Detection of Fake Profiles on Twitter Using Hybrid SVM Algorithm

Sarangam Kodati[1*], *Kumbala* Pradeep Reddy[2], *Sreenivas* Mekala[3], PL Srinivasa Murthy[4], P Chandra Sekhar Reddy[5]

[1]Associate Professor, Department of CSE, Teegala Krishna Reddy Engineering College, Hyderabad, Telangana.
[2]Associate Professor, Department of CSE, CMR Institute of Technology, Hyderabad, Telangana.
[3]Associate Professor, Department of IT, Sreenidhi Institute of Science and Technology, Hyderabad,Telangana.
[4]Professor CSE Department, Institute of Aeronautical Engineering, Hyderabad. Telangana
[5]Professor CSE Department, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad.

Abstract: Establishing and management of social relationships among huge amount of users has been provided by the emerging communication medium called online social networks (OSNs). The attackers have attracted because of the rapid increasing of OSNs and the large amount of its subscriber's personal data. Then they pretend to spread malicious activities, share false news and even stolen personal data. Twitter is one of the biggest networking platforms of micro blogging social networks in which daily more than half a billion tweets are posted most of that are malware activities. Analyze, who are encouraging threats in social networks is need to classify the social networks profiles of the users. Traditionally, there are different classification methods for detecting the fake profiles on the social networks that needed to improve their accuracy rate of classification. Thus machine learning algorithms are focused in this paper. Therefore detection of fake profiles on twitter using hybrid Support Vector Machine (SVM) algorithm is proposed in this paper. The machine learning based hybrid SVM algorithm is used in this for classification of fake and genuine profiles of Twitter accounts and applied the dimension reduction techniques, feature selection and bots. Less number of features is used in the proposed hybrid SVM algorithm and 98% of the accounts are correctly classified with proposed algorithm.

1. Introduction

The growing popularity of social media platforms has not only benefitted the people but also caught the attention of scammers. On one hand social media is bringing people together and on the other hand it has created a guarded space for fraudsters to carry out number of illegal activities. The absence of any authentication process has made it easy for anyone to make a fake account. This serves as an advantage for the scammers encouraging them to use fake account for illegal activities as there is a good chance that the account holder will not get caught. Owing to this the popularity of fake accounts has increased. These phone accounts can either operated by humans or bots. The use of these phantom accounts to impersonate someone in hope of defaming them has become a common issue. At times these accounts serve the bigger purpose of acting as a trusted acquaintance to get personal information from a person. This obtained information can be used to carry out phishing attacks [2].

People often use these dummy accounts to spread fake news which in the worst case can cause riot like conditions. Some people make use of fake accounts to spread hate which can be directed at certain race, religion, country or often at a particular person [3]. This has increased the cases of cyber bullying leading to rise in the cases of depression and anxiety in teenagers. The social media platforms have also seen an increase in the number of accounts which provide services or products in exchange of money[19]. But most of these accounts are fake as a result thousands of people are sold fake products and are promised fake services by these accounts. Sometimes these fake accounts are used by companies to build hype for their bad products and services [4]. Not only scammers but also a lot of influencers also use fake bot followers to appear popular, which help them in gaining more offers from companies asking to publicise their products. At times the fake accounts

* Corresponding author: k.sarangam@gmail.com

can also be used to befriend a person in order to stalk them. Another big issue associated with the fake accounts is the amount of data overload that they are resulting in [5]. With the number of fake accounts being in millions it has become impossible to manually detect them. Luckily the advancement in digital technology can benefit a lot in this situation. Methods like Machine Learning can help in making the stratification process a lot easier and accurate [6]. This project involves use of machine learning model to classify social media accounts as genuine or fake. Spam URLs and spam tweets sending strategy use the attack strategy of social engineering by spammers. Irregular spam accounts proliferation uses an ideal arena of twitter. From defamatory actions a model is developed by researchers from the simulation impacts and this method detects and recovered the fault profiles. Number of fake spam profiles is present in the twitter network which causes the issues in providing security and privacy to normal users. In this research one of the key parts is spam profiles identification which improves the safety of real users.

