IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 5, Issue 4, April 2016*

# Machine Learning Techniques for Fake Profiles Classification in Online Social Network

**Latha P[1], Dr. M.V. Vijaykumar[1]**

Department of Computer Science and Engineering, Dr. AIT, Bengaluru, Karnataka, India[1]

**Abstract:** Online social networks (OSN's) such as facebook, twitter, linked in, weibo have become important aspect of users in daily life. Social network have changed the way people interact, it fulfils user needs by information sharing and appreciation. OSNs suffer from fake profiles creation. An OSN user faces security problems like identity theft, privacy violation, leakage of personal information, identity theft. Fake users may inject spam, modify the online ratings and extract knowledge of genuine user. It is very difficult to detect, verify fake profiles manual. There is a need of automation of this fake profile user's identification process. This paper explores different machine learning techniques which can be used for classification of fake profiles and real users.

**Keywords:** online social networks, fake profiles, classification, machine learning algorithms.

## I. INTRODUCTION

Online Social Networks (OSNs) such as facebook, digg, twitter, linkedin, tuenti has become more popular because of the services they provide for information sharing, knowledge sharing, events organizing, appreciations and respect. This popularity has increased and interest for attacking and modifying user profile data for increase in the business value. One of the problem reported in 2010 that 1.5 million fake facebook accounts were distributed for sale. Fake (Sybil) OSN accounts can be used for various purposes [1]. For instance, they enable spammers to abuse an OSN's messaging system to post spam [2], or waste OSN advertising. Figure 1 shows sample online social networking graph.

OSNs are web-based services that facilitate individuals to construct a profile, which is either public or semi-public. SNS contains list of users with whom we can share a connection, view their activities in network and also converse. SNS users communicate by messages, blogs, chatting, video and music files. OSN also have many disadvantages such as information is public, security problem, cyber bullying and misuse and abuse of OSN platform.

The security issues in SNS are divided in the following four groups: privacy breaches, viral marketing, network structural attacks, and malware attacks. The large number of users and the characteristics of the online social networks make them particularly vulnerable to malicious content propagation. Often performed in the form of URLs contained in messages, malicious attacks can lead to a very large number of people being infected in a very short time. From simple use of a profile for performing attacks, malicious actors have now moved on to a more collective and synchronized way of undertaking actions. This allows them to increase the impact of their attack by increasing the total number of targeted profiles. As a result, the detection of malicious profiles is insufficient to eradicate malicious campaigns and a characterization of them is required.

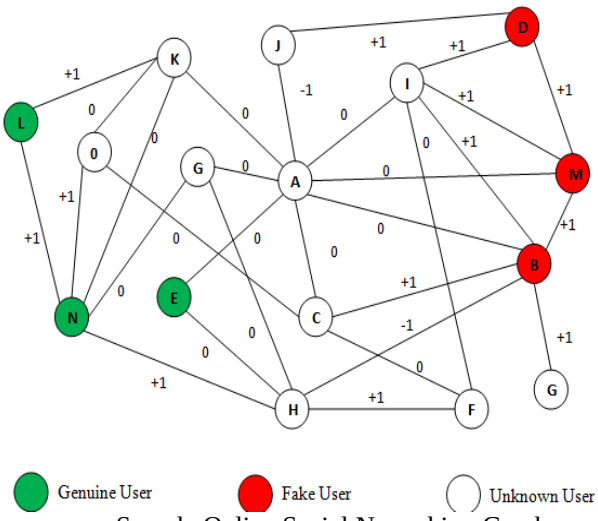


Fig. 1. Sample Online Social Neworking Graph

In section II, we discuss works related to this topic. In section III, different machine learning techniques which can be used for classification of fake profiles are presented. We proposed algorithms in section IV, followed by experimental result in section V. Finally, we give the conclusion in section VI.

