On Video Recommendation over Social Network

Xiaojian Zhao[1], Jin Yuan[2], Richang Hong[3],
Meng Wang[2,3], Zhoujun Li[1], and Tat-Seng Chua[2]

[1] State Key Laboratory of Software Development Environment,
Beihang University, Beijing 100191, China
[2] School of Computing, National University of Singapore, 117590, Singapore
[3] School of Computer and Information,
Hefei University of Technology, Hefei Anhui 230009, China
{zhaoxj01,hongrc.hfut,eric.mengwang}@gmail.com
{yuanjin,chuats}@comp.nus.edu.sg, lizj@buaa.edu.cn

Abstract. Video recommendation is a hot research topic to help people access interesting videos. The existing video recommendation approaches include CBF, CF and HF. However, these approaches treat the relationships between all users as equal and neglect an important fact that the acquaintances or friends may be a more reliable source than strangers to recommend interesting videos. Thus, in this paper we propose a novel approach to improve the accuracy of video recommendation. For a given user, our approach calculates a recommendation score for each video candidate that composes of two parts: the interest degree of this video by the user's friends, and the relationship strengths between the user and his friends. The final recommended videos are ranked according to the accumulated recommendation scores from different recommenders. We conducted experiments with 45 participants and the results demonstrated the feasibility and effectiveness of our approach.

Keywords: Video Recommendation, Relationship Strength, Activity Domain.

1 Introduction

Driven by the proliferation of digital capture and the advent of near-ubiquitous broadband Internet access, videos on the internet are growing at an explosive rate [3]. For example, it is estimated that the most popular video sharing website YouTube[1] stores over 150,000,000 videos in its repository. Therefore, today's online users always face a daunting volume of video clips when they search for interesting videos from the repositories. As a result, there is an increasing demand of video recommendation service which is able to help users to find "*interesting*" or "*highly related*" videos [10].

Currently, there are three prevalent approaches widely used in video recommendation, namely, *content-based filtering* (CBF), *collaborative filtering* (CF), and *hybrid filtering* (HF). The CBF approaches recommend videos based on the similarity between the unseen videos and the past videos viewed by the user [6], while the CF approaches

[1] http://www.youtube.com

K. Schoeffmann et al. (Eds.): MMM 2012, LNCS 7131, pp. 149–160, 2012.

predict video preference of the given user based on the ratings of the other users whose tastes are similar to him/her. Combining both of the above two approaches, HF approaches could achieve a better performance [1].

However, the above approaches ignore an important fact that a user's friends can be a more reliable source to recommend interesting videos rather than strangers. For example, a user is more likely to be interested in the videos recommended by his friends than that from strangers since he and his friends may have more common interests and know each other well. Moreover, for different friends, the user may share interests on different domains. This will affect video recommendation results. For example, a user usually discusses with friend A about the "*sport*" topic, then the sport videos recommended by friend A may be of more interest to the user. While he may share "*diet*" topic with friend B, and thus the diet videos from friend B may be good candidates to be recommended. Not only the relationship strength between users, the video interest is another important factor to be considered in video recommendation task. For example, a user has a strong relationship with a friend on the domain "*sport*", but his friend may be more interested in "*football*" than "*basketball*", thus the videos about "*football*" viewed by his friend are more important than those of "*basketball*". Therefore, the interest degree of video is another important factor in video recommendation task.

Based on these motivations, in this paper, we propose a novel video recommendation framework by considering both user relationship strength and the interest degree of video. For a given user, we calculate a recommendation score for each video candidate. The recommendation score is composed of two parts: the interest degree of this video by the user's friends, and the relationship strengths between the user and his friends. We measure the interest degree of each video based on its textual and visual similarity with the other viewed videos. The relationship strengths between users are inferred through online social network. For each pair of users, we employ a graph model to estimate the relationship strength by considering the users' profile information, the interaction activities as well as the activity domains. The final recommended videos are ranked according to the recommendation scores.

We summarize the main contributions of this paper as follows:

1. To the best of our knowledge, this is the first work that integrates the relationship strength information derived from online social network into a personalized video recommendation framework. We not only utilize relationship strength between users, but also consider them in different activity domains.
2. We also propose an approach to identify the interest degree between a candidate video and the recommender. The interest degree is calculated using the textual and visual similarity between the video and the other videos viewed in the past.

