Finding Tweet Events

Abhaya Kumar Pradhan
School of Computer and Information Sciences,
University of Hyderabad, India

Hrushikesha Mohanty
School of Computer and Information Sciences,
University of Hyderabad, India

ABSTRACT
Finding events from streaming tweets is challenging for needing efficient algorithm enable to scoop events from a tweet with restricted text size and to process fast as at given time large number of tweets are emanate on cyberspace. This paper proposes a syntax based approach and presents a preliminary experimental result.

General Terms
Social Networking, Information Retrieval

Keywords
Event Detection, Social Stream analysis, Clustering, Twitter, Social Media

1. INTRODUCTION
In this 21st century Social media has became an indispensable element of our Social cum Daily life. The wide use of Internet and popularity of smart phones has made our busy life more busier by creating a strong bond with social networking websites like Twitter, Facebook, Google+ and Tumblr etc. Through these types of social networking sites peoples are not only communicating with their friends but also sharing their thoughts and opinions. So there is an increasing flow of information from social networking sites for many purposes like business promotion, governance, security, health and etc. It is necessary to know what people the most are concerned about. This act of finding discussion topics is largely known as event detection. In this paper our objective is to find the events in social steams of Twitter known as tweet.

So far we have seen that how social streams come to play a role in our lives. Now we will talk a little about the properties which make them interesting. Social streams are having 3V features- Velocity, Volume, Veracity. Our habitual and rapid posting in social networking websites and micro-blogs enrich the Velocity and Volume of the social stream. In this world everybody is posting their ideas, personal feelings and opinions which brings Veracity of those posts. Tweets are also having 3v features. Restricted text length (140 characters) and informal way of writing make tweet processing challenging.

Now the question arises that "Why should we process them ?". Here is the answer and motivation for looking events in tweets - we have to agree on a point that now our society demands that everybody should have a social networking account and our social networking friends are quite more in number than real life social friends. By taking the advantage of this current trend Twitter gets popular among the people and can able to attract and add a huge number of users to its account. People are tweeting daily on a variety of topics starting from personal day-to-day life updates to global news, events and urging issues. In a broad way, processing of tweets provides vital information about two things, one is open activity like what is the trends of discussion like Black money, Run For Unity, Rise The Wage, icebucketchallenge etc. and concerns of citizens in governance like - Swachh Bharat Abhiyan ,Make In India, grievance, epidemic and its treatment etc. Secondly helps in unveiling and trapping unusual activities in Cyber Space which will harm our society.

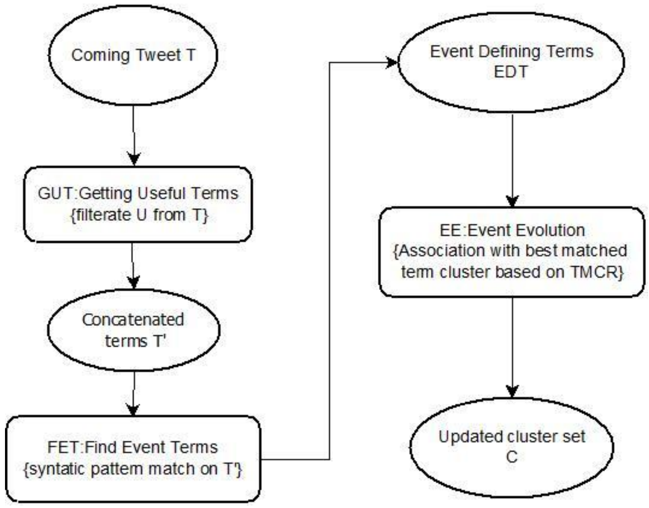


Fig 1: Framework of FTE algorithm

Fig.1 presents the complete framework of our event finding algorithm FTE: Finding Tweet Events. FTE calls GUT: Getting Useful Term, FET: Find Event Terms and EE: Event Evolution algorithms sequentially. Each algorithm is represented by round corner rectangles, inputs and outputs are represented by ovals and the flow is represented by arrow. A tweet T on arrival goes through filtering algorithm GUT which filtered out terms that are not of use for recognising syntactic pattern in T and produces concatenated terms T'. Then T' fed to FET, which makes a syntactic pattern match on T' and provides event defining terms: EDT. Now our EE algorithm takes EDT and associates it to the best matched term cluster based on TMCR: Term Matched Cardinality Ration value and finally we will get updated term clusters C.

