12-2011

Tracking web video topics: Discovery, visualization, and monitoring

Juan CAO

Chong-wah NGO
Singapore Management University, cwngo@smu.edu.sg

Yong-Dong ZHANG

Jin-Tao LI

Tracking Web Video Topics: Discovery, Visualization and Monitoring

Juan Cao, Chong-Wah Ngo, *Member, IEEE,* Yong-Dong Zhang, *Member, IEEE,* and Jin-Tao Li, *Member, IEEE,*

***Abstract*—Despite the massive growth of web-shared videos in Internet, efficient organization and monitoring of videos remains a practical challenge. While nowadays broadcasting channels are keen to monitor online events, identifying topics of interest from huge volume of user uploaded videos and giving recommendation to emerging topics are by no means easy. Specifically, such process involves discovering of new topic, visualization of the topic content and incremental monitoring of topic evolution. This paper studies the problem from three aspects. First, given a large set of videos collected over months, an efficient algorithm based on salient trajectory extraction on a topic evolution link graph is proposed for topic discovery. Second, topic trajectory is visualized as a temporal graph in 2D space, with one dimension as time and another as degree of hotness, for depicting the birth, growth and decay of a topic. Finally, giving the previously discovered topics, an incremental monitoring algorithm is proposed to track newly uploaded videos, while discovering new topics and giving recommendation to potentially hot topics. We demonstrate the application on three months' videos crawled from YouTube during Dec 2008 to Feb 2009. Both objective and user studies are conducted to verify the performance.**

***Index Terms*—Topic Trajectory Mining, Video Recommendation, Visualization,**

I. Introduction

WITH the rapid advancement in web technology, social media website has become a convenient platform for people to assess the world and present their opinion. Among different media forms, web video is becoming increasingly popular for its rich audio-visual content. However, the unprecedented explosion in the volume of web videos has made it difficult for web users to quickly access the video topics of concern and for web administrators to conduct a systematic and thorough monitoring of web activities. In some countries, the wide spread of web videos even has became a "social concern". An interesting statistics by YouTube [2] shows that 50% of users watch web videos through recommendations from friends, while no more than 22% of users indeed initiate search queries to explore videos of interest. Driven by such a strong need for recommendation, numerous news websites such as CNN and Sina have designed a column called "Hot Topic" to manually collect hot articles and videos.

The work described in this paper was fully supported by the grants from the National Basic Research Program of China (973 Program, 2007CB311100), the Research Grants Council of the Hong Kong Special Administrative Region, China (CityU 119610), the National Nature Science Foundation of China (60902090) and National High Technology and Research Development Program of China (863 Program,2009AA01A403).

Juan Cao,Yong-Dong Zhang and Jin-Tao Li are with the Institute of Computing Technology, Chinese Academy of Science, Beijing, China (e-mail: caojuan@ict.ac.cn; zhyd@ict.ac.cn; jtli@ict.ac.cn).

Chong-Wah Ngo is with the Department of Computer Science, City University of Hong Kong, HongKong (e-mail: cwngo@cs.cityu.edu.hk).

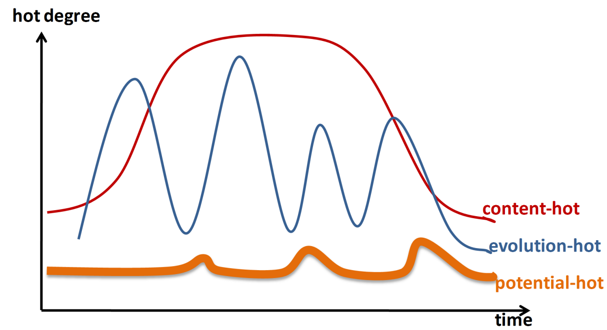


Fig. 1. Trajectories for different types of hot topics. The width of trajectory indicates the strength of correlation among the events of a topic.

In the research community, Topic Detection and Tracking (TDT) [1] is one such effort to automatically structure online news articles into topics. Nevertheless, most approaches in TDT focus mainly on the discovery of popular topics for news browsing [6], [7], [11], [16] while ignoring the evolution trend of topics, which is often a matter of deep concern when performing web video monitoring. Figure 1 shows three types of hot topics [3] often observed in web videos: content-hot, evolution-hot, and potential-hot. The examples of topics include "US presidential election 2008" for content-hot, "Tibet Dalai Lama" for evolution-hot, and "makeup tutorial" for potential-hot. The latter two types of topics are relatively difficult to be discovered and tracked. Evolution-hot topics have a strong evolution trend which repeatedly attracts public attentions through peripheral events. Their contents are often related to some sensational and sensitive news or discussion in the Internet. On the other hand, potential-hot topics are those that are initially confined to a small group of web users at the time of monitoring but is steadily attracting new viewers or participants. They are typically very focused and narrowed in their scope of discussion. Such topics might end up with an erupt trend and are worthwhile to be monitored before they become popular in the Internet.

This paper addresses the discovery, monitoring and visualization of web video topics with various evolution trends. Due to the fact that the textual and visual information of web videos tend to be noisy and sparse, traditional TDT based on full-text analysis is not competent for this problem. Moreover, most approaches in TDT consider topic discovery as the clustering of static dataset [20]. However, considering the massive and dynamic growth of video data in Internet, clustering will be

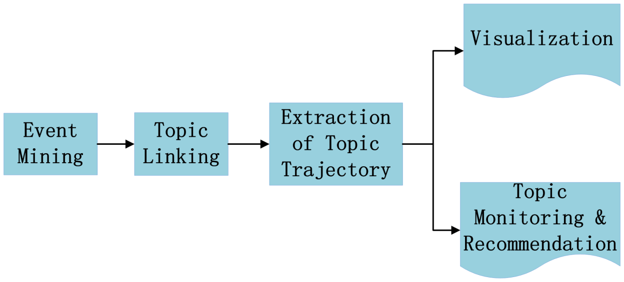


Fig. 2. Overview of proposed framework.

time consuming and furthermore, continuous monitoring and recommendation are more demanding.

In this paper, we model the evolution of video events as a graph, where videos at different time slots are grouped as events and linked via textual-visual similarity. A topic is a salient trajectory extracted from the graph, in which its "hot-degree" change (or evolution) can be vividly depicted along the timeline as shown in Figure 1. The representation allows efficient discovery of topic trajectories, visualization of evolution trends, and monitoring of new, old and potentially hot topics. The contribution of our work includes:

- *Discovery*. We propose techniques for mining evolving topics by detecting bursty tags and events over time. Through a novel inverted-video index, videos of different events are efficiently grouped. The collected events are modeled as a graph and temporally linked via textual-visual similarity. A social-based saliency measure is proposed to extract hot topics with strong development tendency.
- *Visualization*. A topic is presented as a trajectory in two-dimensional space, with one dimension as hot-degree while other as time axis. By attaching tags and videos to a trajectory, the representation not only vividly explores the evolution trend of a topic, but also facilitates the browsing and recommendation of evolution-hot and potential-hot topics.
- *Monitoring and recommendation*. We employ aging theory [6] to depict the evolution of a topic as the change of energies. Three different types of topics: old, new, potential, are separately considered. The developed algorithm based on trajectory extraction and energy modeling is capable of routing newly detected events to an existing topic, signaling bursty events, and recommending potential topics that are likely to exhibit strong evolution trend.

For clarity, we define the following two terminologies. *Event* is a group of related web videos conveying a story and discovered at a time slot. *Topic* is a group of topic related events found over different time slots. Figure 2 shows the overview of proposed framework. The input are web videos each associated with the meta-data including user-supplied tags and video view count. The tags of video normally include few textual keywords while the view count summarizes the number of users who watch the video. Given web videos, the mining of events is conducted by analysis of tag trajectories and clustering of tags (Section III). By considering tag and visual similarities, events are linked for construction of topic evolution graph. A social-based saliency measure is then proposed for the discovery of topics by extracting topic trajectories from the graph (Section IV). By further comparing the newly arrival videos and the previously discovered topics, old, new and potential topics are continuously monitored for video recommendation (Section V). The visualization of hot topics of different types as shown in Figure 1 can also be realized by vividly showing the evolution of topic trajectories in a 2D space (Section VI.D).

