ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?

Rupali Vishnu Mutkule, S.G.Salve

Department of CSE, MGM's College of Engineering, Swami Ramanand Teerth Marathwada University, Nanded, Maharashtra, India

ABSTRACT: Twitter is the web application assuming double parts of online long range informal communication and micro blogging. Clients speak with each other by distributing content based posts. The popularity and open structure of Twitter have attracted in an expansive number of computerized programs, known as bots, which give off an impression of being a twofold edged sword to twitter. Honest to goodness bots create a lot of generous tweets conveying news and refreshing nourishes, while malicious bots spread spam or malignant substance. All the more curiously, in the center amongst human and bot, there has developed cyborg classified to either bot-helped human or human-helped bot. To help human clients in recognizing their identity communicating with, this paper concentrates on the order of human, bot, and cyborg accounts onTwitter. We initially lead an arrangement of substantial scale estimations with a gathering of more than 500,000 records. We watch the distinction among human, bot, and cyborg regarding tweeting behavior, tweet content, and account properties. In view of the estimation comes about, we propose an arrangement framework that incorporates the accompanying four sections: 1) an entropy-based component, 2) a spam detection component, 3) an account properties component, and 4) a decision maker. It utilizes the mix of components removed from an obscure client to decide the probability of being a human, bot, or cyborg. Our trial assessment exhibits the viability of the proposed arrangement framework. Contribution work is Naïve Bayes algorithm for classifying tweets are spam and non-spam and also identifying twitter account may in given category by using all features of twitter account is gives best performance.

I. INTRODUCTION

The growing user population and open nature of Twitter have made itself an ideal target of exploitation from automated programs, known as bots. Like existing bots another web applications (i.e., Internet chat, blogs and online games), bots have been common on Twitter. Twitter does not inspect strictly on automation. It only requires the recognition of a CAPTCHA image during registration. After gaining the login information, a bot can perform most human tasks by calling Twitter APIs. More interestingly, in the middle between humans and bots have emerged cyborgs, which refer to either bot-assisted humans or human-assisted bots. Cyborgs have become common on Twitter. After a human registers an account, he may set automated programs (i.e., RSS feed/blog widgets) to post tweets during his absence. From time to time, he participates to tweet and interact with friends. Different from bots which greatly use automation, cyborgs interweave characteristics of both manual and automated behavior.Automation is a double-edged sword to Twitter. On one hand, legitimate bots generate a large volume of benign tweets, like news and blog updates. This complies with the Twitter's goal of becoming a news and information network. On the other hand, malicious bots have been greatly exploited by spammers to spread spam. The definition of spam in this paper is spreading malicious, phishing, or unsolicited commercial content in tweets. These bots randomly add users as their friends, expecting a few users to follow back. In this way, spam tweets posted by bots display on users' homepages. Enticed by the appealing text content, some users may click on links and get redirected to spam or malicious sites.2 If human users are surrounded by malicious bots and spam tweets, their twittering experience deteriorates, and eventually the whole Twitter community will be hurt.In the paper, we first conduct a series of measurements to characterize the differences among human, bot, and cyborg in terms of tweeting behavior, tweet content, and account properties. By crawling Twitter, we collect over500,000 users and more than 40 million tweets posted by them. Then, we perform a detailed data analysis, and find asset of useful features to classify users into the three classes.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

Twitter in the most important online service for altering the information or messages (tweet), in the world 140 million user created account on twitter and most important thing is that 340 million messages delivered regularly on twitter. The URL shortening services which provide a short alias of a long URL it is useful service for Twitter users who want to share long URLs via tweets (140-character tweets containing only texts). The famous URL shortening services like bit.ly and goo.gl also provide shortened URLs' public click analytics consisting of the number of clicks and referrers of visitors. URL shortening services provide a combined form to protect the privacy of visitors from attacker. Example: Alice, updates her messages using the official Twitter client application for iPhone, "Twitter for iPhone" will be included in the source field of the corresponding metadata. Moreover, Alice may disclose on her profile page that she lives in the USA or activate the location service of a Twitter client application to automatically fill the location field in the metadata. Using this information, we can determine that Alice is an iPhone user who lives in the USA. The simple inference attack that can estimate individual visitors using public metadata provided by Twitter. The main advantage of the preceding inference attack over the browser history stealing attacks is that it only demands public information. In this paper, we propose novel attack methods for inferring whether a specific user clicked on certain shortened URLs on Twitter.