The rest of this paper is organized as follows. In section 2, we review the related work. Section 3 details of the proposed methodology of interest degree estimation and the relationship strength estimation. The initial experiment results and evaluations are provided in Section 4. Finally, we conclude the paper and discuss the directions of future works in Section 5.

2 Related Work

There are three prevalent approaches widely used in video recommender system, namely *content-based filtering* (CBF), *collaborative filtering* (CF), and *hybrid filtering* (HF) that combines the above two approaches. For CBF approaches, videos can be recommended based on the contents of previously viewed videos. For example, Mei et al. [11] presented an online video recommendation system, VideoReach, using multi-modal relevance between videos and users' click-through data. They considered three modalities, textual, visual and aural, and combined the relevance scores from them by using the attention fusion function. The CF approaches compare a user's ratings of videos with those of hundreds of others, find people who share similar preferences, and then recommend videos that are interesting for those people with similar preferences [12]. For example, Baluja et al. [2] built a user-video graph which represents the co-view information among different users and its recommendation was performed by a graph propagation in which the label of each node was obtained from its neighbors. However, the CBF approach neglects the fact that different users may share similar interestingness and the performance of CF approach is decreased when there is a shortage of user's ratings such as user can only view a very small portion of the videos from a large-scale online video database. Thus, HF approaches, which combine both of the above two approaches in a single framework, are proposed. For example, Burke [4] employed mixture models which build the recommendation based on a linear combination of voting, the content-based prediction and the collaborative prediction.

The growth and popularity of online social networks, such as Facebook and Google+[2], have led to a surge in research focusing on estimating the relationship strength between different users in online social network. Gilbert et al. [7] presented a predictive model that maps the social media data to the strength of ties between friends. However, these works only consider the binary prediction task of distinguishing the strong ties from the weak ties. Xiang et al. [16] developed an unsupervised model to estimate the relationship strength between different users from the interaction activity (e.g., communication, tagging) and the user similarity. However, it mixes all the interaction activities from various activity domains together and did not consider the fact that the relationship strengths between the same user pair may be different in various activity domains.

In this paper, we utilize the relationship strength between different users in different domains to help recommend videos to the user. Meanwhile, we also consider the interest degree of video candidates from each recommender's viewing history.

3 Approach

Give a set of users $\mathcal{U} = \{u_1, u_2, \ldots, u_K\}$ associated with a dataset of viewed videos $\mathcal{V} = \{v_1, v_2, \ldots, v_H\}$, the video recommendation service aims to provide a video list for each user u_k. Here, we calculate a recommendation score for each video with respect to the user u_k ($1 \leq k \leq K$), and then return a rank list of videos to this user according to the recommendation scores. As shown in Fig. 1, for a given video v_h ($1 \leq h \leq H$), its

[2] http://plus.google.com/

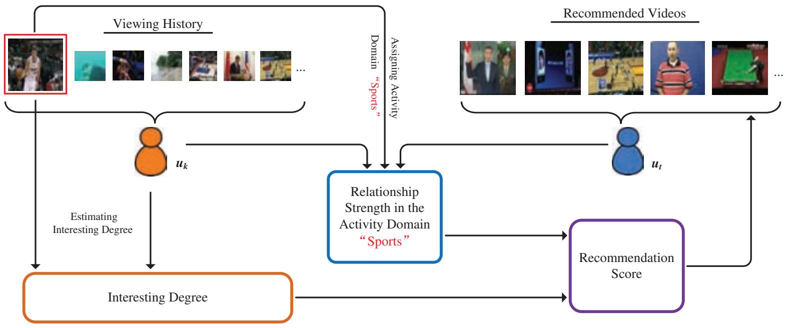


Fig. 1. The schematic illustration of the proposed video recommendation strategy that explores the user's viewing history and the relationship strengths in various activity domains

recommendation score to the user u_k is determined by two factors: the interest degree of the video v_h by the user u_t ($t \neq k$) who has viewed it before, and the relationship strength between the user u_k and u_t in the specific activity domain in which video v_h belongs to. Next, we will describe these two factors in detail.

3.1 Interest Degree Estimation

Give a set of videos $\mathcal{V} = \{v_1, v_2, \ldots, v_H\}$, we estimate the interest degree $I(v_h, u_k)$ of the video v_h by the user u_k as follow:

$$I(v_h, u_k) = \mu(v_h, u_k)U(v_h, u_k) \tag{1}$$

where $\mu(v_h, u_k)$ is an indicator whether u_k has viewed v_h before; $U(v_h, u_k)$ reflects the importance of the video v_h among all the viewed videos by u_k.