II. Related Work

With the explosive growth of web data, timeline analysis plays a vital role in TDT for describing the time-evolving nature of topics. In [11], He *et al.* analyzed the characteristics of word trajectory to classify periodic and aperiodic features for event detection. In [7], Chen *et al.* presented the ideas of mining hot terms by timeline analysis. Using multi-dimensional sentence modeling grounded on hot terms, hot topics are further extracted. In a similar spirit, Fung *et al.* proposed a parameter-free probabilistic model for analyzing time-varying features and detecting bursty events from text streams [10]. The recent work [21] by Wang *et al* presented life-cycle analysis of user attention and media focus to rank the hot degree of topics. These works, however, are based on full-text analysis especially on news articles. Plugging full-text analysis to sparse-text environment such as web video domain remains questionable.

Different from news articles where images and videos play a supplementary role to text, web videos are rich in audio-visual content and surrounding texts are mainly to provide background context. Using sparse texts and limited visual cues for tracking video topics becomes a new challenge for TDT. One such pioneering work tackling the challenge was [15]. Liu *et al.* proposed a bipartite graph reinforcement model to overcome the sparse-text problem. The idea is to densify sparse texts of videos by information propagation through disclosing the bi-directional correlation between videos and texts. However, the model treats videos collected over time as a static database. As a result, the model is not directly extensible for modeling the evolution trend of topics by timeline analysis. More importantly, despite that sparse texts are densified, the amount of texts to describe topics are still limited as evidenced by [15] in the experiments.

While most approaches including [15] deal only with text information, there have been several works which utilize both visual and speech cues for TDT. One early system was developed by Ide *et al* [12]. The system segments news videos into stories and constructs dependencies among stories as a graph structure. A novel interface, named mediaWalker, was also introduced for browsing the development of news topics [13]. In [9], Duygulu *et al.* presented techniques to detect and track repeated sequence of news shots. Topic clusters are discovered based on the association of textual and visual cues. Similarly in [24], Zhai *et al.* proposed techniques to link news stories from different TV channles by textual correlation and keyframe matching. In [22], Wu *et al.* utilized near-duplicate keyrames and speech transcripts to build language

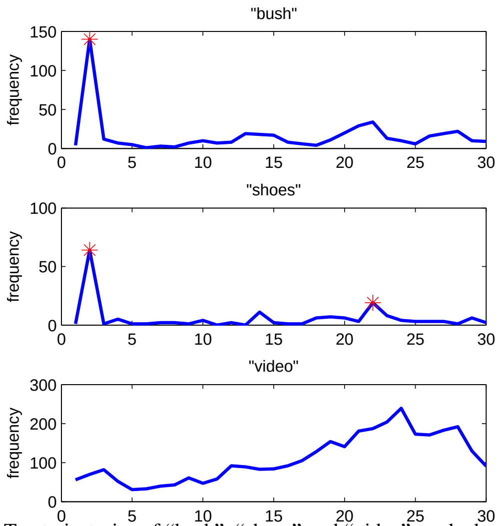


Fig. 3. Tag trajectories of "bush", "shoes" and "video" tracked over a period from 13-Dec-2008 to 13-March-2009. The "*" marks the detected peaks of trajectories. Each time unit has a span of three days.

models for novel news story detection. In a system developed by Neo *et al.*, by leveraging the external sources such as online news articles and blogs, news videos are clustered into a hierarchical structure [18]. With this structure, topic-based video browsing, search and question-answering were explored. All these works, nevertheless, were targeting news videos where speech transcripts and closed captions can be quite reliably extracted by current technologies. Because speech and caption are professionally edited, they indeed provide plenty amount of text-based information to be exploited for TDT. This is in contrast to web videos which are mostly amateur-made, and only limited text information can be derived from speech and caption cues.

III. Event Mining

With vast amount of videos uploaded per day, a fundamental task is the grouping of videos according to the underlying events. This section outlines the clustering of videos at discrete units of time point, with particular focus on the tag trajectory analysis and modeling.

A. Modeling Tag Trajectory

We model the temporal evolution of user-supplied tags as trajectories in a two dimensional space, with one dimension as time, and the other as video frequency. A trajectory is assumed to be encompassed by a history-window starting from a time unit when the tag is observed till the current time unit. Denote $\mathcal{Y}(t)$ as the trajectory of a tag $\mathcal{Y}$ till time unit t_i, then

$$\mathcal{Y}(t) =< y_f(t_{i-n}), y_f(t_{i-n+1}), \ldots, y_f(t_i) > \qquad (1)$$

where $y_f(t)$ is video frequency indicating the number of videos with the tag $\mathcal{Y}$ at time unit t, and $n \geq 0$. A time unit is set to a time span of 3 days, considering the fact that the continuous rolling of interest about a topic will normally last for several days, and so does the number of videos uploaded to social media websites. Intuitively speaking, a bursty $\mathcal{Y}(t)$ will exhibit peak-like trajectory, indicating a large population of users visits an evolving topic centered around a theme. Figure 3 shows examples on the trajectories of three tags "bush", "shoes" and "video" for three months from 17 December 2008 to 14 March 2009. As shown in the figure, the trajectories of "shoes" and "bush" show bursts at several time units, while the trajectory of "video" is relatively flat though with high frequency. It becomes apparent that a tag becomes meaningful at special time of points, and grouping tag trajectory with similar distribution of feature peaks gives clue to arising of events. By this intuition, meaningful tags are first identified by feature peak detection. Simple thresholding of frequency, nevertheless, will give rise to false alarms as the case of "video" where the value of $y_f(t)$ remains high over three months. We thus adopt sliding window technique to determine the set of local peaks $\mathcal{P}^*$ as follows

$$\mathcal{P}^*(t) \geq \mu(y_f(t)) + \alpha \times \sigma(y_f(t)), \text{ for } [t-W, t+W] \qquad (2)$$

where the window size is $2 \times W + 1$, and μ and σ are the mean and standard deviation of $y_f(t)$ within the sliding window. The parameter α controls the burst level of peaks in the frequency dimension, which determine how hot a mined event is expected to be. The window size is typically set to two months (or $W = 10$), based on the observation that the life cycle of a hot topic normally lasts for no more than two months [8]. In Figure 3, three peaks, respectively from "shoes" and "bush" trajectories, are extracted by setting $\alpha = 3$ and $W = 10$.

B. Inverted-Video Index

With the set of peaks $\mathcal{P}^*(t)$ identified at different time points, candidate events at time t can be mined by clustering tags in $\mathcal{P}^*(t)$. For this purpose, we propose a novel tag representation considering the tag-video association and the number of video views by web users. A tag is represented as a vector in m dimensional space, where m is the number of videos at time t. Let v_i be ith video and $\Theta(v_i)$ as the view count of v_i, a tag $\mathcal{Y}$ is represented as

$$\mathcal{Y} = [R_{\mathcal{Y}}(v_1) \times \Theta(v_1), \ldots, R_{\mathcal{Y}}(v_m) \times \Theta(v_m)]^T \qquad (3)$$

where $R_{\mathcal{Y}}(v_i) = \{0, 1\}$ indicating the the presence or absence of $\mathcal{Y}$ in the tag list of a web video v_i. For ease of computation, we also normalize the view count $\Theta(v_i) = [0, 1]$. The user-supplied tags of a video are expected to be subjective and noisy. Generally speaking, videos that are accurately and completely tagged are more searchable and thus are likely to receive more user view counts. So using view count in the representation somewhat hints the accuracy of a tag in describing video content, and also signifies the contribution of a video to a tag. In contrast to conventional representation which describes videos as a vector of tags, Equation (3) encapsulates the association of a tag to videos with social cues, i.e., $\Theta(v_i)$, acquired from social media. Because videos are captured as a feature vector for describing tag distribution, we name the representation as "inverted video index". The idea is similar to inverted file indexing [17] in information retrieval where each term (tag) links to a list of documents (videos) containing the term.

