NAME: AVISHEK MUKHERJEE

EMAIL: amukherjee3@albany.edu

DATE: 2/14/2019

PAPER TITLE: Online Human-Bot Interactions: Detection, Estimation and Characterization

1. What is the main problem or issue that the authors are addressing?

Ans: The paper is concerned mainly with social bots. The experiments are run on Twitter data. As per the paper, social bots are accounts controlled by software, algorithmically generating content and establishing interactions. The authors present a framework to detect social bots on Twitter, they used thousands of features extracted from public data and meta-data about users and their friend network, tweet content, sentiments, network patterns, and activity time-series. The training data is partly enriched with manual annotation and was generated by both humans and social bots. The model yields high accuracy, agreement and is able to detect bots. They estimate that 9%-15% are bots.

They analyzed the flow of content and the interaction with various users or groups of users. They also used K-Means clustering, to cluster various kinds of accounts like spammers, self promoters and account that post content from applications.

1. Provide a short summary of the authors’ approach/argument.

Ans: The author uses the BotOrNot framework from Indiana University’s social network observatory. They collected data from the Twitter API and identified 1150 features and grouped it into 6 different sub-classes.

User-based features: extracted from the Twitter API

Friends Features: They classified the below features of friends/follower by using: retweeting mentioning, being retweeted, and being mentioned.

Network features: network structure contains crucial information regarding flow of information, and characterization of different kinds of communication. They classify three kinds of networks: retweet, mention and hash-tag cooccurrence.

Content and language features: they collect information about length and entropy of tweet text, additionally they also do POS tagging.

Sentiment features: The framework uses tools to extract various sentiments like arousal, valence, dominance scores, happiness score, polarization and strength, emoticon score.

Model Evaluation:

To train the system they used 15K manually verified Twitter bots indentified by honey-pot approach and 16K human accounts. The collection was limited to the 200 public tweets from the user’s timeline and upto 100 most recent public tweets mentioning that user. This yielded a 2.6 million tweets produced by social bots and 3 million tweets produced by humans.

OTS software like scikit-learn library was used to evaluate the model. The ‘Area under the receiver operating characteristics’ with 5-fold cross-validation, and computation of the average AUC score across the folds using Random Forests, AdaBoost, Logistic Regression and Decision Tree Classifiers. The best classification performance of 0.95 AUC was obtained by the Random Forest algorithm. They used Random Forest model trained using 100 estimators and Gini co-efficient to measure the quality of splits.

To obtain an updated evaluation of the accuracy of the model, they constructed an additional, manually-annotated collection of Twitter user accounts.

1. What are the main strengths and/or weaknesses of the approach?

Ans: The results are very impressive at 0.89 accuracy(Annotation), 0.94(Merged) and 0.9-0.94(Mixture at various ratios).

Weaknesses:

1. The study is restricted only to the English Language. But many social bots are active in places like Belgium (Dutch Belgium vs French Belgium) Netherlands, India, Spain etc. And may have effect on swaying election results.
2. The study is restricted to the Twitter API, which has lesser social outreach than Facebook and Instagram. Consequently, they use more complex social bots which were purportedly effective in swaying election results like Brexit, Scotland and United Kingdom etc.
3. The work doesn’t address the ‘sybil’ accounts that pollute online discussion by lending false credibility to their messages.
4. The human data collection of 200 most recent tweets is kind of limited, and may not properly identify the social bots from humans.
5. The study uses ‘scikit-learn’ in-built algorithms, which is kind of not so famous nowadays in the data science arena.
6. Manual annotation of such huge datasets may be erroneous, which were not accounted for.
7. The study ignores the seasonality of the data. For example, social bots are more active during the election season or just before the election season. Or during major social events.
8. Provide at least 1 question regarding the paper that you’d like to address during class discussion.

Ans: I would like to discuss the effectiveness of the datasets, and why there is no mention of popular NLP packages like NLTK or Stanford coreNLP not applied to activities like POS tagging, or in case of Sentiment Analysis. The use of off-the-shelf scikit-learn packages for the machine learning purpose is kind of not favoured, because in 2017 scikit-learn is not the market leading Machine Learning software package.