International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 12 | Dec 2019 www.irjet.net p-ISSN: 2395-0072

Fake Profile Identification using Machine Learning

Samala Durga Prasad Reddy

Mahatma Gandhi Institute of Technology, Computer Science and Engineering, Hyderabad, Telangana, India

---

Abstract - *In the present generation, the social life of everyone has become associated with online social networks. These sites have made a drastic change in the way we pursue our social life. Making friends and keeping in contact with them and their updates has become easier. But with their rapid growth, many problems like fake profiles, online impersonation have also grown. There are no feasible solutions exist to control these problems. In this paper, I came up with a framework with which the automatic identification of fake profiles is possible and is efficient. This framework uses classification techniques like Random Forest Classifier to classify the profiles into fake or genuine classes. As this is an automatic detection method, it can be applied easily by online social networks that have millions of profiles whose profiles cannot be examined manually.*

Key Words: Social Media, Facebook, Random Forest Classifier, Classification, Framework, and Dataset.

1. INTRODUCTION

Social networking site is a website where each user has a profile and can keep in contact with friends, share their updates, meet new people who have the same interests. These Online Social Networks (OSN) use web2.0 technology, which allows users to interact with each other. Social networking sites are growing rapidly and changing the way people keep in contact with each other. The online communities bring people with the same interests together which makes users easier to make new friends.

1.1. History

These social networking sites starting with http://www.sixdegrees.com in 1997 then came http://www.makeoutclub.com in 2000. Sixdegrees.com couldn't survive much and closed very soon but new sites like myspace, LinkedIn, Bebo became successful and Facebook was launched in 2004 and presently it is the largest social networking site in the world.

1.2. Social Impact

In the present generation, the social life of everyone has become associated with online social networks. These sites have made a drastic change in the way we pursue our social life. Adding new friends and keeping in contact with them and their updates has become easier. Online social networks have an impact on science, education, grassroots organizing, employment, business, etc. Researchers have been studying these online social networks to see the impact they make on the people. Teachers can teach the students easily through this making a friendly environment for the students to study, teachers nowadays teachers are getting themselves familiar with these sites bringing online classroom pages, giving homework, making discussions, etc. which improves education a lot. The employers can use these social networking sites to employ the people who are talented and interested in the work, their background check can be done easily using this. Most of the OSN is free but some charge the membership fee and uses this for business purposes and the rest of them raise money by using the advertising. This can be used by the government to get the opinions of the public quickly. The examples of these social networking sites are sixdegrees.com, The Sphere, Nexopia which is used in Canada, Bebo, Hi5, Facebook, MySpace, Twitter, LinkedIn, Google+, Orkut, Tuenti used in Spain, Nasza-Klasa in Poland, Cyworld mostly used in Asia, etc. are some of the popular social networking sites.

2. Objective

In today's online social networks there have been a lot of problems like fake profiles, online impersonation, etc. To date, no one has come up with a feasible solution to these problems. In this project, I intend to give a framework with which the automatic detection of fake profiles can be done so that the social life of people become secured and by using this automatic detection technique we can make it easier for the sites to manage the huge number of profiles, which can't be done manually.

3. LITERATURE SURVEY

Various fake record recognition methodologies depend on the investigation of individual interpersonal organization profiles, with the point of distinguishing the qualities or a combination thereof that help in recognizing the legitimate and the fake records. In particular, various features are extracted from the profiles and posts, and after that Machine learning algorithms are used so as to construct a classifier equipped for recognizing fake records.

For instance, Nazir et al. (2010) [1] describes recognizing and describing phantom profiles in online social gaming applications. The article analyses a Facebook application, the online game "Fighters club", known to provide incentives and gaming advantage to those users who invite their peers into the game. The authors contend that by giving such impetuses the game motivates its players to make fake profiles. By presenting those fake profiles into the game, the

user would increase a motivating force of an incentive for him/herself.

Adikari and Dutta (2014) [2] depict recognizable proof of fake profiles on LinkedIn. The paper demonstrates that fake profiles can be recognized with 84% exactness and 2.44% false negative, utilizing constrained profile information as input. Techniques, for example, neural networks, SVMs, and Principal component analysis are applied. Among others, highlights, for example, the number of languages spoken, training, abilities, suggestions, interests, and awards are utilized. Qualities of profiles, known to be fake, posted on uncommon sites are utilized as a ground truth.

