USE OF MACHINE LEARNING ALGORITHMS IN CYBER SECURITY

Sheha kothari
Computer Science Dept.
Pacific University
Sneha.kothari@gmail.com

Pariniti Gurjar
Computer Science Dept.
Pacific University
pariniti@gmail.com

Abstract: Artificial intelligence (AI) has made incredible progress, resulting in the most sophisticated software and standalone software. Meanwhile, the cyber domain has become a battleground for access, influence, security and control. This paper will discuss key AI technologies including machine learning in an effort to help understand their role in cyber security and the implications of this new technology. This paper discusses and highlights the different uses of machine learning in cyber security.

Keywords: machine learning, Cyber security, attack detection, artificial intelligence.

I. INTRODUCTION

Technologies like Big Data, Cloud Computing, Artificial Intelligence, etc., have been repeated over and over again in many forums, in many cases without a clear understanding of their importance or their system for solving real problems successfully. AI is the creation of intelligent machines that can learn from experience, allowing them to function

is a very important way in which we can measure the detection and classification of malware.

and react the way one can. These technologies enable computers to be processed to process large amounts of data and to identify patterns and patterns. Mechanical learning methods have been used in many areas of science because of their unique properties such as flexibility, disability, and the ability to adapt quickly to new and unknown challenges. Cyber security is a fast-growing field that needs a lot of attention due to significant advances in social media, cloud and web technology, online banking, the mobile environment, smart grid, etc. Machine learning, the AI branch (Figure 1), is used successfully to solve a small part of the problem. Various computer learning systems have been successfully deployed to address a wide variety of computer security problems.

Machine learning - sometimes referred to as Artificial Intelligence (AI) - is a powerful tool used by cyber security companies. Applied Artificial Intelligence (AI-powered Machine Learning) technology

ISSN (Print):22040595
ISSN (Online): 2203-1731

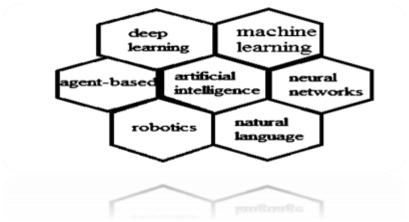

Figure 1: Artificial intelligence branch

II. Machine Learning

Artificial Intelligence [Wan 08] is the field of science that studies the synthesis and analysis of computational agents that act intelligently. Machine learning is a subset of the broader field of Artificial Intelligence. The current applications of AI are mostly restricted to Machine Learning (ML).

Machine Learning and Artificial Intelligence [Mar 18] is more connected to cutting through industries and applications than at any other time recent

memory such as computer power, storage capacity and increased data collection.

Learning machines teach the machine how to answer a question or how to make a decision on your own. It contradicts the traditional system, which requires giving the machine clear instructions to answer certain questions. In fact, every conceivable situation needs to be settled in advance so that it will cover every possible situation. ML can include techniques such as math, statistical preparation, or data mining. ML algorithms try to make decisions about their behaviour and find solutions to problems by dedicating themselves from models based on sample inputs that represent real-life situations.

There are many types of ML & each works differently. If we make the field more general, we can define three main categories of ML (shown in Figure 2): supervised learning, supervised learning and enhanced reading.

A. Supervised Learning:

In supervised learning, the machine is trained using sample data labelled to tell the machine what the data represents. Supervised learning algorithm by inserting inputs defined as P and output defined as Q and algorithms used to create and read map function (f) by output input. The goal of a supervised learning algorithm is to accomplish the map measurement function so that in every new entry (P), a new predicted result (Q) is extracted. In other words, the learning algorithm detects a set of inputs and their corresponding results, and the algorithm learns by measuring its concrete result and relevant results in order to detect errors and the learning model is altered accordingly. Supervised learning algorithms use patterns to predict label values for unlabelled data. This is achieved by division, retreat, prediction, etc.

B. Unsupervised Learning

In unsupervised reading, the machine is trained using labelled data. Uncontrolled reading is where only input data (P) is available without equal output variance. The purpose of unsupervised learning is to model data structure to learn more about details. Algorithms are required to obtain structure, direction and meaning within the data in order to reach a conclusion. These algorithms do not contain any type of historical data to predict the outcome unlike monitored algorithms. That means the machine does not know what the data represents and what responses are expected. The machine will need to find out the patterns and textures of the unlabelled inputs and get the expected result. The division of movie genres on Netflix is an example of unattended reading.

C. Reinforcement Learning

ISSN (Print):22040595
ISSN (Online): 2203-1731

In reinforcement teaching, the machine communicates with its environment to achieve a specific purpose. It is the same with unregulated learning, as the machine is trained using labelled data. However, in reinforcement teaching, the machine receives feedback on the results.

