[25] The relationship between currency exchange and illicit behaviour in an underground forum

Gilberto Atondo Siu (University of Cambridge), Ben Collier (University of Cambridge) and Alice Hutchings (University of Cambridge).

Reviews Phase 1

Review 1

Time: Jun 04, 08:55

Overall evaluation: 1 (weak accept)

This paper presents a measurement study of the types of crimes seen in HackForums over a period of 10 years. It also studies to the type of cryptocurrency used (offered and wanted) in this form. The authors design a method that uses natural language processing to account for the aforementioned characteristics of the measurement.

Strengths: + Relevant problem + Interesting research questions + Large dataset + Annotation extraction not straightforward

Weaknesses: - The crime type classifier is not described - Use of heuristics in the evaluation

Comments:

The motivation of the problem, as well as the contribution, are clear. The paper addresses a relevant problem. I have enjoyed reading this paper as it offers a very interesting overview of how an underground forum has evolved in terms of types of crimes. The paper is well written and easy to follow.

One of the main limitations of the paper is that the classifier itself is not well described. In particular, I was left to wonder what are the actual features used by the classifier. Section IV.C jumps from the Baseline Design (evaluation) to Statistical Model (classifier) without fully explaining what are the features of the classifier. Authors position that they use NLP, but it is unclear how NLP is used. Offering details here is important as the evaluation shows that using "Post content" (*) has a negligible improvement w.r.t. baseline 2. Perhaps this negative result is due to the design of the feature extraction process and the way NLP is used. (* I presume "Post content" means that authors leverage NLP over the content of the posts).

Similarly, it is unclear how authors make use of heuristics to build the baseline 2. Authors explain that "the second baseline proposed assigned the crime type conditionally on the bulletin board name as shown in Table II". What does it mean conditionally in this context? Is the board name hard-coded to the type of crime?

The quality of the paper would improve considerably if authors were to shed light on the type of features used and if they were to offer additional details on how the heuristics are extracted.

I found the analysis of the currency exchange measurement and the qualitative analysis very insightful with valuable nuanced conclusions (e.g., minor currencies, opportunistic cash out). This type of analysis will trigger interesting discussions at WACCO.

Reviewer's confidence: 5 (expert)

Review 2

Time: Jun 10, 16:28

Overall evaluation: 3 (strong accept)

The paper describes a classification of the HackForums underground forum posts through a machine learning classifier and an analysis of the classified post from a currency exchange viewpoint.

It could add to the research to an estimation of the use of these currencies outside the forum to understand if also some fluctuations were due to some external factors or social popularity at the time.

Overall the paper is very well structured, and every step (from data collection, labeling, training, and analysis) is clearly explained. The paper was a great read.

Reviewer's confidence: 3 (medium)

Rebuttal Letter

Title: Follow the money: The relationship between currency exchange and illicit behaviour in an underground forum

This document outlines briefly a tentative plan of changes for this paper. My objective for future revisions is to address the weaknesses highlighted by the reviewers as follows:

I will improve the classifier description to specify the features used. I will also include further information about how NLP was used for pre-processing the data before running the classification models. This information will refer to tokenisation, removing stop-words, lemmatisation and vectorising words by using TF-IDF.

I will also offer additional details about the heuristics used in the baseline design. This information will include more insights obtained through the annotation process that allowed us identify apparent relationships between certain crime types and specific bulletin board titles.

Finally, I will explore further the factors that could have had an impact on the price fluctuations of the relevant cryptocurrencies mentioned in our paper.

Reviews Phase 2

Review 3

Time: Jun 29, 02:27

Overall evaluation: -1 (weak reject)

This paper performs a classification over a dataset of illicit forum posts. They demonstrate that taking advantage of the existing sub-forum structure from which posts are collected is an important feature for classification. The paper looks at subsets of posts related to different types of illicit businesses, and then identifies which observed transactions are advertised in which currencies. From this, the paper draws conclusions about favored currencies over time. The paper also presents a set of representative posts about the facebook effort towards cryptocurrencies to demonstrate how appear in this context in their early stages.

In the comparison of table 4 and table 5, it would potentially be interesting to know how strong of a signal the sub-forums are in isolation. If posts are categorized just by forum without post content, what precision and recall are observed?

The distribution of labeled training data presented figure 1 makes me worried about the limited sample used for training for the less common types of crime. If there were only a few samples, it seems harder to trust the accuracy of the final pool used to generate e.g figures 5 and 6.

Figures 1-9 would be more useful if they were normalized by the number of posts over time. It seems likely that as presented many of the spikes may correspond with fluctuating levels of user activity. It would be more useful, and the point that these figures seem to aim to illustrate, is the relative interest between the different payment mechanisms.

I wasn't really sure what to take away from discussion of the facebook currency. None of the discussion presented seemed to provide any intuition or useful direction, and in general seemed to represent broad common-knowledge speculation.

Reviewer's confidence: 3 (medium)

Review 4

Time: Jul 02, 18:33

Overall evaluation: 1 (weak accept)

=== meta review ===

The reviewers suggest that the paper can be accepted after some minor changes. Besides individual comments from the reviews, authors should at least address the following issues:

1. Provide a numerical break-out of the training data set used. The bulk of section 4, per the indication in figure 1 of extremely small training sets for many of the topics. Since the measures are calculated using a 70/30 split over 4000 posts, but for the trailing categories, especially spam, identity theft, and eWhoring, that means validation will be on something like 10 annotated items. Authors should expand on this limitation, and also provide further details on the oversampling process made and how this partially thwarts the limitations.

2. Provide further details on the feature set used to build the classifiers, and consider adding a new experiment using only the forum names to analyse it in isolation

3. Figures 1 to 7 should be provided on relative terms would make discussion of them much easier.

Reviewer's confidence: 5 (expert)