2. Role of Machine Learning in Detection of Fake profiles

Since last twenty years, there is an enormous improvements are observed in OCIAL networking phenomenon. So number of social networks is introduced different online services which are attracts huge amount of users. The increasing capacity of users is depending on information credibility on Online Social Networks (OSNs) [7]. Online social networks are being a part of every one social life in present generation. Technology usage is widely increased in nowadays. Online social networks are playing an important role in modern society. Social networks are dealing millions of users in present days all over the world. Facebook and twitter are two social networks in which the user interactions are more and daily life can be highly impacted with these social networks[16]. Large amount of fake account creation is the major problem of OSN networks. These fake accounts are does not match with real profiles of humans. Spam, web rating and fake news are representing some fakes [8]. The detection of different resources is currently expended by OSN operators and then fake accounts are closed. Almost 46% of users are operating the twitter account on the mobile phones only [9]. SMS text messages sending and e-mails sending are publishes the tweets. Messages capacity of twitter is 140 characters of message which is used for exchanging and publishes on twitter directly from smart phones using a wide array of Web-based services [10]. Number of users is maintained by the twitter. Better social lives are maintained with these social sites but also there are some disadvantages or issues are existed with these social networks. Online harassment, privacy, trolling, potential for misuse, fake account creation and etc are some of the social networks issues [11]. We will implement machine learning algorithms to predict if an account is controlled by fake user. Unsupervised Learning and supervised learning are two types of machine learning methods. Input data is estimated or mapped with desired output by using the training data labeled set in supervised learning. But there is not providing labeled examples in unsupervised learning and during the learning process there is no idea about output. Input data of supervised learning is called as training data and at a time it has result or known label as spam/not-spam [12].

A training process prepares the model and make the predictions when it required and make them correct if the predictions are wrong. Once the training data can achieves desired accuracy levels then the training process stops. With the algorithm of trained machine learning fake profiles can be detected and it is the main aim of machine learning method [13]. The training data is having the particulars of person as gender, age and friends list. So the fake profiles are detected or predicted with these particulars and data security is enhanced on social networking sites. Naïve Bayes (NB), Decision Tree (DT) and Support Vector Machine (SVM) are used in proposed machine learning algorithms. From prediction result account activities analysis is also provided [14].

The researchers did so far make use of traditional machine learning algorithms like random forest, naive bayes, SVM, and decision tree. These methods are incapable of doing feature selection on their own. Thus the researcher has to study relation between the features and target variable in order to decide which features are to be considered and which can be rejected. Another drawback is being their inability to adapt with the changing patterns in the input dataset which can make them insufficient at times. Hence these methods require constant monitoring. The changing patterns in the input can cause them to give incorrect results thus reducing the accuracy. Also one major issue with them is that they do not perform well if the dataset is too large or is unstructured. This makes traditional methods highly unsuitable for real life scenarios as in such cases where the data is mostly unstructured and often too large[17]. Owing to the drawbacks of the traditional methods it has become necessary to explore advance algorithms like deep neural network [15,18,20].

3. Twitter Fake Profile Detection Using SVM

Fake profile detection model designing for twitter presented in this paper uses the machine learning concept. Training and testing are two main stages in Machine learning framework. Fig. 1 shows the block diagram of proposed detection model for Twitter fake profile detection using SVM.

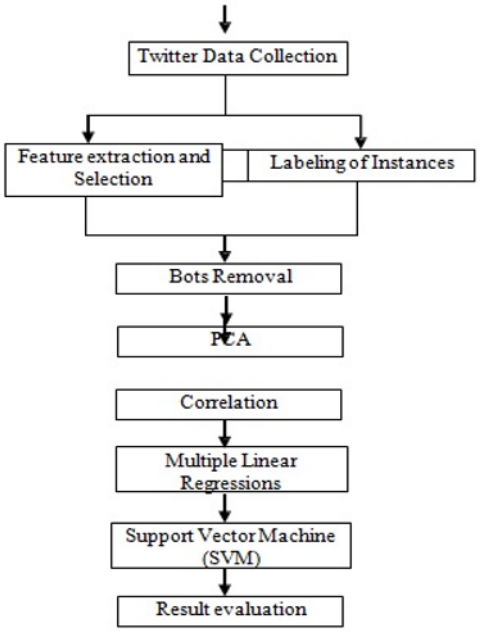

Fig. 1: Framework of Twitter Fake Profile Detection Using SVM

3.1 Twitter data collection

Collection of twitter data is the first process of this method. For research purpose publically available data or API streaming twitter data is used.