Chu et al. (2010) [3] go for separating Twitter accounts operated by humans, bots, or cyborgs (i.e., bots and people working in concert). As a part of the detection problem formulation, the Identification of spamming records is acknowledged with the assistance of an Orthogonal Sparse Bigram (OSB) text classifier that uses pairs of words as features.

Stringhini et al. (2013) [4] analyze Twitter supporter markets. They describe the qualities of Twitter devotee advertises and group the clients of the business sectors. The authors argue that there are two major kinds of accounts who pursue the "client": fake accounts("sybils"), and compromised accounts, proprietors of which don't presume that their followers rundown is expanding. Clients of adherent markets might be famous people or legislators, meaning to give the appearance of having a bigger fan base, or might be cybercriminals, going for making their record look progressively authentic, so they can rapidly spread malware what's more, spam. Thomas et al. (2013) [5] examine black market accounts utilized for distributing Twitter spam.

De Cristofaro et al. (2014) [6] investigate Facebook like cultivates by conveying honeypot pages. Viswanath et al. (2014) [7] identify black market Facebook records based on the examination of anomalies in their like behavior. Farooqi et al. (2015)

[6] explore two black hat online commercial centers, SEO Clerks and My Cheap Jobs. Fayazi et al. (2015) think about manipulation in the online review.

Table 1: Profile-based methods for Identifying Social Media Accounts

Reference	Ground truth	Detection method	Accuracy
Adikari 2015	Known fake LinkedIn profiles, posted on special web sites	Number of languages spoken, education, skills, recommendations, interests, awards, etc. are used as features to train neural networks, SVMs, and principal component analysis.	84% TP, 2.44% FN
Chu et al. 2010	Manually labelled 3000x2 Twitter profiles as human, bots, or cyborgs.	1. Text classification via Bayesian classifier (Orthogonal Sparse Bigram); 2. Regularity of tweets; 3. Frequency and types of URLs; the use of APIs.	100%
Lee et al. 2010	Spam accounts registered by honeypots: 1500 in MySpace and 500 in Twitter	Over 60 classifiers available in Weka are tried. Features include: i) demographics, ii) content and iii) frequency of content generation, iv) number and type of connections. The Decorate meta-classifier provided the best results.	99,21% (MySpace), 88,98% (Twitter)
Stringhini et al. 2010	Spam accounts registered by honeypots: 173 spam accounts in Facebook and 361 in Twitter	Random forest was constructed based on the following features: ratio of accepted friend requests, URL ratio, message similarity, regularity in the choice of friends, messages sent, and number of friends.	2% FP, 1% FN (Facebook); 2.5% FP, 3.0% FN (Twitter)
Yang et al. 2011a	Spam Twitter accounts defined as the accounts containing malicious URLs: 2060 spam accounts	Graph based features (local clustering coefficient, betweenness centrality, and bi-directional links ratio), neighbor-based features (e.g., average neighbors' followers), automation-based features (API ratio, API URL ratio and API Tweet similarity), and timing-based features were used to construct different classifiers.	86% TP, 0,5% FP
Yang et al. 2011b	1000 legit and 1000 fake accounts provided by Renren	Invitation frequency, rate of accepted outgoing and incoming requests, and clustering coefficient were used as features for an SVM classifier.	99%

4. PROPOSED FRAMEWORK

4. 1. Overview

Each profile (or account) in a social network contains lots of information such as gender, no. of friends, no. of comments, education, work, etc. Some of this information is private and some are public. Since private information is not accessible so, we have used only the information that is public to determine the fake profiles in the social network. However, if our proposed scheme is used by the social networking companies itself then they can use the private information of the profiles for detection without violating any privacy issues. We have considered this information as features of a profile for the classification of fake and real profiles. The steps that we have followed for the identification of fake profiles are as follows.

1. First, all the features are selected on which the classification algorithm is applied. Proper care should be taken while choosing features such as features that should not be dependent on other features and those features should be chosen which can increase the efficiency of the classification.

2. After proper selection of attributes, the dataset of previously identified fake and real profiles are needed for the training purpose of the classification algorithm. We have made the real profile dataset whereas the fake profile dataset is provided by the Barracuda Labs, a privately held company providing security, networking and storage solutions based on network appliances and cloud services.

3. The attributes selected in step 1 are needed to be extracted from the profiles (fake and genuine). For the social networking companies which want to implement our scheme don't need to follow the scrapping process, they can easily extract the features from their database. We applied to scrap off the profiles since no social network dataset is available publicly for the research purpose of detecting the fake profiles.

