E^2SGM: Event Enrichment and Summarization by Graph Model

Xueliang Liu[1], Feifei Wang[1], Benoit Huet[2], and Feng Wang[3]

[1] Hefei University of Technology, China
[2] EURECOM, Sophia Antipolis, FRANCE
[3] East China Normal University, China
liuxueliang@hfut.edu.cn, ahfywff@gmail.com, huet@eurecom.fr, fwang@cs.ecnu.edu.cn

Abstract. In recent years, organizing social media by social event has drawn increasing attentions with the increasing amounts of rich-media content taken during an event. In this paper, we address the social event enrichment and summarization problem and propose a demonstration system E^2SGM to summarize the event with relevant media selected from a large-scale user contributed media dataset. In the proposed method, the relevant candidate medias are first retrieved by coarse search method. Then, a graph ranking algorithm is proposed to rank media items according to their relevance to the given event. Finally, the media items with high ranking scores are structured following a chronologically ordered layout and the textual metadata are extracted to generate the tag cloud. The work is concluded in an intuitive event summarization interface to help users grasp the essence of the event.

1 Introduction

With thousands of photos and videos being uploaded every minute to social media sharing websites, there is an increasing need to automatically uncover the structure of online media, so as to help users grasp the gist at a glance. It is often the case that among user contributed media, much content is captured during real-life events and conveys various interesting information about them [3]. However, in most cases users fail to provide detailed annotations about the media before sharing them online. Efficiently and effectively understanding the heterogeneous event semantic behind these multimedia content and summarizing them succinctly for consumer consumption, is a significant challenge in the research community[4, 1].

In this work, we address the event enrichment and summarization problem and propose a framework to summarize events by photos/videos from a large media collection. The problem is solved in three steps. At first, considering the huge amount of media in the dataset, we employ a coarse retrieval method to obtain a summarization candidate set, so as to reduce computation cost. Then, in order to find the most relevant medias for each event, a graph ranking method is proposed to rank them by the relevance to the event, and the ones with high ranking score are kept for summarization purpose. Finally, we provide a vivid interface to users. A graph layout, where all of the media items are ordered in chronological order, is used to visualize and structure the summarized event. The framework of the proposed approach is illustrated in Fig 1.

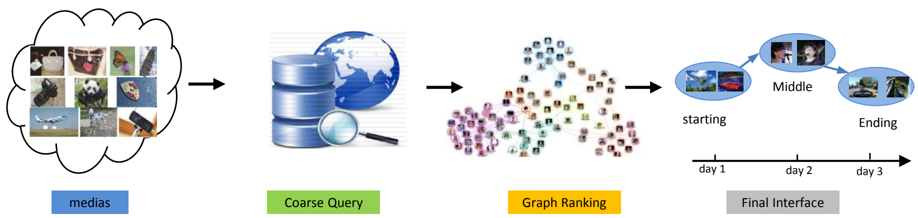

Fig. 1. The framework of the proposed method.

2 Our Solution

The main goal of the proposed event-based media summarization system is to find the relevant media from a big multimedia dataset for a given event and represent it with an informative and engaging layout. In this section, we reveal the details of the proposed approach.

2.1 Dataset

The demonstration is built on the YF100MCC dataset which contains 100M photos and videos compiled from Yahoo! Flickr [5]. For each photo or video, many metadata, such as the owner, taken/uploading time, title, description, location, and the url are provided.

The efficiency of the textual, temporal and spatial feature in event based research has been proven in previous work [2]. Unfortunately the spatial feature is optional in this dataset. Therefore, only the textual and temporal features are selected to model the media data. Specifically, the tf-idf method is used to extract the textural feature and 1000D textual feature is obtained for each media item and the media taken time, which is represented as the number of pas seconds from Unix epoch, is used as the temporal feature.

2.2 Coarse Query

To reduce the computation cost when summarizing an event, we build a text indexing system on the dataset and use the first sentence in an event description as the textual query. The results are further filtered by temporal span to remove the irrelevant ones according to the event taken time.

Considering there are different types of events, we design different time spans to filter the results. For the global multi-day event, we keep the media items that are taken in some days before an event. For the local single-day event, we only keep the media items that are taken on the same day of the event timestamp. In the dataset, there are some media without taken time attribute, and we use the upload time instead.

2.3 Graph-based Ranking

In the coarse query result, the amount of media is still too large, and their importance on representing an event is not evaluated. The users still can not get the event gist directly and quickly by looking at them. In our framework, we propose a graph ranking method to rank the media data and obtain a small subset which are highly relevant to an event.

