Detection of Fake Profiles on Twitter using Random Forest & Deep Convolutional Neural Network

[1]PRIYANKA SHAHANE, [2]DEIPALI GORE

[1]M. E. Scholar, Computer Department, P. E. S. Modern College of Engineering, Pune, Maharashtra

[2]Assistant Professor, Computer Department, P. E. S. Modern College of Engineering, Pune, Maharashtra

ABSTRACT: The number of fake profiles on social media platforms such as Twitter is increasing day by day with the continued growth of these platforms. These Fake profiles are a preferred means for malicious users to commit various cyber crimes. Hence, their detection and removal has become necessary in order to protect legitimate users and maintain trustworthiness. This system describes the scalable approach to classify fake vs. real profiles based on various features such as user name, friends count, followers count, location, profile age, profile image etc. Here, we use supervised machine learning to train our classifiers i.e. Random Forest and Deep Convolutional Neural Network and test our system on Twitter dataset.

KEY WORDS: social media, cyber crimes, machine learning, deep learning, activation functions.

I. Introduction

Now a days, people around the world rely on social media for various purposes such as sharing opinions, knowledge and experiences, expanding their social connectivity, browsing for resources and information. However, the features that make social media valuable for users also makes them susceptible to different types of cyber crimes. Fake profiles are preferred means for offenders to fulfill their malicious intents. The various malicious intents of fake identities includes,

- Disseminating malware:

For example: Misdirecting users to some malicious websites, Stealing credentials by creating false communications.

- Influencing actions of users:

For example: News of bomb blast in the city.

- Cyber bullying:

For example: Grooming victims in order to commit various crimes.

- Manipulation of credit worthiness of account:

For example: Creating fake identities on the name of some reputed person and disseminating some unsolicited content over it in order to defame that person.

- Skewing perceptions:

For example: Fake identities are used to create fake likes for advertisement of a particular product in order to create illusion that their product is better than that of their competitors.

Fake profiles can be generated by humans or computers (bots). Here we focus on identification of fake profiles created by humans as very little research has been done so far to detect fake profiles created by humans.

II. LITERATURE SURVEY

Eastee Van Der Walt et. al.[1] implemented SVM Linear, Random Forest and Adaboost for classification of Fake vs. Real profiles on twitter and they found that Random Forest gives the best result. It has been observed that there are many features available on social media platforms that describes identity of particular profile. For example, location, name, profile image, followers count, friends count, account created date, URL count, status count, number of re-tweets.

Indira Sen et. al.[2] implemented Random Forest, Logistic Regresssion, AdaBoost with Random Forest as base initiator, Multilayer Perceptron with feed forward architecture, Support Vector Machine with RBF Kernel for identification of fake likes on instagram and they found that Multi Layer Perceptron with Feed Forward architecture gives best results. For Multi Layer Perceptron, sigmoid activation function is used here.

Cao Xiao et. al. [3] implemented Support Vector Machine, Logistic Regression and Random Forest for detection of clusters of fake accounts on LinkedIn dataset and they found that Random Forest gives the best results. They have used three different types of features i.e. basic distribution features, pattern features and frequency features.

Christie Fuller et. al. [4] implemented Logistic Regression, Decision Tree and Neural Network while creating automated decision support tool for deception identification and they found that Neural Network gives best result. They have used “person of interest statements” dataset which is officially known as Form 1168.

Sai Peddinti et. al. [5] implemented a classifier designed by themselves which converts four class classification problem into the binary classification problem for Twitter dataset. Then one binary classifier classifies profile as anonymous or non-anonymous while other classifies as identifiable or non-identifiable. Finally, results of these two classifiers are considered in order to classify each profile as ‘unknown’, ‘identifiable’ or ‘anonymous’ for Twitter dataset. They have used Random Forest as a base classifier for binary classifiers. Here, the choice of classifier and trees count is dependent on cross validation results.

This survey states that Random Forest and Neural Networks gives best performance for detection of fake vs. real profiles on social media.

III. SYSTEM ARCHITECTURE

The flow of system is as follows:

A. Collect Data:
The Twitter profiles are extracted from Twitter via Twitter API based on two keywords i.e. “Homework” and “School”. Here, we have selected these two keywords because they are frequently used by minors in their communications and offenders generally target minors on social media.

B. Pre-process Data:
The data is pre-processed with the help of various steps such as,

- Lexical analysis:
  This process separates input data into two categories i.e.
  a) Word characters such as letters, numbers and symbols.
  b) Word separators such as tabs, newlines and spaces.

