Adversarial Data Mining: Big Data Meets Cyber Security

Murat Kantarcioglu
University of Texas at Dallas
Department of Computer Science
Richardson, TX
muratk@utdallas.edu

Bowei Xi
Purdue University
Department of Statistics
West Lafayette, IN
xbw@purdue.edu

ABSTRACT

As more and more cyber security incident data ranging from systems logs to vulnerability scan results are collected, manually analyzing these collected data to detect important cyber security events become impossible. Hence, data mining techniques are becoming an essential tool for real-world cyber security applications. For example, a report from Gartner [4] claims that "Information security is becoming a big data analytics problem, where massive amounts of data will be correlated, analyzed and mined for meaningful patterns". Of course, data mining/analytics is a means to an end where the ultimate goal is to provide cyber security analysts with prioritized actionable insights derived from big data. This raises the question, can we directly apply existing techniques to cyber security applications?

One of the most important differences between data mining for cyber security and many other data mining applications is the existence of malicious adversaries that continuously adapt their behavior to hide their actions and to make the data mining models ineffective. Unfortunately, traditional data mining techniques are insufficient to handle such adversarial problems directly. The adversaries adapt to the data miner's reactions, and data mining algorithms constructed based on a training dataset degrades quickly. To address these concerns, over the last couple of years new and novel data mining techniques which is more resilient to such adversarial behavior are being developed in machine learning and data mining community. We believe that lessons learned as a part of this research direction would be beneficial for cyber security researchers who are increasingly applying machine learning and data mining techniques in practice.

To give an overview of recent developments in adversarial data mining, in this three hour long tutorial, we introduce the foundations, the techniques, and the applications of adversarial data mining to cyber security applications. We first introduce various approaches proposed in the past to defend against active adversaries, such as a minimax approach to minimize the worst case error through a zero-sum game. We then discuss a game theoretic framework to model the sequential actions of the adversary and the data miner, while both parties try to maximize their utilities. We also introduce a modified support vector machine method and a relevance vector machine method to defend against active adversaries. Intrusion detection and malware detection are two important application areas for adversarial data mining models that will be discussed in details during the tutorial. Finally, we discuss some practical guidelines on how to use adversarial data mining ideas in generic cyber security applications and how to leverage existing big data management tools for building data mining algorithms for cyber security.

CCS'16 October 24-28, 2016, Vienna, Austria

ACM ISBN 978-1-4503-4139-4/16/10.

DOI: http://dx.doi.org/10.1145/2976749.2976753

Keywords

Big Data Analytics for Cyber Security; Adversarial Data Mining

1. OVERVIEW

Due to existence of malicious attackers that try to evade detection, data analytics techniques for cyber security need to be resilient against the changing behaviors of the adversaries, and are able to quickly detect previously unknown new attack instances. To develop resilient data analytics techniques, recently, various adversarial data mining techniques (including our own prior research) are being developed to counter adversarial behaviors.

This tutorial is intended for an audience who are interested in applying data mining techniques to cyber security challenges. Basic prior knowledge about statistics and basic data mining knowledge, such as standard supervised learning and unsupervised learning techniques, would be useful.

The three hour long tutorial is divided into three parts. In the first 45 minutes, we introduce the foundations of adversarial data mining. The second part (approximately 45 minutes) is devoted to the novel techniques used in adversarial data mining. The third part (approximately 90 minutes) is focused on the application areas and practical tools. Below are the detailed descriptions of the foundations, the techniques, and the applications of adversarial data mining that is discussed during the tutorial.

1.1 Foundations of Adversarial Data Mining

One intuitive approach to fight the adversary is to let the classifier adapt to the adversary's actions, either manually or automatically. Such a classifier was proposed in [2]. They define the problem as a game between two cost-sensitive opponents: a naive Bayes classifier and an adversary playing optimal strategies. Their adversary-aware algorithm made predictions according to the class that maximizes the conditional utility. The problem is that this becomes a never-ending game between the classifier and the adversary.

Furthermore, in online learning such as [1], a strategic game has been used to learn a concept in real time or make a prediction for the near future by seeing instances one at a time. As the first part of our tutorial, we briefly discuss these various theoretical foundations and their implications for cyber security tasks.

1.2 Adversarial Data Mining Techniques

In the second part of this tutorial, we discuss in detail different approaches developed for adversarial data mining. We discuss the two-player game model of the problem [3], where the adversary tries to maximize its return and the data miner tries to minimize the misclassification cost. We examine the classifier performance and the adversary's behavior at such an equilibrium point. In an adversarial environment where the underlying populations are constantly being transformed, the classifier's initial success has little impact on its long term performance, even if the classifier is constantly adjusted to fight the new threats. A classifier's performance at the equilibrium is a more accurate measure for its eventual effectiveness. We then discuss the implications of these results on how to choose a set of effective features for adversarial data mining applications.

