Document Based Clustering For Detecting Events in Microblogging Websites

[1]K. Prathyusha
[1] PG Student, Department of C.S.E
[1] SSITS, JNTUA University
Andhra Pradesh, India
[1] *usha.kalamalla@gmail.com*

[2] P. Subhan Basha
[2] Assistant Professor, Department of C.S.E
[2] SSITS, JNTUA University
Andhra Pradesh, India
[2] *subhan.mahammad@gmail.com*

Abstract— Social media has a great influence in our daily lives. People share their opinions, stories, news, and broadcast events using social media. This results in great amounts of information in social media. It is cumbersome to identify and organize the interesting events with this massive volumes of data, typically browsing, searching, monitoring events becomes more and more challenging. A lot of work has been done in the area of topic detection and tracking (TDT). Most of these methods are based on single-modality (e.g., text, images) information or multi-modality information. In the single-modality analysis, many existing methods adopt visual information (e.g., images and videos) or textual information (e.g., names, time references, locations, title, tags, and description) in isolation to model event data for event detection and tracking. This problem can be resolved by a novel multi-model social event tracking and an evolutionary framework not only effectively capturing the events, but also generates the summary of these events over time. We proposed a novel method works with mmETM, which can effectively model the social documents, which includes the long text along with the images. It learns the similarities between the textual and visual modalities to separate the visual and non-visual representative topics. To incorporate our method to social tracking, we adopted an incremental learning technique represented as mmETM, which gives informative textual and visual topics of event in social media with respect to the time. To validate our work, we used a sample data set and conducted various experiments on it. Both subjective and quantitative assessments show that the proposed mmETM technique performs positively against a few best state-of-the art techniques

Keywords- *document clustering, event tracking, detection, multi-modality,social networks..*

*****

1. INTRODUCTION

Social media has a great influence in our daily lives. People share their opinions, stories, news, and broadcast events using social media. This results in great amounts of information in social media. Methods to organize social media posts to support more informative views of data to users are needed so that users can easily find groups of posts that they are interested in. For example, clustering relevant topics together allows business users to go directly to the cluster of business related events. Many approaches for data mining and analysis for clustering and event detection in social media have been researched, but most of them consider content-based analysis or analysis using one type of data as a homogeneous network.

Microblogging, as a form of social media, is a fast emerging tool for expressing opinions, broadcasting news, and facilitating the interaction between people. The ease of publishing content on social media sites and the wide spread of various electronic devices (e.g. cellphones, tablets, etc.) have enabled users to report real-life events as they happen around them. One of the most representative examples of social media is Twitter, which allows users to publish short tweets (messages within a 140-character limit) about any subject. The range of widely known events includes community specific events, such as local gatherings, or can be wider-reaching national or even international in significance. For example, the Iranian election protests in 2009 were extensively reported by Twitter users. Another good example, where Twitter was employed as are source for the US government to communicate with citizens, was the outbreak of swine flu when the US Center for Disease Control (CDC) used Twitter to post the latest updates on the pandemic [1].

People tend to comment on real-world events they encounter, both local and global, when a topic suddenly attracts their attention, for example, a sporting event [2,3], adverse weather update [4], or terror attack [5,6,7] etc. The process of social media is shown in Figure 1. It would be of great benefit if we could get the evolutionary trends of social events and visualize the theme pattern overtime, which is the goal of event tracking and event evolution. Therefore, given an event initialized with the first story, we need to recognize which subsequent stories describe the same event and mine the event theme patterns and obtain the evolution process overtime, and then visualize these automatically. Recently, mining and monitoring social event in social media has attracted extensive research interests, such as social event mining [8], social event detection and tracking [9]–[11] and event evolution [12]. Different from the traditional event tracking and evolution problems, which generally involve a single modality such as textual information, social media data include unstructured metadata in multiple modalities.