Graph ranking method [6] is popularly used in media analysis, where the vertices and edges represent medias and their relations respectively. The graph ranking method is derived from the local similarity assumption. Obviously, the images with the similar content should be assigned the similar relevant score on a given event. Mathematically, given a media collection $D = \{x_1, x_2, \cdots, x_n\}$, we denote r an ordering of theses media data, and $r(i)$ be the score of media x_i under the ordering. The similarity between two media can be computed based on Gaussian function,

$$W_{ij} = exp(-\frac{\|x_i - x_j\|^2}{\sigma^2}) \tag{1}$$

where σ is the radius parameter of the Gaussian function.

In addition, the relevance of an image should depend on its "closeness" to the event. In this problem, the event is described using both textual and temporal modality. Hence the images which are closer in the two modalities to an event should be assigned a higher relevant score.

There have been several works addressing the similarity problem. For simplicity, we use Cosine similarity and Gaussian similarity method to measure the textual and temporal similarities respectively.

The similarity of an event e and an image x_i with textual feature can be formulated as follows.

$$\mathbf{Sim}(e, x_i) = \frac{\|Text(e)Text(x_i)\|}{\|Text(e)\| \, \|Text(x_i)\|} \tag{2}$$

where $Text(o)$ is the textual feature extracted from o.

The time similarity can be formulated according to the time span between the image taken time and event held time.

$$\mathbf{Span}(e, x_i) = exp^{-\frac{\|Time(e) - Time(x_i)\|}{\sigma_t}} \tag{3}$$

Based on these assumptions, we formulate the event based media ranking problem in a graph learning framework as follows.

$$\mathbf{\Omega}(\mathbf{r}) = W_{ij}\left(\frac{\mathbf{r}(i)}{\sqrt{D(i)}} - \frac{\mathbf{r}(j)}{\sqrt{D(j)}}\right)^2 + \lambda_1 \sum_{i=1}^{n} (\mathbf{r}(i) - \mathbf{Sim}(e, x_i))^2 + \lambda_2 \sum_{i=1}^{n} (\mathbf{r}(i) - \mathbf{Span}(e, x_i))^2. \tag{4}$$

where $\mathbf{r}(i)$ is the relevance score of x_i, and $D(i) = \sum_{i=1}^{n} W_{ij}$. There are three parts in the above objective function. The first part is the graph smoothness term which means that the similar medias could achieve similar score, the second and third parts mean that the relevance score should be larger if a media is closer to the given event measured by textual or temporal similarity.

Equation 4 can be optimized by derivative method, and the closed form solution could be derived as follows.

$$\mathbf{r}^* = \frac{\lambda_1 \mathbf{Si} + \lambda_2 \mathbf{Sp}}{(\lambda_1 + \lambda_2)I + L} \tag{5}$$

The media are sorted by r^* and the top items which are assigned with high values are used in the following illustration steps.

2.4 Illustration

A collection of high relevant media is obtained by the graph based ranking method as proposed in Section 2.3. To help the users understand the evolution of an event, we represent the data by a graph method. In other words, we link the image pairs according to their similarity and show them according to the chronological order. In detail, we measure the similarity of a image pair by Equation 1. For each image, only two nearest neighbors are connected by an edge to encourage the sparsity of the graph and to provide a neatly-arranged visual to the users. When the graph is built, all the image vertices are presented along the timeline according to their chronological order. Fig. 2 illustrates an example to demonstrate the rationale of this illustrating approach.

Besides visual illustration, we use a tag cloud to organize the textual data to provide a nice and meaningful visualization. Tag cloud is a form of histogram which can represent the amplitude of over a hundred items. In tag cloud, the importance of each tag is shown with different format. This format is useful for quickly perceiving the most prominent terms. For each media illustrating an event, we segment the text description into tags, count the frequency of each tag and generate the tag cloud with tags in different font sizes and colors. By tag cloud, the key concept of an event could be perceived effectively.

We design a pilot system for event summarization to help users visualize and understand the event. For more details, please visit the project website [4].

3 Implement and Result

3.1 Implement Configuration

In the experiment, the YF100MCC dataset is managed by MongoDB[5]. To assist the coarse search, we create the index on taken time and text attributes. The

[4] http://lmc.hfut.edu.cn/eventgraph
[5] http://www.mongodb.org

first sentence and the timestamp in event description is used as the query to retrieve the relevant media roughly. Specifically, we use the first event description sentence to do the full-text search which is supported by MongoDB. As to the temporal attribute, we design different rule to different events. For the global and multi-day events, we query the media which is taken 5 days before the timestamp. For the one-day events, it is reasonable consider the relevant media from the ones that were taken on the same as the event.

The coarse query candidates are ranked by the proposed graph ranking algorithm. The more relevant a media is to an event, the higher score it is assigned by the model. In our experiments, the parameters λ_1 and λ_2 are both set to 0.1 experimentally.

