

In our previous studies, we observed that a considerable proportion of users share their YouTube accounts to their Google+ homepages [25]. Enlightened by this, we first collected 10,500 users with available YouTube accounts from Google+ homepages. Then, we crawled these user data from Google+ and YouTube by the respective APIs. On Google+, the user registration and all the posts are collected including their interactions with text, image and video. On YouTube, all the video-related behaviors are collected such as favorite, rating, commenting, uploading and adding to playlist. As a result, we obtained a dataset with 1,903,107 video-related behaviors and 1,083,485 text- and image-related behaviors for the 10,500 users. Table I summarizes the statistics of the video-related homogeneous behaviors.

We evaluate the performance of the proposed cross-OSN user modeling solutions in the context of personalized video recommendation. In order to better illustrate the significance of social interaction to user modeling, we only keep the users who interacted with no less than 10 users in our dataset in the video recommendation experiments. Besides, the videos rated by less than five users are also filtered out. As a result, we obtained a dataset with 2,181 users and 3,667 videos for our experiments. The detailed statistic information about the experimental data is summarized in Table II. We use different amounts of training data (90%, 70%, 50%, 30%, 10%) and the rest to evaluate the performance.

Two evaluation metrics are utilized, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Specifically, RMSE is defined as

$$RMSE = \sqrt{\frac{\sum_{i,j \in \mathcal{P}} \left(R_{ij} - \hat{R}_{ij}\right)^2}{T}} \tag{18}$$

TABLE III
CORRELATIONS BETWEEN DIFFERENT VIDEO-RELATED BEHAVIORS IN YOUTUBE AND GOOGLE+

	favor	upload	playlist	comm	rate	share+	comm+
favor	–	0.680	0.524	0.371	0.666	0.675	0.662
upload	0.680	–	0.500	0.378	0.655	0.687	0.670
playlist	0.524	0.500	–	0.238	0.457	0.488	0.477
comm	0.371	0.378	0.238	–	0.434	0.387	0.379
rate	0.666	0.655	0.457	0.434	–	0.666	0.648
share	0.675	0.687	0.488	0.387	0.666	–	0.719
comm+	0.662	0.670	0.477	0.379	0.648	0.719	–

TABLE IV
OVERALL CORRELATION FOR EACH VIDEO-RELATED BEHAVIOR IN YOUTUBE AND GOOGLE+

favor	upload	playlist	comm	rate	share+	comm+
3.579	3.572	2.684	2.188	3.526	3.623	3.556

TABLE V
LINEAR WEIGHTS OBTAINED BY MKL-BASED BEHAVIOR QUANTIFICATION

favor	upload	playlist	comm	rate	comm+
0.2162	0.1895	0.0968	0.1211	0.1114	0.3559

where $R_{ij}(\hat{R}_{ij})$ denotes the real (estimated) interaction value user i gives to item j, T denotes the total number of the tested interaction elements.

MAE is defined as

$$MAE = \frac{\sum_{i,j \in \mathcal{P}} |R_{ij} - \hat{R}_{ij}|}{T}. \tag{19}$$

B. Experimental Results for Homogeneous Behavior Quantification

We first introduce the parameter settings. In LDA model, there are two hyperparameters: α and β. We empirically fix the parameters according to the prior expectation about the data. The hyperparameter β is fixed to 0.1 and α is set to $50/N_T$ where N_T is the number of the latent topics. We set $N_T = 25$ according to the perplexity examination [26]. In order to guarantee the effectiveness of the experiment, we only keep the users who have all the video-related behaviors in Table I. The number of clusters in K-means is set to 25 which is equal to the number of the latent topics in LDA.

In heuristic quantification, the average correlations over all users are shown in Table III. We can see that the correlations between behaviors are quite different. The correlations of each behavior with the others are summed up and shown in Table IV. It is shown that the total correlations for these behaviors are not simply proportional to their numbers shown in Table I. Google+ "share" behavior has the largest overall correlation indicating its high overlap with other behaviors. This also validates our motivation to use the "share" behavior as the reference behavior in learning-based quantification. The total correlation scores listed in Table I are normalized to obtain the significance weight for each behavior in the fused user-video matrix for user modeling. We record this method as *Heuristic* in the video recommendation experiment.

