International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Term Document Frequency (TDF) Method for Extracting User Posts and Emerging Events in Social Networks

S.S.Manivannan, M.Iyapparaja, M.Prasanna, T.Velmurugan, J.Prabhu

Abstract: ***Social network is a place where people exchange and share data related to the current trends and events all over the world. This specific behavior of users made us concentrate on the logic that processing these data may lead us to the extracting the current topic of curiosity between the users. Applying data clustering technique like Term-document-Frequency (TDF) based approach over these data may leads us up to the mark but there will be little chance of negatives. We are going to do a likely medium that can give both usual mentioning behaviour of a consumer and also the frequency of users occurring in their mentions. It also works well even the data of the messages are very small information. These extracted emerging topic are shown to the user those who subscribe for the details.***

Keywords ***: Mentions, Mentionee, Emerging Events, Social Network.***

I. INTRODUCTION

Information exchanged over social networks has become a common activity among people all over the world. As this information is real time in nature processing these data can bring us some important information earlier than any other media. As these information are human generated they are full of noise and understanding similar context from the content is really hard. If we failed to understand the context and relationship between the users post properly than your extracted topic will be wrong

II. LITREATURE SURVEY

The proposal of emerging pictures has turned into a prevalent element utilized by business web indexes to pull openly consideration. By surfing through emerging pictures, web index clients can find emerging occasions initially. But, the choice of trending pictures is exceptionally testing but it's a kind of open problem. Most existing work is profoundly reliant on article endeavours; however little to begin with recognize a couple plain components for inclining pictures. This is one of the idea, we research an arrangement of elements that will recognize slanting pictures regular ones. We made an arrangement of emerging mindful elements in view of a few regular criteria, which mirror the qualities of inclining pictures. We additionally develop a physically marked dataset in light of a business web crawler's inquiry for a last two weeks. We get and our idea technique on the dataset and also the outcomes exhibit its adequacy [2].

Twitter is a well known online networking site, where a huge number of tweet about certain theme amongst event of occasions. This outcome in slanting subjects for example #MH370, #MH17, #South Korea ship and so forth. When review slanting point, not every one of the tweets are pertinent to one self. In shows this issue, the experiment concentrates in the mechanized personalization of tweets for mainstream inclining themes. Its especially target on the data which as "Like" or Aversion" specific theme contingent upon individual inclination by highlight choice. For improvement, subject related watchwords are chosen as elements foe speaking to a classification display from client favoured to the posts and also variety of news. At last, after all this process it gives the effective results this gives the idea for preparation in light of little number of posts cans giv snappy adaptive on the sequence approaching posts [4].

People in general nature of client produced content via web-based networking media stages offers the likelihood for pattern checking as an understanding into the subjects that pull in the consideration of an extensive part of clients. While Google Trends and Twitter have as of now been perceived as an important wellspring of pattern data by the experts and researchers, right now there are no reasonable executions or research endeavours in the field of pattern recognition over Face book open posts.

Manuscript published on November 30, 2019.
* Correspondence Author
S.S.Manivannan*, Associate Professor, SITE School, Vellore Institute of Technology, Vellore, Tamilnadu, INDIA [4]Associate professor , SITE School, Vellore Institute of Technology, Vellore, Tamilnadu, INDIA
manivannan.ss@vit.ac.in, iyapparaja.m@vit.ac.in
M.Iyapparaja, Associate Professor, SITE School, Vellore Institute of Technology, Vellore, Tamilnadu, INDIA [4]Associate professor , SITE School, Vellore Institute of Technology, Vellore, Tamilnadu, INDIA
manivannan.ss@vit.ac.in, iyapparaja.m@vit.ac.in
M.Prasanna, Associate Professor, SITE School, Vellore Institute of Technology, Vellore, Tamilnadu, INDIA [4]Associate professor , SITE School, Vellore Institute of Technology, Vellore, Tamilnadu, INDIA
manivannan.ss@vit.ac.in, iyapparaja.m@vit.ac.in
T.Velmurugan, Associate Professor, SITE School, Vellore Institute of Technology, Vellore, Tamilnadu, INDIA [4]Associate professor , SITE School, Vellore Institute of Technology, Vellore, Tamilnadu, INDIA
manivannan.ss@vit.ac.in, iyapparaja.m@vit.ac.in
J.Prabhu, Associate Professor, SITE School, Vellore Institute of Technology, Vellore, Tamilnadu, INDIA [4]Associate professor , SITE School, Vellore Institute of Technology, Vellore, Tamilnadu, INDIA
manivannan.ss@vit.ac.in, iyapparaja.m@vit.ac.in

In this paper we display two commitments towards pattern observing over Face book open posts. We propose and assess a framework for pattern recognition in view of the attributes of the posts share on Face book.

