Automation on Twitter Accounts.

Swati krishna Dehade
Department of Information Technology
Pune Institute of Computer Technology
Pune, India
Swa.dehade@gmail.com

Prof. A.M.Bagade
Department of Information Technology
Pune Institute of Computer Technology
Pune, India
ambagade@pict.edu

Abstract— The Online social networks like Facebook or Twitter have become influential information diffusion platforms as they have attracted hundreds of millions of users. As there is the possibility of reaching millions of users within these networks not only have gained attention standard users, but also cyber-criminals who violate the networks by distributing spam. This is can be carried out either by creating fake accounts, bots, cyborgs or by hacking and compromising accounts. The Fake or compromised accounts are subsequently used to spread spam in the name of their legitimate owner. The popularity of Twitter attracted large number of automated programs known as bots which appear two edged sword. On the One side of which authentic bots are present, generating a huge amount of nonthreatening tweets, e.g. blog updates/news which complies with twitters objective of becoming a news information web. While the other side of this is malicious bots which are used by the spammers for Spreading spam .On top of this, Cyborgs have emerged in between human and bot which is referred to either bot-assisted human or human-assisted bot. This paper focuses on the classification of human, bot and cyborgs accounts on Twitter. We propose a classification system that is divided into following parts: Data Collection, Classifying Twitter and Detecting Account.

Keywords- Automatic identification, bot, cyborgs, Twitter, social networks.

I. INTRODUCTION

Twitter is one of the most red hot online social networking and microblogging services, wherein its users interact with each other by sending and receiving text based posted which are termed as tweets. The tweet size is limited up to 140 characters. These days millions of Users are using twitter in their day to day life to keep in touch with their friends, meeting new peoples and for discussions. As twitter is one of the fastest growing applications, it has also attracted cyber criminals. This new platform has been used by the cyber criminals to achieve their malicious goals of spreading spam. The increasing fame of microblogging sites like Twitter has resulted rise in social networking scams .The Twitter scam uses tempting messages like ‘‘Just saw this photo of you’’ followed by a link after clicking on it will take you to a site that uploads malware onto your computer. Also by exploiting the phishing techniques, the message may appear to come from one of your regular followers, possibly even a friend or relative. Actually Twitter account has been hijacked. The list of trending topics is one of the most well-liked tools in twitter, which capture the most recent emerging trends and topics of discussion. By using this quality of twitter, people can immediately collect news about a particular subject or learn at a quick look which is the most popular topic on which most people speak. Unfortunately, this rising and open structure of microblogging phenomenon allows spammers to distribute malicious tweets. Twitter users are provided with several methods to report spam which are then investigated by the twitter and the accounts are then suspended in case of spam. However, the suspension process is slow. Our approach for the spam hitch focuses on the recognition of tweets consisting spam instead of identifying spam accounts. An enhanced technique to identify spammers would be filtering users who have written many spam tweets.

Commenting and Real-time event reporting are the most popular activities on social-networking sites such as Facebook and Twitter. But the approaches of these two websites are completely different, the former one is about staying in touch with their friends and family hence the messages are generally private. Whereas in later one message are by default public. In addition, to this in built functionalities such as retweeting, trending topics and hashtags makes twitter a very simple way of spreading news rapidly all through the network.

Twitter is hugely well-liked with More than 100 million active users posting about 200 million tweets per day. The simplicity of information diffusion on Twitter and huge users makes it a trendy means to spread outside content like photographs, articles, and videos, by including URLs in tweets. yet, these URLs may also connect to low quality content like, phishing websites , malware or spam websites. The statistics shown recently contains 8% spam tweets and other malicious contents.

Who is tweeting?

Many people are using Twitter, this is the reason for spammer and other mischief’s to break into the network has been rising. It has been shown by the researcher that on click ratio for spam URLs included in tweet posts is much higher than in email Furthermore, there is one more kind of spam: Twitter accounts that are used to spread propaganda even though such accounts have a clear rival agenda, there are also other types of accounts on Twitter that do not act normally:

Broadcaster accounts: Such accounts put on air headlines Non-stop. In this category is also the large number of media organization accounts (CNN, Fox News,NPR, etc.)

Activist accounts: Such accounts belong to political activists who continuously tweet and retweet the same content for political motives.

The ultimate approach of this paper is identify and classify automation feature of Twitter accounts into 3 categories, human, bots, and cyborgs which we will manage. This will help Twitter to have healthy community tweets and also human users to recognize the real tweets.

The paper is organized as follows: Section II presents the literature review. In Section III, a description about proposed system . Section IV mathematical model. Section V gives the experimental results finally concludes the paper.

.

