DARTMOUTH

Institute for Security Technology and Society at Dartmouth

Sockpuppet Research

Sockpuppet accounts refer to multiple accounts controlled by a single individual. Sockpuppets continue to be a major problem in online social network sites as well as review systems on ecommerce platforms. Sockpuppetry may be used to manipulate online discussions with the aim of swaying public opinion or engage in review fraud to increase a business's revenue. With the ease with which users and organizations engage in this malicious and manipulative activity, more research into sockpuppet detection is crucial for preventing their widespread harmful effects. To this end, Dartmouth has led two research projects in sockpuppet detection for the previously mentioned contexts.

Sockpuppets in Online Discussions

As online discussion communities are one of the primary targets of sockpuppet attacks, this Dartmouth-Stanford project focuses on identifying, characterizing, and predicting sockpuppetry in 9 online news and entertainment communities. From a set of 3,656 sockpuppets comprising 1,623 sockpuppet groups, analysis showed that sockpuppets have a variety of distinguishing traits and types, which were utilized in the development of automated tools in sockpuppet detection.

Specifically, sockpuppets are often characterized by a higher similarity to each other than to other users in measures such as average word length, personal pronouns, and number of function words. In addition, sockpuppets use "I" and "you" more often than ordinary users, as well as writing shorter sentences. In terms of interaction, sockpuppets start fewer discussions and post more in existing discussions, participate in discussions with more controversial topics, and interact with each other more if they are in a pair.

Furthermore, sockpuppets vary in their deceptiveness and methods of deception. For instance, a pair of sockpuppets can pretend to be two different individuals (pretenders) or two accounts that an individual uses without masquerading (non-pretenders). The former group tends to have very different display names, while the latter group tends to have virtually identical display names. Sockpuppet pairs may also interact with each other differently, such as being a supporter (positive agreement score), non-supporter (0 agreement score), or dissenter (negative agreement score). Across all communities, 60% of the sockpuppets are non-supporters, while 30% are supporters. Only 10% were dissenters.

"We tested the STARS attack on 7 recent review fraud detectors and found that all of them can be severely compromised by the STARS attack on all 4 datasets that we tested — from Amazon, Epinions, Bitcoin Alpha, and Bitcoin OTC exchanges."

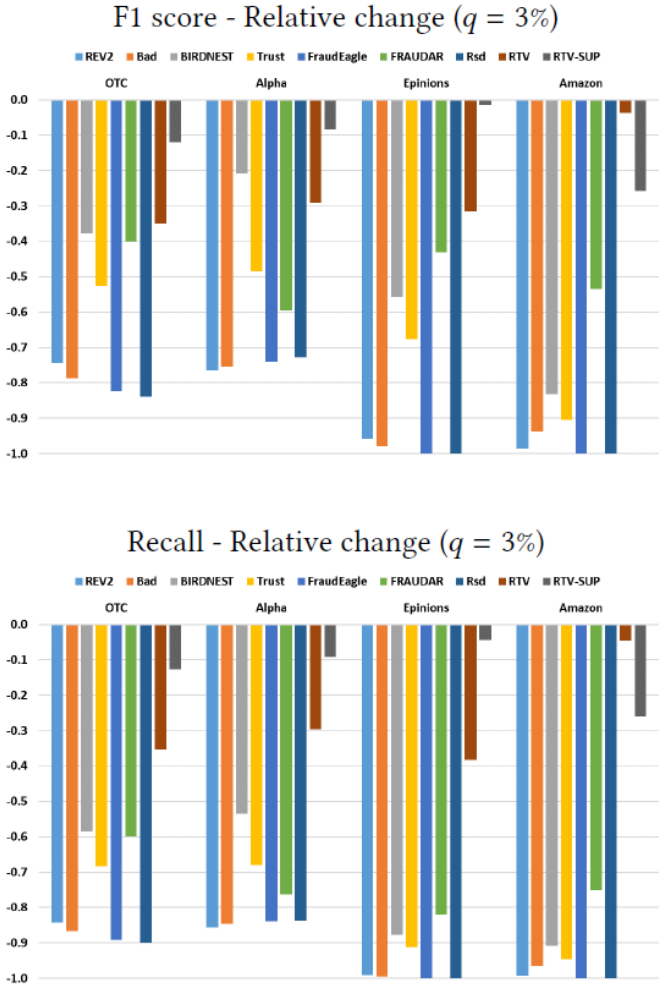
Lastly, two classifiers were developed using activity features, community features, and post features in detecting if an account is a sockpuppet and if two accounts are sockpuppet pairs. Performance evaluation amounted to a 0.68 AUC for the former and a 0.91 AUC for the latter. This work received an honorable mention for best paper of the 2017 World Wide Web Conference.

STARS

Sockpuppets are likewise a problem in online ecommerce review systems, such as engaging in review fraud. The Sockpuppet-based Targeted Attack on Reviewing Systems (STARS) is a Dartmouth project that seeks to understand how attackers might compromise review fraud detectors. STARS was designed as an attack that a fraudster might use to avoid flagging by review detection systems in order to manipulate the star rating of a product. STARS works by having each

account assign the maximum possible score to a target product, then generating a number of other reviews that fit a normal distribution to existing reviews for other randomly chosen products. This way, STARS creates a haystack of random products with the one target product review being the needle. Evaluation of STARS showed that it severely compromises both F1 score and recall.

In order to mitigate the impact STARS can have, Dartmouth also developed the RTV (Reviewer-Trusted-Verified) algorithm, which indicate the three possible statuses a reviewer can have. RTV leverages the use of "verified" reviewers present on certain platforms. RTV also proposes the notion of "trusted" reviewers, who are reviewers that a platform pays to write honest reviews of products. After testing, RTV was found to outperform all other algorithms in all cases for F1 score and recall, as well as obtaining relatively high precision in some cases. RTV is therefore a highly reliable fraud detection system, as long as the cost trusted reviewers are fulfilled.



An example F1 and Recall scores for the RTV algorithm

Additional Information

References

1. Srijan Kumar, Justin Cheng, Jure Leskovec, and V.S. Subrahmanian. "An army of me: Sockpuppets in online discussion communities." In *Proceedings of the 26th International Conference on World Wide Web*, pp. 857-866. International World Wide Web Conferences Steering Committee, 2017.
2. Rui Liu, Runze Liu, Adrea Pugliese, and V.S. Subrahmanian. "STARS: Defending against Sockpuppet-based Targeted Attacks on Reviewing Systems." Under review.

Video

https://www.cs.dartmouth.edu/~dsail/projects/sockpuppet/videos/

Presentation

https://www.cs.dartmouth.edu/~dsail/projects/sockpuppet/slides/

PARTICIPANTS

Lead: V.S. Subrahmanian

Justin Cheng, Srijan Kumar, Jure Leskovec, Rui Liu, Runze Liu, Andrea Pugliese.

DARTMOUTH

Security and Artificial Intelligence Laboratory