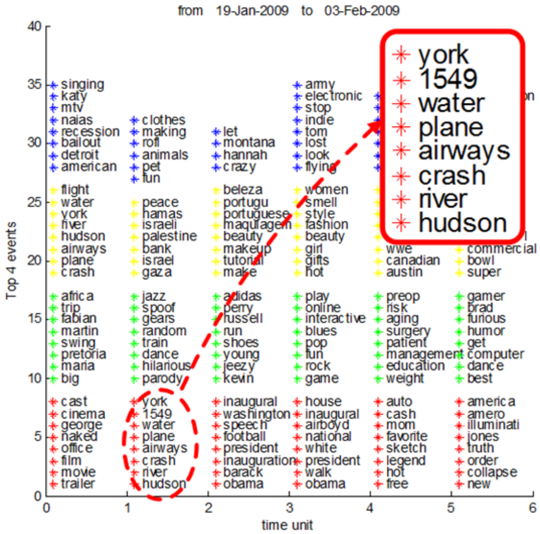


Fig. 4. Events generated from 19-Jan-2009 to 03-Feb-2009. Top-4 clusters of event are shown on each time unit.

C. Event Clustering

Having the tag representation as in Equation (3), events are separately mined from each time unit t. Two major steps are: clustering tags $\mathcal{Y}$ into clusters, and assigning videos to tag clusters. Basically each tag cluster corresponds to a candidate event, and contains a handful of significant terms describing the event. Note that the process involves only bursty tags as identified in Equation (2). In addition, the dimension of a tag vector is not high (the average dimension of tag in 453 time units is 212 in our experiments) since only videos from a time unit are considered for clustering. We adopt a density-based K-means algorithm [4] for tag clustering. An advantage of the algorithm is that the number of clusters K, which can vary across every time unit, can be automatically found without user input. Figure 4 shows the examples of tag clusters mined by the algorithm for YouTube videos from 19 January to 3 February of 2009. The x-axis corresponds to time, and the events at each time unit are mined on a 3-day basis. For illustration purpose, we only show the top-4 largest clusters with the top-8 tags. The results show that most tag clusters are meaningful and basically give clue to the major theme of an event. For instance, the highlighted cluster at time unit 1 contains tags such as "crash", "Hudson", "airways" and so on, which accurately describe an event about the airplane crash occurred on the Hudson river.

The group of tags forming a cluster can intuitively be viewed as a tag pattern uniquely describing an event. Thus the assignment of videos to an event k at time unit t, abbreviated as $\mathcal{E}_k$, can be done by directly matching the tags associated with a video to the tags in a cluster. We implement this by the intersection of two tag sets respectively from a video and a cluster. A video with three or more common tags is then assigned to the corresponding cluster. To this end, we represent an event $\mathcal{E}_k$ in a triple form as following:

$$\mathcal{E}_k =< \mathcal{C}, \mathcal{V}, \Lambda > \tag{4}$$

where $\mathcal{C}$ is the tag cluster, $\mathcal{V}$ is the set of videos uploaded at time t and assigned to $\mathcal{E}_k$. Λ is a value indicating the popularity of an event and derived based on video view counts as following:

$$\Lambda = \sum_{v_i \in \mathcal{V}} \Theta(v_i) \tag{5}$$

The representation takes into account the size of an event in terms of $|\mathcal{C}|$ and $|\mathcal{V}|$, and the popularity of videos. Under this representation, an event containing a large number of videos but does not attract much user attention, is not necessarily more significant than an event with less videos but receives excessive hit counts.

IV. Topic Discovery

The section presents the discovery of topics by grouping the events mined at different time units. The discovery task is basically to inter-link events which share similar theme as a topic through textual and visual analysis. Topics are then ordered by saliency which is defined in terms of their social popularity and life-span evolution.

A. Multi-modal Based Event Linking

We view the set of mined events as in Figure 4 as a graph, denoted as $\mathbf{G} = (\mathsf{V}, \mathsf{E})$. The vertex set $\mathsf{V} = \{\mathcal{E}_k\}_{k=1,2,...}$ is the set of events, and the edge set E is the relationship between any two events. We consider both tag and visual similarities to establish E. Due to the fact that the set of tags describing an event may not be specific, tag similarity itself is not sufficient to deal with the problem of theme-shift phenomenon. For example, an event about basketball star Kobe could share tags such as "match", "score", "goal", "foul" with another event about World Cup. In other words, tag similarity may misleadingly link non-relevant events which may have high similarity in context information as a topic. Extra topic clue based on visual analysis is thus necessary to avoid theme-shift.

In visual analysis, we utilize the fact that videos in a topic may be partially near-duplicate. In web video domain, different versions of an original video are often edited, cropped, or inserted into another video with new material [23]. These versions are often served as a reminder of event milestones of topics, or as a change of opinions or perspectives by adding extra captions or editing effects. By finding and linking partial near-duplicate keyframes, the similarity of two events at different time units can be more objectively measured.

We employ the temporal network algorithm recently proposed in [19] for detection of partial near-duplicates. The algorithm models two frame sequences as a temporal network, and optimally locates a path which partially aligns two sequences by considering visual and temporal consistency. The existence of such path in a network hints the presence of partial near-duplicate in two videos. Figure 5 shows the examples of partial near-duplicates which are mined for the topic " US presidential election 2008".

We define the similarity between two events, $\mathcal{E}_i =< \mathcal{C}_i, \mathcal{V}_i, \Lambda_i >$ and $\mathcal{E}_j =< \mathcal{C}_j, \mathcal{V}_j, \Lambda_j >$, as the linear sum of text and visual similarities, as following:

$$Sim(\mathcal{E}_i, \mathcal{E}_j) = \gamma \times \text{Text}(\mathcal{E}_i, \mathcal{E}_j) + (1-\gamma) \times \text{Visual}(\mathcal{E}_i, \mathcal{E}_j) \tag{6}$$

Fig. 5. Examples of videos about "US presidential election 2008". These videos are related by sharing near-duplicates which are highlighted in the red boxes.

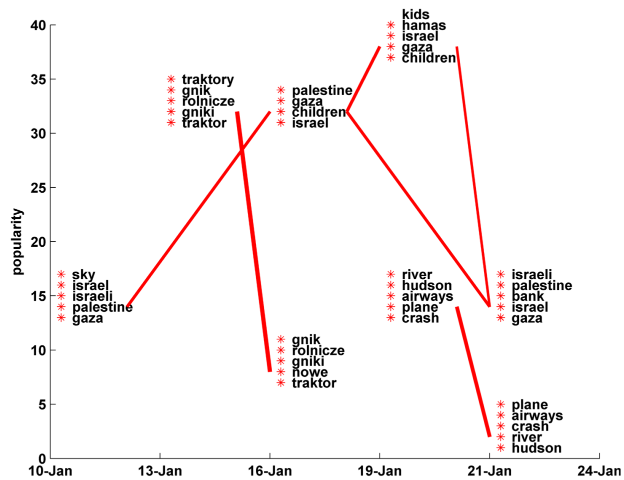


Fig. 6. Topic evolution graph. For ease of illustration, only few examples of events mined from 10-Jan-2009 to 22-Jan-2009 are shown. These events are about the topics "a conflict at Gaza" and "airplane crash". Edge width indicates event similarity.

where

$$\text{Text}(\mathcal{E}_i, \mathcal{E}_j) = \text{Cosine}(\vec{\mathcal{C}}_i, \vec{\mathcal{C}}_j) \quad (7)$$

$$\text{Visual}(\mathcal{E}_i, \mathcal{E}_j) = \frac{1}{\min(|\mathcal{V}_i|, |\mathcal{V}_j|)} \times \sum_{v_m \in \mathcal{V}_i, v_n \in \mathcal{V}_j} \mathcal{ND}(v_m, v_n) \quad (8)$$

and $\mathcal{ND} = \{0, 1\}$ indicating the presence of partial near-duplicate between two videos v_m and v_n. The tag clusters are represented as sparse vectors in vector space model, and the text similarity is computed based on the cosine similarity between tag vectors. Based on Equation (6), two events with similarity score greater than zero will be linked.

Figure 6 shows a graph **G** where events at different time units are linked. The y-axis indicates the popularity of events based on Equation (5). The edge width indicates the similarity of events computed as in Equation (6). By tracing the event links and observing the gradual change of tags in this example, the evolution of a topic can be approximately told. For instance, the topic "a conflict at Gaza" happened at 10-Jan-2009 increasingly captures public attention due to peripheral events shown in videos where kids and children were killed. The topic is eventually cooled down on 21-Jan-2009.