II. LITERATURE SURVEY

Project Name	Author Name	Proposed System	This Paper We Refer to
1) ""You might also like:" Privacy risks of collaborative filtering,"	A. Calandrino, A. Kilzer, A. Narayanan, E. W. Felten, and V. Shmatikov,	In this paper we develop algorithms which take a moderate amount of auxiliary information about a customer and infer this customer's transactions from temporal changes in the public outputs of a recommender system. Our inference attacks are passive and can be carried out by any Internet user.	Idea about Privacy risks of collaborative filtering.
2) "Timing attacks on web privacy,"	E. W. Felten and M. A. Schneider,	This paper presents a novel timing attack method to sniff users' browsing histories without executing any scripts. Our method is based on the fact that when a resource is loaded from the local cache, its rendering process should begin earlier than when it is loaded from a remote website. We leverage some Cascading Style Sheets (CSS) features to indirectly monitor the rendering of the target resource.	The evaluation shows that our method can effectively sniff users' browsing histories with very high precision. We believe that modern browsers protected by script-blocking techniques are still likely to suffer serious privacy leakage threats.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

3) "Tweet, tweet, retweet: Conversational aspects of rewetting on twitter,"	D. Boyd, S. Golder, and G. Lotan,	In the proposed system weexamines the practice of retweeting as a way by which participants can be "in a conversation." While retweeting has become a convention inside Twitter, participants retweet using different styles and for diverse reasons. We highlight how authorship, attribution, and communicative fidelity are negotiated in diverse ways.	We highlight how authorship, attribution, and communicative fidelity are negotiated in diverse ways. Using a series of case studies and empirical data, this paper maps out retweeting as a conversational practice.
4) "I Know the Shortened URLs You Clicked on Twitter: Inference Attack using Public Click Analytics and Twitter Metadata,"	Jonghyuk Song, Sangho Lee, Jong Kim	Only use public information provided by URL shortening services and Twitter; i.e., click analytics and Twitter metadata. We determine whether a target user visits a shortened URL by correlating the publicly available information. Our approach does not need complicated techniques or assumptions such as script injection, phishing, malware intrusion or DNS monitoring. All we need is publicly available information.	practical attack technique that can infer who clicks what shortened URLs on Twitter.
5) **"**Inferring Privacy Information From Social Networks ?**"**	Jianming He1, Wesley W. Chu1, and Zhenyu (Victor) Liu2	take both social network struc-tures and in°uence strength of social relations into consideration.	Investigated the problem of privacy inference in social net-works. Using Bayesian networks
6) **"**Scriptless Timing Attacks onWeb Browser Privacy,**"**	Bin Liang, Wei You, Liangkun Liu, Wenchang Shi	To perform an elaborated investigation to reveal additional exploitable browser mechanisms. With more dynamic and interactive features introduced in browsers in present times	Presented a new timing attack method for sniffing users' browsing histories

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

7) "Protecting Browser State from Web Privacy Attacks"	Collin Jackson, Dan Boneh, Andrew Bortz, John C Mitchell	Propose that a general same-origin principle should be applied uniformly across di_erent types of information stored on a web user's machine. We also develop ways for users to limit track-ing, in the form of browser extensions that are available for download.	presents some more powerful tracking methods based on caching various kinds of _les.
8) "Protecting Browsers from Cross-Origin CSS Attacks,"	Lin-Shung Huang, Chris Evans, Zack Weinberg, Collin Jackson	stricter content handling rules that completely block the attack, as long as the targeted web site does not make certain errors	present a general form of this attack that can be made to work in any browser that supports CSS, even if JavaScript is disabled or unsupported.
9) "Web Browser History Detection as a Real-World Privacy Threat"	Artur Janc1 and Lukasz Olejnik2	the pioneering the data acquisition of history-based user preferences	analyze the impact of CSS-based history detection and demonstrate the feasibility of conducting practical attacks with minimal resources
10) "A Topic-focused Trust Model for Twitter"	Liang Zhao	Experiments on Twitter event detection demonstrated that our method can effectively extract trustworthy tweets while excluding rumors and noise. In addition, a comparative performance analysis demonstrated that our method outperforms existing supervised learning schemes using tweets manually labelled or tweets generated based on keyword matching as the training set.	Utilizing credible news reports to infer trustworthiness of tweets exhibiting contextual similarity in textual, spatial and temporal features

