International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 07 | July 2018 www.irjet.net p-ISSN: 2395-0072

Identifying Spammers in Twitter Using Minimized Feature Set

T.Miranda Lakshmi[1], R.Josephine Sahana[2], V.Prasanna Venkatesan[3]

[1]Department of Computer Science, Research and Development Centre, Bharathiyar University, Coimbatore, India.
[2]M.Phil.Scholar, PG & Research Department of Computer Science, St. Joseph's College of Arts and Science, Cuddalore, Tamil nadu. India.
[3]Department of Banking Technology, Pondicherry University, Puducherry, India.

Abstract - With the rapid growth of information technology and networking, the people around the world are able to share lot of information on the internet. At present millions of users around the world are engaged themselves with online social networking sites. Twitter is one such platform where users can post and share their content called tweet. Spammers are fake users who send unwanted messages or perform fraudulent activities for malicious purpose. Preventing genuine users from these spam account is an important issue in twitter and also in other social network sites. For this purpose many algorithms and methods are proposed by various researchers. These fake identification methods are based on features related to user account and content. There are several attributes available to distinguish the legitimate users from fake users. These features are used to train the machine learning algorithms to detect and predict the fake identities. Training the algorithm with all feature set is much difficult and time consuming process. Instead of using all the features, this paper aims to minimizes the features and finds the machine learning algorithm that detect the spam accounts more accurately.

Keywords: OSN, Twitter, Fake Identities, Feature Engineering, Machine Learning.

I INTRODUCTION

In this digital era Online Social Networks (OSN) is the most preferable way to communicate platform among the people around the globe. Facebook, Twitter, LinkedIn, Instagram, GooglePlus are some of the most popular and trending platforms that instantly connect the people with their audience. Twitter is a micro blogging platform where users can post and share their own opinions and also view other people's view about any topics, events, products, social issues etc., as it is a micro blogging service it supports only 140 characters and it can be used through Pcs, Laptops and Smart Phones[1] . Its openness and ease of use attracts many users and also it draws the attention of spammers or fake accounts. These fake accounts can be generated either by human, computer or cyborgs. The computer generated fake accounts are known as "bots" where cyborg is a combination of human and bot i.e., it is created by human but further handled by bots. However the purpose of these fake identities is,

- To change the actions of an individual or group
- To change perceptions of an individual or group
- To hide the malicious activity of an individual or group
- To spread malware

These fake identities can be detected by many approaches and methods. Machine Learning is one of the methods used to detect false identities in social media platforms. The main objective of machine learning is making the computer automatically learn from the past experiences and predict the future more accurately[2]. Random Forest, Decision Tree, Neural Network, Support Vector Machine, Linear Regression are some of the algorithms that are mostly used to detect the fake identities. By using this machine learning algorithms the spam accounts can be predicted with more accurately and efficiently. The rest of the paper is structured as follows: Section II provides the various works related to detection of fake accounts in Online Social Networks. Section III and Section IV provides the basics of machine learning and methodology. Section V concludes the paper with future work.

II REVIEW OF LITERATURE

Various techniques were proposed by different researchers to detect the fake identities in OSN. The fake accounts can be identified either by the content posted by them or by analyzing the profile characteristics of the account. There are also some third party services that are used to detect and prevent from spam users. Each technique had its own merits and demerits [3]. This section focus on various techniques and works related to fake account detection in social media platforms. The following figure 1 depicts the various techniques available to detect spammers.