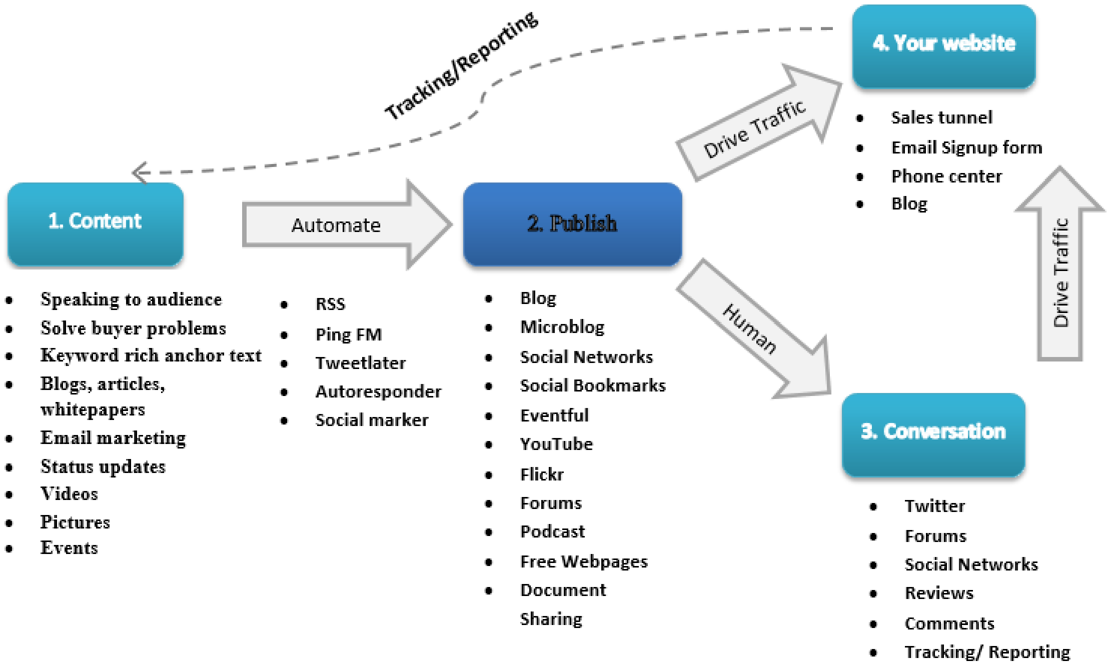

Figure 1: Social Media Process Flow

Generally, almost all existing work focuses on either textual features or images [13] in isolation. Limit efforts have been devoted to analyzing these multi-modality in a unified way to model multimedia event content. In different social media platforms, social media events have rich multi-modal information, such as text, images, and videos, which complement each other and are helpful for the social event analysis [14]. For example, given the same social events, they may have different textual information due to different users, but their visual information may be similar .Therefore, multi-modal feature fusion is useful for social event analysis [13]–[16]. Reuter and Cimiano [15] use multi-modal features to model the similarity of events and media data for their assignment. Though many multi-modal features, such as tag, time, location or visual features, are exploited, the importance and effectiveness of those features [16] have not been studied in details until now. Moreover, most of these previous methods focus on feature designing rather than modeling the textual and visual information jointly and ignoring the semantic relationship among multiple modalities of social events. Thus, it is necessary and challenging to explore an effective multi-modal fusion strategy for social event tracking and evolution analysis.

1.1 Motivation

The utilization of interpersonal interaction in spreading data of rising occasions, (for example, fires, bombings, cataclysmic events and malady flare-ups) and malevolent episodes has propelled an examination challenge for the early location of developing occasions and the following of a specific occasion. Rising occasions, for example, irresistible illnesses and the internet started/plotted assaults/turmoil should be identified in their beginning periods. Developing occasions like cataclysmic events may should be accounted for continuously when they are seen by individuals. In any case, with the extensive measure of short and loud messages at present accessible on informal communities, it is difficult to filter and deal with posts physically and in addition screen developing occasions as they are unfurling. Along these lines, having a component that can consequently play out this undertaking continuously would be extremely beneficial to administrative divisions, for example, debacle reaction and scourge anticipation offices. These profile highlights depict the client's tastes, inclinations, the gatherings he or she has a place with, and so forth. Informal organization Analysis (SNA) is a well known research field in which strategies are created for breaking down 1-mode systems, as friendto-friend1, 2-mode or alliance systems, 3-mode [17] and even multi-mode dynamic networks[18]. By multi-mode systems we mean to be specific such systems where performing artists can be connected with different sorts of substances by edges like those amongst clients and their interests in two-mode case or by hyper edges like those related clients, labels, and assets in three-mode case; now and then such systems are called heterogeneous since various kinds of hubs are included.

