Leveraging Machine Learning to Improve Unwanted Resource Filtering

Sruti Bhagavatula Christopher Dunn
Chris Kanich Minaxi Gupta Brian Ziebart
University of Illinois at Chicago Indiana University Bloomington

UIC Department of Computer Science
UNIVERSITY OF ILLINOIS AT CHICAGO
COLLEGE OF ENGINEERING

SCHOOL OF INFORMATICS AND COMPUTING
INDIANA UNIVERSITY
Bloomington

INTRODUCTION

Advertisements have long been considered extremely annoying and intrusive to most users on the web. Besides just serving ad content, ad networks use the advertisements to retrieve information about the user and their browsing habits. Additionally, many ads are being used as “malvertisements" that serve malware to unsuspecting victims. Hence, advertisement blocking is definitely a non-trivial defense against the security risks present on the internet of today. We propose an alternate approach to manual ad blockers which leverages machine learning to bootstrap a superior classifier for ad blocking with less human intervention. This poster is based on the work to be published in the 2014 ACM Workshop of Artificial Intelligence and Security [1].

APPROACH

Our goal is to overcome the problem of having to manually update the EasyList filter list [2]. To achieve this goal, we use a supervised learning approach to bootstrap the process of learning whether a URL is ad-related by considering EasyList filters as oracles. We train on URLs matched by “old” EasyList filters and test on URLs matched by the “new” EasyList filters. We compute 2 main types of specific accuracy measures in addition to average accuracy.

1. Baseline accuracy = No. of positively classified URLs matched by both filter lists / No. of URLs matched by both lists

2. New-Ad accuracy = No. of positively classified URLs matched by new but not old list / No. of URLs matched by new but not old list

Features

We use a total of 12 features for classification that are described by the feature sets below:

A. Ad-related keywords (2 features)
B. Lexical features (2 features)
C. Related to the original page (2 features)
D. Size and dimensions in URL (2 features)
E. In an iframe container (1 feature)
F. Proportion of external requested resources (3 features)

RESULTS

Table 1 shows the results of the different classifiers on the dataset according to the 3 different performance measures: accuracy, precision and false positive rate. Additionally, figure 1 shows the baseline and new-ad accuracy that the different classifiers achieve on the dataset.

Comparison of Classifiers			
Classification method	**Avg. Accuracy**	**Precision**	**FP Rate**
Naïve Bayes	89.50%	89.09%	14.3%
SVM (linear)	92.10%	92.36%	7.4%
SVM (poly kernel)	90.51%	90.56%	7.34%
SVM (rbk kernel)	92.18%	92.43%	7.7%
l2-Reg. Logistic Regression	92.44%	92.43%	7.5%
k-Nearest Neighbors	97.55%	98.6%	1.3%

Table 1: A comparison of the average accuracy, precision and false positive rate over 6-fold CV of various machine learning approaches to ad classification. The logistic regression was performed with a threshold of 0.5.

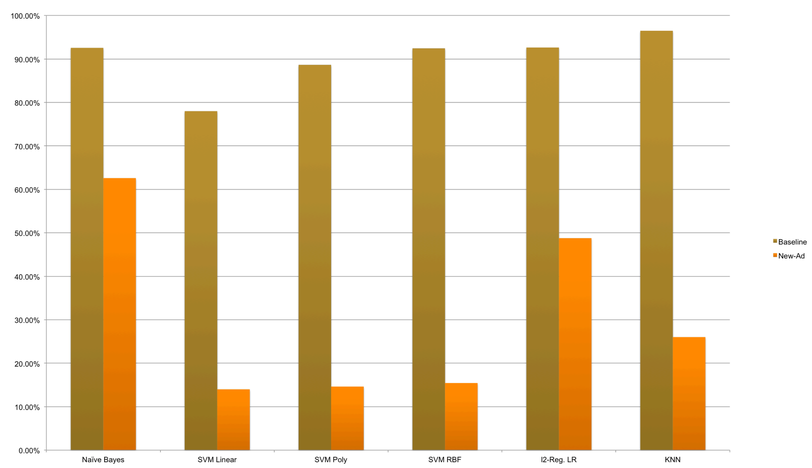

Figure 1: Baseline accuracy and new-ad accuracy for the six different machine learning approaches. New-ad accuracy is out of a total of 123 examples.

kNN Performance

Table 2 shows the results of the kNN classifier after removing the feature sets A-F.

Comparison of Feature sets			
Feature Set (f)	**Accuracy**	**Baseline Acc.**	**New-Ad Acc.**
A	90.21%	81.82%	48.78%
B	97.42%	95.20%	48.78%
C	96.82%	95.16%	34.96%
D	95.94%	93.38%	27.64%
E	96.22%	94.21%	21.95%
F	76.88%	57.50%	9.76%

Table 2: Average accuracy, baseline accuracy, and new-ad accuracy without each feature set (f).

CONCLUSION

Using the kNN supervised classification scheme, we were able to achieve high accuracy and low false positive rate on advertisement URLs by training on historical advertisement regex matches as well as classify ads that were not identified by the new filters [1].

ACKNOWLEDGEMENTS

Special thanks to the Grace Hopper Celebration of Women in Computing for this opportunity and to the anonymous reviewers for their helpful feedback.

REFERENCES

1. S. Bhagavatula, C. Dunn, C. Kanich, M. Gupta, B. Ziebart. *Leveraging Machine Learning to Improve Unwanted Resource Filtering*. Proceedings of the 2014 ACM Workshop on Artificial Intelligence and Security, Nov 2014, Scottsdale, AZ.
2. Adblock Plus community. EasyList. https://easylist.adblockplus.org, 2014.