III. EXISTING SYSTEM

Twitter has been widely used since 2006. To better understand micro blogging usage and communities, studied over 70,000 Twitter users and categorized their posts into four main groups: daily chatter, conversations, sharing information or URLs, and reporting news. Their work also studied 1) the growth of Twitter, showing a linear growth

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

rate; 2) its network properties, showing the evidence that the networks scale-free like other social networks; and 3) the geographical distribution of its users, showing that most Twitter users are from the US, Europe, and Japan. A group of over 100,000Twitter users and classified their roles by follower-to-following ratios into three groups: 1) broadcasters, which have a large number of followers; 2) acquaintances, which have about the same number on either followers or following; and 3) miscreants and evangelists (e.g., spammers),which follow a large number of other users but have few followers. The information diffusion on Twitter, regarding the production, flow, and consumption of information. The quantitative study on Twitter by crawling the entire Twitter sphere. Their work analyzed the follower-following topology, and found manpower-law followerDistribution and low reciprocity, which all mark a deviation from known characteristics of human social networks. Twitter lists as a potential source for discovering latent characters and interests of users. Atwitter list consists of multiple users and their tweets. Their research indicated that words extracted from each list are representative of all the members in the list even if the words are not used by the members. It is useful for targeting users with specific interests. The behaviors of spammers on Twitter by analyzing the tweets originated from suspended users in retrospect. They found that the current marketplace for Twitter spam uses a diverse set of spamming techniques, including a variety of strategies for creating Twitter accounts, generating spam URLs, and distributing spam.

DISADVANTAGES OF EXISTING SYSTEM

1. The existing machine learning-based streaming system, twitter user accounts are not categorized inhuman, bot or cyborg.

EXISTING SYSTEM ARCHITECTURE

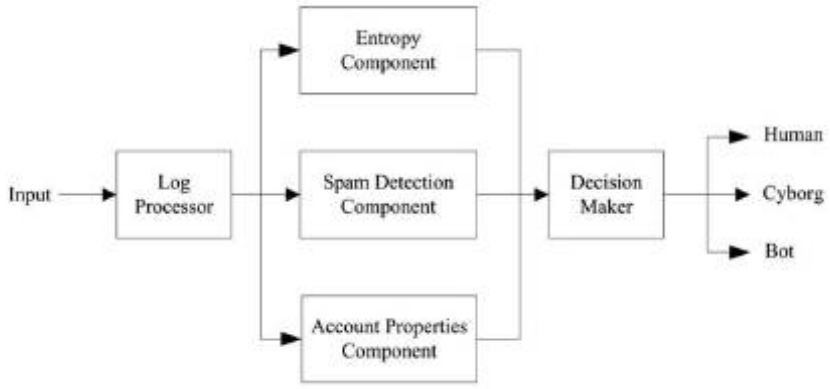

Fig No 01 Existing Approach

IV. PROPOSED SYSTEM

The proposed system using twitters data analysis, and finds a set of useful features to classify users into the three classes. Based on the measurement results, we propose an automated classification system that consists of four major components:

1. The entropy component uses tweeting interval as a measure of behavior complexity, and detects the periodic and regular timing that is an indicator of automation.
2. The spam detection component uses tweet content to check whether text patterns contain spam or non-spam.
3. The account properties component employs useful account properties, such as tweeting device makeup, URL ration, to detect deviations from normal.
4. The decision maker is based on Naïve Bayes, and it uses the combination of the features generated by the above three components to categorize an unknown user as human, bot, or cyborg.

To collected data, we randomly choose different samples and classify them by manually checking their user logs and homepages. The ground-truth set includes 2,000 users per class of human, bot, and cyborg, and thus in total three are

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

6,000 classified samples. The system classifies Twitter users into three categories: human, bot, and cyborg. The system consists of several components: the entropy component, the spam detection component, the account properties component, and the decision maker. The high-level design of our Twitter user classification system is shown in fig.2.

