Chapter 4
The Era of Advanced Machine Learning and Deep Learning Algorithms for Malware Detection

Kwok Tai Chui
Hong Kong Metropolitan University, Hong Kong

Patricia Ordóñez de Pablos
The University of Oviedo, Spain

Miltiadis D. Lytras
Deree College — The American College of Greece, Greece

Ryan Wen Liu
https://orcid.org/0000-0002-1591-5583
Wuhan University of Technology, China

Chien-wen Shen
National Central University, Taiwan

ABSTRACT

Software has been the essential element to computers in today's digital era. Unfortunately, it has experienced challenges from various types of malware, which are designed for sabotage, criminal money-making, and information theft. To protect the gadgets from malware, numerous malware detection algorithms have been proposed. In the olden days there were shallow learning algorithms, and in recent years there are deep learning algorithms. With the availability of big data for training of model and affordable and high-performance computing services, deep learning has demonstrated its superiority in many smart city applications, in terms of accuracy, error rate, etc. This chapter intends to conduct a systematic review on the latest development of deep learning algorithms for malware detection. Some future research directions are suggested for further exploration.

DOI: 10.4018/978-1-7998-7789-9.ch004

INTRODUCTION

Computing tools and smartphones have played an impactful role towards smart city vision in recent decades (Ficco, Esposito, Xiang, & Palmieri, 2017; Rose, Raghuram, Watson, & Wigley, 2021). According to the Statista (Technology Markets: Software, 2021), as shown in Figure 1, there is a steady growth rate of around 7.1-7.7% in the revenue of software development from 2017 to 2025, except the historical low 2.6% in 2020 and bounced back to 9% in 2021 during pandemic. The projection could be altered depending on the deployment of 5G and development of 6G (Stergiou, Psannis, & Gupta, 2020).

Intuitively, the more the number of software linking to gadgets, the more the number of malware attacks. Yet, numerous types of malware have been developed such as scareware, wiper, rogue software, adware, spyware, ransomware, Trojan horses, worms, and computer viruses (Kumar, 2020; Rendell, 2019). Surprisingly, the global yearly malware attacks (SonicWall, 2021) does not follow an increasing trend, as shown in Figure 2. From 2015 to 2017, the percentage changes in the number of malware attacks are -3.7% and 8.9%, respectively. There was a notably increment by 22.1% from 2017 to 2018 and slightly decrement by 5.7% from 2018 to 2019. Compared the last two recorded periods from 2019 to 2020, a significant drop (43.4%) in the number of malware attacks was observed. The key explanation to the drop of the malware attacks is malware detection algorithms which can detect malware and thus avoid the damage of gadgets.

A lot of traditional machine learning algorithms was employed for malware detection in literature, including decision tree, Naïve Bayes, support vector machine, K-nearest neighbour, Bayseian network, multi-layer perception, J48, and random forest, (Jerlin, & Marimuthu, 2018; Li et al., 2018; Narudin, Feizollah, Anuar, & Gani, 2016). There is room for improvement in terms of accuracy. Owning to the fact that a large amount of data is available as training dataset, attention is drawn into deep learning which can further enhance the accuracy of the detection model.

Figure 1. The worldwide statistics on the revenue of software development (by segment).