The learning-based behavior quantification results are demonstrated in Table V. We can see that Google+ "comment" (denoted as *comm+*) behavior has the biggest weight, while

TABLE VI
PERFORMANCE COMPARISON OF LSOCMFCMS WITH DIFFERENT BEHAVIOR QUANTIFICATION STRATEGIES

Training data	Metrics	MKL	Heuristic	Equivalent
90%	RMSE	0.271	0.2848	0.2959
	MAE	0.2205	0.2222	0.2326
70%	RMSE	0.2828	0.3089	0.3197
	MAE	0.2272	0.2350	0.2384
50%	RMSE	0.3127	0.3220	0.3316
	MAE	0.2373	0.2478	0.2496
30%	RMSE	0.334	0.3373	0.358
	MAE	0.2535	0.2617	0.2761
10%	RMSE	0.3613	0.3662	0.3701
	MAE	0.2657	0.2697	0.2844

YouTube "comment" behavior has very small weight. This can be explained by the fact that on Google+, the commented videos are often shared by their friends and the comment behaviors are often positive. Besides, YouTube "favorite" behavior also obtains significant weight, while YouTube "add-to-playlist" behavior has the smallest weight. This result is reasonable as "favorite" behavior explicitly express the user's preference to the video, while the behavior of "add-to-playlist" conveys much fewer preference hints.

We evaluate the performance of homogeneous behavior quantification in the context of video recommendation applications. Specifically, three behavior quantification solutions are compared: 1) *Equivalent*, the different behaviors are assigned with equal significance weight; 2) *Heuristic*, the different behaviors are quantified based on their total correlations with the other behaviors (see Table IV); and 3) *MKL*, the proposed learning-based behavior quantification solution. For user modeling method, we utilize the proposed local social regularized MFCM with strict constraints (denoted as *LSocMFCMS*). The recommendation results in terms of RMSE and MAE are shown in Table VI. We can observe that the performance of personalized video recommendation based on the proposed learning-based behavior quantification strategy is consistently superior than those based on heuristic and equivalent quantification strategies. We explain this result that by kernel learning we can reduce the influence from noisy user-video interactions and discover the underlying correlation between the different behaviors.

C. Experimental Results for Local Social Regularization

In order to measure the effectiveness of local social regularization, we compare our proposed approach with several social recommendation approaches as well as the basic Matrix Factorization-based methods. The examined methods are as follows:

- PMF: the basic probabilistic matrix factorization method [27];
- SocMF1: integrating social relations by global average-based regularization [11];
- SocMF2: integrating social relations by global individual-based regularization [12];
- MFCMS: the basic MFCM method with strict constraints without social regularization;
- MFCML: the basic MFCM method with loose constraints without social regularization;
- GSocMFCMS: the MFCMS with global social regularization;
- GSocMFCML: the MFCML with global social regularization;
- LSocMFCMS: the MFCMS with local social regularization; and
- LSocMFCML: the MFCML with local social regularization.

The global user similarity in SocMF1, SocMF2, MFCMS and MFCML is measured by the cosine value between user vectors, which are obtained by extracting the semantic words in the content information of user posts on Google+.[3]

Three major parameters are involved in social recommendation: K, λ and λ_s. K is the number of the video clusters. The regularization parameter λ is to prevent the overfitting of the model. The parameter λ_s is to control the influence of social interactions. It is impractical for all the examined methods to achieve the best performance by tuning these parameters. Therefore, according to the prior knowledge, we fix the number of video clusters $K = 25$ which is equal to the number of latent topics in LDA, and set $\lambda = 0.05$ for all the models to ensure that PMF and MFCML(S) have the similar performance [28]. Since λ_s has different scales in the examined methods, we set λ_s to a relatively small value of $\lambda_s = 0.0001$ in GSocMFCML, GSocMFCMS, LSocMFCML and LSocMFCMS models, and a relatively big value of $\lambda_s = 0.1$ in SocMF1 and SocMF2 models. This setting guarantees that these models achieve a good performance when $\lambda = 0.05$. In the following, we first conduct experiments based on the above parameter values. Then, we examine the influence of the parameters λ_s and K on the performance of the proposed approaches.

