Detecting Fake Accounts in Media Application Using Machine Learning

Gayathri A
Dept. of Computer Science, VelammalEngg College Chennai, India
Radhika S
Dept. of Computer Science, VelammalEngg College Chennai, India
Mrs. Jayalakshmi S.L.
Assistant Professor, Dept. of Computer Science, VelammalEngg College Chennai, India

--ABSTRACT--

The social network, a crucial part of our life is plagued by online impersonation and fake accounts .Fake profiles are mostly used by the intruders to carry out malicious activities such as harming person , identity theft and privacy intrusion in Online Social Network(OSN).Hence identifying an account is genuine or fake is one of the critical problem in OSN .In this paper we proposed many classification algorithm like Support Vector Machine algorithm and deep neural network .It also studies the comparison of classification methods on SpamUser dataset which is used to select the best.

Keywords - **fake accounts, fake identities, social media, data science, friends, followers, fake profiles**

I. INTRODUCTION

In the present generation, everyone in society has become associated with the Online Social Networks(OSN). These OSN have made a drastic change in the way we pursue our social life. Making new friends, keeping in contact with them and knowing their updates has become easier. But with the rapid growth of social media many problems like fake profiles, online impersonation have also grown.There are no feasible solution existing to control these problems .Fake accounts can be either human-generated, computer-generated(also referred as “bots”), or cyborgs[1]. A cyborg is half-human, half-bot account [1]. Such an account is manually opened by a human, but from then onwards the actions are automated by a bot.

To become member of the OSN the user has to create his profile by entering information like name, photo, date of birth, Email ID, graduation details, place of work, home town, interests and so on [2][3]. Some of the fields are mandatory and some are optional and it varies from one OSN to the another. These websites are popular because of people’s interest in finding friends, sharing pictures, tagging people in group photos,sharing their ideas and views on common topics, maintain good business relationship and general interest with others.

In this paper we came up with a framework in which automatic detection of fake profiles is possible and is efficient. This framework uses classification techniques like Support Vector Machine, Random Forest and Deep Neural Networks to classify the profiles into fake or genuine classes.As it is an automatic detection method, it can be applied easily by OSN which has millions of profile where profiles cannot be examined manually .We evaluate whether readily available and engineered features that are used for the successful detection ,using machine learning models.

II. RELATED WORK

This paper presents some filtering algorithms that rely on classification to decide whether the profile is genuine or fake.

III. SUPPORT VECTOR MACHINE

Support Vector Machine is a binary classification algorithm that finds the maximum separation hyper plane between two classes. It is a supervised learning algorithm that gives enough training examples, divides two classes fairly well and classifies new examples .It offers a principle approach to machine learning problems because of their mathematical foundation in statistical learning theory [10]. SVM construct their solution as a weighted sum of SVs ,which are only a subset of the training input .It is effective in cases where number of dimensions is greater than the number of samples given

IV. RANDOM FOREST

Random Forest is versatile method performing both classification and regression tasks[8]. It has nearly same hyperparameters as a decision tree or a bagging classifier .It creates many variations of trees .The best outcome will be used to predict identity deception .Each outcomes from the classifier represents different section of a tree.

V. SPAM FILTERING

The research study by Simranjit Kaur et al [4] is based on implementing a k-mean clustering algorithm on vector set to increase efficiency .To detect spam emails using neural networks the two phases namely training and testing are needed to be done. The process of detecting spam and phishing emails using feed forward neural network .The paper has 11 features have been implemented as a binary values 0 or 1 with value 1 indicating this feature appeared in the tested email and value 0 indicating non-appearance case.

VI. Activity Pattern

The research study by Jiang et al used Catch Sync to detect suspicious behavior in Twitter based on synchronized and abnormal user activity. They were able to show that their approach resulted in high efficiency of detecting fake accounts in Twitter.

VII. Supervised Machine Learning

Garadi et al [7] evaluates whether the readily available and engineered features that are used for the successful detection using machine learning algorithms of fake identities created by bots or computers can be use to detect the fake identities created by humans. It is done by considering that similar features can serve as a catalyst for uncovering identity deception by humans on online social networks.

VIII. einforcement Machine Learning

Venakatesan et al [5] presented a reinforcement proof-of-concept model that rewards itself for detecting bots successfully. Reinforcement machines learning models require feedback from the environment to adjust and improve. This is not readily available in social media network.

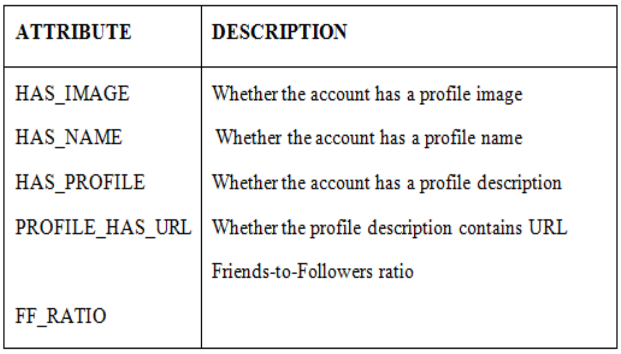
IX. Unsupervised Machine Learning

Miller et al [6] proposed that supervised machine learning models require a label included in the corpus to predict the expected outcome. With unsupervised machine learning the data isunlabeled and data are being grouped based on the similarity of the data considered. It is not practical to search the class consisting of fake accounts. The norm is to train a one class support vector machine on the minority class.