## II. LITERATURE SURVEY

Several research works has been carried out in the field of online social networks fake profile classification. Cai et al. [3] addresses the issue of detecting Sybil attacks in social networks. In collaborative and recommendation based computer systems, attackers may create fake or malicious identities to gain more influence in the system. This work proposes a statistic model (the latent community model) and associated learning algorithms for detecting Sybil attacks. The latent community model groups the nodes in a network into closely linked communities that are linked relatively loosely with the rest of the graph.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 5, Issue 4, April 2016*

Privacy and confidentiality are the key attributes which needs to be taken care in online social networks for providing the services effectively. Many research works has been proposed to decrease the threats on confidentiality and user privacy for the user group containing profiles in OSNs. For instance 800 million for facebook users one of most liked OSNs. Conti et.al planned the concept of Virtual Private Social Network (VPSN). VPSN fundamentally is a sign of the concept of Virtual Private Network (identified in computer networks) contained by online social networks: only friends surrounded by the VPSN are capable to observe the valid and actual information of an individual [4]. Other users in the online social network, as well as the online social network manager, do not have permission to access and observe the same information. The majority of the effort in the literature intended at preventing the information of the online social network in lack of awareness for unauthorized users, which mean to be accessed merely by the authorized users in the authorized way. For instance, Mahmood et.al showed how an attacker may obtain right of entry to the information that the victim distributes and shares in the profile, not in favor of the victim's wishes [5].

Yang et. al. [6] does a case study of cyber criminal ecosystem on Twitter. The objective is to detect spam accounts individually and analyze Twitter spam account's social relationships. Twitter criminal accounts usually publish or link to malicious content, which intends to damage or disrupts users' browsers or computers, or to compromise user's privacy. According to the analysis, criminal accounts are even more socially connected than legitimate users. Some accounts are in the center; some are in the edge. This work designs an algorithm to infer more criminal accounts and exploiting the properties of their social relationships and semantic correlations

## III. MACHINE LEARNING TECHNIQUES

This section discussed about different machine learning algorithms

### A. Support Vector Machine

The Support Vector Machine is a learning procedure based on statistical learning theory. SVMs have originally been developed to solve classification problems but can be extended to regression problems as well [7]. Hereto, an alternative loss function that includes a distance measure is introduced [8]. In this case a ε-insensitive loss function is used where,

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \le \varepsilon, \\ |\xi| - \varepsilon & \text{otherwise.} \end{cases}$$

This means that we do not accept any deviations that are larger than ε. In other words, the goal is to find a function that has at most ε deviation from the actual target values yi for every point xi in the training set (with n data points and m variables)

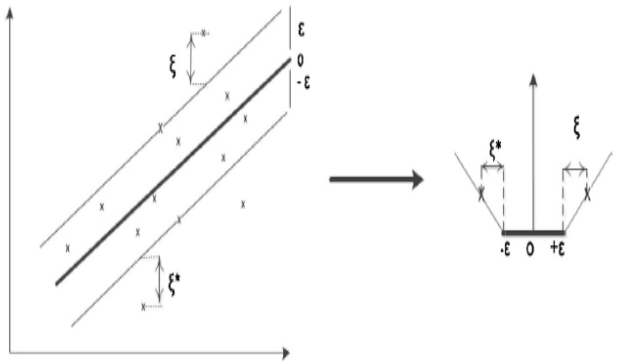


Fig. 2. SVM classifier

### B. Random Forest

Random Forest (RF) is an ensemble method constructing a collection of univariate tree classifiers [9]. The first step is to take L samples of size n at random from the original data using bootstrap sampling. Next, from every sample, a tree is built by splitting on k (≤ m) input variables randomly selected from the total amount of input variables in the original data. The number of selected input variables is a hyper parameter of the learner and stays constant during the forest growing process [10]. After inducing a large number of trees, a majority voting procedure is performed to decide on the final class. Although Random Forest builds upon multiple decision trees which are comprehensible, the ensemble loses this advantage and creates an output that is hard to interpret (a set of hundreds of trees).

### C. Naive Bayes

A Naive Bayes Classifier can be defined as an independent feature model that deals with a simple probabilistic classifier based on Bayes' theorem with strong independence assumptions [11]. There are several models which assume different fitting for Naive Bayes. The most common models are: Bernoulli Event Model characterized as Boolean weight which uses binary feature occurrences; another one is the multinomial model which uses feature occurrence frequencies. Consider the bug classification into n different classes C = {C1,C2…….,Cn}.The unseen fake user(Bi) will be classified using (2) to class with higher posterior probability.

$$P(Ck,Bi)=P(Bi|Ck)(P(Ck)/P(Bi)$$

P(Ck) is the prior probability of class Ck calculated using (3), N is the number of fake users in the training data and Nk is used to denote total number of fake users from training data which belong to class Ck.

$$P(Ck)=Nk/N$$

Instead of taking word information as input we are using feature information for fake profile specific features and for features of natural language type we are considering word information. Words are unigram features but extracted features from fake information may be Bi-gram, Trigram or Multigram. Fake specific features may be a combination of number of words as in Trace Decode and Commands.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 5, Issue 4, April 2016*

