ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

Malicious tweet detection and disaster reporting using topic and user behavior analysis

[1] Dhanashri R.Deosarkar, [2] Sandeep Gore
[2] Assistant Professor
[1][2] Department Of Computer Science And Engineering of G.H.Raisoni College of Engineering and Management, Pune, India.

***Abstract* – The social networking sites received much more attention recently in that tweeter is used more than others. In tweeter micro blogging services received more attention. Micro blogging is used to blog the words which are related to that topic, that depend on three behavioral factors which are t user virality, topic virality and user susceptibility. In proposed system Malicious tweets identifies using traffic patterns in that this system uses click traffic analysis method and for the malicious URL uses the URL shortening websites to identifies blacklisted URLs. The URL shorteners are used to sharing URLs on Twitter, because tweeter having 140 character tweet limit per message. Spammers uses the URL shorteners to improve the user quality of their spam URLs. Our proposed system provides integrated approach for the spam detection from different tweets. The integrated approach includes the different machine learning techniques, spam URL detection and NLP. Firstly this system identifies the sensitivity of tweet depend on the topic varality or user virality after that micro blogging is used to calculate the tensor factor which means tensor factor is used to calculate the user impact on that tweet. Last module is disaster event reporting in this if the event like earth quick is occurred then it sends the mail or message to the people witch are present in that area.**

Keywords- click traffic analysis, URLs, natural language processing, earthquakes, micro blogging.

I. INTRODUCTION

. Among the different social networking sites TWITTER has received much more attention recently. Tweeter is used by billions of people to connect socially to their closedone, friends and coworkers through their mobile and computer. Twitter having limit of 140 character, if tweeter ask any question then answer must be less than or equal to 140 character. For that reason different spammers using the URL shortner to short their URL into 140 character limit[1], spammers are also used the URL shorteners to improve the quality of spam URLs[2]. Firstly this system identifies the sensitivity of tweet depend on the topic varality or user virality after that micro blogging is used to calculate the tensor factor which means tensor factor is used to calculate the user impact on that tweet[2]. If the disaster event occurred in any area then alert people in that area related to that disaster event via mail or message[3].

PROJECT IDEA

In proposed system twitter data analysis can be performed on basis of user behavior analysis in terms user communication on twitter with aspect of topic and user virality, user sensitivity. In which topic virality measures ratio of discussion subject attraction by twitter users. User virality demonstrate user tweet in various aspect for micro blogging. User sensitivity analysis in terms user behavior on social media like facebook, twitter blogs. Proposed system modulate tensor model for blog propagation according user tweets. In proposed system a tensor factor is used to find sets of behavioral factors. Based on this framework, proposed system develops a probabilistic factorization model and numerical factorization mode[2]. Proposed work uses efficient algorithm based on large dataset and synthetic dataset for model propagation. This work enhances the existing work for blog propagation for user tweet with tweet dataset online. Online micro-blogging propagation model used stemming operation over twitter data to generate blog statement. Naïve bayes classifier differentiates the proposed work for blogging classification in the tweeter data.

The given paper is organized as follows. Motivation of the problem is in section 2. Section 3 contains the existing system and section 4 contains the proposed system in which it explains the three modules of proposed system. We conclude and acknowledge the paper in section 5.

II. MOTIVATION

The challenges come from the large and heterogeneous nature of the social media application like twitter, face book, social media blogging, which makes it difficult to generate and recommend situational and social blogging.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, March 2018

Online tweet analysis to generation tweet summary for various kind of analysis like vote prediction, public survey, usefulness, trustworthy content or information needed.

1. Twitter content propagation in micro-blogging from users tweet analysis from their followers to followers. Tweet are exposed based on
   (a) Popularity of the tweet sender
   (b) Malicious tweet detection.
   (c) Virality of the topic
   (d) Sensitivity of the receiver.
2. Content summarization for micro-blogging from user tweet online.
3. User tweet event analysis from their tweet data.
4. Tweet behavior classification using tensor model.
5. Tweet micro-blogging.
6. Situational tweet event analysis to prediction and analysis of disaster event.

III. EXISTING SYSTEM

Existing method based on two different aspects: the identification of malicious tweets without checking background of the user; and the other based on analysis of language for detecting spam on twitter in such topics that are in trending at that time[4]. Trending topics are the topics of discussion that are popular at that time. We first collected the tweets related to many trending topics, labeling them on the basis of their content which is either malicious or safe. After a labeling process we extracted a many features based on the language models using language as a tool. We also calculate the performance and classify the given tweets as spam or not spam[7]. Thus our system designed for detecting spam on Twitter and disaster event reporting , focusing mainly on analyzing of tweets instead of the user accounts. The microblocking concept is used in Japan for earthquake detection they developed the system which is used for earthquake detection in japan. The previous system is designed with the SVM classification algorithm but the proposed system using the Naïve Bayes classifier which is more efficient than the previous work.

IV. PROPOSED SYSTEM

Proposed system is designed to automatically malicious tweet detection and disaster event reporting based user specific and topic specific behavior analysis from user tweet.