Some Questions?

We can automatically identify real-world events including disruptive events as they happen from posts on the social media in a particular place and for a predefined time period to improve public safety and decision support.

The research questions assist in understanding the scope of the work in this thesis. There are four main research questions:

Q1 Can we detect "events" in real time from the streaming media and introduce a strategy to integrate this knowledge into a Decision Support System (DSS)?

Q2 Can we identify sub-event details including disruptive events and their context within the streaming media (topic clustering)?

Q3 Since not all features are expected to improve a system's performance, can we investigate the dynamics of event/topic identification of three kinds of influential feature: temporal, spatial and textual, in order to optimize feature selection and to improve the effectiveness of topic clustering?
Q4 Can we summarize events to enable decision makers to read effectively only high quality summaries of most representative posts from Twitter?

2. Related Work

In this Section, we briefly review previous methods which are most related to our work including event tracking and topic model methods.

2.1 Social Event Detection and Tracking:

With the massive growth of social events in Internet, how to recognize and monitor social event becomes more and more challenging. A lot of work has been done in the area of topic detection and tracking (TDT) [18], [17], [10], [9], Most of these methods are based on single-modality (e.g., text, images) information or multi-modality information. In the single-modality analysis, many existing methods adopt visual information (e.g., images and videos) or textual information (e.g., names, time references, locations, title, tags, and description) in isolation [10], [9] to model event data for event detection and tracking

2.2 Event Summarization:

Multi-document summarization which addresses the information overload problem has drawn much attention in the past two decades. Gong and Liu [19] propose a generic text summarization method that creates text summaries by ranking and extracting sentences from the original documents. Haghighi and Vanderwende [20] present an exploration of generative probabilistic models by utilizing a hierarchical LDA-style model to represent content specificity as a hierarchy of topic vocabulary distributions for multi-document summarization Zhouetal.[21] present an ovel two-layer summarization framework to summarize multiple disaster-related documents. Wang et al. [22] propose a new multi-document summarization framework based on sentence-level semantic analysis and symmetric non-negative matrix factorization. However, most existing document summarization methods generate short summaries by selecting sentence from the text streams and ignore the rich visual information.

2.3 Social Event Analysis:

Sentiment analysis based on social event content has drawn much attention in determining sentiment from underlying text streams. Hu et al. [23] investigate whether social relations can help sentiment analysis by proposing a sociological approach to handle noisy and short texts for sentiment classification. In [8], the authors consider the problem of identifying the segments and topics of an event that garnered praise or criticism according to aggregated Twitter responses, and propose a flexible factorization framework to learn factors about segments, topics, and sentiments for event analytics via Twitter sentiment. The above methods mostly classify sentiments of document sources of the tweets around the event or provide insights into the event's segments and topics through the aggregated Twitter sentiment. In this paper, different from the traditional methods, which conduct the sentiment analysis based on social event content, we focus on social event tracking and evolution analysis via multi-modal event topic model learning.

2.4 Topic Model:

Topic models, such as Latent Dirichlet Allocation (LDA) and Probabilistic Latent Semantic Analysis (PLSA) have been widely applied to various applications and have many extensions, such as supervised Latent Dirichlet Allocation (SLDA), dynamic latent Dirichlet allocation and its variations. Blei and Lafferty propose the dynamic topic model method which uses state space models on the natural parameters of the multinomial distributions that represent the topics to analyze the time evolution of topics in large document collections. AlSumait et al. propose an online LDA method, which extends the Gibbs sampling method and utilizes it to derive hyper parameters of the topic-word distribution at the next time slice.