We then discuss various attack models against data mining techniques: a *free-range* attack model that permits arbitrary data corruption and a *restrained* attack model that anticipates more realistic attacks that a reasonable adversary would devise under penalties. We introduce optimal support vector machine (SVM) [9, 8] learning strategies against the two attack models. We also discuss other resilient learning techniques developed to counter adversarial adaptation and behavior.

1.3 Cyber Security Applications

We discuss intrusion detection and mobile malware detection as two important application areas for adversarial data mining. Especially, we discuss how some of these data mining techniques could be implemented on recent big data management systems such as Spark.

Intrusion Detection: Detecting the presence of information flows and flow changes is a fundamental problem in complex and multi-scale data networks with important applications in different aspects of network design. In network security and intrusion detection, for example, adversaries can hide their identity by launching the so-called stepping-stone attack where compromised hosts are used as stepping stones to relay attacking commands. For the network defender, the problem is to detect stepping-stone connections—a flow of information—and trace such connections to the attacker. In other forms of intrusion, the attacker may hijack the normal traffic and create a hidden pathway. The problem here is to detect the presence of irregular traffic flow or the change of flow dynamics from the normal traffic flows. Adversarial data mining techniques provide an alternative approach for intrusion detection. In this part, we discuss how adversarial data mining theory and tools could be applied to intrusion detection. In addition, we survey the previous data mining approaches used for intrusion detection (e.g., [7]) and how they could be updated in the light of recent adversarial data mining developments.

Mobile Malware Detection: As smart phone software become more complex, more malware programs are created to exploit specific weaknesses in smart phone software. Smart phones end users all together constitute a large portion of the powerful mobile network. Having access to the enormous amount of personal information on this network is a great incentive for adversaries to attack the smart phone mobile world.

Malicious activities on mobile phones are often carried out through lightweight applications, scrupulously avoiding detection while leaving little trace for malware analysis. Over the years many malware detection techniques have been proposed. These techniques can be roughly divided into two groups: static analysis and dynamic analysis. Static analysis techniques (e.g., [5]) discover implications of unusual program activities directly from the source code. Although static analysis is a critical component in program analysis, its ability to cope with highly dynamic malware is usually unsatisfactory. Dynamic analysis (also known as behavioral analysis) (e.g., [6]) identifies security holes by executing a program and closely monitoring its activities. Information such as system calls, network access, and files and memory modifications is collected from the operating system at run-time. Since the actual behavior of a program is monitored, threats from dynamic tactics such as obfuscation are not as severe in dynamic analysis. However, dynamic analysis can not guarantee a malicious payload is always activated every time the host program is executed. Meanwhile some dynamic analysis techniques leverage data mining techniques.

We discuss the weaknesses of traditional data mining techniques against mobile malware that randomizes its action. For example, data mining based malware detection technique could easily degrade in performance, if a malware that captures private phone calls randomizes the amount of data transferred. Adversarial data mining models with carefully chosen utility functions of the malware developers offer a more effective approach. We discuss how those approaches could be applied for mobile malware detection.

2. ACKNOWLEDGMENTS

The research reported herein was supported in part by ARO award W911NF-12-1-0558, NIH awards 1R0- 1LM009989, & 1R0- 1HG006844, NSF awards CNS-1111529, CNS-1228198, & CICI-1547324.

3. REFERENCES

[1] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[2] N. N. Dalvi, P. M. Domingos, Mausam, S. K. Sanghai, and D. Verma. Adversarial classification. In *10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004*, pages 99–108, 2004.

[3] M. Kantarcioglu, B. Xi, and C. Clifton. Classifier evaluation and attribute selection against active adversaries. *Data Min. Knowl. Discov.*, 22(1-2):291–335, 2011.

[4] N. MacDonald. Information security is becoming a big data. https://www.gartner.com/doc/1960615/ information-security-big-data-analytics.

[5] A. Moser, C. Kruegel, and E. Kirda. Limits of static analysis for malware detection. In *23rd Annual Computer Security Applications Conference (ACSAC 2007), December 10-14, 2007, Miami Beach, Florida, USA*, pages 421–430, 2007.

[6] A. Shabtai, U. Kanonov, Y. Elovici, C. Glezer, and Y. Weiss. "andromaly": a behavioral malware detection framework for android devices. *J. Intell. Inf. Syst.*, 38(1):161–190, 2012.

[7] R. Sommer and V. Paxson. Outside the closed world: On using machine learning for network intrusion detection. In *31st IEEE Symposium on Security and Privacy, S&P 2010, 16-19 May 2010, Berleley/Oakland, California, USA*, pages 305–316, 2010.

[8] Y. Zhou, M. Kantarcioglu, and B. M. Thuraisingham. Sparse bayesian adversarial learning using relevance vector machine ensembles. In *12th IEEE International Conference on Data Mining, ICDM 2012, Brussels, Belgium, December 10-13, 2012*, pages 1206–1211, 2012.

[9] Y. Zhou, M. Kantarcioglu, B. M. Thuraisingham, and B. Xi. Adversarial support vector machine learning. In *18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12, Beijing, China, August 12-16, 2012*, pages 1059–1067, 2012.