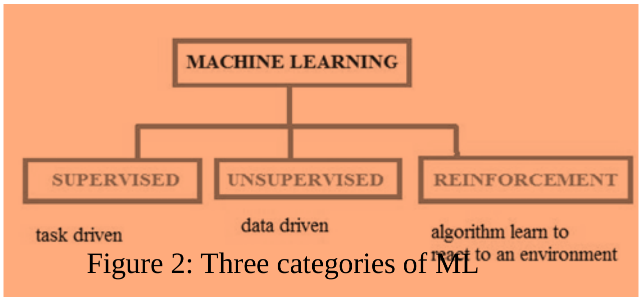

Figure 2: Three categories of ML

III. Cyber Security

Security becomes one of the most important topics in industrial IT and Operational Technology (OT), which is the hardware and software used in the manufacturing environment.

Cyber security is defined as technologies and processes designed to protect computers, computer hardware, software, networks and data from unauthorized access, vulnerability provided by the Internet by cyber criminals, terrorist groups and hackers. Cyber security is related to protecting your internet and network for digital devices and information from unauthorized access and modification. One of the biggest challenges cyber security elements are a fast and flexible environment for security risks. Business a network with mainframes, a customer-server model, a group of closed programs and attacks are severely limited by viruses, worms and Trojan horses which are major cyber threats. The focus was mainly on malwares such as viruses, worms and Trojans with the aim of causing systemic damage. Cyber threats occasionally target computers connected to the Internet.

Artificial Intelligence methods are powerful and flexible; as a result increasing security performance and a better protection system from a growing number of previous cyber threats. Various AI methods can be used in cyber security such as intelligent agent, neural networks, expert system, data mining, machine learning and in-depth learning.

IV. Machine Learning in Cyber Security

Machine learning is an effective tool that can be employed in many areas of information security. There exist some robust anti-phishing algorithms and network intrusion detection systems. Machine learning [Jor 15] can be successfully used for developing authentication systems, evaluating the protocol implementation, assessing the security of human interaction proofs, smart meter data profiling, etc.

Machine learning [Kan 17] has identified an important opportunity in the cyber security industry. New methods of machine learning can greatly improve the accuracy of the threat detection and improve network visibility due to the large number of computer analytics they can handle. They are also announcing a new era of independent response, in which the machine system is smart enough to understand how and when to fight the threats of continuity.

Various electronic learning methods have been used successfully to address a wide range of computer security issues. We will have to separate the three areas where most cyber ML algorithms detect the application: spam detection, malware analysis and intrusion detection.

A. *Spam and phishing detection*

Spam detection and data theft involving sensitive identity theft includes a large set of strategies aimed at minimizing time wastage and risk posed by unwanted emails. Nowadays, unsolicited emails, that is, phishing scams, represent the preferred method by which an attacker sets a precedent within a business

ISSN (Print):22040595
ISSN (Online): 2203-1731

network. Phishing spam emails include malware or links to compromised websites. Spam detection and criminal identity theft are becoming increasingly difficult because of the escape tactics used by attackers over traditional filters. ML methods can improve the process of spam detection.

Filtering spam based on the text content of email messages can be seen as a special case of text sorting, categories are spam and non-spam. Today the most effective spam filters are based on the mathematical principles of Learning Machines. Spam filters using Machine Learning [Bla 08] also train while in use and reduce manual effort while delivering higher filtering accuracy.

Although the work of text-sharing has been extensively researched, its specific e-mail and spam detection system in particular has recently been developed. Some of the first research studies focused on the problem of spam filtering where the Naïve Bayes (NB) was used to address the problem of building spam filters. Naive Bayes is an old machine learning algorithm where we can use all of our feature to find out if they are a bad file or not and use it for a purpose of categories. The NB was named because of its earlier strengths in the field of documentation and because of its ease of use in a cost-effective decision-making framework. Although high performance levels were obtained using only the wording features, it has been observed that by adding additional non-text features and other domain information, filter performance can be greatly improved. Theft of sensitive information is therefore intended to steal sensitive personal information.

Investigators [No 06] identified three main groups of anti-crime information theft methods: investigator (monitoring, content filtering, anti-spam), prevention (authentication, mark management and transformation), and remediation (site reduction, technical).

B. E-mail Spam Filtering

Automatic email segmentation uses mathematical methods or machine learning techniques and aims to create a model or specifically separate spam filtering function from user mail stream. Model construction or separation requires a set of pre-configuration. The process of building a model is called training. Machine learning algorithms have found greater success among all previous techniques employed in the spam filtering process. In fact, the success stories of Gmail, can be attributed to their timely conversion and effective use of Learning Filters to filter not only incoming spam but other abuse such as Denial-of-Service (DoS), virus delivery, and other thoughtful attacks.

