

ProminenceRank based Event Timeline Generation in Twitter using Hashtag Clustering

Nilesh Gambhava¹, Dr. Ketan Kotecha²,

¹Computer Engineering Department, Nirma University, Ahmedabad-382481, India. nileshgambhava@gmail.com

²Parul University, Waghodia, Vadodara-391760, India. drketankotecha@gmail.com

Abstract: Communication through social media network such as Twitter has facilitated the fastest and richest platform for information spreading and opinion sharing. Intelligent exploration of this information can generate valuable records as they represent the essence of real-world societal aspects. At the same time, huge volume of short unstructured tweets, noisy and unordered format are certain major challenges to be dealt with in order to generate meaningful outcomes. In this paper, we propose a novel heuristic function called *prominenceRank* based event timeline generation using hashtag clustering technique. Event timeline generation task requires generating an ordered sequence of tweets representing important incidents about the entity or an event in chronological order. To determine a set of related tweets we use hashtag clustering using a genetic algorithm. The *prominenceRank* based algorithm extracts the most relevant tweets that further can be used to generate timeline. Experiments are performed using a real-life dataset that consists of approximately 2,00,000+ tweets from 24 September 2017 to backward. We evaluate the proposed algorithm and generated the timeline for three different major events using the selected tweets. The timeline generated shows the efficiency of our approach in terms of considering the substantial diversity, relevance, and effectiveness of the selected tweets.

Keywords: Event Timeline, Social Media, Hashtag Clustering, Prominence Rank.

1. Introduction

Due to the increasing use of Social media services, such as Twitter, a treasurable repository of otherwise unavailable information is created in this era of digital communication. Intelligent exploration of this information can generate valuable records as they represent the essence of real-world societal aspects. Accessibility to these information has become easy using public APIs or various mining tools. The social media has an intractable impact on each and every happening event. The increasing use of Twitter has revolutionized the way to create and spread information anywhere in the world at any period of time. Many such parameters have restructured the framework for communication and created a great motivation to reveal numerous research dimensions. Still, the obstructions are many for automatically gaining some useful outcomes from the data available on Twitter. The first limitation is length constraint of 140 characters in Tweet. Because of which users try to represent as much as they can by using short unstructured words. As a whole Tweet gives the clear sense of what it is made for. But to derive the core material related to specific event requires intelligent efforts and exploitation tools. Also, a substantially sharpened approach is required to discover the necessary information from the noisy and irrelevant data of Twitter.

In this research, we propose a novel approach to generate a timeline of an event using hashtag clustering. Our work in [9] presents the distinctive model using a genetic algorithm for hashtag clustering. Hashtag denotes a word or a phrase preceded by a hash sign (#). It is used to tag the messages on a specific theme or domain. The hashtag symbol (#) before some keyword or phrase in tweets signifies the importance of tweet during the search. Twitter also uses hashtags to index keywords. Because of the hashtag, it's possible that your tweet is seen by hundreds or even millions of users who are not following you but have searched using a particular hashtag. Tweets without hashtag have a very short life. Usually, the most of all relevant information is posted by different users world-wide in the form of tweets and retweets. But the task of generating event or entity related sequence of incidents to frame the timeline is an increasingly complex task. Problems like an enormous amount of data, redundancy of information, the presence of noisy data etc... are to be resolved for the generation of a precise timeline. Moreover, an intelligent search is required to get the accurate information for creating an exact sequence of information in chronological order. Currently, if we try to search on Tweeter regarding some event or incident then it simply shows all the possible Tweets sorted in date order. To overcome this problem, we propose a novel heuristic called as the *prominenceRank* based selection of Tweets from the available data to generate the precise and efficient event timeline.

Reframing the details of an event from the available collection of tweets is a significant task. Our research work mainly focuses on time-specific events. Events such as demonetization took place and spread over a specific region in a short duration of time. Tweets and retweets are posted about every incident occurred for that event with very high frequency. After some duration of the time, things disappear. Our aim is to extract the relevant sequences of incidents enlightening the glimpse of an event from the available details and to assemble them for the precise timeline generation. Event timeline generation by analyzing social media data has a wide application area. Some of the major applications are, creating an article in Wikipedia, writing a newspaper story, forming a historical report, generating a documentary etc... Timeline generation is one of the most powerful techniques which can make analysis tasks easier and faster considering the events happening in any or every corner of the world.

The rest of the paper is organized as follows. The review of literature is presented in section 2. Section 3 presents our proposed approach and algorithm for event timeline generation using hashtag clustering. Case studies and observations are discussed in section 4. Section 5 concludes the paper with an insight into the future work.

