The Language of Legal and Illegal Activity on Tor

Anonymous ACL submission

Abstract

1 Introduction

2 Related Work

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. **?** have used an SVM classification to identify sexual predators in chatting message systems. The model includes both lexical and behavioral features.

3 Datasets

DUTA (Nabki et al., 2017)

Cleaning.

4 Domain Analysis

4.1 Distances between Domains

Leshem

4.2 NER and Wikification

Dan: experimental setup, results and

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.

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

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.