An Effective Multi-Clue Fusion Approach for Web Video Topic Detection

Tianlong Chen[1], Chunxi Liu[2], Qingming Huang[1,2]

[1] Key Lab of Intell. Info. Process., Inst. of Comput. Tech., CAS, Beijing, 100190, China

[2] Graduate University of Chinese Academy of Sciences, Beijing, 100049, China
{tlchen, cxliu, qmhuang}@jdl.ac.cn

ABSTRACT

The efficient organization and navigation of web videos in the topic level could enhance the user experience and boost the user's understanding about the happened events. Due to the potential application prospects, topic detection attracts increasing research interests in the last decade. On one hand, the user concerned real world hot topic always leads to a massive discussion in the video sharing sites, such as YouTube, Youku, *etc*. On the other hand, the search volume of the topic related keywords are growing explosively in the search engine such as Google, Yahoo, *etc.* These keywords are the queries formulated by the users to search their concerned topics. They reflect the users' intention and could be used as a clue to find the hot topics. In this paper, different from the traditional topic detection methods, which mainly rely on data clustering, we propose a novel multi-clue fusion approach for web video topic detection. In our approach, firstly by utilizing the video related tag information, a maximum average score and a burstiness degree are proposed to extract the dense-bursty tag groups. Secondly, the near-duplicate keyframes (*NDK*) are extracted from the videos and fused with the extracted tag groups. After that, the hot search keywords from the search engine are used as guidance for topic detection. Finally, these clues are combined together to detect the topics hidden in the web video data. Experiment is conducted on the YouTube video data and the results demonstrate that the proposed method is effective.

Categories and Subject Descriptors

H.3.1 [Information Storage and Retrieval]: *Content Analysis and Indexing - Abstracting methods.*

General Terms

Algorithms, Design, Experimentation.

Keywords

Topic detection, multi-clues fusion, tag group

1. INTRODUCTION

With the rapid development of the Internet and multimedia technologies, the video data are growing explosively. In the last decade much effort has been devoted to make these videos accessed more easily by the users. Previous research in the web video domain mainly focused on improving video retrieval accuracy or near-duplicate detection [4, 7, 14]. However, without effective abstracting technologies when facing the huge amount of video data, it is hard and time-consuming for the users to obtain their concerned hot topics information. Integrating these videos into topics provides the users with an easy way to understand the happened real world events, and therefore could enhance the user experience. Topic detection is such a technology to summarize information from the unstructured web video data.

The objective of topic detection is to discover new or previously unidentified event which refers to some unique thing that happens at a specific time and place [9]. Topic detection rises from detecting topics in news articles or blog posts [2, 3, 6]. Chen *et al.* [6] extract hot terms from the text by combining *TF*PDF* and the Aging Theory. Then, based on the extracted hot terms, key sentences are identified and grouped into clusters. Each cluster is deemed as a topic. The method in [3] extracts the documents that are highly related to the bursty features based on timestamp. The extracted documents are then segmented to construct the event hierarchy. Sun *et al.* [2] integrate event-related queries, news articles, and blog posts through the notion of query profile, and group the query profiles into event fragments.

Although many methods have been proposed for text topic detection, they are not suitable for web video topic detection. In order to exact hot topics from the web videos efficiently, some research effort has been contributed to it. The web videos not only contain abundant textual information (title, tag, *etc*), but also consist of rich visual content. Cao *et al* [10] propose an algorithm based on salient trajectory extraction from a topic evolution link graph for topic discovery. Hong *et al.* [13] summarize the content of videos by analyzing the tags associated with key-shots which are established and ranked based on near-duplicate keyframe detection. Shao *et al.* [12] propose a Star-structured K-partite Graph based co-clustering and ranking framework for web video topic discovery and visualization. These video topic detection methods are mostly based on clustering the tags or keyframes. However, direct clustering result may not be very effective due to the deficiency of text information (noisy, incomplete and inconsistent) and low accuracy of visual content analysis. Moreover, due to the lack of efficient clustering guidance information, how to decide the right cluster number is an insurmountable problem for the unsupervised clustering topic detection approach. The users' queries recorded in the search engine could be used as such a clue to guide us to detect topics. In order to get more information about their concerned topic, the users usually formulate the topics into queries and search them through the search engine. As the users' queries reflect users' intention and hot topics always attract high attention of users, the hot topic related keywords will be reflected by the users' queries recorded in the search engine. For example, the topic "*Bush was attacked by shoes during press conference in Iraq*" reflects by some hot search queries, like "*bush*", "*shoes*", "*Iraq*", and so on.