2. Literature Review

Wang et al. [14] presented summarization and timeline Generation for evolutionary tweet streams using novel continuous summarization framework called Sumblr. Authors propose an online tweet stream clustering algorithm to cluster tweets, a TCV-Rank summarization technique for generating online summaries and historical summaries of arbitrary time durations and designed an effective topic evolution detection method, which monitors summary-based/volume-based variations to produce timelines automatically from tweet streams. Yao et al. [7] present an approach based on determinantal point processes (DPPs) by jointly modeling the topical relevance of each selected tweet and overall selection diversity. For the purpose of better balancing relevance and diversity, authors introduced two novel strategies, namely spectral rescaling and topical prior. Zhou et al. [13] proposed a four stage approach which includes classification of real-world events reporting tweets, online incremental clustering, post-processing and sub-events summarisation.

Chang et al. [12] investigated the problem of timeline summarization and proposed a novel framework Timeline-Sumy, which consists of episode detecting and summary ranking. Magdi et al. [11] examined the dependency of Tweet Timeline Generation on retrieval quality and its effect on having a biased evaluation. Li and Cardie [6] address the problem of reconstructing users' life history based on their Twitter stream and proposed an unsupervised framework that creates a chronological list for personal important events (PIE) of individuals. The authors introduce a non-parametric multi-level Dirichlet Process model to recognize four types of tweets: personal time-specific (PersonTS), personal time-general (PersonTG), public time-specific (PublicTS) and public time-general (PublicTG) topics, which, in turn, are used for further personal event extraction and timeline generation. Fan et al. [3] proposed a graph-based dynamic greedy clustering approach, which considers the coverage, relevance, and novelty of the tweet timeline. Authors constructed the tweet semantic graph using tweet embedding representation. Based on the graph, the coverage of timeline is estimated according to the graph connectivity. Also, they have integrated a noise tweet elimination component to remove noisy tweets with the lexical and semantic features based on relevance and novelty.

3. Event Timeline Generation using Hashtag Clustering

The frequent availability and conveniences have increased the popularity of social media services as a common platform for sharing information and knowledge. But due to the massive volume of the database, it becomes challenging to acquire significant information about an entity or an event. Especially, mining the relevant information from all the available messy data requires excessive efforts. A generic approach to provide an effective and efficient method to understand an event/entity by presenting a sequence of incidents in chronological order leads to the growing requirement of event Timeline generation. The objective of Event timeline generation from Twitter is logically identifying considerable tweets that represent the core essence and can generate a timeline which provides enough emphasis on major incidents. In the subsequent subsections, we describe the detailed algorithm for event timeline generation using hashtag clustering.

3.1 Hashtag Clustering using Genetic Algorithm

Extracting information from a tweet database is extremely difficult because of its unstructured nature. To show the importance or emphasis on a specific word in Twitter, it is preceded by a hash sign (#). The word or phrase called hashtag then highlights the relevant category for tweets which help them include more easily in Twitter search. A hashtag is also used as an index of searching and it's possible that your tweet is seen by hundreds or even millions of

users who are not following you but have searched using a particular hashtag. Single hashtag searching may not give optimal result hence grouping of similar hashtag plays a vital role in the improvement of twitter hashtag search result. Grouping of a similar hashtag is a typical clustering problem. Hashtag clustering has various applications like [1], [4], [10]. We have exploited clustering capability of genetic algorithm in our previous research work [9] for the clustering of hashtags representing the similar category or theme. Genetic Algorithm (GA) [2], [5], [8] is an adaptive heuristic search algorithm that mimics the evolutionary process of natural selection and survival of the fittest. Fig. 1 shows the complete genetic algorithm framework for hashtag clustering. From the selected corpus of tweets, hashtags are extracted and the co-occurrence frequency of each possible pair is computed. The initial population contains randomly generated solution chromosomes. Until the predefined number of generations is completed, the GA loop, inclusive of fitness calculation and application of selection, crossover and mutation operators, is continued. During every generation, fitness function measures the effectiveness of solution and GA operators strengthen the solution chromosomes. As the generations increase, the solutions are evolved and after a predefined number of generations the resultant clusters contain relevant hashtags within them.

