Chapter 12

Integrating Adversarial Training Techniques to Enhance Cybersecurity Resilience Against Machine Learning Threats

Firas Tarik Jasim
Northern Technical University, Al-Dour Technical Institute, Iraq

Alnoman Mundher Tayyeh
https://orcid.org/0000-0001-7651-0240
Institute of Technology, Middle Technical University, Iraq

Abdulsattar Abdullah Hamad
College of Education, University of Samarra, Iraq

Omar Azeez Abbas
College of Administration and Economics, University of Samarra, Iraq

Noor Mohammed kadhim
College Education for Human Sciences, Wasit University, Iraq

Ahmed Ibrahim Turki
University of Samarra, Iraq

Elaf Sabah Abdulwahid
College of Education for Women, Tikrit University, Iraq

Khalid Saeed Lateef Al-Badri
https://orcid.org/0000-0003-3678-4954
College of Education, Samarra University, Iraq

Luma Saad Abdalbaqi
College of Education for Women, Tikrit University, Iraq

ABSTRACT

The speedy improvement of gadget mastering technology has considerably trans-

DOI: 10.4018/979-8-3373-0330-7.ch012

formed the landscape of cybersecurity. However, those structures are increasingly more liable to antagonistic assaults that make the maximum their weaknesses, posing enormous dangers to their effectiveness. This look investigates the mixture of hostile training strategies into device studying models to decorate their resilience against evolving cybersecurity threats. Our findings display a remarkable decline in overall performance while models are uncovered to adverse examples, with benign detection quotes within the IDS losing from 90% to 80%. In evaluation, the phishing detection tool demonstrates an ability to evolve through retraining, with accuracy increasing from 88% to 93% after imposing non-stop learning strategies. We advise a feedback loop for non-stop gaining knowledge. The outcomes underscore the need for ongoing variation in gadget studying fashions to protect in opposition to ultra-modern cyber threats, supplying treasured insights for future studies and realistic applications in cybersecurity.

INTRODUCTION

The rapid advancements in machine learning (ML) technologies have significantly transformed various domains, particularly cybersecurity, where automated systems are increasingly relied upon to detect and mitigate cyber threats. These technologies enable the development of systems such as Intrusion Detection Systems (IDS) and Phishing Detection Systems, which are crucial for identifying and responding to potential cyberattacks. However, alongside these advancements, Nguyen et al (2015), adversarial attacks have emerged as a critical challenge, exploiting vulnerabilities in ML models to compromise their effectiveness. These attacks introduce intentionally crafted inputs designed to mislead models, resulting in severe consequences for cybersecurity systems, Tramèr et al., (2017).

Adversarial attacks not only threaten the reliability of ML-based cybersecurity systems but also highlight the need for robust strategies to ensure their resilience against evolving threats, Papernot et al., (2016). Traditional ML models often struggle to maintain their performance when exposed to adversarial examples, leading to significant declines in detection accuracy, Zhang et al., (2021). For instance, studies have shown that benign detection rates in IDS models can drop from 95% under normal conditions to 80% when adversarial examples are introduced. Similarly, Szegedy, (2013), while phishing detection systems exhibit a degree of adaptability through retraining, their initial vulnerability underscores the limitations of existing methodologies.

This study aims to address these challenges by exploring the integration of adversarial training techniques into ML models to enhance their robustness against adversarial threats.

A Power Control Strategy for IoT Sensors Developed for 5G Networks

Weston Mwashitaand Marcel Ohanga Odhiambo (2020). *International Journal of Smart Sensor Technologies and Applications (pp. 22-41).*
www.irma-international.org/article/a-power-control-strategy-for-iot-sensors-developed-for-5g-networks/272126