Generally, for a given user u_k, the importance of the video v_h $U(v_h, u_k)$ could be estimated according to the textual and visual information of all the viewed videos by u_k. Take the textual information as an example, if some words of the video v_h, such as "*football*" etc, frequently appear in the other viewed videos by u_k, then this video v_h is very important for user u_k since user u_k likes this topic "*football*". Based on this idea, in our approach, we adopt a linear function to calculate the importance value $I(v_h, u_k)$ based on textual and visual sources as:

$$U(v_h, u_k) = \alpha U_t(v_h, u_k) + (1 - \alpha)U_v(v_h, u_k) \tag{2}$$

where $U_t(v_h, u_k)$, $U_v(v_h, u_k)$ denote the importance of v_h measured by the textual and visual information respectively, and α is the balance weight which we empirically set in experiments.

Textual Importance Estimation. For a user u_k, the textual importance of video v_h is calculated by averaging the textual similarities between v_h and the other viewed videos by u_k, which we express it as:

$$U_t(v_h, u_k) = \frac{1}{num(v_o|u_k)} \sum_{o=1}^{L} \mu(v_o, u_k) S_t(v_h, v_o) \tag{3}$$

where $num(v_o|u_k)$ is the number of videos viewed by the user u_k, and $S_t(v_h, v_o)$ is the similarity between the video v_h and v_o measured based on textual information [14]. Here, the textual information includes video's title, description, tag and category etc. We represent each video v_o as a set of words $\mathcal{W}_o$, and the the similarity between video v_h and v_o is calculated as:

$$S_t(v_h, v_o) = \frac{1}{|\mathcal{W}_o||\mathcal{W}_l|} \sum_{w_o \in \mathcal{W}_o, w_l \in \mathcal{W}_l} exp(-\frac{NGD(w_o, w_l)}{\sigma}) \tag{4}$$

where $NGD(w_o, w_l)$ is the normalized Google distance [5] between the word w_o and w_l, and σ is a scaling parameter.

Visual Importance Estimation. The visual importance $U_v(v_h, u_k)$ is calculated by averaging the visual similarities between v_h and the other viewed videos by u_k, which we express as:

$$U_v(v_h, u_k) = \frac{1}{num(v_o|u_k)} \sum_{o=1}^{L} \mu(v_o, u_k) S_v(v_h, v_o) \tag{5}$$

where $num(v_o|u_k)$ is the number of videos viewed by user u_k, and $S_v(v_h, v_o)$ is the similarity between video v_h and v_o measured based on visual information. We represent each video as a set of key-frames. For each key-frame, we extract 428-dimensional global visual features, including 255-dimensional block-wise color moment, 128-dimensional wavelet texture, and 75-dimensional edge direction histogram [8][17][18]. The visual similarity $S_v(v_h, v_o)$ is calculated by averaging the similarities between their key-frames:

$$S_v(v_h, v_o) = \frac{1}{|v_h||v_o|} \sum_{\mathbf{x}_i \in v_h, \mathbf{x}_j \in v_o} (1 - \cos(\mathbf{x}_i, \mathbf{x}_j)) \tag{6}$$

where $\mathbf{x}_i$, $\mathbf{x}_j$ are key-frames in v_h and v_o respectively, $|v_h|$, $|v_o|$ represent the key-frame numbers contained in the corresponding videos, and $\cos(\mathbf{x}_i, \mathbf{x}_j)$ is the cosine distance between these two key-frames [13].

3.2 Relationship Strength Estimation

As Fig. 2 shows, the relationship strength estimation is composed of three steps: data preprocessing, activity domain assignment and graph-based relationship strength estimation. We will introduce these three steps in the rest of this subsection.