1. **Entropy Component:**
   The entropy component detects periodic or regular timing of the messages posted by a Twitter user. On one hand, if the entropy or corrected conditional entropy is low for theater tweet delays, it indicates periodic or regular behavior, a sign of automation. More specifically, some of the messages are posted via automation, i.e., the user may be potential bot or cyborg. On the other hand, a high-entropy indicates irregularity, a sign of human participation.
2. **Spam Detection Component:**
   The spam detection component examines the content of tweets to detect spam. We have observed that most spam tweets are generated by bots and only very few of them are manually posted by humans. Thus, the presence of spam patterns usually indicates automation. Since tweets are text, determining if their content is spam can be reduced to a text classification problem.
3. **Account Properties Component:**
   Twitter account-related properties are very helpful for the user classification. The first property is the URL ratio. Thus, high ratio (e.g., close to 1) suggests a bot and a low ratio implies a human. The second property is tweeting devicemakeup.The third property is the followers to friends ratio. The fourth property is link safety, i.e., to decide whether external links in tweets are malicious/phishing URLs or not. We run a batch script to check a URL in five blacklists: Google Safe Browsing, Phishing Tank, URIBL, SURBL, and Spamhaus. If the URL appears in any of the blacklists, the feature of link safety is set as false. The fifth property is whether a Twitter account is verified. The sixth property is the account registration date. The last two properties are the hash tag ratio and mention ratio.Hashtag ratio of an account is defined as the number of hash tags included in the tweets over the number of tweets posted by the account. Mention ratio is defined similarly.
4. **Decision Maker:**
   We select Random Forests the machine learning algorithm, and implement the decision maker based on it. Random Forest creates an ensemble classifier consisting of a set of decision trees. We denote the number of features in the data set as M, and the number of features used to make the decision at a node of the tree as m(<<M). Each decision tree is built top-down in a recursive manner. Forever node in the construction path, m features is randomly selected to reach a decision at the node. The node is then associated with the feature that is the most informative. Entropy is used to calculate the information gain contributed by each of the m features (namely, how informative a feature is). In other words, the recursive algorithm applies a greedy search by selecting the candidate feature that maximizes the heuristic splitting criterion.

ADVANTAGES OF PROPOSED SYSTEM

1. Extraction of user account features and categories as Tag based features and URL based features.
2. The system implements a method which calculates maximum conditional entropy.
3. The system implements a method which will use spot filter mechanism to detect whether the post is spam or not.
4. The system implements application can also be hosted online for its use and the data will be stored and fetched from server.
5. User with maximum number of spam can be blocked from the system.
6. Performance evaluation done on Dataset by using TPR, FPR, Precision, Recall and F-measure.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

