Adversarial Learning in Credit Card Fraud

Stephen Adams, Tyler Cody, Don Brown, Peter Beling
University of Virginia, System Engineering Department

MOTIVATION

- In the United States in 2013 alone, credit card fraud cost companies almost $7.1 billion dollars. Given these enormous costs, fraud detection and classification has become a very active area of research in machine learning and data mining domains.
- Although the power of machine learning techniques for fraud detection has greatly increased over the past decades, the incentives for fraudsters to circumvent and adapt to these classification algorithms has also grown.

GOALS AND OBJECTIVES

- The goal of this project is to create a highly scalable and efficient pipeline for detecting fraud in an adversarial environment.
- Develop simulated fraud attacks.
- Develop simulated lender defenses.
- Develop environment in which attack-defense pairs can be efficiently tested.

NEED & INDUSTRIAL RELEVANCE

- Project has the potential to make explicit modeling of adversary behavior the norm in credit fraud detection and to lead to revamped strategies for model rebuilding and rollouts.
- While the proposed work is set in the context of credit card fraud detection, concepts of adversarial learning can be applied to many other domains that involve similar adversarial pairings, including for example medical and tax fraud.

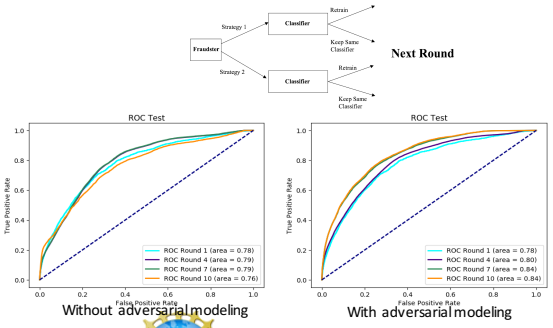
APPROACH (RESEARCH METHODS)

- Explicit modeling of adversary behavior in a manner that fuses game theory and machine learning.
- **Adversarial classifier reverse engineering.**
- Model active experimentation by the fraudsters.
- Adversary shifts the distribution of its transactions to oversample effective attacks.
- When should the lender rebuild its classifier? Randomize its strategy? Employ disinformation?

DELIVERABLES/OUTCOMES

- Engineered features for fraud data set, with implementation on AWS
- Definition and implementation of adversary strategies
- Definition and implementation of defense strategies
- Analysis of ROCs under attack-defense combinations
- Conclusions and final report

PRELIMINARY RESULTS



Without adversarial modeling

With adversarial modeling

www.nsfcvdi.org

A National Science Foundation Industry/ University Cooperative Research Center