## IV. IMPLEMENATION

Here to identify the real profile and fake profile their behaviour are modeled based on their activity. Real profile is tagged with a value 1 and fake profile is tagged with the value -1 and value with 0 is considered are as unknown which need to be identity. Each relation is weighted with +1 if user connected tend to have similar behavior and -1 if they tend to have opposite behavior. The value 0 represents that the behavior of a user is independent of other connected user. This information is used in the application for classifying the Unknown users as real or fake. Based on the small number of known reak and Fake users in the large social network and the number of connections that the users have, a initial prior class label is assigned. The initial class label is set to real or fake depending on the potential of that user becoming of reak or Fake by its neighbors. It sets the initial probability for this person becoming a real or Fake.

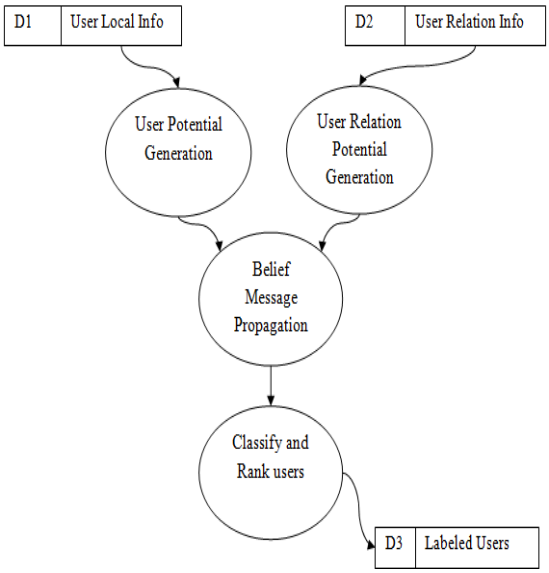


Fig. 3. Work flow of classifying and ranking

User Relation Potential Generator computes the coupling strength of the relation between two users connected. Belief Propagation Inference implements the Belief Propagation Technique to infer the class of each uses in the social network. Each nodes/users pass message to their neighboring nodes with their probability distribution for being real and Fake. This is propagated to all the users and their probability of being real and Fake gets updated whenever a message from their neighbor is received. This process is repeated for predefined iterations or till the probability doesn't change between iterations. These modules outputs the prediction of class labels (real or Fake) with their probability for all users. Classify and Rank sorts the users with their probability of Fake and ranks them. The highest probable Fake user gets top rank.

## V. RESULTS

We collected publicly available data sets on online social network. The algorithms need to be trained using the training dataset and should be evaluated using the testing dataset. Implementation is done using java.