3. Clustering

Bunching is the undertaking of collection an arrangement of articles such that items in a similar gathering (called a group) are more like each other than they are to those in different gatherings (bunches), as indicated by some separation measure. From a machine learning point of view, the look for groups is the unsupervised learning of concealed examples in vast informational collections or databases. When all is said in done, bunching is a typical procedure for factual information examination, and is utilized as a part of numerous fields, including machine learning, design acknowledgment, picture investigation, data recovery, bioinformatics, information pressure, and PC illustrations.

Algorithm1: Online Clustering Algorithm
Input: n set of documents (D1,...,Dn)
Threshold τ
Output: m clusters (C1... Cm)
Step1: Fora given τ, compute the centroid similarity function E (Di,Cj) of each cluster Cj
Step2: If centroid similarity E(Di,Cj)$\geq\tau$
Do:
1) A new cluster is formed containing Di ;
2) The new centroid value = Di.
Step3: If centroid similarity E(Di,Cj) < τ
Do:
1) Assign it to the cluster which gives the maximum value of E(Di,Cj) ;
2) Add Di to cluster j and recalculate the new centroid value Cj.

More specifically, it plays an outstanding role in data mining applications such as scientific data exploration, information retrieval and text mining, spatial database applications, Web analysis, and many others. Clustering based methods are also widely used for the task of news topic detection. They assume that each news story talks about one news event which corresponds to one topic .A large number of methods for topic detection in the Topic Detection and Tracking (TDT) research. We have shown a sample model of generic online clustering algorithm in Algorithm 1.

3.1 Multi Model Clustering

Multimodal clustering is an unsupervised technique for mining interesting patterns in n-adic binary relations or n-mode networks. Among different types of such generalized patterns one can find bi clusters and formal concepts (maximal bi cliques) for 2-mode case, tri clusters and tri concepts for 3-mode case, closed n sets for n-mode case, etc. Object-attribute bi clustering(OA-bi clustering) for mining large binary data tables (formal contexts or 2-mode networks) arose by the end of the last decade due to intractability of computation problems related to formal concepts; this type of patterns was proposed as a meaningful and scalable approximation of formal concepts. In this paper, our aim is to present recent advance in OA bi clustering and its extensions to mining multi-mode communities in SNA setting. We also discuss connection between clustering coefficients known in SNA community for 1-mode and 2-mode networks and OA-bi cluster density, the main quality measure of an OA-bi cluster. Our experiments with 2-, 3-, and 4-mode large real world networks show that this type of patterns is suitable for community detection in multi-mode cases within reasonable time even though the number of corresponding n-cliques is still unknown due to computation difficulties. An interpretation of OA-bi clusters for 1-mode networks is provided as well.

3.2 Story Clustering

These days a lot of TV channels communicate news 24 hours per day. We can without much of a stretch get news by turning on the TV. Be that as it may, ordinary just many stories are new to the group of onlookers, and in most time similar stories rehash and once more, while interleaving with varieties of old and crisp substance. In this way, bunching point related news stories is fascinating and direly requesting as it is the central advance for news perusing, recovery, and rundown it merits clearing up some related wordings. A news shot is a video shot containing some portion of news content, which might be shots with the anchorperson, meeting, or occasions like motorcade or auto. A news story may incorporate a few news shots to totally pass on a message. It frequently contains a succession of shots with the anchorperson and the occasion itself, and finishes until the point when the anchorperson reports the following message. A news theme may contain a few news stories portraying development of an occasion after some time, for example, "Congressman shooting" or "Steve Jobs' wiped out leave.