II. RELATED WORK

Measurement and Classification of Humans and Bots in Internet Chat [2]. In this paper, the authors carried out measurements on a big commercial chat network that captured a total of 14 types of chat bots ranging from simple to advance. Furthermore, it is has been seen that human activities are more complex than bot activities. Based on the measurement study, they proposed a classification system to correctly differentiate chat bots from human users. The proposed classification system consists of two components: (1) an entropy-based classifier and (2) a machine-learning-based classifier. The two classifiers go together in chat bot detection.

Detecting Spam in a Twitter Network [3]**.**In this paper, the author has studied, to which extent spam has got into social networks. To gather the data about spamming bustle, they created a big and varied set of "honey-profiles" on three large social networking sites, and logged the kind of contacts and messages that they acknowledged. Further then, conducted analysis on the gathered data to recognize irregular behavior of users who contacted their profiles. And proposed a method that identified spammers in social networks based upon the analysis of the irregular behavior.

Usability of CAPTCHAs or usability issues in CAPTCHA design [4].Generally most websites adopt (CAPTCHA) Completely Automated Public Turing test to tell Computers and Human Apart. A challenging approach to isolate web robots from humans . But, CAPTCHA is difficult for user and research illustrate that it can be evade by machine learning algorithm. As a result it is not a practical and safe way out for stopping Spambot.

Jaber et al. [5] Proposes an automatic ways of generating web pages data as train data set for web spam categorization using semi-supervised learning algorithm. However it is a physical process of web page labeling which is time requiring and labour intensive but still it is just an approach to categorize web spam rather than Spambot detection which is the basis of Web spam data.

Cailing Dong and Bin Zhou, [6] Studied a range of content-based and link-based features spam classification model. They carried a detailed analysis of content spam on the web using topic models and propose a number of new current variety measures for content spam detection. This is still a development on content recognition as contrasting to host detection which is the spambot.

Mohammed et al. [7] examine four types of classification algorithms (naïve Bayes, decision tree, SVM and K-NN) to detect Arabic web spam pages in search engine, based on content and came to the result that decision tree is the best. This is a good work at removing the biasness in search result as a result of web spam intrusion.

III. PROPOED SYSTEM

This paper describes our automated system for classification of users of twitter. The system classifies users into three categories: human, bot, and cyborg. The system consists of following components: Classifying Twitter, Data Collection and Detecting Account. The high-level design of our Twitter user classification system is shown in Fig no 1.

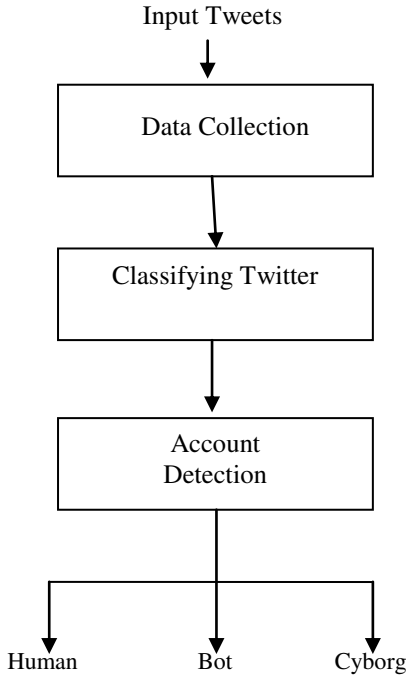

Fig no 1 Proposed Work

In classification component, the data set contains tweets posted by the users in their account time spam, with the help of which we can extract useful features for classification, like tweeting behaviors and text patterns. User API will gather the information of active users in the data collection component. The crawler calls the timeline API to collect the authors of the tweets included in the timeline. Lastly in detecting account uses tweet content, and account properties i.e account-related properties to catch bot variation from the regular human distribution.

Modules:

Data Collection:

The information of active users will be collected using the User API. The crawler calls the timeline API to gather the authors of the tweets incorporated in the timeline. The crawler can repeatedly call the timeline API as the Twitter

timeline regularly updates. at same time our proposed approach will spot the illegal user account properties. A general approach shared by bots is following a large number of users.

Classifying Twitter:

To build up an automatic classification system, we need a ground-truth set that contains known samples of human, bot, and cyborg. That we are doing by arbitrarily choose different samples and categorize them by manually examining their user logs and homepages. In classification, the data set contains tweets posted by the users in their account life span, which we canextract useful features for classification, such as tweeting behaviors and text patterns.

Detecting Account:

In this module, we will describe the distinction among human, bot, and cyborg in terms of tweeting behavior, tweetcontent, and account properties. The data analysis is carried out on the tweets posted by the users. This will authenticate the tweet content and their account details. All details are captured in the dataset to validate the account properties for the user.

