To Detect Fake Identities in Twitter using Machine Learning Models.

Pedada Chetanroop
Student
p.chetanroop0509@gmail.com

Ruthwik Preetham
Student
rutwikritu@gmail.com

Dr. S Gowri
Associate Professor
gowri.it@sathyabama.ac.in

Abstract: **Social bots are considered to be the most popular type of spamming like spamming malware links, produce fake news, spread rumours to manipulate public opinion. Recently large-scale social bots have been created and are wide spread on social which have a bad impact on public and internet users' safety in all social media platforms. Bot detection aims to distinguish bots from humans to aid understanding the news or opinions. In recent times, classification of bots in social media have become more as they are populated everywhere. In this paper, we propose a decision tree classifier and a deep learning method to classify bots and humans in Twitter using Twitter API. This proposed model uses what an account has tweeted and cross reference against a bag of words model. These methods are unique that applies deep learning concepts to classification. Using real world data from twitter shows the validity of the model we proposed.**

Keywords: Social Media, Twitter, decision trees, Neural networks, accuracy metrics.

Ⅰ. INTRODUCTION

Bots are automated software that scrape or spam, interact with human on a social media platform. Users often tend to use social bots to manipulate opinions, rapid spam of rumours and also rate a product with fake reviews. According to twitter, over 15% accounts are bots. As the social media evolve, the issues caused by bots are obvious.

Bots in twitter control the account via twitter API, to automate actions like follow users, unfollow, tweet, retweet, and reply to direct messages. Few bots in this media are bots that help the user to schedule a tweet, help the assist by providing valuable insights over API. Such bots are termed as hybrid bots. For these specific reasons researchers are in design of advanced methods to classify social media bots, hence bot detection is a very valuable research problem.

The challenges in bot detection have been identified by many teams in the past. Common methods used in classifying are: Graph based, crowdsourcing and machine learning.

Graph based uses social graphs of social networks to get the network information and links across multiple accounts to detect bot activities. The crowdsourcing involves expert annotator to identify, evaluate and classify accounts. The machine learning involves training large datasets with multiple algorithms and statistical methods to distinguish human or computer led activities across social media platforms.

Ⅱ. OBJECTIVES

- Survey current strategies on Machine learning and deep learning to classify bots.
- Classify twitter accounts from human and bots
- The authenticity of people constantly questioning if it is a fake or not, this classifier authenticates and solves the problem.

Ⅲ. Methodology

There are papers focusing on different techniques and methods that are commonly targeted to detect social media bots/ fake accounts on twitter media platform.

Following are the methods used:

- Graph Based Approach
- Crowdsourcing
- Machine Learning Algorithms

Graph based approach uses nodes and samples of benign nodes for classification. Three social media based methods are employed namely trust propagation, graph clusters, and graph metrics.

Crowdsourcing approach is a detection method that involves an annotator to identify patterns across profiles or the tweets shared by human and social accounts. The main aim of human is here to find the differences in bots and human accounts.

This method involves a lot of guesses from the annotators.

Machine Learning techniques uses a large amount of datasets that have many features. Using these models, patterns on feature of twitter accounts can be found and the probability that those accounts being humans or bots can be calculated.

Ⅳ. LITERATURE SURVEY

A. Graph based Detection

Malicious accounts: dark of the social networks.

K. S. Adewole, N. B. Anuar, A. Kamsin, K. D. Varathan, and S. A. Razak

In this paper, the authors proposed three methods to detect accounts. The first method is on trust propagation, to evaluate the relationship that exists between two nodes are strong or weak based on trust. The next method is clustering, in which the nodes of social graph are clustered with similar characteristics. The last method is based on properties, where probability, scale free graph structure are calculated.

Random walk based fake account detection in online social networks.

J. Jia, B. Wang, and N. Z. Gong

The authors proposed a Sybil detection method that employs a random walk based method in an undirected graph. This method calculates scores into two classes: legitimate and Sybil. The authors assumed the graph satisfies the property, for which two linked nodes tend to share the same label. The Sybil nodes with score 1 and the legitimate node is of score 0. With AOC as an evaluation metric, the quality of ranking is 0.96.

B. Crowdsourcing

The darpa twitter bot challenge

V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, and F. Menczer

DARPA held a Twitter bot contest to identify bots that support vaccination discussions on Twitter. All teams used judgement to classify bots using their own implementation techniques.