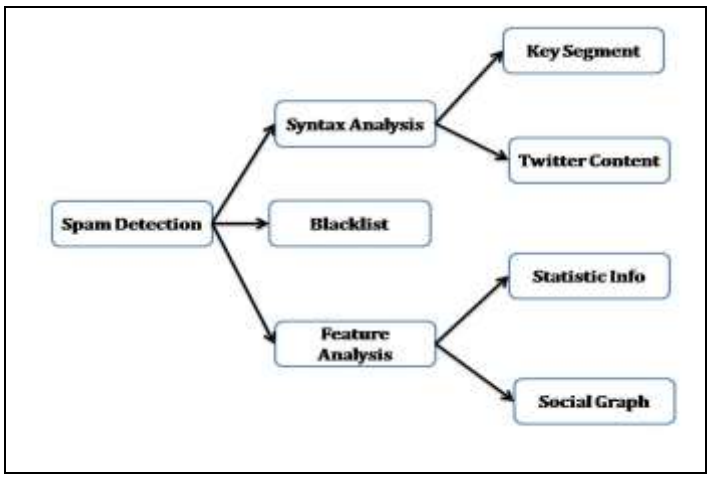

Figure 1: Spam Detection Techniques

Spam Detection:

Online Social Networks needs fast and new rising techniques to detect the spammers. The following section discusses the techniques that already available to detect the spammers.

Syntax Analysis:

Analyzing spam based on syntax is categorized into two types.

i) Key Segment and

ii) Twitter Content.

Key Segment based method relies on the truth that spammers generally make use of sensitive information to attract people. So analyzing the URL, Username and Keyword will detect the spammers successfully.

Twitter Content spammers usually post content with similar malicious words. These malicious words can be examined to identify the spammers. The techniques available to characterize the textual content of tweet are TF- IDF (Term Frequency – Inverse Domain Frequency), Bag of Words, Sparse Learning.

Blacklist:

It is a technique that applies third party services such as Google's Safe Browsing API. These blacklist technique blocks the malicious links before it reaches the receiver.

Feature Analysis:

In the literature most of the researchers use features based analyze method to detect spam accounts. It is categorized into two types.

i) Statistic Information:

It denotes statistic information related to user account and tweet content. These features are sufficient to differentiate legitimate users from malicious users. For example, spam tweet contains more hash tags, spam words; more digits and they broadcast more messages than legitimate users. In the same way features like number of followers, life time of the account etc., are useful to differentiate spam and non-spam accounts.

ii) Social Graph:

Social graph based method use features from social graphs of twitter according to their follower and following relationship. It is further divided into two types, i) Graph based method that focus on macroscopic attributes of graph nodes and ii) Neighborhood based method that focus on microscopic relationships of graph nodes.

The following Table 1 depicts the list of objects that are mostly used in the previous work to detect fake identities in Twitter platform[4]–[6].

Table 1: Features Used Frequently in Previous Work to Detect Fake Identities

User Account	User Content
• Id • Name • Screen Name • Location • Url • Description • Followers_Count • Friends_Count • Listed_Count • Favourites_Count • Statuses_Count • Created_At • Utc_Offset • Time_Zone • Profile_Image_Url	• Created_At • Id • Text • Source • Retweeted_Status • Reply_Count • Retweet_Count • Favorite_Count • Entities favorited

From this study it can be stated that feature based methods are used frequently to detect the fake identities. The following section gives attention to the related works and feature analysis method. When compared to the common twitter users, the spam users usually aim to belittlement of ideas. When we equate the genuine users with fake users they spread their profile with inconsistent information. Usually spam message contains lot of grammatical error, emotional content that affect the readers view and they often post fake content with limited set of words. [7].

There are many attributes that are available to describe the identity of an account in social media platform. For example, name, profile image, screen name, description etc., Using these feature set many researchers successfully detect fake identities in social media platforms [8]. These fake identities are created by humans, bots or cyborgs. To differentiate among these accounts series of experiments were conducted in terms of tweeting behavior, tweet content and account properties.