The rest of the paper is organized as follows. Section 2 reviews related work on event finding in tweets. Section 3 introduces the proposed syntax for events. Section 4 describes proposed event finding algorithm in detail. Section 5 presents the experimental results and observation of our proposed technique with an example. Lastly Section 6 concludes with summary and future work.

2. RELATED WORK
Event detection in social media streams is a new area of research. Charu in [1] discussed both the supervised and unsupervised methods for the event detection problem (dynamic and associated network structure of the stream) using topical content of the document, their temporal distribution and the geographical structure of the dynamic network of interaction. In another work [2] researchers proposed an automated system that detects events from tweets using its textual and temporal components. And the system ET uses a hierarchical clustering technique based on the common co-occurring features of the extracted event

representative keywords for determining events. In another work Maximilian et al. [3] have proposed an algorithm for detecting geo-spatial events in tweets. First they identified places that shows a high amount of activity from the tweets issued in a given geographical region and secondly analysed the spatio-temporal clusters to find possible presence of real-world events using a machine learning component. An interactive interface is developed [4] to detect, track and summarize events in real-time from tweets. And also gives category, location information and temporal distribution. Luca Maria Aiello et al. [5] gave a good survey on six topic detection methods and have carried out a study of utility of algorithms for three twitter datasets. They also have proposed a topic detection method based on n-grams co-occurrence and document frequency-inverse document frequency (df-idf) topic ranking. Another work [6] describes an automated method for detecting event, involving known entities from twitter using NLP technique. Mario Cataldi et al. [7] proposed a real-time topic detection technique for finding the most emergent topics expressed by the community. In this three phase based technique, first they are using a novel aging theory for modelling term life cycle. Next they determined the authority of the users by analysing the social relationship with the Page Rank algorithm and finally coming out with the emerging topics using a navigable topic graph which connects the emerging terms under the user specified time constraints. In another paper [8] researchers address the issue of detecting influenza epidemics. Their proposed system extract the tweets related to influenza using Tweet API. Then SVM (Support Vector Machine) based classifier used to extract the only tweets that are about actual influenza patients. Chenliang Li et al. [9] proposed a segment-based event detection system for tweets, called Twevent. Twevent first clusters the bursty tweet segments into candidate events considering both their frequency distribution and content similarity. After clustering, Wikipedia is exploited to identify the realistic events and to derive the most newsworthy segments to describe the identified events. David Corney et al. [10] evaluated an automated topic detection system using a ground truth derived from main stream media about the FA Cup Final (2012-13).Ahmed et al. [11] proposed a real-time summarization approach which is applied on soccer games. A novel system Twitter-based Event Detection and Analysis System (TEDAS) proposed [12] having functionality like detects new events, ranks events according to their importance and analyses the spatio-temporal pattern for events. Xiangmin Zhou et al. [13] proposed a framework to detect composite social events over streams, covering multiple dimensions of social data. They proposed a graphical model called location-time constrained topic (LTT) to capture the content, time, and location of social messages. Similarity between two messages can be calculated by measuring the distance between their distributions and events can be identified by performing similarity join over social streams. In another work [14] researchers proposed a method for multi-tweet summarization of the real time events which provide a quick overview about the important moments of the events to the search user. They have proposed a graph-based retrieval algorithm that identifies tweets with popular discussion points among the set of tweets returned by Twitter search engine in response to a query comprising the event related keywords. And they did topical clustering of the tweets before applying the retrieval algorithm to acquire maximum coverage of topical diversity.

This work proposed here falls under unsupervised method of detection of events in tweets. Neither, we would like to claim it as a NLP based technique. Rather, our method is based on a heuristic adopting a framework for event representation that follows the syntax to find out patterns describing an event in a tweet. The details of the algorithm follows in next section.

