Name: Hyomin Seo
Date: 18 Nov 2020
Paper Title: Detection of Novel social Bots by Ensembles of Specialized Classifiers
Author Names: Onur Varol, Kai-Cheng Yang, Alessandro Flammin, and else.
Year Published: 2020

Open questions:
Future work on the social dot detection program?

The topic areas covered by the paper are:
The paper introduces and explains the machine learning system that can aid the detection of social bots. Social bots are agents that perform autonomous communications on various social media platforms. Most of the task is to influence a course of discussion and create a strong bias on an opinion. There are several kinds of such bots, including the ones introduced in the paper: traditional spambots, social spam-bots, and fake followers.

Needless to say, such application can heavily affect a debate, if not something more critical, by playing a disproportionate role in spreading repetitive, low credibility, argumentative, and biased posts/articles. Therefore, the availability of tools to identify social bots is still essential for protecting the authenticity and health of the information ecosystem.

The Machine Learning system introduced in the paper aims to detect such social bots on social media platforms by adopting an Ensemble of specialized classifiers - ESC, deployed in the newest version of Barometer. The introduced model is an improved version of the previous attempts to achieve such a goal of detecting social bots.

The previous approaches to this problem were:
There have been attempts on social bot detection methods based on machine learning. Some of the techniques introduced in the paper are the following.

Crowdsourcing
- A practical method for collecting annotated data
- Limited availability of annotated data hinders Crowdsourcing from being a suitable answer

Supervised / Unsupervised Machine Learning
(apart from the improved one introduced in the paper)
- Trained with annotated data set of both social bots and human
- Unsupervised models is a bit more effective in finding coordination among social bots (getting the features overseen by supervised learning)
- The continuously-adaptive characteristics of social bots make this approach quickly obsolete.

Overall, these attempts meet significant performance deterioration when behaviors undetected (unprecedented) from training data are hooked. The annotation dataset essential to such models' training is either extremely limited or expensive due to scalability and user privacy.

Outline the basic new approach or approaches to this problem:

The paper categorizes the data they have collected to conduct the study and feed it to the system.

1. **Dataset:** The training is executed with various labeled datasets gained from Bot Repository. The data is either annotated by human or automated techniques based on several different approaches (measurement), such as account behavior, filters on metadata. The extracted information is then processed and divided into six categories.

2. **Categories:** Metadata from the accounts and friends, Retweet/mention networks, Temporal features, Content information, Sentiment

3. **In Domain Cross-Validation**: The in-domain cross-validation process is quoted below since this is the most concise way to explain the method.

   "We use Random Forest classifiers with 100 decision trees (similar to the baseline model described in § 3.1). The fraction of trees outputting a positive label is calibrated using Platt's scaling [31] and binarized with a threshold of 0.5. We use 5-fold cross-validation for in-domain classification; for consistency, we split training and test samples into cross-domain cases as well, reporting average precision and recall."

   The result from the in-domain cross-validation shows that the social bots mostly yield conspicuous lower cross-domain scores, with lower recall. In contrast, social accounts deliver consistent left-skewed score results across the dataset. This distinguishable difference detected in the two versions paves the way for improving generalization.

4. **Proposed Method**:

   Training on both accounts of humans and social bots: knowing that human accounts are considerably more consistent than that of the bots, the models can be trained to distinguish the difference between the two account

   Building specialized models for distinct bot classes: since different bot classes have different sets of informative features, each bot class has t be considered separately for greater accuracy and adaptability.

5. **In domain/ Cross-domain performance.**

   In domain: ESC can detect bots with reasonable accuracy in the classic (in-domain) scenario.

   In In-Domain, ESC is proven to detect bots with reliably accurate in-domain scenarios, meaning that it can identify impurities within the comprehensive broad training data. The presented ESC shows a more robust analysis of errors, meaning that the incorrect labels do not hinder the overall process; it might only confuse classifier subsets. ESC is capable of detecting bots with reasonable accuracy in the classic (in-domain) scenario. This advanced aspect of the ESC model will let it be accountable for presenting mislabeled accounts impairing the traditional machine learning model.

Cross-Domain: In cross-domain, the ESC's generalization performs even better than the current version of the Barometer.

In this set of experiments, some datasets are held out in the training phase and are then used as cross-domain test cases*: Cresci-stock, gilani-17, cresci-robust, kaiser-1, kaiser-2, and kaiser-3.* The result shows that the cross-domain performance is more dataset-sensitive (dataset used for training and testing) than in the in-domain.

6. **Model adaptation**

The traditional machine learning model with obsolete bot classes and many classifiers is not apt for learning a new domain. The presented ESC is improved on the aspect, learning new bots through a new classifier rather quickly. The process happens from scratch with not necessarily erasing the old-bot classifiers but preserved. This feature lets the ESC use fewer labeled examples to train a new specialized classifier when novel types of bots are observed in the wild.

7. **Conclusion**

The paper empirically demonstrates that the proposed approach generalizes better than a monolithic classifier and is more robust to mislabeled training examples. The proposed architecture is highly modular as each specialized classifier works independently, so one can substitute any part with different models as needed. Also, by including additional technical classifiers when new annotated datasets become available, the system can learn about new domains efficiently, since then only fewer annotated examples are necessary.

Critical assumptions made include:
Even though the paper employs various methods to perform this prediction, it is admitted that the entire logic is highly dependent on the validity and availability of data. Aside from that, the reason for the study's motivation and the evaluation is built upon mostly acceptable and general assumptions.

The performance of the techniques discussed in the paper was measured in what manner:
The data selection for the study was mainly from approved platforms. The model evaluation was flawless and verified, with a highly detailed technical layout and testing the model on different platforms as well. Similar to the paper discussed previously (New york city tenant help with machine learning), this system also provides an efficient aid to the current dilemma, expediting the process of improving a defect of previous attempts.

New background techniques are used in the paper:
Social bot: agents that perform autonomous communications on various social media platforms, where most of the task is to influence a course of discussion and create a strong bias on an opinion.
Platt's scaling: is an algorithm to solve the aforementioned problem. It produces probability estimates.
In-Domain and Cross-Domain: explained in the method section.

I rate and justify the value of this paper as:

The paper points out a convenient and prevalent problem in the current, digitally carried-information driven society. Its analysis of the current/ previous method to alleviate such a problem is well documented, clearly pointing out what has to be improved. The model suggested in the paper, as far as it is explained, does satisfy all of the needs of a social-bot detection machine learning model, especially in the sense that it is quickly adaptable and apt at learning new classifiers with minimum effort. The practical use of this model on the various central social platforms may successfully encourage more clean, unbiased, and unaffected discussion, which can result in a healthier conclusion for critical matters.