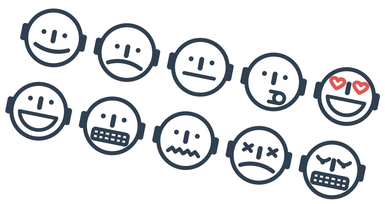
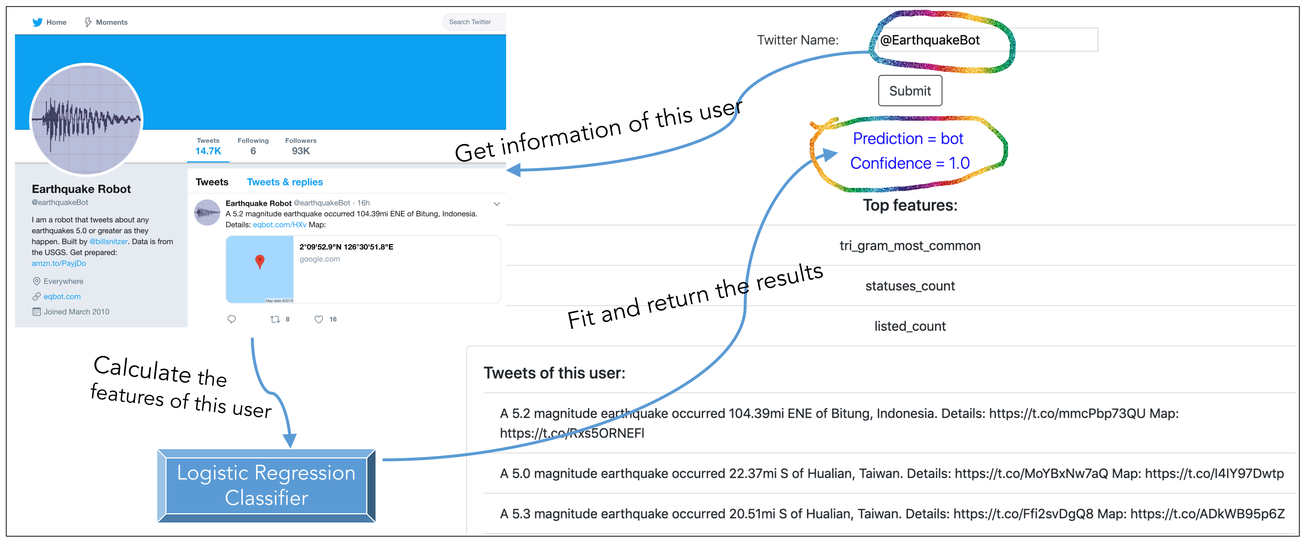
Using machine learning to spot a Twitter Bot

Deyuan Chen, Southern University of Science and Technology
Chunjiang Li, University of Posts and Telecommunications
Chenguang Tang, University of Posts and Telecommunications



Introduction

Twitter bots can automatically perform actions such as tweeting, re-tweeting, or direct messaging other accounts. There are proper usages of bots, however, there are also a lot of improper usages such as violating user privacy, spamming or spreading fake news. So, we delved into detecting twitter bots to prevent malicious acts in the future, which can also be applied into other social media platforms.



Conclusions

We observed that many bots tweet the same content as other users. What's more, there is often no biography, or indeed a photo, associated with bot Twitter accounts.

Interestingly, we also found that Twitter has brought in more stringent policies regarding automation on the platform.

Data and Methods

Training data size:

	Count
Bot	8841
Human	4585
Total	13426

We considered three classifiers. Firstly, we extracted a few features with our data and used the ***Logistic Regression classifier*** to fit the model. Then we calculated the accuracy with ***cross-validation*** and compared the accuracy with two additional classifiers: ***Multi-layer Perceptron*** and ***Random Forest***. Finally, we decided to use ***Logistic Regression*** because of its outstanding performance.

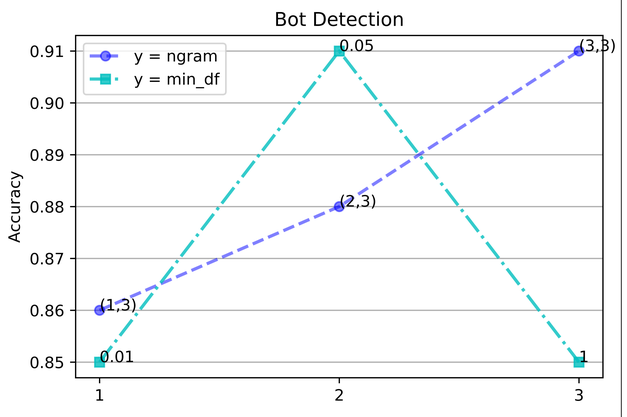
We mainly used two types of features to classify bots. One type is the attributes of twitter users, such as the followers count, verified or not. Another type is based on the text analysis of user tweets.

Results

	F1	Precision	Recall
Logistic Regression	0.91	0.91	0.91
Multi-layer Perceptron	0.84	0.84	0.84
Random Forest	0.89	0.89	0.89

The most predictive features of bots were:

- **Most common trigram**, bots tend to tweet the same contents.
- **Default profile**, indicates that user has not altered the background of their profile. A high percentage of bots use default profile.
- **Statuses count**, the number of tweets (including retweets) issued by the user. Since many bots are used to spread fake news or something, bots have a higher statuses count.



Notes:

1. ngram (min_n, max_n): an *n-gram* is a contiguous sequence of *n* items from a sentence. Here all *n-grams* with lower boundary *min_n* and upper boundary *max_n* will be extracted.
2. min_df: when building the vocabulary, we ignore terms that have a document frequency strictly lower than the given threshold. Correspondingly, there is a parameter named *max_df* which is used to ignore terms with high frequency.

Limitations

The features we choose are limited, many of which are relevant to the contents of the tweets. Also, the training dataset is not big enough, so the the parameters we chose may not be optimal. Therefore, we decide to explore more features and train more data to improve the classifier.

Related Work

Kudugunta, S., & Ferrara, E. (2018). Deep neural networks for bot detection. Information Sciences, 467, 312-322.

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.

Scan to view our repository!
github.com/tapilab/elevate-osna-bots