B. Extracting Salient Topic Trajectory

To extract topics from the graph **G** as shown in Figure 6, a straightforward way is to decompose **G** into multiple connected subgraphs. Each subgraph is then regarded as a candidate topic. This simple method, however, does not work in practice since **G** could be noisy with both false and missing links. Figure 7(a) shows a case of false link, where the subgraph indeed contains two topics, and each topic crosses a central node Z. To recover topics from a subgraph, a saliency measure is proposed to extract topic trajectories. We define a topic trajectory as an order set of events which are linked chronologically, denoted as $\mathsf{T} =< \mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_{|\mathsf{T}|-1}, \mathcal{E}_{|\mathsf{T}|} >$ where events are placed in time order. The saliency of T is measured as

$$\text{Saliency}(\mathsf{T}) = \sum_{\mathcal{E}_t \in \mathsf{T}} \Lambda(\mathcal{E}_t) + \sum_{\{\mathcal{E}_{t-1}, \mathcal{E}_t\} \in \mathsf{T}} Sim(\mathcal{E}_{t-1}, \mathcal{E}_t) \quad (9)$$

The first term measures the social popularity of a topic based on the view count of events as defined in Equation (5). The second term measures the topic compactness and evolution trend based on the event similarity defined in Equation (6). Since false events are usually loosely connected to the relevant events of a topic, this measure is effective in pruning false links. Basically, a topic with higher saliency score indicates larger number of popular videos, most videos are tightly linked, and the topic evolves a longer period of time. To extract the set of topic trajectories from a subgraph $\mathcal{G} \in \mathbf{G}$, the following is incrementally done till $\mathcal{G}$ is empty:

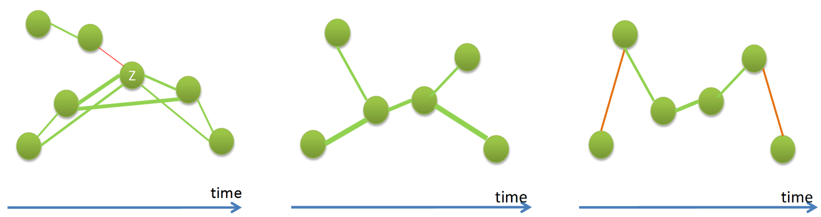
$$\mathsf{T}^* = \arg\max_{\mathsf{T} \in \mathcal{G}} \text{Saliency}(\mathsf{T}) \quad (10)$$

Equation (10) can be engineered by extracting the top-k most salient trajectories, or by discarding topics with few events and low saliency score.

While the problem of false link can be effectively dealt with by the saliency measure, the missing links may cause a topic be split into multiple trajectories. Figure 7(b) shows an example of subgraph $\mathcal{G}$ containing one topic, but events in the topic are sparsely linked, resulting in multiple candidate trajectories. To solve the missing link problem, edge densification is performed on $\mathcal{G}$. The algorithm runs by adding or enhancing edges to two temporally adjacent events $\mathcal{E}_p$ and $\mathcal{E}_q$which can be backtracked to a common event $\mathcal{E}_o$ with strong event links. Figure 7(c) shows the trajectory extracted from 7(b) after edge densification. In brief, we summarize the procedure of topic trajectory extraction in Algorithm 1.

C. Visualization

An extracted topic trajectory can be straightforwardly projected to a 2D space of time and hot-degree. The hot-degree can be defined based on topic popularity as in Equation (5). Figures 10, 11 and 12 show the example trajectories of potential-hot, content-hot and evolution-hot respectively. The trajectories offer a glimpse of topic evolution and compactness. More discussion on topic visualization will be given in Section VI-D.



(a) (b) (c)

Fig. 7. Topic subgraphs: (a) multiple topics in a subgraph, (b) a topic with more than one trajectory, (c) salient trajectory extracted from (b). Each node means an event discovered at time t. The edge width indicates event similarity.

Algorithm 1 Salient topic discovery.

Input: A dataset of web videos and their tags
Output: Topic set $\mathbb{T}$

1) Mine bursty tags with sliding window using Eqn (2).
2) Cluster tags with inverted-video index (Section III-B).
3) Generate events at each time unit, $\mathcal{E} =< \mathcal{C}, \mathcal{V}, \Lambda >$.
4) Multi-modal event linking via Eqn (6) to form a topic graph $\mathbf{G}$.
5) Decompose $\mathbf{G}$ into subgraphs, $\mathbf{G} = \mathcal{G}_1 \cup \mathcal{G}_2 \cup \ldots$, by depth first search (DFS) algorithm.
6) For each subgraph $\mathcal{G}_i$
   a) Perform edge densification.
   b) Incrementally extract salient topic trajectory T^* via Eqn (10), and let $\mathsf{T}^* \in \mathbb{T}$.

V. Incremental Monitoring

Topic monitoring aims to track the most recent topic development from web videos with the knowledge of historical data. This is different from the topic discovery which performs offline mining of topics from a dataset. In other words, topics of videos are monitored continuously, while the number and size of topics may grow or decay as time flies. We investigate three tasks in this section: routing an event discovered at time t to a topic previously found; reporting a new topic; and signaling a potentially hot topic, of either evolution-hot or potential-hot, for recommendation.

A. Modeling Topic Energy

A topic can be considered as a life cycle with stages of birth, growth, decay and death. At different stages of this cycle, a topic is logged (birth), routed and probably recommended (growth), and discarded (decay and death) from historical data. To monitor the evolution, we adopt the aging theory [6] to characterize a topic T as a function of energy as following

$$0 \le F(\mathbf{E}_\mathsf{T}) \le 1 \tag{11}$$

where $\mathbf{E}_\mathsf{T}$ is the energy of T, and $F()$ is a strictly increasing function, i.e., $F(0) = 0$ and $F(\infty) = 1$. We use the following non-linear function to depict the energy change of a life cycle:

$$F(\mathbf{E}_\mathsf{T}) = \frac{10 \times \mathbf{E}_\mathsf{T}}{1 + 10 \times \mathbf{E}_\mathsf{T}} \tag{12}$$

Equation (12) grows steeply in the beginning and fades away gradually since then, which depicts a sigmoid function. This equation has indeed mimically depicted the evolution patterns of most topics of web videos, where the number of video views starts sharply, exhibits periodic peaks with arisen of different peripheral events, and then grows slowly and becomes quiet.

Based on Equation (10), we model the energy of a topic since its appearance till at a time t, as a saliency measure depending on video view count and topic compactness. In other words, the measure is a function monitoring the saliency degree of a topic trajectory T over time, i.e., $\mathsf{T}(\mathsf{t})$. For notation abbreviation, we denote $\mathbf{E}_\mathsf{T}^{(t)}$ as the saliency of T observed up to time t, and $\mathbf{E}_\mathsf{T}^{(t)}$ is defined as,

$$\mathbf{E}_\mathsf{T}^{(t)} = \alpha \times \mathsf{Saliency}(\mathsf{T}) - \beta \tag{13}$$

where α is an energy transfer factor and β is a decay parameter. In the extreme case of no decay, i.e., $\beta = 0$, no existing topics will be removed from monitoring as time progresses. On the other extreme, when no new videos are routed to a topic T after time t, $\mathsf{Saliency}(\mathsf{T})$ converges to a constant value. Consequently $\mathbf{E}_\mathsf{T}^{(t+n)}$ is decayed by a factor of $n \times \beta$, and the topic will be excluded from inspection when the energy becomes zero. By aging theory [6], the optimal values of α and β can be empirically estimated. We will further elaborate the parameter estimation in Section VI-C.

B. Incremental Algorithm

Let $\Delta^{(t)}$ be a time slot at time t and recall that each time slot spans for 3 days. Algorithm 2 presents the procedure for incremental monitoring of topics at $\Delta^{(t)}$. The algorithm is basically an extension of Algorithm 1, with aging theory being taken into account to monitor different stages of topics. The outputs include the newly discovered topics in set $\mathbb{N}$, and the potentially hot topics for recommendation in set $\mathbb{R}$. The step-1 to step-3 of the algorithm detect events $\mathcal{L}^{(t)}$ at $\Delta^{(t)}$. Step-4 expands the topic graph to $\mathbf{G}^{(t)}$ by performing linking between the events in $\mathbf{G}^{(t-1)}$ and $\mathcal{L}^{(t)}$. For efficiency, a sliding window of historial data is imposed such that only the most recent events in $\mathbf{G}^{(t-1)}$ is required to perform linking. To monitor topics, the algorithm only needs to consider the subgraphs in $\mathbf{G}^{(t)}$ which include events in $\mathcal{L}^{(t)}$. In step-6, new topics are discovered by including events in $\mathcal{L}^{(t)}$ which are not linked to any events in $\mathbf{G}^{(t-1)}$. The energies of new topics are initialized with saliency measure, and the new topic set $\mathbb{N}$ at $\Delta^{(t)}$ is then collected.