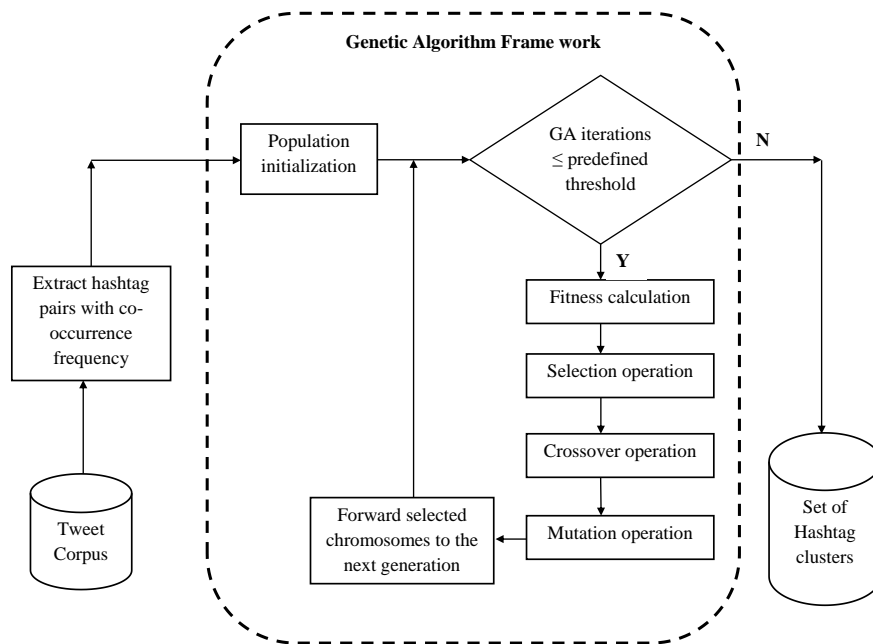


Figure 1. The flow diagram for hashtag clustering using genetic algorithm

3.2 Event Timeline Generation using Hashtag Clustering

Tweets are generally short and informal sequence of unstructured text. But proper extraction of relevant tweets helps in characterizing timeline for an event/entity. Moreover, certain significant details locally or globally are available only from the information-rich database created online through user's posts on social media. Table 1 shows the algorithm for event timeline generation using hashtag clustering. Relevant hashtags are clustered together using a genetic algorithm as explained in section 3.1. Thereafter, the event or entity is queried to the system for timeline generation. The hashtag representing the query is determined and the cluster to which that hashtag is a member is searched. Once the cluster is found, the tweets concerning all the hashtags of that cluster are selected as an entirely available tweet data for the timeline generation. We propose a novel *prominenceRank* based heuristic function, which calculates the rank defining how effectively the particular tweet enlightens the incident, which significantly contributes to the timeline generation. The *prominenceRank* uses the parameters that are considered as important indicators for a tweet such as, the number of retweets of a tweet. Also, the retweet pattern of a user defines the level of eminence justifying the opinion posted by the user. The *prominenceRank* is defined as the ratio of a number of retweets of a tweet to the user's average retweet count. Higher the *prominenceRank* of a tweet is, more it contributes to the effective sequence of incidents for the timeline generation. After sorting all selected tweets in decreasing order of their *prominenceRank*, the top N most substantial tweets are considered and re-organized in chronological

order for the timeline generation. Fig. 2 shows the schematic diagram of event timeline generation algorithm from the hashtag clusters. The key searching element used in our approach is a cluster containing the correlated hashtags. Hashtags are created for all the relevant incidents describing the event and frequently they are cited in the tweets.

Table 1: Algorithm for Event Timeline Generation

Input:	
i.	Collection of tweets $T = \{t^1, t^2, \dots, t^n\}$
i.	Clusters of Hashtags $C = \{C_1, C_2, \dots, C_m\}$
i.	Hashtag query H_q for which the timeline is to be generated
Output: Event timeline $E_q = \{t_q^1, t_q^2, \dots, t_q^k\}$	
1.	Search the cluster C_s which contains the Hashtag H_q
2.	Store all hashtags within the cluster C_s in the set rel_Htag
3.	Locate all possible tweets $\{t_p^1, t_p^2, \dots, t_p^l\}$ from T with all the hashtags of rel_Htag
4.	Find the relevant set of tweets as follows:
4.1	For all tweets tp^1, tp^2, \dots, tp^l find the <i>prominenceRank</i> as
4.1.1	Find retweet value rtv_p^i for tp^1, tp^2, \dots, tp^l
4.1.2	$prominenceRank^i = rtv_p^i / \text{average retweet count of owner}$
5.	Sort all the tweets $t_p^1, t_p^2, \dots, t_p^l$ in decreasing order of <i>prominenceRank</i> ⁱ
6.	Select top N tweets, re-arrange in chronological order and store in E_q
7.	Return E_q

The *prominenceRank* defines the measure of how significantly a particular tweet describes the incident related to the event for which the timeline is to be generated. The user can specify the bottom value for it and can decide up to what level of depth the tweets are to be included in timeline generation.