From the analysis it can be identified that bots post more duplicate and automatic tweets. So they are lack in intelligence. So designing an automated classification system can ease the process of separating humans from bots (Chu,

Gianvecchio and Wang, 2010). Sentiments also play a vital role in identification of fake accounts. Sentiment features based on syntax and semantics of tweet, behavior of user and user's neighborhood can also be used to detect fake accounts [10]. To increase the accuracy of detecting fake identities in social networks machine learning can be combined with natural language processing. Using this model fake identity can be identified with advanced features [11]. Most of the existing methods depend on user profile attributes. But in reality in many cases the attributes are unavailable to process because of privacy. So instead of analyze the user profile features analyzing the activities of same users in different OSNs based on the content they generated may produce better result in identifying fake accounts [12]. Instead of using classification for fake identitification clustering can also be used. They can be identified using similarity. Instead of making a prediction for each individual account clusters can be used to determine whether they have been created by the same actor [13]. Both classification and clustering showed same accuracy. But the advantage of using clustering is it does not required labeled data [14]. And also preprocessing the data set using supervised discretization technique namely entropy minimization discretization on numerical features can enhance the accuracy [15].

The Table 2 represents the previous works based on feature analysis related to user account and content. Almost all the authors use similar algorithm for classification of fake accounts. In which random forest, support vector machine and neural network algorithms produce the best accuracy.

Random Forest (RF) algorithm is best machine algorithm for a large number of datasets. It is a collection of decision tree. This method divides the given dataset into several subsets of same size then decision tree is trained for each subset. Finally all the trees are arranged in a descending order and select the average value to prediction [16]. Neural Network (NN) is a classification method that produces a less error rate than the decision tree. A neural network is usually a layered graph with the output of one node feeding into one or more nodes in the next layer. [17]. Support Vector Machine (SVM) finds the division that exploits the boundary between two data populations. This exploitation reduce the over fitting of the learning data. The main advantage is that it does not require huge memory. It provides more precise result on small and clean dataset [18]. From the study it can be stated that techniques and methods to detect the spammers in Online Social Networks is an emerging process. So it is essential to build a classifier model with advanced features. Our further work is based on building a classifier model that can identify the spammers with minimized feature set.

III. MACHINE LEARNING ALGORITHM

Machine learning is a subset of artificial intelligence that focus on designing of system thus it allows them to learn and make predictions based on the past experiences. It has three types of learning and number of algorithm associated with it. [2], [19]–[21] The Figure 2 illustrates the steps involved in machine learning process.

Table 2: Previous work related to detect fake identities in Twitter.

S.No	Author	Features	Comparison of Machine Learning Algorithms	Result
1	Al-janabi, Quincey and Andras, 2017	Tweet and user account based features	Compare Random Forest, Linear Regression, Naïve Bayes, K Nearest Neighbor	Random forest shows better accuracy with 92%
2	Alsaleh and Alarifi, 2014	Tweet based features	Compare Decision tree J48, random forest, SVM, NN	Neural network shows better accuracy with 95%
3	Azab *et al.*, 2016	Tweet and account	Random forest, Decision Tree, Naïve Bayes, Neural Network, Support Vector Machine	SVM shows better accuracy with 99.90 % with mimimum feature set
4	Kamoru *et al.*, 2017	User account	Mimimize the feature Support Vector Machine, Neural Network	SVM shows better accuracy with 84.04% with selected features

5	Gupta *et al.*, 2017	Tweet and account	Support Vector Machine, Neural Network, Gradient Boosting, Random Forest	NN shows better accuracy with 91%
7	An-garc and Omez, 2015	Tweet and account	Random Forest,J48,k-nearest neighbor, Sequential Minimal Optimization,	Sequential Minimal Optimization shows better accuracy with 68.47%
8	Walt and Eloff, 2018	User	Random Forest, Boosting, Support Vector Machine	RF shows better accuracy with 87.11 %

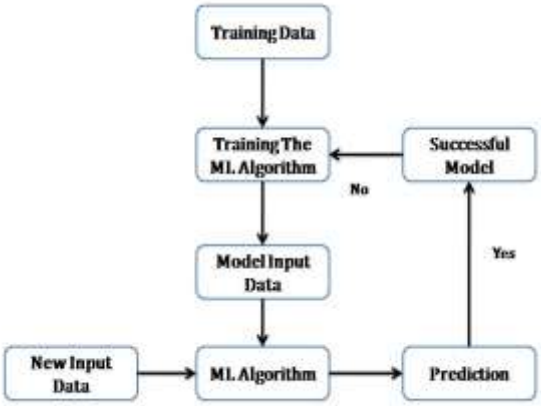

Figure 2: Steps in Machine Learning

First step in building the model is a collection of the dataset. After preprocessing, the data set should be divided as training and test data. Training data is used to train and build the classifier and test data is used to evaluate the performance of the classifier. The developed model is trained several times for better prediction of the data. After successful building of the classifier the test data can be given as an input to verify the precision of the classifier. By repeating the process the successful model can be build to better predict the data. In machine learning training can be done by three ways. The Figure 3 simply explains the types of learning in machine learning.