1) Comparison of Different Methods: The experimental results of different video recommendation methods are demonstrated in Table VII and Table VIII in terms of RMSE and MAE, respectively. We have the following observations. 1) MFCM with local social regularization (LSocMFCML and LSocMFCMS) outperforms the other strategies consistently in different experimental settings. 2) The performance of LSocMFCMS (GSocMFCMS) is slightly better than that of LSocMFCML (GSocMFCML) due to the fact that LSocMFCMS strictly subjects to the constraints. However, LSocMFCMS is more time-consuming than LSocMFCML in the training process since the user distribution needs to be completely updated at each iteration step. 3) The performance of SocMF1 and SocMF2 is much better than that of the basic MFCMS (PMF), which demonstrates the effectiveness of social regularization. 4) LSocMFCMS achieves superior performance than the global social regularization model GSocMFCMS. Compared with the state-of-the-art global social recommendation methods, SocMF1 and SocMF2, the relative improvement of LSocMFCMS is around 3%.

Furthermore, in order to examine the influence of each behavior on the performance of our proposed approach, we conduct experiments with each behavior missing, respectively. Table IX demonstrates the experimental results. We can see

[3]We only keep noun words which are the least noisy representations for user interests.

TABLE VII
PERFORMANCE COMPARISON OF DIFFERENT STRATEGIES BY RMSE

Training data	Metrics	PMF	SocMF1	SocMF2	MFCML	MFCMS	GSocMFCML	GSocMFCMS	LSocMFCML	LSocMFCMS
90%	RMSE	0.2998	0.2802	0.2869	0.3025	0.2976	0.2907	0.2825	0.2742	**0.2710**
	Improve	9.61%	3.28%	5.54%	10.41%	8.94%	6.78%	4.07%		
70%	RMSE	0.3157	0.2928	0.3045	0.315	0.3198	0.3103	0.2941	0.2852	**0.2828**
	Improve	10.42%	3.42%	7.13%	10.22%	11.57%	8.86%	3.84%		
50%	RMSE	0.3371	0.3168	0.3282	0.3226	0.3215	0.3247	0.3277	0.3203	**0.3127**
	Improve	7.24%	1.29%	4.72%	3.07%	2.74%	3.70%	4.58%		
30%	RMSE	0.3568	0.3443	0.3542	0.3476	0.3435	0.347	0.3391	0.3452	**0.3340**
	Improve	6.39%	2.99%	5.70%	3.91%	2.77%	3.75%	1.50%		
10%	RMSE	0.3761	0.374	0.376	0.3768	0.3673	0.3764	0.3639	0.3767	**0.3613**
	Improve	3.94%	3.40%	3.91%	4.11%	1.63%	4.01%	0.71%		

TABLE VIII
PERFORMANCE COMPARISON OF DIFFERENT STRATEGIES BY MAE

Training data	Metrics	PMF	SocMF1	SocMF2	MFCML	MFCMS	GSocMFCML	GSocMFCMS	LSocMFCML	LSocMFCMS
90%	MAE	0.2362	0.2255	0.2289	0.2405	0.2337	0.2304	0.2273	0.2218	**0.2205**
	Improve	6.65%	2.22%	3.67%	8.32%	5.65%	4.30%	2.99%		
70%	MAE	0.239	0.2377	0.2308	0.2483	0.2343	0.2448	0.2304	0.2332	**0.2272**
	Improve	4.94%	4.42%	1.56%	8.50%	3.03%	7.19%	1.39%		
50%	MAE	0.2542	0.2382	0.2474	0.2513	0.2522	0.253	0.2395	0.253	**0.2373**
	Improve	6.65%	0.38%	4.08%	5.57%	5.91%	6.21%	0.92%		
30%	MAE	0.2695	0.2594	0.2682	0.2672	0.2612	0.2668	0.2572	0.2685	**0.2535**
	Improve	5.94%	2.27%	5.48%	5.13%	2.95%	4.99%	1.44%		
10%	MAE	0.286	0.285	0.2868	0.2854	0.2716	0.2869	0.2683	0.2873	**0.2657**
	Improve	7.10%	6.77%	7.36%	6.90%	2.17%	7.39%	0.97%		