Automated agents classification with human users.

Z. Gilani, E. Kochmar, and J. Crowcrof

Construction of ground truths datasets with four annotators to classify and labelling twitter profiles into two categories bot or human. They were provided with features to consider, such as date, number of tweets, and favoured tweets, researchers used Cohen's Kappa (*k*) to annotated accounts in twitter.

C. Machine Learning

Pre-processing methods for classifying bots and humans.

M. Kantepe and M. C. Ganiz

The framework for bot classification on twitter was proposed in this paper by Kantepe and Ganiz, who trained Machine Learning algorithms repeatedly after pre-processing and feature extraction. Using twitter API and Apache Spark to collect 1,800 twitter accounts and extract 62 features grouped into three types: User features, tweet features, and periodic features. The highest classifiers among logistic regression, multinomial naïve-bayes, support vector and gradient boosted trees was 86% for gradient boosted trees with f1 score of 83%.

Behavior enhanced deep bot detection in social media

C. Cai, L. Li, and D. Zeng

Behaviour enhanced deep learning model for bot classification implements convolutional neural networks(CNN) with LSTM long-shot-term memory and hidden layers to classify history and behaviour. Used public dataset in experimentation and an additional tweets using Twitter API. The main aim for BeDM is to capture latent features by fusing content and behaviour information. This paper is first to apply deep learning techniques to classify bots.

Botornot: A system to evaluate social bots

C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer

Proposed a system termed BotorNot, which uses random forest classifier to classify social media bots. Their system used 1000 features from 6 classes by which they analysed network, users, friends, temporal, content and sentiment features. The AUC score was 95%. Further in 2014, the authors improved their work, with extended training data, with less AUC than previous results.

Twitter turing test: Identifying social machines

A. Alarifi, M. Alsaleh, and A. Al-Salman

Collected a data of 1.8 million accounts then randomly selected 2000 accounts for the sample manually labelling into human or bot with annotators. Optimal extraction using two

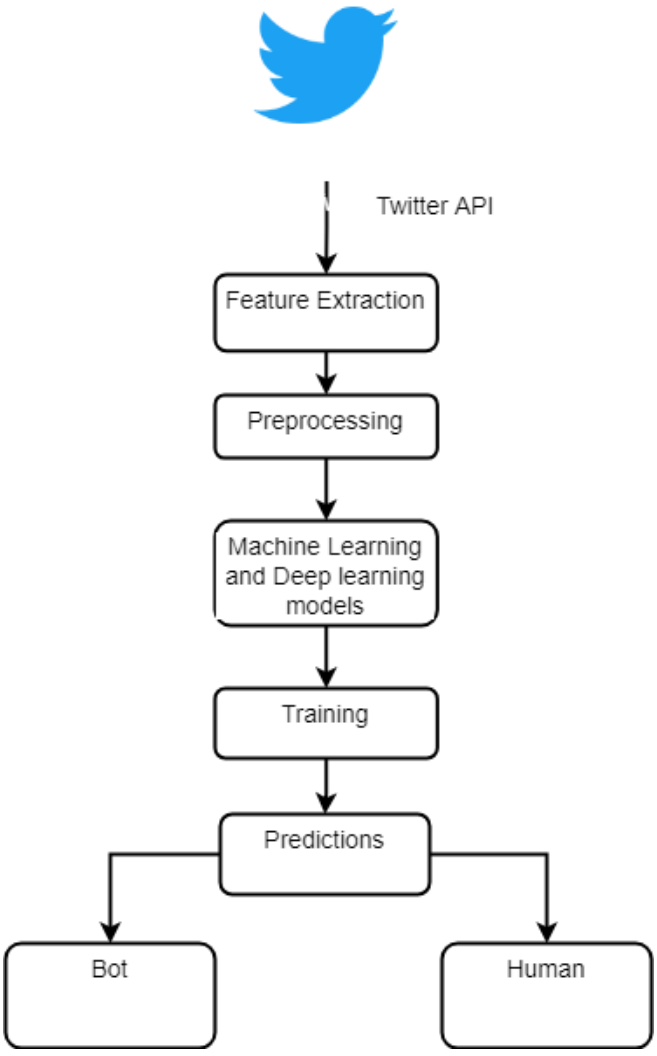
functions. Characteristics based on the correlation method. Finally, they extracted 8 traits to score. A model using a learning algorithm decision tree, Bayesian networks, support vector machines and multi-layer artificial neural network. the best performance was scored by a random forest classifier 88% of accuracy.