Data Preprocessing. Given the set of the interaction activities (messages, news feeds and events) downloaded from the Facebook website, there are three main sequential steps in our data preprocessing: word correction by Aspell[3], stop word removed, and stemming by WordNet[4]. After that, we obtain the dataset composed of the interaction

[3] http://aspell.net
[4] http://wordnet.princeton.edu/

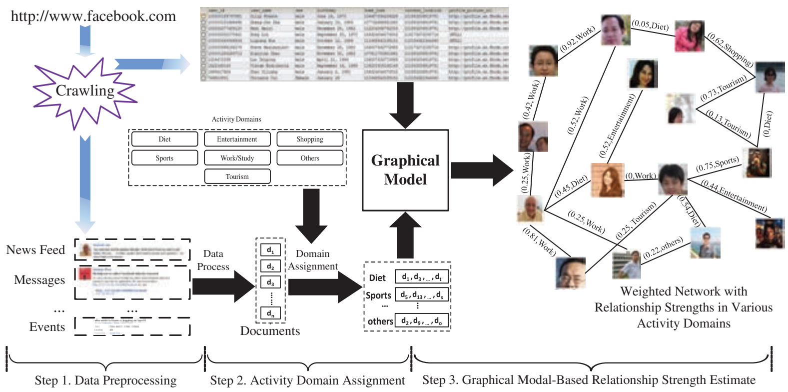


Fig. 2. A general framework to measure the relationship strength between different users in various activity domains in social network, where the (number,domain) pairs on the edges of the right network represent the value of relationship strength in that activity domain

activity documents $\mathcal{D} = \{d_1, d_2, \ldots, d_N\}$, where N is the number of documents. For each interaction activity document d_n, its related users refer to those users sending or receiving this document. We record this user-document relationship by a matrix $\mathbf{UD} = \{ud_{kn}\}_{k,n=1}^{k=K,n=N}$, where ud_{kn} indicates whether document d_n is related to user u_k.

Activity Domain Assignment. Given the activity domains $\mathcal{A} = \{A_1, A_2, \ldots, A_L\}$, we assign an activity domain A_l to each document d_n in $\mathcal{D}$ before estimating the relationship strength. Here, we define seven activity domains as "*diet*", "*entertainment*", "*shopping*", "*sports*", "*work*", "*tourism*", and "*others*". We represent each document d_n as a set of words. Let $Sem(d_n, A_l)$ be the relatedness degree between document d_n and activity domain A_l, which is calculated as:

$$Sem(d_n, A_l) = \sum_{w_n \in \mathcal{W}_n} tf_n * NGD(w_n, A_l) \tag{7}$$

where tf_n is the normalized frequency of the word w_n in $\mathcal{W}_n$, and $NGD(w_n, A_l)$ is the normalized Google distance [5] between w_n and the domain name A_l. For each document d_n, the domain A_l with the highest relatedness degree is assigned only if $Sem(d_n, A_l)$ is larger than a threshold, otherwise, this document belongs to "others".

Graphical Model Based Relationship Strength Estimation. To estimate the relationship strength between different users in various activity domains, we build a graphical model (described in Fig. 3) based on two observations:

1. For two users u_i and u_j, given a specific activity domain A_l, the relationship strength $T_l^{(i,j)}$ in this domain is determined by $S^{(i,j)}$, the profile similarity between these two users.
2. The relationship strength $T_l^{(i,j)}$ between u_i and u_j in activity domain A_l impacts their interaction activities on this domain (denoted as $D_l^{(i,j)}$).

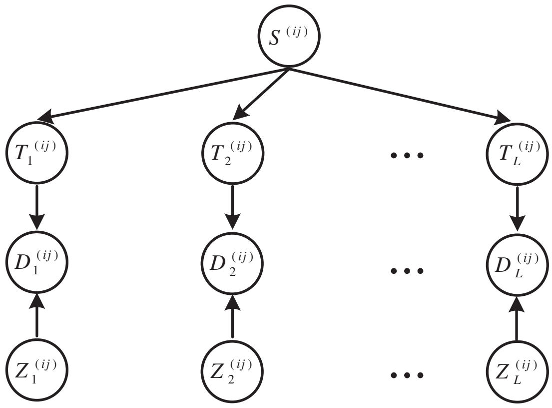


Fig. 3. Graphical model for estimating the relationship strength in various activity domains

Furthermore, to increase the accuracy of the graphical model, we introduce an auxiliary variable $Z_l^{(i,j)}$ for each $D_l^{(i,j)}$. The detailed descriptions of these variables in Fig. 3 are as follow:

– $S^{(i,j)} = (s_1^{(ij)}, s_2^{(ij)}, \ldots, s_P^{(ij)})$ is the similarity vector between the users u_i and u_j, where P is the number of attributes in the profile. For the p-th attribute f_p with discrete values, we set $s_p^{(ij)} = 1$ if u_i and u_j have the same values on f_p, and $s_p^{(ij)} = 0$ otherwise. On the other hand, if the values on f_p are continuous, $s_p^{(ij)}$ is determined according to:

$$s_p^{(ij)} = 1 - \frac{|f_p^i - f_p^j|}{\max\limits_{1 \le k_1, k_2 \le K} |f_p^{k_1} - f_p^{k_2}|} \tag{8}$$

where f_p^i represents the value of user u_i on the p-th attribute.
– $T_l^{(ij)}$ is the relationship strength between users u_i and u_j in activity domain A_l.
– $D_l^{(ij)}$ is the strength of interaction activities between users u_i and u_j in activity domain A_l. We measure it based on their related documents in this domain, which is calculated as:

$$D_l^{(ij)} = \sum_{n=1}^{N} Sem(d_n, A_l) * ud_{in} * ud_{jn} \tag{9}$$

– $Z_l^{(ij)}$ is an auxiliary variable. We set $Z_l^{(ij)}$ to 1 in our experiment.

As illustrated in Fig. 3, our graphical model represents the likely causal relationships among all the variables by modeling their conditional dependencies. Based on these dependencies, the joint distribution decomposes as follows:

$$\begin{aligned} &P(T_1^{(ij)}, \ldots, T_L^{(ij)}, D_1^{(ij)}, \ldots, D_L^{(ij)} | \mathbf{S}^{(ij)}, Z_1^{(ij)}, \ldots, Z_L^{(ij)}) \\ &= \prod_{l=1}^{L} P(T_l^{(ij)} | \mathbf{S}^{(ij)}) P(D_l^{(ij)} | T_l^{(ij)}, Z_l^{(ij)}) \end{aligned} \tag{10}$$

In this work, we adopt the widely-used Gaussian distribution to model the conditional probabilities $P(T_l^{(ij)}|\mathbf{S}^{(ij)})$ and $P(D_l^{(ij)}|T_l^{(ij)}, Z_l^{(ij)})$, which are expressed as:

$$\begin{aligned} P(T_l^{(ij)}|\mathbf{S}^{(ij)}) &= \mathcal{N}(\mathbf{w}_l^T\mathbf{S}^{(ij)}, v) \\ P(D_l^{(ij)}|T_l^{(ij)}, Z_l^{(ij)}) &= \mathcal{N}(\alpha_l T_l^{(ij)} + \beta_l Z_l^{(ij)}, v) \end{aligned} \tag{11}$$

where $\mathbf{w}_l$ is a P-dimensional weight vector to be estimated, α_l, β_l are two coefficients, and v is the variance in Gaussian model, which is configured to be 0.5 in our experiments. To avoid over-fitting, we put L_2 regularizes on parameters $\mathbf{w}_l$ and α_l, β_l, which can be regarded as Gaussian priors:

$$\begin{aligned} P(\mathbf{w}_l) &\propto e^{-\frac{\lambda_1}{2}\mathbf{w}_l^T\mathbf{w}_l} \\ P(\alpha_l, \beta_l) &\propto e^{-\frac{\lambda_2}{2}(\alpha_l^2+\beta_l^2)} \end{aligned} \tag{12}$$

Among all the variables, $\mathbf{S}^{(ij)}$, $D_l^{(ij)}$, $Z_l^{(ij)}$ are all visible, and $\mathbf{w}_l$, α_l, β_l are to-be-learned parameters. Given the samples of the user pairs $\mathcal{P} = \mathcal{U} \times \mathcal{U}$, the joint probability is expressed according to Eq. (10)-Eq. (12):

$$\begin{aligned} &\prod_{l=1}^{L} P(\mathcal{P}|\mathbf{w}_l, \alpha_l, \beta_l)P(\mathbf{w}_l)P(\alpha_l, \beta_l) \\ &= \prod_{l=1}^{L} \prod_{(i,j)\in\mathcal{P}} P(D_l^{(ij)}|Z_l^{(ij)}, T_l^{(ij)}, \alpha_l, \beta_l)P(T_l^{(ij)}|\mathbf{S}^{(ij)}, \mathbf{w}_l)P(\mathbf{w}_l)P(\alpha_l, \beta_l) \\ &\propto \prod_{l=1}^{L} \prod_{(i,j)\in\mathcal{P}} e^{-\frac{1}{2v}(\mathbf{w}_l^T\mathbf{S}^{(ij)}-T_l^{(ij)})^2} e^{-\frac{1}{2v}(\alpha_l T_l^{(ij)}+\beta_l Z_l^{(ij)}-D_l^{(ij)})^2} e^{-\frac{\lambda_1}{2}\mathbf{w}_l^T\mathbf{w}_l} e^{-\frac{\lambda_2}{2}(\alpha_l^2+\beta_l^2)} \end{aligned} \tag{13}$$