TABLE IX
PERFORMANCE COMPARISON OF LSOCMFCMS WITH ONE OF THE BEHAVIORS MISSING

Training data	Metrics	Missing behaviors							all behaviors
		favor	upload	playlist	comment	rate	share+	comment+	
90%	RMSE	0.2798	0.2744	0.2711	0.2723	0.2717	0.2743	0.2741	0.2710
	MAE	0.2299	0.2261	0.2209	0.2279	0.2207	0.2247	0.2270	0.2205
70%	RMSE	0.2918	0.2846	0.2832	0.2844	0.2834	0.2865	0.2854	0.2828
	MAE	0.2344	0.2352	0.2275	0.2344	0.2279	0.2347	0.2324	0.2272
50%	RMSE	0.3239	0.3178	0.3149	0.3187	0.3157	0.3184	0.3194	0.3127
	MAE	0.2503	0.2407	0.2385	0.2408	0.2384	0.2400	0.2412	0.2373
30%	RMSE	0.3418	0.3382	0.3354	0.3391	0.3368	0.3431	0.3385	0.3340
	MAE	0.2584	0.2539	0.2541	0.2533	0.2543	0.2604	0.2550	0.2535
10%	RMSE	0.3752	0.3639	0.3619	0.3630	0.3625	0.3733	0.3710	0.3613
	MAE	0.2796	0.2664	0.2663	0.2662	0.2659	0.2787	0.2745	0.2657

that the behavior "favorite" has the highest significance, while the behavior "add-to-playlist" has the lowest significance. Although the "share+" behavior has the largest correlations (see Table IV) with other behaviors, it has less influence than the "favorite" behavior. This can be explained by the fact that the "favorite" behavior has much larger number of videos than the "share+" behavior (see Table I) with adequately high correlation with other behaviors (see Table IV). The experimental results are basically consistent with our quantification results in the previous section.

2) Influence of Parameters: In order to illustrate the influence of the parameters on the video recommendation performance in the proposed models, we conduct experiments on different parameter settings. We fix the parameter values as stated in the previous section and examine the performance change of LSocMFCMS by tuning the parameters one by one. Firstly, we fix the other parameters and conduct experiments on different settings of parameter λ_s. Fig. 4 demonstrates the influence of λ_s on the performance of our proposed approach LSocMFCMS. We can see that the performance remains steady when λ_s changes from 0.00001 to 0.1 and deteriorates sharply after 0.1. This indicates the performance is insensitive to the change of λ_s when $\lambda_s \leq 0.1$. This indicates that the social relation helps personalized video recommendation lot within a certain range.

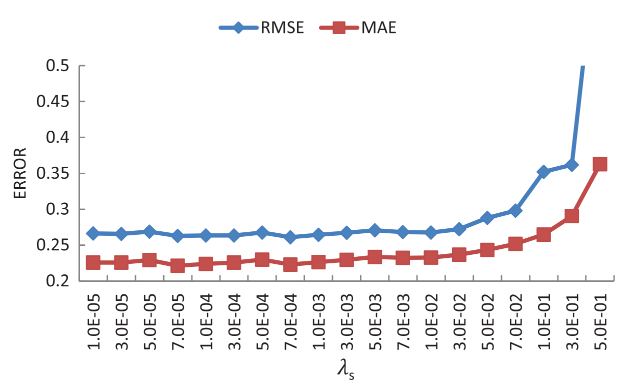


Fig. 4. Influence of social interactions.

We also investigate the influence of iteration steps on the performance of LSocMFCMS. We conduct experiments with the different settings of iteration steps and the experimental results

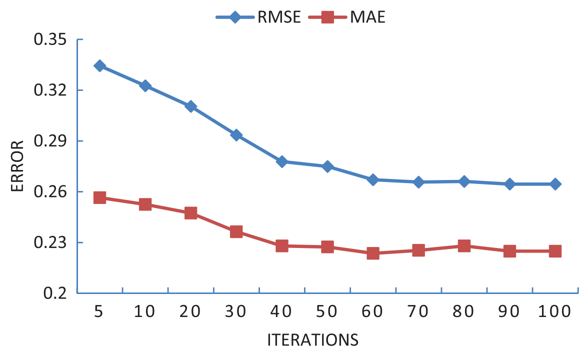


Fig. 5. Influence of iteration steps.

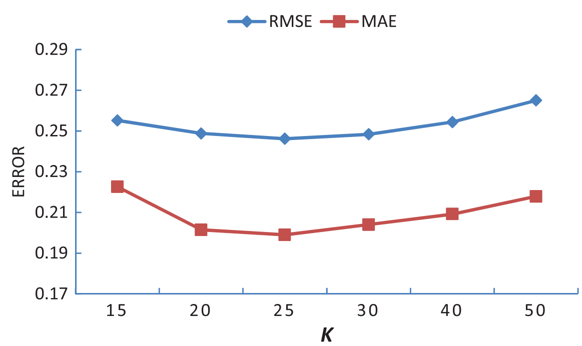


Fig. 6. Influence of the number of video topics.

are shown in Fig. 5. It is observed that the performance improves with the increasing of the number of the iteration steps. However, when the iteration step gets close to 100, the performance remains relatively steady. This indicates that the proposed model can converge within acceptable iteration steps.

