Twitter is Faster: Personalized Time-aware Video Recommendation from Twitter to YouTube

ZHENGYU DENG[1,2], MING YAN[1,2], JITAO SANG[1,2] and CHANGSHENG XU[1,2],
[1]National Lab of Pattern Recognition, Institute of Automation, CAS, China
[2]China-Singapore Institute of Digital Media, Singapore

Traditional personalized video recommendation methods focus on utilizing user profile or user history behaviors to model user interests, which follows a static strategy and fails to capture the swift shift of the short-term interests of users. According to our cross-platform data analysis, the information emergence and propagation is faster in social textual stream-based platforms than that in multimedia sharing platforms at micro user level. Inspired by this, we propose a dynamic user modeling strategy to tackle personalized video recommendation issue in the multimedia sharing platform YouTube, by transferring knowledge from the social textual stream-based platform Twitter. In particular, the cross-platform video recommendation strategy is divided into two steps: (1) Real-time hot topic detection: the hot topics that users are currently following are extracted from users' *tweets*, which are utilized to obtain the related videos in YouTube. (2) Time-aware video recommendation: for the target user in YouTube, the obtained videos are ranked by considering user profile in YouTube, time factor and quality factor to generate the final recommendation list. In this way, the short-term (hot topics) and long-term (user profile) interests of users are jointly considered. Carefully designed experiments have demonstrated the advantages of the proposed method.

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

General Terms: Algorithms, Experimentation, Performance

Additional Key Words and Phrases: Short-term interest, Personalization, Video recommendation, Time-aware, Cross-platform

1. INTRODUCTION

The User Generated Content (UGC) is propagated online tremendously with the arising of Web 2.0, which leads to the arrival of "Big Data Age". Taking YouTube [1], the most popular online-video sharing website as an example, there are two billion videos on this website and more than 60 hours of new videos are uploaded every minute [2]. Faced with the critical information overload, the exploration and discovery of interesting resources for network users from the

[1] http://www.youtube.com/.
[2] http://en.wikipedia.org/wiki/youtube/.

This work is supported in part by National Basic Research Program of China (No. 2012CB316304), National Natural Science Foundation of China (No. 61225009, 61332016, 61303176), and Beijing Natural Science Foundation (No. 4131004). This work is also supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.
Author's address: Zhengyu Deng, Ming Yan, Jitao Sang, Changsheng Xu (the corresponding author), National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing 100190, China (e-mail: zydeng, ming.yan, jtsang, csxu@nlpr.ia.ac.cn).

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

tremendous data become extremely difficult. The personalized services, including personalized search, subscription, recommendation, etc., stand out for solution and play a vital role in tackling the issue of information overload [Sang and Xu 2012] [Gao et al. 2013].

Most of the traditional personalized video recommendation methods are devoted to static user modeling, which utilizes the user profile or the history behaviors to understand the long-term interest of the user. However, user interest often distributes dynamically, which differs from time to time. Especially, when surrounded by the tremendous fresh messages every day, user's short-term interest may change continuously with the current hot events [3]. For example, the fact that a user has read the news about "US presidential election" may lead to the consequent action of searching for videos on "US presidential election debate" to gain further details about this event. In this case, the user may not really have great interest in politics from the long-term perspective, but his/her short-term interest is largely influenced by the popularly acquired information around him/her. Therefore, the personalized video recommendation strategies which cannot capture the swift drift of the user interest will fail to push the timely videos to desired users.

Existing personalized video recommendation work referring to the short-term interests of users mainly focuses on the single multimedia sharing platform itself. The limitations of the short-term interest extraction in single multimedia sharing platform are as follows. 1) Firstly, the users' behaviors in single platform are often limited, resulting in that it is difficult to exactly capture the swift drift of user interest. 2) Besides, recommendation based on the users' short-term interests inferred from their current behaviors in the same platform may make the recommendation always lags behind users' actual behaviors, leading to duplicated recommendation. 3) Moreover, information emergence and propagation in multimedia sharing platforms is slower than that in social textual stream platforms and users' short-term interests extracted from multimedia sharing platforms are less time-aware. Therefore, we propose to enrich users' short-term interests by introducing social textual stream platforms and we are dedicated to investigating whether a user has consistent behaviors between different platforms and their temporal relations.

Notably, social textual stream-based platforms (such as Twitter, Weibo) are widely regarded as a source of real-time breaking news, and information emergence and propagation is faster in these platforms than multimedia sharing platforms (such as YouTube, Flicker). It was reported that the news about "Virginia earthquake" appeared in Twitter almost at the same time when the earthquake happened, and it propagated throughout America in the following five minutes even faster than the earthquake waves [4]. Existing work analyzed the temporal patterns of user behaviors between different platforms on global level. In this paper, we are interested in investigating whether there is a consistent conclusion on a micro user level, e.g., for a specific user, is there any activity pattern that he/she has come across a piece of news on Twitter before they search the related videos on YouTube? To answer this question and explore the temporal characteristics of different platforms on user level, we further investigate into the temporal patterns across different platforms based on each single user and find that information emergence and propagation is also faster in social textual stream-based platforms than multimedia sharing platforms on user level. From the perspective of information inquiry, it is highly possible that users have come across a piece of news in Twitter before they search the related videos in YouTube. In other words, if we know which topic a user is following currently in social textual stream platforms, we can recommend the relevant videos to him/her on YouTube to help get deeper insight into this topic. Enlightened by this, we designed a personalized time-aware video recommendation solution for the multimedia sharing platforms by exploiting users' activities in social textual stream platforms to capture users' short-term interests.

In this paper, we address the time-aware personalized video recommendation issue by cross-platform collaboration from Twitter to YouTube: we use YouTube as the video sharing platform to perform the recommendation task, and Twitter as the social textual stream platform to extract the real-time hot topics users followed. We first conduct a cross-platform data analysis to examine the evolution of topics between Twitter and YouTube and conclude that information propagation in Twitter is faster than that in YouTube on both global level and user level. Based on this observation, we

[3] Hot event is defined as a subject discussed and shared frequently in many documents and platforms. Examples are like "Olympic opening ceremony 2012", "US election day 2012", "Super bowl game", etc. In this paper, hot event is equivalent to hot topic.
[4] http://domaingang.com/short-news/tweetquake-twitter-moving-faster-than-an-earthquake/.

design a cross-platform video recommendation strategy which is generally divided into two steps: (1) Real-time hot topic detection: the hot topics that users are following currently are extracted from users' *tweets*, which are utilized to obtain the related videos on YouTube. (2) Time-aware video recommendation: for the target user on YouTube, the obtained videos are ranked by considering the user profile in YouTube, time factor and quality factor to generate the final video recommendation list. In this way, the short-term (hot topics) and long-term (user profiles) interests of users are jointly considered, which is illustrated in Fig. 1. The inputs include the *tweets* users shared/reshared in Twitter and the user profile in YouTube; whereas the output is the generated video recommendation list. To summarize, the main contributions of this paper are as follows:

(1) We propose to address the personalized time-aware video recommendation issue by exploiting cross-platform collaboration to integrate the short- and long-term interests of users.
(2) We perform exploratory data analysis on cross-platform user activity data, and validate that information emergence and propagation in Twitter is faster than that in YouTube on both global level and user level.
(3) We present a novel framework of cross-platform collaboration based on the temporal patterns of user behaviors between social textual stream platforms and multimedia sharing platforms.

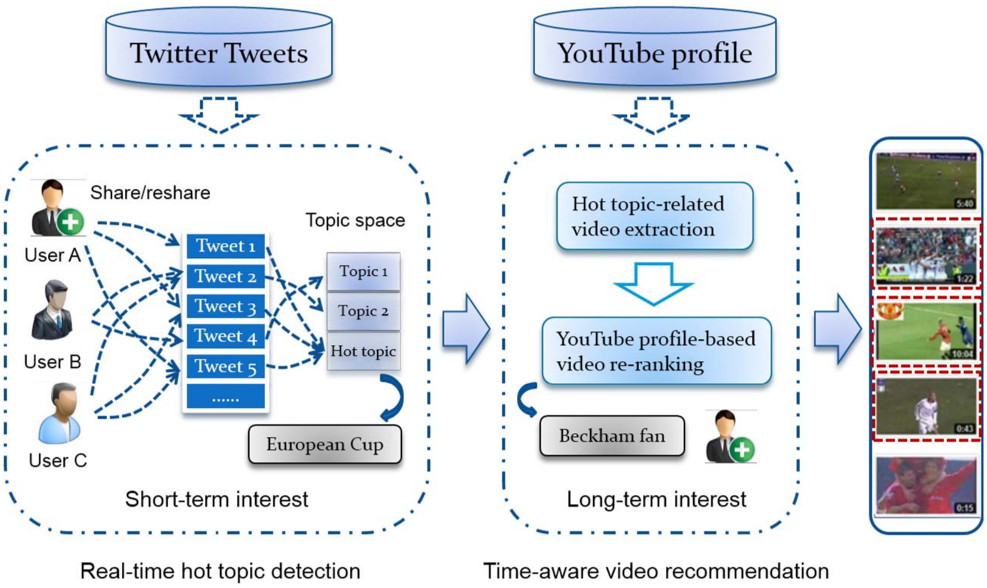

Fig. 1. The proposed framework.

The rest of this paper is organized as follows. We first review related work in Section 2. Then in Section 3, we give a description of data collection and present the data analysis results on the data collection. Inspired by the data observations, the cross-platform video recommendation solution is introduced in Section 4. To evaluate the performance of our proposed approach, experimental results and analysis are reported in Section 5. Finally, conclusions and future work are presented in Section 6.

