International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
Available at http://inpressco.com/category/ijcet
©2021 INPRESSCO®, All Rights Reserved

Research Article

Terrorist Activities Detection via Social Media Using Big Data Analytics

Miss.Manisha Sagar Prof.M.B.Vaidya

Department of Computer Engineering AVCOE, Sangamner. Savitribai Phule Pune University Pune, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Social media are perhaps the richest source of human generated text input. Opinions, feedbacks and criticisms provided by internet users reflect attitudes and feelings towards certain topics and issues. The large volume of such information makes it impossible for any group of people to read it. Thus, social media has become an important tool for spreading their opinions and influencing or attracting people in general to join their terrorist activities. Twitter is the most common and simple way to reach many people in a short time. In this paper, focused on the development of a system that can automatically detect terrorism-supporting tweets by real-time analytics using machine learning framework. Proposed system is entirely dependent on training data and tries to improve accuracy. This system will help to block the terrorist accounts from twitter so that they can't promote their views or spread fear among ordinary people.

Keywords*: Psychological pressure, text mining, sentiment analysis, social media, machine learning.*

Introduction

Social media, today, plays arguably a very essential part in a life of not only an individual but also in functioning of a government. Statistics shows that the worldwide population is about 7.72 billion where internet users are not less than billions. There are about 3.397 billion active social media accounts with 5.54 social media accounts per person. The impact of social media on society and its people and the data that generates in the wake of it has been discussed. The progress of 21st century can barely be anticipated without the indication of the part of social media in it. It wouldnt be overstating to say that social media is ubiquitously present in all spheres of life, be it education, health care, business, disaster management, politics, tourism industry and of course the use of media sharing and entertainment needs no mention. In the wake of all such convenience provided by the social media, it too, does have a darker side to cast. Misuse of social media, the other side of the coin, also needs to be accounted. On one hand this may seem to be abridging a communication gap and faster news delivery among people; however, on other hand, it is being heavily misused by many. Misuse on a level of genocide, murders, bombings, conspiracies etc. In a study conducted by [6], he states that about 90% of terrorist activities online are con- ducted via social media platforms while 76% of U.K. terrorists engage in internet to research and strategizing their actions. In 2013, ASG, a Salafi Jihadist terrorist organization, kidnapped Australian Warren Rodwell and held him for about 472 days. The group made use of YouTube and Facebook for ransom videos and to demonstrate the proof of life. A robust literature conducted by, revealed the extensive use of social media by Islamic State (ISIS) to advertise their ideology and enrole members and supporters. The usage of pro-active social media tactics by three extremistrelated groups operational in Asia Pacific namely, Abu Sayyaf in the Philippines; Jamaat-e-Islami in Bangladesh; the Uyghurs in China has been examined by. This study explains how these groups affords numerous opportunities to exploit their reach, impact and effect using social media. Another such misuse of social media has been examine who presents the social media activity and online presence of media mujahedeen who are deemed to be the supporters of jihadist groups and circulate propaganda content online.

A. Objectives

- To find terrorist activities using social media.
- To block terrorist accounts.
- Try to stop terrorists activities.
- To try to improve detection accuracy using machine learning algorithms.

Review of Literature

A model to predict information flow size and survival was developed by [1], with the help of data fetched from one of the popular social networking website Twitter. The information flow size and survival were modeled using zero truncated negative binomial (ZTNB) regression method and Cox re- gression technique respectively. Using a sample of 427,330 Twitter data, they reported a novel outcome that identified the sentiment expressed in the tweet which was found to be statistically predictive of both size as well as survival of information flows of such nature. Additionally, co-occurrence of URLS and the time lags among retweets and hashtags also proved substantial.

In a similar study, [2] reported how investigation of an open source communications data gathered through social media platforms could elucidate the inter- and intracommunity con- flict dynamics, surfacing in the wake of such unfortunate events. They claimed that the Twitter data gathered after the murder of Fusilier Lee Rigby, convincingly supports the Collins three phases of conflict dynamics. They also analysed two key claims, first regarding the interactive nature of the conflict, and second on how the detail provided by the digital data give compelling insights into the complex network of relationships that come forth and develop over the course of such dispute.

Another interesting study of this event was reported by [3], who worked this case study as a part of computational criminology. They showed the temporal variation in cyberhate that is relative to concepts of much criminological theory, like the diffusion, duration, escalation and de-escalation of crime.

Analysis of social reactions to the murder of Lee Rigby, was studied by [4] using data collected by systematic monitoring of twitter. They investigated a number of online behaviors with offline effects.