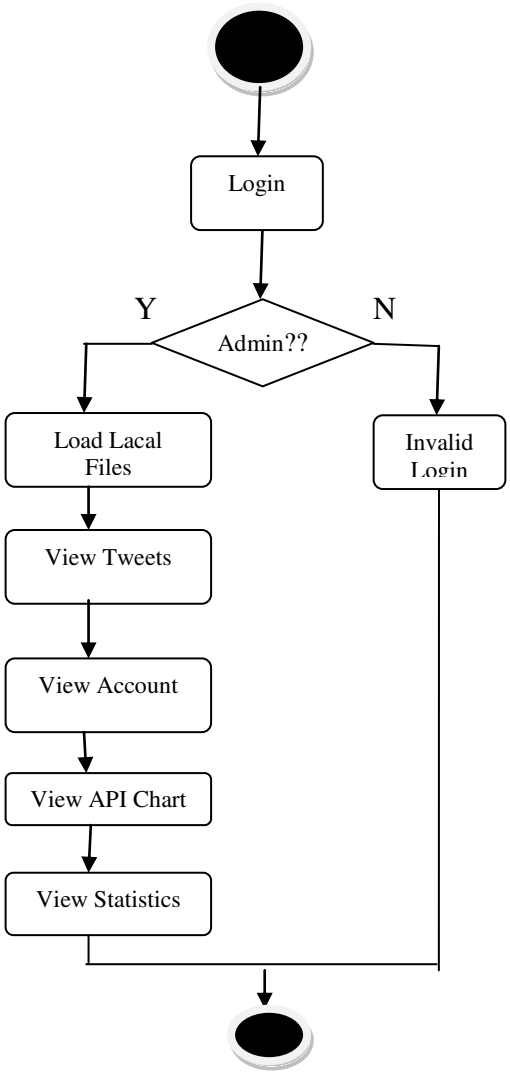

Fig no 2 General Flow

The above Fig no 2 shows the flow of the system.first the admin will login he/she load files from the dataset.He can view the tweets posted by the users. Administrator can select a particular twitter account then he can view the tweets from that twitter account and the type of application(like mobile or system etc) they used to do that particular tweet. Administrator can view the tweet count and reputation count. Admin view the API chart and Statistics.

IV MATHEMATICAL MODEL

The Bayesian classifier uses the content of chat messages to identify chat bots. Since chat messages (including emoticons) are text, the identification of bots can be perfectly fitted into the domain of Bayesian text classification. Within the Bayesian paradigm, the text classification problem can be formalized as $f : T \times C \rightarrow \{0,1\}$, where f is the classifier, $T = \{t_1, t_2, \ldots\ldots t_n\}$ is the texts to be classified, and $C = \{c_1, c_2, \ldots\ldots c_k\}$ is the set of predefined classes . Value 1 for indicates that text is in class , and value 0 indicates the opposite decision. There are many techniques that can be used for text classification, such as naïve Bayes, support vector machines, and decision trees. Among them, Bayesian classifiers have been very successful in text classification, particularly in e-mail spam detection. we choose Bayesian classification for our text classifier for detecting bots.. If the probability is equal to or greater than a predefined threshold, then message is classified as a bot message. According to Bayes theorem ,

$$P(bot \mid M) = \frac{P(M \mid bot)P(bot)}{P(M)} = \frac{P(M \mid bot)P(bot)}{P(M \mid bot)P(bot) + P(M \mid human)P(human)}$$

A message M is described by its feature vector<f1,f2…,fn> .A feature f is a single word or a combination of multiple wordsin the message. To simplify computation, in practice it is usually assumed that all features are conditionally independent with each other for the given category. Thus, we have

$$P(bot \mid M) = \frac{P(bot)\prod_{x=1}^{n} p(f_x \mid bot)}{P(bot)\prod_{x=1}^{n} p(f_x \mid bot) + P(human)\prod_{x=1}^{n} P(f_x \mid human)}$$

V EXPERIMENTAL RESULTS

Administrator providing credentials in order to access the application. Home page after successful login. click on load from file to read the twitter accounts from local machine info.all the twitter account details i.e friends details are stored into friends folder and their corresponding tweets are stored into tweets folder. Administrator can select a particular twitter account then he can view the tweets from that twitter account and the type of application(like mobile or system etc) they used to do that particular tweet.