Figure 3: Machine Learning Training Types

This paper gives attention to supervised algorithm such as Random Forest, Neural Network, and Support Vector Machine.

IV. METHODOLOGY

Data set used in this paper is a Twitter user account details collected from the GitHub data repository which contains both real and fake identities. This research selects only user account features such as Screen Name, Email Id, Domain Name, Latitude, and Longitude and classification algorithms that include Random Forest, Support Vector Machine and Neural Network.

Dataset Used

The proposed method is implemented using Rapid Miner tool. It needs dataset with a mixture of real and false accounts labeled accordingly. The algorithms need to be trained using the training dataset and should be evaluated using the testing dataset. The data set can be obtained from publicly available data repositories. But availability of user account dataset is limited because of privacy issues. The details of the dataset and its features are given in Table 4.1 and 4.2 respectivley. The screenshots of the sample dataset is shown in Fig. 4.1-4.2 respectively.

Random Forest (RF):

RF is a classification method that constructs a set of decision tree at training phase. It basically constructs a set of decision trees at training phase and then each tree operates on randomly chosen attributes. First we are going to load the training data set into Random Forest to train the algorithm that produces a decision tree. Then using these decision tree spammers can be identified. The objective of using RF in this work is, it uses random sampling method for selection of features so it can produce the best result.

Neural Network (NN):

Classification using NN consists of three steps. First step is, Data preprocessing means selection of features. Second step is, Data training in which the selected features in previous step are fed into NN. Third step is testing which evaluate the efficiency of performance. The main advantage of using NN in this work is, it can be used to train the complex data as well.

Support Vector Machines (SVM):

The objective of SVM is to find the hyper plane that optimally separates with maximum margin the training data into two partitions. So in this work it is very much useful to separates the real users from the fake users.

These selected algorithms are trained by the selected features and the performance is evaluated by giving new input data. By comparing the results of the selected algorithms we could find the algorithm that predicts the fake identities accurately with minimized features.

IV. CONCLUSION

The main purpose of this study is validation of profile characteristics of user to detect fake identities in social media platforms. This paper also proposes a minimized feature set to detect fake accounts in twitter platform. This minimized feature set can give more precise result. The result will be evaluated by comparing the performance of Random Forest, Neural Network and Support Vector Machine algorithm. The future work is based on applying this fake detection method to other social media platforms.

REFERENCES

[1] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter , a Social Network or a News Media ? Categories and Subject Descriptors,” pp. 591–600, 2010.

[2] C. Burch, “A survey of machine learning,” 2001.

[3] T. Wu, S. Wen, Y. Xiang, and W. Zhou, “Twitter spam detection : Survey of new,” *Comput. Secur.*, 2017.

[4] V. S. Subrahmanian *et al.*, “The DARPA Twitter Bot Challenge,” 2016.

[5] F. Dqwhsh *et al.*, "3UHSURFHVVLQJ) UDPHZRUN IRU 7ZLWWHU % RW â€™ HWHFWLRQ," 2017.

[6] B. Y. E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The Rise of Social Bots," 2016.

[7] M. Granik and V. Mesyura, "Fake News Detection Using Naive Bayes Classifier," pp. 900–903, 2017.

[8] E. V. A. N. D. E. R. Walt and J. A. N. Eloff, "Using Machine Learning to Detect Fake Identities : Bots vs Humans," vol. 6, 2018.