Step-7 routes new events to existing topics, while making recommendation if a topic is potentially hot and likely to grow further. In step-7c, the energy of a topic T^* is updated by accumulating the saliency of a new event $\mathcal{E}$ included to T^*. The saliency of $\mathcal{E}$ is similar to Equation (9), which is equal to the sum of view counts $\Lambda(\mathcal{E})$ and the edge weight of $\mathcal{E}$. Energy function $F(.)$ is used to monitor the evolution of a topic, by inspecting $F(\mathbf{E}_{\mathsf{T}}^{(t)}) - F(\mathbf{E}_{\mathsf{T}}^{(t-1)}) \geq \theta$, where θ is a user parameter to decide whether a topic should be recommended. Once all events are processed, step-8 performs energy decay to all topics in $\mathbb{T}^{(t)}$. Step-9 further removes insignificant topics from monitoring.

Algorithm 2 Incremental monitoring of topics.

Input: $\mathbf{G}^{(t-1)}$: topic graph; $\mathbb{T}^{(t-1)}$: topic set
Output: $\mathbb{T}^{(t)}$; $\mathbb{N}$: new topics; $\mathbb{R}$: recommended topic set

1) Mine bursty tags at time slot Δ_t using Eqn (2).
2) Cluster tags with inverted-video index (Section III-B).
3) Generate an event list $\mathcal{L}^{(t)} = \{\mathcal{E}_1, \mathcal{E}_2, \ldots\}$ at $\Delta^{(t)}$.
4) Update $\mathbf{G}^{(t-1)}$ by $\mathcal{L}^{(t)}$ to $\mathbf{G}^{(t)}$ via Eqn (6).
5) Decompose $\mathbf{G}^{(t)}$ into subgraphs, $\mathbf{G}^{(t)} = \mathcal{G}_1 \cup \mathcal{G}_2 \cup \ldots$, by DFS.
6) For any $\mathcal{E}_i \in \mathsf{V}$ not linked to any $\mathsf{T} \in \mathbb{T}^{(t-1)}$
   a) Let $\mathsf{T}^* = \mathsf{V} \in \mathbb{N}$ and $\mathsf{T}^* \in \mathbb{T}^{(t)}$
   b) $\mathbf{E}_{\mathsf{T}^*}^{(t)} = \alpha \times \text{Saliency}(\mathsf{T}^*)$
7) For a subgraph $\mathcal{G}_i = (\mathsf{V}, \mathsf{E})$ and $\exists \mathcal{E}_i \in \mathsf{V}$
   a) Perform edge densification.
   b) Update $\mathbb{T}^{(t)} = \mathbb{T}^{(t-1)}$ and remove all $\mathsf{T} \in \mathbb{T}^{(t-1)}$ where $\mathsf{T} \cap \mathsf{V} \neq \emptyset$.
   c) For a topic trajectory T^* extracted via Eqn (10)
      i) Let $\mathsf{T}^* \in \mathbb{T}^{(t)}$
      ii) $\mathbf{E}_{\mathsf{T}^*}^{(t)} = \mathbf{E}_{\mathsf{T}^*}^{(t-1)} + \alpha \times \text{Saliency}(\mathcal{E}_i)$
      iii) If $F(\mathbf{E}_{\mathsf{T}^*}^{(t)}) - F(\mathbf{E}_{\mathsf{T}^*}^{(t-1)}) \geq \theta$, let $\mathsf{T}^* \in \mathbb{R}$
8) $\mathbf{E}_{\mathsf{T}}^{(t)} = \mathbf{E}_{\mathsf{T}}^{(t)} - \beta, \forall \mathsf{T} \in \mathbb{T}^{(t)}$.
9) Remove topics $\mathsf{T} \in \mathbb{T}^{(t)}$ where $\mathbf{E}_{\mathsf{T}}^{(t)} \approx 0$.

VI. Experiment

A. Data Set

We use the version 1.0 of MCG-WEBV dataset [5] for experiments. MCG-WEBV has a total of 80,031 web videos, which is one of the largest available video dataset made publicly available in the literature. These videos are popular YouTube videos uploaded from 13 December of 2008 to 13 March of 2009. The video crawling starts from a set of seed videos which are downloaded from the "Most viewed" videos of "This month" and covers fifteen YouTube categories. This set of seed videos, containing 3,282 videos, are named as the "core dataset". The videos in core dataset are used as seeds to expand the dataset by crawling their "Related videos" and the videos uploaded by the same authors. This eventually forms a pool of 80,031 videos including their associated tags for the MCG-WEBV dataset, and the time period of core dataset covers from Dec. 2008 to Feb. 2009. Figure 8(a) shows the distribution of videos in the dataset. Figure 8(b) shows the corresponding distribution of bursty tags detected via Equation (2), and Figure 8(c) shows the distribution of salient topics discovered by Equation (10).

B. Topic Discovery

We split the evaluation into objective and subjective measures. The objective evaluation is based on the core dataset

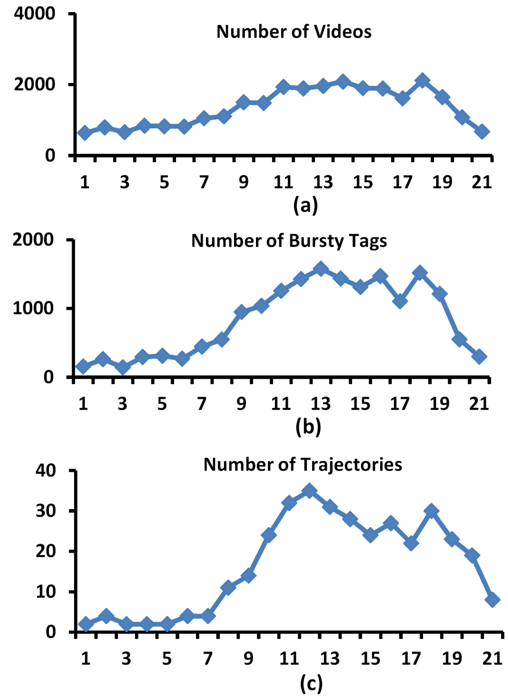


Fig. 8. The number of web videos (a), bursty tags (b), and salient topic trajectories (c) mined from videos during 13-Jan-2009 to 13-March-2009 in MCG-WEBV dataset. Each time unit is a 3-day interval.

of MCG-WEBV. The ground-truth topics of the dataset are manually labeled and generated by 10 assessors. There are a total of 73 ground-truth topics being manually identified. In subjective measure, we perform topic discovery on the whole dataset of 80,031 videos. There is no ground-truth data, thus we conduct user study by inviting assessors to grade the mined topics.

1) Objective Evaluation: We compare our trajectory-based approach in Algorithm 1 to a clustering-based algorithm implemented in the CLUTO toolkit [14]. CLUTO includes various algorithms as well as criterion functions to optimize clustering performance. Because tags are represented as vectors as in Equation (3), we use vcluster program and its default parameters for video clustering. By vcluster, k-way clustering is performed by $k-1$ repeated bisecting of video data, until the k desired clusters are found. During each step, a cluster is bisected so that the resulting clustering solution optimizes a selected criterion function.