MM'12, October 29–November 2, 2012, Nara, Japan.
Copyright 2012 ACM 978-1-4503-1089-5/12/10...$15.00.

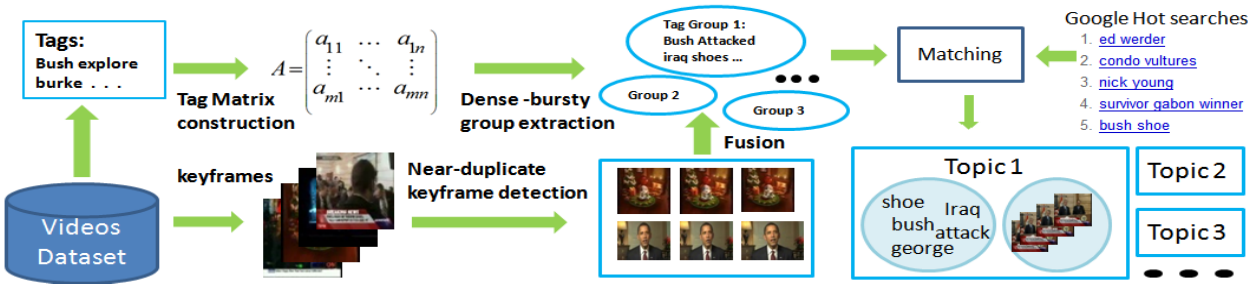

Figure 1. Illustration of the flowchart of our approach

In summary, hot search queries provide us additional information for topic detection.

In this paper, we propose a novel multi-clue fusion approach to detect hot topics from the web videos, which is shown in Figure 1. First, by utilizing the video related tag information, a maximum average score and burstiness degree is proposed to find dense-bursty tag groups from the tag similarity matrix. The similarity between two tags is measured by their co-occurrence number and temporal trajectories. Then, near-duplicate keyframes (*NDK*) are extracted from the web videos. The *NDKs* are further fused with extracted tag groups. The fused tag groups are deemed as the candidate topics. Finally, the tag groups are matched with hot keywords obtained from the search engine to find the hot topics. Compared with the existing web video topic detection methods, the main contributions of the proposed approach are summarized as follows:

1. A novel topic detection method is proposed based on tag groups. Unlike previous clustering methods, the dense-bursty tag groups could overcome the noisy and incomplete problem of the tags. Several bursty and tightly interlinked tags can provide a better description of the topics.
2. The fusion of near duplicate key frames with dense-bursty tag groups further refine the tag group extraction result and makes the topic detection result more accurate.
3. The utilization of the users' intention is presented as the guidance for topic detection. The hot search queries in the search engine reflect the users' intention and could be used as a guidance to find the hot topics.

2. TAG MINING

In this section, the dense-bursty tag group extraction method is described in detail. First, the relationship between the tags is analyzed by tag co-occurrence and tag trajectory. Then, a maximum average score criteria is proposed to extract the dense tag groups which are further filtered by the burstiness degree.

2.1 Tag Similarity Matrix Construction

Co-occurrence information is particularly helpful for mining the relationship between the tags. For example "*boxing*" and "*bush*" are unlikely to appear in the same video, while the tag "*shoe*" and "*bush*" co-occur a lot due to the event "*Bush was attacked by shoes*". An event always leads to several co-occurred tags which could be used as the descriptive information of the event.

Given two tags f_1 and f_2, the co-occurrence of them is defined as:

$$C(f_1,f_2)=\frac{co\_occur(f_1,f_2)}{occur(f_1)+occur(f_2)-co\_occur(f_1,f_2)} \qquad (1)$$

where *occur(f)* represents the number of videos tagged by *f*, $co\_occur(f_1, f_2)$ is the total number of videos containing both f_1 and f_2. The maximum value of $C(f_1,f_2)$ is 1 which represents f_1 and f_2 always appear together and the minimum value is 0 which means these two tags never co-occur.

Besides the co-occurrence similarity of the two tags, we measure the temporal similarity of them. We model tags as trajectories in a two dimensional space, with one dimension as time and the other as feature weight:

$$Y_f=[y_f(1),y_f(2),...y_f(T)] \qquad (2)$$

where each element $y_f(t_i)$ is the weight of tag *f* at time unit t_i,.$y_f(t_i)$ is defined as the normalized *DF/IDF* score:

$$y_f(t_i)=\frac{DF_f(t_i)}{N(t_i)}\times\log(\frac{N}{DF_f}) \qquad (3)$$

where $DF_f(t_i)$ represents the number of videos containing the tag *f* at time unit t_i, DF_f is the total number of videos containing tag f over the time duration *T*, $N(t_i)$ represents the number of videos at time unit t_i, and *N* is the total number of videos over *T*.