2. RELATED WORK

This section provides a review of areas related to our work. We first report a survey on cross-platform data analysis. Then some existing user modeling mechanisms and personalized video recommendation research that are closely relevant to our work are introduced.

2.1 Cross-platform Data Analysis

Multi-platform analysis and application has recently drawn attentions on the academia communities due to the increasingly vast accessibility of various account fusion and management tools such as “Friendfeed” and “Aboutme”. In general, the work can be mainly divided into two parts, i.e., the static analysis and the dynamic analysis.

The work of static analysis mainly focuses on analyzing and comparing the long-term social structures [Ahn et al. 2007] [Lerman and Ghosh 2010] [Magnani and Rossi 2011] or social behaviors [Szomszor et al. 2008] [Abel et al. 2011] across different social platforms. For example, [Ahn et al. 2007] made a comparison on the structures of three online social networking services: Cyworld, MySpace and Orkut. In order to explore the different user behavior patterns in multiple social platforms, [Abel et al. 2011] analyzed and compared the user tag clouds from multiple folksonomies: Flickr, Twitter and Delicious and alleviated the cold-start problem by leveraging multiple platform collaboration.

While the other part puts emphasis on exploring the temporal dynamic patterns and evolutions among different social platforms [Yang and Leskovec 2011] [Osborne et al. 2012] and this is also what we are concerned with. [Yang and Leskovec 2011] investigated into the temporal patterns associated with online content from multiple online sources and analyzed how the popularity of content grows and fades over time. They found that weblogs trail main stream media by one or two hours for most of the considered events. [Osborne et al. 2012] aimed at identifying events more precisely by utilizing the content from auxiliary platform Wikipedia which indirectly resulted in the finding that Wikipedia lags behind Twitter by about two hours. However, existing dynamic analysis work only analyzed and compared the temporal patterns across different social platforms on global level, and did not turn into micro individual level. In contrast to the these work, our dynamic analysis work focuses on the comparison of the temporal characteristics across different social platforms on user level.

2.2 User Modeling and Personalized Video Recommendation

Based on the cross-platform temporal analysis on user level, we further propose a time-aware user modeling mechanism which aims at better inferring and capturing users’ short-term interests. Next, some existing work on user modeling will be introduced. Afterwards, related studies on personalized video recommendation are summarized.

[Fortuna et al. 2011] proposed a SVM-based segment approach to model users of large websites based on different data sources: access logs, page content and user registration data. Szomszor and Alali [2008] utilized user tag cloud to model user interests and identified significant benefits of cross-platform user modeling. Besides, [Abel et al. 2011a] analyzed user model in Twitter, where Twitter user profile was enriched with the similar news. Most of these user modeling methods focus on utilizing constant user profile, demography or context to model user interests, which follows a static strategy. Moreover, [Koren 2010] proposed a time-aware factor model where the feature vectors of user and item are changing along the whole time period. [Xiong et al. 2010] proposed a Bayesian probabilistic tensor factorization model, incorporating time as an additional feature factor. [Koenigstein et al. 2011] proposed to use session factors, inferred from time-stamps associated with items, to model temporal user behavior. Besides, [Xiang et al. 2010] proposed a session-based temporal graph model to capture the long- and short-term preference over time. [Bennett et al. 2012] investigated the interaction between short- and long-term behaviors, and how this information can be combined to learn effective models. [Yang et al. 2012] proposed a local implicit feedback model, where local and global information, represented by implicit feedback, are combined to capture users’ stable and local changeable preferences. All of these models are limited in single multimedia sharing platform and did not promptly captured users’ short-term interests. In our work we capture a full set of user activities in social textual stream platform, which ensures dynamic updates to catch the swift drift of users’ short-term interests. Furthermore, some researchers analyzed the characteristics of different social tagging platforms and proposed some multi-platform user modeling methods, but such work [Abel et al. 2011] [Abel et al. 2011b] was mainly devoted to coarse combination of multiple sources.

With the time-aware user modeling mechanism, we design a personalized video recommendation solution with cross-platform collaboration. Traditional video recommendation strategies focus on three typical approaches, namely, Collaborative Filtering (CF) [Baluja et al. 2008] [Davidson et al. 2010], Content-Based Recommendation (CBR)

[Wang et al. 2007] [Liu et al. 2009] [Mei et al. 2011], and Hybrid Recommendation (HR) [Jin et al. 2010] [Park et al. 2011] [Zhao X. et al. 2011]. Most of these work is based on static user modeling, which suffers from the swift drift of user interest. According to our data analysis, the change of user interest is less timely in multimedia sharing platform than that in social textual stream platform. Therefore, in this paper, we focus on extracting users' short-term interests from social textual stream platform for personalized video recommendation in multimedia sharing platform. In academic communities, social textual stream analysis has drawn lots of attentions. Some researchers studied the network structures and how information propagates through the Twitter network [Weng et al. 2010] [Kwak et al. 2010] [Lerman and Ghosh 2010]. Furthermore, [Gao et al. 2011] investigated the interplay of individual interests and the public trends in Twitter. [Abel et al. 2011c] studied the characteristics of Twitter profile and investigated how to utilize user behaviors in Twitter for personalization. However, their research only focuses on textual stream applications. Moreover, [Roy et al. 2012] assumed an intermediate topic space can be built across Twitter and YouTube and proposed to learn from social stream to facilitate multimedia applications. Yet none of this work has been done on user level. Inspired by the temporal patterns of user behaviors across platforms on user level, we propose a dynamic user modeling strategy for personalized video suggestion on multimedia sharing platform utilizing users' real-time interests inferred from social textual stream platform.

3. CROSS-PLATFORM DATA ANALYSIS

In this section, we examine the characteristics of user behaviors between Twitter and YouTube. Firstly, we describe how we collect our cross-platform dataset. Afterwards, in order to maintain the reliability of data analysis, we manually select some widely-known hot topics which are frequently talked about both in Twitter and YouTube. Thereafter we present a global temporal dynamic analysis on the popularity of certain topics regarding all the users' behavior data in our dataset. Finally we further investigate whether there are some temporal behavior patterns across Twitter and YouTube on user level and category level, respectively.

3.1 Data Collection

The fact that many network users create and maintain multiple accounts on different Web 2.0 platforms provides possibilities for the cross-platform collaboration. With the trend of information aggregation, many users are willing to manage their separate accounts using social media aggregation tools such as FriendFeed [5], About.me [6]. We have also found that users tend to provide their accounts of other platforms when registering into social network sites. For instance, we have observed from our Google+ dataset that a considerable proportion of users share their accounts like YouTube, Flickr and Twitter at their Google+ homepages [Deng et al. 2013] [Yan et al. 2013]. Inspired by this, to obtain a collection of users with both accounts in Twitter and YouTube, we started from Google+ platform where about 10.5% of the users' homepages contains the accessible URL links of their YouTube and Twitter accounts, and collected 126,971 users in total. Then we only kept the users who have both Twitter and YouTube accounts and removed those who have less than ten videos in their uploading list, resulting in the final cross-network dataset with 7,686 users. The users' registration information and behavior data from Jan. 2012 to Apr. 2013 were downloaded from both Twitter and YouTube, respectively. In Twitter, the behavior data contains all the users' posted *tweets/retweets* with timestamp; whereas in YouTube, it contains all the users' video-related behaviors such as commenting, rating, favoring and uploading a video. As a result, we got a dataset with more than 8 million *tweets* and 0.75 million video-related behaviors for our 7,686 users. The following experiments and analysis are based on this dataset.

3.2 Topic Selection

In order to find hot topics which are widely spread between Twitter and YouTube in our dataset, we combine the official statistics of the trending topics in 2012 with a simple sorting method by word frequency. In Twitter we use

[5] http://friendfeed.com/.
[6] http://about.me.com/.

tweet words and hashtags to identify a topic, while in YouTube the video tags are utilized to identify a topic. We started from the top trending searches of 2012 revealed by Google [7] and collected all the trending topics with different locations and different categories the official statistics mentioned. Then we aggregated all the tags of the YouTube videos and all the *tweet* words as well as the hashtags involved in our dataset, respectively. We further counted the word and tag frequencies upon all the behavior data in Twitter and YouTube, respectively and sorted the words and tags according to their frequencies. Finally, we selected the trending topics with high frequencies in both networks as our ultimate topics. As a result, we obtain 20 trending topics shown in Table I. The following cross-network data analysis work in this section is all based on these 20 topics and we will only use the topic index "T#" in the subsequent subsections for the sake of brevity.

Table I. The final selected hot topic list

	Topic		Topic		Topic
1.	US presidential election 2012	2.	gangnam style	3.	Super bowl 2013
4.	Olympic games 2012	5.	Justin Bieber	6.	Star wars film
7.	The Dark Knight Rises	8.	Minecraft Game	9.	Samsung Galaxy S III
10.	Michael Jackson 2012	11.	Christmas 2012	12.	Google Nexus 4 release
13.	Iphone5 release	14.	Black Ops II	15.	Doctor Who TV Series
16.	Prometheus	17.	Google glasses	18.	Call me maybe
19.	Spider Man	20.	Skyfall		

Thereafter, we will describe how we identify and represent the selected trending topics in Twitter and YouTube separately since different social networks tend to use different terms to indicate the trending topics. For instance, people may use "Obama election 2012" or "Mitt Romney election" to indicate the topic "US presidential election 2012" in YouTube, while they may adopt the hashtags such as #USelection, #voteobama or #Obama2012 to indicate this topic in Twitter. To capture all the terms which can represent the trending topics in Twitter and YouTube, respectively, we use the YouTube Search API to search for all the 20 trending topics in YouTube engine and aggregate the tags of the returned videos; while in Twitter we search for the related *tweets* from the downloaded *tweet* dataset since the Twitter Search API can only retrieve the *tweets* currently posted. Then we adopt the same word frequency sorting method as in the previous topic extraction procedure and manually select the high frequency terms which can represent the topic as our indicator for the topic. In this way, we count how many in the 7,686 users have referred to the different selected topics via their user tag clouds in Twitter alone, in YouTube alone and in both Twitter and YouTube, respectively. The result is shown in Table II. We can see that users are more active in Twitter and for all these topics the number of involved users is larger in Twitter than YouTube. Moreover, many users pay attention to certain topics in both Twitter and YouTube.