For evaluation metric, we use the standard Precision, Recall and F-Measure to assess the performance. Let $\mathbb{T}_G$ and $\mathbb{T}_D$ as the sets of ground-truth and detected topics. We first define the following three functions, respectively, the recall, precision and F-Measure of any two topics $\mathsf{T_G} \in \mathbb{T}_G$ and $\mathsf{T_D} \in \mathbb{T}_D$:

$$R(\mathsf{T_D}, \mathsf{T_G}) = \frac{|\mathsf{T_D} \cap \mathsf{T_G}|}{|\mathsf{T_G}|} \tag{14}$$

$$P(\mathsf{T_D}, \mathsf{T_G}) = \frac{|\mathsf{T_D} \cap \mathsf{T_G}|}{|\mathsf{T_D}|} \tag{15}$$

$$F(\mathsf{T_D}, \mathsf{T_G}) = \frac{2 \times R(\mathsf{T_D}, \mathsf{T_G}) \times P(\mathsf{T_D}, \mathsf{T_G})}{R(\mathsf{T_D}, \mathsf{T_G}) + P(\mathsf{T_D}, \mathsf{T_G})} \tag{16}$$

By Equation (16), each detected topic $\mathsf{T_D}$ is matched to a ground-truth topic where

$$\mathsf{T}_\mathsf{G}^* = \arg\max_{\mathsf{T}_\mathsf{D} \in \mathbb{T}_D} F(\mathsf{T}_\mathsf{D}, \mathsf{T}_\mathsf{G}) \tag{17}$$

The F-Measure of T_D is then measured as

$$\text{Recall}(\mathsf{T}_\mathsf{D}) = \frac{|\mathsf{T}_\mathsf{D}|}{Z} R(\mathsf{T}_\mathsf{D}, \mathsf{T}_\mathsf{G}^*) \tag{18}$$

$$\text{Precision}(\mathsf{T}_\mathsf{D}) = \frac{|\mathsf{T}_\mathsf{D}|}{Z} P(\mathsf{T}_\mathsf{D}, \mathsf{T}_\mathsf{G}^*) \tag{19}$$

$$\text{F-Measure}(\mathsf{T}_\mathsf{D}) = \frac{|\mathsf{T}_\mathsf{D}|}{Z} F(\mathsf{T}_\mathsf{D}, \mathsf{T}_\mathsf{G}^*) \tag{20}$$

where Z is the total number of videos in $\mathbb{T}_D$. The three measures basically bias the topics with larger size of videos.

In the experiments, trajectory-based approach discovers 50 topics. The input to clustering algorithm requires the parameter for number of clusters. We experiment two settings by inputting the number as 50 (the same as trajectory approach) and 73 (ground-truth cluster number) respectively. The discovered topics are matched to the 73 ground-truth topics. The top-10 best-match topics ranked by F-Measure is selected for evaluation. Table I shows the performance comparison by averaging the precision, recall and F-Measure of top-10 topics of both approaches. Overall, when the number of clusters is 50, trajectory-based approach outperforms clustering-based approach on precision with 53.9% and F-Measure with 19.1%. The recall nevertheless is lower than that of clustering approach. The lower recall rate is indeed not surprised because trajectory-based approach considers the saliency of topics in terms of event burst and evaluation trend. As a result, videos which are sparsely uploaded and scattered over time might not meet the criterion and thus are missed. On the other hand, while clustering-based approach offers better recall rate, there are excessive number of false positives in a topic, resulting in difficulty in browsing videos and observing the topic evolution.

We also compare the trajectory-based approach with [15] which is based on bipartite graph model (BGM) with weight updating strategy. This algorithm is proposed for dealing with web videos containing only sparse text. However, the performance of BGM is not satisfactory (precision$= 0.2553$, recall$= 0.6525$) when the number of clusters is set to 73. The result is even worse when weight updating is considered. We observe that this is mainly because the dataset is diverse with noise and many clusters. The model, which is originally designed as a filtering step for topic reranking, cannot effectively deal with noisy data.

In the experiments, we set the parameter $\gamma = 0.8$ in Equation (6). In other words, the textual feature plays a major role compared to visual cue for trajectory-based approach. The setting of γ indeed depends on the nature of topics. For topics related to news, visual cue is effective in dealing with the problem of theme-shift. Specifically, the tags for describing a news topic may change according to the development of a topic, but near-duplicate clips may be inserted into web videos as a reminder of event miletopics. In this case, tags are not as effective as visual cues for topic discovery. On the other hand, visual cue could also generates false links, for example, when the editing effects imposed on two unrelated videos are the same. In this case, tags are likely to perform better than visual cues.

TABLE I
COMPARISON BETWEEN TRAJECTORY-BASED AND CLUSTERING-BASED TOPIC DISCOVERY ON CORE DATASET. (): NUMBER OF CLUSTERS.

	Precision	Recall	F-Measure
Clustering(50)	0.4791	0.8856	0.6068
Clustering(73)	0.5965	0.91	0.7084
Trajectory(50)	0.7376	0.7744	0.7230

TABLE II
SUBJECTIVE USER STUDIES ON TOPIC DISCOVERY. THERE ARE 122 TOPICS EXAMINED BY 5 ASSESSORS. X: A TOPIC, %: PERCENTAGES OF TOPICS, AND (): NUMBER OF TOPICS.

	Strongly Agree	Agree	Disagree	Average Score
X is hot?	60.6% (74)	22.1% (27)	17.2% (21)	0.72
X is precise?	71.3% (87)	18.8% (23)	9.8% (12)	0.81

2) Subjective Evaluation: We further verify the trajectory-based approach in Algorithm 1 on the whole dataset of MCG-WEBV. There are a total of 122 topics being discovered. We invited 5 assessors to grade the topics, by asking the two questions shown in Table II:

1) Topic-X is a hot and important topic, and I would like to learn how the topic evolves.
2) The events in topic-X are accurate enough to convey the main theme(s) of the topic.

Each topic is rated with three-degree scores: strongly agree (1.0), agree (0.5), and disagree (0). The assessors were asked to watch all the events in a topic, and explicitly mark events which are regarded as irrelevant to the topics. Before jumping to the next topic, the assessors were requested to give scores to question-1. For question-2, the score was automatically computed by the system. A topic where no event is marked as irrelevant is regarded as "strongly agree". Meanwhile, a topic which has (more)less than 30% of events being indicated as irrelevant is graded as (dis)agree.

Table II summarizes the result of user evaluation. The last column indicates the score after averaging the scores of five assessors for 122 topics. The result shows that the average scores are 0.72 and 0.81 for question-1 and question-2 respectively, which is very encouraging. The statistics in 2nd to 4th columns are based on majority voting of assessors. In other words, a topic receives score of "strongly agree" if majority assessors cast this vote. In the case of drawn, we consider the best-case scenario. In other words, if "agree" and "disagree" each receives two votes, the topic is graded as "agree". This is reasonable since the other vote is given to "strongly agree" and it is fair to rate the topic as "agree" in this case. Overall, the result shown in Table II is encouraging. There are more than 80% of topics being agreed or strongly agreed to be hot, while more than 90% being regarded as accurate or precise. The topics which are not regarded as hot are mainly those topics containing excessive number of movie clips and TV series.

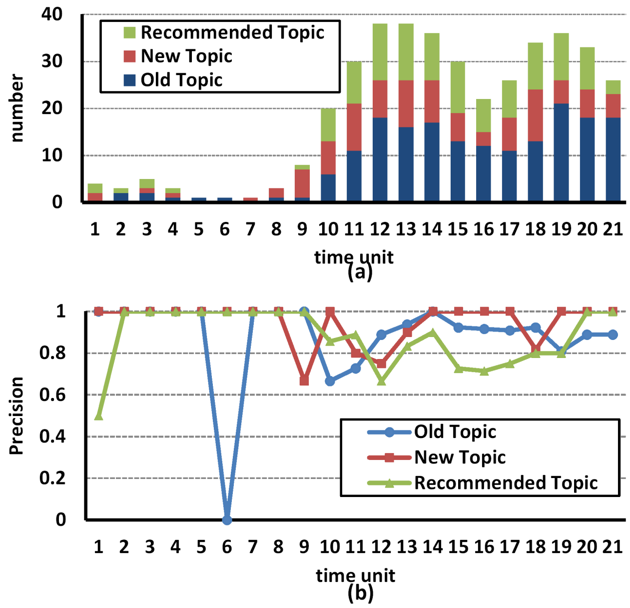


Fig. 9. Performance of topic monitoring: (a) the number of old, new and recommended topics monitored over three months; (b) the precision of tracking different topics.

C. Topic Monitoring

We use the whole dataset of MCG-WEBV for experimenting topic monitoring. We set the length of sliding window to one month and the step to one time unit with 3 days. Then the 3-month dataset is divided into 21 time slots, and we incrementally monitor the topics in each slot. Because topics detected at every time slot are a subset of the topics discovered in the whole dataset, we use the results of user study in Table II as the ground-truth for performance evaluation.

