

# Scalable and Unbiased Solution to Generate Social Event Storyboard using Smooth Nonnegative Matrix Factorization(SNMF)

Prof. Dakshayani Ijeri<sup>1</sup>, Miss. Sushma Singaraddi<sup>2</sup>, Miss. Sweta Jogur<sup>3</sup>,  
Miss. Vidyashree Valasang<sup>4</sup>, Mr. Ragu Konnur<sup>5</sup>

<sup>1</sup>Assistant Professor, <sup>2,3,4,5</sup>B.E Students, B.L.D.E.A'S V.P Dr P.G Halakatti College of Engineering and  
Technology, Vijayapura, Karnataka, India

<sup>1</sup>cse.ijeri@bldeacet.ac.in, <sup>2</sup>singaraddis@gmail.com, <sup>3</sup>shwetajogur64@gmail.com,  
<sup>4</sup>vidyashreevalasang@gmail.com, <sup>5</sup>rajukonnur@gmail.com

**Abstract** – Generally people are more interested to know about the social events such as news about entrepreneur, celebrity, sports star, politicians, scientist, etc. Most of the news website and some of the search engines provides the information of current affairs but have certain limitations like the coverage of news is under the control of editor where he can modify as per his requirement, the arena of news coverage is too small and unscalable. Story board generation system will overcome these limitations through scalable and unprejudiced solution to automatically detect social events especially related to celebrities along a timeline by using SNMF(Smooth Non-Negative Matrix Factorization) algorithm. This algorithm has two approaches namely Event Detection and Photo Selection. Event Detection will help to categorize the topics through user queries and it will also take care of topic fusion of similar events. Photo selection will enable users to get required information just by clicking on interested image. The story board generation system will satisfy the users by displaying the results in appropriate manner. It will display the search results with its respective category along with appropriate images and user has to just click on image to get the related story board.

**Keywords** — Smooth Nonnegative Matrix Factorization (SNMF), Unprejudiced Solution, Event Detection, Topic fusion.

## I. INTRODUCTION

Nowadays, people are more curious to know about public figures like entrepreneur, celebrity, sports star, politicians, scientist, etc. Most of the news website and some of the search engines provides the information of current affairs. Recent search engines often shows the summaries of famous person as simple story, through that summarization people can get the celebrity basic information like portrait, nationality, birth date, representative works and honors etc seamlessly. These short stories from websites just provide a quick introduction of a person but it cannot satisfy the curiosity of the people who are in need to get more detailed information of the particular celebrity.

These existing websites leads to several limitations like First, the arena of news coverage is too 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 unscalable. Even for specific domains, only a few top stars are covered, as the editing effort to cover more celebrities is not financially viable.

Third, the coverage of news is under the control of editor where he can modify as per his requirement. 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. Meanwhile, about 30% of search queries aim to search for real-world events according to statistics from a commercial search engine data [5]. A further-70% of these queries are related to celebrities, including artists, sports stars, politicians, scientists, entrepreneurs, et al. Thus, 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.

Therefore, to achieve the goal a novel approach is proposed using Smooth Nonnegative Matrix Factorization (SNMF)[6] for event detection, by fully leveraging information from query semantics[7], 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.

## II. RELATED WORK

[1]As social creatures, people are by nature curious about others' activities. Information on famous persons have often been of particular interest. 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. 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. From such a summarization, people can easily get a celebrity's basic information like portrait, nationality, birthday, representative works, and awards.

[2]News articles are not good enough for mining events considering users interests, as most reports from mainstream media are dominated by breaking news and influential social events. Weblog is not an ideal choice as blog posts are mainly about individual stories covering regular people rather than interesting events for all general users. Besides news and weblog data, there have been some recent research efforts attempting to extract events from web search logs. Search log data is a good data resource for detecting those events and gaining user attention instantly, because 1) search logs may cover a wide variety of real-world events.2) search logs directly reflect users' interests, as they are in essence a majority voting over billions of internet users; and 3) search logs respond promptly to events happening in real time.

[3]The method for photo selection from image search logs is presented. In this way, photos that have more duplicates returned for queries of the same event, while at the same time they do not appear in search results of those popular queries, will be selected to describe that event.

[4]To generate a vivid storyboard for social event which is similar to image selection to some distance. Traditionally, photos are selected according to their local and global features to judge the photo quality. Photo selection is final step to help summarize the events from photo collection. To complete timeline for storyboard which is quite different with common photo selection job.

## III. SYSTEM ARCHITECTURE

The system will require two databases, one to store all event details and another to maintain user query details. The system architecture is divided in two parts as admin architecture(Fig 3.1) and User Architecture(Fig 3.2). As shown in Fig 3.1the admin has to login with valid username and password. After login successful he can do some operations such as ,view all users and authorize, view all friend request and response, add images to Storyboard with

image category, image name, image description, URL, add image, view all images with date and time, with rank and reviews. Admin will perform matrix factorization based on the highest frequency queries to detect event. Matrix Factorization is done by considering log data that has n number of queries. By applying matrix factorization we can group the images based on the category. Admin can also view search key request and generate using RSA.