Malicious Tweet Detection: An integrated approach for malicious tweet detection based on natural language processing for spam detection to find malicious tweet online. First it collect the tweets and then label them[1]. In labeling it also check the URL spamming[3][6].

There are four types of spammers are identified on tweeter these are phishers, Malware Propagators, Marketers, Adult content Propagators[6].After that Natural language processing is used for feature normalization, case folding, remove stop words[5].

Content Propagation in microblogging: Content propagation model is implemented for micro blogging topic virality and user impact factorization on twitter among follower and followed user[2]. Naïve Bayes classifier is used this gave the accurate result as compared to other existing classifiers. Here Tensor factorization is used to find users impact over tweeter network.

Disaster Reporting: Disaster event detection and reporting based on real time tweet from twitter[4]. In this proposed system is designed to report for disaster event to tweeter user.

Proposed disaster events can be earthquake, tsunami, cyclone and social media impact on twitter is reported to users. After checking of disaster event alert people of that area via mail or message.

In fig process of the given proposed system is explained. There are three modules are present in this system that are sentiment analysis, microbolgging and disaster event reporting. Algorithms used in given system are NLP, Naïve Bayes classifier.

Architecture

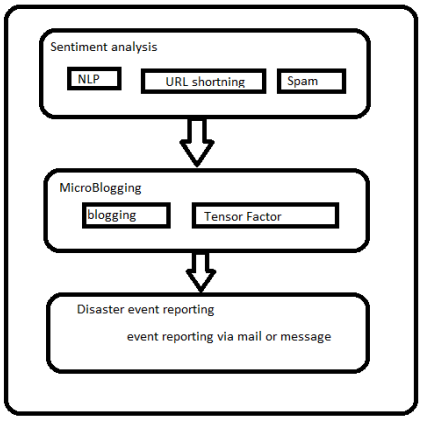

Fig. 1 Architecture of Malicious Tweet detection

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, March 2018

V. CONCLUSION

Proposed system focuses on user and content factors for content malicious tweet detection based on user and topic behavior analysis from user tweet. Enhancement to existing model proposes textual summarization of user tweet to generate blog generation with help of user review after consideration for user tweet data analysis.

Proposed system evaluate the spam tweets and depend on topic and user varality check if any disaster event occurred, If event occurred then system alert people in that area via mail. In the future, this work can be expressed with multi factor data mining by user behavior analysis from different social media sentiment. These factors include users' tweet as sentiment for positive and negative feedback for malicious tweet detection and disaster event reporting.

ACKNOWLEDGEMENT

I would like to thanks my Guide Mr. Sandeep Gore for their continuous help and encouragement. Also a word of appreciation for Department of CSE, G. H. Raisoni College of Engineering and Management for providing useful resources.

REFERENCES

1] Sagar Gharge, Manik Chavan "An Integrated approach for Malicious Tweets detection using NLP" 978-1-5090-5297-4/17 2017 IEEE.

2] Ee-Peng Lim, Tuan-Anh Hoang "Microblogging Content Propagation Modeling Using Topic-Specific behavioral Factors" in ieee transactions, vol. 28, no. 9, september 2016

3] Danesh Irani, De Wang, "Click Traffic Analysis of Short URL Spam on Twitter" conference on collaborative computer networking, application and worksharing 2013.

4] Makoto Okazaki, Takeshi Sakaki, Yutaka Matsuo and Makoto Okazaki, "Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development" 1041-4347/13 _ 2013 IEEE Published by the IEEE Computer Society.

5] Diotra Henriyan, Mochamad Vicky Ghani Aziz , Rifki Wijaya and Ary Setijadi Prihatmanto, "Design and Implementation of Natural Language Processing with Syntax and emantic Analysis for Extract Traffic Conditions from Social Media Data" 978-1-4673-6713-4/15/$31.00 ©2015 IEEE

6] Chao Chen, Jun Zhang, Yi Xie, Yang Xiang Wanlei Zhou "A Performance Evaluation of Machine Learning-Based Streaming Spam Tweets Detection" 2329-924X © 2016 IEEE

7] Kamalanathan Kandasamy and Preethi, "An Integrated Approach to Spam Classification on Twitter Using URL Analysis, Natural Language Processing and Machine Learning Techniques" 978-1-4799-2526-1/14/$31.00 ©2014 IEEE

8] Zi Chu, Steven Gianvecchio, Haining Wang, Sushil Jajodia "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?" 1545-5971/12/$31.00 _ 2012 IEEE Published by the IEEE Computer Society

9] William Deitrick, Brian Dickinson , Wei Hu, Alex Hai Wang and Zachary Miller, "Twitter spammer detection using data stream clustering" 2013 Published by Elsevier Inc.

10] Juan Martinez-Romo , Lourdes Araujo "Detecting malicious tweets in trending topics using a statistical analysis of language" 0957-4174/$ - see front matter _ 2012 Elsevier Ltd. All rights reserved.