- Remove stopwords:
  This step removes most common words in the documents such as pronouns, conjunctions, prepositions and articles.

- Porter stemmer:
  This algorithm replaces all the variants of word such as gerunds, plurals, past tense suffixes and third person suffixes by a single stem word.

- Feature selection:
  Here, we select various features such as user name, screen name, account created date, number of retweets, listed count, location, followers count, friends count, tweets count, number of url and profile image for classification of fake vs. real identities.

- Bots removal:
  Since in this system we focus on detection of fake identities created by humans, bots are removed during pre-processing based on certain parameters such as absence of punctuation, absence of profile name and image.

C. Create Fake Profiles:
Here, we create some fake profiles in order to train our classifiers i.e. Random Forest and Deep Convolutional Neural Network.

D. Validate profiles:
Fake profiles that we have created manually are evaluated using Chi Square Test and Mann Whitney U Test in order to prove that fake profiles that we have injected manually belongs to the same corpus that we have previously extracted from Twitter via Twitter API.

E. Inject Fake profiles:
Fake profiles that pass above two tests are injected into the system.

F. Create new features:
Here, some new features are engineered from the existing features in order to improve the classification accuracy of the system. For example, (Number of URLs)/(Total number of tweets), (Friends Count)/(Followers Count)^2 ratios are high for fake profiles.

G. Classification:

We have experimented with 4 Fold, 10 Fold and 12 Fold cross validation techniques in order to train our classifiers. We have tested for two different classifiers,

- Random Forest (RF):
  This algorithm creates number of decision trees by randomly picking up the features from our feature set in order to test the class label of desired Twitter profile. Finally, majority voting is performed on the output of all the decision trees to predict final output of the Random Forest algorithm,(Figure 1).

- Deep Convolutional Neural Network (DCNN):
  It is fully connected feed forward network with one input layer, two hidden layers and one output layer. Here, we have tested for linear, sigmoid and tan h activation functions. For deep learning, Deeplearning4j libraries are used.

H. Evaluate Results:
Finally, performance of the system is evaluated based on various performance metrics such as accuracy, precision, recall and f1 score.

- Accuracy : $\frac{(TP+TN)}{(TP+TN+FP+FN)}$ (1)

- Precision : $\frac{TP}{(TP+FP)}$ (2)
- Recall : $\frac{TP}{(TP+FN)}$ (3)
- F1 score :2*(Precision* Recall)/(Precision + Recall) (4)

Where,

TP is True Positive

TN is True Negative

FP is False Positive

FN is False Negative

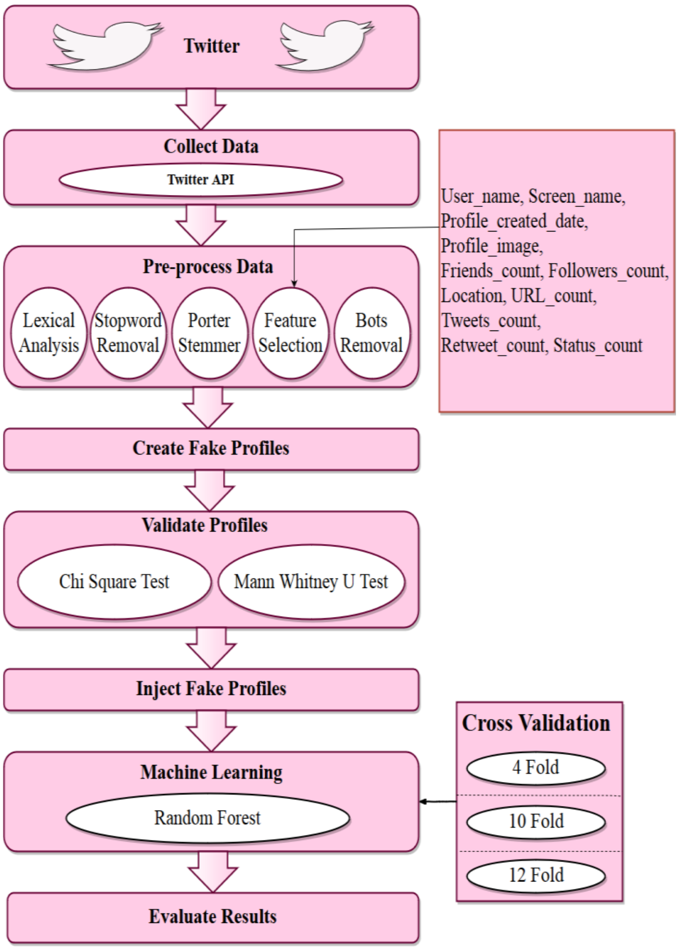


Figure1. Detection of Fake vs. Real Profiles on Social Media using Random Forest algorithm.

