INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 2)
Available online at: www.ijariit.com

Event detection and identification using social media data stream

Uday Pradip Patare
patareuday@outlook.com
Sanjivani College of Engineering, Kopargaon, Maharashtra

Abhishek Gajre
abhishekgajare17@gmail.com
Sanjivani College of Engineering, Kopargaon, Maharashtra

Ankita Mahajan
ankitamahajan84808@gmail.com
Sanjivani College of Engineering, Kopargaon, Maharashtra

Pradnya Patil
pradnyapatil303@gmail.com
Sanjivani College of Engineering, Kopargaon, Maharashtra

Y. S. Deshmukh
Yogeshdeshmukhit@sanjivani.org.in
Sanjivani College of Engineering, Kopargaon, Maharashtra

ABSTRACT

Today Social Media plays an important role in our life. Some popular social media platform such as Facebook, Twitter etc. Using these we can be connected with each other. We Create and share the Events on social media. Events such as Earthquake, conference, News, Weather news. To detect and identify the specific event from a group of the event is the crucial task. Traditional methods are not efficient for the detecting events and it requires more human intervention and having less accuracy. Our Aim is to detect and identify the events with less human intervention and higher the accuracy using LDA, HITS method.

Keyword— *Event detection, LDA, HITS, Twitter*

1. INTRODUCTION

From a few years, social media is rapidly growing up now lots of services existing such as Facebook, Twitter etc. Today Billions of people uses facebook, twitter hence the huge data is generated. These data consist the Images, Videos, Text Messages, and Events. On every minute's Number of events created and shared across the world through facebook, twitter. Detecting and monitoring the events is one of the most challenging tasks. Twitter is one of the most widely used services in just 780 days it crosses the 10 million users. According to this thousand of events created and shared across the world in just minutes. So that detecting an event is not easy for the traditional method [2] because this methodsdepend on human intervention and the accuracy is also less. According to Statistain India, Twitter has more than 30 million active users until December 2018.

2. EVENT DETECTION METHODS

In the field of data analysis, event detection plays an important role in monitoring the top hot events. Due to the older methods, internet watch officer can't track the hot events in a timely manner. To address this problem we use Hypertext-Induced Topic Search [HITS] based topic decision method and Latent Dirichlet Allocation [LDA][3] based three-step model.