Once the user is registered, he will be able to login successfully by providing username and password as shown in Figure 3.2. The authorized user will be provided with search key by admin through which he will be able to search for images of his interest.

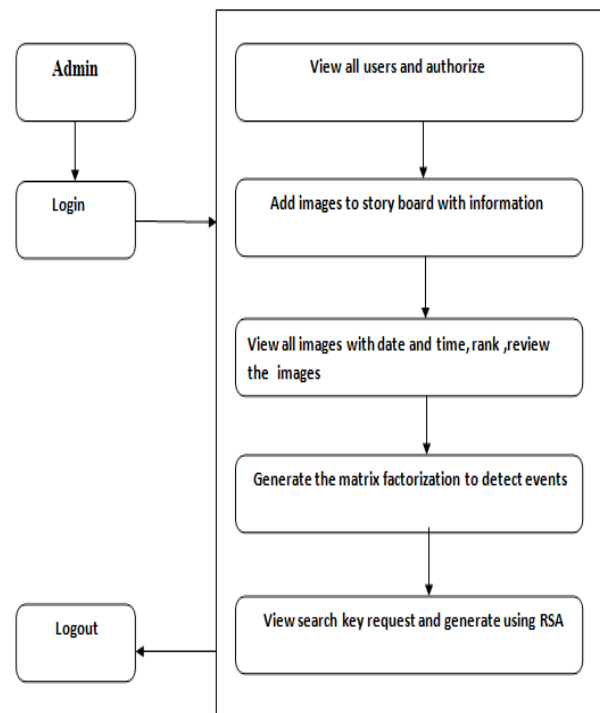


Figure 3.1Admin Architecture

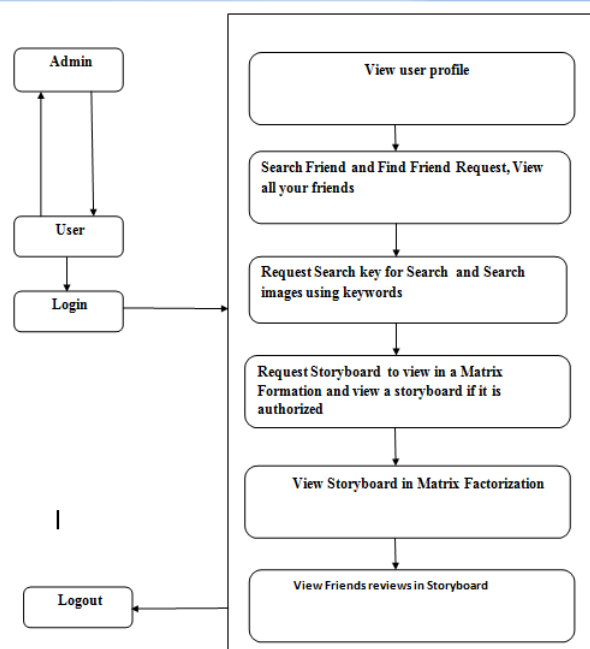


Figure 3.2 User Architecture

User will then select event of his choice by clicking on image and finally storyboard of that event will be displayed in the form of matrix as a result.

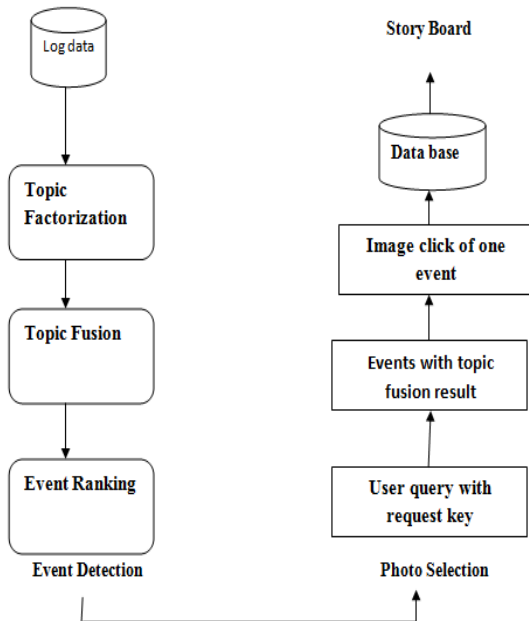


Fig 3.3 Event Detection and Photo Selection

Fig 3.3 shows the two approaches of SNMF algorithm.

**Event detection:** Initially events are searched from log data. Then it discovers groups of queries that have high frequency (frequent queries sent by user) which is known as Topic factorization. Topics with similar behaviors are merged together along a timeline which is called topic fusion. Ranking to each event is assigned which is given by user.