Since the joint probabilities of L activity domains in Eq. (13) are independent, we can divide Eq. (13) into L independent joint probabilities, and infer the solution for each activity domain separately. In our implementation, we use a gradient-based method to optimize it over the parameters $\mathbf{w}_l^T$, α_l, β_l, and variable $T_l^{(ij)}$. Due to the limited space, the detailed implementation is not presented here.

3.3 Video Recommendation

Given a user u_k, in this step, we calculate the recommendation score $R(v_h, u_k)$ for each video v_h. Aforementioned, the recommendation score $R(u_k, v_h)$ is determined by two factors: the interest degree of video v_h by user u_t ($t \neq k$) (denoted as $I(v_h, u_t)$, see Section 3.1), and the relationship strength between u_t and u_k in the special domain A_l that v_h belongs to (denoted as $T_l^{(tk)}$, see Section 3.2). We multiple these factors as:

$$R(v_h, u_k) = \sum_{t=1, t\neq k}^{K} I(v_h, u_t)T_l^{(tk)}\mu(v_h, A_l) \tag{14}$$

where $\mu(v_h, A_l)$ is an indicator to represent whether v_h belongs to domain A_l, $\mu(v_h, A_l) = 1$ indicates that video v_h belongs to domain A_l, and $\mu(v_h, A_l) = 0$ otherwise. In our approach, we represent video v_h as a word set, where the words inside are collected from the textual description associated with v_h. Based on the approach in Section 3.2, we can assign a domain to video v_h. According to the recommendation scores in Eq. (14), we return a rank list of videos to each user.

4 Experiments

4.1 Experimental Settings

The dataset is downloaded from the Facebook website, which is a popular online social network site with over 600 million members worldwide. To download data from Facebook, we first selected 9 active users from three countries (Singapore, China and America) as the seed nodes. After obtaining their consents, we collected all the friends of these 9 users, which results in a total of 632 people. Since it is hard to collect the viewing history of all people, we only sampled 45 persons who have at least three common acquaintances. We downloaded each user's profile information, including location, language, religion, interests and etc. Besides, we downloaded the interaction activities (messages, news feed, etc.) for each user between Sep. 2010 and Oct. 2010. This results in a total of 22,500 interaction activity documents. Meanwhile, the video viewing behaviors of these users on YouTube were tracked in a one-month period (from Dec. 2010 to Jan. 2011). Video links from each user's log were extracted. The video itself and the corresponding title, tag, category, description information were downloaded and stored in our database. It is shown that there are about 150 videos viewed per user on average.

For each user, we split the viewed videos into two parts, the first part is the videos viewed in the previous two weeks and the second part is the videos viewed in the next two weeks. The second part is used for testing. In other words, we regard videos in the second part as the relevant samples for recommendation. We assign the relevant score of 1 to the videos in the second part, and 0 to the other videos for a user. It is worth noting that this setting actually underestimates the performance, as the user may also be interested in the videos out of the second part. Though a more rigorous approach for ground truth establishment is to let users label all videos with interestingness, our approach is still reasonable for comparing different algorithms. For performance evaluation metric, we adopted the normalized discounted cumulative gain (NDCG) [9] [15].

4.2 Experimental Results

To comprehensively evaluate our approach, we consider two types of classical recommendation methods as baselines:

1. Content-based filtering method (CBF): the videos are recommended based on the similarity between the unseen video and the videos previously viewed by the user. The similarity between these two videos is calculated using Eq. (4) and Eq. (6).
2. Collaborative filtering method (CF): it predicts the preference to a video based on the ratings of similar users. The distance of vectors which represent the user's viewing history is adopted to measure the similarity between different users.

