*IOSR Journal of Computer Engineering (IOSR-JCE)*
*ISSN : 2278-0661, ISBN : 2278-8727, PP : 64-69*
*www.iosrjournals.org*

# Web Video Discovery, Visualization and Monitoring

Mr.V.C.Patil[1], Miss.B.K.Ugale[2]
[1](Computer Engg,JSPM ,Pune University,India)
[2](Computer Engg,ICOER ,Pune University,India)

**ABSTRACT:** *Now a day's the massively growth of web-shared videos in Internet (such as Face book , YouTube users), efficient organization and monitoring of videos remains a practical challenge. Because these days the cyber crime security is mostly important issue over Internet .While now a days broadcasting channels and social sites are keen to monitor online events, identifying topics of interest from very big volume of user uploaded videos, pictures, images and giving recommendation to emerging topics are by no means easy , such process involves discovering of new topic, visualization of the topic content and incremental monitoring of topic evolution. The studies of web shared videos 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 videos crawled from YouTube or Face book during three months. Both objective and user studies are conducted to verify the performance.*

**Keywords -** *Topic trajectory, Video recommendation, Visualization.*

## I. INTRODUCTION TO DATA MINING

The analysis step of the "Knowledge Discovery in Databases" process, or KDD), a field at the intersection of computer science and statistics, is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data preprocessing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. The term is a buzzword, and is frequently misused to mean any form of large-scale data or information processing (collection, extraction, warehousing, analysis, and statistics) but is also generalized to any kind of computer decision support system, including artificial intelligence, machine learning, and business intelligence. In the proper use of the word, the key term is discovery, commonly defined as "detecting something new". Even the popular book "Data mining: Practical machine learning tools and techniques with Java"(which covers mostly machine learning material) was originally to be named just "Practical machine learning", and the term "data mining" was only added for marketing reasons.[6] Often the more general terms "(large scale) data analysis", or "analytics" – or when referring to actual methods, artificial intelligence and machine learning – are more appropriate. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). This usually involves using database techniques such as spatial indexes. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting are part of the data mining step, but do belong to the overall KDD process as additional steps. The related terms data dredging, data fishing, and data snooping refer to the use of data mining methods to sample parts of a larger population data set that are (or may be) too small for reliable statistical inferences to be made about the validity of any patterns discovered. These methods can, however, be used in creating new hypotheses to test against the larger data populations.

### *A. Overview of Data Mining*

The development of Information Technology has generated large amount of databases and huge data in

various areas. The research in databases and information technology has given rise to an approach to store and

manipulate this precious data for further decision making. Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis.

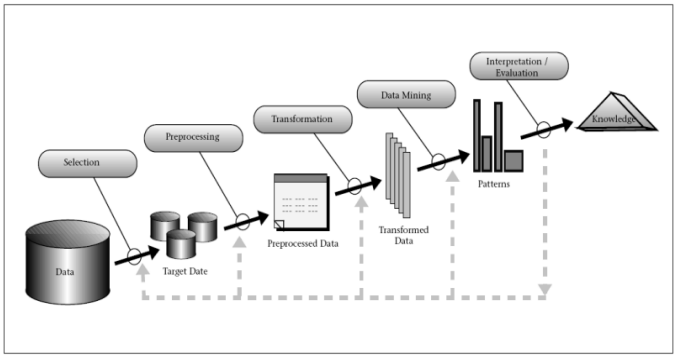


Fig. 1 Knowledge discovery Process

Data mining is a logical process that is used to search through large amount of data in order to find useful data. The goal of this technique is to find patterns that were previously unknown. Once these patterns are found they can further be used to make certain decisions for development of their businesses.

Three steps involved are

1)**Exploration:** In the first step of data exploration data is cleaned and transformed into another form, and important variables and then nature of data based on the problem are determined.

2) **Pattern Identification:** Once data is explored, refined and defined for the specific variables the second step is to form pattern identification. Identify and choose the patterns which make the best prediction.

3) **Deployment:** Patterns are deployed for desired outcome.

### *B. Multimedia Data Mining*

"What is a multimedia database?" A multimedia database system stores and manages a large collection of multimedia data, such as audio, video, image, graphics, speech, text, document, and hypertext data, which contain text, text markups, and linkages. Multimedia database systems are increasingly common owing to the popular use of audio video equipment, digital cameras, CD-ROMs, and the Internet. Typical multimedia database systems include NASA's EOS (Earth Observation System), various kinds of image and audio-video databases, and Internet databases. Our study of multimedia data mining focuses on image data mining. Mining text data and mining the World Wide Web are studied in the two subsequent sections. Here we introduce multimedia data mining methods, including similarity search in multimedia data, multidimensional analysis, classification and prediction analysis and mining associations in multimedia data.

## II. TRACKING WEB VIDEO

These days 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 statistic by YouTube 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.

### *A. Topic Detection and Tracking*

In the research community, topic detection and tracking (TDT) 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 while ignoring the evolution trend of topics, which is often a matter of deep concern when performing web video monitoring. Three types of hot topics often observed in web videos: content hot, evolution- hot, and potential-hot. The examples of topics include "U.S. 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 contrary, 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.