To estimate the α and β parameters of topic energy model in Equation (13), we adopt the same approach as in [6]. We select ten discovered topics from Section VI-B2 for estimation. For each topic, two points (r_1,s_1) and (r_2,s_2) are selected to calculate α_i and β_i. We set $r_1 = s_1 = 0.2$ and $r_2 = s_2 = 0.85$. We average the values of α_i and β_i over ten topics, and set the parameters as: $\alpha = 0.0789$, $\beta = 0.3342$.

Figure 9 shows the results of monitoring old and new topics, while recommending potentially hot topics based on Algorithm 2. As observed in Figure 9(a), there are only few topics being detected in the first eight time slots. This is mainly because there are less videos during these periods as shown in Figure 8(a). For example, there is only one old topic at time unit 6. This topic is missed while another topic is falsely detected, as a result, the precision drops sharply to 0. The number of topics, nevertheless, grows quite significantly since then. The precision of monitoring different types of topics is shown in Figure 9(b). The precision at the initial few time slots fluctuates because there are only few topics discovered during these periods. The precision is maintained at around 0.8 level on average since after the 8th time slot. From our analysis, the topic energy model is effective in excluding topics which are lack of continuing and strong evolution trend. These topics usually decay rapidly, and are eliminated from monitoring once degrading to close zero energy value. Overall, the average precision (AP) of old and new topics are 0.879 and 0.949 respectively, which is very encouraging.

The potentially hot topics are recommended based on the step-7c(iii) of Algorithm 2. To evaluate the performance, a recommended topic is marked as appropriate if the energy of the topic indeed increases in the subsequent time slot. In other words, future evolution trend of a topic is pre-computed for evaluation. As shown in Figure 9(b), the AP of recommended topics is 0.877, which indicates that most of the recommended topics continue evolving in the next time unit.

D. Visualization

To visualize the evolution of topics, a topic browsing system is developed. Figure 10 shows the system interface where a topic is depicted as a trajectory in two-dimensional space as in 10(a). Each event (node) in the topic trajectory is attached with keyframes and bursty tags. By clicking an event, a second-level interface as shown in Figure 10(b) will be popped up. The page basically ranks the list of videos as well as the user-supplied tags under the event. The tags are displayed in different colors and sizes representing the frequency and relevancy of the tags to the events. Videos are ranked according to the social popularity. Generally speaking, the most view video of an event is likely, though not absolutely, to be more representative. Each video is represented with a keyframe and its title. By clicking the keyframe, the video will be displayed as shown in Figure 10(c). From user point of view, the system facilitates efficient browsing of topics by tracing the life cycle and hot degree of topics over different times. Clicking events of interest will further bring users to the group of videos they may wish to browse. From monitoring point of view, the system allows a web monitor to rapidly glimpse the evolution trend of topics. Future trend can possibly be predicted, and actions such as whether to recommend or to stop a topic from further spreading can be made accordingly.

The example trajectories of evolution-hot, content-hot and potential-hot being discovered are shown in figures 10(a), 11 and 12 respectively. In Figure 11, the topic "USA presidential election 2008" became hot because the uploaded videos received many hit counts over times. This kind of video topics captures short-term hot issues and can be recommended to users who are querying "*What's hot now?*". On the other hand, in Figure 12, the topic about the discussion of "Islamic belief" did not keep hot throughout the whole life-span. Instead, the topic has strong evolution trend and has been repeatedly concerned by the public. This kind of topics is generally about sensitive politic issue or super-star, which periodically triggers public concerns. These topics are often welcomed by TV broadcasters who concern about "*What's going on?*". The potential-hot topic shown in Figure 10(a) is about the public show "Resident Evil 5". The topic was concerned by a small group of users initially, and suddenly broke out after the official announcement of the show. Another good example discovered by our approach is about the topic "Makeup tutorial", which started with few girls uploading videos about makeup, and then turned hot when some of the girls became hot web stars for their uploaded videos. This kind of topics can be recommended to web monitors who concern "*What's going to be hot?*".

TABLE III
TOP-10 HOT TOPICS AMONG THE 120 TOPICS DISCOVERED BY OUR SYSTEM IN MCG-WEBV DATASET.

Topic ID	Description
1	US presidential election 2008.
2	A Conflict between Israel.
3	Tutorial of Apple's Iphone.
4	Drama of flower boys.
5	Discussion of Islamic belief.
6	Resident Evil 5.
7	Angeles Lakers.
8	Films of Alex. Jones.
9	Discussion of Game Xbox.
10	American Idol Season 8.

TABLE IV
SUBJECTIVE STUDIES FOR TRAJECTORY-BASED VISUALIZATION.

	Strongly Agree	Agree	Disagree	Average Score
User-friendly?	2	7	1	0.55
Informative?	6	3	1	0.75
Efficient?	8	2	0	0.9

To access the feasibility of trajectory-based visualization based on the developed system as shown in Figure 10, we sample the top-10 topics for evaluation. The topics are listed in Table III. Ten assessors were invited to evaluate the system based on the following criteria:

1) Is the interface user-friendly and attractive?
2) Is the display information-rich?
3) Can you locate the videos of interest efficiently using the interface?

The assessors rate each question with a three-degree score: strongly agree (1.0), agree (0.5) and disagree (0). Table IV summarizes the result of user evaluation. Overall, the result is very encouraging. All assessors agreed that the trajectory-based visualization interface is novel and the presented information is more concise if compared to traditional way of browsing videos in sequential order. Based on the feedbacks from assessors, the interface is especially efficient for them to find the videos of interest. First, the topic trajectory can precisely tell the evolution and important events at different time units. Second, important events can be located effortlessly from the keyframes and short descriptions attached to peaks of trajectories. Some assessors indicated that they indeed preferred browsing evolving topics such as evolution-hot and potential-hot, and expressed their opinion about uncertain issues, rather than the videos posted by officials in content-hot topics. Under the platform of Web 2.0, most videos under evolution-hot and potential-hot are indeed created and uploaded by Internet users rather than by officials.The vivid display of trajectories allows users to quickly predict and locate the types of hot topics which interest them. Nevertheless, some assessors did comment that a gist of events on a trajectory with sparse texts and keyframes is insufficient to fully describe a topic. We will further study this issue as future extension of our browsing system.

VII. CONCLUSION

To meet the massive and dynamic growth of web videos, we have presented a trajectory-based approach to efficiently discover, track, monitor and visualize web video topics. In discovery, we model the set of events as a evolution graph, and salient topics are mined in the form of trajectory. Besides the traditional content-hot topics, the proposed approach is also capable of discovering the evolution-hot and potential-hot topics. In monitoring, we consider the aging of topics as a energy function. Events discovered at different time units can thus be continuously tracked and monitored. Meanwhile, a video browsing system is developed for topic visualization. Topic trajectory is vividly displayed in a 2D space of time and hot degree. With the display, the evolution trend of a topic can be efficiently browsed and traced. The "three-hot" topics as shown in Figure 1 can also be identified.

REFERENCES

[1] http://www.nist.gov/speech/tests/tdt/.

[2] *YouTube Report 2009*, http://youtubereport2009.com/, 2009.

[3] J. Cao, C. W. Ngo, Y. D. Zhang, D. M. Zhang, and L. Ma, "Trajectory-based visualization of web video topics," in *Proceedings of the international conference on Multimedia*, 2010.

[4] J. Cao, T. Xia, J. T. Li, Y. D. Zhang, and S. Tang, "A density-based method for adaptive lda model selection," *Neurocomputing*, vol. 72, no. 7-9, pp. 1775–1781, 2009.

[5] J. Cao, Y. D. Zhang, Y. C. Song, Z. N. Chen, X. Zhang, and J. T. Li, "Mcg-webv: A benchmark dataset for web video analysis," in *Technical Report,ICT-MCG-09-001*, 2009.

[6] C. C. Chen, Y. T. Chen, Y. Sun, and M. C. Chen, "Life cycle modeling of news events using aging theory," in *European Conference on Machine Learning*, 2003.

[7] K. Y. Chen, L. Luesukprasert, and S. T. Chou, "Hot topic extraction based on timeline analysis and multi-dimensional sentence modeling," *IEEE Trans. on Knowledge and Data Engineering*, vol. 19, no. 8, pp. 1016–1025, 2007.

[8] T. C. Chou and M. C. Chen, "Using incremental PLSI for threshold-resilient online event analysis," *IEEE Trans. on Knowl. and Data Eng.*, vol. 20, pp. 289–299, March 2008.