3. SYNTAX FOR EVENTS

The assumptions made here on tweets includes:

—Tweets are written in English.

—A tweet is about an event, not of multiple events.

An event is seen as a manifestation in three dimensional space with coordinates Actor, Time and Place. Based on this concept describing an event could be in some of these ways e.g.:

<event> <at> <time> | <place>

<event> <by> <actor>

<event> <at> <time> <in> <place>

Keeping these structures in mind we plan to scoop out connecting terms in a tweet with a hypothesis that connecting terms are of certain events. A connecting term initiates a cluster of terms related to an event. Thus, the proposed event finding process includes finding out connecting terms and associating them with existing term clusters. In case no such cluster is found then a new cluster (for an event) is formed. We propose a generalised syntax that is followed to search for syntactic patterns in a tweet.

Syntax:

<connected-term> ::= <term> <link-term> <term>

<link-term> ::= <and> | <&> | <,> | <with> | <in>|

<of> | <form> | <for> | <to> | <at> | <on>

Some examples of such syntactic patterns are:

<term> <and> <term> : India and China

<term> <&> <term> : location & clean

<term> <,> <term> : monasteries , monks

<term> <with> <term> :interacted with Chinese

<term> <in> <term> : Make in India

<term> <from> <term> : picture from launch

<term> <of> <term> : spirit of patriotism

<term> <for> <term> : enthusiasm for swachh

<term> <to> <term> : community to tackle

<term> <at> <term> : action at U.N.

<term> <on> <term> : show on Sunday

Thus, we make a point on capability of these syntactic patterns in representing an event. On this hypothesis, in the next section we present algorithms for finding events.

4. EVENT FINDING

This section formalises the procedure for event finding with algorithmic presentation. A tweet T is viewed as a sequenced composition of terms. First, T is put through a filter to filter out terms that are not of use for recognising syntactic patterns in T. These are

U = {@ ,_, () , [] ,{}, - , : , ; , / , \ , . , ? , ! , a , an , the ,this , that , my , his , her , our , your}

The algorithm that performs this filtering is GUT: Getting Useful Terms. The resultant pre-processed T, be called T' is put for the processing by FET(T',S) where S is the syntax.

Algorithm 1 GUT(T , U)

Input: T:coming tweet, U:unnecessary term Set

Output: T':concatenated terms

Begin

T'=nil

for !empty(T) **do**

token=Term_Picker(T)

if token $\notin$ U **then**

Concatenate(T',token)

De_Concatenate(T,token)

end if

end for

return T'

End

The algorithm picks up each term and done by function Term_Picker( ) and its return into variable 'token' is matched with set U. In case of a match the term is filtered out else the term is concatenated at the front end of T'. So, the resultant processed T' that is a sequence of term devoid of term unnecessary for further processing. Now T' is fed to algorithm FET: Find Event Terms explores T' to find connected terms satisfying the syntax used to describe an event from a selected connected term, it collects the terms and stores in a set EDT called event defining terms.

Algorithm 2 FET (T' , S)

Input: T':concatenated terms, S:syntax

Output: EDT:event defining terms

Variables: tp:set of terms in a pattern

Begin

EDT=nil

for !empty(T') **do**

tp=Find_Pattern(T')

if !tp **then**

EDT=EDT $\cup$ tp

T'=Truncate(T')

end if

end for

return EDT

End

Primarily, the algorithm has broken up the sequenced terms in T' to a set of terms that describes an event. This has done by function Find_Pattern( ) and it returns EDT matched with syntactic pattern after removal of link-term. Now EDT is fed to algorithm EE: Event Evolution which associates EDT to its most relevant existing term cluster or may create a new term cluster having only this EDT.

Basically the EE algorithm finds the best matched among the existing term cluster set for a EDT, done by function Best_Match( ).This function calculates TMCR(i.e. ratio of number of term matched verses total number of terms in a term cluster) for a EDT on matching with a cluster of the set of clusters C. It returns a term cluster as BMC which having maximum TMCR value because TMCR gives the percentage of match of terms. Then EE associates the EDT to the BMC otherwise form a new term cluster and update the existing cluster set C.