[9] Z. Chu, S. Gianvecchio, and H. Wang, "Who is Tweeting on Twitter : Human , Bot , or Cyborg ?," pp. 21–30, 2010.

[10] J. P. Dickerson, V. Kagan, and V. S. Subrahmanian, "Using Sentiment to Detect Bots on Twitter : Are Humans more Opinionated than Bots ?," no. Asonam, pp. 620–627, 2014.

[11] S. R. Pulluri, J. Gyani, and N. Gugulothu, "A Comprehensive Model for Detecting Fake Profiles in Online Social Networks," pp. 385–394, 2017.

[12] Y. Li, Z. Zhang, Y. Peng, H. Yin, and Q. Xu, "Matching user accounts based on user generated content across social networks," *Futur. Gener. Comput. Syst.*, vol. 83, pp. 104–115, 2018.

[13] D. M. Freeman and T. Hwa, "Detecting Clusters of Fake Accounts in Online Social Networks Categories and Subject Descriptors," pp. 91–101, 2015.

[14] M. A. Fernandes, P. Patel, and T. Marwala, "Automated detection of human users in Twitter," *Procedia - Procedia Comput. Sci.*, vol. 53, pp. 224–231, 2015.

[15] B. Er, Ö. Akta, K. Õ. Õ. Deniz, and C. Akyol, "Twitter Fake Account Detection," pp. 2–6, 2017.

[16] M. Al-janabi, E. De Quincey, and P. Andras, "Using supervised machine learning algorithms to detect suspicious URLs in online social networks," pp. 1104–1111, 2017.

[17] B. A. Kamoru, A. Bin, J. Omar, and M. A. Jabar, "SPAM DETECTION ISSUES AND SPAM IDENTIFICATION," vol. 95, no. 21, pp. 5881–5895, 2017.

[18] S. Adikari and I. Systems, "Identifying fake profiles in linkedin," 2011.

[19] S. Kaur and S. Jindal, "A Survey on Machine Learning Algorithms," vol. 3, no. 11, pp. 6–14, 2016.

[20] K. Das and R. N. Behera, "A Survey on Machine Learning : Concept ," pp. 1301–1309, 2017.

[21] S. Vljqlilfdqw, S. Iru, D. O. O. Frpsdqlhv, and R. U. Rujdql, "$ 5hylhz rq &\ehu 6hfxulw\ 'dwdvhwv iru 0dfklqh /hduqlqj $ojrulwkpv," pp. 2186–2193, 2017.

[22] D. Radovanović and B. Krstajić, "Review Spam Detection using Machine Learning," 2018.

[23] M. Alsaleh and A. Alarifi, "TSD : Detecting Sybil Accounts in Twitter," 2014.

[24] A. El Azab, A. M. Idrees, M. A. Mahmoud, and H. Hefny, "Fake Account Detection in Twitter Based on Minimum Weighted Feature set," vol. 10, no. 1, pp. 13–18, 2016.

[25] H. Gupta, M. S. Jamal, S. Madisetty, and M. S. Desarkar, "A Framework for Real-Time Spam Detection in Twitter," pp. 380–383, 2017.

[26] P. G. A. L. An-garc and C. L. G. Omez, "Supervised machine learning for the detection of troll profiles in twitter social network : application to a real case of cyberbullying," vol. 24, no. 1, 2015.

Table 4.1 Dataset Description

Data set	Source	#of instances	#of attributes	# of class	Bankrupt/Non-Bankrupt
User profile	Git Hub	1525	16	2	1025/500

Table 4.2 Dataset Features

screen_name	user_tweeted	user_retweeted	user_favourited
user_replied	likes_per_tweet	retweets_per_tweet	lists_per_user
follower_friend_ratio	tweet_frequency	favourite_tweet_ratio	age_of_account_in_days
sources_count	urls_count	cdn_content_in_kb	source_identity