We adopt the histogram intersection to measure the temporal similarity of two tags:

$$T(Y_{f_1},Y_{f_2})=\frac{\sum_i \min(y_{f_1}(t_i),y_{f_2}(t_i))}{\sum_i \max(y_{f_1}(t_i),y_{f_2}(t_i))} \qquad (4)$$

Then the tag similarity matrix A is as follows:

$$A(i,j)=\begin{cases}0, f_i=f_j \\ w_{ij}=\log(\theta\times C(f_i,f_j)+(1-\theta)\times T(f_i,f_j)), f_i\neq f_j\end{cases} \qquad (5)$$

where θ is a tradeoff parameter with value ranging from 0 to 1. This definition fuses the symbiotic relationship and temporal similarity of the tags. It is an effective method to describe the dependence of two tags.

2.2 Dense-Bursty Tag Group Extraction

The dense-bursty tag group has two properties: denseness and burstiness. Denseness means the tags in the group closely interlink with each other. Burstiness refers to the number of videos containing those tags growing explosively in short time duration. Based on the above observations, we propose a maximum average score to extract dense tag groups and then filter the result by bursty degree. Algorithm 1 outlines the proposed method. First we select two tags with the largest weight and add them to the group (lines 1-3). Then the tag with the maximum average weight is added into the group and repeats the steps (lines 3-6). When a dense tag group is extracted, we further filter it by the burstiness degree which is calculated as follows. Firstly we gather videos containing no less than three common tags with the tag group; then we count the number of the videos at each time slot to form a histogram; finally we calculate the variance of the histogram as burstiness degree, and abandon the tag groups with small burstiness degree.

Before adding a new tag into the group, we ensure that the tag has the maximum average weight linked with the tags already contained in the group. Therefore the result group keeps dense. The burstiness degree reflects the distribution of the tag group. The uneven distributions are more likely generated by the bursty tag groups which are accordance with the prosperity of the event.

Algorithm 1. Dense-bursty Tag Groups Extraction

```
Input:
    $A$ – Tag similarity matrix
    $F = ( f_1 , f_2 \dots f_M )$ – tag dictionary. $M$ is the total number of tags
Output:
    $G = ( g_1 , g_2 \dots)$ – the set of dense-bursty tag group
while ( $W_{\max} = \max(w_{ij})$ & $W_{\max} > \partial$ ) do
    add $f_i$ and $f_j$ to $g*$
    while ( $avg_{\max} = \max_{k \in T} \frac{\sum_{t \in g*} w_{tk}(t)}{|g*|}$ & $avg_{\max} > \beta$ ) do
        add $f_k$ to $g*$
        $w_{tk} = 0$ Where $f_t \in g*$
    end while
    if Burstiness( $g*$ ) > $\eta$ do add $g*$ to $G$
    end if
end while
Return $G$
```

3. FUSING TAG GROUP WITH NDK

Although web videos contain textual information, they are noisy, incomplete and inconsistent. Besides textual information, Near-Duplicate Keyframes (*NDK*) are commonly used in web videos to discuss the same event. In this section, we convert the *NDK* into tag groups, and then merge them with the former dense tag group extraction result. The *NDK* conversion strategy is described as:

$$Tg(NDK_P) = \{ f \mid size[\{V \mid f \in V, V \in NDK_P\}] >= 3\} \quad (6)$$

where *Tg* refers to tag group, *f* is a tag and *V* is a video. We first adopt the method in [8] to extra *NDKs*. Then for each *NDK*, we gather the videos containing the *NDK*. For each tag in the video set, we count the number of videos which contain that tag. The tags whose number is not less than 3 times are selected as the tag group of the *NDK* empirically. Finally, we merge the tag groups with the result in section 2. Two tag groups will be merged if they contain more than two common tags. The tag groups in section 2 are extracted from video related textual information, while the tag groups extracted from *NDKs* reflect the visual content linking of the videos. The combination will overcome the imperfection of these two types of information.