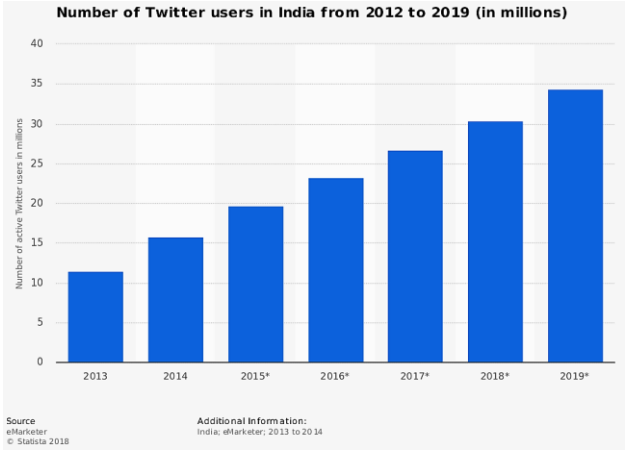

Fig. 1: Twitter user in India

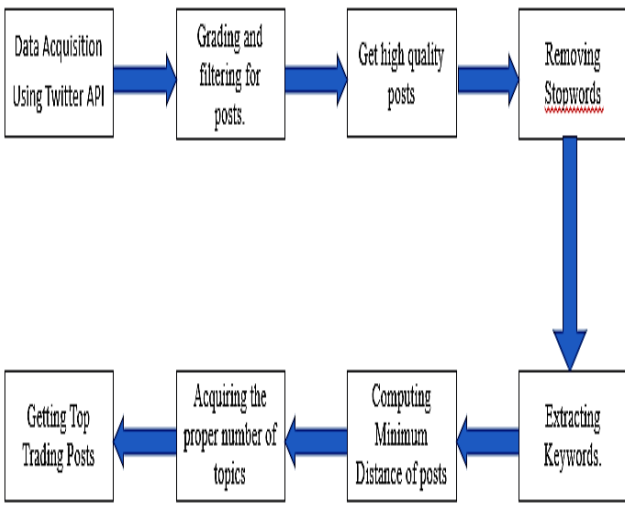

Fig. 2: HITS Procedure

2.1. Hypertext-Induced Topic Search [HITS]

It is based on topic decision method that can automatically detect the number of topics and identify key posts from among a large number of posts. creates a smaller high-quality training dataset by selecting high-quality posts and influential users from among a collection of users and posts, which largely reduces the impact of irrelevant posts and ordinary users, and improves the efficiency and accuracy of event detection compared with those of existing methods.

2.2. Latent Dirichlet Allocation [LDA]

LDA [3] was discovered by David Blei, Andrew Ng, and Michael l. Jordan in 2003 as a graphical model for topic discovery. LDA is a generative statistical model that allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar.

PLSA [2] and LDA are both widely used topic models for extracting hidden variables from a collection of posts. There are some similarities between these two models in their detection of events from among a large number of posts.

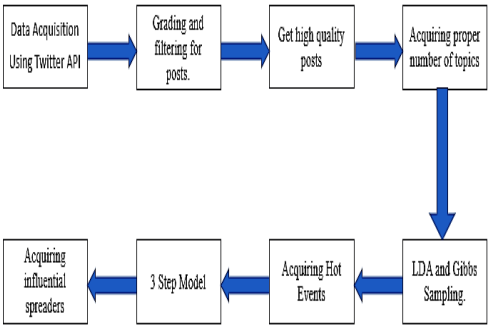

Fig. 3: LDA Procedure

2.2.1 Posts generated using LDA

- Choose a post di according to the prior probability P.di
- From the Dirichlet distribution, select the topic distribution I of the generated post di. In other words, the topic distribution i is generated by the super parameter in the Dirichlet distribution;
- Sample the topic zij of the jth word in the post di from the topic's polynomial distribution I;
- The Dirichlet distribution generates the
- distribution of the words d;
- Sample the distribution of words in the final set of words

Table 1: Notations Used in the LDA model

$\propto$	The super parameter of θm
β	Super parameter of ϕ_k
θ	Topic distribution
ϕ	Word distribution
d	Post
w	Word
Z	Number of topics
K	Number of words

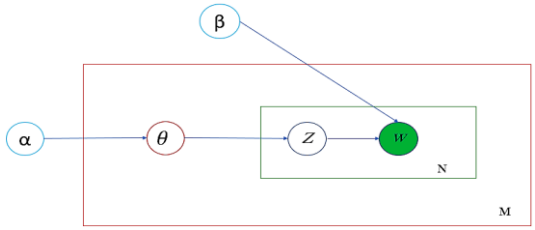

Fig. 4: Graphical Representation of LDA

2.3. Gibbs Sampling

Gibbs sampling is a Markov chain Monte Carlo (MCMC) algorithm for obtaining a sequence of observations which are approximated from a specified multivariate probability distribution when direct sampling is difficult. Gibbs sampling exhibits two exceptionally advantageous features. First, it provides a reliable level of accuracy as it asymptotically approaches the correct distribution. Second, it is more memory-efficient since they need only maintain the counters and state variables, which makes it the preferred method for dealing with large-scale datasets. A more detailed comparison of these methods can be found in Ref. [9].

3. RELATED WORK

As social media rapidly growing up researcher, learner, scientists take more interest in Event detection, Identification due to its openness (Twitter API, Facebook API). Social Media Data is open for through API.

3.1 Expert Finding

The task of expert finding has attracted extensive attention of information retrieval community since it was included in enterprise track

3.2 Identification of influential spreader

Most existing work on expert finding in social media focuses on the identification of influential spreaders.