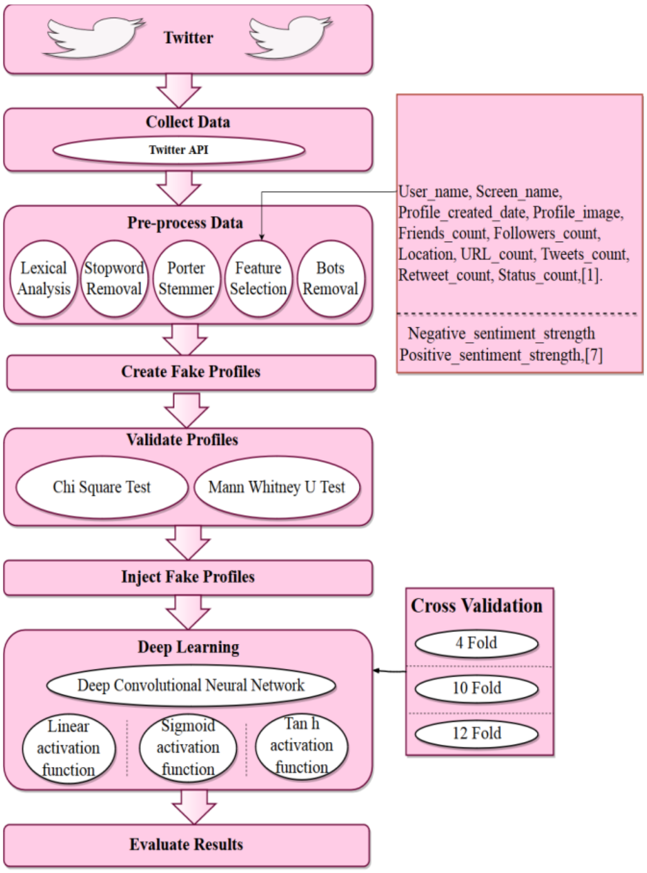

Twitter
Collect Data
Twitter API
Pre-process Data
Lexical Analysis
Stopword Removal
Porter Stemmer
Feature Selection
Bots Removal
User_name, Screen_name, Profile_created_date, Profile_image, Friends_count, Followers_count, Location, URL_count, Tweets_count, Retweet_count, Status_count,[1].
Negative_sentiment_strength Positive_sentiment_strength,[7]
Create Fake Profiles
Validate Profiles
Chi Square Test
Mann Whitney U Test
Inject Fake Profiles
Cross Validation
4 Fold
10 Fold
12 Fold
Deep Learning
Deep Convolutional Neural Network
Linear activation function
Sigmoid activation function
Tan h activation function
Evaluate Results

Figure 2. Detection of Fake vs. Real Profiles on Social Media using Deep Convolutional Neural Network.

IV. DATASET

Dataset consists of profile information as well as Tweets of Twitter users. Number of Twitter profiles extracted via Twitter API is around 4000 and the number of fake profiles injected manually is about 3000. So, the data set consists of near about 7000 Twitter accounts. Data is saved in MySQL database as well as in .txt file.

V. MATHEMATICAL MODEL

Let S be a system having Input (I), Functions (F) and Output (O).
S = {I, F, O}

Where,
I is a set of Twitter profiles from different users.

I = {Profile 1, Profile 2,….., Profile N}

O is the authenticity of account.

O = {Fake, Real}

F is the set of functions used to predict authenticity of the account.

F = { F1, F2, F3, F4 }

Where,
F1 is a function for Random Forest.
F2 is a function for Linear activation of DCNN.
F3 is a function for Sigmoid activation of DCNN.
F4 is a function for Tan h activation of DCNN.

F1 is a function for Random Forest and it is given by,

$$w = \sum_{i=0}^{n}(inp[i]) = (hid[i]) \quad (5)$$

w > t: 1;

w < t: 0

Where,

w = Weight assigned based on equality of input and hidden identities.

t = Threshold kept on calculated weight to detect fake identities

Here, input (inp) corresponds to the profiles whose class label is to be detected and hidden (hid) profiles corresponds to the training data whose class label is already known.