4. QUERY-GUIDED EVENT DETECTION

The hot search queries recorded in the search engine reflect the users' concerns. As indicated by Sun [1], though event-related hot searches increase rapidly when an event occurs, it does not signify that all hot search queries are event-related. First of all, the website names account for a large part of the hot search queries, like *Google*, *Yahoo*. Second, different hot search queries may refer to the same event. For instance, the event "*Bush was attacked by shoes*" leads to multiple hot searches: "*bush*", "*shoes*", "*Iraq*", and so on. Third, person name takes a litter part of the hot search queries.

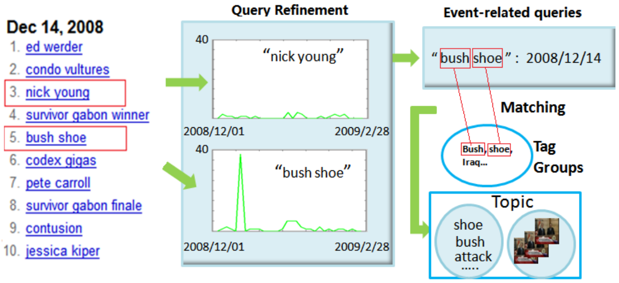

Figure 2. An example of matching strategy

By the above observation, we need to refine the queries. If a hot query is event-related, lots of videos tagged with the query should appear during that query occurrence period. In query refinement step, the query trajectory is calculated. The horizontal axis is the time line while the vertical axis is the frequency of the videos containing the query. If the number of videos during the occurrence period of the query is more than a threshold, we consider it as a candidate event related query. As shown in Figure 2, there are lots of video related to "*bush shoe*" on Dec. 14, while little videos are related to "*nick young*". Therefore "*bush shoe*" is more likely to be an event related hot search query. After the event related hot search queries are extracted, we match them with the tag groups. The matching strategy is to find the overlap between the hot search queries and the tag groups. Finally, a video with no less than three common tags with a tag group is assigned to that group. The tag group and the corresponding videos form the final topic.

5. EXPERIMENTS

To demonstrate the effectiveness of the proposed method, we conduct experiments in the MCG-WEBV dataset [11]. MCG-WEBV consists of 80,031 web videos from December 2008 to February 2009 on YouTube. The whole dataset starts from a "core dataset" containing 3,282 videos which are downloaded from the "Most viewed" videos of "This month". We conduct our experiments on the "core dataset" which consists of a total of 73 ground-truth topics being manually annotated. Besides the video dataset, the hot search queries are obtained from *Google trends* [5]. We download the hot search queries of each day from December 2008 to February 2009, which is the corresponding period of the MCG-WEBV dataset.

We adopt *Precision*, *Recall* and *F-Measure* to evaluate the performance of the proposed method, which is defined as:

$$\Pr ecision = \frac{|E^*|}{|E_D|} \qquad \text{Re}\, call = \frac{|E^*|}{|E_T|} \quad (7)$$

$$F = \frac{2 \times \Pr ecision \times \text{Re}\, call}{\Pr ecison + \text{Re}\, call} \quad (8)$$

where E_T is the video set of ground-truth topic and E_D is the detected topic. $E*$ is the correctly detected videos in E_D.

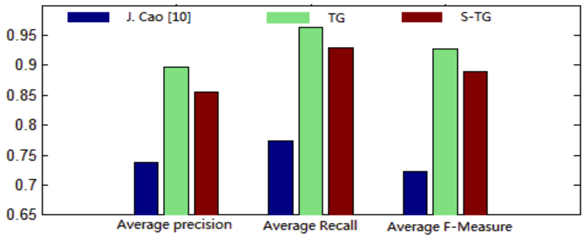

Figure 3. Result of method comparison

For each detected topic, we find the best-match topic in the ground-truth and sort all detected topics by the F-measure. The average precision, recall and F-measure of the top-10 topics are calculated for evaluation which is the same as in [10]. We test the performance of the proposed method by matching tag groups and hot search queries (S-TG). Besides that, in order to prove the effectiveness of hot search queries, we test the performance of the method using only tag group without queries (TG). We fix the parameters to θ =0.2, β =-0.25 and η = 0.43 empirically. The method proposed in [10] is treated as the baseline which is based on salient trajectory extraction from a topic evolution link graph. A total of 50 topics are detected by Cao's method, 83 topics by TG and 53 topics by S-TG. The comparison result is shown in Figure 3. From the figure we can see that the performance of TG and S-TG method are both much better than Cao's method. It may be due to the reason that Cao's method is based on clustering of tags at different time slots to detect events and topics. The same topic related tags may be split to the adjacent time slots which increase the error rate. However, this problem is solved in our method by the accurate tag similarity definition and the global strategy to find tag groups.