Fig. 4 illustrate the comparison of average NDCG@20 in various activity domains. We can see that our approach outperforms the other methods in almost every domain except for "*work*" domain. One possible reason is that the representative words in "*work*" domain are diverse. On the other hand, "*sports*" domain contains very highly repetitive words such as "*basketball*", "*swimming*" and "*jogging*", and thus the relationship

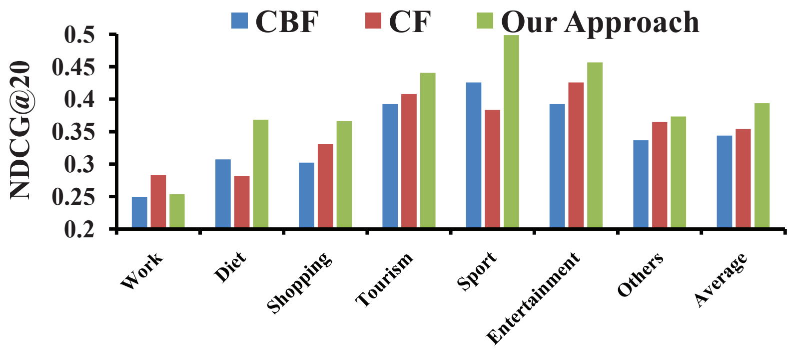


Fig. 4. The performance of the three video recommendation methods in different active domains

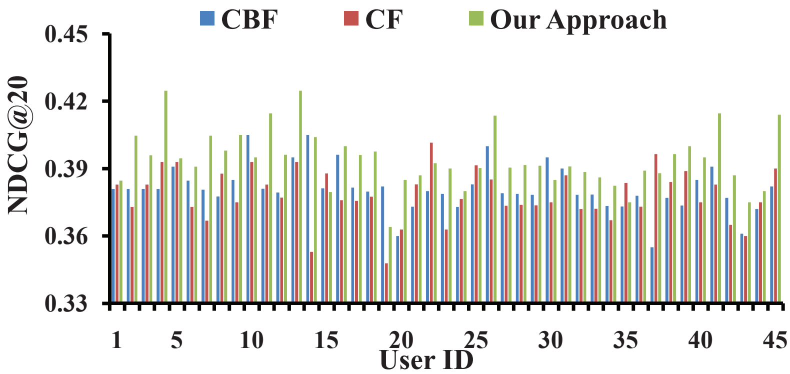
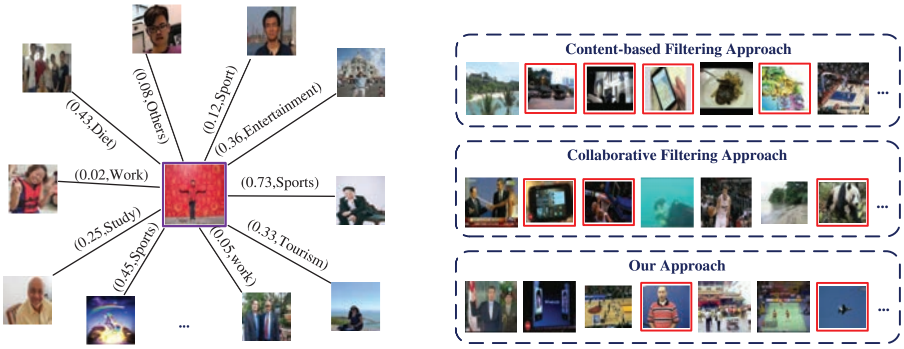


Fig. 5. The performance of the three recommendation methods for each user

strength in the "*sports*" domain can be estimated more accurately. In addition, people rarely use the social network to discuss work related topics.

Fig. 5 illustrates detailed NDCG@20 results for the 45 users. We can see that for most users our approach achieves better results than the other two methods. The reason is that the two main components in the ranking score function Eq. 14 integrates the positive aspects of both CBF and CF methods. One component is the relationship strength in the activity domain in which the video to be recommended belongs to. It is estimated based on user's profile information and the interaction activities between different users. The other component is the interest degree to the video given by the recommender. It is estimated based on the recommender's viewing history. So its performance is obviously better than the other two methods. Fig. 6 illustrates the top relationship strength in different activity domains for a user in a social network and the different ranking lists produced by three video recommendation strategies for this user. We can see that the user's main interest is "*sports*", which is estimated by our proposed relationship strength measurement. From Fig. 6(b), we can see that our proposed approach recommends more relevant videos given the user's interest.



