Journal of The Institute of Electronics and Information Engineers Vol.54, NO.2, February 2017 https://doi.org/10.5573/ieie.2017.54.2.076
ISSN 2287-5026(Print) / ISSN 2288-159X(Online)

논문 2017-54-2-10

악성코드로부터 빅데이터를 보호하기 위한 이미지 기반의 인공지능 딥러닝 기법

(Image-based Artificial Intelligence Deep Learning to Protect the Big Data from Malware)

김 혜 정*, 윤 은 준**

(Hae Jung Kim and Eun Jun Yoon©)

요 약

랜섬웨어를 포함한 악성코드를 빠르게 탐지하여 빅데이터를 보호하기 위해 본 연구에서는 인공지능의 딥러닝으로 학습된 이미지 분석을 통한 악성코드 분석 기법을 제안한다. 우선 악성코드들에서 일반적으로 사용하는 2,400여개 이상의 데이터를 분석하여 인공신경망 Convolutional neural network 으로 학습하고 데이터를 이미지화 하였다. 추상화된 이미지 그래프로 변환하고 부분 그래프를 추출하여 악성코드가 나타내는 집합을 정리하였다. 제안한 논문에서 추출된 부분 집합들 간의 비교 분석을 통해 해당 악성코드들이 얼마나 유사한지를 실험으로 분석하였으며 학습을 통한 방법을 이용하여 빠르게 추출하였다. 실험결과로부터 인공지능의 딥러닝을 이용한 정확한 악성코드 탐지 가능성과 악성코드를 이미지화하여 분류함으로써 더욱 빠르고 정확한 탐지 가능성을 보였다.

Abstract

Malware, including ransomware to quickly detect, in this study, to provide an analysis method of malicious code through the image analysis that has been learned in the deep learning of artificial intelligence. First, to analyze the 2,400 malware data, and learning in artificial neural network Convolutional neural network and to image data. Extracts subgraphs to convert the graph of abstracted image, summarizes the set represent malware. The experimentally analyzed the malware is not how similar. Using deep learning of artificial intelligence by classifying malware and It shows the possibility of accurate malware detection

Keywords : Malware, Artificial Intelligence, Deep Learning, Convolutional Neural Network, Security

1. 서 론

현대는 인공지능, 생명과학, 로봇기술 등 기업들의 제조업과 정보통신기술(ICT)을 융합한 빅데이터 기반의 4차 산업혁명 시대로서 새로운 시대를 선도하고 있다. 차세대 4차 산업혁명을 주도하기 위해서는 사람과 사물, 사물과 사물이 인터넷 통신망으로 연결되는 사물인터넷 시대에 막대한 빅데이터를 분석할 필요가 있다. 빅데이터를 분석하여 데이터 속에 숨겨진 의미 있는 패턴을 파악하고 분석하여 미래를 예측하는 것이 가능하기 때문이다. 그러나 이러한 빅데이터의 활용에 있어 가장 장애가 되고 있는 악성코드의 공격과 같은 사이버 보안 문제 또한 심각해지고 있다. 더구나 악성코드 가운데 랜섬웨어는 감염시킨 PC를 암호화하여 볼모로 돈을 요구하는 것으로서 보다 많은 사이버 공격을 받을 것으로 예상된다. 또한 랜섬웨어의 공격이 일반사용자에서 기업으로 공격목표를 바꾸어 가고 있다. 따라서 랜섬웨어와 같은 악성바이러스로부터 빅데이터를 보호하는 방법이 절실해지고 있다[1~2]. 악성바이러스의 공격에 대비하기 위해서는 바이러스 방지 프로그램을 이용

* 정회원, 경일대학교 사이버보안학과(Department of Cyber Security, Kyung-il University)
** 정회원, 경일대학교 사이버보안학과(Department of Cyber Security, Kyung-il University)
© Corresponding Author (E-mail : ejyoon@kiu.kr)
Received ; November 3, 2016 Revised ; December 8, 2016
Accepted ; January 24, 2017

하고, 파일에 대한 주기적 백업과 알 수 없는 메일에 첨부된 파일을 열지 않는 등의 기본적인 방법이 있지만 보다 정확하고 판단하기 쉬운 방법이 필요하다. 따라서 데이터베이스에 존재하는 링크된 파일의 헤더 정보와 해싱값 정보를 파악하고 탐색하여 악성코드를 찾아내는 기존의 해결법과는 달리 인공지능의 딥러닝 기법을 적용하고자 한다.

