International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 25th Dec 2018. Link :http://www.ijiemr.org/downloads.php?vol=Volume-07&issue=ISSUE-13

Title: **AUTOMATIC GENERATION OF SOCIAL EVENT STORYBOARD FROM IMAGE CLICK-THROUGH DATA**

Volume 07, Issue 13, Pages: 715–720.

Paper Authors
[1]M.SURYAKUMARI, [2]MD.SAMEERUDDIN KHAN

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

AUTOMATIC GENERATION OF SOCIAL EVENT STORYBOARD FROM IMAGE CLICK-THROUGH DATA

[1]M.SURYAKUMARI, [2]MD.SAMEERUDDIN KHAN

[1]Mtech student, Sree Dattha Institute of Engineering and Science

[2]Professor, Sree Dattha Institute of Engineering and Science

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.

1. INTRODUCTION

The events are detected from search log data and generate story boards where events are arranged along a time line. It is found that search log data is a good data resource for event detection because: (1) search logs cover a wide varietyof real world events (2) search log directly reflect user's interests (3) search logrespond to real time events.

To discover events from log data, an approach called Smooth Non-negative Matrix Factorization (SNMF) framework is used. There are two basic ideas for SNMF:

(1)It promotes event queries

(2) It differs events from popular queries. SNMF guarantee weights for each topic to be nonnegative and considers time factor for event development. To make event detection easier, relevant images are attached for each event.

There are two phases for the proposed approach: Event detection by SNMF and Event photo selection. In event detection, initially events are searched from log data. Then it discovers groups of queries that have high frequency which is known as topic factorization. Next topics with similar behaviors are merged together along a timeline which is called topic fusion. Event ranking happens in which topics like social events are highlighted. After ranking top topics are called social events and non top topics are called profile topics. In event photo selection, both the social events and profile topics are sent to search engines like Google or Bing. The search engines generate two sets of image thumbnails which contains relevant images to social events. Image similarity measures occur in which similarity between events and images are measured. Image ranking is done which is sorting of images in the social event image set. Finally all social events together with their images constructa storyboard.

2. PROBLEM STATEMENT

GOALS

- We propose 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
- We have conducted comprehensive evaluations on largescale real-world click through data to validate the effectiveness.

ALGORITHMS USED:

SNMF Topic Factorization: In classic topic modeling, the inputs are text documents consisting of words and the outputs are decompositions of these documents into topics. Here, each topic is a distribution over the word vocabulary. Analogically, we treat one day's log data as a "document" and each query as a "word". The "vocabulary" consists of all the unique queries of a celebrity in his/her log records, i.e., the set Q defined in Section III A. The assumption is, various stories (potentially interesting events or other representative aspects) of a celebrity are considered as "latent topics" leading to different search queries. It should be noted that we choose a whole query as a "word" but not break each query into real English words. This is because a query is more like a short phrase having specific semantic meanings compared to single word. Breaking a query into words may introduce unexpected ambiguities to topic factorization. For example, the word "love" in the queries "love story" and "love Harry Styles" of Taylor Swift has completely different semantics – the former is about one of her famous songs and the latter is about her ex-boyfriend. Widely used algorithms for topic factorization include probabilistic latent semantic indexing (PLSI), latent Dirichlet allocation (LDA), singular value decomposition (SVD) , non-negative matrix factorization (NMF) , and their variants. In this paper, we choose NMF as it has a nice advantage – data must be decomposed into a sum of additive components. In other words, both the coefficients of "documents' distributions over topics" and the coefficients of "topics' distributions over queries" must be non-negative. This makes sense, especially for event modeling, as it is hard to accept the explanation that we observe a certain query just because some events didn't happen that day. In addition, the non-negative coefficients also improve event mining in the next subsections. The log data is first converted into a matrix D of the size

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

|Q| × |D|. Each row represents a query and each column indicates one day. Every item Dij is the number ith query that was observed on the j th day. NMF aims to find two nonnegative matrices W and H satisfying D ≈ W × H.

W = [w1, . . . wK] in which every column wk(1 ≤ k ≤ K) denotes a topic, and K is the pre-defined number of topics. H = [h1, . . . h|D|] in which each column hd(1 ≤ d ≤ |D|) is the decomposition coefficients of topics for the d th day. According to [17], the decomposition problem converts to minimizing the cost function.

$$D_{KL}^{g}(\mathbf{A}\|\mathbf{B}) = \sum_{ij}\left(\mathbf{A}_{ij}\ln\frac{\mathbf{A}_{ij}}{\mathbf{B}_{ij}} - \mathbf{A}_{ij} + \mathbf{B}_{ij}\right). \quad (3)$$

arg min W,H D g KL(DkW × H) s.t.W ≥ 0, H ≥ 0.