PROPOSED SYSTEM ARCHITECTURE

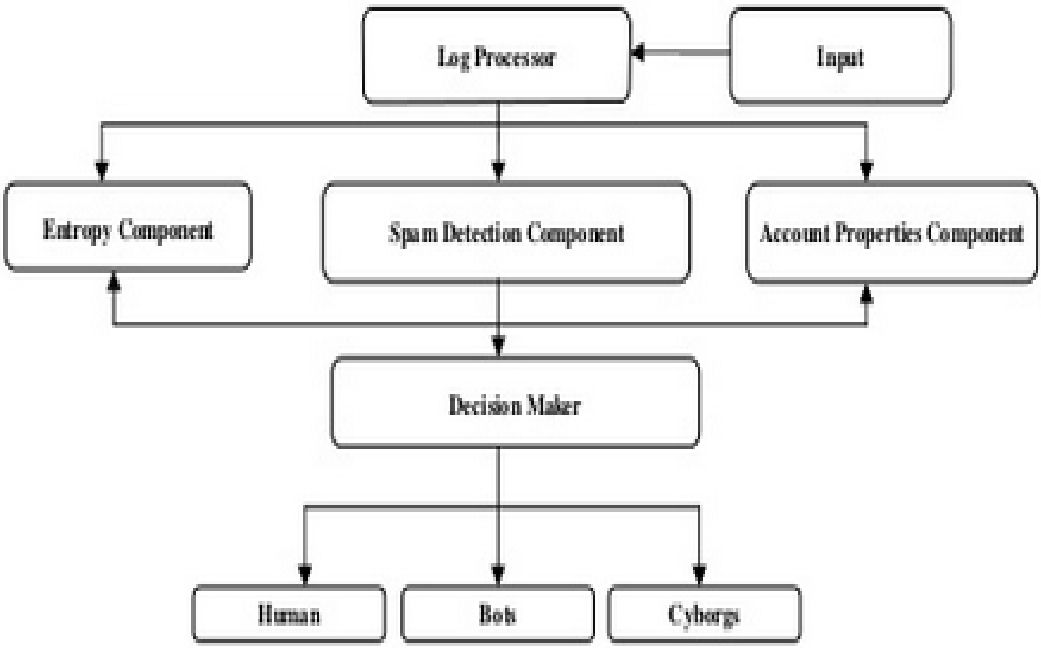

Fig.2 Block diagram of proposed system

V. MATHEMATICAL MODEL

The entropy rate is defined as either the average entropy per random variable for an infinite sequence or as the conditional entropy of an infinite sequence.
A random process $X = \{X_i\}$ is defined as a sequence of random variables. The entropy of such a sequence of random variables is defined as:

$$H(X_1, \dots, X_m) = -\sum_{i=1}^{m} P(x_i) log\ P(x_i) \qquad (1)$$

The corrected conditional entropy, denoted as CCE, is computed as

$$CCE(X_m|X_1, \dots, X_{m-1}) = CE(X_m|X_1, \dots, X_{m-1}) + perc(X_m).EN(X_1) \quad (2)$$

The spam detection component examines the content of tweets to detect spam.
The probability that a message M is spam, P(spam|M), is computed from Bayes theorem:

$$P(spam|M) = \{p(spam) \prod_{i=1}^{n} P(f_i|spam)\} / \{p(spam) \prod_{i=1}^{n} P(f_i|spam) + P(not\ spam) \prod_{i=1}^{n} P(f_i|not\ spam)\} \quad (3)$$

Given an unknown user U represented by the feature vector, the decision maker determines the class C to which U belongs to. Namely,

$$U =< f1, f2, \dots, fn > \rightarrow C = \{human, bot, cyborg\} (4)$$

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

VI. PERFORMANCE EVALUATION

Complexity:

Before defining the actual term complexity, let us discuss about few real life scenarios. Take an example of railway reservation counter. People go there to book their tickets. The time to book tickets will depend on how many service windows are available, queue size and time taken by each representative. A person will either go in a lane which is very short or will stand in queue where representative is very fast. Take another example of sitting plan of 40 students. The first thing to be considered is the no. of chairs needed and second thing is their order of sitting. If they sit roll.no wise then it will be easy for teacher to take attendance and time required will be lesser. Thus every process in the real world depends on how much **time** it takes to execute and how much space it consumes. **Complexity** is defined as the **running time** to execute a **process** and it depends on **space** as well as **time.**

It is of two types:

- Space Complexity
- Time Complexity

Basic Operation

Basic operation means the main operation that will be **required** to solve particular problem. For example the basic operation in searching is comparison. The complexity depends on the basic operation.

The focus to determine the **cost** is done on running time i.e **time complexity** and it depends on the following factors

- Size of Input Data
- Hardware
- Operating System
- Programming Language used

Space Complexity

The amount of computer **memory** required to solve the given problem of particular **size** is called as **space** complexity. The space complexity depends on two components

- **Fixed Part** – It is needed for instruction space i.e byte code. Variablespace, constants space etc.
- **Variable Part** – Instance of input and output data.
- **Space(S)** = **Fixed Part + Variable Part**

Time Complexity

The **time** required to **analyze** the given problem of particular **size** is known as the time complexity. It depends on two components

- **Fixed Part** – Compile time
- **Variable Part** – Run time dependent on problem instance. Run time is considered usually and compile time is ignored.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

