Abstract

Spam, phishing, and fraud detection are security applications that impact most people. Challenges for building spam, phishing, and fraud detection models include difficulty in obtaining annotated data, increased computational complexity for robust detection methods, annotation errors, and changes in the underlying data distribution. We address the above challenges as follows. Clustering and active learning are combined to make efficient use of annotated data, yielding state of the art spam detection performance using only 10% of the annotated data employed by previously published methods. Social Network Analysis (SNA) reputation features for mail transfer agents are introduced to evaluate paths from sender to receiver, increasing the detection rate by 70% (with the same false positive rate) for state of the art spam detection. Random projections with boosting achieve state of the art spam detection with a 75% reduction in computational cost for message classification. The Randomized Hough Transform - Support Vector Machine updates training set annotations, increasing the (precision, recall) F measure by 9.3% compared to a state of the art method for handling adversarial noise. Spectral clustering of URL n-grams and transductive semi-supervised learning are used to increase the detection rate by 100% (doubling the detection rate with the same false positive rate) for state of the art phishing detection under adversarial modification of message text. Reputation and similarity features are used to enhance the ability to withstand changes in underlying data distributions, producing a 13.5% increase in cost savings for state of the art fraud detection. Future research possibilities include the application of these methods to identify deception in social media channels.