Machine Learning in Cyber Security

Akriti Saxena
Dept of Computer Applications,
Invertis University, Bareilly, India.

Abstract- *Within the ever-growing and quickly increasing field of cyber security, it is nearly impossible to quantify or justify the explanations why cyber security has such an outsized impact.The development of effective techniques has been an urgent demand in the field of the cybersecurity community. Machine learning for cybersecurity has become an issue of great importance recently due to the effectiveness of machine learning and deep learning in cybersecurity issues. Cyber security may be a necessary thought for people and families alike, also for businesses, governments, and academic establishments that operate inside the compass of world network or net. With the facility of Machine Learning, we will advance cyber security landscape.Machine learning techniques have been applied for major challenges in cybersecurity issues like intrusion detection, malware classification and detection, spam detection and phishing detection. Although machine learning cannot automate a complete cybersecurity system, it helps to identify cyber-security threats more efficiently than other software-oriented methodologies, and thus reduces the burden on security analysts. In this paper, we present the current state of art machine learning applications and their potential for cybersecurity.*

Keywords- Cybersecurity, Machine Learning, Spam Detection, Malware Detection.

I. INTRODUCTION

Since the invention of the internet technology, cyberspace has emerged as a central hub for the creation of cyber-attacks. The advances in technologies further facilitate hackers to discover vulnerabilities and to create viruses and malware continuously challenging the cyber security industry. Cyber security refers to the science of technologies, processes, and practices designed to shield networks, devices, programs, and information from attacks, damage, or unauthorized access. Cyber security involves providing secure computing and communicative environment with proper innovations and procedures intended to shield PCs, systems, projects. These frameworks are made out of network security and host security systems with firewalls, anti-virus software, Intrusion detection systems etc. Machine Learning is proven to be capable of solving the most common problems in different domains like image processing, Health informatics applications, physical sciences, Computational Biology, Robotics, Financial prediction, Audio Processing, Medical Diagnostics, Video Processing, and Text Processing. Specifically Machine learning techniques are also applied successfully in the field of cyber security to develop effective solutions. Machine learning has excellent potential for detecting various types of cyber-attacks and thus has become an important tool for the defenders. ESET conducted a survey on ―usage of machine learning for cyber security, in which 80% of the participants believed that Machine Learning will help their organization to detect and respond faster to threats.

II. MACHINE LEARNING TECHNIQUES

2.1 Regression:

In regression, the value of a dependent feature is estimated based on the values of the independent features by learning from the existing data related to past events and such knowledge is used to handle new events. In cybersecurity, Fraud detection can be solved by regression. Once a model is learnt from the past transaction database, based on observed features of the current transactions, it determines fraudulent transactions. Machine learning provides Linear regression, Polynomial regression, Support vector machine, Decision tree, Random forest and other regression methods for regression analysis.

2.2 Classification:

Classification is another extensively used supervisory machine learning task. In cybersecurity, spam detection is successfully implemented by machine learning based classifiers which involves discriminating a given email messages as spam or not. The spam filter models are able to separate spam messages from non-spam messages. Machine learning techniques for classification include Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Decision Tree, and Random Forest Classification. Applicability of the above supervisory machine learning techniques is conditioned based on the availability of large collection of labeled data.

2.3 Clustering:

Both regression and classification are supervised learning models, for which labeled data is essential. Clustering is an unsupervised learning model, which extracts general patterns from the data even when the data is not labeled. Groups of similar events constitute a cluster as they share common features that define a specific(behavior)pattern. In cybersecurity, clustering can be used for forensic analysis, anomaly detection, malware analysis, etc. K-means, KMedoids, DBSCAN clustering are some of the machine learning clustering techniques used in cybersecurity.

III. CYBER SECURITY ISSUES:

The four major areas where Machine Learning algorithms play a crucial role are Intrusion Detection Systems, Malware analysis, Mobile (Android) malware detection and Spam Detection.

3.1 Intrusion Detection:

Whenever secure information compromised by malicious software or policy violations, then Intrusion Detection Systems comes into the picture. Detection of an intrusion can be done in several ways. Broadly the methods are classified into either signature-based or anomaly-based. In the signature-based approach, all packets are compared with the signatures of known malicious threats. In the anomaly-based approach, network traffic is monitored against an established baseline of normality. Saroj Kr. Biswas [1] showed that machine learning based feature selection techniques play an important role in a good intrusion detection system. They applied a combination of feature selection techniques and achieved good results. R. Vinaya Kumar [2] applied traditional machine learning algorithms as baselines for comparisons. Several experiments conducted with different learning rate and hidden layer sizes and achieved Detection Rate of 98.88%. N. Shone [3] A deep learning model was proposed for Network Intrusion Detection System operation with combination of machine learning methods.

3.2 Malware Detection:

Malware is coined from malicious software in short, is a specific type of cyber threat software. Generally it is used for illegal activities like compromising the system by stealing data or bypass access control or cause harm to the host computer. The term malware is broadly used for various types of malicious programs like viruses, Trojan horses, worms, bugs, rootkits, spyware, Ransomware. Each of these malware types contains several families. For example, ransomware can be classified as Charger family, RansomBO family, Svpeng family, Simplocker family, etc. These malicious programs can be embedded in different formats like UNIX ELF (Executable and linkable) files, windows PE files (Portable Executables with .exe). Document-base malware programs can be embedded in .doc, .pdf, files. Malware can also be in the form of extensions and plugins for popular software platforms like web browsers, web frameworks. Mozammel Chowdhury [4] proposed Neural Network-based approach for malware detection. They extracted features from PE headers using the n-gram method and conducted experiments with the extended set of features and achieved 97% accuracy with ANN. Mahmoud Kalash [5] conducted experiment with two well-known datasets 'Malimg' and 'Microsoft malware' for malware detection and reported that they achieved 98.52%, 99.97% accuracy on the two datasets respectively.