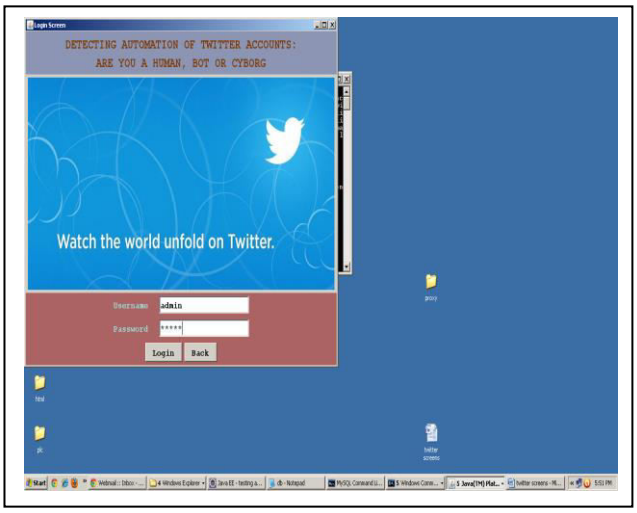

Fig no 3 Main screen

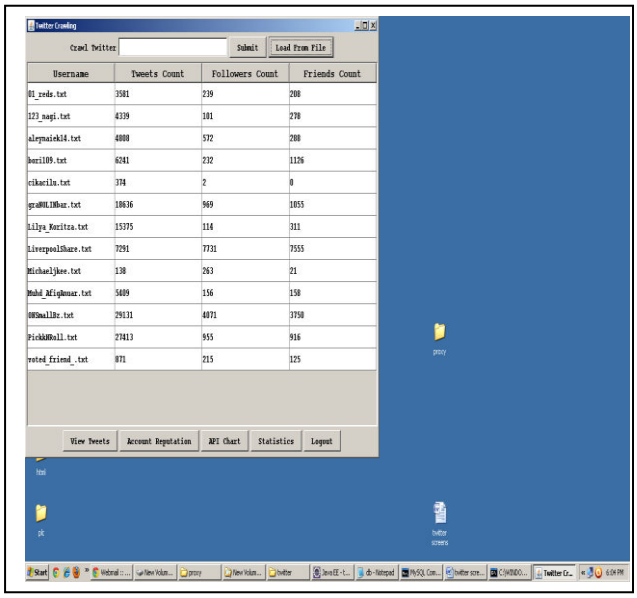

Fig no 4 Load Files

CONCLUSION

Here we have proposed a system which will help us classify human,bot or cyborg. We collected dataset and based on this data recognized features that can differentiate humans, bots, and cyborgs on twitter. It has been observed that human have complex performance i.e. high entropy, while bots and cyborgs are often given away by their standard or periodic timing, i.e., low entropy. While examining the text of tweets, it has been encountered that a large amount of bot tweets contains spam content. At last we found that certain account properties, like tweeting device makeup are also helpful in detecting automation.

REFERRENCES

[1] Zi Chu, Steven Gianvecchio, Haining Wang, and SushilJajodia, "Detecting Automation of Twitter Accounts:Are You a Human, Bot, or Cyborg?",*IEEE Transaction on Dependable and Secure Computing*, Vol. 9, 2012.

[2] S. Gianvecchio, M. Xie, Z. Wu, and H. Wang, "Measurement and Classification of Humans and Bots in Internet Chat," Proc 17th USENIX Security Symp, 2008.

[3] S. Yardi, D. Romero, G. Schoenebeck, and D. Boyd, "Detecting Spam in a Twitter Network," First Monday, vol. 15, no. 1, Jan. 2010.

[4] Y. Jeff and A. Ahmad Salah El, "Usability of CAPTCHAs or usability issues in CAPTCHA design," in Proceedings of the 4th symposium on Usable privacy and security Pittsburgh, Pennsylvania: ACM, 2008.

[5] Jaber Karimpour, Ali A Noroozi and Somayeh Alizadeh. Article: Web Spam Detection by Learning from Small Labeled Samples. International Journal of Computer Applications 50(21):1-5, July 2012. Published by Foundation of Computer Science, New York, USA.

[6] Cailing Dong, Bin Zhou, "Effectively Detecting Content Spam on the Web Using Topical Diversity Measures," wi-iat, vol. 1, pp.266-273, 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, 2012.

[7] Mohammed Al-Kabi , Heider Wahsheh, Izzat Alsmadi , Emad Al-Shawakfa, Abdullah Wahbeh , Ahmed Al-Hmoud. "Content-based analysis to detect Arabic web spam"Article: Journal of Information Science June 2012 vol. 38 no. 3 284-296.

[8] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5-32, 2001.

[9] T.K. Ho, "The Random Subspace Method for Constructing Decision Forests," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 8, pp. 832-844, Aug. 1998.

[10] J. Yan, "Bot, Cyborg and Automated Turing Test," Proc. 14th Int'l Workshop Security Protocols, Mar. 2006.