The Future of Machine Learning in Cybersecurity

By Dr. May Wang

Machine learning (ML) is a commonly used term across nearly every sector of IT today. And while ML has frequently been used to make sense of big data—to improve business performance and processes and help make predictions—it has also proven priceless in other applications, including cybersecurity. This article will share reasons why ML has risen to such importance in cybersecurity, share some of the challenges of this particular application of the technology and describe the future that machine learning enables.

Why Machine Learning Has Become Vital for Cybersecurity

The need for machine learning has to do with complexity. Many organizations today possess a growing number of Internet of Things (IoT) devices that aren't all known or managed by IT. All data and applications aren't running on-premises, as hybrid and multicloud are the new normal. Users are no longer mostly in the office, as remote work is widely accepted.

Not all that long ago, it was common for enterprises to rely on signature-based detection for malware, static firewall rules for network traffic and access control lists (ACLs) to define security policies. In a world with more devices, in more places than ever, the old ways of detecting potential security risks fail to keep up with the scale, scope and complexity.

Machine learning is all about training models to learn automatically from large amounts of data, and from the learning, a system can then identify trends, spot anomalies, make recommendations and ultimately execute actions. In order to address all the new security challenges that organizations face, there is a clear need for machine learning. Only machine learning can address the increasing number of challenges in cybersecurity: scaling up security solutions, detecting unknown attacks and detecting advanced attacks, including polymorphic malware. Advanced malware can change forms to evade detection, and using a traditional signature-based approach makes it very difficult to detect such advanced attacks. ML turns out to be the best solution to combat it.

What Makes Machine Learning Different in Cybersecurity

Machine learning is well understood and widely deployed across many areas. Among the most popular are image processing for recognition and natural language processing (NLP) to help understand what a human or a piece of text is saying.

Cybersecurity is different from other use cases for machine learning in some respects. Leveraging machine learning in cybersecurity carries its own challenges and requirements. We will discuss three unique challenges for applying ML to cybersecurity and three common but more severe challenges in cybersecurity.

Three Unique Challenges for Applying ML to Cybersecurity

Challenge 1: The much higher accuracy requirements. For example, if you're just doing image processing, and the system mistakes a dog for a cat, that might be annoying but likely doesn't have a life or death impact. If a machine learning system mistakes a fraudulent data packet for a legitimate one that leads to an attack against a hospital and its devices, the impact of the mis-categorization can be severe.

Every day, organizations see large volumes of data packets traverse firewalls. Even if only 0.1% of the data is mis-categorized by machine learning, we can wrongly block huge amounts of normal traffic that would severely impact the business. It's understandable that in the early days of machine learning, some organizations were concerned that the models wouldn't be as accurate as human security researchers. It takes time, and it also takes huge amounts of data to actually train a machine learning model to get up to the same level of accuracy as a really skilled human. Humans, however, don't scale and are among the scarcest resources in IT today. We are relying on ML to efficiently scale up the cybersecurity solutions. Also, ML can help us detect unknown attacks that are hard for humans to detect, as ML can build up baseline behaviors and detect any abnormalities that deviate from them.

Challenge 2: The access to large amounts of training data, especially labeled data. Machine learning requires a large amount of data to make models and predictions more accurate. Gaining malware samples is a lot harder than acquiring data in image processing and NLP. There is not enough attack data, and lots of security risk data is sensitive and not available because of privacy concerns.

Challenge 3: The ground truth. Unlike images, the ground truth in cybersecurity might not always be available or fixed. The cybersecurity landscape is dynamic and changing all the time. Not a single malware database can claim to cover all the malware in the world, and more malware is being generated at any moment. What is the ground truth that we should compare to in order to decide our accuracy?

Three ML Challenges Made More Severe in Cybersecurity

There are other challenges that are common for ML in all sectors but more severe for ML in cybersecurity.

Challenge 1: Explainability of machine learning models. Having a comprehensive understanding of the machine learning results is critical to our ability to take proper action.

Challenge 2: Talent scarcity. We have to combine domain knowledge with ML expertise in order for ML to be effective in any area. Either ML or security alone is short of talent; it is even harder to find experts who know both ML and security. That's where we found it is critical to make sure ML data scientists work together with security researchers, even though they don't speak the same language, use different methodologies, and have different ways of thinking and different approaches. It is very important for them to learn to work with each other. Collaboration between these two groups is the key to successfully applying ML to cybersecurity.

Challenge 3: ML security. Because of the critical role cybersecurity plays in each business, it is more critical to make sure the ML we use in cybersecurity is secure by itself. There has been research in this area in academics, and we are glad to see and contribute to the industry movement in securing ML models and data. Palo Alto Networks is driving innovation and doing everything to make sure our ML is secure.

The goal of machine learning is to make security more efficient and scalable in an effort to help save labor and prevent unknown attacks. It's hard to use manual labor to scale up to billions of devices, but machine learning can easily do that. And that is the kind of scale organizations truly need to protect themselves in the escalating threat landscape. ML is also critical for detecting unknown attacks in many critical infrastructures. We can't afford even one attack, which can mean life or death.

How Machine Learning Enables the Future of Cybersecurity

Machine learning supports modern cybersecurity solutions in a number of different ways. Individually, each one is valuable, and together they are game-changing for maintaining a strong security posture in a dynamic threat landscape.

Identification and profiling: With new devices getting connected to enterprise networks all the time, it's not easy for an IT organization to be aware of them all. Machine learning can be used to identify and profile devices on a network. That profile can determine the different features and behaviors of a given device.

Automated anomaly detection: Using machine learning to rapidly identify known bad behaviors is a great use case for security. After first profiling devices and understanding regular activities, machine learning knows what's normal and what's not.

Zero-day detection: With traditional security, a bad action has to be seen at least once for it to be identified as a bad action. That's the way that legacy signature-based malware detection works. Machine learning can intelligently identify previously unknown forms of malware and attacks to help protect organizations from potential zero-day attacks.

Insights at scale: With data and application in many different locations, being able to identify trends across large volumes of devices is just not humanly possible. Machine learning can do what humans cannot, enabling automation for insights at scale.

Policy recommendations: The process of building security policies is often a very manual effort that has no shortage of challenges. With an understanding of what devices are present and what is normal behavior, machine learning can help to provide policy recommendations for security devices, including firewalls. Instead of having to manually navigate around different conflicting access control lists for different devices and network segments, machine learning can make specific recommendations that work in an automated approach.

With more devices and threats coming online every day, and human security resources in scarce supply, only machine learning can sort complicated situations and scenarios at scale to enable organizations to meet the challenge of cybersecurity now and in the years to come.

paloalto®
NETWORKS

3000 Tannery Way
Santa Clara, CA 95054

Main: +1.408.753.4000
Sales: +1.866.320.4788
Support: +1.866.898.9087

www.paloaltonetworks.com

© 2022 Palo Alto Networks, Inc. Palo Alto Networks is a registered trademark of Palo Alto Networks. A list of our trademarks can be found at https://www.paloaltonetworks.com/company/trademarks.html. All other marks mentioned herein may be trademarks of their respective companies.
parent_a_the-future-of-machine-learning-in-cybersecurity_081222