Here, D g KL(AkB) is the generalized Kullback-Leibler divergence of two matrices

$$\arg\min_{\mathbf{W},\mathbf{H}}\{D_{KL}^{g}(\mathbf{D}\|\mathbf{W}\times\mathbf{H}) + \lambda\times S(\mathbf{H})\},$$
$$S(\mathbf{H}) = \sum_{d=2}^{|\mathcal{D}|}\|\mathbf{h}_d - \mathbf{h}_{d-1}\|_2 \quad s.t.\mathbf{W}\geq 0, \mathbf{H}\geq 0. \quad (4$$

Like most other topic modeling algorithms, the standard NMF ignores the orders of input documents. In other words, permutation of the order of columns in D would not affect the decomposition results. However, for log mining, the temporal order is a critical factor which needs to be taken seriously. That is to say, there shouldn't be significant difference between queries (and related topics) from two adjacent days. Similar constraints also arise when decomposing time-series signals such as audio stream [12]. To embed such constraints, Smooth Non-Negative Matrix Factoriazation(SNMF) was proposed by introducing an extra regularization factor S(H) to the cost function.

$$dist_Q(t_k, t_l) = KL_Q(t_k, t_l)$$
$$= \frac{1}{2}\sum_{i=1}^{|Q|}(P_Q(q_i|t_k)\ln\frac{P_Q(q_i|t_k)}{P_Q(q_i|t_l)} + P_Q(q_i|t_l)\ln\frac{P_Q(q_i|t_l)}{P_Q(q_i|t_k)}). \quad (5)$$

3. PROBLEM SOLUTION

DISADVANTAGES:

- 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 covered1, as the editing effort to cover more celebrities is not financially viable.
- Third, reported event news may be biased by editors' interests.
- Discovering events from a search log is not a trivial task.
- Existing work on log event mining mostly focus on merging similar queries into groups, and investigating whether these groups are related to semantic events like "Japan Earthquake" or "American Idol". Basically, their goals are to distinguish salient topics from noisy queries. Directly applying their approaches will fail as the discovered topics are more likely related to vast

and common topics, which may be familiar to most users.

PROPOSED SYSTEM:

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. we would like to detect those more interesting social events to entertain users and fit their browsing taste, which could be supplementary to some current knowledge bases. A novel approach is proposed in this paper using Smooth Nonnegative Matrix Factorization (SNMF) for event detection, by fully leveraging information from query semantics, temporal correlations, and search log records. We use the SNMF method rather than the normal NMF method or other MF method to guarantee that the weights for each topic are non-negative and consider the time factor for event development at the same time. The basic idea is two-fold: 1) promote event queries through by strengthening their connections based on all available features; 2) differentiate events from popular queries according to their temporal characteristics.

4. CONCLUSION

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, because we can distinguish the topics in a way that gets the events which are most appealing to common users. The associatedimages were selected to make up the storyboard in a timelineto present a good representation of the mined events using the image search results features and relationships.

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.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

[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 a 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 thresholdresilient 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. Fevotte. 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 clickthrough 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.

[21] H. Liu, J. He, Y. Gu, H. Xiong, and X. Du. Detecting and tracking topics and events from web search logs. ACM Transactions on Information Systems (TOIS), 30(4):21, 2012.

[22] D. G. Lowe. Object recognition from local scale-invariant features. In Computer vision, 1999. The proceedings of the seventh IEEE international conference on, volume 2, pages 1150–1157. Ieee, 1999.

[23] Q. Mei, C. Liu, H. Su, and C. Zhai. A probabilistic approach to spatiotemporal theme pattern mining on weblogs. In Proceedings of the 15th international conference on World Wide Web, pages 533–542. ACM, 2006.

[24] T. Mei, Y. Rui, S. Li, and Q. Tian. Multimedia search reranking: A literature survey. ACM Computing Surveys (CSUR), 46(3):38, 2014.

[25] M. Platakis, D. Kotsakos, and D. Gunopulos. Searching for events in the blogosphere. In Proceedings of the 18th international conference on World wide web, pages 1225–1226. ACM, 2009.

[26] S. D. Roy, T. Mei, W. Zeng, and S. Li. Towards cross-domain learning for social video popularity prediction. Multimedia, IEEE Transactions on, 15(6):1255–1267, 2013.

[27] Y. Rui, T. S. Huang, and S.-F. Chang. Image retrieval: Current techniques, promising directions, and open issues. Journal of visual communication and image representation, 10(1):39–62, 1999.

[28] E. Sadikov, J. Madhavan, L. Wang, and A. Halevy. Clustering query refinements by user intent. In Proceedings of the 19th international conference on World wide web, pages 841–850. ACM, 2010.

[29] S. Song, Q. Li, and N. Zheng. Understanding a celebrity with his salient events. In Active Media Technology, pages 86–97. Springer, 2010.

[30] Y. Suhara, H. Toda, and A. Sakurai. Event mining from the blogosphere using topic words. In ICWSM, 2007.

[31] S. Tan, C.-W. Ngo, J. Xu, and Y. Rui. Celebrowser: An example ofbrowsing bigdata on small device. In Proceedings of International Conference on Multimedia Retrieval, page 514. ACM, 2014.

[32] T. C. Walber, A. Scherp, and S. Staab. Smart photo selection: Interpret gaze as personal interest. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 2065–2074. ACM, 2014.