3.3 Challenges in Event Detection

There are various research challenges characteristic in occasion recognition. We legitimize why off-the rack information and content mining approaches are not reasonable for handling occasion identification. Volume and Velocity. Information from online networking come in awesome volume and speed. In this way, calculations ought to be on the web and adaptable in memory and computational assets. High information volume makes bunch handling computationally infeasible. Ongoing Event Detection. Occasions ought to be identified at the earliest opportunity, particularly when the approach is expected to be utilized as a part of basic applications like crisis reaction. For this situation, techniques for occasion recognition ought to be assessed as far as Precision and Recall as well as regarding how quick they can distinguish a specific sort of occasion. Online networking are filled with spam messages, notices, bot accounts that distribute huge volumes of messages, fabrications, and web images. Another deterrent is that literary data in internet based life is extremely constrained. Clients typically distribute short messages a reality that makes off-the-rack Text Mining and NLP techniques inadmissible. Highlight Engineering. Choosing the most appropriate highlights to use in directed or unsupervised learning segments is certainly not a minor undertaking. Printed portrayals, for example, Term-Document grids are not sufficient. The same number of specialists have seen, there are specific attributes that show up in occasion related messages. These highlights could be content-based traits, for example, TF-IDF scores, number of labels and emojis or basic highlights like the quantity of devotees (Twitter) or companions (Facebook). Administered approaches for the most part center around content highlights keeping in mind the end goal to prepare classifiers, for example, Naive Bayes or Support Vector Machines. Numerous specialists have presumed that the usage of the right list of capabilities is extremely pivotal for the occasion location process. For instance, Becker et al. presents a correlation between basic highlights and Term-Document lattices. Be that as it may, the nature of this information is very different (point location and following) and subsequently it serves just as final resort. Results got from TDT5 could significantly differ from those acquired from Twitter or Facebook. TDT5 originates from news-wire articles and contains all around framed superb content. Then again, web based life content has one of a kind printed attributes including shortenings, utilization of slang dialect and incorrect spellings. An open dataset assembled from online networking sources is imperative since it could be utilized to prepare managed classifiers and furthermore assess the calculations as far as Precision or Recall.

3.4 Event Detection and Categorization

The term 'occasion' can be defined in various routes, contingent upon the spaces and the enthusiasm of the clients or leaders. The definition of an occasion changes in granularity too, contingent upon the manner by which the occasion recognition will be connected. Our concentration is to define and associate these definitions to our assignment of distinguishing and portraying occasions in the internet based life. The objective of recognizing occasions and their related records via web-based networking media locales is to screen constant internet based life streams and concentrate data. We have demonstrated the free internet based life checking instruments in Table 1.

Table 1: Free Social Monitoring and Tracking Tools [24]

Tools	Description	URL	Measures
Google	Monitors Web for brand and keyword mentions. Sends email alerts.	www.google.com/alerts	Social Engagement Social Advocacy
Social mention	Monitors Social Web and blogs for brand mentions. Send email alerts.	www.socialmention.com	Social Engagement Social Advocacy
Facebook Insights	Track interactions with fans and content within Facebook. Must be page admin to view.	www.facebook.com/insights	Social Views Social Engagement Social Advocacy
Klout	Measures an individual social influence. Want high klout score folks recommending brand.	www.klout.com	Social Advocacy
Bitly	Link shortening tool that makes it easy to track and analyze clicks and referral sources.	www.bitly.com	Social Referrals
Tweetbeep	Monitors Twitter for brand and keyword mentions. Sends email alerts.	www.tweetbeep.com	Social Engagement Social Advocacy
Google Analytics	Website and blog analytics that tracks unique visits; and inbound clicks from social referrals.	www.google.com/analytics	Social Audience Social Views Social Referrals
YouTube Insights	Website and blog analytics that tracks unique visits; and inbound clicks from social referrals.	www.youtube.com/t/advertising_insight	Social Audience Social Views

3.5 Feature Selection

Highlight determination is a central issue in mining huge informational indexes. The issue isn't constrained to the aggregate handling time yet includes dimensionality lessening to accomplish better speculation. Highlight choice is a compelling method for lessening dimensionality, evacuating superfluous information, and expanding learning exactness. We register the highlights of Twitter message bunches so as to uncover qualities that may help identify the groups that are related with occasions, especially troublesome occasions. Not all highlights are required to enhance the framework's execution or prompt more exact separation of the bunching calculation.

3.6 Temporal Features

Time is a critical measurement of any data space. It can be exceptionally valuable for an extensive variety of data recovery assignments, for example, record investigation, comparability hunt, outline, and bunching.

Spatial Features (Geo-spatial, Regional)

Several algorithms have been proposed to estimate the location of Twitter users by means of a content analysis of tweets.