Algorithm 3 EE (EDT , C)

Input: EDT:event defining terms, C:existing term cluster set

Output: updated term cluster set

Variables: BMC:best matched cluster,

TMCR:maximum term matched cardinality ration,

NTMCR:newly calculated term matched cardinality ratio,

tmatch:number of terms matched

Begin

if (EDT = NULL) **then**

Do Nothing

else

BMC=Best_Match (EDT , C)

if (BMC!= NULL) **then**

BMC = BMC $\cup$ EDT

else

C = C $\cup$ Create_Cluster(EDT)

end if

end if

End

Best Match (EDT , C)

{

TMCR = 0

for each c $\in$ C **do**

tmatch = Term_Match(EDT , c)

if (tmatch == 0) **then**

return BMC as NULL

else

NTMCR = (tmatch / cardinality(c))

if (NTMCR > TMCR) **then**

TMCR = NTMCR

BMC = c

else

continue;

end if

end if

end for

return BMC;

}

The FTE: Finding Tweet Events algorithm calls Algorithm-1 (GUT), Algorithm-2 (FET) and Algorithm-3 (EE) sequentially and the working of it pictorially represented in figure-2.2.

Algorithm 4 FTE(T,U,S,C)

Input: T:coming tweet,

U:unnecessary term Set,

S:syntax ,

C:existing term cluster set

Output: updated term cluster set

Begin

T' = GUT(T,U)

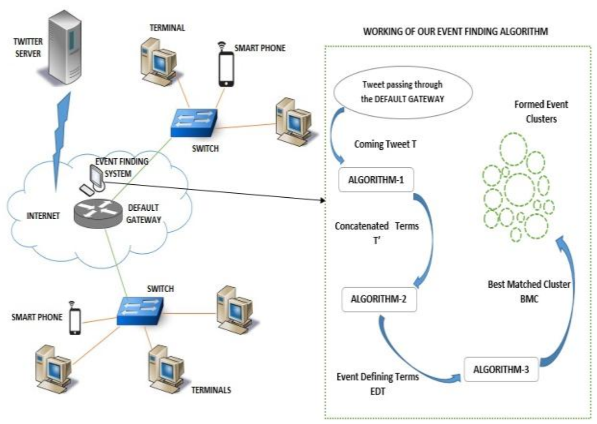
EDT = FET (T' , S)

C = EE (EDT , C)

End

4.1 System Framework

A possible real world implementation of the concept proposed here is shown in fig.2.1 and fig.2.2. A set of communicating nodes extend their connectivity to external world on Internet where possibly a twitter server is providing the tweet services. The proposed event finding algorithm sits at a gateway through which tweets on being issued at nodes, pass through. The use of the framework can be well understood with an example as follows. Suppose any Intelligence Department want to track some unusual activity which is going between different terminals then they can easily trap them using FTE algorithm.



2.1 **2.2**

Fig. 2. FTE algorithm Implementation

5. EXPERIMENT AND RESULTS

We have collected Tweets from Twitter of different profile covering different topics. Then systematically calculated them using our FTE algorithm and got meaningful Event clusters satisfying the topics. Due to the lack of space we are only showing the step by step execution of our FTE algorithm on a single topic **MAKE IN INDIA** collected from PMO, INDIA Twitter profile.

COLLECTED TWEETS:

Tweets Source: PMO, INDIA TWITTER PROFILE

T_1: Make in India is not a slogan but a mission to be accomplished with single minded commitment,about new processes: Minister @nsitharaman

T_2: We welcome the PM's call to Make in India.We realised the potential of India very early: Mr. Kenichi Ayukawa,Maruti Suzuki,India

T_3: I can tell you,my next destination is Madhya Pradesh. We are going to invest there: Mr. YC Deveshwar during Make in India programme

T_4: The PM will unveil the global Make in India initiative!