본 논문에서는 인공지능의 딥러닝 방법으로 랜섬웨어를 포함한 악성코드를 학습하여 악성코드에 대한 패턴을 인식하고, 인식된 악성코드의 특징을 사용하여 이미지화하는 방식의 딥러닝 기반 악성코드 검사 시스템을 제안하고자 한다. 인공지능의 딥러닝을 이용하여 악성코드를 학습하고 악성코드의 일부분이 달라지더라도 비슷한 이미지가 형성되는 특성을 이용하여 새로운 방법의 악성코드 검사 방법을 제안하고자 한다. 제안한 논문의 구성은 다음과 같다. II장의 본론에서는 랜섬웨어를 포함한 악성코드의 정의, 탐색방법과 관련연구에 대해 기술한다. III장에서는 제안한 인공지능의 딥러닝을 적용한 학습방법과 악성코드 이미지화 방법을 설명하며 IV장에서는 시스템의 검증을 위해 다른 기계학습 기반 시스템과 비교 실험, 분석을 한다. V장에서는 결론 및 향후 연구에 대해 설명한다.

II. 관련 연구

1. 악성코드와 탐색 방법

악성코드(Malware)란 돈이 되는 바이러스 랜섬웨어를 포함하여 컴퓨터에 악영향을 끼칠 수 있는 모든 소프트웨어로서 예전에는 단순히 컴퓨터 바이러스만이 활동하였지만 네트워크가 발달하면서 이메일이나 웹으로 감염되는 경우가 많아졌다. 더구나 랜섬웨어와 같은 악성바이러스는 백신 프로그램으로 악성코드를 제거해도 암호화된 파일은 복구되지 않기 때문에 최악의 악성코드라 불린다. 랜섬웨어와 같은 악성바이러스는 메일 송신과 메일 열기 시 첨부파일에 의해 감염의 대부분이 발생하기 때문에 악성코드를 탐지하는 방식으로 주요 확장자를파일은 메일 첨부에 금지하거나 Virustotal 안티바이러스 프로그램의 엔진을 이용하여 검색하는 방법을 주로 사용하고 있다. 다양한 악성코드 분석 및 탐지 연구가 활발하게 진행되고 있지만, 시간이 갈수록 지능화되고 정교해지는 랜섬웨어와 같은 악성코드를 대응하기에는 많은 한계가 있다. 현재 랜섬웨어 악성코드 탐지 및 대응을 위해 안티 바이러스 업체에서 가장 일반적으로 사용하고 있는 방법은 시그너쳐 기반의 탐지 방법이다. 그림 1 은 기존의 시그너쳐 기반 악성코드 검사시스템의 구조를 도식화하였다. 그림 1 과 같이 기존의 악성코드 검사는 입력의 바이너리 파일로 해싱된 값을 사용하여 데이터베이스에서 동일한 해싱값을 검색하여 탐지하는 방식이다[2].

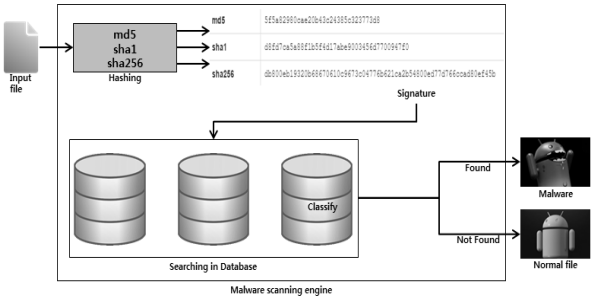


그림 1. 기존의 Malware 탐지 시스템의 동작 원리
Fig. 1. The operation principle of the conventional detection systems Malware.

이러한 시그너쳐 기반 악성코드 탐지 시스템은 악성코드의 일부분이 조금만 달라져도 해싱값이 전혀 달라 시스템을 속이끄기 쉬운 문제가 발생한다[2~3]. 그림 2에서는 16진수로 표현된 바이너리 파일의 예시를 들었다.