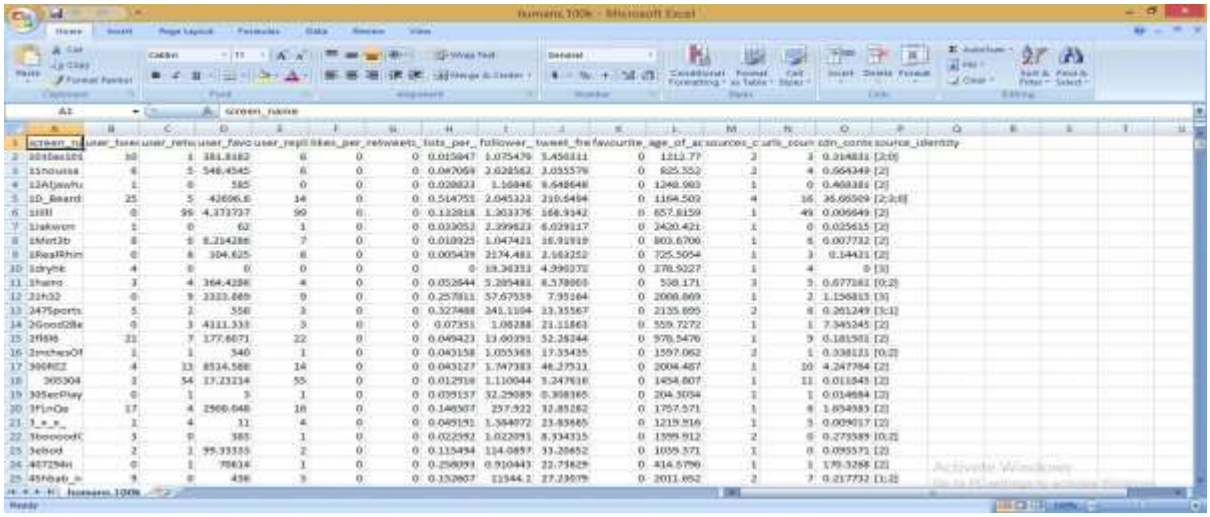

Fig. 4.1. Sample dataset –Part 1

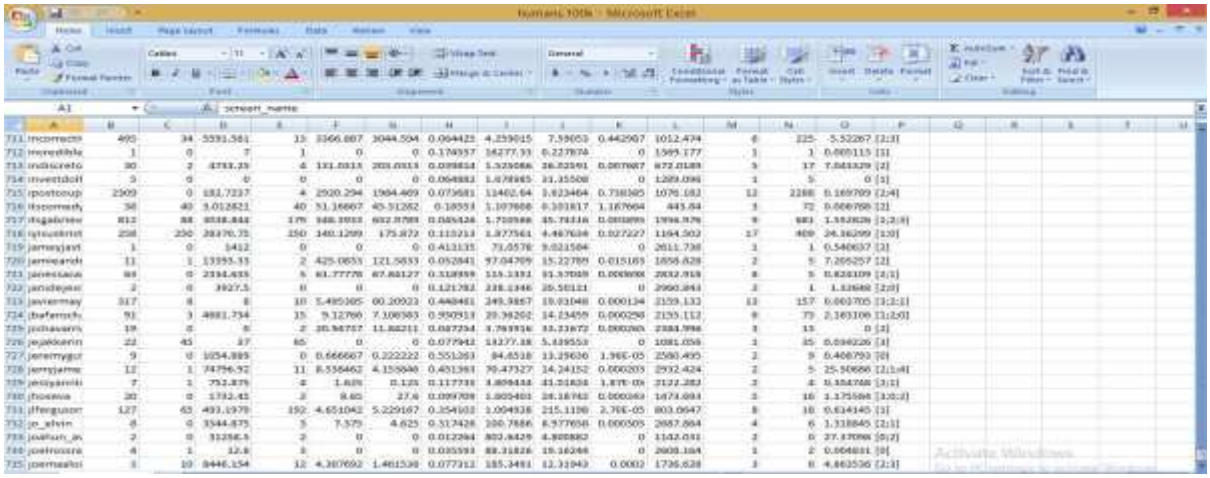

Fig. 4.2. Sample dataset –Part 2