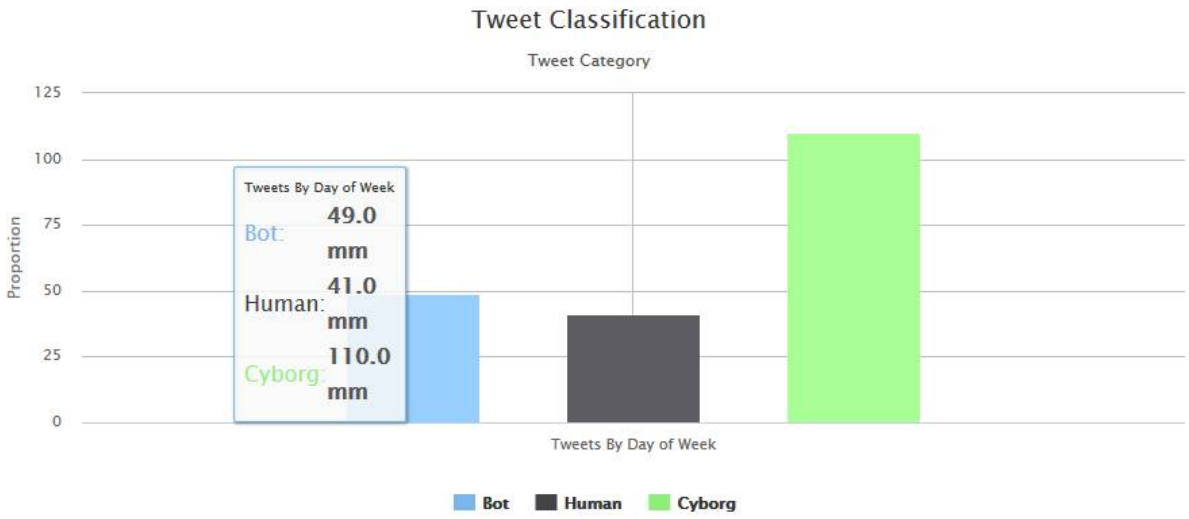

Fig3. Tweets Posted

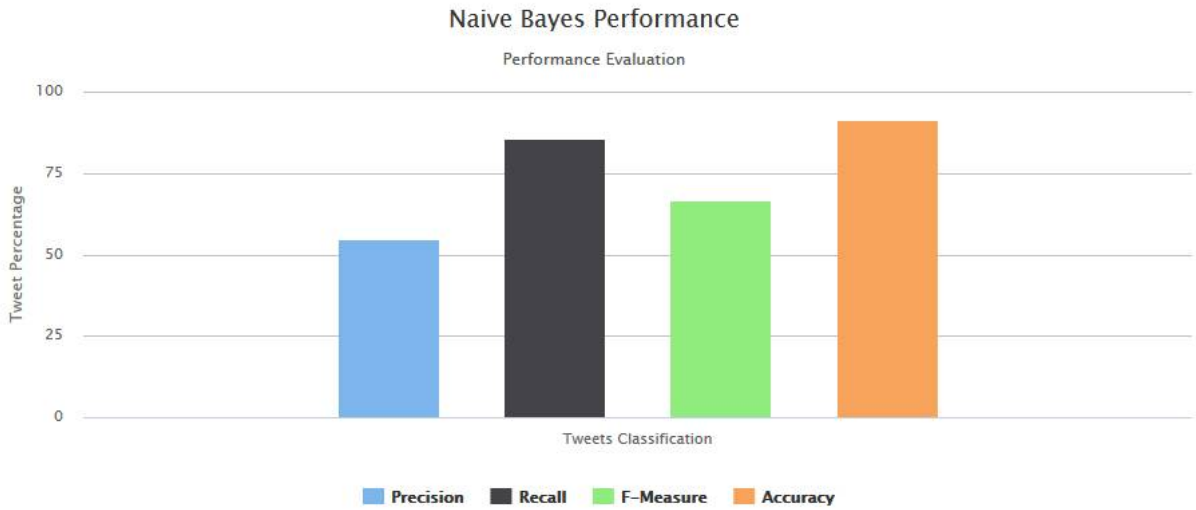

Fig.4. Performance of Tweets Categorization

VII. CONCLUSION

In this paper, we have studied the problem of automation bybots and cyborgs on Twitter. As a popular web application, Twitter has become a unique platform for information sharing with a large user base. However, its popularity and very open nature have made twitter a very tempting target for exploitation by automated programs, i.e., bots. The problem of bots on Twitter is further complicated by the key role that automation plays in everyday Twitter usage. To better understand the role of automation on Twitter, we have measured and characterized the behaviors of humans, bots, and cyborgs on Twitter. By crawling Twitter, we have collected one month of data with over 500,000 Twitter users with more than 40 million tweets. Based on the data, we have identified features that can differentiate humans, bots, and cyborgs on Twitter. Using entropy measures, we have determined that humans have complex timing behavior, i.e., high entropy, whereas bots and cyborgs are often given away by their regular or periodic timing, i.e., low entropy. In examining the text of tweets, we have observed that a high proportion of bot tweets contain spam content. Lastly, we