In the annual Boston Marathon on April 15, 2013, two homemade pressure cooker bombs detonated in the vicinity of the finish line of the race. This killed three people and several hundred others were injured. About sixteen people lost limbs. The social media posts uploaded instantly after the bombings were examined by [5]. They found specific keywords to appear regularly before the official public safety and media reports. People adjacent to the explosions posted messages within minutes via Twitter. This helped to detect the location and details of events. This showcases the role of social media in the ahead of time identification and portrayal of emergency events.

Study of this event was also undertaken by [6]. They examined the influence of features of tweet on the dispersion of two categories of messages marathon tragedy namely, real and rumor related (both in the context of the Boston tragedy). Negative binomial analysis showed that tweet features like usage of hashtag, number of followers and reaction time have an impact on tweet message diffusion during the bombing. The number of followers illustrated a positive relationship with message dispersion. However, the relationship between reaction time of tweet and message dispersion was, however, negative. Interestingly, messages that without hashtags were spread more than those with hashtags.

A similar study of this attack was reported by [7], who gathered more than 18 million tweets from 15,509 tweeter users in Paris on November 13, 2015. They measured the level of their anxiety, anger, and sadness post-attacks. The authors proposed the use of computational focus groups and a completely novel investigation framework to evaluate a social media stream which archives user location and history. The study resulted in outcomes that would be unlikely to manifest through other media or methods.

This study[8] identified the Helpers to be the Convergence Behavior Archetypes who were the most to retweet all over the crisis. On other hand, the Mourners had the highest impact by retweeting the most. This indicated that those users who create emotional content tend to retweet the most. Additionally, the Detectives disseminated information into other communities the most. The authors not only extended the knowledge on how users converge on social media in crisis situations, but also help the crisis managers to get more insight in users behavior. Knowing which type of behavior on social network has a effective impact, might help in controlling the amount of data that is generated during a crisis situation.

Twitter became an essential channel of communication between the emergency responders, government and the public. This highly facilitated the emer- gency management of the crises. A comprehensive analysis of understanding crisis communication trends mediated by social media was studied by [9]. TwitterMate was used to collect the data generated during tweets and also to analyze it. It also identifies the main hashtags surfaced by the crowd and specific Twitter accounts of individuals, NGOs and emergency respon- ders. A total of 67,849 tweets were gathered and examined. Four primary types of hashtags were detected: social support, terror attack, geographical locations and organizations.

This terror attack has been also examined by [10]. They investigated the number of tweets, geographical location of tweets and users demographics. They also evaluated if users in developing countries are inclined to tweet, retweet or reply during the event of a terrorist attack. They Define new metrics which about reach and impression of the tweet. They reported that, users from developing countries are inclined to tweet more initially and at critical period of the terrorist occurrence. Furthermore, hefty number of tweets originated from the Kenya with 23% from women and 73% from men. Also, original posts had a highest number of tweets followed by replies and retweets.

Proposed Methodology

With all the real time data retrieved from twitter, system aimed to relate some semantics with the data. The system wanted to analyse the data and find out if there were any patterns. We aimed to find out what were the words and Hashtags that were trending the most and who was at the center of the network graphs.

The system wanted to see how many unique users data had and what the popularity on different parameters was for different crises. Yet another aim of our proposed system is to find out the crisis orientation of the users on Twitter. To look at which who focuses on which behaviour. The system will analyse the main parameters like anxiety, anger, and sadness about postattacks. Since a lot of data is getting generated every day, there need a system which could generate daily analysis. The system wanted to be able to see the tweets related to terrorism attacks as and when they happened. There was a need to keep a track of the terrorism attacks related tweets.

A. Architecture

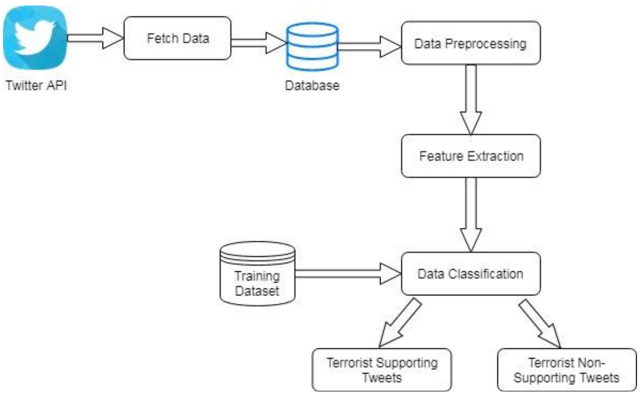

Fig. 1. Proposed System Architecture

B. Algorithm

Preprocessing Algorithm

Data preprocessing removes redundancy and ambiguity inherit in the data and transforms the reviews into sentences to facilitate sentence-level aspect-based classification. First, sentences are extracted by identifying the delimiters (e.g. dot, exclamation or question mark). Next, redundant information, e.g. duplicate sentences, is removed. Finally, ambiguous, vague or misspelled terms are corrected.