Watch it LIVE http://www.narendramodi.in/watch-live/

T_5: PM is launching http://www.makeinindia.com

T6: After seeing all this,hearing these people I do not need to tell you anything more to convince you to Make in India: PM @narendramodi

T_7: Biggest issue is trust. Why do we not trust our fellow countrymen? I want to change this: PM @narendramodi

T_8: A decision on self certification...many may feel this is a small step but what can be bigger than trusting 125 crore Indians: PM

T_9: People are talking about FDI but I see things differently.

FDI is also a responsibility for the people of India: PM

@narendramodi

T_{10}: My definition of FDI for the people of India is First Develop India: PM @narendramodi

T_{11}: We have to create opportunities of employment. If the poor get jobs the purchasing power of families will increase: PM @narendramodi

T_{12}: We have to increase manufacturing and at the same time ensure that the benefits reach the youth of our nation: PM @narendramodi

T_{13}: This is the step of a Lion...Make in India: PM @narendramodi

T_{14}: Industry does not come when there are too many incentive schemes. We have to create a development & growth oriented environment: PM

RESULTS AFTER APPLYING GUT ALGORITHM:

T'_1: Make in India is not slogan but mission to be accomplished with single minded commitment,about new processes Minister nsitharaman

T'_2: We welcome PM's call to Make in India We realised potential of India very early Mr Kenichi Ayukawa, Maruti Suzuki,India

T'_3: I can tell you,next destination is Madhya Pradesh We are going to invest there Mr YC Deveshwar during Make in India programme

T'$_4$: PM will unveil global Make in India initiative Watch it LIVE http www narendramodi in watch live

T'$_5$: PM is launching http www makeinindia com

T'$_6$: After seeing all,hearing these people I do not need to tell you anything more to convince you to Make in India PM narendramodi

T'$_7$: Biggest issue is trust Why do we not trust fellow countrymen I want to change this PM narendramodi

T'$_8$: A decision on self certification many may feel is small step but what can be bigger than trusting 125 crore Indians PM

T'$_9$: People are talking about FDI but I see things differently FDI is also responsibility for people of India PM narendramodi

T'$_{10}$: My definition of FDI for people of India is First Develop India PM narendramodi

T'$_{11}$: We have to create opportunities of employment If poor get jobs purchasing power of families will increase PM narendramodi

T'$_{12}$: We have to increase manufacturing and at same time ensure benefits reach youth of nation PM narendramodi

T'$_{13}$: is step of Lion Make in India PM narendramodi

T'$_{14}$: Industry does not come when there are too many incentive schemes We have to create development & growth oriented environment PM

RESULTS AFTER APPLYING <u>FET</u> ALGORITHM:

EDT_1 : { Make , India , mission , be , accomplished , single , commitment , about , }

EDT_2 : { call , Make , India , potential , Ayukawa , Maruti , Suzuki }

EDT_3 : { you , next , going , invest , Make , India }

EDT_4 : { Make , India , narendramodi , watch }

EDT_5 : { NULL }

EDT_6 : { all , hearing , need , tell , more , convince , you , Make ,India }

EDT_7 : { want , change }

EDT_8 : { decision , self }

EDT_9 : { responsibility , people , India }

EDT_{10} : { definition , FDI , people , India }

EDT_{11} : { have , create , opportunities , employment , power , family }

EDT_{12} : { have , increase , manufacturing , same , youth , nation }

EDT_{13} : { step , Lion , Make , India }

EDT_{14} : { have , create , develop , growth }

RESULT AFTER APPLYING <u>EE</u> ALGORITHM:

<u>Iteration 1</u>:

Collected Term Set: EDT_1

Existing Cluster Set: NULL

Best matched Cluster: NULL

Remarks: Create a new cluster c_1 with EDT_1

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single , commitment , about }

<u>iteration 2</u>:

Collected Term Set: EDT_2

Existing Cluster Set: c_1

Best matched Cluster: c_1

Remarks: Add EDT_2 to cluster c_1 and Update

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single commitment , about , call , potential , Ayukawa , Maruti , Suzuki }