Table II. The number of users who have referred to each of the selected trending topics in Twitter alone, YouTube alone and both the two, respectively

Topic No.	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Twitter	2908	3850	1107	1376	1071	2385	2251	857	1164	519
YouTube	1203	1391	726	369	448	1394	757	616	525	367
Both Two	667	706	312	139	88	644	376	243	216	68

Topic No.	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20
Twitter	4155	1434	2708	890	1114	791	1704	897	951	1254
YouTube	1476	416	595	188	717	291	923	640	305	333
Both Two	856	220	294	68	317	97	367	143	96	112

[7] http://www.marketingcharts.com/wp/topics/entertainment/google-reveals-2012s-top-trending-searches-25381/.

3.3 Attention of Topics in Global Level

To verify the effectiveness of our method to represent and track hot topics, we select the popular topic "Super bowl 2013" from the trending topic list in Table I and count how many users in our dataset begin to pay attention to this topic each day in both Twitter and YouTube during an observation of two months from Jan. 1^{st} to Feb. 28^{th}, 2013 since the final round of super bowl happened on Feb. 4^{th}, 2013. The temporal dynamic results are shown in Fig. 2. We further track this topic with the real-world timeline from its Wikipedia page [8] and the real events are labeled in this figure. We find that the statistics from our dataset in Twitter well capture the real sub-events during the timeline of "Super bowl 2013", since we can see from Fig. 2 that a peak of user attention occurs near each of the important sub-events which in turn verifies the effectiveness of our mechanism to represent and track the topics.

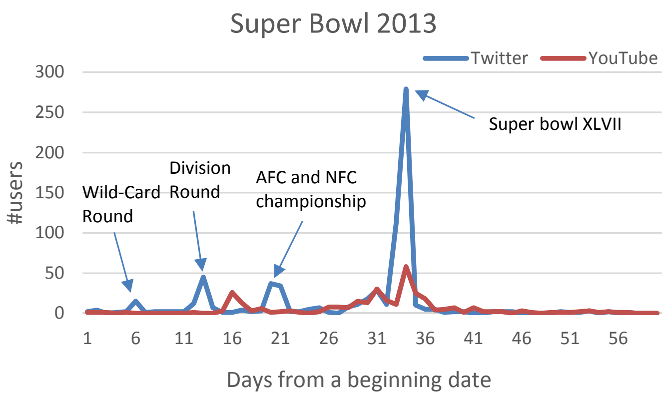

Fig. 2. The temporal dynamics of global user attention on the topic "Super bowl 2013".

Moreover, we can see from the figure that many users in Twitter began to follow this topic since the Wild-Card Round and Division Round, while in YouTube users mainly began to follow this topic after the Division Round ends. The attention in both networks achieves the peak when the final day of "Super Bowl 2013" came. However, the attention in Twitter rises up earlier and more steeply than that in YouTube during the last several days of the "Super bowl 2013", which further indicates the time-awareness of Twitter platform. The different characteristics of the platforms can partly explain this phenomenon. On one hand, it is more convenient to upload or transmit texts than videos, because 1) uploading videos is more time-consuming and 2) the videos about some events are not always available. On the other hand, textual stream platforms are user-centric, while multimedia sharing platforms are content-centric [Benevenuto et al. 2009]. The interactions between users are often more frequent than those between user and content. Therefore, information emergence and propagation in Twitter is faster than that in YouTube on global level.

3.4 Attention of Topics in User Level

After the global temporal analysis, we further investigate into the temporal patterns across Twitter and YouTube on user level. In other words, we try to figure out when a topic emerges, whether the majority of users are first involved in this topic in Twitter and then go to YouTube for more details or vice-versa. Therefore, we first collect the users who have referred to the topics in Table I in both Twitter and YouTube (the number of the available users can be found in Table II). Then we analyze the users' behavior data to find when the users first referred to the topics in Twitter and YouTube, respectively. We use the same method to identify the topic as in Section 3.2 and if the topic indicator occurs in user's *tweets* or video tags, we consider that the user refer to the topic. We examine in which network the user pay attention to the selected topic earlier and count the votes from the users who have referred to the topics both in Twitter

[8] http://en.wikipedia.org/wiki/Super_Bowl_XLVII.

and YouTube. Besides, we calculate the ratio between the Twitter faster votes and YouTube faster votes. The result is shown in Table III. We can see that the number of user votes for "Twitter is faster" is far larger than that for "YouTube is faster" almost on all topics. In other words, a larger proportion of users tend to first follow trending topics in Twitter and then go to YouTube for more details on these topics. Therefore, the local temporal analysis based on each single user also meets the global patterns demonstrated in Section 3.3, indicating that information emergence and spread in Twitter is also faster than that in YouTube on user level.

Table III. The number of user votes for "Twitter is faster" and "YouTube is faster" and their ratio

Topic No.	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
#Twitter faster votes	456	479	159	93	57	379	242	156	157	40
#YouTube faster votes	211	227	153	46	31	265	134	87	59	28
The ratio	2.16	2.11	1.04	2.02	1.84	1.43	1.81	1.79	2.66	1.42
Topic No.	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20
#Twitter faster votes	561	179	211	51	202	59	236	78	55	52
#YouTube faster votes	295	41	83	17	115	38	131	65	41	60
The ratio	1.93	4.56	2.54	3.00	1.76	1.55	1.80	1.20	1.34	0.87

3.5 Reaction to Certain Topics in Category Level

In section 3.4, we also find the user vote ratios are different for different topics where for the topic "Skyfall", the number of the YouTube faster votes is even a little larger than that of the Twitter faster votes. Therefore, a further question naturally arises: whether the temporal order between Twitter and YouTube has something to do with the topic category. To further investigate it, we classify the examined topics into five categories: "Technology", "Political", "Movie", "Entertainment" and "Sport". We further calculate the average user vote ratio in each category and the result is shown in Table IV. We can see that the ratios on different categories differ a lot, where on the "Technology" category is the largest while the ratio on the "Movie" category is relatively small. It may be partly due to the fact that users are more likely to discuss and get to know full comments on electronic products before they search for the related videos, while a lot of users may share their opinions on the newly-released movies after they have completely watched them in YouTube.

Table IV. The user vote ratio between Twitter and YouTube on different categories

Category	Technology	Political	Entertainment	Sport	Movie
The ratio	2.84	2.16	1.90	1.53	1.46

4. PERSONALIZED TIME-AWARE VIDEO RECOMMENDATION

According to the data analysis in Section 3, we conclude that user behaviors in Twitter are ahead of those in YouTube. Therefore, it is reasonable to recommend related videos on YouTube based on the hot topics users followed on Twitter. In this section, we will elaborate two issues: 1) how to extract the hot topics users are focusing on currently in Twitter, and 2) how to recommend relevant videos to the target user in YouTube based on the extracted hot topics.

4.1 Real-time Hot Topic Detection

In order to capture the short-term interest of a user, it is a vital problem how to extract the hot topics he/she followed based on the *tweets* he/she shared/reshared currently in Twitter. The most obvious characteristic of the *tweet* is its short text format, making each tweet usually expresses one single topic. Traditional topic models often do not work well on *tweets*. To address this issue, Zhao W. et al. [2011] proposed Twitter Latent Dirchlet Allocation (Twitter-LDA) model as an extension to the standard LDA. In the following we first briefly describe Twitter-LDA. Then we will discuss the extension to this model to extract hot topics in our scenario.

4.1.1 *Twitter-LDA.* In Twitter-LDA, it is assumed that each *tweet* is produced by a single topic and a background model. Let $\theta^u \sim Dir(\alpha)$ denotes the topic distribution of user u. Let $\phi^t \sim Dir(\beta)$ denotes the word distribution for topic t and $\phi^b \sim Dir(\beta)$ denotes the word distribution for background model. Let $\pi \sim Dir(\gamma)$ denotes a Bernoulli distribution that governs the choice between the topic words and the background words. When writing a *tweet*, a user u first chooses a topic t from a topic set T based on his/her topic distribution θ^u. Then he/she selects to use the topic t or the background model based on the Bernoulli distribution π. In this way, he/she chooses a bag of words one by one based on the chosen topic or the background model. The details of the generative process are as follows:

1 Draw $\phi^b \sim Dir(\beta), \pi \sim Dir(\gamma)$;
2 For each topic $t = 1, ..., T$,
  (a) draw $\phi^t \sim Dir(\beta)$;
3 For each user $u = 1, ..., U$,
  (a) draw $\theta^u \sim Dir(\alpha)$,
  (b) for each *tweet* $s = 1, ..., N_u$,
    i. draw $z_{u,s} \sim Multi(\theta^u)$,
    ii. for each word $i = 1, ..., N_{u,s}$,
      A. draw $y_{u,s,i} \sim Multi(\pi)$,
      B. draw $w_{u,s,i} \sim Multi(\phi^\beta)$ if $y_{u,s,i} = 0$ and $w_{u,s,i} \sim Multi(\phi^{z_{u,s}})$ if $y_{u,s,i} = 1$

where Dir (Multi) indicates Dirichelt (Multinomial) distribution.