In view of our outcomes we propose three classes of inclining points: 'troublesome occasions', 'famous subjects' and 'day by day schedules'. We investigate and analyze the qualities of the proposed classifications as far as appropriation and data dispersion keeping in mind the end goal to expand the comprehension of rising patterns on Face book. At last we reach inferences from our discoveries regarding difficulties and open doors for future work towards this path [5].

Identification of rising subjects is currently accepting restored intrigue inspired by this fast growth of this social networks and sites.

The conventional-term-frequency is a technique which is not fit in this unique environment, in light of the fact that this information is being exchanged in social network data that are mixed as a text and also a pictures. Online links and recordings. We give attention for the get up of recurring ideas in the social partition of the theories planning. In this specific, we give attention on notices of client's connection that are produced powerfully through notices and also posts. We give a new model for saying control of the social organize user and new idea to identify the development of the different point from the model which is measured. Aggregating anomaly rates from various consumers, we demo that can be pinpoint from the developing points that will shows some answers say links in some data in the posts that we can exhibit few works in the genuine informational collections that can be accumulated in the social network. Investigations demos the new idea say peculiarity that is built as the method to be identified updated themes in the given experience as along of the scheduled data irregularity focused methods and in a little cases considerably advanced the particular subject is insufficient distinguished by the substances in the posts [6].

Connect Anomaly identification is a standout amongst the most imperative themes in informal community. A number of the informal organizations for example FB, Google+ require a compelling and productive structure to distinguish veered off information. Insistency recognition strategies are commonly actualized in social stream mode, and hence can't be effectively reached out to substantial scale issues without giving up calculation where the client's connection is created progressively (answers, notices and retweets). Another approach show i.e. likelihood display, this model to the catch typical connecting conduct of an informal community clients and propose to identify the emerging point from the interpersonal organizations through the likelihood demonstrate. We gather peculiarity score from the diverse client. What's more, collected score encourage to change-point investigation or with burst recognition, at last demonstrate that to distinguish emerging themes proper in light of the specify in everyday community share. Our strategy is to gather number of genuine information from ongoing twitter account [7].

III. OBJECTIVE:

The objective of this project is to find the emerging events in the social media. The posts which are posted by the users can be extracted from the database and analysed from that we can find the emerging events or topic.

A. *Existing System:*

In existing trend or topic extracting process is done based on the higher interaction of the particular post in social network as an emerging topic. If a particular post has been shared by multiple users and commented by a multiple users than that particular post is considered an emerging topic.

It supports all kinds of post, It can be an image, textual data or a video.

B. *Disadvantages:*

An emerging or a trending topic is not to be a single user and share and commented by other, any user can post same topic. Because the process doesn't check the relationship of one post with the other, there will be a deviation in the result. Sometime a main emerging topic becomes unnoticed when it is posted by multiple users and share and comment by small groups of people.

IV. PROPOSED SYSTEM:

In this we are going to propose a new idea that can show the usual mentioning activity of the consumer and also verifies the relationship between different posts. It gets the number of mentions in the particular post and also the consumer occurring in that mentions it also gives the interaction of all the post. As emerging topics rely on time we consider only mentions (comments and share) at a given time instance. We use term document frequency algorithm for understanding the relationship or context between posts