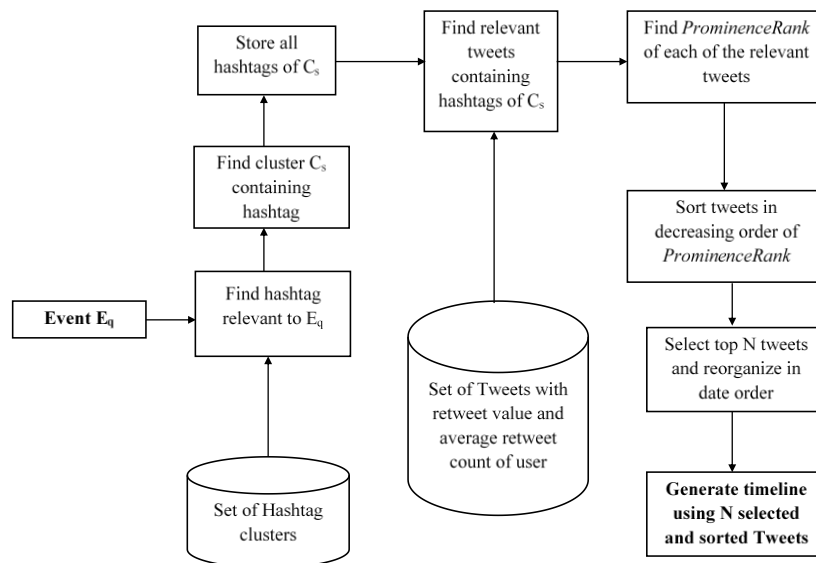


Figure 2. The schematic diagram for event timeline generation

4. Experiments and Observations

4.1 Dataset

We perform experiments using three different case study events. The dataset consists of tweets taken from selected top 76 Indian media twitter accounts. Twitter gives recent 3200 tweets of each user. Our dataset is consisting of approximately 2,00,000+ tweets from 24 September 2017 to backward. These tweets contain 1,00,000+ hashtags and 27,000+ pairs of hashtags. We have removed hashtags with co-occurrence frequency less than 5 to make it computable. Our clustering model using GA generated 56 clusters of hashtags. We consider three current events for generating timeline using our proposed algorithm.

4.2 Experiments and Discussions

We found following three major events from our dataset;

1. RamRahim of DeraSachaSauda
2. Pradyuman Murder case of Ryan International School
3. Japan PM's visit to India for Bullet Train

We have applied our proposed novel approach to these events. Hashtag Clusters generated using the procedure shown in fig. 1 for the events are shown in table 2. These hashtags are used to extract the relevant tweets from the available dataset.

Table 2: Case study events for Timeline Generation

Events	Related Hashtags
RamRahim of DeraSachaSauda	#DERASACHASAUDA #RAMRAHIM #RAMRAHIMSINGH #RAMRAHI MVERDIC#HONEYPREET #HARYANA
Pradyuman Murder case of Ryan International School	#PRADYUMAN #PRADYUMANMURDERCASE #RYANINTERNATIONA LSCHOOL #GURUGRAM
Japan PM's visit to India for Bullet Train	#BULLETRAIN #PMMODI #SHINZOABE;

Table 3 shows the event timeline generation for the first case study considered, RamRahim of DeraSachaSauda. The first column shows the *prominenceRank*(pR) of the specific tweet. The selected tweets are re-organized in chronological order. *prominenceRank* based selection allows us to determine those tweets which are been paid maximum attention of users. Such re-organized set of tweets generate the list of incidents which together summarize the event in a whole.

Table 3: Event timeline generated for RamRahim and DeraSachaSauda

pR	User	Date	Social media post presenting the incident
10	thetribunechd	17-Aug-17	Court reserves verdict in rape case against #DeraSachaSauda chief @Gurmeetramrahim https://t.co/BrK3OJb74a https://t.co/FvziLqTg0A
36	Abpnewstv	25-Aug-17	#RamRahimVerdict: Should chief minister Manohar Lal Khattar resign post #Haryana violence?
18	Abpnewstv	25-Aug-17	#RamRahimSingh's properties to be seized to compensate the loss incurred in violence by his followers: High Court... https://t.co/KY022GmZ4b
12	Abpnewstv	25-Aug-17	BREAKING NEWS - Seize all properties of #RamRahimSingh, orders High Court of Punjab and Haryana... https://t.co/7gG9FZ3zKr
11	Abpnewstv	25-Aug-17	BREAKING NEWS - #BabaRamRahim followers set on fire two railway stations in Punjab. One is named Malaout rly stn. https://t.co/JQJIKZeiOk
10	Abpnewstv	25-Aug-17	BREAKING NEWS -#RamRahimSingh convicted by CBI court in rape case, quantum of punishment on Aug 28... https://t.co/024qz3E16r
16	AmarUjalaNews	25-Aug-17	या कोर्ट जाना मक्तिद के साथी इनाम को कैसे बुला सकता है, कैसे उन्होंने एक सीधे सादे इनाम को बुलाया??- साथी महाराज#RamRahimCase
48	AmarUjalaNews	25-Aug-17	भाजपा संसद साथी महाराज ने किया बाबा राम रहीम का बचाव, कहा- वो एक 'पवित्र आत्मा'>> https://t.co/zj7sRrcup #RamRahimSingh #RamRahimVerdic