4.1.2 *Handling high-frequency words.* Generally speaking, the background model of Twitter-LDA mainly contains some high-frequency words which are common presented in *tweets*. Actually, the words which represent hot topics also have high occurrence frequencies. In other words, the hot words will have high probability to be assigned as background words in original Twitter-LDA, which will make the hot topic detection unsuccessful. To handle this issue, we explore some extensions to Twitter-LDA. In original Twitter-LDA model, the local background model is changeable with different inputs. However, we know that the high-frequency words are relatively steady in a global level. Inspired by this, we first collect all the *tweets* shared and reshared by users in our dataset from Jan. 2013 to Apr. 2013 and obtained a global background model by Twitter-LDA. Then, we utilize the global background model as a prior probability distribution of local background model when applying Twitter-LDA each time. We call the extended Twitter-LDA model "G-Twitter-LDA". Specifically, when initializing, each word is randomly assigned to a topic or local background model in original Twitter-LDA, while this distribution will be guarded by the global background model in G-Twitter-LDA. Besides, for a *tweet* s, whether each word i in s is generated by local background model or a topic t is also modified by the global background model. For a word i, the probabilities that it is generated by local background model or a topic t are calculated by:

$$p(s = t, z_{u,s,i} = b, y_{u,s,i} = 0) \;\propto\; \lambda_{b_i} * \frac{n_b^{-i} + \gamma}{n_b^{-i} + \sum_{z \in T} n_z^{-i} + 2 * \gamma} * \frac{n_{b,w_i}^{-i} + \beta}{n_b^{-i} + \beta * V} \tag{1}$$

$$p(s = t, z_{u,s,i} = t, y_{u,s,i} = 1) \;\propto\; \lambda_{t_i} * \frac{\sum_{z \in T} n_z^{-i} + \gamma}{n_b^{-i} + \sum_{z \in T} n_z^{-i} + 2 * \gamma} * \frac{n_{t,w_i}^{-i} + \beta}{n_t^{-i} + \beta * V} \tag{2}$$

where λ_{bi} is the generating probability of word i in global background model; λ_{ti} is the generating probability of word i in all topics; n_b^{-i} is the number of words generated by local background model; n_z^{-i} (n_t^{-i}) is the number of words generated by topic z (t); n_{b,w_i}^{-i} is the number of occurrences of word i assigned to local background model and n_{t,w_i}^{-i} is the number of occurrences of word i assigned to the topic t.

Since we target at time-aware recommendation, a time window r is defined. All the *tweets* shared or reshared by users within r are collected, and then G-Twitter-LDA is applied to obtain 1) latent topics and 2) users' distributions on these topics. The time window should be properly chosen because if it is too wide, the short-term interest might not be detected; whereas if it is too narrow, some interesting topics might not be captured. In our experiments, we set the time window r as one day since a day is a normal time unit and users' behaviors often present regular patterns within one day. In other words, the *tweets* users shared or reshared within one day are collected. Note that there are a lot of typos and meaningless words in *tweets*. Therefore, we build up a noun vocabulary using WordNet to keep only noun words and filter out the stopwords. Afterwards, we adopt G-Twitter-LDA to extract the hot topics users followed.

4.1.3 *Short-term interest detection.* Actually, not all of the topics extracted by G-Twitter-LDA are hot topics. Table V shows the ten topics obtained by G-Twitter-LDA on Feb. 4^{th}, 2013 (Super bowl game day). We can find that "Topic 3" exactly presents the hot event "Super bowl", but other topics are not widely-known hot events. Therefore, we need to distinguish the hot topics from the general ones by their characteristics. We know that hot topics are often shared or reshared frequently; hence, the words that represent the hot topics will have high frequencies. Actually, from Table V we can observe that the frequencies of the top-words of the hot topic 3 are much higher than those of other topics, which verifies our assumption. Therefore, we can infer whether a topic is popular by the frequencies of the top-words of this topic. For each topic t, its popularity score $Hot(t)$ is calculated by: $Hot(t) = \sum_{i=1}^{N} p(w_i)$, where w_i is the word which has the ith highest occurrence probability in topic t and N is set to 3 in our model.

After we obtain the popularity score of each topic, we sort these topics by their scores and take the top-3 topics as hot topics. As video watching is time-consuming, we assume that a user shall only care about one single topic in a short while. Therefore, we only focus on the favorite hot topic of each user from the obtained three hot topics and take it as the most accurate expression of the user's short-term interest. Then we retrieve the most relevant videos on this topic in YouTube and recommend them to the target user.

Table V. Topics obtained by G-Twitter-LDA on Super bowl final day

Topic No.	Topic-word learnt distributions by G-Twitter-LDA
T0:	travel–130 tweet–82 talk–70 monday–57 daily–57 pm–56 join–49 art–49 chat–44...
T1:	thanks–232 love–84 february–82 morning–76 free–76 coffee–63 monday–63 people–54...
T2:	social–362 marketing–269 business–232 thanks–205 google–176 twitter–141 content–133...
T3:	**superbowl–2951 bowl–1086 super–1024** power–565 game–368 twitter–330 blackout–279...
T4:	maar–82 iii–79 king–70 foursquare–67 nog–57 mayor–55 lot–50 food–41 parking–40...
T5:	know–279 love–243 people–239 right–239 think–210 game–204 god–144 man–136...
T6:	se–117 man–72 est–68 mi–62 hat–49 apple–46 pas–45 som–42 ne–41 jag–39 mm–36...
T7:	photo–208 photography–133 news–94 dead–59 daily–54 hostage–53 alabama–46 child–44...
T8:	people–94 love–77 quote–77 life–67 obama–62 think–58 work–57 school–52 read–48...
T9:	android–235 google–118 twitter–111 windows–93 iphone–86 mobile–80 phone–75...

4.2 Personalized Video Recommendation

In the above subsection we obtained the hot topics users followed in Twitter within a time window r. In this subsection we will elaborate how to recommend the relevant videos in YouTube to the target user: 1) We will first introduce how to extract the videos relevant to the hot topics, and then 2) we will re-rank these videos by user profile in YouTube.

4.2.1 *Hot topic-related video extraction.* In Section 4.1, we know that each hot topic can be represented by some weighted words which have high frequencies in the topic-word distribution obtained by G-Twitter-LDA; hence we only keep the top-10 words sorted by their frequencies and use Vector Space Model (VSM) to describe hot topics. Besides, for a video, the "title", "category", "tags" and "description" are regarded as its significant semantic expression. However, these data annotated by web users contains plenty of noises such as meaningless words or typos. To tackle this issue, we adopt WordNet to filter out the noises and only keep noun tags which are the least noisy representations for a video. Given a video v, its similarity with a hot topic t is calculated by cosine value:

$$Score(v|t) = \frac{\mathbf{v}^T \mathbf{t}}{\sqrt{\mathbf{v}^T \mathbf{v}} \sqrt{\mathbf{t}^T \mathbf{t}}} \tag{3}$$

where $\mathbf{v(t)}$ is a feature vector of v (t); $\mathbf{v} \in \mathrm{R}^d, \mathbf{t} \in \mathrm{R}^d$, d is the dimension of the vocabulary.

After obtaining the score of a video, we can judge whether the video is hot topic-related by setting a threshold $f_H \in (0, 1)$. If the score of the target video is no less than f_H, it will be regarded relevant to the hot topic and put into the candidate video pool, which may be recommended to the target user. Given a video v, the indicator $I(v)$ is:

$$I(v) = \begin{cases} 1 & Score(v|t) \geq f_H \\ 0 & otherwise \end{cases}$$

where 1 (0) indicates that the video v is relevant (irrelevant) to the hot topic t.

Moreover, as analyzed in Section 3, the information emergence and propagation in Twitter is very fast and the event outbreaking in the real world will be posted on Twitter immediately and transmitted quickly. The Twitter users could get to know the event almost without any delay. Thus the uploading time of the recommended videos shall be as close as possible to the time the target user attends to the event; otherwise some irrelevant videos may be involved in and recommended to the user. Therefore, we define a penalty factor based on the time interval (denoted as f_T) between the video uploaded time (denoted as T_u) and the user attention time (denoted as T_a), i.e. $f_T = T_u - T_a$. The score of the video v is updated with this factor and expressed as follows:

$$Score(v|t, f_T) = Score(v|t) * 2^{-|T_u - T_a|^{\lambda_T}} \tag{4}$$

where control parameter $\lambda_T \in [0, 1]$.

Furthermore, the quality of the video is very important in recommendation. Many videos have the similar associated tags and content but their qualities may differ a lot, leading to different popularities. Thus we need to consider the quality of each video when recommending. We know that videos are often rated by the users who watched them. The ratings can be regarded as an effective factor to measure the qualities of videos. Besides, people tend to watch it if the view count of a video is very large. Therefore, we take the rating (denoted as f_R) and view count (denoted as f_C) as quality factors to refine the score of each video.

$$Score(v|t, f_T, f_R, f_C) = Score(v|t) * 2^{-|T_u - T_a|^{\lambda_T}} * f_C^{\lambda_C} * e^{\lambda_R (f_R - 5)} \tag{5}$$

where control parameters $\lambda_C \in [0, 1]$, $\lambda_R \in [0, 10]$.