Textual Features

Textual features can be used as individual features (e.g. n-grams), but many studies have combined them to optimize the solution to data mining challenges, such as information diffusion, opinion mining, spam and spammer detection, and identifying the most knowledgeable posts and influential users .

4. Proposed Work

The proposed work is contributed as following:

1) Our multi-modular occasion following and advancement system is reasonable for sight and sound records from different online networking stages, which can viably catch their multi-modular themes, as well as get the transformative patterns of get-togethers and produce viable occasion synopsis points of interest after some time.

2) Our proposed mmETM model can abuse the multi-modular property of get-together, which can successfully display online life archives incorporating long content with related pictures and take in the relationships amongst's printed and visual modalities to isolate the visual-agent point sand non-visual-delegate subjects.

3) We propose a novel incremental mmETM show for get-together following, which can get the entire transformative procedure of occasions with literary and visual subjects after some time and help comprehend the occasions.

4) We gather a substantial scale dataset for examine on get-together following with multi-methodology data, and will discharge it for scholastic utilize. We assess our proposed display and exhibit that it accomplishes much preferable execution over existing techniques [25].

4.1 Multi-Modal Event Topic Model

It is well observed that the dependency relationships between textual and visual modalities of different semantic concepts are different. In most of event documents containing long text and the corresponding images, some semantic topic descriptions can be represented well by both textual and visual content, and some topic descriptions are appropriately represented by textual content and have no clear visual correspondence, such as economy, history ,etc. For the social event documents, Incremental Parameter Inference to track multiple events over time, we present an incremental inference method for our proposed mmETM that sequentially updates the model at each epoch using the newly obtained event documents and the parameters of the previous epoch. Since an event consists of many stories over time, the incremental updating strategy allows our proposed mmETM model to work in an online mode. Specifically, in the mmETM at an epoch, the count of words in topics is used to construct priors at the following epoch.

4. 2 Event Tracking

Through the incremental mmETM model learning, we can obtain the document-topic distributions for each social event over times. In order to track multiple events, we apply the similarity computing identification method in to our incremental mmETM model. Based on the learned mmETM model at epoch, multiple event documents can be classified into their corresponding events at epoch using the similarity computing identification method. As a result, multiple events can be tracked over time. The complete flow of this work is shown in Figure 2.