3.2 Feature selection

The collected dataset feature selection is processed in this step. Spam account detection uses the different feature parameters, in that some are useless. So from extracted features only useful features are selected. Spam account detection effective results are dependent on the selected features. The estimated threshold value is 0.8, and below this correlation levels to the class variable feature pairs are eliminated by using Spearman's Rank-Order correlation.

Total 11 sets of correlated feature pairs are selected in this step as output. Relevance analysis step selects the features and are used as the inputs of redundancy analysis step. If two values are correlated completely then that features are said to be redundant each other but the features determination is not straightforward in reality when one feature is correlated with set of features.

Hence, redundant features are eliminated by using Markov Blanket technique. In a Bayesian network Markov Blanket for a node A, MB (A) consists of group of nodes with A's parents, its children and other parents of its children. Neighboring nodes set forms the node Markov Blanket in Markov random field. Non-redundant features of two output sets with two different versions are obtained when applying Markov blanket on correlated feature pairs of MB (Fi) and MB (Fj). These redundancy step output features has further used in classification section. Two feature sets are selected and 1 represents the contributing features whereas 0 represents the ignored features. The purpose of training uses the labeling of set of samples which is small with spam or non-spam. The services of spam filtering or manual process the labeling action. Only spam free instances are allowed by spam filtering so label the instances as non-spam on the other hand spam effective instances are eliminated so label the instances as spam.

3.3 Principal Component Analysis "PCA"

The dimensions of feature vectors are reduced by using the dimension reduction technique of principle component analysis (PCA). The best features are discovered and can efficiently describe the data as well as unnecessary features are stripping by assignment of lower weights which results that mining process cannot impact. From 16 PCAs total 10 PCAs are selected in this work. 92% of data is covered with selected 10 PCAs.

3.4 Spearman's Rank-Order Correlation

Most used feature selection method is Spearman's Rank-Order Correlation. In between X and Y quantitative variables there exists a monotonic relationship and its direction and strength are measured by this correlation method. If X and Y are independent then 0 is output measure of this correlation and if the values are in between -1 and +1 which indicates the direction and level of correlation. Each and every variable correlation coefficient is the outputs of this algorithm which are represented in the form of table.

3.5 Multiple Linear Regression

The relationship in between dependent variable or independent variable as predictor input and response output is described by the models of Linear Regression. Two linear variables are considered in simple linear regression as x and y, in this one variable is dependent on others as shown in below equation 1 as;

$$y = a + bx \quad (1)$$

Where, a is a constant, regression coefficient is denoted with b. two or more independent variables are considered in multiple linear regression in which dependent variable value is predicted as shown in below equation 2.

$$y = a + bx1 + cx2 + dx3 \quad (2)$$

One independent variable with 16 dependent variables dataset is used in multiple linear regressions. Multi co-linearity problem is raised in the multiple linear regressions. Here multiple factors are correlated not in terms of response variable but also to each other. Standard coefficient errors are increased with Multi co-linearity problem and making some variables as statistically insignificant and some as significant. Out of 16 predictions total 12 predictions

are obtained as result after removing the redundant variables.

3.6 Support Vector Machine (SVM) Wrapper method is mostly used feature selection model. A learning model selects and qualifies the different feature subsets in this method. Highest predictive performance with subset features is selected. Bit manipulation can calculates the all subsets and for given set 2 , n subsets are calculated; here n is the feature number for F set. For instance, there will be 2 3 are subsets of set {1, 2, 3}. The feature set with best performance is provided in this method and there is a requirement intensive computation for large feature space. 16 feature vectors are existed in baseline dataset so 2^{16}-1 = 65,535 subsets are possible without empty subset. Particular data instance class is identified by testing after the training of detection models based on machine learning with labeled samples. Complex algorithms and models are developed by using machine learning in data analytics field for prediction themselves. Detection model performance influences the prediction accuracy. Single classifier or group of classifiers is present in detection model of spam account detection system. Hybrid SVM (Support Vector Machine) is used in development process of these classifiers. Data mining algorithms stability, different labeled instances and features proper selection are the factors which influence the detection model performance.