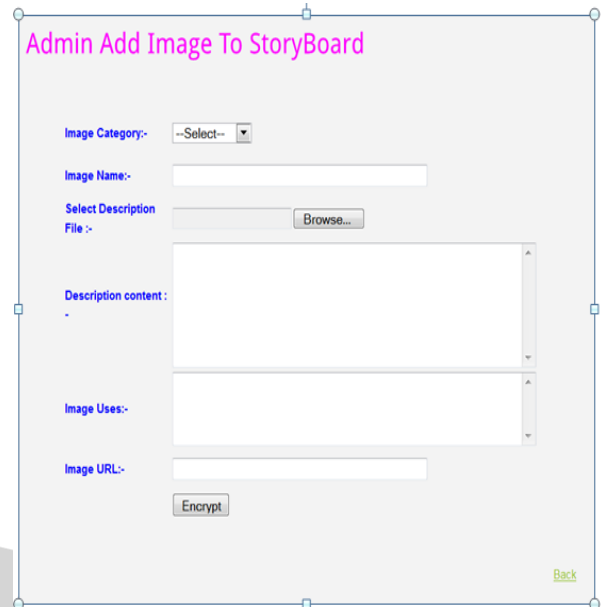
**Photo Selection:** User requests for event with the search key. As a result of topic fusion all related events will be displayed with an image. Then user will click on any one image. The storyboard of that image will be displayed as a result.

**ADVANTAGES**

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

**IV. EXPERIMENTAL RESULTS**

The admin will add all required information to the database such as image category, name description URL as shown in Figure 4.1. This information will be treated as storyboard for that particular image.



The form includes fields for: Image Category (dropdown), Image Name (text input), Select Description File (with a Browse... button), Description content (text area), Image Uses (text area), and Image URL (text input). There is an Encrypt button and a Back link.

Figure 4.1 Add images to storyboard



Figure 4.2 Social event Storyboard

Once the images are added by admin, he can view social event storyboard as shown in figure 4.2.

**Matrix Factorization Of Images**

Image ID	Image	Image Category	Image Name	Image Rank
36		Actors	<a href="#">aish</a>	2

**Matrix Factorization=1\*5**

Image ID	Image	Image Category	Image Name	Image Rank
11		Animals	<a href="#">Elephant</a>	11
12		Animals	<a href="#">Elephant</a>	4
23		Animals	<a href="#">Elephant</a>	5

**Matrix Factorization=3\*5**

Figure 4.3 Matrix Factorization of images

Based upon on user views of images the ranks for images will be assigned to each image through matrix factorization as shown in Figure 4.3. Matrix Factorization is done by considering log data that has n number of queries. Through this the events of same category will be grouped, as shown in Fig 4.3 the category name as animal consists only animal events even though database consists of events from different category such as celebrity, electronic gadgets etc.



Figure 4.4 User Search

User will get search key by admin once he gets authorized and with the help of this key he will be able to search through this system as shown in Figure 4.4. As a search result of topic fusion, it will display series of events for that particular search along with time line. User has to select any one event of his choice.

Once user selects event by clicking on particular image the story board will be displayed related to that event as shown in Figure 4.5



Figure 4.5 Story board

This system has been evaluated for 75 celebrities over 2 months of period and has shown good performance. In user study photo selection has high relevance as compared to google and bing search.

The usage of topic factorization method is successful in

delivering user satisfaction results as compared to traditional websites. It mainly depends on user query frequency to categorize events. Topic fusion will display the different events of same category along the time line which will make user to select the event chronologically.

## V. CONCLUSION

This system uses search logs as data source to generate social event storyboards automatically. With contrast to traditional websites where they display the search results in disorder manner, this system produces the results in chronological order. Users are allowed to choose the event through image click which produces the respective soryboard. The system uses Smooth Nonnegative Matrix Factorization (SNMF) which depends on event detection and photoselection. The event detection uses three methods namely Topic Factorization, Topic Fusion and Event Ranking which allows the information in databases to be arranged according to user query frequency and user rankings given to each event. Photo selection will allow user to get search result in chronological order and generate story board with respect to image click.

## REFERENCES

- [1] W. Weerkamp et al., "In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval" ACM, 2011
- [2] H. Liu et al., "Detecting and tracking topics and events from web search logs. ACM Transactions on Information Systems (TOIS)" ACM, 2012.
- [3] C.-C. Wue et al., "Learning to personalize trending image search suggestion. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval" ACM, 2014.
- [4] T. C. Walber et al., "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.
- [5] 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.
- [6] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In Advances in neural information processing systems, pages 556–562, 2001.
- [7] H. L. Chieu and Y. K. Lee. Query based event xtraction 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