4.2.2 *YouTube profile-based video re-ranking.* So far we only consider the short-term interests of users in Twitter to obtain the video list and it is necessary to re-rank the videos using the long-term interests of users in YouTube. There are two reasons for that: On one hand, the hot topics obtained from Twitter are still in high-level and if we consider the user interests in YouTube, we can get more specific results in finer level. Imagine when a user is detected focusing on "European Championship", if all the videos about this event are recommended to him/her, it is inelegant and could not capture the user at his/her first glimpse. However, if the user is further detected as a fan of David Beckham from his/her YouTube profile and we recommend the videos Beckham joined in to him/her, it is more likely capture the user immediately. On the other hand, user behaviors differ from platform to platform, and user interests in multimedia sharing platforms may not be fully expressed by user behaviors in social textual stream platforms. Therefore, mining the long-term interests of users in YouTube is also significant.

The registration information (like "AboutMe", "Hobbies", "Movies") of users is very useful to express their preferences. Besides, users' active actions (like "upload", "favor", "add to playList") on videos strongly indicate their attentions and preferences as well. Therefore, the users' profiles can be built up by extracting the "titles", "tags", "categories" and "descriptions" associated with those videos as well as the registration information. Then we also utilize WordNet to filter out the noises and only keep noun tags to built up user profiles in YouTube. The similarity of each video with the user profile in YouTube is measured by cosine value:

$$Score(v|u) = \frac{\mathbf{v}^T \mathbf{u}}{\sqrt{\mathbf{v}^T \mathbf{v}} \sqrt{\mathbf{u}^T \mathbf{u}}} \tag{6}$$

where **u** is a feature vector of user u in YouTube; $\mathbf{u} \in \mathrm{R}^d$, d is the dimension of the vocabulary.

Afterwards, we obtain the final score of each video considering all the factors: short- and long-term interests of users, as well as time factors and quality factors of videos.

$$\begin{aligned} &Score(v|u,t,f_T,f_R,f_C) \\ =&(\lambda_U Score(v|u) + (1-\lambda_U)Score(v|t)) * 2^{-|T_u-T_a|^{\lambda_T}} * f_C^{\lambda_C} * e^{\lambda_R(f_R-5)} \\ =&(\lambda_U \frac{\mathbf{v}^T\mathbf{u}}{\sqrt{\mathbf{v}^T\mathbf{v}}\sqrt{\mathbf{u}^T\mathbf{u}}} + (1-\lambda_U)\frac{\mathbf{v}^T\mathbf{t}}{\sqrt{\mathbf{v}^T\mathbf{v}}\sqrt{\mathbf{t}^T\mathbf{t}}}) * 2^{-|T_u-T_a|^{\lambda_T}} * f_C^{\lambda_C} * e^{\lambda_R(f_R-5)} \end{aligned} \tag{7}$$

where $\lambda_U \in [0,1]$ is the weight parameter.

Finally, we re-rank these videos according to their final scores and recommend the top-10 videos to the target user.

5. EXPERIMENTS

The above section has elaborated the process of the proposed approach. Next we will evaluate the performance of the approach and demonstrate the impact of the short-term interest on the performance of personalized recommendation. Meanwhile, we will clarify the question: what influence do the parameters have on the performance of the proposed approach?

5.1 Experimental Settings and Evaluation Metrics

Through data analysis in Section 3, we found that "Twitter is faster" phenomenon is more significant on short-standing topics than long-standing ones. Moreover, long-standing topics usually involve with several attention peaks (see Fig. 2 for illustration), making the corresponding short-term user interests difficult to measure. Enlightened by this, we are firstly interested to investigate the performance of the proposed personalized time-aware video recommendation solution on short-standing topics. Therefore, to validate the effectiveness of our proposed method, we selected 10 popular short-standing hot topics which happened within one day for the experiments. Some of these topics are from the 20 ones in Table I (e.g. "Google Nexus 4 release"); some others are a portion of the long-standing topics among the 20 ones, for example, "US presidential election 2012, 1st debate" and "US presidential election day 2012" are from the topic "US presidential election 2012". The selected topics are presented in Table VI.

Table VI. The final selected hot topic list for experiments

	Topic		Topic		Topic
1.	US presidential election, 1st debate	2.	Obama inauguration 2013	3.	US election day 2012
4.	Super bowl final 2013	5.	Olympic opening ceremony 2012	6.	Star wars film release
7.	Dark Knight Rises release	8.	Google Nexus 4 release	9.	Iphone5 preach
10.	Google glass release				

To simplify the process of the experiments, we only focus on the tested hot topics. We take the topic "Olympic opening ceremony 2012" as an example to elaborate the experimental procedure. We collect all the *tweets* tweeted and retweeted on July 27^{th}, 2012 by the users in our dataset. Afterwards G-Twitter-LDA is adopted to get the hot topics and user-topic distributions on that day. Next, the users whose favorite topic (i.e., having the biggest distribution value on this topic) is "Olympic opening ceremony 2012" on that day will be recommended related videos in YouTube. The videos uploaded on or before July 28^{th}, 2012 in our dataset are taken as test videos and the total number of the videos in the dataset is 277,932. In order to evaluate the effectiveness of our recommendation method, the hot topic-related videos commented, rated, favored or uploaded by the users in the day or after are taken as the ground truth [9].

[9] The user's video-viewing behavior in YouTube may not be stimulated by the hot topic he/she follows in Twitter if the viewing behavior happens far behind the following behavior. However, the viewing behavior indicates the user's preference on this topic and can be regarded as ground truth in the experiments.

We compute the recall and precision of the top-N recommended videos and utilize the average F-score as our final evaluation metrics.

$$F - score(N) = 2 \cdot \frac{Precision(N) \cdot Recall(N)}{Precision(N) + Recall(N)} \tag{8}$$

The examined strategies include:

(1) recommend by random (Random);
(2) recommend by trending videos [10] (Trend);
(3) recommend by only Twitter hot topic (HT);
(4) recommend by only YouTube user profile [11] (UP);
(5) recommend by Twitter hot topic and YouTube user profile (HT+UP);
(6) recommend by Twitter hot topic and YouTube user profile considering time factor (HT+UP+TF);
(7) recommend by Twitter hot topic and YouTube user profile considering quality factor (HT+UP+QF);
(8) recommend by Twitter hot topic considering time and quality factors (HT+TF+QF);
(9) recommend by YouTube user profile considering time and quality factors (UP+TF+QF);
(10) recommend by Twitter hot topic and YouTube user profile considering time and quality factors (HT+UP+TF+QF).

5.2 Parameter Settings

In G-Twitter-LDA model, there are three hyperparameters: α, β and γ. We empirically fix the parameters according to the prior expectation about the data. The hyperparameter α controls the mixing degree of user-topic distribution and big value of α encourages high mixing of topics. As our goal is to identify the dominant hot topic, the user-topic hyperparameter α is fixed to a relatively small value of $\alpha = 1$ to discourage topic mixing. In the similar way, we set the hyperparameters $\beta = 0.05$, $\gamma = 0.5$. Besides we assume that the number of hot topics is no more than 10 each day. Therefore, we set the number of latent topics as $N_T = 10$ in G-Twitter-LDA.

In our proposed personalized time-aware video recommendation model, five parameters are involved: f_H, λ_U, λ_T, λ_C and λ_R. The parameter f_H decides whether the candidate videos are hot topic-related or not. Small f_H indicates that more videos will be regarded as hot topic-related and less videos will be filtered out. Since we have calculated the similarities of the candidate videos with hot topics, we do not need to filter out too many videos and set f_H to a small value of $f_H = 0.1$. The parameter λ_U controls the weight of user YouTube profiles which express the long-term interests of users on video domain. Therefore, the similarities of the candidate videos with YouTube profiles will have a bigger influence on the performance than those with hot topics after we filter out the irrelevant videos by f_H. Enlightened by this, we set λ_U to a big value of $\lambda_U = 0.7$. The parameter λ_T controls the weight of time factor. We set λ_T to a relatively small value of $\lambda_T = 0.3$ to allow the recommended videos in a broader time range. The parameters λ_C and λ_R control the weights of quality factors of videos. Since the view counts of popular videos are very big, we set λ_C as a relatively small value of $\lambda_C = 0.4$ to avoid that only the popular videos are recommended to users. On the contrary, the differences of the video ratings are small. Therefore, we set λ_R to a big value of $\lambda_R = 5.0$ to encourage the videos which have higher ratings. The following experiments will be launched based on the above parameter values.

5.3 Experiment Results and Analysis

In this subsection, we will first compare the performance of Twitter-LDA and our G-Twitter-LDA and demonstrate the obtained hot topics by G-Twitter-LDA. Then, we will compare the performance of different strategies. Following that the influence of parameters on the performance will be investigated.

[10] Trending videos, measured by the time and quality factors, are not limited to the ones related to the hot topics user currently is focusing on.
[11] The information associated with the ground truth has not been included in user profile.

Table VII. The comparison of background models obtained by original Twitter-LDA and G-Twitter-LDA

Model	Background model	Hot topic
Twitter-LDA	**olympics** thanks google love know **opening** people think right us twitter photo **ceremony**...	Olympic opening ceremony
G-Twitter-LDA	us oh world check nice post home photo weekend free night morning cool blog life pic...	
Twitter-LDA	**vote election voting obama** know **romney** people think us thanks right love work google...	US election day 2012
G-Twitter-LDA	us photo post news check win free oh ballot world blog line morning nice real voting...	