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

have discovered that certain account properties, like external URL ratio and tweeting device makeup, are very helpful on detecting automation. Based on our measurements and characterization, we have designed an automated classification system that consists of four main parts: the entropy component, the spam detection component, the account properties component, and the decision maker. The entropy component checks for periodic or regular tweet timing patterns; the spam detection component checks for spam content; and the account properties component checks for abnormal values of Twitter-account-related properties. The decision maker summarizes the identified features and decides whether the user is a human, bot, or cyborg.

REFERENCES

[1] "Top Trending Twitter Topics for 2011 from What the Trend," http://blog.hootsuite.com/top-twitter-trends-2011/, Dec. 2011.
[2] "Twitter Blog: Your World, More Connected," http://blog. twitter.com/2011/08/your-world-more-connected.html, Aug. 2011.
[3] Alexa, "The Top 500 Sites on the Web by Alexa," http:// www.alexa.com/topsites, Dec. 2011.
[4] "Amazon Comes to Twitter," http://www.readwriteweb.com/archives/amazon_comes_to_twitter.php, Dec. 2009.
[5] "Best Buy Goes All Twitter Crazy with @Twelpforce," http:// twitter.com/in_social_media/status/2756927865, Dec. 2009.
[6] "Barack Obama Uses Twitter in 2008 Presidential Campaign," http://twitter.com/BarackObama/, Dec. 2009.
[7] J. Sutton, L. Palen, and I. Shlovski, "Back-Channels on the Front Lines: Emerging Use of Social Media in the 2007 Southern California Wildfires," Proc. Int'l ISCRAM Conf., May 2008.
[8] A.L. Hughes and L. Palen, "Twitter Adoption and Use in Mass Convergence and Emergency Events," Proc. Sixth Int'l ISCRAM Conf., May 2009.
[9] 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.
[10] B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna, "Your Botnet Is My Botnet: Analysis of a Botnet Takeover," Proc. 16th ACM Conf. Computer and Comm. Security, 2009.
[11] S. Gianvecchio, Z. Wu, M. Xie, and H. Wang, "Battle of Botcraft: Fighting Bots in Online Games with Human Observational Proofs," Proc. 16th ACM Conf. Computer and Comm. Security, 2009.
[12] A. Java, X. Song, T. Finin, and B. Tseng, "Why We Twitter: Understanding Microblogging Usage and Communities," Proc. Ninth WebKDD and First SNA-KDD Workshop Web Mining and Social Network Analysis, 2007.
[13] B. Krishnamurthy, P. Gill, and M. Arlitt, "A Few Chirps about Twitter," Proc.First Workshop Online Social Networks, 2008.
[14] S. Yardi, D. Romero, G. Schoenebeck, and D. Boyd, "Detecting Spam in a Twitter Network," First Monday, vol. 15, no. 1, Jan. 2010.
[15] A. Mislove, M. Marcon, K.P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," Proc. Seventh ACM SIGCOMM Conf. Internet Measurement, 2007.
[16] S. Wu, J.M. Hofman, W.A. Mason, and D.J. Watts, "Who Says What to Whom on Twitter," Proc. 20th Int'l Conf. World Wide Web, pp. 705-714, 2011.
[17] H. Kwak, C. Lee, H. Park, and S. Moon, "What Is Twitter, a Social Network or a News Media?"Proc. 19th Int'l Conf. World Wide Web, pp. 591-600, 2010.
[18] I.-C.M. Dongwoo Kim, Y. Jo, and A. Oh, "Analysis of Twitter Lists as a Potential Source for Discovering Latent Characteristics of Users," Proc. CHI Workshop Microblogging: What and How Can We Learn From It?, 2010.
[19] D. Zhao and M.B. Rosson, "How and Why People Twitter: The Role that Micro-Blogging Plays in Informal Communication at Work," Proc. ACM Int'l Conf. Supporting Group Work, 2009.