14	Outlookindia	25-Aug-17	Police resort to lathi charge. Water cannons and tear gas are being used to control #RamRahimSingh's followers in... https://t.co/kVMSqPv13p
10	punjabkesari	25-Aug-17	पंजाब हाईकोर्ट का फैसला: हिंसा के चलते राम रहीम की सारी संपत्ति होगी जब्त#RamRahimVerdict
13	thetribunechd	25-Aug-17	#BreakingNews: Water cannon, tears gas shells fired, OB Vans damaged by #RamRahimSingh followers
13	thetribunechd	25-Aug-17	#Sirsa Dera chief #GurmeetRamRahimSingh held guilty of rapeTo be taken to Rohtak jail in a chopper Army deployed... https://t.co/CDaU4jpoC9
20	Outlookindia	26-Aug-17	#RamRahimVerdict: Retired #CBI Investigator Reveals How Seniors, Politicians Put Pressure On Him To Drop The Case... https://t.co/fgF8ObtkSz
28	suchetadalal	26-Aug-17	RT @ghaywan: Lest we forget: 15 yrs ago, journalist Ram Chander Chhatrapati, was shot dead for bringing out #RamRahimSingh's rape case out...
63	punjabkesari	26-Aug-17	राम रहीम का बचाव करने वालों पर चढ़के विचार, कल- शर्म को https://t.co/Wlb2lhG7Td #RamRahimGuilty #KumarVishwas https://t.co/lieHnRU72k
13	WIONews	26-Aug-17	Political parties need 'gurus' like #RamRahimSingh to win elections & stay relevant in #politics. By @kartikeya 1975 https://t.co/2CxNPQJD9z
20	Outlookindia	27-Aug-17	RT @Outlookindia: #RamRahimVerdict: Retired #CBI Investigator Reveals How Seniors, Politicians Put Pressure On Him To Drop The Case https://t.co/fgF8ObtkSz
15	Aajtak	28-Aug-17	बलात्कारी बाबा का परिणाम#RamRahimSentencing https://t.co/xZSvqSWual
20	Live_Hindustan	28-Aug-17	#RamRahimVerdict - आज़म में सीली दुनिया: फासवर्ड से खुलते हैं #RamRahimSingh के कर्म https://t.co/jAmLrM0vW2 #RapistRamRahim
11	NewsWorldIN	28-Aug-17	It Is Not 10, But #20Years Of Jail Term For Rapist #RamRahim #RamRahimVerdict https://t.co/rJoC5Djmol https://t.co/RdJ7UtaC8B
17	Abpnewstv	28-Aug-17	BREAKING NEWS: Dera chief #RamRahimSingh sentenced to 10 years in jail by CBI Judge Jagdeep Singh in 15-yr-old rape... https://t.co/AL1B9VcmPn
46	thetribunechd	28-Aug-17	#Breaking: It is 20 years in jail for #Dera chief #GurmeetRamRahimSingh, not 10 years https://t.co/9aWYne1n5R https://t.co/7MKll7M3pj
20	JagranNews	2-Sep-17	डेटा प्रेमियों के धर्मरक्षण की सखि, मोके का 'फासवर्ड' उदरे को निरनरियां सखि https://t.co/ZxbMFLvO1N #RamRahimSingh
11	thetribunechd	20-Sep-17	Over 600 skeletons buried in Dera, admits senior vice chairman @sushilmanav #DeraSachaSauda

Table 4 depicts the timeline generated for the event of RyanInternationalSchool. The event timeline generated using our algorithm effectively represents a sequence of timeline incidents and automatically recapitulates the central theme of each incident from the relevant post. We have represented the timeline directly showing the finally selected set of tweets in chronological order. This can be further used as a diversified and suggestive resource for the generation of a Wikipedia article or report etc... by users.