[9] P. Duygulu, J. Y. Pan, and D. A. Forsyth, "Towards auto-documentary: Tracking the evolution of news stories," in *ACM Multimedia*, 2005.

[10] G. Fung, J. X. Yu, P. S. Yu, and H. Liu, "Parameter free bursty events detection in text streams," in *International Conference on Very Large Data Bases*, 2005.

[11] Q. He, K. Chang, and E. P. Lim, "Analyzing feature trajectories for event detection," in *ACM SIGIR Conference*, 2007.

[12] I. Ide, H. Mo, N. Katayama, and S. Satoh, "Topic threading for structuring a large-scale news video archive," in *ACM Conference on Image and Video Retrieval*, 2004.

[13] I. Ide, T. Kinoshita, T. Takahashi, S. Satoh, and H. Murase, "mediawalker: a video archive explorer based on time-series semantic structure," in *ACM Multimedia*, 2007.

[14] G. Karypis, *CLUTO a Clustering Toolkit*. http://www.cs.umn.edu cluto: Technical Report 02-017, Dept. of Computer Science, University of Minnesota, 2002.

[15] L. Liu, L. Sun, Y. Rui, Y. Shi, and S. Yang, "Web video topic discovery and tracking via bipartite graph reinforcement model," in *International World Wide Web Conference*, 2008.

[16] Q. Mei and C. Zhai, "Discovering evolutionary theme patterns from text: an exploration of temporal text mining," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2005.

[17] A. Moffat and J. Zobel, "Self-indexing inverted files for fast text retrieval," *ACM Trans. Inf. Syst*, vol. 14, no. 4, pp. 349–379, Oct. 1996.

[18] S. Y. Neo, T. Ran, and et. al., "The use of topic evaluation to help users browse and find answer in news video corpus," in *ACM Multimedia*, 2007.

[19] H. K. Tan, C. W. Ngo, and T. S. Chua, "Efficient mining of multiple partial near-duplicate alignments by temporal network," *IEEE Trans. on Circuits and Systems for Video Technology*, 2010.

[20] C. H. Wang, M. Zhang, S. P. Ma, and L. Y. Ru, "Automatic online news issue construction in web environment," in *International World Wide Web Conference*, 2008.

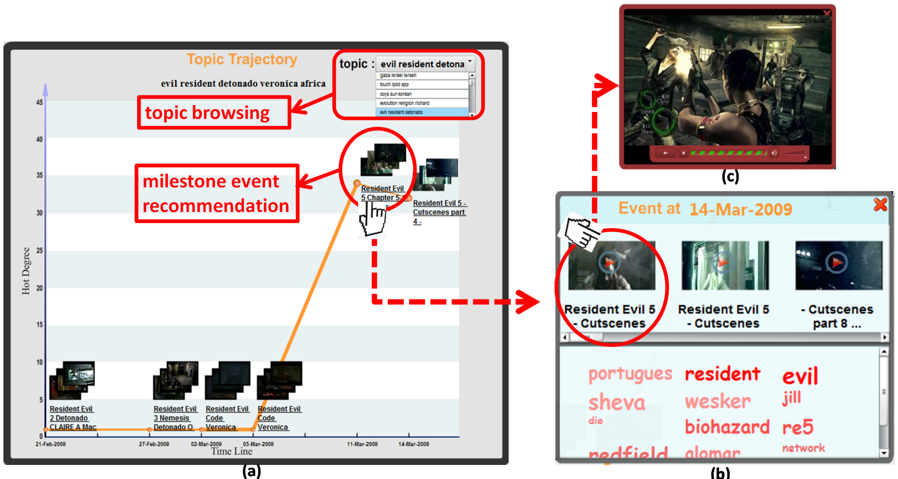


Fig. 10. An example of browsing the web video topic "Resident Evil 5" using our interface. (a) Topic-level interface displays a trajectory showing the topic evolution in a 2D space of time and hot degree. The trajectory gives a glimpse of events happened at different time units. (b) Event-level interface provides a localized view of individual events with videos and tags. Videos are ranked according to popularity based on their view counts, while tags are displayed in different colors and sizes signifying their frequency and relevancy respectively. (c) Video-level interface plays the videos selected by users.

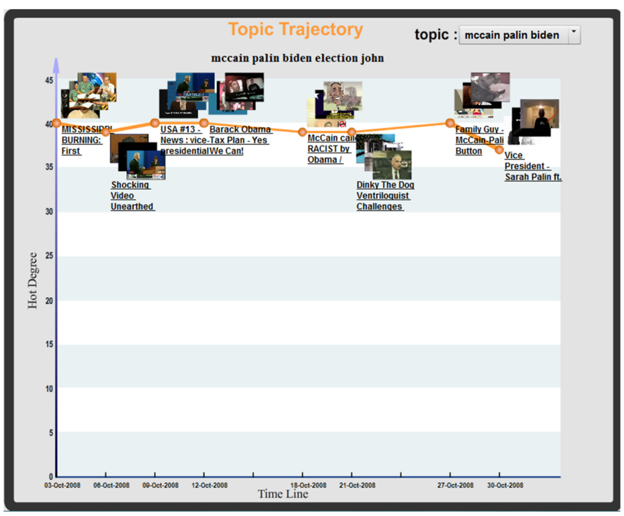


Fig. 11. A content-hot topic "US presidential election 2008".

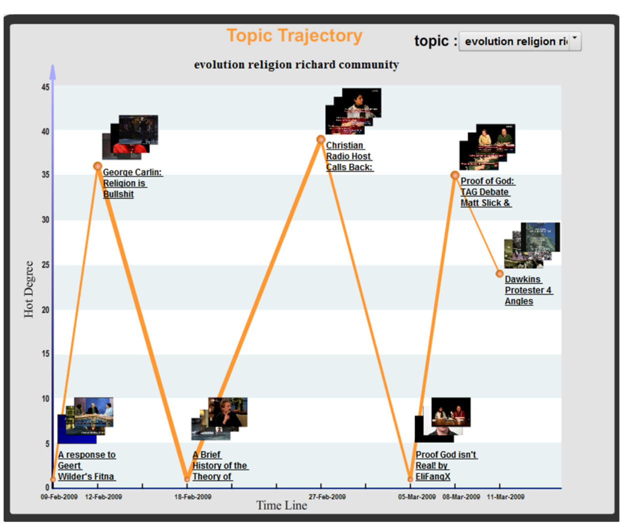


Fig. 12. Trajectory visualization of the evolution-hot topic "Islamic belief".

[21] C. H. Wang, M. Zhang, L. Y. Ru, and S. P. Ma, "Automatic online news topic ranking using media focus and user attention based on aging theory," in *ACM Conference on Information and Knowledge Management*, 2008.
[22] X. Wu, A. G. Hauptmann, and C. W. Ngo, "Novelty detection for cross-lingual news story with visual duplicates and speech transcripts," in *ACM Multimedia*, 2007.
[23] X. Wu, A. G. Hauptmann, and C. W. Ngo, "Practical elimination of near-duplicates from web video search," in *ACM Multimedia*, 2007.
[24] Y. Zhai and M. Shah, "Tracking news stories across different sources," in *ACM Multimedia*, 2005.

Juan Cao is an Associate Professor of the Institute of Computing Technology(ICT), Chinese Academy of Science(CAS). Her research interests focus on large web video process and multimedia retrieval. She received her Ph.D. degree from ICT-CAS in 2008. She was a Senior Research Associate of City University of Hong Kong in 2009, and an visiting scholar of Columbia University in 2010.

Chong-Wah Ngo (M02) received his Ph.D in Computer Science from the Hong Kong University of Science & Technology in 2000. He received his MSc and BSc, both in Computer Engineering, from Nanyang Technological University of Singapore. Before joining City University of Hong Kong in 2002, he was with Beckman Institute of University of Illinois in Urbana-Champaign. He was also a visiting researcher of Microsoft Research Asia in 2002. His research interests include video computing and multimedia information retrieval.

YongDong Zhang received his Ph.D. from Tianjin University in 2002, Tianjin, China. Now he is a Professor of the Institute of Computing Technology(ICT), Chinese Academy of Science(CAS). His major field includes video coding and incoding, multimedia information retrieval.

JinTao Li is a professor and Ph.D. Supervisor of the Institute of Computing Technology(ICT), Chinese Academy of Science(CAS). His major field includes multimedia processing and virtual reality technology.