1. Stop word Removal-This technique removes stop words like is, are,they,but etc.

Initialize i,j

for i=1 to no of words in documents for j=1 no of words in stopword list if

Words(i)==Stopwords(j) then eliminate words(i) end if end for

2. Tokenization-This technique removes Special character and images.

Initialize feature vector bg feature =[0,0..0] for token in text.tokenize() do if token in dict then token idx=getindex(dict,token) bg feature[token idx]++

else

continue end if end for

3. Stemming– Removes suffix and prefix and Find Original words for e.g.- 1. played – play 2.Clustering - cluster

The word w

Input = Normalize(input) if normalizeValidate(input) then return input; for each rule in rules do if input match with rule then

Stem = ExtractStem(input,rules) if not TestStemLength(Rule) then end for return input

4. Random Forest Machine Learning Algorithm:

Step 1: Let the number of training cases be N, and the number of variables in the classifier be M.

Step 2: The number m of input variables to be used to determine the decision at a node of the tree; m should be much less than M.

Step 3: Choose a training set for this tree by choosing n times with replacement from all N available training cases (i.e. take a bootstrap sample). Use the rest of the cases to estimate the error of the tree, by predicting their classes.

Step 4: For each node of the tree, randomly choose m variables on which to base the decision at that node. Calculate the best split based on these m variables in the training set.

Step 5: Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier).

For prediction a new sample is pushed down the tree. It is assigned the label of the training sample in the terminal node it ends up in. This procedure is iterated over all trees in the ensemble, and the average vote of all trees is reported as random forest prediction.

C. Mathematical Model

The mathematical model for Terrorist activities Detection System is as-

$S = \{I, F, O\}$

Where,

I = Set of inputs

The input consists of set of Words. It uses Twitter dataset.

F = Set of functions

$F = \{F1, F2, F3 \ldots FN\}$

F1:Tweets Extraction

F2:Tweets Preprocessing

F3: Feature Extraction F4: Classification

O:Terrorist activities detection

Results and Discussion

Experimental evaluation is done to compare the proposed system with the existing system for evaluating the performance. The simulation platform used is built using Java framework (version jdk 8) on Windows platform. The system does not require any specific hardware to run; any standard machine is capable of running the application.

A. Comparison with Existing System

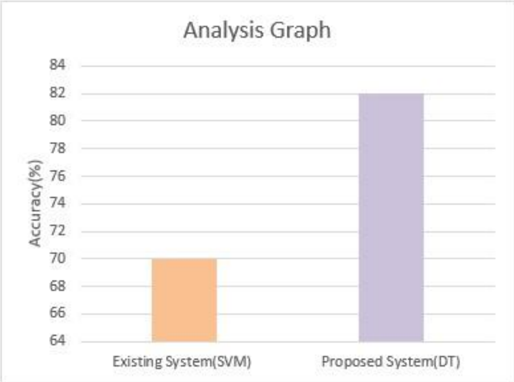

Fig. 2. Comparative Graph

Table 1:Comparative Result

Sr. No.	Existing System(Naive Bayes)	Proposed System(Random Forest)
1	70%	82%

B. Classification Performance

Tweets are retrieved in a streaming way, and Twitter dataset provides the Streaming API for developers and researchers to access public tweets in real time. The aim of this paper is to bridge the gap by carrying out a performance evaluation, which was from two different aspects of NLP and machine leaning algorithms.

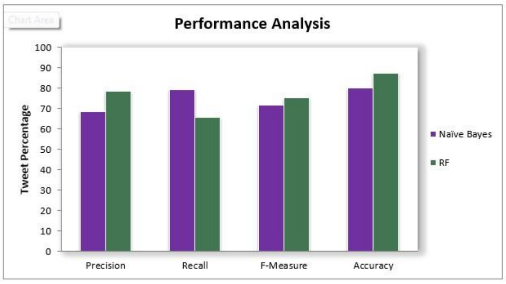

Fig. 3. Classification Performance

Conclusion

With the increased use of social media the current work focused mainly on use of social media as a tool for terrorism. India which is known to be one of the wide countries in the world with having more than 65% of its youth below age-group of 35;Social media plays vital role in the life of this young youth. The proposed systems will try to analyze the a common platform to manifest the progress of counterterrorism strategies in this digital world; There is scope for verifying the changed sentiments of the user before and after an attack.