Table 4: Event timeline generated for Pradyuman Mu rder case of Ryan International School

pR	User	Date	Social media post presenting the incident
3	punjabkesari	2-Sep-17	हनीप्रीत इन्सा की तलाश के लिए गुच्छाम में छापेभारी https://t.co/vPxrZ2SKGC #HoneyPreetInsan #GuruGram... https://t.co/J0SKFCCBfh
3	punjabkesari	8-Sep-17	रवॉन इंटरनेशनल स्कूल में कल के आरोपी बस कंडक्टर ने कबूला जुर्म#RyanInternationalSchool #Gurugram
10	punjabkesari	8-Sep-17	रवॉन इंटरनेशनल स्कूल में कल से पहले बच्चे के साथ यौन शोषण की कोशिश,10 लोग हिरासत में##RyanInternationalSchool
3	RajatSharmaLive	9-Sep-17	Shocked at the brutal murder of a Class 2 student in #RyanInternationalSchool #Gurugram.This murder shd be treated as 'rarest of rare' crime
4	NewsX	9-Sep-17	#Gurugram: Conductor confesses to killing 7-year-old school boy#RyanInternationalSchool @gurgaonpolice #Gurgaon https://t.co/rVfGabhJbe
4	abpnewstv	9-Sep-17	#RyanInternationalSchool Murder: 'School Could Not Even Ensure Basic Safety For My Son,' Says Mother https://t.co/FLH0Vvkc0C
40	NewsWorldIN	9-Sep-17	RT @NewsWorldIN: Innocent Pradyuman Murdered: Know What Happened That Day! #RyanInternationalSchool #Gurugram https://t.co/pyKlaLeYh5

40	NewsWorldIN	9-Sep-17	Innocent Pradyuman Murdered: Know What Happened That Day! #RyanInternationalSchool #Gurugram https://t.co/pyKJLaYh5
3	NewsWorldIN	9-Sep-17	#Gurugram: Members of Bar Association of Sohna decide not to appear on behalf of accused in... https://t.co/4tCHIDWF1C
4	NewsX	10-Sep-17	#SchoolMurder: FIR filed against the management of #RyanInternationalSchool, Gurugram https://t.co/nDUav34cG3
3	NewsX	10-Sep-17	#SchoolMurder: Cops lathi charge parents protesting outside #RyanInternationalSchool https://t.co/4pNq3B4nXV
5	AmarUjalaNews	10-Sep-17	प्रद्युमन मर्डर केस: आरोपी कंडक्टर की बहन बोली- झूठा बयान देने के लिए बनाया गया दबाव https://t.co/tW9SJnQCTT #pradyuman
4	AmarUjalaNews	10-Sep-17	#RyanInternationalSchool ये केंद्रीय मंत्री बोले सवा सौ करोड़ की आबादी में अपराध नहीं रूक सकते!#AmarujalaTV https://t.co/Z9FAyFQuJY
16	abpnewstv	10-Sep-17	#RyanInternationalSchool: Police lathi charge on parents https://t.co/SS2eXFxMg1
4	NavbharatTimes	10-Sep-17	गुरुग्राम: रायन इंटरनेशनल स्कूल के बाहर प्रदर्शन कर रहे पैट्रोल पर पुलिस ने लाठीचार्ज किया#RyanInternationalSchool https://t.co/KS9r6twUTT
3	NavbharatTimes	10-Sep-17	रायन इंटरनेशनल स्कूल के बगल में मौजूद शराब की दुकान में अभिभावकों ने आग लगाई (ANI)#RyanInternationalSchool https://t.co/mwg7j6XHdk
3	NavbharatTimes	11-Sep-17	नोएडा: #RyanInternationalSchool ट्रिपलिका का ड्राइवर बिना आईडी और बर्तन परिसर में घूम रहा था, अभिभावकों ने पकड़ा https://t.co/9pXzARObtl
3	NavbharatTimes	11-Sep-17	#RyanInternationalSchoolMurder केस: मां बोली खुदकुशी कर लूंगी, पिता SC के लिए निकले https://t.co/flTV1VNhc9 via... https://t.co/CC0NfIKlo
5	indiatvnews	13-Sep-17	RT @RajatSharmaLive: Vital disclosures made in #Pradyuman murder case. Watch #AajKiBaat Tonight at 8.57 on @indiatvnews https://t.co/Xci5Ux...
5	indiatvnews	13-Sep-17	RT @RajatSharmaLive: More revelations in #Pradyuman murder Watch Special Show Tonight at 7.57 https://t.co/HIFtMu79mz
4	indiatvnews	14-Sep-17	RT @RajatSharmaLive: Clinching facts about #Pradyuman 's killer Tonight at 7.55 on @indiatvnews https://t.co/pnlH9mbP8z
3	indiatvnews	15-Sep-17	Haryana CM @mlkhattar visits family of 7-yr-old #Pradyuman who was killed in Gurugram's #RyanInternationalSchool https://t.co/ms2mrWGXoZ
5	thetribunechd	15-Sep-17	This case is handed over to CBI for investigation: Haryana CM @mlkhattar on 7-year-old #Pradyuman murder-ANI
3	IndiaToday	18-Sep-17	School of horror reopens. Should #RyanInternationalSchool waited for a CBI probe to begin? @preetichoudhry https://t.co/XQpi6p4Dp0
5	abpnewstv	23-Sep-17	#Pradyuman's father says he contacted PM Modi and various Min., but none replied https://t.co/dZca8Dle3H
3	indiatvnews	24-Sep-17	#Gurugram Municipal Corporation Election Result: BJP wins 13 out of 35 wards, 21 seats won by independent candidates, INLD wins 1 seat

