International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor (2018): 7.426

Review on Generating Social Event Storyboard

Pournima R. Patil[1], Amol B. Rajmane[2], Rohit D. Mane[3]

[1]M.Tech.-II (Computer Science & Engineering), Ashokrao Mane Group of Institutions, Vathar. Kolhapur, Maharashtra

[2]Ashokrao Mane Group of Institutions, Vathar, Kolhapur, Maharashtra

[3]Dr. J. J. Magdum College of Engineering, Jaysingpur, Kolhapur, Maharashtra

Abstract: ***Traditional websites have been driven by means of human-edited events which cause large web seeks visitors. This paper is a survey performed for identifying the various event detection strategies that are beneficial for event mining. Moreover this work indicates an automatic system to detect events from search log data and generates a storyboard where events are arranged chronologically. To detect events from search log data, an approach known as Smooth Nonnegative Matrix Factorization framework (SNMF) is used. In addition, to describe events set of relevant images are attached to each event. These relevant images are taken from image log data by analyzing both local and global image content feature.***

Keywords: Storyboard, click-through data, factorization, image search.

1. Introduction

Storyboard is a collection of images representing the events where events are arranged in chronological order and each event is attached with relevant image.

Traditional websites had some limitations like human edited and small coverage area, nonsalable services, and biased reports news, where event generation was not automatic. But some latest techniques taken under consideration here would work as perfect solution to automatically generate storyboard for those events. To generate a storyboard, the events are detected from search log data. It is observed that search log data is a good data resource for event detection because: 1) Search logs may cover a variety of real-world events; 2) Search logs clearly gives users interests, over many Internet users; and 3) Search logs respond to events happening in real time correctly.

To detect events from search log data, a method called Smooth Non-negative Matrix Factorization (SNMF)is used. SNMF contains two phases, Event detection by SNMF and Photo selection. There are three steps for event detection.In event detection, initially events are detected from search log data. After that it discovers groups of queries having high co-occurrence frequency,is known as topic factorization.

2. Literature Survey

The technique used to perform topic factorization is Smooth Nonnegative Matrix Factorization(SNMF).This method discovers group of topics having highest co-occurant frequency from search log data. The log data as taken converted to matrix of size Q*D.The NMF ignores the order of input documents. This would not affect the decomposition result but when decomposing time series signals similar constraints arises. To fix such constraints SNMF was proposed by introducing extra regularization factor S(H).

1) Probabilistic topic models

Machine learning researchers have developed probabilistic topic modeling [1], a suite of algorithms that aim to discover & annotate large archives of documents with thematic information. Topic modeling algorithms are statistical methods that analyze the words of original texts to discover the themes that run through them. Through topic modeling we are able to organize and summarize electronic document at a scale that would be impossible by human annotation. Latent Dirichlet Allocation is the simplest topic model.

2) Query Based Event Extraction along a Timeline

The author H.L. Chieu presents a framework and a system that extracts [2] events relevant to a query from a collection of documents and places events along a timeline. There are two different measures for ranking sentences: 1) Interest 2) Burstiness. Interesting sentences are sentences reporting interesting events. Burstiness involves extraction of sentences that are strongly associated to the date duration. The advantages of this approach are: 1) It is efficient and not require any expensive operations. 2) Sentences are better units of information as they allow quick access to their source documents.

3) Representative Photo Selection For Restaurants in Food Blogs

In the paper, proposed a system to analyze and select representative photos [3] for restaurant based on blog platform. A strong food detection model retrieves food photos and an aesthetic quality assessment method is used to select representative photos. These representative photos create impression of the restaurant on users. Here two main components are included: photo quality assessment and representative photo selection. Photo quality assessment rank the photos based on lots of aesthetic rules, such as composition, color distribution and intensity distribution. In selection process it selects representative photos for each restaurant.

4) Time Dependent Event Hierarchy Construction

In this paper, an algorithm called Time Driven Documents-partition (TDD) [4] is proposed to construct an event hierarchy in a text corpus based on a user query. An event is an object which contains: 1) Collection of documents having similar contents 2) Set of representative features 3) Event period. As event hierarchy resulted, the TDD algorithm

Licensed Under Creative Commons Attribution CC BY

tackles the problem through three steps: 1) Identify the features that are related to the query according to both timestamps and contents of the document.2) Extracting documents related to the burst features based on time 3) Partition the extracted documents to form event hierarchy.

5) Multimedia search reranking: A literature survey

Multimedia search reranking, which reorders visual documents [5] to improve initial text-only searches. The purpose of this paper is to categorize and evaluate numerous techniques have been developed for visual search re-ranking. These also discuss relevant issues such as data collection, evaluation metrics, and benchmarking.

6) Event Detection from Evolution of Click-through Data

In this paper, author proposed the approach [6] to detect events from the click-through data, which is nothing but log data of web search engines. Given the click-through data to proposed approach, at each time it is represented by different graphs such as bipartite Graphs, vector based graph etc. Event detection problem is equivalent to clustering process. The clustering process is typically based on a two-phase graph cut algorithm. In the first phase, query page pairs are clustered based on the semantic-based similarity.In the second phase, query-page pairs are clustered based on the evolution pattern-based similarity.

