



SOCIAL NETWORK BASED EVENTS AND INFORMATION FROM IMAGE CLICKTHROUGH DATA USING DATAMINING

¹P.VIJAYA MOUNIKA, ²Dr.B.HARIBABU

¹M.TECH DEPT OF CSE, KAKINADA INSTITUTE OF TECHNOLOGICAL SCIENCES, RAMACHANDRAPURAM,
ANDHRAPRADESH, INDIA, 533255

²ASSISTANT PROFESSOR, KAKINADA INSTITUTE OF TECHNOLOGICAL SCIENCES, RAMACHANDRAPURAM,
ANDHRAPRADESH, INDIA, 533255

ABSTRACT

Recent studies have shown that a noticeable percentage of web search traffic is about social events. While traditional websites can only show human-edited events, in this paper we present a novel system to automatically detect events from search log data and generate storyboards where the events are arranged chronologically. We chose image search log as the resource for event mining, as search logs can directly reflect people's interests. To discover events from log data, we present a Smooth Nonnegative Matrix Factorization framework (SNMF) which combines the information of query semantics, temporal correlations, search logs and time continuity. Moreover, we consider the time factor an important element since different events will develop in different time tendencies. In addition, to provide a media-rich and visually appealing storyboard, each event is associated with a set of representative photos arranged along a timeline. These relevant photos are automatically selected from image search results by analyzing image content features. We use celebrities as our test domain, which takes a large percentage of image search traffics. Experiments consisting of web search traffic on 200 celebrities, for a period of six months, show very encouraging results compared with handcrafted editorial storyboards.

I. INTRODUCTION

As social creatures, people are by nature curious about others' activities. Information on famous persons have often been of particular interest. This tendency has remained true in the internet era [35]. Since common search engines as well as news websites often experience massive search demands about a myriad of current affairs, a great amount of news and events are collected from the web. However, most social events originate from professional

editors. In this case, it is quite meaningful to detect such events for users automatically instead of manual efforts.

Current search engines often show the summaries of famous persons as a simple profile. From such a summarization, people can easily get a celebrity's basic information like portrait, nationality, birthday, representative works, and awards. The search engine summaries can be considered a concentrated



Fig. 1. Screen shot of www.people.com, a website for celebrity news. The marked region shows recent news of Britney Spears, arranged along timeline.

version of a person's larger relevant event collection. Although such a short profile is very helpful for quickly introducing a person, it cannot satisfy people's curiosity for more detailed and timely information of celebrities. By contrast, some professional websites provide comprehensive and up-to-date information on famous persons. Fig. 1 shows a screen shot of www.people.com, a website well-known for celebrity news and photos. In the marked region of Fig. 1, it shows Britney Spears's recent news (events) arranged along a timeline.

This is a very nice feature for fans to trace their idols' activities. Almost all these websites are powered by human editors, which inevitably leads to several limitations. First, the coverage of human center domains is small. Typically, one website only focuses on celebrities in one or two domains (most of them are entertainment and sports), and to the best of our knowledge, there are no general services yet for tracing celebrities over various domains. Second, these

existing services are not scalable. Even for specific domains, only a few top stars are covered, as the editing effort to cover more celebrities is not financially viable. Third, reported event news may be biased by editors' interests. In this paper, we aim to build a scalable and unbiased solution to automatically detect social events especially related to celebrities along a timeline. This could be an attractive supplement to enrich the existing event description in search result pages. In this paper, we will focus on those events happening at a certain time favored by users as our celebrity-related social events.



Fig. 2. Example of Adele storyboard, from July 2012 to December 2012. The first event is the expectation that she would give birth soon. The second event is the release of her new album "Skyfall". The third event is about her weight.

Meanwhile, about 30% of search queries aim to search for real-world events according to statistics from a commercial search engine data [23]. A further-70% of these queries are related to celebrities, including artists, sports stars, politicians, scientists, entrepreneurs, et al. Thus, we will focus on events related to celebrities because of the volume of related search queries and the ability to obtain ground truth events from professional websites.



II. EXISTING SYSTEM

In the existing system, web search log is another data source which has attracted the interests of many researchers. Search log data contains useful information like user queries and clicked search result URLs. It has been successfully exploited in various areas like relevance ranking, query expansion and query alternation. Besides, search log data is an unbiased statistic showing user intention. It is therefore a good resource for event detection, especially for those events attracting the interests of internet users.

Zhao et al. [40] and Liu et al. [21] have done lots of work in this area. In [40], a bipartite graph is constructed based on query and click URL pairs, and two similarity measurements are proposed for event clustering.

Disadvantages

There is no Event story board to generate group images to gather. It is only fuzzy keyword Search Scheme and not based on story board.

III. PROPOSED SYSTEM

The system proposes a novel framework to detect interesting events by mining users' search log data. The framework consists of two components, i.e., Smooth Non-Negative Matrix Factorization event detection and representative event related image photo selection.

The systems have conducted comprehensive evaluations on large scale real-world click through data to validate the effectiveness.

Advantages

The images are generating based on non-negative matrix factorization. The image click more easy due to SNMF Topic Factorization.