5.3.1 *Comparison of Twitter-LDA and G-Twitter-LDA.* To illustrate the effectiveness of G-Twitter-LDA model in our scenario, we compared it with the original Twitter-LDA model. Table VII demonstrates the background models obtained by Twitter-LDA and G-Twitter-LDA, respectively. We can see that some key words of the hot topics "London Olympics opening ceremony" and "US presidential election 2012" are assigned to the background models with high frequencies in original Twitter-LDA, while in our G-Twitter-LDA, there are just some normal high-frequency words in the background models. This is due to the fact that "London Olympics opening ceremony" was only popular around Jul. 27^{th}, 2012 and "US presidential election 2012" was only popular around Nov. 6^{th}, 2012, but from a long-term perspective, the words that represent the two topics do not have high frequencies so that they will not be regarded as background words in our G-Twitter-LDA.

Table VIII. The comparison of hot topics obtained by original Twitter-LDA and G-Twitter-LDA on "Super bowl final" day

Model	Latent topics
Twitter-LDA	T0: se est man mi ne hat pas som jag ma men war oder fr apple pr ir os morgen pi son pinter...
	T1: android google mobile iphone windows phone tech chrome jailbreak apple carbon news...
	T2: commercial god farmer paul dodge **superbowl** ha harvey ram bell taco black **super bowl**...
	T3: **superbowl game** power ray halftime lewis football blackout commercial baltimore outage...
	T4: photo maar art nog hoe dead comic nu ben caw walking artist cat social door valentine...
	T5: travel photography snow photo monday foursquare february pm join talk chat chicago...
	T6: social thanks marketing business blog content post google job iii free email mobile search...
	T7: **super bowl** power **superbowl** oreo twitter outage game blackout problem commercial...
	T8: obama health gun abc america news government world brain human local school edu...
	T9: love quote life morning food coffee monday feel drink oh pain penny heart bit lunch mind...
G-Twitter-LDA	T0: travel tweet talk monday daily pm join art chat cst love chicago caw photo grand castle...
	T1: thanks love february morning free coffee monday people month think work social know...
	T2: social marketing business thanks google twitter content know job daily mobile data think...
	T3: **superbowl bowl super power game** twitter blackout oreo outage halftime watching think...
	T4: maar iii king foursquare nog mayor lot food parking theatre car russian nu bones park...
	T5: know love people right think game god man look work oh hope thanks shit feel watch...
	T6: se man est mi hat apple pas som ne jag mm war oder ha fr google pr os men ma morgen...
	T7: photo photography news dead daily hostage alabama child abc safe nature wind space...
	T8: people love quote life obama think work school read america gun god human know years...
	T9: android google twitter windows iphone mobile phone jailbreak think carbon tech chrome...

Besides, the hot topics extracted by G-Twitter-LDA model shall be more compact since the background model is well obtained. To verify this assumption, we examine the hot topics extracted by Twitter-LDA and G-Twitter-LDA on Feb. 4^{th}, 2013 (Super bowl game day) which is demonstrated in Table VIII. We can see that both the two models have detected the hot event "Super bowl game". However, the key words of the event are distributed in serval extracted topics in Twitter-LDA, while in our model, they are just in the top positions of topic 3, which further demonstrates the effectiveness of G-Twitter-LDA in our application.

Furthermore, the automatically detected hot topics by G-Twitter-LDA are shown in Table IX. The first column shows the date when the hot topics happened and the second column presents the known hot topics which happened during the corresponding date; and the third column demonstrates the detected hot topics, where the topic order and top words are presented. We can observe that all of the tested hot topics in Table VI are detected (highlighted in blue color). Besides, the top topic-words obtained by G-Twitter-LDA can describe the known events correctly. In other words, G-Twitter-LDA is capable of extracting the latent hot events in *tweets* and applicable in our scenario.

Table IX. The automatically detected hot topics based on G-Twitter-LDA

Date	Hot topic	The detected three hot topics each day
2012-07-20	Dark Knight Rises release	**T0: aurora shooting batman colorado dark people movie knight...** T4: google social know thanks people business think work mobile... T6: thanks know love work people right think friday hope weekend...
2012-07-27	Olympic opening 2012	**T5: olympics opening ceremony watching queen 2012 think watch...** T6: thanks social twitter know love people think marketing business... T9: google apple android lion mountain iphone mobile fiber news nexus...
2012-09-12	Iphone5 preach	**T3: apple iphone iphone5 android phone announcement galaxy think...** T1: thanks twitter google social people know think data business web... T2: know think love work thanks right people feel coffee read look...
2012-10-03	US election, 1st debate	T2: photography social google twitter business thanks content mobile... **T6: debate obama romney tonight watch know athletics people vote...** T5: apple iphone google android mobile galaxy mini store amazon phone...
2012-10-29	Nexus 4 release	T7: hurricane safe stay coast east weather wind power rain hope map... **T2: nexus google android windows apple phone iphone se amazon mi...** T4: thanks social twitter know think people marketing work business...
2012-10-31	Star wars film release	T9: halloween 2012 costume pumpkin badge boo google foursquare... **T7: disney star wars halloween book man movie lucas know episode...** T5: social google business thanks twitter people marketing data content...
2012-11-06	US election day 2012	**T3: vote election voting obama romney polls people know right...** T0: 2012 foursquare badge voting mayor cup yelp altering school... T4: know people vote think right thanks love work hope election...
2013-01-21	Obama inauguration 2013	**T1: obama inauguration president people inaugural years watching...** T2: know people love think right bowl game thanks work look hope... T9: social thanks twitter business marketing google content search...
2013-02-04	Super bowl final 2013	**T3: superbowl bowl super power game twitter blackout oreo outage...** T2: social marketing business thanks google twitter content know job... T5: know love people right think game god man look work oh hope...
2013-02-21	Google glass release	T5: social thanks twitter marketing google business content know love... **T3: google android apple pixel glass mobile iphone chrome nexus...** T4: think know people thanks love right look work hope game man...

5.3.2 *Comparison of different strategies.* The comparison of average F-score of all topics at different depths by different strategies is illustrated in Fig. 3. We can see that the method which combines hot topic in Twitter and user profile in YouTube considering time factor and quality factor (HT+UP+TF+QF) has the best performance. Besides, it is in accordance with the expectation that the performance of the strategy that combines hot topic with user profile (HT+UP) is better than the one that only uses hot topic (HT) or user profile (UP). Moreover, the performance of the strategy HT+TF+QF is much better than the strategy UP+TF+QF, indicating that HT has greater influence than UP on the performance, i.e., the short-term interest contributes the majority of gains in personalized video recommendation. In addition, the performance of the strategy UP+HT+TF+QF is far better than the strategy UP+TF+QF, demonstrating that the performance gain from HT is remarkable. Furthermore, the time factor (TF) and quality factor (QF) have great positive influence since the performance of the strategy HT+UP+TF+QF is much better than the strategies: HT+UP+QF or HT+UP+TF, respectively, and from the improvements we can conclude that TF is more important than QF in our scenario.

Next, in order to investigate the differences of performances on different hot topics and different categories declared in Section 3.5, the average F-score of all of the tested topics on all the hot topic-related strategies are shown in Fig.

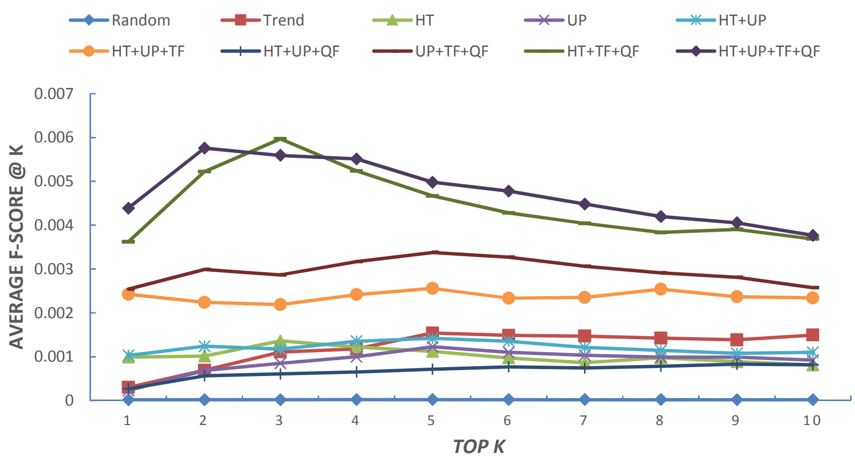

Fig. 3. Different strategies comparing with F-score.

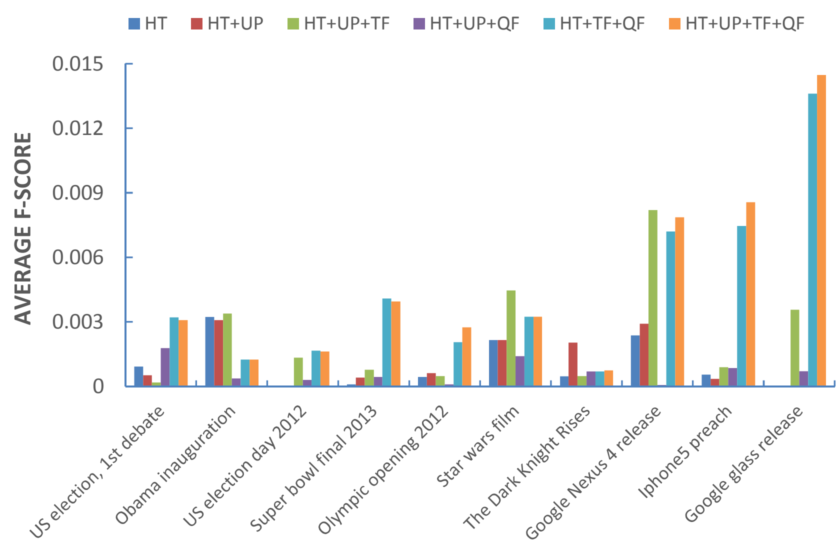

Fig. 4. The comparison of performance on different topics.