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. However, considering the massive and dynamic growth of video data in Internet, clustering will be time consuming and furthermore, continuous monitoring and recommendation are more demanding.

For 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 Fig. 2. The representation allows efficient discovery of topic trajectories, visualization of evolution trends, and monitoring of new, old, and potentially hot topics.

***1.*** 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.

**2.** Visualization:

A topic is presented as a trajectory in 2-D 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.

**3.** Monitoring and Recommendation:

We employ aging theory to depict the evolution of a topic as the change of energies. Three different types of topics, old, new, and 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 signalling bursty events, and recommending potential topics that are likely to exhibit strong evolution trend.

***B. Video Recommendation System***

Personalized recommendations are a key method for information retrieval and content discovery in today's information-rich environment. Combined with pure search(querying) and browsing (directed or non-directed), they allow users facing a huge amount of information to navigate that information in an efficient and satisfying way. As the largest and most-popular online video community with vast amounts of user-generated content, YouTube presents some unique opportunities and challenges for content discovery and recommendations.

Founded in February 2005, YouTube has quickly grown to be the world's most popular video site. Users come to YouTube to discover, watch and share originally-created videos. YouTube provides a forum for people to engage with video content across the globe and acts as a distribution platform for content creators. Every day, over a billion video plays are done across millions of videos by millions of users, and every minute, users upload more than 24 hours of video to YouTube. In this paper, we present our video recommendation system, which delivers personalized sets of videos to signed in users based on their previous activity on the YouTube site (while recommendations are also available in a limited form to signed out users, we focus on signed in users for the remainder of this paper). Recommendations are featured in two primary locations: The YouTube home page (http://www.youtube.com) and the "Browse" page at http://www.youtube.com/videos.

1. Goals

Users come to YouTube for a wide variety of reasons which span a spectrum from more to less specific:

To watch a single video that they found elsewhere (direct navigation), to find specific videos around a topic (search and goal-oriented browse), or to just be entertained by content that they find interesting. Personalized Video Recommendations are one way to address this last use case, which we dub unarticulated want. As such, the goal of the system is to provide personalized recommendations that help users find high quality videos relevant to their interests. In order to keep users entertained and engaged, it is imperative that these recommendations are updated regularly and reflect a user's recent activity on the site. They are also meant to highlight the broad spectrum of content that is available on the site. In its present form, our recommendation system is a top-N recommender rather than a predictor We review how we evaluate the success of the recommendation system in section 3 of this paper. An additional primary goal for YouTube recommendations is to maintain user privacy and provide explicit control over personalized user data that our backend systems expose. We review how we address this goal.

2. Challenges

There are many aspects of the YouTube site that make recommending interesting and personally relevant videos to users a unique challenge: Videos as they are uploaded by users often have no or very poor metadata. The video corpus size is roughly on the same order of magnitude as the number of active users. Furthermore, videos on YouTube are mostly short form (under 10 minutes in length). User interactions are thus relatively short and noisy. Compare this to user interactions with movie rental or purchase sites such as Netflix or Amazon where renting a movie or purchasing an item are very clear declarations of intent. In addition, many of the interesting videos on YouTube have a short life cycle going from upload to viral in the order of days requiring constant freshness of recommendation

## III. FRAMEWORK TRACRKING WEB VIDEO

We first define the following two terminologies used in this paper. *Event E* is a group of related videos conveying a story and discovered at a time unit. *Topic* T is a group of topic related events found over time. The system composes of the following three steps.

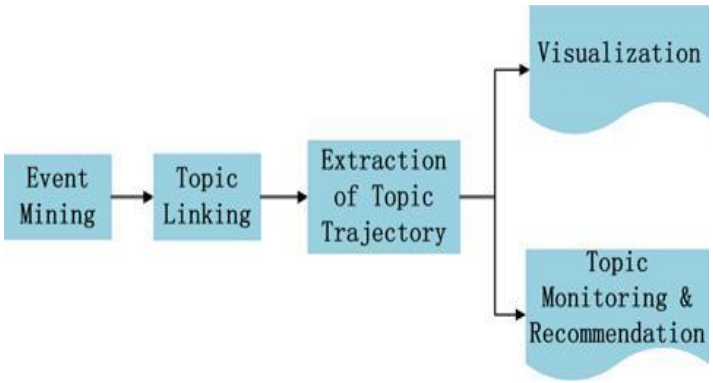


Fig. 2 Overview of proposed framework.

### *A. Topical Linking*

A hot topic wills evolutes from one event to another with time. So we measure the similarities between events in different time units based on tags and visual near-duplicates. Events with more common tags and near-duplicate segments receive higher weights. Then a topic evolution graph **G** = (V*;* E) is generated, where each node is an event and the edge between nodes signifies the similarity between two events.