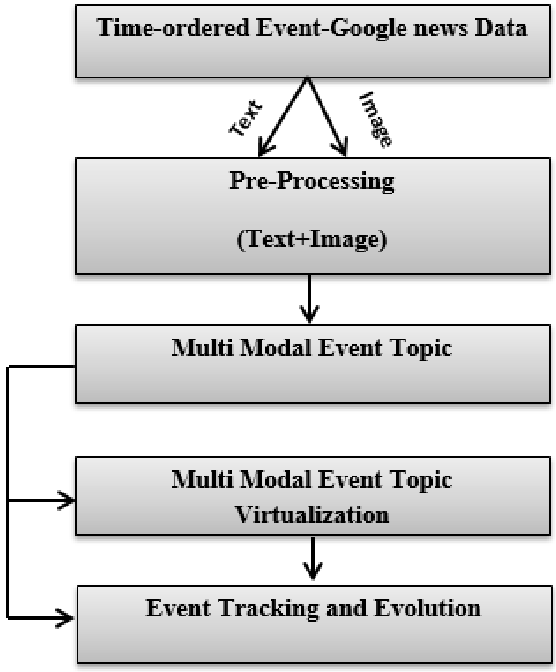

Figure 2: Multi-modal event tracking and evolution framework

5. Results and Discussion

In this section, we are going to describe our experimental study in brief. The work is done on a sample dataset, which contains different set of documents of three groups. The set up was done with the help of java programming language with a back end of MYSQL and Apache tomcat is used as a server. The work is comprised in three modules and each module can do different functionalities. The aim of our work is to cluster the events or upcoming events on social media website on different user's accounts. For controlling the activities between the media and user we used as intermediate software/person as admin. The admin is able to add the media and he is also able to accept the membership of the user and can delete the membership too. For example a news channel can be registered at the server and is now capable of uploading any events or upcoming events conducted in real time. The users can register in to the social media website and they can able to get the events from the registered media. The challenging task here is when any document is uploaded by the media then this document is placed in to the respective cluster in the website. For example if there are three categories like cricket, FIFA, Olympics. There may 'n' number of sub categories in each of these main categories. For example, in cricket there will be T20 World Cup, One Day International World Cup, and Indian Premiere League etc. In the user's site, there will be three categories, and the document is placed in correct category. From the user end, when a document is uploaded by the media about IPL CSK team specific event, then it should come under the category of sub cluster of the Cricket category. For example take a category of swimming in Olympics it is categorized in to three sub clusters and it is shown in Figure 3.

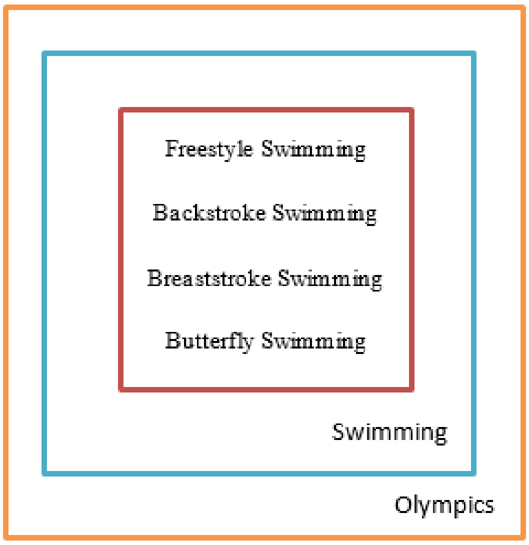

Figure 3: Sample categories of clusters of our work.

6. Conclusion

Microblogging, as a form of social media, is a fast emerging tool for expressing opinions, broadcasting news, and facilitating the interaction between people. The ease of publishing content on social media sites and the wide spread of various electronic devices (e.g. cellphones, tablets, etc.) have enabled users to report real-life events as they happen around them. In the first section, we have presented an overview of the importance of event clustering in social media and motivation of our work along with few research issues. In the next section, we presented the methods for event tracking

ISSN: 2454-4248
650 – 656

and detection along with literature work. Then, the clustering and the need of multi-modality clustering is introduced. Our proposed work comprises that the documents are clustered based on the multi-modality property. It is an unsupervised clustering technique for mining interesting patterns in n-adic binary relations or n-mode networks. We considered the sample documents as our input and the in future we apply our prototype to the real time dataset and will compare our results with the state of the art techniques.

REFERENCES

[1]. Joshua Ritterman, Miles Osborne, and Ewan Klein. Using prediction markets and twitter to predict a swine flu pandemic. In In Proceedings of the 1st International Workshop on Mining Social, 2009.

[2]. Nasser Alsaedi, Pete Burnap, and Omer Rana. A combined classification clustering framework for identifying disruptive events. In Proceedings of the 7th ASE International Conference on Social Computing, Stanford University, CA., USA, Social Com '14, 2014.

[3]. Munmun De Choudhury, Nicholas Diakopoulos, and Mor Naaman. Unfolding the event land scape on twitter: Classification and exploration of user categories. In Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW '12, pages 241–244, New York, NY, USA, 2012. ACM.

[4]. Alexandra Olteanu, Carlos Castillo, Nicholas Diakopoulos, and Karl Aberer. Comparing events coverage in online news and social media: The case of climate change. In Proceedings of the Ninth International Conference on Web and Social Media, ICWSM '15, pages 288–297, 2015.

[5]. Alexandra Olteanu, Sarah Vieweg, and CarlosCastillo. What to expect when the unexpected happens: Social media communications across crises. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15, pages 994–1009, New York, NY, USA, 2015. ACM.

[6]. Onook Oh, Manish Agrawal, and H. Raghav Rao. Information control and terrorism: Tracking the Mumbai terrorist attack through twitter. Information Systems Frontiers, 13(1):33–43, mar 2011.