4. We can see that the performances of different topics differ a lot. The performance of the topic "Google glass release" which has the highest F-score is far better than that of the topic "Dark Knight Rises release". Besides, the strategy that combines hot topic and user profile considering time factor and quality factor (HT+UP+TF+QF) has the best performance for most of the topics. However, for the topic "Obama inauguration 2013", the performance of the strategy HT+UP+TF+QF is far inferior to HT+UP+TF. In other words, the quality factors have rather negative influence on this topic. This may be explained by the fact that some users focus on politics does not indicate that they really feel interested in it, but politics has a great influence on their life. Therefore, the political videos which have large view counts may not really be favored by users.

Moreover, we observe from Table IV that the temporal characteristic of user behaviors across Twitter and YouTube is category sensitive. In certain categories such as "Technology", the user behaviors in Twitter are obviously ahead of those in YouTube; while it is not evident in some categories such as "Movie". From Fig. 4, we find that the top-3 topics (“Google glass release”, “Iphone5 preach”, “Google Nexus 4 release”) which have the best performance belong to the “Technology” category and the topic (“Dark Knight Rises release”) which has the worst performance belongs to the “Movie” category, indicating that the recommendation performance is also category sensitive. This phenomenon further verifies the rationality of our motivation.

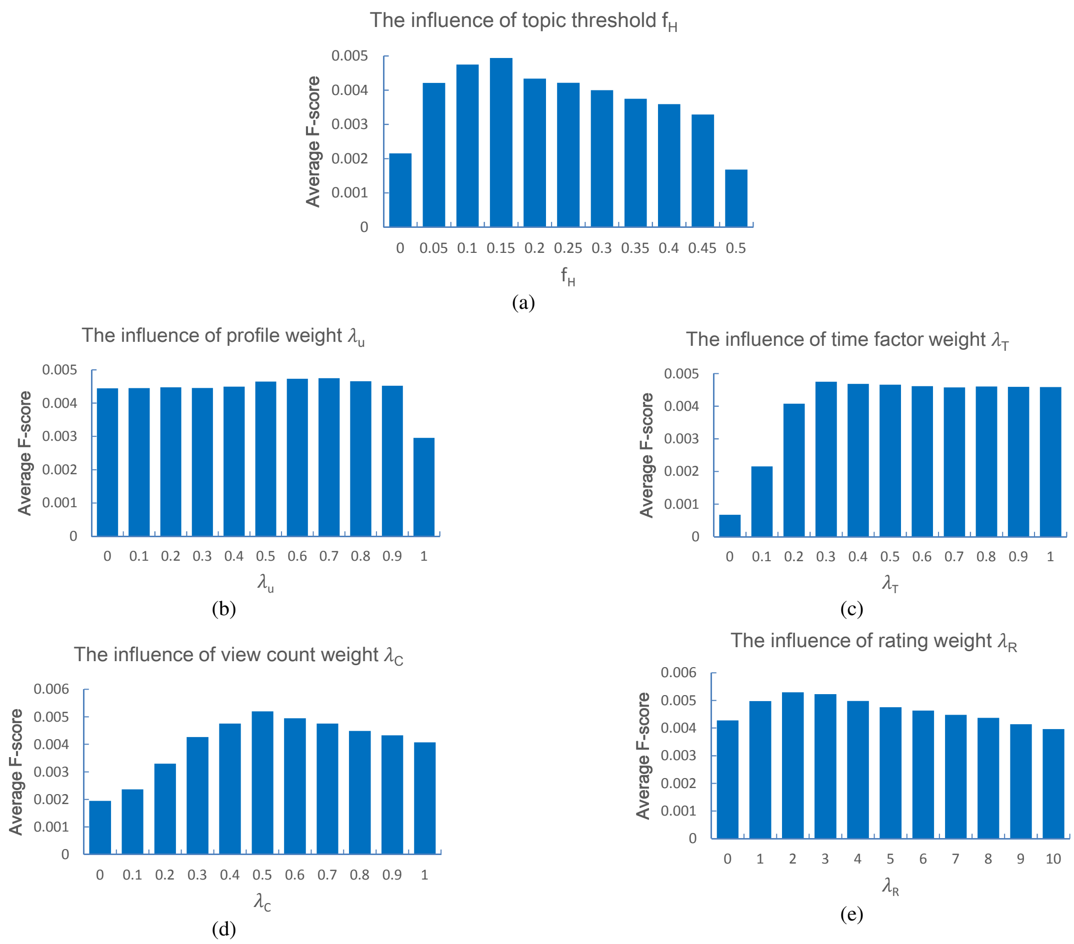

Fig. 5. The influence of the parameters on the performance.

5.3.3 *Influence of parameters on the performance.* In order to illustrate the influence of the five parameters on the recommendation performance in our model, we launch experiments on different parameter settings. We empirically fix these parameters as stated in Section 5.2. Afterwards we observe the change of the performance by tuning the parameters one by one. Specifically, we first fix $\lambda_U = 0.7$, $\lambda_T = 0.3$, $\lambda_C = 0.4$ and $\lambda_R = 5.0$ and tune f_H in [0, 0.5] with the interval of 0.05 to get different results. Then we tune λ_U in [0, 1] with the interval of 0.1 with the rest parameters fixed. In the same way λ_T, λ_C and λ_R are tuned one by one [12] within the domain of definition. Fig. 5

[12] $\lambda_T \in [0, 1]$ with the interval of 0.1; $f_C \in [0, 1]$ with the interval of 0.1; $f_R \in [1, 10]$ with the interval of 1.

demonstrates the impact of f_H, λ_U, λ_T, λ_C and λ_R on the average F-score of top-10 recommended videos of the strategy HT+UP+TF+QF.

We can observe from Fig. 5 (a) that the performance varies a lot with f_H changing from 0 to 1 and it achieves the optimal F-score when $f_H = 0.15$. This figure demonstrates that it is necessary to filter out those videos which are irrelevant to the target hot topics while keeping adequate number of videos in the candidate pool.

Fig. 5 (b) shows that the performance is improving as λ_u increases from 0 to 0.7 and it achieves the optimal F-score when $\lambda_u = 0.7$, indicating that user profile in YouTube has a dominant influence on the performance after we filter out the irrelevant videos by f_H. In other words, re-ranking the videos by YouTube profiles of users is non-trivial as YouTube profiles represent the long-term interests of users in video domain. Besides, the performance at $\lambda_u = 1.0$ (i.e., the short-term interests of users in Twitter is not considered) is much inferior to those at other points, which further confirms the necessity to consider the short-term interest in personalized recommendation.

From Fig. 5 (c) we can see that the F-score at $f_T = 0$ (not considering the time factor) is far lower than those at other points, indicating that the time factor plays a key role in improving the performance. This phenomenon is consistent with our motivation of time-aware recommendation.

Fig. 5 (d) demonstrates that the performance also varies with different f_C values. The optimal F-score at $f_C = 0.5$ is much greater than the F-score at $f_C = 0$ which does not consider the influence of view count. In other words, the video popularity also has a great influence on the performance.

Fig. 5 (e) shows that the performance remains relatively steady when f_R changes within [1, 10]. The reason lies in the fact that the popular videos often have high ratings, i.e., the view count and the rating are consistent in most occasions. Since the view count has been considered, the rating will not have significant impact on the performance.

In general, we can observe that the performance is relatively steady when the values of these parameters 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 curse of parameters.

5.4 Discussions

It is known that users' behaviors are influenced by both their short- and long-term interests. In our work we only focus on users' short-term interests in Twitter and assume that users will get deeper insight into their favorite hot topics in YouTube. However, the question still remains on whether users' further video view behaviors in YouTube are influenced by their long-term interests in Twitter, i.e., whether the consistency of users' short- and long-term interests in Twitter can influence their video view behaviors in YouTube. To investigate into this issue, we conduct an analysis briefed as below.

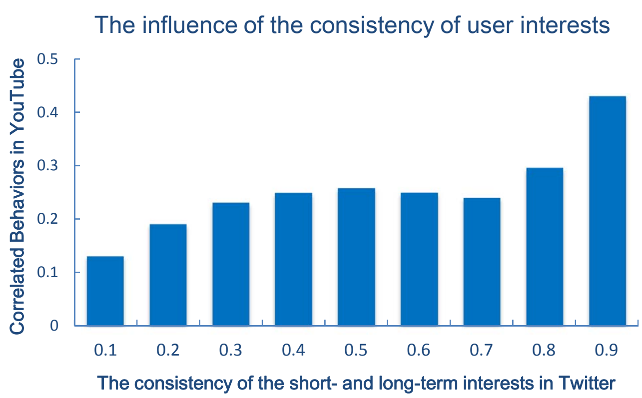

Fig. 6. The correlation of users' behaviors in YouTube with the consistency of their short-term and long-term interests in Twitter.