X. Filtering

When a new threat is identified and verified will that sender be added to a blacklist [9]. Similar methods of dealing with spam have been proposed on twitter to blacklist known malicious URL content and quarantine known as bots.

Table 1: Attributes used in previous study.



ATTRIBUTE	DESCRIPTION
HAS_IMAGE	Whether the account has a profile image
HAS_NAME	Whether the account has a profile name
HAS_PROFILE	Whether the account has a profile description
PROFILE_HAS_URL	Whether the profile description contains URL
FF_RATIO	Friends-to-Followers ratio

XI. Motivation Work

In today’s online social networks there have been a lot of problems like fake profiles, online impersonation etc., Till date , no one has come up with a feasible solution to these problems .In this project we 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 using automatic detection technique we can make it easier for the sites to manage the huge number of profiles which can’t be done manually.

XII. Proposed Work

This paper proposes the detection process starts with the selection of the profile that needs to be tested. After selection of the profile the suitable attributes ie., features are selected on which the classification algorithm is being implemented ,the attributes extracted is passed to the trained classifier .

The classifier is being trained regularly as new training data set is feed into the classifier. The classifier determines whether the profile is fake or real. The classifier may not be 100 % accurate in classifying the profile so the feedback obtained from the result is being given back to the classifier. For example if the profile is identified as fake ,social networking sites can send notification to the profile to submit details.

Classification is the process of learning a target function f that maps each record consulting of set of attributes to one of the predefined classes models from an input data set. Classification technique is a approach of building classification models from an input data set. This technique uses a learning algorithm to identify a model that best fits the relationship between the attribute set and class label of the training set.

The model generated by the learning algorithm should both fit the input data correctly and correctly predict the class labels of the learning algorithm is to build the model with good generality capability.Different steps are executed to classify an account as fake or genuine profiles. They are:

Data set of both fake and genuine profiles with various attributes like number of friends ,followers, status count. Dataset is divided into training and testing data. Classification algorithm are trained using training dataset and testing data set is used to determine the efficiency of algorithm .From the dataset used 80% of both (real and fake) are used to prepare a training data set and 20% of both profiles are used to prepare a testing dataset.

Features are selected to apply classification algorithms. The classification algorithm is being discussed further. Attributes are selected as features if they are not dependent on other attributes and they increase efficiency of the classification.

Table 2: **FEATURES EXTRACTED**

S.NO	FEATURES
1	Number of friends
2	Number of
3	followers
4	Favorite Count
5	Languages
6	Known
7	Sex code
	Listed Count
	Status Count

After selection of attributes, the dataset of profiles that are already classified as fake or genuine are needed for the training purpose of the classification algorithm. We have used a publicly available dataset of 1337 fake users and 1481 genuine users consisting of various attributes including listed count, status count, number of friends, followers count, favourites, languages known, sex code.

Classification is the process of categorizing a data object into categories called classes based upon features/attributes associated with that data object. Classification uses a classifier, an algorithm that processes the attributes of each data object and outputs a class based upon this information. In this project, we use Support Vector Machine as a classifier. Support Vector Machine is an elegant and robust technique for classification on a large data set not unlike the data sets of Social Network with several millions of profiles. Algorithm used for classification are Support Vector Machine, Random Forest and Deep Neural Networks.

Confusion Matrix is a technique for describing the performance of a classification algorithm. Confusion Matrix is used to give you a better idea of what your classification model is getting right and what types of errors it is making. All the algorithm results are plotted in confusion matrix to know where the error has occurred.

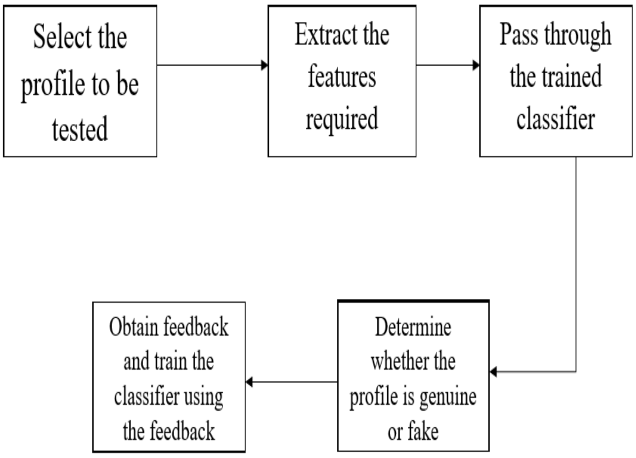

Select the profile to be tested
Extract the features required
Pass through the trained classifier
Determine whether the profile is genuine or fake
Obtain feedback and train the classifier using the feedback

Fig 1: RESEARCH STEPS

XIII. Experimental Results

We used Keras with TensorFlow backend to implement the Multi-Layer Perceptron model. We have used a 1-hidden layer neural network with 500 hidden units. The output obtained from the neural network is a single value which we pass through the sigmoid non-linearity to squish it in the range [0, 1] The sigmoid function is defined by the output from the neural network gives the probability (positive tweet) i.e. the probability of the tweets sentiment being positive. At the prediction step, we round off the probability values to class labels 0 (negative) and 1 (positive). Red hidden layers represent layers with sigmoid non-linearity. We also conducted experiments using SGD + Momentum weight updates and found out that it takes too long to cover the entire data set. We ran our model up to 20 epochs after which it began to over fit. Thus identifying the profile is real or fake.