F2 is a function for Linear activation of DCNN and it is given by,

$$y = a + k \quad (6)$$

Where,

k = $\sum w_i \, x_i$

x_i = Set of features.

w_i = Weights associated with features.

a = Bias.

F3 is a function for Sigmoid activation of DCNN and it is given by,

$$y = 1/(1 + e^{-k}) \quad (7)$$

Where,

k = $\sum w_i \, x_i$

x_i = Set of features.

w_i = Weights associated with features.

F4 is a function for Tan h activation of DCNN and it is given by,

$$y = \tanh(x) = 2/(1 + e^{-2k}) - 1 \quad (8)$$

Where,

k = $\sum w_i \, x_i$

x_i = Set of features.

w_i = Weights associated with features.

VI. RESULTS AND ANALYSIS

In order to evaluate the performance of a system we have analyzed accuracy, f1 score, precision and recall with which fake identities on social media can be detected,

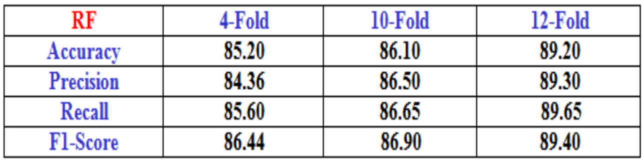
A. Using different cross validation techniques:

- 4-fold
- 10-fold
- 12-fold

B. Using different activation functions of DCNN:

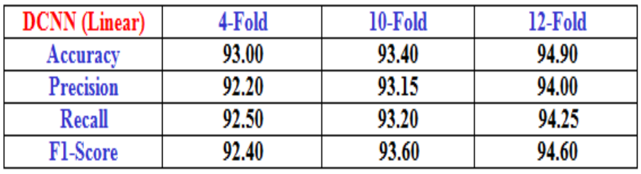
- Linear
- Sigmoid
- Tan h

TABLE 1: CLASSIFICATION RESULTS FOR FAKE VS. REAL PROFILES USING RANDOM FOREST WITH 4 FOLD, 10 FOLD and 12 FOLD CROSS VALIDATION.



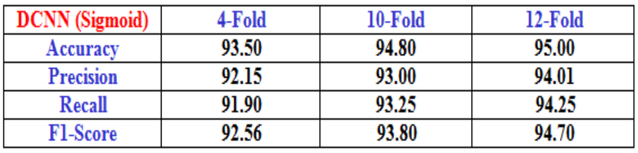
RF	4-Fold	10-Fold	12-Fold
Accuracy	85.20	86.10	89.20
Precision	84.36	86.50	89.30
Recall	85.60	86.65	89.65
F1-Score	86.44	86.90	89.40

TABLE 2: CLASSIFICATION RESULTS FOR FAKE VS. REAL PROFILES USING DCNN (LINEAR ACTIVATION FUNCTION) WITH 4 FOLD, 10 FOLD and 12 FOLD CROSS VALIDATION.



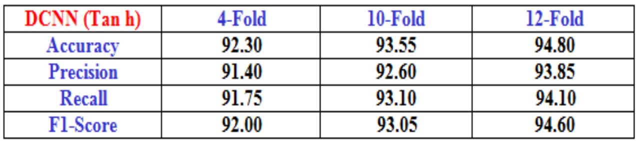
DCNN (Linear)	4-Fold	10-Fold	12-Fold
Accuracy	93.00	93.40	94.90
Precision	92.20	93.15	94.00
Recall	92.50	93.20	94.25
F1-Score	92.40	93.60	94.60

TABLE 3: RESULS FOR CLASSIFICATION OF FAKE VS. REAL IDENTITIES USING DCNN (SIGMOID ACTIVATION FUNCTION) WITH 5 FOLD, 10 FOLD and 15 FOLD CROSS VALIDATION.