3.6 Results Evaluation

Detection rate, accuracy, false negative, true positive, f-measures, precision, recall and etc are evaluation parameters which are evaluates the performance of detection models.

4. Results

Once the proposed SVM classifier had been trained, their effectiveness was evaluated. Using confusion matrix we can describe the performance of the classification model. The most fundamental terms used with a confusion matrix for a binary classifier are:

- True-positive (TP): the number of accounts correctly identified as Faked.
- False-positive (FP): the number of accounts incorrectly identified as Faked.
- True-negative (TN): the number of accounts correctly identified as Trusted.
- False-negative (FN): the number of accounts incorrectly identified as Trusted

These can be further used to find following metrics to determine the effectiveness of each model:

Precision: Precision is the ratio of true positives to the values predicted correctly. It is defined in Eq.3, given as follows:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Recall: Recall is the ratio of true positives to the total number of positives. It is defined in Eq.4, given as follows:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negative}$$

Accuracy: the correct identification of accounts from corpus are determined by using the parameter accuracy.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

The MIB Datasets are used which are of fake and legitimate Twitter accounts published by the Institute of Informatics and Telematics (IIT), Italian National Research Council's (CNR). The data set included 11,737 Twitter accounts. This dataset is divided in to two parts as training set with 70% of data and testing set with 30% of data. Total feature subsets in the dataset are then trained and also tested with the use of

the proposed hybrid SVM technique. The following Table 1 shows the precision and Recall metrics of the different ML techniques such as Logistic Regression, random forest (RF), SVM, and proposed hybrid SVM in predicting of identity deception by humans on Twitter.

Table 1: Comparative Analysis of Different Classified Techniques

Parameter	Precision	Recall
RF	58.52%	63.21%
Logistic Regression	83.51%	86.63%
SVM	73. 89%	76.98%
Hybrid SVM	98.16%	85.87%

These results indicate that the different supervised machine learning models such as SVM, and proposed hybrid SVM in predicting of identity deception by humans on Twitter. At best, an Accuracy of 98.16% was achieved from the proposed hybrid SVM machine learning model.

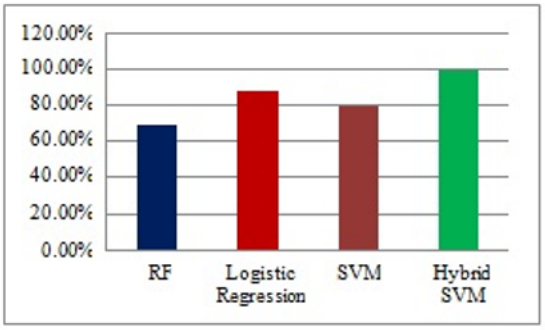

Fig. 2: Detection Accuracy of Different Machine Learning Techniques

5. Conclusion

The techniques of machine learning modules usage is increasing day to day. The usage of datasets with fake profiles efficiently eliminates the difficulty in finding fake profiles. The detection technique of fake accounts that are created by humans is described in this paper. A new hybrid classification algorithm is introduced for detecting fake profiles efficiently on social networks. Multi linear regression method is used for finding the values of SVM trained model. Spearman's Rank-Order correlation is used to reduce the feature vector. Remarkable accuracy is then obtained with correlation feature set among different features. Best features are selected in correlation technique and redundancy is removed. From the result analysis it has been observed that the proposed hybrid SVM achieved an accuracy of 98% in fake profile detection on twitter and acquired a better performance as well as efficient one compared to the other existing machine learning techniques.