Considering the studies reported and analysis performed hitherto, it to be the need of the hour to increase the magnitude of data analysis on a much larger scale and more so on a regular basis. This should be done not only to identify the acts of terrorism on social media but also as a safety tool, preventive measures and post-attack examination. The study can also include a machine learning approach to train a system to automatically classify the tweets and do a sentiment analysis of the tweets/comments.

References

[1] M. Mirbabaie, D. Bunker, A. Deubel and S. Stieglitz, Examining Convergence Behaviour During Crisis Situations in Social Media-A Case Study on the Manchester Bombing 2017, In International Working Conference on Transfer and Diffusion of IT, Springer, Cham. pp. 6075, June 2018.

[2] M. L. Williams and P. Burnap, Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data, British Journal of Criminology, vol. 56(2), pp. 211238, 2015

[3] Roberts, C., Innes, M., Preece, A., and Rogers, D. (2017). After Woolwich: Analyzing open source communications to understand the interactive and multi-polar dynamics of the arc of conflict. The British Journal of Criminology, 58(2), 434-454.

[4] Martin Innes, Colin Roberts, Alun Preece and David Rogers (2018) Ten Rs of Social Reaction: Using Social Media to Analyse the Post-Event
Impacts of the Murder of Lee Rigby, Terrorism and Political Violence, 30:3, 454-474, DOI: 10.1080/09546553.2016.1180289

[5] C. A. Cassa, R. Chunara, K. Mandl, and J. S. Brownstein, Twitter as a sentinel in emergency situations: lessons from the Boston marathon explosions, PLoS currents, vol. 5, 2013

[6] J. Lee, M. Agrawal and H. R. Rao, Message diffusion through social network service: The case of rumor and non-rumor related tweets during Boston bombing 2013, Information Systems Frontiers, vol. 17(5), pp. 9971005, 2015

[7] P. Burnap, M. L. Williams, L. Sloan, O. Rana, W. Housley, A. Edwards and A. Voss, Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack, Social Network Analysis and Mining, vol. 4(1):206, June 2014.

[8] Y. R. Lin, D. Margolin, and X. Wen, Tracking and analyzing individual distress following terrorist attacks using social media streams, Risk analysis, vol. 37(8), pp. 15801605, 2017.

[9] T. Simon, A. Goldberg, L. Aharonson-Daniel, D. Leykin, and B. Adini, Twitter in the cross fire the use of social media in the Westgate Mall terror attack in Kenya, PloS one, vol. 9(8), pp. e104136, 2014.

[10] F. R. Ishengoma, Online social networks and terrorism 2.0 in developing countries, arXiv preprint arXiv, 1410.0531. 2014

[11] M. Mirbabaie, D. Bunker, A. Deubel and S. Stieglitz, Examining Convergence Behaviour During Crisis Situations in Social Media-A Case Study on the Manchester Bombing 2017, In International Working Conference on Transfer and diffusion of IT, Springer, Cham. pp. 6075, June 2018

[12] Bunker, D., Mirbabaie, M., and Stieglitz, S. (2017). Convergence be haviour of bystanders: an analysis of 2016 Munich shooting Twitter crisis communication. In Proceedings of the Australasian Conference on Information Systems.

[13] M. Mirbabaie, and E. Zapatka, Sensemaking in Social Media Crisis Communication A Case Study on the Brussels Bombings in 2016, 2017

[14] A. Gupta, T. zyer, J. Rokne, and R. Alhajj, Social Network Analysis to Combat Terrorism: 2015 Paris Attacks, Social Networks and Surveillance for Society, pp.165-179

[15] T. Simon, A. Goldberg, L. Aharonson-Daniel, D. Leykin, and B. Adini, Twitter in the cross fire the use of social media in the Westgate Mall terror attack in Kenya, PloS one, vol. 9(8), pp. e104136, 2014

[16] C. A. Cassa, R. Chunara, K. Mandl, and J. S. Brownstein, Twitter as a sentinel in emergency situations: lessons from the Boston marathon explosions, PLoS currents,vol. 5, 2013