4. After this, the dataset of fake and real profiles are prepared. From this dataset, 80% of both profiles (real and fake) are used to prepare a training dataset and 20% of both profiles are used to prepare a testing dataset. We find the efficiency of the classification algorithm using the training dataset containing 922 profiles and a testing dataset containing 240 profiles.

5. After the preparation of the training and the testing dataset, the training dataset is feed to the classification algorithm. It learns from the training algorithm and is expected to give correct class levels for the testing dataset.

6. The levels from the testing dataset are removed and are left for determination by the trained classifier. The efficiency of the classifier is calculated by calculating the no. of correct predictions divided by total no. of predictions. We have used three classification algorithms and have compared the efficiency of the classification of these algorithms.

4.2 Proposed framework

The proposed framework in figure 1 shows the sequence of processes that need to be followed for continues detection of fake profiles with active learning from the feedback of the result given by the classification algorithm. This framework can easily be implemented by social networking companies.

1. The detection process starts with the selection of the profile that needs to be tested.

2. After the selection of the profile, the suitable attributes (i.e. features) are selected on which the classification algorithm is implemented.

3. The attributes extracted is passed to the trained classifier. The classifier gets trained regularly as new training data is feed into the classifier.

4. The classifier determines whether the profile is fake or genuine.

5. The classifier may not be 100% accurate in classifying the profile so; the feedback of the result is given back to the classifier.

6. This process repeats and as the time proceeds, the no. of training data increases and the classifier becomes more and more accurate in predicting the fake profiles.

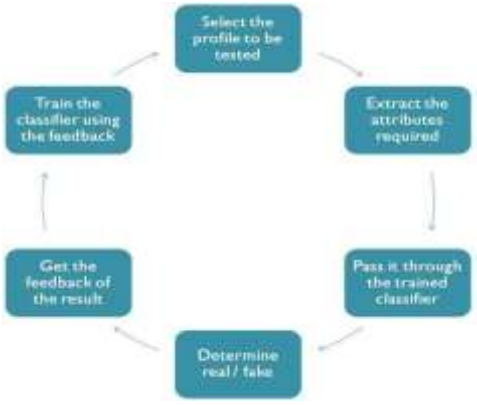

Figure 1: Framework for Identification of fake profiles

4.3 Classification

Classification is the process of learning a target function f that maps each record, X consisting of a set of attributes to one of the predefined class labels, Y. A classification technique is an approach of building classification models from an input data set.

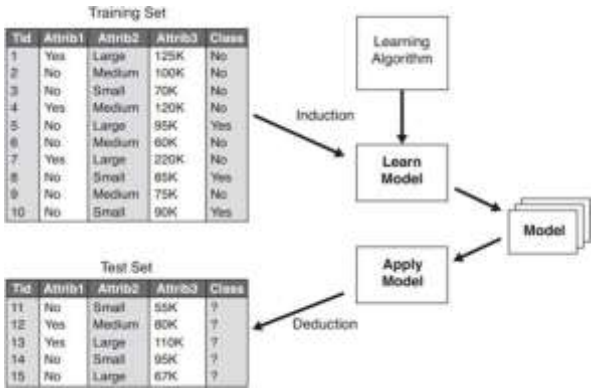

Figure 2: General approach for building a classification model

This technique uses a learning algorithm to identify a model that best fits the relationship between the attribute set and a class label of the training set.

Figure 2 shows the general approach for building a classification model. The model generated by the learning algorithm should both fit the input data correctly and correctly predict the class labels of the test set with as high accuracy as possible. The key objective of the learning algorithm is to build the model with good generality capability.

The classifier that I have implemented for classifying the profiles is Random Forest.

4.4. Random Forest

Random forest is a supervised learning algorithm that is used for both classifications as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees mean more robust forests. Similarly, the random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method that is better than a single decision tree because it reduces the over-fitting by averaging the result.

We can understand the working of the Random Forest algorithm with the help of following steps:

Step 1 – First, start with the selection of random samples from a given dataset.

Step 2 – Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.

Step 3 – In this step, voting will be performed for every predicted result.

Step 4 – At last, select the most voted prediction result as the final prediction result.