References

1. Koyel Chakraborty, Siddhartha Bhattacharyya, Rajib Bag, **"**A Survey of Sentiment Analysis from Social Media Data", IEEE Transactions on Computational Social Systems, Volume: **7**, Issue: 2, (2020)

2..Muhammad Adil, Rahim Khan, M. Ahmad Nawaz Ul Ghani, "Preventive Techniques of Phishing Attacks in Networks", 2020 *3rd International Conference on Advancements in Computational Sciences (ICACS),* (2020)

3.Fatih Cagatay Akyon, M. Esat Kalfaoglu, "Instagram Fake and Automated Account Detection", *2019 Innovations in Intelligent Systems and Applications Conference (ASYU),* (2019)

4.Ranojoy Barua, Rajdeep Maity, Dipankar Minj, Tarang Barua, Ashish Kumar Layek, "F-NAD: An Application for Fake News Article Detection using Machine Learning Techniques", 2019 *IEEE Bombay Section Signature Conference (IBSSC)*, (2019)

5.Ebtihal A. Hassan, Farid Meziane, "A Survey on Automatic Fake News Identification Techniques for Online and Socially Produced Data", 2019 *International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE),* (2019)

6.Estée Van Der Walt, Jan Eloff, "Using Machine Learning to Detect Fake Identities: Bots vs Humans", IEEE Access, Volume: 6, (2018)

7.Naeimeh Laleh, Barbara Carminati, Elena Ferrari, "Risk Assessment in Social Networks Based on User Anomalous Behaviors", IEEE Transactions on Dependable and Secure Computing, Volume: **15**, Issue: 2, (2018)

8.Md. Arafatur Rahman, Vitaliy Mezhuyev, Md Zakirul Alam Bhuiyan, S. M. Nazmus Sadat, Siti Aishah Binti Zakaria, Nadia Refat, "Reliable Decision Making of Accepting Friend Request on Online Social Networks", IEEE Access, Volume:**6**, (2018)

9.Myo Myo Swe, Nyein Nyein Myo, "Fake Accounts Detection on Twitter Using Blacklist",

10.Estée Van Der Walt, Jan Eloff, "Using Machine Learning to Detect Fake Identities: Bots vs Humans", IEEE Access, Volume: 6, (2018)

11.Nafiseh Sedaghat, Mahmood Fathy, Mohammad Hossein Modarressi, Ali Shojaie, "Combining Supervised and Unsupervised Learning for Improved miRNA Target Prediction", IEEE/ACM Transactions on Computational Biology and Bioinformatics, Volume: **15**, Issue: 5, (2018)

12.Naman Singh, Tushar Sharma, Abha Thakral, Tanupriya Choudhury, "Detection of Fake Profile in Online Social Networks Using Machine Learning", 2018 *International Conference on Advances in Computing and Communication Engineering (ICACCE)*, (2018)

13.Simon Fong, Yan Zhuang, Jiaying He, "Not every friend on a social network can be trusted: Classifying imposters using decision trees", *The First International Conference on Future Generation Communication Technologies*, (2012)

14.Huaizu Jiang, Jinjun Wang, Yihong Gong, Na Rong, Zhenhua Chai, Nanning Zheng, "Online Multi-Target Tracking With Unified Handling of Complex Scenarios", IEEE Transactions on Image Processing, Volume: **24**, Issue: 11,(2015)

15. Kui Wu, Xuancong Wang, Nina Zhou, AiTi Aw, Haizhou Li, "Joint Chinese word segmentation and punctuation prediction using deep recurrent neural network for social media data", 2015 International Conference on Asian Language Processing (IALP), (2015).

16. Raghunadha Reddy, T., Vishnu Vardhan, B., Vijayapal Reddy, P,"A survey on Authorship Profiling techniques", International Journal of Applied Engineering Research, 11 (5), (2016),pp. 3092-3102.

17. Swaraja K , "Medical image region based watermarking for secured telemedicine", Multimedia Tools and Applications, 77 (21), (2018) ,pp. 28249-28280.

18. Kumar, P., Singhal, A., Mehta, S., Mittal, A,"Real-time moving object detection algorithm on high-resolution videos using GPUs", Journal of Real-Time Image Processing, 11 (1), (2016) , pp. 93-109.

19.Kumar, S.K., Reddy, P.D.K., Ramesh, G., Maddumala, V.R."Image transformation technique using steganography methods using LWT technique", Traitement du Signal (2019)
20. Mahalle, G., Salunke, O., Kotkunde, N., Gupta, A.K., Singh, S.K,"Neural network modeling for anisotropic mechanical properties and work hardening behavior of Inconel 718 alloy at elevated temperatures", Journal of Materials Research and Technology, 8 (2), pp. 2130-2140.(2019)