Detecting automation of twitter accounts: Are you a human, bot, or cyborg

Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia

Constructed ground truth samples of 6000 accounts over 521,407 accounts using the Twitter API. Eight features extracted and implemented Random forest classifier with tenfold-cross inspection. The confusion matrix to measure the system performance achieved a score of 96%, which is very high compared to other models.

Ⅴ. SYSTEM ARCHITECTURE



Ⅵ. EXISTING SYSTEM

The survey on existing system shows that the three existing methods to classify bots as discussed above are Graph Approach, crowdsourcing and Machine Learning algorithms. Machine learning methods are optimal when compared to the rest of the methods. Most of the papers use tree-based approach, Bayes theorem and random classifier. Random classifier was the most commonly used classifier as it is less complex in tuning and achieving more accurate results. Although it is prone to overfitting, Bayes theorem and Random forest classifiers is better when the data has relatively low number of features. SVM depends on selective kernel, as it reduces the error rate in classification process. Likewise, Neural networks effectiveness depends on sample size; when the samples are large, the model performs well

Ⅶ. PROPOSED SYSTEM

In this section, we propose a machine learning models with data collected using twitter API. The algorithms used are decision tree classifier, neural network.

Model with neural network is found to be more accurate with the training accuracy of 90% and testing accuracy of 96%.

Ⅶ. Classification Techniques

Neural Networks:

A. Dataset

The dataset used is taken using honeypot method from Twitter API. This dataset contains accounts that labelled bot or not including some featues.

B. Metrics

Evaluation criteria are precision, recall and F1 score which are commonly used in bot detection. For each split, trained models with 80% data, cross-validation with 10% and the remaining is used for testing.

C. Performance

In the experiments, we conducted a series of tests with different parameter and layers with activation functions. The experiment results are shown below:

Number of hidden layers*	**Output layer activation function**	**Precision**	**Recall**	**F1**
1	Sigmoid	74.06	75.1	73.2
2	Sigmoid	86.21	84.63	84.61
2	Softmax	68.76	67.33	67.80
3	Sigmoid	88.12	86.90	**86.54**

*activation function in hidden layers is relu

Decision Tree Classifier:

The dataset and evaluation metrics are same as previous model (neural network). A tree can be categorized into two variables: direction and fiendishness. Choices are made based on data sharing.

Performance:

The Area under the curve for training is 95.6 and for testing is 0.93 when trained with decision tree classifier. The metrics are given below:

Training accuracy	**Testing accuracy**	**Precision**	**Recall**	**F1**
88.24	87.85	91.11	83.69	97.725

Ⅷ. Conclusion

In this paper we proposed two methods to classify accounts in social media accounts by bots or human. The results show the performance and effectiveness with high f1 scores. In future work, we can plan to study other social media platforms with behaviour modelling for bot classification.

References:

[1] M. Kantepe and M. C. Ganiz, "Preprocessing framework for Twitter bot detection," 2017 International Conference on Computer Science and Engineering (UBMK), 2017, pp. 630-634, doi: 10.1109/UBMK.2017.8093483.

[2] J. Jia, B. Wang and N. Z. Gong, "Random Walk Based Fake Account Detection in Online Social Networks," 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), 2017, pp. 273-284, doi: 10.1109/DSN.2017.55.

[3] V. S. Subrahmanian et al., "The DARPA Twitter Bot Challenge," in Computer, vol. 49, no. 6, pp. 38-46, June 2016, doi: 10.1109/MC.2016.183.

[4] Z. Gilani, E. Kochmar and J. Crowcroft, "Classification of Twitter Accounts into Automated Agents and Human Users," 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2017, pp. 489-496.

[5] C. Cai, L. Li and D. Zengi, "Behavior enhanced deep bot detection in social media," 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), 2017, pp. 128-130, doi10.1109/ISI.2017.8004887.

[6] Z. Chu, S. Gianvecchio, H. Wang and S. Jajodia, "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?," in IEEE Transactions on Dependable and Secure Computing, vol. 9, no. 6, pp. 811-824, Nov.-Dec. 2012, doi: 10.1109/TDSC.2012.75.