Figure 3 will illustrate its working :

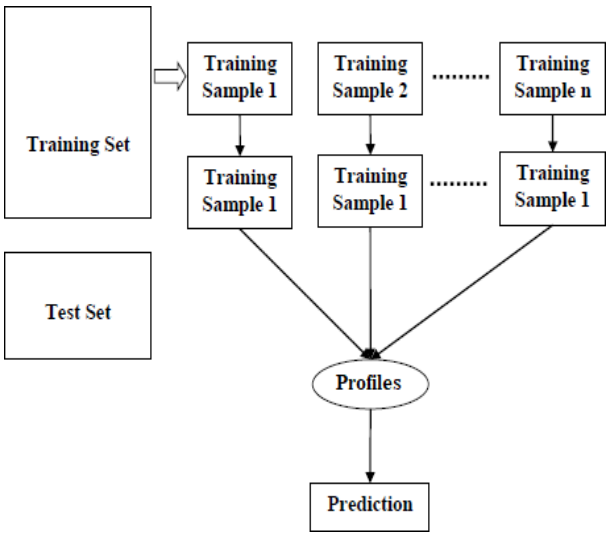

Figure 3: working of Random Forest

5. IMPLEMENTATION

5.1. Dataset

We needed a dataset of fake and genuine profiles. Various attributes included in the dataset are a number of friends, followers, status count. Dataset is divided into training and testing data. Classification algorithms are trained using a training dataset and the testing dataset is used to determine the efficiency of the algorithm. From the dataset used, 80% of both profiles (genuine and fake) are used to prepare a training dataset and 20% of both profiles are used to prepare a testing dataset.

5.2. Attributes Considered

Table 2 shows the Attributes considered for fake profile identification and the description for each of the attributes is provided.

Table 2: Attributes Considered for the fake profile Identification

S. No	Attribute	Description
1.	Profile ID	The Profile ID of account holder
2.	Profile Name	The name of the account holder
3.	Status Count	The number of tweets made by the account
4.	Followers Count	The number of followers for the account
5.	Friends Count	The number of friends for the account
6.	Location	The location of the account holder
7.	Created Date	The date the account was created
8.	Share count	The number of shares done by account holder
9.	Gender	The Gender of the account holder
10.	Language Code	The language of account holder

5.3. Evaluation Parameters

Efficiency/Accuracy = Number of predictions/Total

Number of Predictions Percent Error = (1-Accuracy)*100

Confusion Matrix - Confusion Matrix is a technique for summarizing the performance of a classification algorithm. Calculating a confusion matrix can give you a better idea of what your classification model is getting right and what types of errors it is making.

TPR- True Positive Rate TPR=TP/(TP+FN)

FPR- False Positive Rate FPR=FP/(FP+TN)

TNR- True Negative Rate TNR=TN/(FP+TN)

FNR- False Negative Rate FNR=1-TPR

Recall- How many of the true positives were recalled (found), i.e. how many of the correct hits were also found.

Recall = TP / (TP+FN)

Precision- Precision is how many of the returned hits were true positive i.e. how many of the found were correct hits.

Precision = TP / (TP + FP)

F1 score- F1 score is a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score.

ROC Curve- The Receiver Operating Characteristic is the plot of TPR versus FPR. ROC can be used to compare the performances of different classifiers.

6. RESULTS

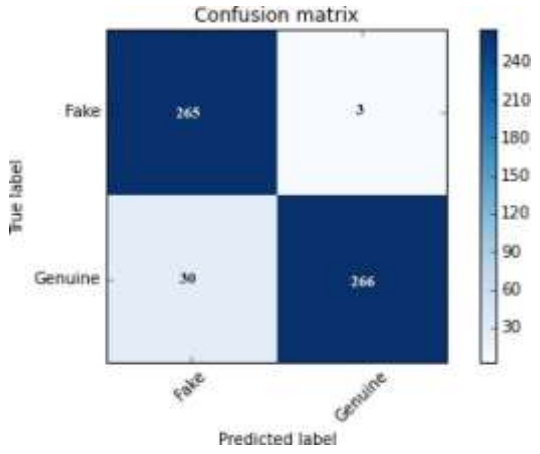

Figure 4: Confusion Matrix

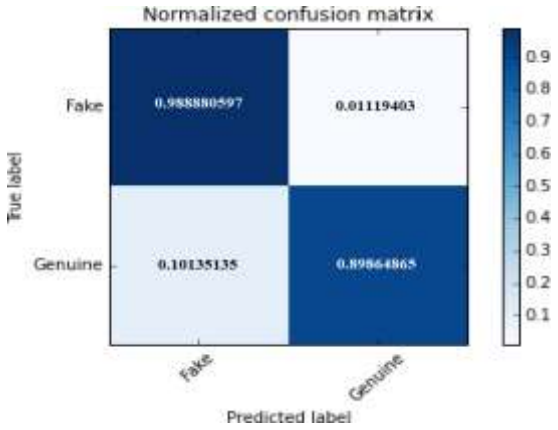

Figure 5: Normalized Confusion Matrix

	precision	recall	f1-score	support
Fake	0.85	0.98	0.91	268
Genuine	0.98	0.84	0.90	296
avg / total	0.91	0.90	0.90	564