We used sparse vector representation of tweets for training the classifier. We identify that the presence of bigrams features significantly improved the accuracy. The overall accuracy across all machine learning models was very high with the highest being 94.43% using Deep Neural Networks and 94% using Random Forest method and finally 90.01% using Support Vector Machine algorithm. These results are just below what one would expect from getting the prediction right by chance.

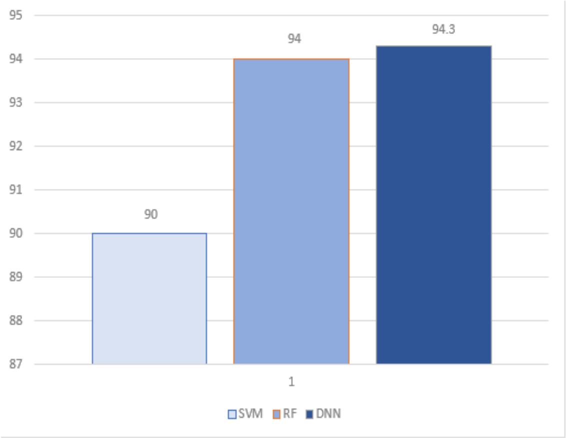


Fig 2: RESULTS

XIV. CONCLUSION

In this Project we have presented a machine learning pipeline for detecting fake accounts in online social networks. Rather than making a prediction for each individual account, our system classifies clusters of fake accounts to determine whether they have been created by the same actor. Our evaluation on both in-sample and out-of-sample data showed strong performance, and we have used the system in production to find and restrict more than 250,000 accounts. In this work we evaluated our framework on clusters created by simple grouping based on registration date and registration IP address. In future work we expect to run our model on clustering that are created by grouping on other features, such as ISP and other time periods, such as week or month.

Another promising line of research is to use more sophisticated clustering algorithms such as k-means or hierarchical clustering. While these approaches may be fruitful, they present obstacles to operating at scale: k-means may require too many clusters (i.e., too large a value of k) to produce useful results and clustering of data may be too intensive for classifying millions of accounts in Online Social Network.

From a modeling perspective, one important direction for future work is to apply feature sets used in other spam detection models, and hence to realize multi-model ensemble prediction. Another direction is to make the system robust against adversarial attacks, such as a botnet that diversifies all features, or an attacker that learns from failures.

REFERENCES

[1] Estee Van Der Walt and Jan Eloff,"Using Machine Learning to Detect Fake Identities:Bots vs Humans"IEEE Trans. Emerg.TopicsComput. Intell., vol. 1, no. 1, pp. 61–71 March 2018.

[2] Loredana Caruccio,DomenicoDesiato and Giuseppe Polese"Fake Account Identification in Social Networks" IEEE International Conference on Big Data.,, vol. 9, no. 6, pp. 811–824,2018.

[3] SarahKhaled,Neamat El-Tazi and Hoda M. O. Mokhtar"Detecting Fake Accounts on Social Media" IEEE International Conference on Big Data.., vol.6 pp 101-110 ,2018.

[4] SuyashSomani and Somya Jain "Resolving Identities on FaceBook and Twitter" Tenth International Conference on Contemporary Computing (IC3), 10-12 August 2017.

[5] FrancescoBuccafurri, Gianluca Lax,Denis Migdal, Serena Nicolazzo, Antonino Nocera and Christophe Rosenberger"Contrasting False Identities in Social Networks by Trust Chains and Biometric Reinforcement " International Conference on Cyberworlds vol 5,2017.

[6] MohamedTorky, Ali Meligy and Hani Ibrahim"Recognizing Fake Identities In Online Social Networks Based on a Finite Automaton Approach"International Journal of Computer Applications, 2016.

[7] SuprajaGurajala, Joshua S White, Brian Hudson,Brian R Voter and Jeanna N Matthews"Profile characteristics of fake Twitter accounts"Big Data &Society,July–December 2016: 1–13.

[8] Simranjit. Kaur. Tuteja, ‘‘A survey on classification algorithms for email spam filtering,’’ International Journal Eng. Sci., vol. 6, no. 5, pp. 5937–5940, 2016.

[9] M.A.Devmane and N.K.Rana "Detection and Prevention of Profile Cloning in Online Social Networks"IEEE International Conference on Recent Advances and Innovations in Engineering,May 09-11, 2014.

[10] SaraKeretna ,Ahmad Hossny and Doug Creighton "Recognising User Identity in Twitter Social Networks via Text Mining"IEEE International Conference on Systems, Man, and Cybernetics,2013.