The Language of Legal and Illegal Activity on DarkNet

Anonymous ACL submission

Abstract

1 Introduction

Omri

– darknet has much illegal activity (https://www.profwoodward.org/2016/02/how-much-of-tor-is-used-for-illegal.html). people use it because it's easier to be anonymous, hard to track etc.

– scalably monitoring activity in darknet uses NLP tools, but little is known about what characteristics the text in Tor has, and how well do off-the-shelf NLP tools do on this domain.

– X et al. (2017) published a corpus and looked into text classification on texts from Tor, and Y et al. (2019) looked into how influential criminal sites are.

– We investigate how legal and illegal activity taken from DarkNet is different, comparing to a clearnet website with similar content as a control condition.

– We also explore methods for classifying texts from DarkNet into legal and illegal. This is important both for understanding whether these two types are different in terms of their text and in what ways, and as a practical tool.

– A bit on the experiments and results.

2 Related Work

Elior

Detecting illegal activities in the Web The detection of illegal activities in the Web is sometimes derived from a more general topic classification. For example, Biryukov et al. (2014) used the software Mallet (McCallum, 2002) and the web service uClassify (Kågström et al., 2013) for a classification of the content of Tor hidden services into 18 categories, which allows the distinction between illegal or contreversial content on one hand and human rights or freedom of speech content on the other hand. Graczyk and Kinningham (2015) combined unsupervised feature selection and an SVM classfier for the classification of drug sales in an anonymous marketplace. However, although the detection of illegal activities can be easily deduced in some cases, the legal status of a given product can change (Graczyk and Kinningham, 2015) and a given topic could cover both legal and illegal content. For example, in the recent work of Avarikioti et al. (2018) on Tor content clasification, in most of the categories both legal and illegal content appear.

Some works have directly addressed a specific type of illegality and a particular communication context. Morris and Hirst (2012) have used an SVM classification to identify sexual predators in chatting message systems. The model includes both lexical features, including emoticons and behavioral features that correspond to conversational patterns. Another example is the detection of pedophile activity in peer-to-peeer networks (Latapy et al., 2013) where a predefined list of keywords was used to detect child-pornography queries.

3 Datasets

DUTA (Nabki et al., 2017)

Cleaning. Elior + Daniel

4 Domain Analysis

4.1 Distances between Domains

Leshem

4.2 NER and Wikification

In order to analyze the named entities in each domain and the differences between them we used a

Table 1: Percentage of Wikifiable Named Entites per Domain

	Percentage
Legal Onion	50.8 ±2.31
Illegal Onion	32.5 ±1.35
eBay	38.6 ±2

"wikification" technique, searching for said entities in public datasets such as Wikipedia.

Using SpaCy's named entity recognition, we first found all the named entities in each site in the datasets. After finding the named entites we searched for relevant Wikipedia entries for each named entity using the DBpedia Ontology API. For each domain we counted the total number of named entites and what percentage of them had corresponding Wikipedia articles.

According to our results (Table 1) the wikification percentages of eBay sites and illegal Onion sites are comparable and relatively low. However, sites selling legal drugs on Onion have a much higher wikification percentage.

Presumably the named entities in Onion sites selling legal drugs are more easily found in public databases such as Wikipedia because they are mainly well known names for legal pharmaceuticals. However, in both illegal Onion and eBay sites, the list of named entities include many nicknames for illicit drugs and paraphernalia. These nicknames are usually not well known by the general public and are therefore less likely to be found on Wikipedia or other public databases.

In addition to the differences in wikification percentages between the domains, we found that SpaCy had trouble correctly identifying named entities in both Onion and eBay sites. There were a fair number of false positives (words and phrases that were found by SpaCy but were not actually named entities), especially in illegal Onion sites. We believe the informal language used in our datasets makes it harder for SpaCy to function optimally in this capacity.

These findings lead us to believe that the popular tools used today for named entity recognition and analysis are not ideal for processing informal language on the internet, especially language dealing in illicit activities.

5 Classification Experiments

Daniel

5.1 Results

Legal vs. Illegal

Legal Onion vs. Ebay

6 Discussion

The legal and illegal are pretty distant, which is evident in a few ways: word distribution is different, NER and Wikification work less well for illegal. This has practical implications: we need to adapt our tools to deal with illegal Tor data.

Looking at specific sentences, we see that it's hard distinguishing them based on the identity of the words, which means that looking at the wordforms is a very poor solution for tackling this. However, using modern text classification, they can be distinguished in a 72%.

Looking at how different types of language influence results: given that replacing all words with their POS tags gives the same performance, this tells us their syntax is different as well.

Methological: Legal and illegal in Onion are distinct enough to be considered different domains (as distant as Legal and Ebay). Therefore, Tor could be a good testbed for working on legal and illegal classification.

7 Conclusion

References

Georgia Avarikioti, Roman Brunner, Aggelos Kiayias, Roger Wattenhofer, and Dionysis Zindros. 2018. Structure and content of the visible Darknet. ArXiv:1811.01348v2 [cs:CY].

Alex Biryukov, Ivan Pustogarov, Fabrice Thill, and Ralf-Philipp Weinmann. 2014. Content and popularity analysis of Tor hidden services. In *2014 IEEE 23rd International Conference on Distributional Computing Systems Workshops (ICDCSW)*, pages 188–193.

Michael Graczyk and Kevin Kinningham. 2015. Automatic product categorization for anonymous marketplaces.

J. Kågström, R. Karlsson, and E. Kågström. 2013. uClassify web services.

Matthieu Latapy, Clemence Magnien, and Raphael Fournier. 2013. Quantifying pedophile activity in a large P2P system. *Information Processing & Management*, 49(1):248–263.

Andrew Katchites McCallum. 2002. MALLET: A machine learning for language toolkit.

Colin Morris and Graeme Hirst. 2012. Identifying sexual predators by SVM classification with lexical and behavioral features. In *CLEF (Online Working Notes/Labs/Workshop)*, volume 12, page 29.

Mhd Wesam Al Nabki, Eduardo Fidalgo, Enrique Alegre, and Ivan de Paz. 2017. Classifying illegal activities on Tor network based on web textual contents. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 35–43.