3.3 Topic-specific expert finding

Several proposals approach the problem of topic-specific expert finding. Weng et propose an approach called Twitterrank which works in two steps.

4. APPLICATION OF EVENT DETECTION

Uses of event detection are as follows:

4.1. Network Monitoring

In today's informative world, network monitoring is of paramount importance. For instance, businesses are often interested in the frequency of visits to their websites and the general geographic locations of the visitor.

4.2. Traffic Monitoring

4.3. Health Monitoring and Management

The detection and prediction of conditions or events are also of extreme importance in healthcare applications. The Center for Disease Control and Prevention (CDC), for instance, continuously monitors medical and public health information from physicians and hospitals across the country.

4.4. Environmental Monitoring and Prediction

Environmental monitoring and prediction is another common area for the application of Event detection methods. The earth's environment can be extremely violent and early warnings of impending natural disasters such as hurricanes, tsunamis, earthquakes, floods are critical for the safety and security of populations within the affected regions.

5. CONCLUSION

This article defined event detection in the context of both natural and artificial systems. The primary challenges, methods, and applications of event detection were examined through research, examples, and literary references. Additionally, this article described the relationships between event detection and modelling and simulation. While much explanation and

numerous examples were provided, this paper only serves to introduce each of these topics. The domain of applicability of event detection and its associated methods is expansive and continually growing.

6. REFERENCES

[1] Event Detection and Identification of Influential Spreaders in Social Media Data Streams Leilei Shi, Yan Wu, Lu Liu_, Xiang Sun, and Liang Jiang IEEE BIG DATA MINING AND ANALYTICS ISSN 2096-0654 03/06 pp34– 46 Volume 1, Number 1, March 2018 DOI: 10.26599/BDMA.2018.9020004

[2] T. Hofmann, Probabilistic latent semantic indexing, in Proc. 22nd Annu. Int. ACM SIGIR Conf. Research and Development in Information Retrieval, Berkeley, CA, USA, 1999, pp. 50–57.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan, Latent Dirichlet allocation, J. Mach. Learn. Res., vol. 3, pp. 993–1022, 2003.

[4] L. AlSumait, D. Barbara, and C. Domeniconi, On-Line LDA: Adaptive topic models for mining text streams with applications to topic detection and tracking, in Proc. 8th IEEE Int. Conf. Data Mining, Pisa, Italy, 2008, pp. 3–12.

[5] IEEE TRANSACTIONS ON CYBERNETICS 1 Event Detection in Twitter Microblogging Nikolaos D. Doulamis, *Member, IEEE*, Anastasios D. Doulamis, *Member, IEEE*, Panagiotis Kokkinos, and Emmanouel (Manos) Varvarigos

[6] Location-based and Preference-Aware Recommendation Using Sparse Geo-Social Networking Data Jie Bao1∗Yu Zheng2 Mohamed F. Mokbel1† 1Department of Computer Science & Engineering, University of Minnesota, Minneapolis, USA2Microsoft Research Asia, No. 5 Danling Street, Haidian District, Beijing, China {baojie,mokbel}@cs.umn.edu, yuzheng@microsoft.com

[7] TopicSketch: Real-time Bursty Topic Detection from Twitter Wei Xie, Feida Zhu, Jing Jiang, Ee-Peng Lim, and Ke Wang. IEEE Transactions on Knowledge and Data Engineering.

[8] Towards a standard sampling methodology on online social networks: collecting global trends on Twitter C. A. Piña-García1*, Carlos Gershenson1,2,3,4,5 and J. Mario Siqueiros-García1. Springer 2016.

[9] A. Asuncion, M. Welling, P. Smyth, and Y. W. Teh, On smoothing and inference for topic models, in Proc.25th Conf. Uncertainty in Artificial Intelligence, Montreal, Canada, 2009, pp. 27–34.