7) Event Detection in Twitter

This paper focuses on detecting events [7] by analyzing the text stream in Twitter. Detecting events in Twitter has some challenges and these challenges are tackled with EDCoW (Event Detection with Clustering of Wavelet-based Signals). EDCoW builds signals for individual words, filters away the trivial words Cluster signals. The signal such as time domain is converted into the time-scale domain by wavelet transformation. There are two types of Wavelet transformation, continuous wavelet transformation (CWT) and discrete wavelet transformation (DWT). Redundant representation provided by CWT of the signal under analysis. DWT provides a non-redundant, highly efficient wavelet representation of the signal. EDCoW's three main components: (1) Signal construction, (2) Cross correlation computation, and (3) Modularity-based graph partitioning. The advantage of EDCoW is that it achieves a fairly good performance.

8) CeleBrowser: An example of browsing big data on small device

Here author demonstrate the system called CeleBrowser which is mobile-based celebrity video browsing system. This system provides four views: people-centric, timeline-centric, month-centric and topic-centric, for browsing celebrity-related videos.

9) Event mining from the blogosphere using topic words

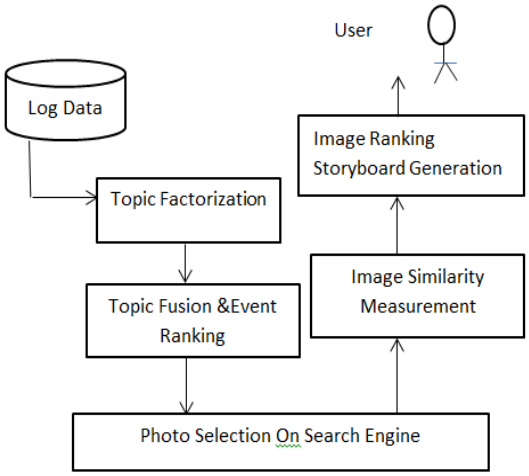
Every blog post in the Blogosphere has a well-defined timestamp, this feature of the Blogosphere used by the author to discover bursty terms and correlations between them during a time interval. To discover bursty terms Kleinberg's automaton technique is applied and employed a Euclidean-based distance in order to discover potential correlations between terms without taking into account their context.

The above reviewed papers suggest us some fine techniques to overcome the limitations seen with traditional techniques. These new techniques are used for the storyboard generation.

3. Proposed Methodology

As per review it is clear that the used method Smooth Non-negative Matrix Factorization has some limitations such as result generated by NMF is semantically not correct. As it is complex one too, the better solution is Latent Dirichlet Allocation (LDA). LDA is the simplest topic model which is a three level hierarchical Bayesian model, where documents are represented as random mixture of latent topics and where each topic is characterized by a distribution over words. So we will use this approach for factorization purpose.

The proposed work will contain four modules as shown in following figure



In the first module topic factorization, we will collect data from search log data and discover groups of queries that have high co-occurrent frequency. After first phase completion, we will perform merging of topics having similar behavior and rank the them in topic fusion and event ranking module. Ranking top topics will be referred as social events and nontop topics will as profile topics. Relevant images of specific topic will be selected through search and they are given to next phase where the content similarity among images is measured by using both local as well as global image features. The assumption is that event related photos should have duplicate images in social event image set but should not in profile set. So that simple ranking function is proposed to sort the photos in social event image set. That makes it easy to identify relevant photos to describe the detected event.

4. Conclusion

This survey is conducted to find out various methods for detecting events. Search logs are good data source for finding events which contains user interest. Smooth Non Negative Matrix Factorization (SNMF) is one of better event detectionmethod. Moreover attaching representative images to events along a timeline generate a storyboard.

Licensed Under Creative Commons Attribution CC BY

We will use LDA (Latent Dirichlet Allocation) technique to generate storyboard, which is better one than used technique.

References

[1] D. M. Blei, "Probabilistic topic models," Commun. ACM, vol. 55, no. 4, pp. 77–84, 2012.
[2] H. L. Chieu and Y. K. Lee, "Query based event extraction along a timeline," in Proc. 27th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., 2004, pp. 425–432.
[3] Y.-J. Chang, H.-Y. Lo, M.-S. Huang, and M.-C. Hu, "Representative photo selection for restaurants in food blogs," in Proc. IEEE Int. Conf. Multimedia Expo. Workshops (ICMEW), Jun./Jul. 2015, pp. 1–6.
[4] G. P. C. Fung, J. X. Yu, H. Liu, and P. S. Yu, "Time-dependent event hierarchy construction," in Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2007, pp. 300–309.
[5] T. Mei, Y. Rui, S. Li, and Q. Tian, "Multimedia search reranking: A literature survey," ACM Comput. Surv., vol. 46, no. 3, 2014, Art. no. 38.
[6] Q. Zhao, T.-Y. Liu, S. S. Bhowmick, and W.-Y. Ma, "Event detection from evolution of click-throughdata," in Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2006, pp. 484–493.
[7] J. Weng and B.-S. Lee, "Event detection in Twitter," in Proc. ICWSM, vol. 11. 2011, pp. 401–408.
[8] S. Tan, C.-W. Ngo, J. Xu, and Y. Rui, "CeleBrowser: An example of browsing big data on small device," in Proc. Int. Conf. Multimedia Retr., 2014, p. 514.
[9] Y. Suhara, H. Toda, and A. Sakurai, "Event mining from the blogosphere using topic words," in Proc. ICWSM, 2007, pp. 1–2.
[10] "Automatic Generation of Social Event Storyboard From Image Click-Through Data" Jun Xu, Tao Mei, Senior Member, IEEE, RuiCai, Member, IEEE, Houqiang Li, Senior Member, IEEE, and Yong Rui, Fellow, IEEE. IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 28, NO. 1, JANUARY 2018

Licensed Under Creative Commons Attribution CC BY