Moreover, in order to examine the influence of the number of video latent topics (clusters) on the performance of LSocMFCMS, we conduct experiments with settings of different video topic numbers. To guarantee the convergence of the training process of LSocMFCMS with big cluster number, we set λ to a relatively small value of 0.005. We can see from Fig. 6 that the performance keeps improving before the number of clusters increases up to 25. This indicates that it is reasonable to set the number of video clusters according to the perplexity metric.

In general, we can observe that the performance is relatively steady when these parameter values change within certain ranges, i.e., the performance is not sensitive to parameter changes, which indicates that our proposed method is practical and will not be immersed in the parameter curse.

D. Discussions

In this subsection, we introduce the limitations of the proposed approach and discuss on the possible solutions.

The first limitation lies in the homogeneous behavior quantification that the derived behavior significance is unique to all the users. In fact, users may have different perceptions for the same type of behavior in reflecting their preferences. For example, the behavior "favorite" clearly reflects the preference of users who have very sparse behaviors, but captures less preference information for those who favor videos very frequently. However, it is quite challenging to discover the user-specific significance weight for each type of behavior. The huge number of parameters to be learned exacerbates the data sparsity problem. Moreover, it is very difficult to evaluate the performance due to lack of ground-truth user data. We leave the discovery of user-specific behavior quantification as one of our future work.

TABLE X
PERFORMANCE COMPARISON WITH DIFFERENT INITIALIZATION STRATEGIES

Training data	Metrics	Random		LDA	
90%	RMSE	0.2599	(87)	0.2527	(70)
	MAE	0.2263		0.1989	
70%	RMSE	0.2688	(101)	0.2637	(77)
	MAE	0.2343		0.2079	
50%	RMSE	0.2762	(122)	0.2734	(86)
	MAE	0.2383		0.2140	
30%	RMSE	0.2894	(127)	0.2849	(81)
	MAE	0.2474		0.2244	
10%	RMSE	0.3159	(73)	0.3120	(53)
	MAE	0.2680		0.2485	

TABLE XI
TIME COMPLEXITIES OF DIFFERENT MODELS

Model	Complexity in training	Complexity in testing
BasMF	$O(R \times K)$	$O(K \times N)$
SocMF	$O(R \times K \times M)$	$O(K \times N)$
LSocMFCM	$O(R \times K \times M \times N)$	$O(K \times N)$

The second limitation is that both behavior quantification and local regularization solutions are towards the interactions with the same type/modality of items, e.g., video-related cross-OSN behaviors. The discrepant significance of these behaviors are explored and fused. In real world, such application scenarios are very common, like the interactions with photos on Flickr and Instagram, etc. However, in case of interactions with different types of items, e.g., the Twitter tweeting and YouTube watching behaviors, the approaches proposed in this work are not readily applied. Our previous work in ACM Multimedia 2014 [8] and ACM ICMR 2015 [2] are devoted to mining the association between users' cross-OSN interactions with different types of items. Therefore, another line of future work is to integrate both the homogeneous and heterogeneous cross-OSN behaviors in a unified framework.

Another major limitation of the proposed cross-OSN user modeling solution is the high computational complexity of local social regularization. The training process of local social regularization is very time-consuming. In order to accelerate the iteration process, we initialize our model by the outputs of LDA model. To examine the effectiveness of this speedup strategy, we compare the performance of LSocMFCMS initialized by random and LDA, respectively. Table X demonstrates the experimental results where the numerics in brackets indicate the iteration steps before convergence. We can observe that LSocMFCMS initialized by LDA converges more quickly with generally superior performance than that initialized randomly. This demonstrates that it is helpful to provide a structured start-point for the iteration process of our model and validates the effectiveness of the convergence acceleration strategy.

