Twitter Bot Detection using Machine Learning

Ruchitha M. Shanmugha Sundar, Chirayu Desai
Northeastern University
{midigarahallishanm.r, desai.ch}@husky.neu.edu

1 The Problem

In recent times social media has become an indispensable part of our daily social communication and networking. There are approximately 2.5 billion users on Facebook and twitter combined. It's estimated that nearly 50% Twitter activity comes from "bots" [1]: automated accounts that advertise products, post spam, influence public opinion and more.

These bots can have drastic political, social and emotional impacts. For example bots can be used to sway the public opinion with respect to their political choices. They are also used in cyber-bullying that results in emotional trauma to people. We aim to come up with Machine Learning solutions that can detect bots on twitter.

2 Related Works and Challenges

A successful Machine Learning solution will help in removing such bots from twitter or flagging them as such to the users. Initial works by Yang et al., focused on detecting bots, which look to mimic human behavior and spread unwelcome advertising and malware [2]. They used features like URLs per tweet and ratio of friends to followers .Cresci et al. worked on identifying fake followers and social bots[3][4]. Lee et al. came up with solutions that used features like the fact that bots are usually created at the same time, and many more. They implemented the Random Forest classifier with boosting and bagging[5].

Identifying twitter bots can be quite challenging on a social media platform as there is no way to fully identify what a bot looks like. Unlike other social media platforms, twitter allows automation and semi-automation in their platform. This just makes identification of bots a gruel-some task. Twitter being a social media has varied number of data points which makes feature selection for model construction a difficult task. Lastly, What if the bots that we are trying to identify are intelligent?

3 Proposed Solution

We are looking to use the data-sets available at https://botometer.iuni.iu.edu/bot-repository/datasets.html for our project. Our plan involves starting with a Naive Bayes or similar simple classifier to to come up with a baseline supervised learning model with limited features, then we will look to explore not just textual/word based features but also implied/computable features like URLs/tweets, variance in tweet rate, etc. and build other types of Classifiers and Regressors. The next step will be working towards a solution using Neural Networks and then Deep Learning if time permits.

4 Milestones

Table 1: Project Milestones

Weeks	Tasks
1-3	1) Data Selection, Prepossessing, Cleaning and Transformation. 2) Selection of limited set of features and Implementation of Baseline Model.
4-6	1) Derivation of implied/calculated features, Selection of complete set of features. 2) Implementation of other classification as well as regression models like Random Forest, Logistic Regression etc. 3) Start with Neural Networks. 4) Analysis, Reflection and Milestone.
7-10	1) Neural Network solution, optimization and Hyper-parameter tuning including Grid Search Cross Validation. 2) Deep Learning Models. 3) Poster and Project Write-Up.

References

[1] Gorwa, Robert. "Twitter has a bot problem and Wikipedia might be the solution". Quartz Media, 2017. https://qz.com/1108092/twitter-has-a-serious-bot-problem-and-wikipedia-mighthave-the-solution/

[2] C. Yang, R. C. Harkreader, and G. Gu, Die Free or Live Hard? Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 318–337. [Online]. Available: https://link.springer.com/content/pdf/10.1007%2F978-3-642-23644-0$_1$7.*pdf*

[3] S. Cresci, R. D. Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "Fame for sale: efficient detection of fake twitter followers," CoRR, vol. abs/1509.04098, 2015. [Online]. Available: http://arxiv.org/abs/1509.04098

[4] S. Cresci, R. D. Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "he paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," CoRR, vol.abs/1701.03017, 2017. [Online]. Available: http://arxiv.org/abs/1701.03017

[5] K. Lee, B. D. Eoff, and J. Caverlee, "Seven months with the devils: a long-term study of content polluters on twitter," in In AAAI Intl Conference on Weblogs and Social Media (ICWSM), 2011.