IV. MODULES

4.1 Admin

In this module, admin has to login with valid username and password. After login successful he can do some operations such as ,View all users and authorize,give click option to view all users locations in GMap using Multiple Markers,View all Friend Request and Response ,Add images to Storyboard with image cat,image name,image desc(enc),image users,URL,Add Image ,View all images with date and time, with rank,reviews ,Generate Matrix Factorization to detect event(see below),View search key request and generate using RSA ,View Storyboard request and Authorize ,View all user Search keyword transactions,view all images rank in chart

4.2 User

In this module, there are n numbers of users are present. User should register before doing some operations and also add your location while registration . After registration successful he can login by using valid user name and password and location. After Login successful he will do some operations like ,Search Friend and Find Friend Request ,View all Your Friends and give click option to view all users locations in GMap using Multiple Markers show Route path using GMap,Request Search key for search and view the same,Auto fill Search key and enter Search keyword,view

all images with imagename,rank and click on image or image name to view similar images in StoryBoard and click on specified image to view its details,click URL to visit corresponding website,Request Storyboard to view in Matrix Formation and view story board if it is authorized,View Friends Reviews on images in Storyboard

V. SCREENSHOTS

Image Name	Image Description	Image Uses	Image URL	Uploaded Date	Image Rank	Image Rate
Elephant	This is Elephant and it is very Large Animal and it has long teeth. There is different kinds of Animals among India And foreign countries also In India Karnataka has highest number of Elephant compare to Other State.	this will shows how much all it is	http://www.elephant.com	04/11/2017 15:52:36	7	★ ★

Username	Review	Review Date
rahesh	this is very wide animal	04/11/2017 15:50:34
rameah	this is one of very tall and large animal.	06/11/2017 10:47:21
mahesh	this is one of very tall animal	06/11/2017 12:33:20

Sl.No	Keywords	Keyword Score
1	animals	32
2	parrot	17
3	is	9
4	animal	7
5	birds	106
6	parrot	17
7	yellow parrot	5
8	eagle	2
9	anima	5
10	ani	5
11	elephant	18
12	apple	5
13	mobile	2
14	fruit	3

VI CONCLUSIONS

In this paper, we use search logs as data source to generate social event storyboards automatically. Unlike common



text mining, search logs have short, sparse text queries and the data size is much bigger than some news websites or blogs. Based on these features, we do not use the query text information to do the analysis. Structure and statistic information are used to get the topics and event detection in our work, which can fit the data well. Furthermore, we add time information in our approach to SNMF to make it easier to discover social events compared with traditional NMF methods. Our work performs better than traditional works in this area, e.g. [40], because we can distinguish the topics in a way that gets the events which are most appealing to common users. The associated images were selected to make up the storyboard in a timeline to present a good representation of the mined events using the image search results features and relationships.

VII. REFERENCES

- [1] C. Alexander, B. Fayock, and A. Winebarger. Automatic event detection and characterization of solar events with iris, sdo/aia and hi-c. In AAS/Solar Physics Division Meeting, volume 47, 2016.
- [2] J. Allan, J. G. Carbonell, G. Doddington, J. Yamron, and Y. Yang. Topic detection and tracking pilot study final report. 1998.
- [3] S. Arora, R. Ge, and A. Moitra. Learning topic models—going beyond svd. In Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on, pages 1–10. IEEE, 2012.
- [4] N. Babaguchi, S. Sasamori, T. Kitahashi, and R. Jain. Detecting events from continuous media by intermodal collaboration and knowledge use. In Multimedia Computing and Systems, 1999. IEEE International Conference on, volume 1, pages 782–786. IEEE, 1999.
- [5] P. N. Bennett, R. W. White, W. Chu, S. T. Dumais, P. Bailey, F. Borisyuk, and X. Cui. Modeling the impact of short-and long-term behavior on search personalization. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, pages 185–194. ACM, 2012.
- [6] D. M. Blei. Introduction to probabilistic topic models. *Comm. ACM*, 55(4):77–84, 2012.
- [7] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *The Journal of machine Learning research*, 3:993–1022, 2003.
- [8] Y.-J. Chang, H.-Y. Lo, M.-S. Huang, and M.-C. Hu. Representative photo selection for restaurants in food blogs. In Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on, pages 1–6. IEEE, 2015.
- [9] H. L. Chieu and Y. K. Lee. Query based event extraction along timeline. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, pages 425–432. ACM, 2004.
- [10] T.-C. Chou and M. C. Chen. Using incremental plsi for threshold resilient online event analysis. *Knowledge and Data Engineering, IEEE Transactions on*, 20(3):289–299, 2008.
- [11] H. Cui, J.-R. Wen, J.-Y. Nie, and W.-Y. Ma. Probabilistic query expansion using



query logs. In Proceedings of the 11th international conference on World Wide Web, pages 325–332. ACM, 2002.

[12] S. Essid and C. F´evotte. Smooth nonnegative matrix factorization for unsupervised audiovisual document structuring. *Multimedia, IEEE Transactions on*, 15(2):415–425, 2013.

[13] G. P. C. Fung, J. X. Yu, H. Liu, and P. S. Yu. Time-dependent event hierarchy construction. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 300–309. ACM, 2007.

[14] T. Hofmann. Probabilistic latent semantic indexing. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pages 50–57. ACM, 1999.

[15] T. Joachims. Optimizing search engines using click through data. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 133–142. ACM, 2002.

[16] N. Kawamae. Trend analysis model: trend consists of temporal words, topics, and timestamps. In Proceedings of the fourth ACM international conference on Web search and data mining, pages 317–326. ACM, 2011.

[17] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.

[18] J. Li and C. Cardie. Timeline generation: Tracking individuals on twitter. In Proceedings of the 23rd international

conference on World wide web, pages 643–652. ACM, 2014.

[19] Z. Li, B. Wang, M. Li, and W.-Y. Ma. A probabilistic model for retrospective news event detection. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pages 106–113. ACM, 2005.

[20] A. Liu, W. Lin, and M. Narwaria. Image quality assessment based on gradient similarity. *Image Processing, IEEE Transactions on*, 21(4):1500–1512, 2012.