### *B. Trajectory Extraction*

In this step, the topic discovery has been transferred into path selection from the above topic evolution graph. Firstly, based on the context of graph, we can optimize the graph by adding the missing edges and remove the isolate weak links. Then, we decompose the huge topic graph into sub graphs with conventional depth first search algorithm, where the closely linked sub graph generally represents a topic or several related topics. Among the graph, every path is a candidate topic evolution trajectory with an order set of events which are linked chronologically, denoted as T =< *E*1*; E2; : : : ; Et-1; Et* > where events are placed in time order. Then a saliency measurement is proposed to extract meaningful topic trajectories from these sub graphs. The saliency of T is as

Saliency (T)=∑^(E1)+∑sim(Et-1,Et) (1)

The first term measures the social popularity of a topic based on the view count of events. The second term measures the topic compactness and evolution trend based on the event similarity. 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.

*C. Topic Visualization*

Each extracted trajectory can be vividly viewed in a 2D space, where the time axis indicates duration of a topic, and the hot-degree axis measures the popularity of events based on video view counts. With reference to Figure 2, the interface supports different levels of topical browsing.

1. Topic level browsing

Includes a scroll box which summarizes the list of detected topics ranked with their trajectory saliency in a video collection. Clicking a topic will show the corresponding trajectory in a 2D space of time and hot degree. Events are distributed along the trajectory and tagged with key frames and a short description of texts, allowing users to rapidly trace the event sequence. The trajectory-based visualization gives a glance of the whole topic evaluation, while signifies the importance of events at different time points. For example, in the topic "Resident Evil 5" initially keeps a low profile, and reaches a peak on 11-March-2009 for an official announcement that this game would be released on 13-March-2009. Users can easily locate the events of interest with the displayed trajectory, or conveniently track backward and forward to see surrounding events.

2. Event level browsing

Supports the efficient means of visualizing tags and videos as shown in Figure 3. The tags are displayed in different colors and sizes representing the frequency and relevancy of the tags to the events. Tag relevancy is determined based on the number of videos with this tag in the corresponding event. By catching the attention with large and deep red tags, the display allows users to efficiently grasp the representative or key tags about the main content of an event. In addition to tags, 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 key frame and its title.

*D. Topic Recommendation*

Topic trajectory display the birth, growth, decay and death of a topic along the timeline. By employing simple classification scheme based on primitive features such as number of peaks, degree of hotness, event compactness (similarity) and topic duration, the trajectories can be broadly categorized into three-hots as following.

**Content-hot** includes topic trajectories which keeps a high-level of hot degree for a certain period of time. This type of topics is concerned by most web users, and the content is often surrounding about a center theme with different peripheral events.

**Evolution-hot** typically includes topics with strong evolution trend, where the trajectories exhibit up-and-down trend over different periods of time. The contents of these topics are usually related to some sensational and sensitive news or discussion in the Internet.

**Potential-hot** includes topics initially concerned by a small group of web users, but increasingly capture the public attention and eventually end up with an erupt trend. This kind of topics is typically very focused andnarrowed in the scope of discussion. Web monitors are especially interested in predicting the tendency of these topics. In addition to automatic classification of three-hot topics, the developed interface indeed also offers efficient means of monitoring and recommending topics. Figure 2 shows a content-hot topic"US presidential election 2008". This topic becomes hot because the uploaded videos received many view counts over times. This kind of video topics captures short-term hot issues and could be recommended to users who are querying "*What's hot now*?". Figure 3 shows a evolution-hot topic about "Islamic belief". The topic did not keep hot throughout the whole life span. Instead, it has strong evolution trend and was repeatedly concerned by the public. This kind of topics is generally about sensitive political issues or super-stars, which periodically trigger public concerns. These topics are often welcomed by TV broadcasters who care about "*What's going on?*". The trajectory shown in Figure 3 is indeed an example of potential-hot topic

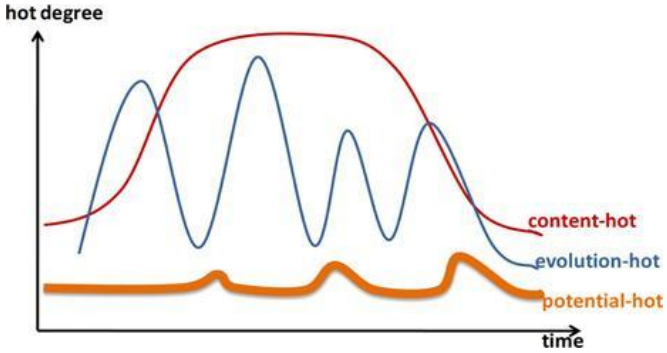


Fig. 3. Trajectories for different types of hot topics.

The width of trajectory indicates the strength of correlation among the events of a topic

## IV. CONCLUSION

The massive and dynamic growth of web videos, have presented a trajectory-based approach to efficiently discover, track, monitor, and visualize web video topics. In discovery, 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, 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 generally displayed in a 2-D space of time and hot degree. With the display, the evolution trend of a topic can be efficiently browsed and traced.

## REFERENCES

[1] . *J. Cao, C. W. Ngo, Y. D. Zhang, D. M. Zhang, and L Ma, "Trajectory based visualization of web video topics," in Proc. Int. Conf. Multimedia,2010,*

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

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

[4]. *Q. Mei and C. Zhai, "Discovering evolutionary theme patterns from text: An exploration of temporal text mining," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining, 2005, pp. 198–207.*