Detecting Trending Topics in Cybersecurity Forum Discussions

Jack Hughes, Seth Aycock, Andrew Caines, Paula Buttery, Alice Hutchings, *University of Cambridge*
jack.hughes@cl.cam.ac.uk

1. Abstract

We present a lightweight method for identifying currently trending terms in relation to a known prior of terms, using **a weighted log-odds ratio with an informative prior**. We apply this method to a dataset of posts from an English-language underground hacking forum, spanning over ten years of activity, with posts containing misspellings, orthographic variation, acronyms, and slang. Our statistical approach supports analysis of linguistic change and discussion topics over time, **without a requirement to train a topic model for each time interval for analysis**. We evaluate the approach by comparing the results to TF-IDF using the discounted cumulative gain metric with human annotations, finding our method outperforms TF-IDF on information retrieval.

2. Dataset

Using forum posts from an underground hacking forum, part of CrimeBB
Available from the Cambridge Cybercrime Centre
https://www.cambridgecybercrime.uk/process.html

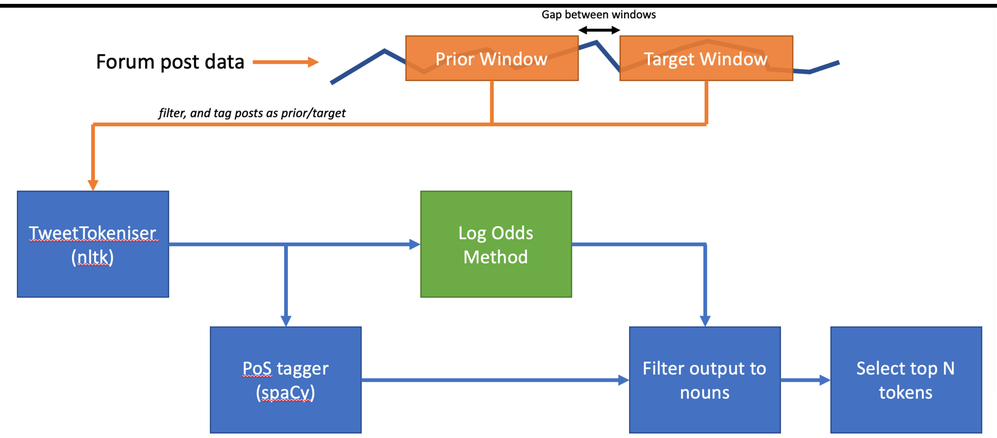
Noise includes:
Orthographic variation: "rat" – animal, or Remote Access Trojan?
Acronyms: nhs (National Health Service), btc (Bitcoin)
Misspellings: ransomeware
Slang: idk, tf

User1	Ransomware infects hospitals all over UK: `link`
User2	anyone think they made some money from this?
User1	They might of done but idk they'll get caught eventually, it's stupid to commit crimes like this
User3	Who tf targets hospitals for ransomeware
User1	I dont believe they actually went for the nhs.. the ransom would be more $$$ lol
User4	I looked up a few btc addresses and can confirm they made money

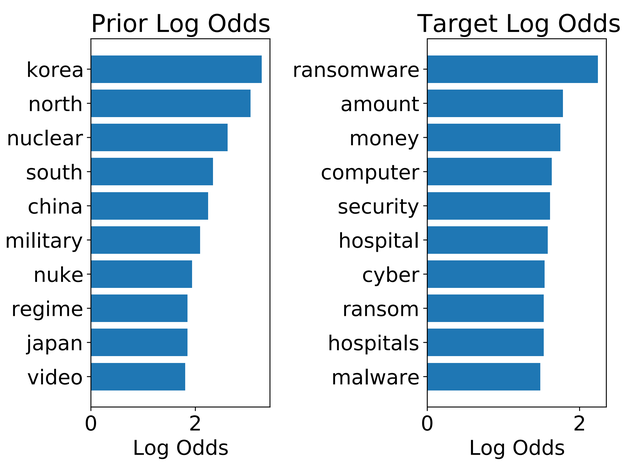
3. Method

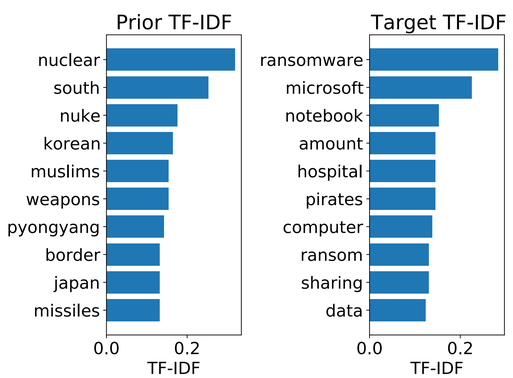
Prior window used to "learn" the distribution of vocabulary, for filtering commonly occurring slang

Detects "bursty" trending topics, depending on the window sizes and gap selected



4. Results





Log odds approach detects bursty trending topics, focused on terms related to the WannaCry ransomware event affecting the National Health Service (NHS)

TF-IDF detected a broader mix of topics

5. Evaluation

Used annotations of a known event (WannaCry) and a randomly selected date range, to the measure saliency of detected topics, using Discounted Cumulative Gain from information retrieval. For the WannaCry event, our log-odds tool scored 0.979 compared to TF-IDF of 0.877. For the random event, the log-odds tool scored 0.978 compared to TF-IDF of 0.753.