Firstly, we collect all the *tweets* users shared or reshared in our dataset to obtain the long-term interest distributions of users by Latent Dirchlet Allocation (LDA). Afterwards, we calculate the consistencies of their short- and long-term interests by cosine similarity and inspect whether they have behaviors in YouTube relevant to the target hot topic (short-term interest in Twitter). The statistic results are shown in Fig. 6. We can see that the number of the related behaviors in YouTube is proportional to the consistency of short- and long-term interests in Twitter, which indicates that the recommendation performance will improve if we can obtain the "real short-term interests" of users by considering their long-term interests in Twitter. We will further explore this work in the future.

Besides, it is a common issue for the evaluation of personalized YouTube video recommendation that the users' view histories in YouTube are not accessible via public API. In our experiments, we utilize the available users' behaviors (commenting, rating, favoring or uploading) to imitate their preferences, which occupy only a small percentage of users' real relevant activities, resulting in the relative low average F-score in our experimental results. However, it will not affect the comparison results of our method with the baselines, since the examined methods can still be evaluated by comparing with the ideal best F-score, which is achieved by recommending all the available videos commented, rated, favored or uploaded by users.

6. CONCLUSIONS

In this paper, we have proposed a personalized time-aware video recommendation solution for multimedia sharing platforms (e.g. YouTube) based on cross-platform collaboration from the social textual stream-based platforms (e.g. Twitter). The emergence and propagation of popular topics in Twitter have been observed ahead of that in YouTube on micro user level. Enlightened by this, we employed simple but effective methods to integrate user's short-term interest from Twitter and long-term interest from YouTube so as to realize personalized time-aware video recommendation. Experimental results on ten short-standing hot topics show the considerable improvement over the examined baselines. In the future we will work towards conducting in-depth experimental evaluation by considering the consistency between short- and long-term interests of users in Twitter. Moreover, a unified framework enabling seamless topic detection and video recommendation will be designed.

REFERENCES

ABEL, F., ARAÚJO, S., GAO, Q., AND HOUBEN, G. 2011. Analyzing cross-system user modeling on the social web. *Web Engineering*, 28–43.

ABEL, F., GAO, Q., HOUBEN, G., AND TAO, K. 2011a. Analyzing user modeling on twitter for personalized news recommendations. *User Modeling, Adaption and Personalization*, 1–12.

ABEL, F., GAO, Q., HOUBEN, G., AND TAO, K. 2011b. Semantic enrichment of twitter posts for user profile construction on the social web. *The Semanic Web: Research and Applications*, 375–389.

ABEL, F., GAO, Q., HOUBEN, G.-J., AND TAO, K. 2011c. Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. In *Proceedings of ACM WebSci '11, 3rd International Conference on Web Science, Koblenz, Germany*. ACM.

AHN, Y.-Y., HAN, S., KWAK, H., MOON, S., AND JEONG, H. 2007. Analysis of topological characteristics of huge online social networking services. In *Proceedings of the 16th international conference on World Wide Web*. ACM, 835–844.

BALUJA, S., SETH, R., SIVAKUMAR, D., JING, Y., YAGNIK, J., KUMAR, S., RAVICHANDRAN, D., AND ALY, M. 2008. Video suggestion and discovery for youtube: taking random walks through the view graph. WWW '08. ACM, 895–904.

BENEVENUTO, F., RODRIGUES, T., ALMEIDA, V., ALMEIDA, J., AND ROSS, K. 2009. Video interactions in online video social networks. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP) 5,* 4, 30.

BENNETT, P. N., WHITE, R. W., CHU, W., DUMAIS, S. T., BAILEY, P., BORISYUK, F., AND CUI, X. 2012. Modeling the impact of short- and long-term behavior on search personalization. SIGIR '12. ACM, New York, USA, 185–194.

DAVIDSON, J., LIEBALD, B., LIU, J., NANDY, P., VAN VLEET, T., GARGI, U., GUPTA, S., HE, Y., LAMBERT, M., LIVINGSTON, B., AND SAMPATH, D. 2010. The youtube video recommendation system. RecSys '10. ACM, New York, USA, 293–296.

DENG, Z., SANG, J., AND XU, C. 2013. Personalized video recommendation based on cross-platform user modeling. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE.

FORTUNA, B., MLADENIC, D., AND GROBELNIK, M. 2011. User modeling combining access logs, page content and semantics. *Arxiv preprint arXiv:1103.5002*.

GAO, Q., ABEL, F., HOUBEN, G., AND TAO, K. 2011. Interweaving trend and user modeling for personalized news recommendation. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on*. Vol. 1. 100 –103.

GAO, Y., WANG, M., ZHA, Z.-J., SHEN, J., LI, X., AND WU, X. 2013. Visual-textual joint relevance learning for tag-based social image search. *Image Processing, IEEE Transactions on 22,* 1, 363–376.

JIN, Y., HU, M., SINGH, H., RULE, D., BERLYANT, M., AND XIE, Z. 2010. Myspace video recommendation with map-reduce on qizmt. In *Proceedings of the 2010 IEEE Fourth International Conference on Semantic Computing*. ICSC '10. Washington, DC, USA, 126–133.

KOENIGSTEIN, N., DROR, G., AND KOREN, Y. 2011. Yahoo! music recommendations: modeling music ratings with temporal dynamics and item taxonomy. In *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 165–172.

KOREN, Y. 2010. Collaborative filtering with temporal dynamics. *Communications of the ACM 53,* 4, 89–97.

KWAK, H., LEE, C., PARK, H., AND MOON, S. B. 2010. What is twitter, a social network or a news media? In *WWW*. 591–600.

LERMAN, K. AND GHOSH, R. 2010. Information contagion: an empirical study of the spread of news on digg and twitter social networks. *CoRR abs/1003.2664*.

LIU, A., ZHANG, Y., AND LI, J. 2009. Personalized movie recommendation. In *Proceedings of the seventeen ACM international conference on Multimedia*. ACM, 845–848.

MAGNANI, M. AND ROSSI, L. 2011. The ml-model for multi-layer social networks. In *2011 International Conference on Advances in Social Networks Analysis and Mining*. IEEE, 5–12.

MEI, T., YANG, B., HUA, X.-S., AND LI, S. 2011. Contextual video recommendation by multimodal relevance and user feedback. *ACM Trans. Inf. Syst. 29,* 2, 10:1–10:24.

OSBORNE, M., PETROVIC, S., MCCREADIE, R., MACDONALD, C., AND OUNIS, I. 2012. Bieber no more: First story detection using twitter and wikipedia. In *Proceedings of the SIGIR Workshop on Time-aware Information Access*.

PARK, J., LEE, S., KIM, K., CHUNG, B., AND LEE, Y. 2011. An online video recommendation framework using view based tag cloud aggregation. *Multimedia, IEEE* 99, 1–1.

ROY, S., MEI, T., ZENG, W., AND LI, S. 2012. Socialtransfer : cross-domain transfer learning from social streams for media applications. In *Proceedings of the 20th ACM international conference on Multimedia*. ACM, 649–658.

SANG, J. AND XU, C. 2012. Right buddy makes the difference: An early exploration of social relation analysis in multimedia applications. In *Proceedings of the 20th ACM international conference on Multimedia*. ACM, 19–28.

SZOMSZOR, M., ALANI, H., CANTADOR, I., OHARA, K., AND SHADBOLT, N. 2008. Semantic modelling of user interests based on cross-folksonomy analysis. *The Semantic Web-ISWC 2008*, 632–648.

SZOMSZOR, M., CANTADOR, I., AND ALANI, H. 2008. Correlating user profiles from multiple folksonomies. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*. ACM, 33–42.

WANG, J., CHNG, E., XU, C., LU, H., AND TIAN, Q. 2007. Generation of personalized music sports video using multimodal cues. *Multimedia, IEEE Transactions on 9,* 3, 576–588.

WENG, J., LIM, E.-P., JIANG, J., AND HE, Q. 2010. Twitterrank : finding topic-sensitive influential twitterers. In *WSDM*. 261–270.

XIANG, L., YUAN, Q., ZHAO, S., CHEN, L., ZHANG, X., YANG, Q., AND SUN, J. 2010. Temporal recommendation on graphs via long- and short-term preference fusion. In *KDD*. 723–732.

XIONG, L., CHEN, X., HUANG, T.-K., SCHNEIDER, J. G., AND CARBONELL, J. G. 2010. Temporal collaborative filtering with bayesian probabilistic tensor factorization. In *SDM*. Vol. 10. SIAM, 211–222.

YAN, M., SANG, J., MEI, T., AND XU, C. 2013. Friend transfer: Cold-start friend recommendation with cross-platform transfer learning of social knowledge. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE, 1–6.

YANG, D., CHEN, T., ZHANG, W., LU, Q., AND YU, Y. 2012. Local implicit feedback mining for music recommendation. In *Proceedings of the Sixth ACM Conference on Recommender Systems*. RecSys '12. ACM, New York, NY, USA, 91–98.

YANG, J. AND LESKOVEC, J. 2011. Patterns of temporal variation in online media. In *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 177–186.

ZHAO, W. X., JIANG, J., WENG, J., HE, J., LIM, E.-P., YAN, H., AND LI, X. 2011. Comparing twitter and traditional media using topic models. In *ECIR*. 338–349.

ZHAO, X., LI, G., WANG, M., YUAN, J., ZHA, Z.-J., LI, Z., AND CHUA, T.-S. 2011. Integrating rich information for video recommendation with multi-task rank aggregation. In *Proceedings of the 19th ACM international conference on Multimedia*. ACM, 1521–1524.