In addition, we find that the performance of TG seems better than S-TG. A further experiment is conducted to compare these two methods, and the result is shown in Table 1. In the table, NDT refers to the number of detected topics. If the F-measure of a detected topic is more than 0.5, we recognize it as correctly detected topic. NCDT represents the number of correctly detected topics. CP is overall correct percentage. From Table 1, we can see that the CP of TG is 0.373 while S-TG is 0.453, which demonstrates S-TG is more effective than TG. The results of S-TG contain less correctly detected topics but at the same time contain less falsely detected topics. This is because until now only top-20 hot search queries could be obtained everyday from *Google*, and not all events could be included in this list. At the same time our experiment shows that hot search queries as an informative clue for topic detection could filter most false topics.

Table 1. The comparison result of TG and S-TG

Method	NDT	NCDT	CP
TG	83	31	0.373
S-TG	53	24	0.453

6. CONCLUSION

In this paper, we present an effective multi-clue fusion approach for web video topic detection. First of all, dense-bursty tag groups are extracted by maximum average score and burstiness degree from the tag similarity matrix. The tag co-occurrence number and temporal trajectories are integrated to establish the matrix. Then, the visual near duplicate keyframe information is fused into tag groups. Furthermore, hot search queries are used as additional information to guide the video clustering. Finally, these clues are combined together to detect the hot topics. The experiment results demonstrate that hot search queries are an effective guidance for topic detection and the proposed topic detection method is effective. In future work, we will try to develop more effective multi-clue fusion strategy for web video topic detection.

7. ACKNOWLEDGMENTS

This work was supported in part by National Basic Research Program of China (973 Program): 2012CB316400, in part by National Natural Science Foundation of China: 61025011, 60833006 and 61070108.

8. REFERENCES

[1] A. X. Sun, M. S. Hu, and E. P. Lim, Searching blogs and news: a Study on popular queries. *In ACM SIGIR,* 2008.

[2] A. X. Sun, and M. S. Hu. Query-guided event detection from news and blog streams. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans,* 41(5): 834-839, 2011.

[3] G. P. C. Fung, J. X. Yu, H. Liu, and P. S. Yu, Time-dependent event hierarchy construction. *In KDD*, 2007.

[4] H. T. Shen, J. Shao, Z. Huang, and X. F. Zhou. Effective and efficient query processing for video subsequence identification. *IEEE Trans. Knowl. Data Eng.* 21(3): 321-334, 2009.

[5] http://www.google.com/trends

[6] K. Y. Chen, L. Luesukprasert, and S. T. Chou. Hot topic extraction based on timeline analysis and multi-dimensional sentence modeling. *IEEE Trans. Knowl. Data Eng.* 19(8): 1016-1025, 2007.

[7] L. F. Shang, L. J. Y. F. Wang, K. P. Chan, and X. S. Hua. Real-time large scale near-duplicate web video retrieval. *In ACM Multimedia,* 2010.

[8] L. X. Xie, A. Natsev, J. R. Kender, M. Hill, J. R. Smith. Visual memes in social media. *In ACM Multimedia,* 2011.

[9] J. Allan, J. G. Carbonell, G. Doddington, J. Yamron and Y. Yang. Topic detection and tracking pilot study: Final report. *In DARPA Broadcast News Transcription and Understanding Workshop,* 1998.

[10] J. Cao, C. W. Ngo, Y. D. Zhang, Y. D. Zhang, and J. T. Li. Tracking web video topics: discovery, visualization and monitoring. *IEEE Transactions on Circuits and Systems for Video Technology,* 21(12): 1835-1846, 2011.

[11] 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.

[12] J. Shao, S. Ma, W. M. Lu, and Y. T. Zhuang. A unified framework for web video topic discovery and visualization. *Pattern Recognition Letters,* 33(4): 410-419, 2012.

[13] R. C. Hong, J. H. Tang, H. K. Tan, C. W. Ngo, S. C. Yan, and T. S. Chua. Beyond search: event driven summarization for web videos. *ACM Transactions on Multimedia Computing, Communications, and Applications,* 2011.

[14] T. L Chen, S. Q Jiang, L. Y Chu, Q. M Huang. Detection and location of near-duplicate video sub-clips by finding dense subgraphs. *In ACM Multimedia,* 2011.