Figure 6: Classification Report

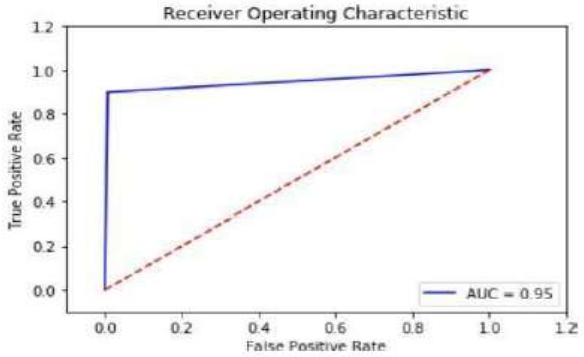

Figure 5.4: ROC curve

The efficiency of the Random Forest Classifier in classifying data is 95%. We have taken 80% of the data for the training dataset and 20% for the testing dataset.

7. CONCLUSION

We have given a framework using which we can identify fake profiles in any online social network by using Random Forest Classifier with a very high efficiency as high as around 95%. Fake profile Identification can be improved by applying NLP techniques and Neural Networks to process the posts and the profiles. In the future, we wish to classify profiles by taking profile pictures as one of the features.

ACKNOWLEDGEMENT

I am grateful to numerous people who have contributed to shaping this research paper. At the outset, I would like to express my sincere thanks to Professor Dr. A. Nagesh for his advice during my research paperwork. As my supervisor, he has constantly encouraged me to remain focused on achieving my goal. I sincerely thank everyone who has provided us with new ideas, constructive criticism, and their invaluable time. Last but not least; I would like to thank all my colleagues and friends for their help and cooperation.

REFERENCES

[1] Nazir, Atif, Saqib Raza, Chen-Nee Chuah, Burkhard Schipper, and C. A. Davis. "Ghostbusting Facebook: Detecting and Characterizing Phantom Profiles in Online Social Gaming Applications." In *WOSN*. 2010.

[2] Adikari, Shalinda, and Kaushik Dutta. "Identifying Fake Profiles in LinkedIn." In *PACIS*, p. 278. 2014.

[3] Chu, Zi, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. "Who is tweeting on Twitter: human, bot, or cyborg?." In Proceedings of the 26th annual computer security applications conference, pp. 21-30. ACM, 2010.

[4] Stringhini, Gianluca, Gang Wang, Manuel Egele, Christopher Kruegel, Giovanni Vigna, Haitao Zheng, and Ben Y. Zhao. "Follow the green: growth and dynamics in twitter follower markets." In Proceedings of the 2013 conference on Internet measurement conference, pp. 163-176. ACM, 2013.

[5] Thomas, Kurt, Damon McCoy, Chris Grier, Alek Kolcz, and Vern Paxson. "Trafficking Fraudulent Accounts: The Role of the Underground Market in Twitter Spam and Abuse." *In* Presented as part of the 22nd {USENIX} Security Symposium ({USENIX} Security 13), pp. 195-210. 2013.

[6] Farooqi, Gohar Irfan, Emiliano De Cristofaro, Arik Friedman, Guillaume Jourjon, Mohamed Ali Kaafar, M. Zubair Shafiq, and Fareed Zaffar. "Characterizing Seller-Driven Black-Hat Marketplaces." arXiv preprint arXiv: 1505.01637 (2015).

[7] Viswanath, Bimal, M. Ahmad Bashir, Mark Crovella, Saikat Guha, Krishna P. Gummadi, Balachander Krishnamurthy, and Alan Mislove. "Towards detecting anomalous user behavior in online social networks." In 23rd {USENIX} Security Symposium ({USENIX} Security 14), pp. 223-238. 2014.

BIOGRAPHY

Samala Durga Prasad Reddy Currently pursuing a Bachelor of Technology in Computer Science and Engineering from Mahatma Gandhi Institute of Technology, Hyderabad, Telangana, India.