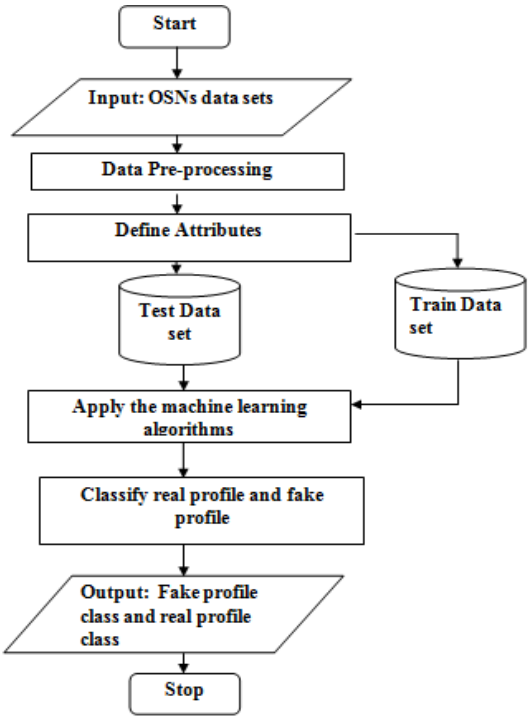


Fig. 4. Work flow of overll classificaiton process

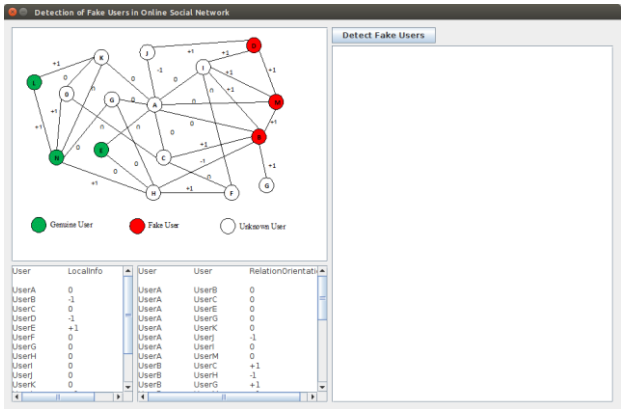


Fig. 5. Snapshot of Data set loading for classificaiton process

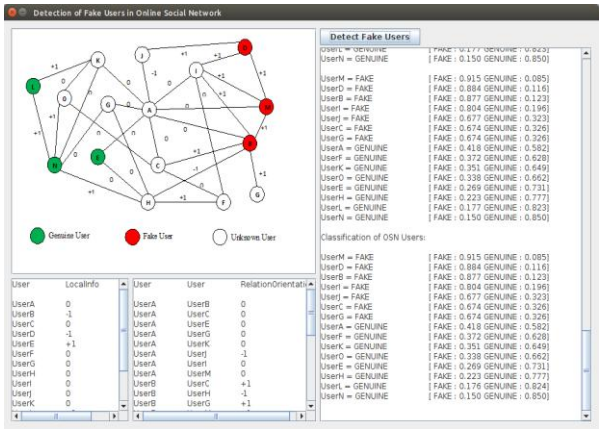


Fig. 6. Snaopshot of data preprocessing classificaiton process


DOI 10.17148/IJARCCE.2016.54283

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 5, Issue 4, April 2016*

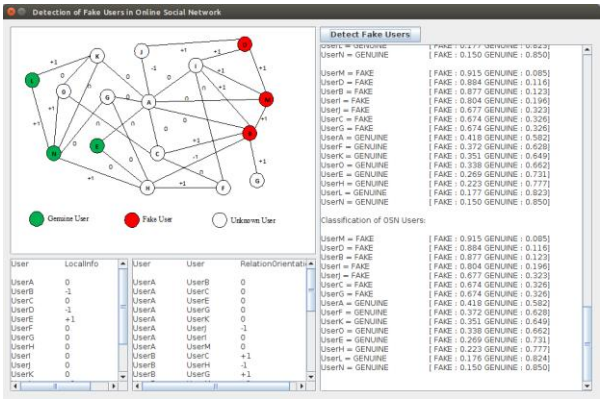


Fig. 7. Snapshot of classifued real profiles and fake profiles

## VI. CONCLUSION

Nowadays OSNs have become more popular in level of ages and many activities user does will be presented in these sites. There is a serious threat in OSNs because many business and personal interactions which will be linked with OSNs. When the identity of the real user is theft, attacker may use his/her personal information for making friendship, access images and tagged unwanted post. The real user will be the victim of cause. So in order to overcome this problem there is a need of new approaches for automating fake profile classification process. We explored some of the machine learning algorithms for fake profile classification. The experimental result shows effectiveness of the proposed approach. In future work, more machine learning algorithms will be explored on different data sets.

## REFERENCES

[1] Fake Accounts in Facebook - How to Counter it. http://tinyurl.com/5w6un9u, 2010.
[2] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao. Detecting and Characterizing Social Spam Campaigns. In IMC, 2010.
[3] Z. Cai and C. Jermaine. The latent community model for detecting sybils in social networks. In NDSS, 2012.
[4] Conti, M, Hasani, A and Crispo, B, Virtual Private Social Networks, Proceedings of the First ACM CODASPY (2011).
[5] Mahmood, S, and Desmedt, Y, Your Facebook Deactivated Friend or A Cloaked Spy, In Proceedings of the Proceedings of the 4th IEEE International Workshop on SESOC (2012).
[6] C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu. Analyzing spammers' social networks for fun and pro_t. In WWW, pages 71-80, 2012.
[7] Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer, New York.
[8] Schölkopf, B., Smola, A.J., 2002. Learning with Kernels: Support Vector Machines,Regularization, Optimization, and Beyond. MIT press.
[9] Tan, P., Steinbach, M., Kumar, V., 2006. Introduction to Data Mining. Pearson Educ-tion, Boston, USA
[10] Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
[11] S. P. Bingulac, "On the compatibility of adaptive controllers (Published Conference Proceedings style)," in Proc. 4th Annu. Allerton Conf. Circuits and Systems Theory, New York, 1994, pp. 8–16.