3.3 Android Malware Detection:

Android is the most widely used mobile platform and hence highly targeted by the mobile malware creators. As the numbers of android malware types are increasing day by day, it has become more and more challenging to detect and classify mobile malware variants. A large number of attempts are made by the researchers towards mobile malware detection.DroidMat [6] applied k-means clustering algorithm on static features from android apps. Sharma and Dash [7] extracted static features from android apps and achieved good results by applying machine algorithms like Random Forest, Decision Trees.AntiMalDroid [8], Droid Dolphin [9] applied Support Vector Machines on dynamic features extracted from malware apps (logged behavior sequence as features) and achieved good accuracy.

3.4 Spam Detection:

Spam Detection is also one of the major challenges in cybersecurity. Spam is an unsolicited bulk messaging generally used for advertising. Generally, spam indicates email spam, but it could be a message on social networking sites and other blogging platforms also. Spam messages waste a lot of valuable time. Sometimes, users get spam emails that disguised themselves as authentic message from a bank to trap the users. Responding to such spam messages may lead to incur heavy financial losses. Machine learning techniques have been applied by many researchers for spam detection. Muhammad N. Marsono[10]applied the Naïve Bayes classification technique for identification of spam messages among incoming email and achieved good results. James Clark [11] applied the K-NN model for automated email classification problem. Mehul Gupta [12] compared various machine learning and deep learning techniques for SMS spam detection on two different datasets. They compared the performance of eight different classifiers and showed that

CNN Classifier achieved the accuracy of 99.19% and 98.25% for the two datasets.

IV. CONCLUSIONS

Machine Learning approaches are widely applied to solve various types of cybersecurity problems. Advances in the field of machine learning offers promising solutions to cybersecurity issues. But it is important to identify which algorithm is suitable for which application. Multi-Layered approaches are needed to keep the solution resilient against malware attacks and to achieve high detection rates. The selection of a particular model plays a vital role in solving cybersecurity issues. In this paper the authors explored the state of art mechanisms for cyber security problems. The autonomous capabilities of machine learning and algorithms must not be overestimated. The combination of human supervision and Machine learning techniques results in achieving the desired goals of cybersecurity.

REFERENCES

[1] Saroj Kr. Biswas, ―Intrusion Detection Using Machine Learning: A Comparison Study‖, International Journal of Pure and Applied Mathematics, Volume 118 No. 19 2018, 101-114.

[2] R. Vinayakumar, MamounAlazab, (Senior Member, IEEE), K. P. Soman, Prabaharan Poornachandran, AmeerAlNemrat, A.N. Venkatraman, ―Deep Learning Approach for Intelligent Intrusion Detection System‖, IEEE Access, VOLUME 7, 2019, Digital Object Identifier 10.1109/ACCESS.2019.2895334.

[3] N. Shone, V. D. Phai, T. N. Ngoc, Q. Shi, "A deep learning approach to network intrusion detection", IEEE Transactions on Emerging Topics in Computational Intelligence-Feb-2018(41-50).

[4] MozammelChowdhury, AzizurRahman, Rafiqul Islam, ―Protecting Data from Mal-ware Threats using Machine Learning Technique‖, 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA).

[5] Mahmoud Kalash, MrigankRochan, Noman Mohammed, Neil D. B. Bruce, Yang Wang, FarkhundIqbal, ―Malware Classification with Deep Convolutional Neural Net-works‖, 978-1-5386-3662-6/18/$31.00 ©2018 IEEE.

[6] D.-J. Wu, C.-H.Mao, T.-E.Wei, H.-M.Lee, and K.-P. Wu, ―DroidMat: Android mal-ware detection through manifest and API calls tracing,‖ in Proc. 7th Asia Joint Conf. Inf. Security (Asia JCIS), 2012, pp. 62–69.

[7] A. Sharma and S. K. Dash, ―Mining API calls and permissions for Android malware detection,‖ in Cryptology and Network Security. Cham, Switzerland: Springer Int., 2014, pp. 191–205.

[8] M. Zhao, F. Ge, T. Zhang, and Z. Yuan.,‖ An efficient SVMbased malware detection framework for Android,‖ in Communications in Computer and Information Science, vol. 243, Springer, 2011, pp. 158–166.

[9] W.-C. Wu, S.-H. Hung, ―A dynamic Android malware detection framework using big data and machine learning,‖ in Proc. ACM Conf. Res. Adapt. Convergent Syst. (RACS), Towson, MD, USA, 2014, pp. 247–252.

[10] Muhammad N. Marsono, M. Watheq El-Kharashi, FayezGebali, ―Targeting spam control on middleboxes: Spam detection based on layer-3 e-mail content classification‖ Elsevier Computer Networks, 2009.

[11] James Clark, Irena Koprinska, Josiah Poon, ―A Neural Network Based Approach to Automated E-mail Classification‖, Proceedings IEEE/WIC International Conference on Web Intelligence, 0-7695-1932-6, Oct. 2003.

[12] Mehul Gupta, AdityaBakliwal, Shubhangi Agarwal & Pulkit Mehndiratta, ―A Comparative Study of Spam SMS Detection using Machine Learning Classifiers‖, Eleventh International Conference on Contemporary Computing (IC3), 2-4 August, 2018, Noida, India, 978-1-5386-6835- 1/18,2018 IEEE.