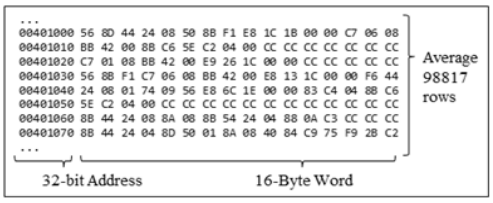


그림 2. 16진수로 표현된 바이너리 파일
Fig. 2. The binary files in hexadecimal representation.

따라서 악성코드를 탐지하기 위한 새로운 방법이 필요하다. 본 논문에서는 기존의 시그니쳐 기반 악성코드 탐지 시스템과는 달리 악성코드의 일부가 달라져도 악성코드임을 구별해 낼 수 있도록 악성코드를 이미지화하였다. 악성코드 내부의 일부분이 달라져도 비슷한 이미지가 형성되는 특성을 이용하여 이미지를 이용한 새로운 방법의 악성코드 탐지 시스템을 제안한다. 또한 인공지능의 딥러닝을 이용하여 악성코드를 학습함으로써 악성코드 탐색과 새로운 변종에 대한 정확도를 높일 수 있다.

2. 선행 연구의 문제점

이전의 연구들에서는 악성코드를 이미지화하여 분류

해내지 않고 악성코드의 해싱값을 시그너쳐로 사용하여 데이터베이스에서 검색하는 방법을 사용하였다. 따라서 변종 악성코드의 정확한 탐지를 위해 악성코드의 일부분이 달라져 해싱값이 전혀 달라지더라도 데이터베이스로부터 검색될 수 있도록 하는 선행 연구가 진행되었다[1~2]. 다른 방법으로는 기계학습에 기반한 방법을 사용해서 악성코드에서 추출해낸 특징을 학습시키고, 새로운 명령어들에 대해 학습된 특징을 찾아내는 방법의 연구가 진행되었다. 추출해낸 특징에 대해 학습 및 검증을 하는 기계학습에 기반한 분류기로 k-nn 를 사용한 사례나 XGBoost 을 사용한 사례에서는 기계학습 기법을 사용하기 위한 기본적인 전처리 과정과 추출해내는 특징에 대해 소개하고 있다[4~5]. 이러한 방법에서는 악성 코드로부터 추출해낸 특징의 학습에 따라 분류기의 성능이 크게 달라지기 때문에 추출시킬 특징을 바꾸어 가면서 분류기의 성능을 확인하였다.

반면에 악성코드를 이미지화하여 특징으로 사용하는 경우에는 악성코드로부터 추출해낼 특징을 선택할 필요 없이 이미지를 특징으로서 사용하게 된다[6]. 기계학습에 기반한 분류기를 사용하는 악성코드 검사 시스템의 사례 중에서 지나치게 많은 특징을 사용하는 경우 불필요한 정보가 분류기의 입력으로 사용되어 분류 성능을 하락시키는 문제가 있다[7~8]. 최근 인공지능의 인공신경망을 이용한 딥러닝 학습기법들이 고해상도 이미지 분야에서 그러한 불필요한 정보를 포함한 고차원적인 입력의 특징을 성공적으로 추출해내고 있다[9]. 따라서 본 논문에서는 인공지능 가운데 이미지 분류에 뛰어난 성능을 보이는 심층 컨볼루션 신경망(Deep Convolutional Neural Network : CNN)을 이용하여 악성코드를 학습시키고 이미지화한 악성코드를 대상으로 탐색하는 시스템을 제안한다.

3. 인공지능의 딥러닝을 적용한 학습

딥러닝이란 경험(데이터)에 의해 스스로 현명해지는 알고리즘인 기계학습 기술의 일종으로 수많은 데이터 속에서 패턴을 찾아내어 인간이 사물을 구분하듯 컴퓨터가 데이터를 구분해 내는 것이다[3]. 인간의 뇌가 동작하는 방식, 특히 뉴런과 시냅스의 신경 네트워크 구조를 본떠서 심층 신경망 알고리즘을 모델링한 것이다. 심층 신경망은 신경망 알고리즘 중에서 여러 개의 층으로 이루어진 신경망을 의미한다. 심층 신경망이 일반적인 기계 학습과 다른 점은 특징 추출(feature extraction)이 자동적으로 이루어지는 점이다. 그림 3은 인공 신경망의 학습 원리를 그림으로 나타내었다.