With R denoting the number of training samples, K denoting the dimension of user (video) latent factor vector, $M(N)$ denoting the number of users (videos), the time complexities for different models are shown in Table XI. Regarding the time complexity in training, traditional MF model without social regularization is $O(R \times K)$, and the traditional MF model with global social regularization is $O(R \times K \times M)$. In our proposed LSocMFCM model, the topic space is dynamically changing during the iteration process, and the topic-level user similarity needs to be updated at each iteration. As a result, the time complexity of LSocMFCM in training is as high as $O(R \times K \times M \times N)$, which makes the online training for large-scale dataset inapplicable. As shown in Table XI, the time complexity of LSocMFCM in testing is the same as the traditional method. Since the goal of model training is to obtain user and video representations which are then used for personalized service, we can train our model off-line and then only need to update the models for the users who contributed new data.

VI. Conclusion

In this paper, we have proposed a novel framework for cross-OSN user modeling to 1) aggregate homogeneous user behaviors, and 2) integrate user social interactions at local topic level. We observed that there are adequate overlaps between different video-related behaviors on YouTube and Google+. A preliminary learning-based behavior quantification strategy is introduced to investigate the significance that each type of behavior contributes to user preferences. Furthermore, we have proposed to integrate the social interactions observed on Google+ into user model on the fused behavior matrix as local social regularization. Promising experimental results have demonstrated the effectiveness of the derived behavior significance and the local social regularization in the context of personalized video recommendation problem. In the future, we will be working towards user-specific behavior quantification, integration of both homogeneous and heterogeneous cross-OSN behaviors, and more efficient local social regularization solution.

References

[1] Y. Gao *et al.*, "Visual-textual joint relevance learning for tag-based social image search," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 363–376, Jan. 2013.

[2] M. Yan, J. Sang, and C. Xu, "Unified youtube video recommendation via cross-network collaboration," in *Proc. 5th ACM Int. Conf. Multimedia Retrieval*, 2015, pp. 19–26.

[3] Y. Koren, "Factorization meets the neighborhood: A multifaceted collaborative filtering model," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2008, pp. 426–434.

[4] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Proc. 4th Int. Conf. Web Search Web Data Mining*, Feb. 2011, pp. 287–296.

[5] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, *Recommender Systems Handbook*, 1st ed. New York, NY, USA: Springer, 2010.

[6] Y. Gao, R. Ji, W. Liu, Q. Dai, and G. Hua, "Weakly supervised visual dictionary learning by harnessing image attributes," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5400–5411, Dec. 2014.

[7] J. Sang, "User-centric cross-OSN multimedia computing," in *Proc. ACM Int. Conf. Multimedia*, Oct. 2015, pp. 1333–1334.

[8] M. Yan, J. Sang, and C. Xu, "Mining cross-network association for YouTube video promotion," in *Proc. ACM Int. Conf. Multimedia*, Nov. 2014, pp. 557–566.

[9] M. Jamali and M. Ester, "Trustwalker: A random walk model for combining trust-based and item-based recommendation," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 397–406.

[10] M. Jiang *et al.*, "Social recommendation across multiple relational domains," in *Proc. 21st ACM Int. Conf. Inf. Knowl. Manage.*, 2012, pp. 1422–1431.

[11] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proc. 4th ACM Conf. Recommender Syst.*, 2010, pp. 135–142.

[12] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Proc. 4th ACM Int. Conf. Web Search Data Mining*, 2011, pp. 287–296.

[13] X. Zhao *et al.*, "Integrating rich information for video recommendation with multi-task rank aggregation," in *Proc. 19th ACM Int. Conf. Multimedia*, 2011, pp. 1521–1524.

[14] M. Szomszor, H. Alani, I. Cantador, K. O'Hara, and N. Shadbolt, "Semantic modelling of user interests based on cross-folksonomy analysis," in *Proc. Semantic Web—ISWC*, 2008, pp. 632–648.

[15] F. Abel, S. Araújo, Q. Gao, and G. J. Houben, "Analyzing cross-system user modeling on the social web," in *Proc. 11th Int. Conf. Web Eng.*, Jun. 2011, pp. 28–43.

[16] F. Abel, E. Herder, G.-J. Houben, N. Henze, and D. Krause, "Cross-system user modeling and personalization on the social web," *User Modeling User-Adapted Interaction*, vol. 23, no. 2–3, pp. 169–209, 2013.

[17] N. J. Yuan *et al.*, "We know how you live: Exploring the spectrum of urban lifestyles," in *Proc. 1st ACM Conf. Online Social Netw.*, 2013, pp. 3–14.

[18] M. Yan, J. Sang, T. Mei, and C. Xu, "Friend transfer: Cold-start friend recommendation with cross-platform transfer learning of social knowledge," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2013, pp. 1–6.