<u>iteration 3</u>:

Collected Term Set: EDT_3

Existing Cluster Set: c_1

Best matched Cluster: c_1

Remarks: Add EDT_3 to cluster c_1 and Update

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single commitment , about , call , potential , Ayukawa , Maruti , Suzuki , you , next , going , invest }

<u>iteration 4</u>:

Collected Term Set: EDT_4

Existing Cluster Set: c_1

Best matched Cluster: c_1

Remarks: Add EDT_4 to cluster c_1 and Update

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single commitment , about , call , potential , Ayukawa , Maruti , Suzuki , you , next , going , invest, narendramodi , watch }

<u>iteration 5</u>:

Collected Term Set: $EDT_5 = NULL$

Existing Cluster Set: c_1

Best matched Cluster: NULL

Remarks: Do Nothing

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single commitment , about , call , potential , Ayukawa , Maruti , Suzuki , you , next , going , invest, narendramodi , watch } iteration 6:

Collected Term Set: EDT_6

Existing Cluster Set: c_1

Best matched Cluster: c_1

Remarks: Add EDT_6 to cluster c_1 and Update

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single commitment , about , call , potential , Ayukawa , Maruti , Suzuki , you , next , going , invest, narendramodi , watch, all , hearing , need , tell , more , convince }

iteration 7:

Collected Term Set: EDT_7

Existing Cluster Set: c_1

Best matched Cluster: NULL

Remarks: Create a new cluster c_2 with EDT_7

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single commitment , about , call , potential , Ayukawa , Maruti , Suzuki , you , next , going , invest, narendramodi , watch, all , hearing , need , tell , more , convince }

c_2 : { want , change }

iteration 8:

Collected Term Set: EDT_8

Existing Cluster Set: c_1 , c_2

Best matched Cluster: NULL

Remarks: Create a new cluster c_3 with EDT_8

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single commitment , about , call , potential , Ayukawa , Maruti , Suzuki , you , next , going , invest, narendramodi , watch, all , hearing , need , tell , more , convince }

c_2 : { want , change }

c_3 : { decision , self }

iteration 9:

Collected Term Set: EDT_9

Existing Cluster Set: c_1 , c_2 , c_3

Best matched Cluster: c_1

Remarks: Add EDT_9 to cluster c_1 and Update

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single commitment , about , call , potential , Ayukawa , Maruti , Suzuki , you , next , going , invest, narendramodi , watch, all , hearing , need , tell , more , convince , responsibility , people }

c_2 : { want , change }

c_3 : { decision , self }

iteration 10:

Collected Term Set: EDT_{10}

Existing Cluster Set: c_1 , c_2 , c_3

Best matched Cluster: c_1

Remarks: Add EDT_{10} to cluster c_1 and Update

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single commitment , about , call , potential , Ayukawa , Maruti , Suzuki , you , next , going , invest, narendramodi , watch, all , hearing , need , tell , more , convince , responsibility , people , definition , FDI }

c_2 : { want , change }

c_3 : { decision , self }

iteration 11:

Collected Term Set: EDT_{11}

Existing Cluster Set: c_1 , c_2 , c_3

Best matched Cluster: NULL

Remarks: Create a new cluster c_4 with EDT_{11}

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single commitment , about , call , potential , Ayukawa , Maruti , Suzuki , you , next , going , invest, narendramodi , watch, all , hearing , need , tell , more , convince , responsibility , people , definition , FDI }

c_2 : { want , change }

c_3 : { decision , self }

c_4 : { have , create , opportunities , employment , power , family }

iteration 12:

Collected Term Set: EDT_{12}

Existing Cluster Set: c_1 , c_2 , c_3 , c_4

Best matched Cluster: c_4

Remarks: Add EDT_{12} to cluster c_4 and Update

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single commitment , about , call , potential , Ayukawa , Maruti , Suzuki , you , next , going , invest, narendramodi , watch, all , hearing , need , tell , more , convince , responsibility , people , definition , FDI }

c_2 : { want , change }

c_3 : { decision , self }

c_4 : { have , create , opportunities , employment , power , family , increase , manufacturing , same , youth , nation }

iteration 13:

Collected Term Set: EDT_{13}

Existing Cluster Set: c_1 , c_2 , c_3 , c_4

Best matched Cluster: c_1

Remarks: Add EDT_{13} to cluster c_1 and Update

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single commitment , about , call , potential , Ayukawa , Maruti , Suzuki , you , next , going , invest, narendramodi , watch, all , hearing , need , tell , more , convince , responsibility , people , definition , FDI , step , Lion }

c_2 : { want , change }

c_3 : { decision , self }

c_4 : { have , create , opportunities , employment , power , family , increase , manufacturing , same , youth , nation }

`iteration 14`:

Collected Term Set: EDT_{14}

Existing Cluster Set: c_1 , c_2 , c_3 , c_4

Best matched Cluster: c_4

Remarks: Add EDT_{14} to cluster c_4 and Update

Updated Cluster Set:

c_1 : { Make , India , mission , be , accomplished , single commitment , about , call , potential , Ayukawa , Maruti , Suzuki , you , next , going , invest, narendramodi , watch, all , hearing , need , tell , more , convince , responsibility , people , definition , FDI , step , Lion }

c_2 : { want , change }

c_3 : { decision , self }

c_4 : { have , create , opportunities , employment , power , family , increase , manufacturing , same , youth , nation , develop , growth }

RESULT:

After applying our FTE algorithm to the collected tweets, we got four term clusters-c_1 , c_2 , c_3 and c_4.Two of them (c_2, c_3) are not clear. The terms in c1 cluster are related to many facades of event 'Make In India' and very clear to make out whereas fourth cluster c_4 also has some terms that partially describes the event. Terms in c_1 and c_4 are mutually exclusive. This shows that proposed algorithm needs a refinement so that all terms with respect to an event can be put into one cluster.

OBSERVATION:

The above experiment carried out on collected tweets from different profiles on different topics. At the end of each experiment, term clusters are formed. Among them some are clearly defining the topic and some are not clear to make out. Table 1 contains experimental results with error rates. Here each term cluster is considered as an event. Event found column get the values of the total term clusters formed at the end of each experiment. Among them some are Emergent Events. Procedure for calculating Emergent Event and error as follows:

For each experiment –

1. Find the number of unique terms in all the term clusters as

UT

2. Take the number of Term clusters found as NC

3. Threshold of number of words to describe an Emerging Event is UT/NC = δ_{ee} (where δ_{ee} : minimum number of terms required to decide an Emergent Event).

4. Count the term clusters whose cardinality ≥ δ_{ee} as NEE

5. Calculate Error = (NEE/NC)*100

All the experiments are carried out on small number of tweets collected from different profiles with the predefined 11 link-terms of syntax and the corresponding error rates are shown in the table 1. The error rate can be minimized to some extend using some more link-terms. And more concrete results can be obtained by experimenting FTE algorithm on large Data Sets. But as we have told before the FTE algorithm needs a refinement so that all the terms related to a topic can be put into one cluster and error rate will be minimized.

Table 1: Experimental results with error rates

Twitter Profile	Topic	Emergent Events	Events found	Error (in %)
PMO, India	Make In India	**2**	**4**	**50**
Narendra Modi	MyCleanIndia	**2**	**6**	**33.33**
President of India	Swachh Bharat Abhiyan	**2**	**6**	**33.33**
Barack Obama	Act on Climate	**2**	**7**	**28.57**
Narendra Modi	Xi Jinping's Visit	**3**	**6**	**50**

6. CONCLUSION AND FUTURE WORK

Here, this paper proposed a technique for finding events in tweets. The technique is detailed with algorithms. We have experimented these algorithms on tweets on 'Make In India','Swachha Bharat','Act On Climate','Xi Jinping visit' etc. and have found efficiency of our technique satisfactory. And the experiment has promising result as good as the result due to statistical techniques that are carried out extensively. This work is a pilot project of our intended research in this upcoming field. We sense a success with this preliminary research on the problem of event finding in tweets.