C. Malware detection:

Malware detection is a very relevant problem because modern malware can automatically create different novels with the same harmful effects but appear as completely different files. These polymorphic and metamorphic features override traditional methods of identifying a computer-resistant system. Malware can be categorized into several categories depending on its purpose: virus, worm, Trojan, adware, spyware, root kit, backdoor, key logger, Ransom ware and Remote Administration Tools. ML techniques can be used to differentiate malware analysis and place it in the correct malware family.

D. *Intrusion Detection:*

The Intrusion Detection System (IDS) is a security mechanism that monitors computer network activities and reports malicious activities to a network administrator. Intercessors make many attempts to gain access to the network and try to damage the organization's data. Security is therefore a very important factor in any type of organization. Intrusion detection aims to detect illegal activities within a computer or network using Intrusion Detection Systems (IDS). Network IDS is still widely distributed in current

ISSN (Print):22040595
ISSN (Online): 2203-1731

business networks. These systems traditionally relied on known attack patterns, but today's deployments include alternatives for malicious detection, threat detection [Tor 16] and classification based on machine learning. Within the broader access point, there are two issues related to our analysis: bottle detection and Domain Generation Algorithms (DGA). A botnet is a network of infected machines owned by attackers and misused to perform many illegal activities. Botnet detection aims to identify the connection between infected devices within the monitored network and external control and control services. Despite the many research proposals and marketing tools that deal with this threat, there are still many botnet. DGA automatically generates domain names, and is often used by an infected machine to communicate with external (s) by creating hostel names from time to time. They represent a real threat to organizations because, with DGA relying on language processing techniques, it is possible to avoid defending them based on a black domain name list.

V. Conclusions:

Mechanical learning methods are increasingly being used in many systems and are also widely accepted for cyber security, which is why it is important to check which and which class of algorithms can obtain sufficient results. We analyse these methods with three relevant cyber security issues: intrusion detection, malware analysis and spam detection. Machine learning as a technology has exploded widely throughout the cyber space. These decision-making algorithms are known to solve several problems. There are many opportunities for information security using machine learning to deal with various challenges in such a complex domain. Spam detection, virus detection, and surveillance camera surveillance are just a few examples Mechanical learning methods have been used in many areas of science because of their unique properties such as flexibility, disability, and the ability to adapt quickly to new and unknown challenges.

REFERENCES

[1] X. B. Wang, G. Y. Yang, Y. C. Li, D. Liu, (2008) “Review on the application of Artificial Intelligence in Antivirus Detection System”, IEEE Conference on Cybernetics and Intelligent Systems, pp. 506 509

[2] Marty, R. AI and Machine Learning in Cyber Security – Towards Data Science. March 16, 2018, from https://towardsdatascience.com/ai-and-machine-learning-in-cyber-security
Applications of Artificial Intelligence (AI) to Network Security

[3] Kanal, E. (2017, January). Machine Learning in Cybersecurity. Carnegie Mellon University– Software Engineering Institute.March 9, 2018

[4] E. Tyugu. Algorithms and Architectures of Artificial Intelligence. IOS Press. 2007

[5] M. Chandrasekaran, K. Narayanan, and S. Upadhyaya, “Phishing Security Conference, 2006

[6]M. I. Jordan and T. M. Mitchell*, “Machine learning: Trends, perspectives, and prospects,”* Science, 2015

[7]E. Blanzieri and A. Bryl*, “A survey of learning-based techniques of email spam filtering,”* Artificial Intelligence Review, 2008

[8]A. Javaid, Q. Niyaz, W. Sun, and M. Alam, *“A deep learning approach for network intrusion detection system,”* in EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS), 2016

[9] G. Tzortzis and A. Likas, *“Deep belief networks for spam filtering,”* in IEEE International Conference on

ISSN (Print):22040595
ISSN (Online): 2203-1731

Tools with Artificial Intelligence (ICTAI), 2007

[10]A. Khan, B. Baharudin, L. H. Lee, and K. Khan, *“A review of machine learning algorithms for textdocuments classification,”* Journal of advances in information technology, 2010.

[11] P. Torres, C. Catania, S. Garcia, and C. G. Garino, “An analysis of Recurrent Neural Networks for Botnet detection behavior,” in IEEE Biennial Congress of Argentina (ARGENCON), 2016.

ISSN (Print):22040595
ISSN (Online): 2203-1731