**Identifying Social Bot Accounts on Twitter**

Lili Xu, NYU Tandon School of Engineering Brooklyn, USA email

Li Wang, NYU Tandon School of Engineering Brooklyn, USA email

Ruowen Tan, NYU Tandon School of Engineering Brooklyn, USA rt1095@nyu.edu

**Introduction**

Nowadays, social bots have been widely used on the internet. Social bots can perform spamming, such as constant messaging account on twitch (streaming site); Social bots can make an account famous, such as tons of bot followers on Twitter; Social bots can also mislead public opinion, such as adding many likes/dislikes on Youtube. Social bot exists because they are fairly easy to be controlled by a program, thus giving a lot of problem on the internet.

Fortunately, because those social bot accounts are created and managed by computer program, they tend to have non-human account behaviors. In this paper, we are going to explore how to identify social bot on Twitter by using Machine Learning Algorithms.

**Motivation**

Most of the social bots are existing on the web silently. Very few people know their existence. Our team want to find out how powerful those social bots can be. For example, if we run a program to check Donald Trump’s twitter, we find many bot followers retweeting his post, that can be very interesting and meaningful.

Also we want to apply the knowledge we just learned in class. Using while learning can provide the best understanding. Twitter.com provide very rich API for data harvesting people like us. With such organized data, we can easily run Machine Learning Algorithms to train our Classifier.

**Related Work**

Ferrara, Emilio, et al. "The rise of social bots." *arXiv preprint arXiv:1407.5225* (2014).

Lee, Kyumin, et al. "Who will retweet this?: Automatically identifying and engaging strangers on twitter to spread information." *Proceedings of the 19th international conference on Intelligent User Interfaces*. ACM, 2014.

**Data**

* Bot accounts

We purchased 1000 USA followers online for our Twitter account @Mikupa55. The followers of this account are our training set for bot users.

Listing 50 examples:

@MarieLaureClerg

@QcKucyk

@TimmyJymmy

@cristinashatto

@Moresse1

@sexyscos

@Juan\_SANC2

@MiArsenio

@filimontesfay

@CesarRico350z

@ingridsimioni

@errika2110

@Jane\_Weston

@alin\_ferdinand

@NickiDiamond1

@sergei89kor

@ARocketToDani

@CheyzGledis

@joseflores98

@filhote0071

@KlarinTab

@aiesan

@Donlucho1708

@JustinB\_LOV

@G47Ronald

@stevenjohnson09

@brunaoliverd

@Prinszz1

@ArthurPedroti

@lolodurival

@BossLady51612

@OldOrxan

@RhymesWithMagic

@Zgmagik

@Ever3109

@gggiiia

@romeus321

@LJMENINAMOLECA

@HayLesleyann

@el\_capitan\_mo

@yngvare

@ruth\_salvad

@HoangViet365

@heywardbrianmcl

@denizrobin7

@paola\_con

@GeorgYanaGeO1

@ailoveny

@sherlyyen97

@UnicoHijo78

* Non-Bot accounts

Those accounts are harvested on Twitter.com manually. We have verified those account’s activity, they all look like active human user.

@bonkurasugi

@gunt3001

@\_namori\_

@profnote

@asuka111

@Ely\_eee

@Miv4t\_

@EaterNY

@superdaniela

@jennygzhang

@alanna

@ssdai

@dz

@Babylonian

@AlolanMeowth

@shabazzybaz

@adultmomband

@Waxmice

@jk\_rowling

@DaleBeran

@KnowledgeAli@adriancrosby371@TrentVessel7@LeeRecioy@WaryJT2@sondra\_barlow@yezhang86@CherieKuang@Jacklynyan@kaliedascope786@Ziyansha@wenqing\_yao@sobeksull@theakkoolaidman@cellsmile@JinghaoLu

@Hanna\_Denea

@KevinChinARO

@gabbileal1

@kelseyylittle

@Ann\_Kimberlyyy

@kaitlynbharris

@britlyn\_leigh

@lauren\_harnear

@\_longlyghosts

@jonathanporte19

@jakeborge

@sterdw

@elw119

@Bigdaou

@ubaldohcamposjr

@RyanMolwele

@LeblancDidi

@ClarkLacy17

@lucidtrip24

@selenagomez

@zaynmalik

@LiamPayne

**Algorithms**

As Instructed in class, we firstly take Naive Bayes Classifier and Decision Tree Classifier into consideration. What’s more, support vector machines and basic regression model (Logistic Regression) will also be used.