Further, the proposed technique is to be enriched in many dimensions to upgrade it to a useful system addressing most of the challenging problems the problem is endowed with. The first problem is of making the algorithm computationally very fast for on-line processing. The second one is with veracity in tweet representations. Until now, all works consider textual tweets only. Handling a large number of tweets for processing is the most challenging task. Associating person, location and time to tweet events is challenging as well as useful particularly for security as well as governance. In our future research, we would like to take up these challenges in tweet processing.

7. REFERENCES

[1] Charu C. Aggarwal, and Karthik Subbian. Event Detection in Social Streams. In Proceedings of the SIAM International Conference on Data Mining, pages 624-635, 2012.

[2] Ruchi Parikh, and Kamalakar Karlapalem. ET: Events from Tweets. In Proceedings of the 22nd international conference on World Wide Web companion, pages 613-620, 2013.

[3] Maximilian Walther, and Michael Kaisser. Geo-spatial Event Detection in the Twitter Stream. Advances in Information Retrieval Lecture Notes in Computer Science, Volume 7814, pages 356-367, 2013.

[4] Andrew J. McMinn, Daniel Tsvetkov, Tsvetan Yordanov, Andrew Patterson, Rrobi Szk, Jesus A. Rodriguez Perez, and Joemon M. Josey. An Interactive Interface for Visualizing Events on Twitter. In Proceedings of the 37th international ACM SIGIR conference on R&D in information retrieval, pages 1271-1272, 2014.

[5] Aiello L.M., Petkos G., Martin C., Corney D., Papadopoulos S., Skraba R., Goker A., Kompatsiaris I., and Jaimes A. .Sensing Trending Topics in Twitter. Multimedia, IEEE Transactions,Volume 15,issue 6, pages 1268-1282, 2013.

[6] Ana-Maria Popescu, Marco Pennacchiotti, and Deepa Arun Paranjpe. Extracting Events and Event Descriptions from Twitter. Proceedings of the 20th international conference companion on World wide web, pages 105-106, 2011.

[7] Mario Cataldi, Luigi Di Caro, and Claudio Schifanella. Emerging topic detection on Twitter based on temporal and social terms evaluation. In Proceedings of the Tenth International Workshop on Multimedia Data Mining (MDMKDD '10), Article 4, 2010.

[8] Eiji Aramaki, Sachiko Maskawa, and Mizuki Morita. Twitter catches the flu: detecting influenza epidemics using Twitter. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11), pages 1568-1576, 2011.

[9] Chenliang Li, Aixin Sun, and Anwitaman Datta. Twevent: segment-based event detection from tweets. In Proceedings of the 21st ACM international conference on Information and knowledge management (CIKM '12), pages 155-164, 2012.

[10] David Corney, Carlos Martin, and Ayse G¨oker.Spot the Ball: Detecting Sports Events on Twitter. Advances in Information Retrieval Lecture Notes in Computer Science, Volume 8416, pages 449-454, 2014.

[11] Ahmed A. A. Esmin, Rômulo S. C. Júnior, Wagner S. Santos, Cássio O. Botaro, and Thiago P. Nobre. Real-Time Summarization of Scheduled Soccer Games from Twitter Stream. Natural Language Processing and Information Systems Lecture Notes in Computer Science, Volume 8455, pages 220-223, 2014.

[12] Rui Li, Kin Hou Lei, Khadiwala R., Chang, K.C.-C.. TEDAS: A Twitter-based Event Detection and Analysis System. Data Engineering (ICDE), IEEE 28th International Conference, pages 1273-1276, 2012.

[13] Xiangmin Zhou, Lei Chen. Event detection over twitter social media streams. The DLVB Journal, Volume 23, issue 3, pages 381-400, 2014.

[14] Khan, M.A.H., Bollegala, D., Guangwen Liu, Sezaki, K.. Multi-tweet Summarization of Real-Time Events. International Conference on Social Computing (SocialCom), pages 128-133, 2013.