(a) The relationship strengths in different activity domains for a user (the center one).

(b) The ranking lists of the three video recommendation methods for the user in subfigure (a).

Fig. 6. The relationship strength network for a user and the recommended videos to him by the three video recommendation strategies

5 Conclusion and Future Works

In this paper, we proposed a novel approach to improve the accuracy of video recommendation by utilizing the relationship strength information from social network. First, the interest degree of each viewed video by a user's friends was calculated. Second, the relationship strengths between different users were measured, taking into consideration not only the user's profile information, interaction activities, but also the activity domains. Third, the recommended videos viewed by the friends of a user were ranked based on their interest degree of each video and the user's relationship strengths with the friends in different domains. We conducted experiments with 45 participants and the results demonstrated the feasibility and effectiveness of our approach. In our future work, we will conduct experiments with more users and will also consider integrating more contextual factors in video recommendation, such as the time and location.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (61170189, 60973105), the Fund of the State Key Laboratory of Software Development Environment under Grant No. SKLSDE-2011ZX-03 and the Singapore National Research Foundation & Interactive Digital Media R&D Program Office, MDA under research grant (WBS:R-252-300-001-490).

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 734–749 (2005)
2. Baluja, S., Seth, R., Sivakumar, D., Jing, Y., Yagnik, J., Kumar, S., Ravichandran, D., Aly, M.: Video suggestion and discovery for youtube: taking random walks through the view graph. In: ACM WWW, pp. 895–904 (2008)

3. Boll, S.: Multitube–where web 2.0 and multimedia could meet. IEEE Multimedia 14(1), 9–13 (2007)
4. Burke, R.: Hybrid Web Recommender Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 377–408. Springer, Heidelberg (2007)
5. Cilibrasi, R., Vitanyi, P.: The google similarity distance. IEEE Transactions on Knowledge and Data Engineering, 370–383 (2007)
6. Gibas, M., Canahuate, G., Ferhatosmanoglu, H.: Online index recommendations for high-dimensional databases using query workloads. IEEE Transactions on Knowledge and Data Engineering, 246–260 (2008)
7. Gilbert, E., Karahalios, K.: Predicting tie strength with social media. In: ACM CHI, pp. 211–220 (2009)
8. Hong, R., Wang, M., Xu, M., Yan, S., Chua, T.: Dynamic captioning: video accessibility enhancement for hearing impairment. In: ACM MM, pp. 421–430 (2010)
9. Hu, X., Tang, L., Liu, H.: Enhancing accessibility of microblogging messages using semantic knowledge. In: ACM CIKM (2011)
10. Mei, T., Aizawa, K.: Video recommendation. In: Chapter of Internet Multimedia Search and Mining. Bentham Science Publisher (2011)
11. Mei, T., Yang, B., Hua, X., Yang, L., Yang, S., Li, S.: Videoreach: an online video recommendation system. In: ACM SIGIR, pp. 767–768 (2007)
12. Park, J., Lee, S., Kim, K., Chung, B., Lee, Y.: An online video recommendation framework using view based tag cloud aggregation. IEEE Multimedia (99), 1 (2010)
13. Wang, M., Hua, X., Tang, J., Hong, R.: Beyond distance measurement: constructing neighborhood similarity for video annotation. IEEE Transactions on Multimedia 11(3), 465–476 (2009)
14. Wang, M., Hua, X., Tang, J., Qi, G., Song, Y.: Unified video annotation via multi-graph learning. IEEE Transactions on Circuits and Systems for Video Technology 19(5) (2009)
15. Wang, M., Yang, K., Hua, X.-S., Zhang, H.-J.: Towards a relevant and diverse search of social images. IEEE Transactions on Multimedia, 12 (2010)
16. Xiang, R., Neville, J., Rogati, M.: Modeling relationship strength in online social networks. In: ACM WWW, pp. 981–990 (2010)
17. Yang, Y., Xu, D., Nie, F., Yan, S., Zhuang, Y.: Image clustering using local discriminant models and global integration. IEEE Transactions on Image Processing (2010)
18. Yang, Y., Zhuang, Y., Wu, F., Pan, Y.: Harmonizing hierarchical manifolds for multimedia document semantics understanding and cross-media retrieval. IEEE Transactions on Multimedia, 10 (2008)