[19] Z. Deng, M. Yan, J. Sang, and C. Xu, "Twitter is faster: Personalized time-aware video recommendation from Twitter to YouTube," *TOMCCAP*, vol. 11, no. 2, p. 31, 2014.

[20] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[21] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "Grouplens: An open architecture for collaborative filtering of netnews," in *Proc. ACM Conf. Comput. Supported Cooperative Work*, 1994, pp. 175–186.

[22] G. R. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *J. Mach. Learn. Res.*, vol. 5, pp. 27–72, 2004.

[23] N. Cristianini, J. Kandola, A. Elisseeff, and J. Shawe-Taylor, "On kernel target alignment," in *Proc. Adv. Neural Inf. Process. Syst. 14*, 2002, pp. 367–373.

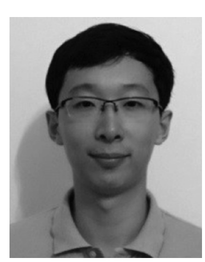
[24] C. Cortes, M. Mohri, and A. Rostamizadeh, "Two-stage learning kernel algorithms," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 239–246.

[25] Z. Deng, J. Sang, and C. Xu, "Personalized video recommendation based on cross-platform user modeling," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2013, pp. 1–6.

[26] T. Hofmann, "Latent semantic models for collaborative filtering," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 89–115, 2004.

[27] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 1257–1264.

[28] J. Wu and T. Li, "A modified fuzzy c-means algorithm for collaborative filtering," in *Proc. 2nd KDD Workshop Large-Scale Recommender Syst. Netflix Prize Competition*, 2008, pp. 1–4.



Jitao Sang received the B.E. degree from the SouthEast University, Nanjing, China, in 2007, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing, China, in 2012.

He is currently an Assistant Professor with the National Laboratory of Pattern Recognition, CASIA. He has authored one book and coauthored more than 40 peer-referenced papers in multimedia-related journals and conferences. He has filed three patents. His research interests include social multimedia computing, user modeling, and web data mining.

Prof. Sang is an Associate Editor for *Neurocomputing*, and is Guest Editor for the *Multimedia System Journal* and *Multimedia Tools and Applications*. He was a Tutorial Speaker at MM 2014 and MM 2015, MMM 2015, ICME 2015, and ICMR 2015. He was Program Co-Chair at PCM 2015 and ICIMCS 2015, Publicity Chair at MMM 2015, Publication Chair at ICIMCS 2013 and 2014, Special Session Organizer at ICME 2015, MMM 2013, and ICIMCS 2013, and Program Committee Member at many conferences (MM 2013, MM 2014, CIKM 2014, etc.). He was the recipient of Best Paper Finalist Award at MM 2012 and MM 2013, the Best Student Paper Award at MMM 2013, and the Best Student Paper Award at ICMR 2015.

Zhengyu Deng received the B.E. degree from the Central South University, Changsha, China, in 2009, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2015.

He is a Data Engineer with Huawei Technologies, Nanjing, China. His research interests include social network analysis, user modeling, and personalized services.

Donguan Lu received the B.S. degree from Beijing Normal University, Beijing, China, in 2007, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2012.

She is a Lecturer with the School of Information Technology and Management, University of International Business and Economics, Beijing, China. Her research interests include social media analysis, information retrieval, and data mining.

Changsheng Xu (M'97–SM'99–F'14) is a Professor with the National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, and Executive Director of China–Singapore Institute of Digital Media, Singapore. He has authored or coauthored over 200 refereed research papers. He holds 30 granted or pending patents. His research interests include multimedia content analysis/indexing/retrieval, pattern recognition, and computer vision.

Dr. Xu is a Fellow of the IAPR and an ACM Distinguished Scientist. He is an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA, the *ACM Transactions on Multimedia Computing, Communications and Applications*, and the *ACM/Springer Multimedia Systems Journal*. He served as Program Chair of ACM Multimedia 2009. He has served as Associate Editor, Guest Editor, General Chair, Program Chair, Area/Track Chair, Special Session Organizer, Session Chair, and TPC Member for over 20 IEEE and ACM prestigious multimedia journals, conferences, and workshops. He was the recipient of the Best Associate Editor Award of the *ACM Transactions on Multimedia Computing, Communications and Applications* in 2012 and the Best Editorial Member Award of the *ACM/Springer Multimedia Systems Journal* in 2008.