[7]. Jim Maddock, Kate Starbird, Haneen J. Al-Hassani, Daniel E. Sandoval, Mania Orand, and Robert M. Mason. Characterizing online rumoring behavior using multi-dimensional signatures. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15, pages 228–241, New York, NY, USA, 2015. ACM.

[8]. D.Patel,W.Hsu, and M.L.Lee,"Mining relationships among interval based events for classification,"inProc.SIGMOD,2008, pp.393–404.

[9]. J.Allan,R. Papka, andV.Lavrenko,"On-line new event detection and tracking," in Proc. SIGIR, 1998, pp. 37–45.

[10]. M. Merler, B. Huang, L. Xie, G. Hua, and A. Natsev, "Semantic model vectors for complex video event recognition," IEEE Trans. Multimedia,vol.14,no.1,pp.88–101,Feb.2012.

[11]. T. Zhang and C. Xu, "Cross-domain multi-event tracking via CO-PMHT," ACM Trans. Multimedia Comput. Commun. Appl., vol. 10, no. 4, pp. 31:1–31:19, 2014.

[12]. Singh,A.Gupta, and A.A.Efros, "Unsupervised discovery of mid-level discriminative patches, "in Proc. ECCV, 2012 ,pp. 73–86.

[13]. L. Xie et al., "Discovering meaningful multimedia patterns with audio-visual concepts and associated text," inProc. ICIP, 2004, pp. 2383–2386.

[14]. S. Qian, T. Zhang, R. Hong, and C. Xu, "Cross-domain collaborative learning in social multimedia,"in Proc. ACMMM, 2015, pp.99–108.

[15]. T.Reuter and P.Cimiano, "Event-based classification of social media streams," in Proc. ICMR, 2012, pp. 22:1–22:8.

[16]. X. Liu and B. Huet, "Heterogeneous features and model selection for event-based media classification," in Proc .ICMR, 2013, pp.151–158.

[17]. X. Ding, B. Liu, and P. S. Yu. A holistic lexicon-based approach to opinion mining. In Proceedings of the International Conference on Web Search and Web Data Mining, WSDM 2008, Palo Alto, California, USA, February 11-12, 2008, pages 231–240, 2008.

[18]. News Poll. Two party preferred. http://polling.newspoll.com.au.tmp.

[19]. Y.Gongand X.Liu, "Generic text summarization using relevance measure and latent semantic analysis," in Proc. SIGIR, 2001, pp. 19–25.

[20]. A. Haghighi and L. Vanderwende, "Exploring content models for multi-document summarization," in Proc. NAACL, 2009, pp.362–370.

[21]. W. Zhou, C. Shen, T. Li, S. Chen, and N. Xie, "Generating textual storyline to improve situation awareness in disaster management," in Proc. IRI, 2014, pp. 585–592.

[22]. D.Wang,T. Li,S.Zhu, and C.Ding, "Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization," in Proc. SIGIR, 2008, pp. 307–314.

[23]. X.Hu,L.Tang,J. Tang,andH.Liu, "Exploiting social relations for sentiment analysis in microblogging, " in Proc. WSDM, 2013, pp.537–546.

[24]. Wei-Ta Chu, Chao-Chin Huang, Wen-Fang Cheng. "News story clustering from both what and how aspects", Proceedings of the 2011 ACM international workshop on Automated media analysis and production for novel TV services - AIEMPro '11, 2011.

[25]. Shengsheng Qian, Tianzhu Zhang, Changsheng Xu. "A Generic Framework for Social Event Analysis", Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval - ICMR '17, 2017.

Authors Profile:

K. Prathyusha is a PG scholar in the department of Computer science and Engineering in SSITS, JNTUA. She received her Bachelor's degree in 2015 from JNTUA University in the stream of Computer Science and Engineering. Her research interests include Data mining and computer networks.

P. Subhan Basha is currently working as Assistant Professor in the department of Computer Science and Engineering in SSITS, JNTUA. He received his master's degree from Anna University in 2010 and Bachelor's degree in 2005 from JNTUH. He has 9 years of teaching experience and has taught various subjects in computer science stream His research interests are Computer Networks, Data mining and Network Security. He also published various National and International Journals.