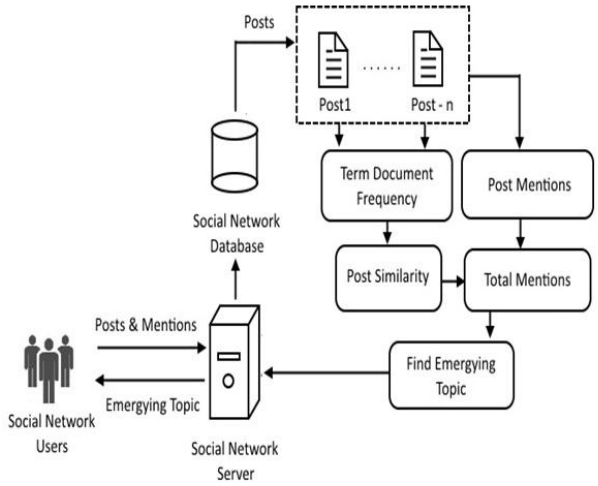
TERM FREQUENCY

```
Input:  Set of Mention Content ( C_post ) belong to Post P_i
        Dictionary Word ( D )
Output: Mention_Vectpr (V_post )
Start
        String finalContent C_F
        getDictionaryWord( )
        foreach word Wpost in C_post
           foreach W_dic in D
           // where W_dic is element of Dictionary Word
              if  equals(W_dic , Wpost) Then
                  concat(C_F, Wpost)
             end If
          end foreach
        end foreach
        arraylist Mention_vector V_post
        foreach W_F in C_F
        // where W_F is element of finalContent C_F
           If  Mention_vector.Contains(W_F) then
                 Mention_vector[W_F] = Mention_vector[W_F]+1
           End if
           Else then
                 Mention_vector[W_F] = 1
          End else
        End foreach
        Return Mention_vector
End
```

V. ADVANTAGES:

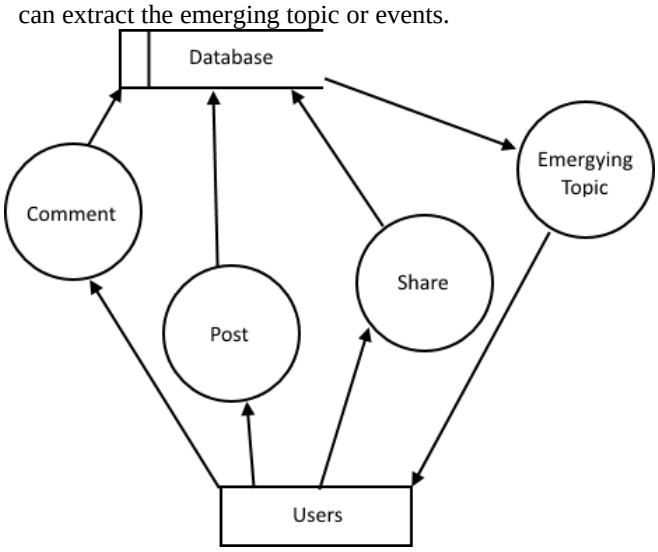
Our proposed system understands the context between the post which avoids the deviation of extracting the emerging topic. It can also handle heterogeneous post like text, images and video. Time instance are strictly followed to find the current emerging topics which means older mentions are not considered for topic extraction process.

VI. SYSTEM ARCHITECTURE:



VII. DATA FLOW DIAGRAM:

This flow diagram shows the data flowing between the users and the database in these as many events can be commented, posted, shared by many users that will be stored in the database. From that analysing all the post we can extract the emerging topic or events.



VIII. CONCLUSION:

Our proposed work finds an emerging topic in social media by calculating the interaction of the all particular posts and also the users who are all interacted and commented in the post. We had demonstrated that using our proposed scheme we can find emerging topic not only in the text but in images and videos. As a further enhancement we had used term document frequency algorithm to combine a different post with similar context by doing this we avoid deviation in the text based emerging topic extraction.

REFERENCES:

1. Event identification in social media through latent dirichlet allocation and named entity recognition by G. Abinaya Department of Computer Science and Engineering, Saveetha Engineering College, Chennai, Tamil Nadu, India and S. Godfrey Winster Department of Computer Science and Engineering, Saveetha Engineering College, Chennai, Tamil Nadu, India.
2. On the selection of trending image from the web by Dongfei Yu University of Science and Technology of China, China, Xinmei Tian University of Science and Technology of China, China and Tao Mei Microsoft Research, United States.
3. Cluster-discovery of Twitter messages for event detection and trending by Shakirabanu Kaleel Masters in Computer Science at Ryerson University, Toronto, Canada and Abdolreza Abhari Department of Computer Science at Ryerson University in Toronto, Canada.
4. Personalization of Trending Tweets using Like-dislike category model by Lu Weilin,Gan Keng Hoon, School of Computer Sciences, University Sains Malaysia.
5. Iyapparaja M, Tiwari. M, Security policy speculation of user uploaded images on content sharing sites , IOP Conf. Series: Materials Science and Engineering 263 (2017) 042018 doi:10.1088/1757-899X/263/4/042019,pp-1-8
6. Discovering Emerging Topics in Social Streams via Link-Anomaly Detection by Toshimitsu Takahashi, Ryota Tomioka and Kenji Yamanishi, Member, IEEE.
7. Iyapparaja M, Bhanupriya Sharma, Augmenting SCA project management and automation Framework, IOP Conf. Series: Materials Science and Engineering 263 (2017) 042018 doi:10.1088/1757-899X/263/4/042018,pp-1-8 [8] J. Allan et al., "Topic Detection and Tracking Pilot Study: Final Report," Proc. DARPA
8. Broadcast News Transcription and Understanding Workshop, 1998.
9. Iyapparaja M et.al. 2012 Coupling and Cohesion Metrics in Java for Adaptive Reusability Risk Reduction IET Chennai 3rd International Conference on Sustainable Energy and Intelligent Systems (SEISCON 2012),52-57.
10. Y. Urabe, K. Yamanishi, R. Tomioka, and H. Iwai, "Real-Time Change-Point Detection Using Sequentially Discounting Normalized Maximum Likelihood Coding," Proc. 15th Pacific-Asia Conf. Advances in Knowledhe Discovery and Data Mining(PAKDD'11), 2011.
11. S. Morinaga and K. Yamanishi. "Tracking Dynamics of Topic Trends Using a Finite Model" Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 811-816, 2004.

AUTHORS PROFILE

Dr.S.S.Manivannan, is currently working as Associate Professor in the School of Information Technology and Engineering in Vellore Institute of Technology (VIT), Vellore. He holds B.E Computer Science and M.E.Computer Science Engineering from University of Madras and College of Engineering (CEG) Anna University, Chennai. He has obtained his Ph.D in the area of Network and Information Security in the Information Technology domain in VIT, Vellore. He has more than 25 standard Scopus Indexed Publications and 5 referred Impact Factor Journals. He is the reviewer in many Scopus indexed and Science Citation Indexed Journals. He has also served as Technical Session Chair in many IEEE International Conferences.

Dr. Iyapparaja Meenakshisundaram, He received Ph.D in Anna University, Chennai, BE degree from Anna University, Chennai and ME degree from Anna University of Technology, Coimbatore. Presently, He is an Associate Professor in School of Information Technology and Engineering, VIT, Vellore. He has 11 years of experience in Teaching and Big data, Software testing and software Engineering field.

He received University Rank holder award for his ME degree. His research interests include Software Testing, Software Engineering, Big data, Networking and Agile Testing. He is life time member of ISTE.

Dr. Prasanna Mani, has completed his MS in Computer Science from Anna University. He received his doctorate in Software Engineering from Anna University. Presently he is working as Associate Professor in Vellore Institute of Technology Vellore (Deemed to be University), Vellore, Tamil Nadu. He has published nearly 25 papers in various national and international journals. He is guiding Research scholars in the area of Software testing and is an eminent reviewer of various international journals. He also authored a book for cracking interview questions of C programming. His area of interest includes Software Engineering, Software Testing, Internet of Things etc.

T. Velmurugan received his B.E degree in Electronics and Communication Engineering in 2000, M.Tech. in advanced communication systems from SASTRA University, Thanjavur India in 2002 and Ph.D. from VIT University, Vellore, India. He is currently serving as a Faculty in the School of Electronics Engineering, VIT University. He has published several papers in reputed journals. His research interest includes Handoff in Wireless Networks, Cognitive Radio Networks and IoT.

Dr. Prabhu J, received his Phd (Computer science and Engineering) and Master of Computer Science and Engineering from Sathyabama University Chennai, Tamil Nadu, India and he received the B.Tech degree from the Department of Information Technology, Vellore Engineering College, Affiliated to Madras University located at Vellore ,Tamil Nadu, India. . He worked as an Assistant Professor in various Engineering colleges for more than 13 years. Now he is working as an Associate Professor in School of Information Technology and Engineering VIT, Vellore, Tamil Nadu, India. His Research interests are software testing, Software Quality Assurance, Software metrics, IoT, Machine Learning and Big data Analytics.