DCNN (Sigmoid)	4-Fold	10-Fold	12-Fold
Accuracy	93.50	94.80	95.00
Precision	92.15	93.00	94.01
Recall	91.90	93.25	94.25
F1-Score	92.56	93.80	94.70

TABLE 4: CLASSIFICATION RESULTS FOR FAKE VS. REAL IDENTITIES USING DCNN (TAN H ACTIVATION FUNCTION) WITH 4 FOLD, 10 FOLD and 12 FOLD CROSS VALIDATION



DCNN (Tan h)	4-Fold	10-Fold	12-Fold
Accuracy	92.30	93.55	94.80
Precision	91.40	92.60	93.85
Recall	91.75	93.10	94.10
F1-Score	92.00	93.05	94.60

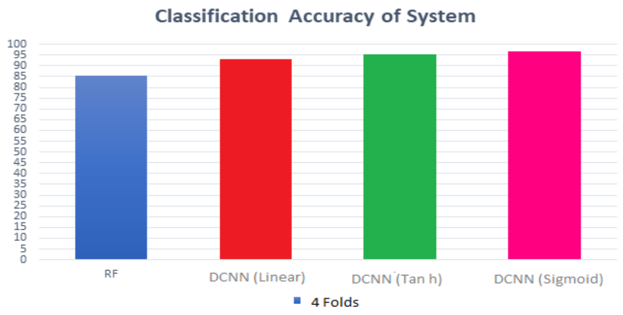


Figure 3. Comparative analysis of accuracy for RF, DCNN(Linear), DCNN(Sigmoid) and DCNN(Tan h) using 4 Fold cross validation.

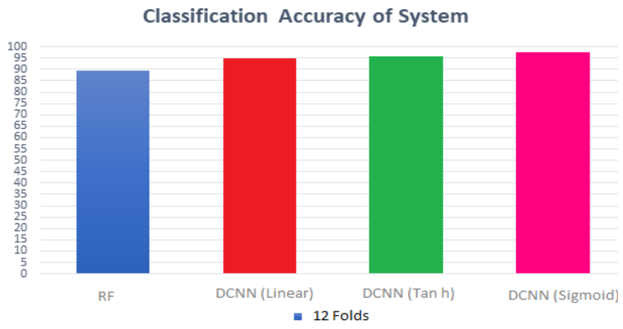


Figure 4. Comparative analysis of accuracy for RF, DCNN(Linear), DCNN(Sigmoid) and DCNN(Tan h) using 12 Fold cross validation.

VII. Conclusion

- The maximum accuracy with which problem of classification of fake vs. real identities on social media can be solved is 95 percent and it is achieved by DCNN with Sigmoid activation function.
- Using Sigmoid activation function (DCNN), gave an increase of 5.8 % in accuracy as compared to RF.
- Also, sigmoid activation function (DCNN) gave an improved accuracy of 0.10 % and 0.20 % respectively as compared to Linear activation function and Tan h activation function.
- The performance of given system varies with dataset used for it.
- We found that the classification accuracy of system increases as the number of folds used in system increases.

References

[1] Estee Van Der Walt and Jan Eloff, "Using Machine Learning to Detect Fake Identities: Bots vs Humans," IEEE, 2018

[2] Indira Sen et. al. "Worth its Weight in Likes: Towards Detecting Fake Likes on Instagram," ACM,2018.

[3] Cao Xiao, David Freeman and Theodore Hwa, "Detecting Clusters of Fake Accounts in Online Social Networks," ACM, 2015.

[4] C. Fuller, D. Biros and R. Wilson "Decision Support for Determining Veracity via Linguistic based Cues," ELSEVIER, 2009.

[5] S. Peddinti, K. Ross and J. Cappos "Mining Anonymity: Identifying Sensitive Accounts on Twitter," ARXIV, 2016.

[6] B. Viswanath et. al. "Towards Detecting Anomolous User Behaviour in Online Social Networks," USENIX, 2014.

[7] J. Dickerson,V. Kagan and V. Subhramanian,"Using Sentiment to Detect Bots on Twitter: Are Humans more Opinionated than Bots?," IEEE,2014

[8] Surendra Sedhai and Aixin Sun, "Semi-Supervised Spam Detection in Twitter Stream," IEEE, 2018.

[9] Ikram et. al., "Combating Fraud in Online Social Networks: Detecting Stealthy Facebook Like Farms," ARXIV, 2016.

[10] R. Oentaryo et. al. "On Profiling Bots in Social Media," ARXIV, 2016.

[11] Priyanka Shahane, Deipali Gore "A Survey on Classification Techniques to Determine Fake vs. Real Identities on Social Media Platforms,"IJRDT, 2018.

[12] Priyanka Shahane, Sneha Kasbe, Rohini Kasar, M.P. Navle "Ontology Based Information Retrieval System Using Multiple Queries For Academic Library,"IRJET, 2018.

[13] Manuel Egele, Gianluca Stringhini, Christopher Kruegel, Giovanni Vigna "Towards Detecting Compromised Accounts on Social Networks," IEEE, 2017.