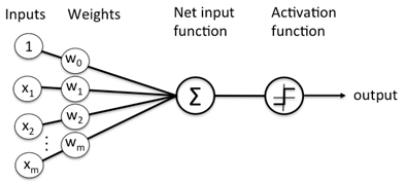


그림 3. 인공 신경망의 학습 원리
Fig. 3. Artificial Neural Networks learning principles.

한 층은 여러 개의 노드로 이루어져 있다. 이러한 인공지능의 기계학습 방법 가운데 딥러닝을 활용하여 악성코드를 학습하고, 학습으로 패턴을 분류하고 정확하게 찾아낼 수 있다면 랜섬웨어와 같은 악성코드의 감염으로 부터의 보안을 위한 시간과 노력, 자금을 줄일 수 있다. 또한 증가하고 있는 사이버보안의 공격으로 부터 일반 사용자, 기관 사용자들도 쉽게 대응할 수 있는 방법이 될 것이다.

Ⅲ. 제안하는 악성코드 분석 기법

1. 악성코드 데이터 셋

제안하는 인공지능 딥러닝 기반의 악성코드 검사 시스템의 학습 및 검증 데이터로 사용한 Microsoft malware classification challenge dataset 은 2015년 Microsoft 사에서 기계학습 기반 데이터 분석 대회인 Kaggle machine learning challenge 에 공개하였다[2]. 아래 표 1 에서 서로 다른 9가지 악성코드의 종류를 기재하였으며 총합 10868개, 약 200GB 용량으로 이루어져 있다.

각각의 악성코드 데이터는 .asm 확장자인 어셈블리 파일과 .bytes 확장자인 바이너리 파일로 구성되어 있다. 각 행은 32비트의 주소와 16 바이트의 워드로 이루어져 있으며 데이터는 평균 98817 행으로 구성되어 있고 길이가 서로 다르다. 이러한 고차원 공간상의 데이터를 벡터화하여 기계학습 기법 분류기의 학습 데이터로 사용하는 경우, 차원 축소 알고리즘을 통해 연관 없고 중복된 정보를 제거하여 분류기의 성능을 향상시키고 계산 복잡도를 하락시킬 수 있다. 본 논문에서는 주어진 고차원의 악성코드 데이터들을 이미지로 간주하여 간단히 이미지를 축소하는 과정을 통해 데이터의 차원을 축소하였다.

표 1. 데이터셋에 포함된 악성코드 종류와 기능
Table1. Malware type and features in the data set.

Class index	Malware name	Description
1	RAMNIT	Strong botnet function
2	LOLLIPOP	Adware
3	KELIHOS v.3	p2p botnet using polymorphism Encrypted
4	VUNDO	Multi-component malware family: trojan, worm
5	SIMDA	Most complex malware Multi-component malware family: botnet,trojan,backdoor,password-stealing
6	TRACUR	Trojan
7	KELIHOS v.1	Botnet
8	OBFUSCATOR.ACY	Combination of methods: Encryption,Compression,Anti-debugging,Anti-emulationtechniques
9	GATAK	Trojan

2. 제안하는 모델의 구조

악성코드 검사를 위해 기계학습에 기반한 분류기를 사용한 연구에서는 악성코드로부터 추출해 낼 특징을 선택하기 때문에 특징의 일부만을 사용하게 된다. 이러한 문제를 극복하기 위해 기존 연구 중 악성코드의 탐지를 위해 악성코드를 이미지화하고 이미지의 질감을 기계학습에 사용할 특징으로 사용한 연구가 있었다. 그러나 이러한 시도는 여전히 고차원의 특징을 학습시키는 경우 성능이 하락하는 기계학습 기법의 특성 때문에 이미지 전체를 특징으로서 사용하지 못하였다. 따라서 본 논문에서는 고차원의 이미지에 대해 자동으로 최적의 특징을 선택해 내는 인공지능 딥러닝 기반의 심층 컨볼루션 인공신경망을 사용하는 악성 코드 검사 시스템을 제안한다.