Table 5 below shows the timeline generated for the third event of Japan PM's visit to India for Bullet Train. The selected tweets present sufficient emphasis on the major incidents describing the event. However, there is no obvious approach to validate the results but a quantitative analysis of tweet sequence justifies the relevancy of the timeline generated.

Table 5: Event timeline generated for Japan PM’s visit to India for Bullet Train

pR	User	Date	Social media post presenting the incident
5	punjabkesari	19-Jul-17	PM मोदी के एक फैसले ने छीन लिया 60 लाख लोगों का निवाला! https://t.co/A6O7HPEVci #PMModi .@narendramodi... https://t.co/n3qN4DtZ7r
4	ZeeBusiness	2-Aug-17	#BulletTrain to be ready by 2023; PM @narendramodi , Japan's Abe to lay foundation stone next month @sureshprabhu... https://t.co/sdq7EgCNeV
4	punjabkesari	4-Aug-17	उपराष्ट्रपति चुनाव से पहले आज #NDA सांसदों के साथ मीटिंग करेंगे #PMModi https://t.co/VUqvOUoztN... https://t.co/Hc0O1sRC5S
7	punjabkesari	7-Sep-17	ट्विटर पर चली PM मोदी को ब्लॉक करने की मुहिम https://t.co/kSagb44sIU #Twitter #PMModi #BlockNarendraModi https://t.co/MyOKkWLAPU
6	EconomicTimes	10-Sep-17	PM @narendramodi, @AbeShinzo to kick-start #India's first #BulletTrain project https://t.co/Ad6bvzPPgT https://t.co/PQv9H9qgID
24	EconomicTimes	11-Sep-17	India's first ever #bullettrain project to start on September 14. Check details https://t.co/r2HzNN5yR5 https://t.co/JuacycMFVc
5	EconomicTimes	11-Sep-17	Work on #bullettrain project to start on September 14: @PiyushGoyal https://t.co/tufO8y3LMf
8	EconomicTimes	11-Sep-17	.@narendramodi, @AbeShinzo to kick-start India's first #BulletTrain project https://t.co/FdGpv5RuhY
4	indiatvnews	13-Sep-17	PM @narendramodi and Japanese PM #ShinzoAbe's road show to Sabarmati Ashram in #Ahmedabad https://t.co/qSieTIwy3p
4	WIONews	13-Sep-17	#Gravitas 9PM with @palkisu, find out how #Japan PM #ShinzoAbe's visit to #India could be a game-changer in power e... https://t.co/3gcq0PH7X
4	WIONews	13-Sep-17	#Japanese PM #ShinzoAbe arrives in India. He is personally greeted by Indian PM #Modi https://t.co/vyIHeuNOtI
5	EconomicTimes	13-Sep-17	350 kmph speed, undersea #travel and a boost to #Modi's job promise: This is what #bulletTrain brings for #India https://t.co/RG3MWgyPhh
4	EconomicTimes	14-Sep-17	In Pics: #India's first #bullettrain will also run under #ArabianSea https://t.co/2EvZJhUseS https://t.co/dTwOeGUOD3
6	htTweets	14-Sep-17	#ShinzoAbe visit: India, Japan sign 15 agreements, PM Modi says they will strengthen partnership. Live updates here... https://t.co/S4DcvSGOek
8	IndiaToday	14-Sep-17	#TTP #BulletTrain project Why is Congress so much against Gujarat & Gujaratis?: @sambitswaraj... https://t.co/bOerxRT9E5
55	BTVI	15-Sep-17	RT @PIB_India: How will the #BulletTrain look like? A first look... https://t.co/xZVyxijWLV
8	NavbharatTimes	15-Sep-17	Funny Video: पटरी पर हल्का होने वालों ने किया #BulletTrain का विरोध, पटरियों के पास Western Toilet बनाने की मांग की... https://t.co/qCDLcrz9YB
4	NewsX	17-Sep-17	#Congress' Manish Tewari uses expletives for #PMModi & 'bhakts' in a tweet. @INCIndia @BJP4India @ManishTewari... https://t.co/ouNrpflPEn
4	indiatvnews	17-Sep-17	9 unknown facts about #SardarSarovarDam which #PMModi inaugurated on his 67th birthday #HappyBirthdayPM https://t.co/FAWeS27n6u
11	NewsX	22-Sep-17	#Varanasi: Protest in #BHU ahead of #PMModi's visit @narendramodi @BJP4India https://t.co/rjWtmaR975
4	EconomicTimes	24-Sep-17	Watch FM @arunjaitley gets interrupted while speaking on #BulletTrain at a seminar in Delhi, reprimands heckler (... https://t.co/0GoUO88Vdz