To provide a readable result to users, we keep the top 50 media items to illustrate an event. The 50 media items of each event are connected by a graph, where each media is linked by two successive items in chronological order. The textual metadata from each media item is extracted and used to generate the tag cloud according to their frequencies.

3.2 Results

In this section, we evaluate the quality of the two types of summarization result. Here, we take an query as example: the event "Occupy Movement" at 1 Nov 2011 12:00am UTC. The event summarization is visualized in Fig 2. As shown in this figure, there are two parts in this summarization, while the left part is the visual graph and the right part is the tag cloud.

In the visual graph, there are 50 medias to illustrate this event. These medias are linked to another item according to their similarity by textual metadata, which provides a clue on the event evolution in the past days. These medias are presented along the timeline according to their chronological order. Thanks to such a graph, the user is able to find out how the event updates at each short period, so that they can comprehend the event gist at a glance.

Besides the visual graph, we also use the tag cloud to present the gist of the event. In the tag cloud, we can find out many terms such as "occupy","banker","economy", "action". With these words, it could be perceived that this event relates to bankers and economy.

4 Conclusions

In this paper, we address the event enrichment and summarization challenge and propose a system that can help users obtain the gist of an event at different timestamps. In the proposed demonstration, the event-relevant candidates are first queried by coarse search approach and are ranked by the proposed graph ranking method to find the most relevant samples for summarizing an event. Finally the event summary is presented by a media graph in chronological order to show the event trend, and by tag cloud to present the gist of the event.

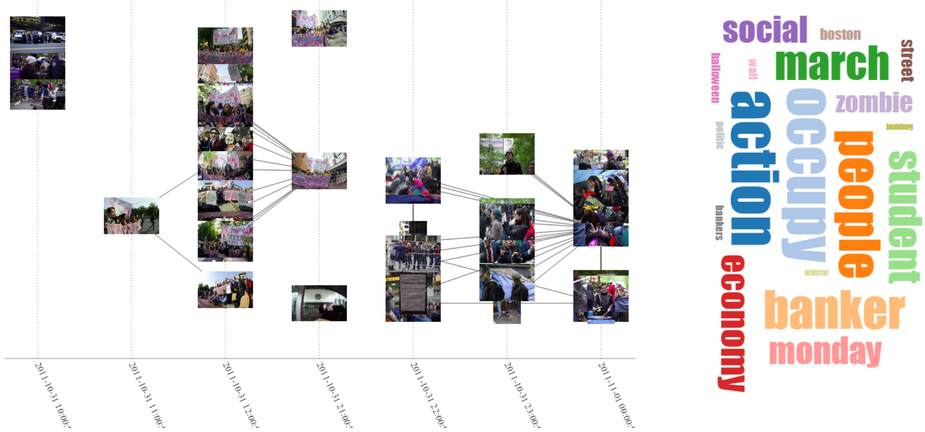

Fig. 2. The Visual Example of the Event “Occupy Movement” at timestamp 1 Nov 2011 12:00am UTC

5 ACKNOWLEDGMENTS

This research is supported in part by the National High Technology Research and Development Program of China(Grant No.2014AA015104) and in part by the National Nature Science Foundation of China under Grant 61502139, Grant 61125106, Grant 61272393, Grant 61322201, and Grant 61432019.

References

1. R. Hong, J. Tang, H.-K. Tan, C.-W. Ngo, S. Yan, and T.-S. Chua. Beyond search: Event-driven summarization for web videos. *ACM Trans. Multimedia Comput. Commun. Appl.*, 7(4):35:1–35:18, Dec. 2011.
2. X. Liu and B. Huet. Heterogeneous features and model selection for event-based media classification. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, April,2013.
3. V. Mezaris, A. Scherp, R. Jain, and M. S. Kankanhalli. Real-life events in multimedia: detection, representation, retrieval, and applications. *Multimedia Tools Appl.*, 70(1):1–6, 2014.
4. G. Petkos, S. Papadopoulos, V. Mezaris, R. Troncy, P. Cimiano, T. Reuter, and Y. Kompatsiaris. Social event detection at mediaeval: a three-year retrospect of tasks and results. In *Proc. ACM ICMR 2014 Workshop on Social Events in Web Multimedia*, 2014.
5. B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.-J. Li. The new data and new challenges in multimedia research. *arXiv preprint arXiv:1503.01817*, 2015.
6. M. Wang, K. Yang, X.-S. Hua, and H.-J. Zhang. Towards a relevant and diverse search of social images. *Multimedia, IEEE Transactions on*, 12(8):829–842, Dec 2010.