5. Conclusion and Future work

In this research work, we proposed an event timeline generation for highly influenced events. Our approach requires the hashtag clustering as a prior knowledge. The hashtag clustering is performed using a genetic algorithm without accessing any domain specific information. We have collected real-life tweet streams extracted from 76 Indian media twitter accounts and evaluated the performance of the proposed algorithm. Our approach is simple and generalized so, it can be applied to the generation of timeline for major events. The novel proposed *prominenceRank* based heuristic automatically identifies the tweets that contribute the necessary and diversified information to the incidents about the specific event. The timeline generated becomes a meaningful resource for many applications like generating a Wikipedia article, writing a newspaper report or forming a historical story or documentary etc... To prove the efficiency of the approach, we consider three most influenced events and generated the timeline. The quantitative analysis of the experimental result demonstrates the applicability and efficacy of the proposed work. Currently, we use the ratio of retweet count and average user's retweet value to define the *prominenceRank*. In the future work, this can be further advanced to include the number of likes received by a tweet in order to emphasize on the popularity of tweet as well.

Acknowledgment

We are very much thankful to the Nirma University for providing resources and other facilities to carry out this research work.

References

- [1]. A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities". In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pp. 56-65, ACM, August 2007.
- [2]. D. E. Goldberg. Genetic algorithms in search, optimization, and machine learning, 1989. Reading: Addison-Wesley, 1989.
- [3]. F. Fan, R. Qiang, C. Lv, W. X. Zhao, and J. Yang, "Tweet timeline generation via graph-based dynamic greedy clustering". In: AIRS (2015)
- [4]. H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?". In Proceedings of the 19th international conference on World wide web, pp. 591-600. ACM, April 2010.
- [5]. J. H. Holland. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press, 1992.
- [6]. J. Li and C. Cardie, "Timeline generation: tracking individuals on twitter". In WWW, 2014.
- [7]. J. Yao, F. Fan, W. X. Zhao, X. Wan, E. Y. Chang, and J. Xiao, "Tweet timeline generation with determinantal point processes". In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA., pages 3080–3086, 2016.
- [8]. M. Mitchell. An introduction to genetic algorithms. MIT press, 1998.
- [9]. N. Gambhava and K. Kotecha, "Social Media Hashtag Clustering Using Genetic Algorithm". International Journal of Advanced Research in Engineering and Technology, 9(1), pp 12–25, 2018
- [10]. W. Feng, C. Zhang, W. Zhang, J. Han, J. Wang, C. Aggarwal, and J. Huang, "STREAMCUBE: hierarchical spatio-temporal hashtag clustering for event exploration over the twitter stream". In Data Engineering (ICDE), 2015 IEEE 31st International Conference, pp. 1561-1572. IEEE, April 2015.
- [11]. W. Magdy, T. Elsayed, and M. Hasanain, "On the Evaluation of Tweet Timeline Generation Task", Springer International Publishing, pp. 648–653, 2016.
- [12]. Y. Chang, J. Tang, D. Yin, M. Yamada, and Y. Liu, "Timeline summarization from social media with life cycle models". In IJCAI, pages 3698–3704, 2016.
- [13]. Y. Zhou, N. Kanhabua N, and A. Cristea, "Real-time timeline summarisation for high-impact events in Twitter". In: 22nd European conference on artificial intelligence, ECAI'16, pp 1158–1166, 2016.
- [14]. Z. Wang, L. Shou, K. Chen, G. Chen, and S. Mehrotra, "On Summarization and Timeline Generation for Evolutionary Tweet Streams", IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 5, May 2015.