그림 4는 제안하는 방법을 도식화하였다.

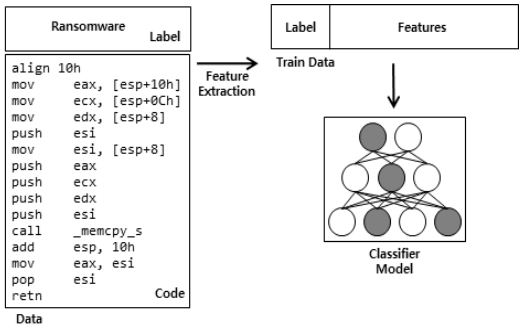


그림 4. 제안하는 방법
Fig. 4. The proposed method.

Convolutional Neural Network은 Convolution, Pooling 연산을 수행하며 이미지에서 최적의 Feature map을 추출해내는 기법으로서 수식으로 나타내면 다음과 같다.[10~12]

$$V_{xy}^{'1} = \sum_{a=0}^{m-1}\sum_{b=0}^{m-1} w_{ab} y^{l-1}{}_{(x+a)(y+b)} \quad (1)$$

수식 (1) 에서 l 번째 Convolutional 층에서 출력되는 벡터 $V^{'}$의 원소 $V_{xy}^{'}$는 이전 층의 출력 벡터 y^{l-1}과 $m \times m$크기의 벡터인 필터 w에 의해 3차원 Convolution 연산을 수행한다. Pooling 층에서는 입력된 $N \times N$크기의 벡터의 $k \times k$영역으로부터 하나의 최대값을 대표값으로 고르는 Max-pooling 연산을 수행하고 $\frac{N}{k} \times \frac{N}{k}$크기의 벡터를 출력한다[11].

아래 그림 5에서는 제안하는 시스템의 전반적인 구조를 도식화하였다. 좌측 바이너리 파일을 이미지화한 악성코드에 대해 반복적인 2×2 Convolution과 2×2 Pooling 연산을 거친다. 마지막으로 하나의 Fully-connected layer를 사용하였다. Convolution 과 Pooling 연산을 사용하는 디자인은 이미지 분류에서 주로 사용되며 대회에서 최고의 성능이 입증되었다[9, 12]. 학습에 의해 이미지화한 악성코드가 입력되었을 때 제안하는 시스템을 통해 자신의 악성코드 종류를 출력하도록 하였다.

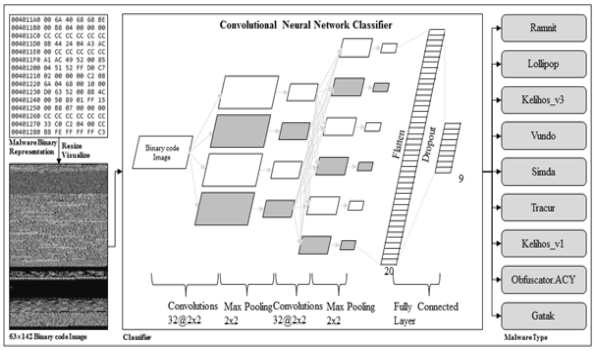


그림 5. 제안하는 딥러닝 기반의 Convolutional 인공신경망 악성코드 검사 시스템
Fig. 5. Proposal malware inspection system.

3. 인공지능을 활용한 데이터 시각화

각각의 악성 코드에는 제작자가 끼워놓은 서명이 존재한다. Arecibo message 라고 부르는 이러한 서명은 적당한 크기로 바이너리 파일을 재배치하고 16 바이트

로 이루어진 명령어의 각 바이트를 정수로 대체 픽셀값으로 사용하여 시각화할 수 있다. 아래의 그림 6은 악성코드 바이너리 파일의 좌측 메모리 주소를 제외하고 행과 열을 계산하여 재배열한 것이고 각 바이트 값을 정수로 환산하여 이미지화 한 것이다.

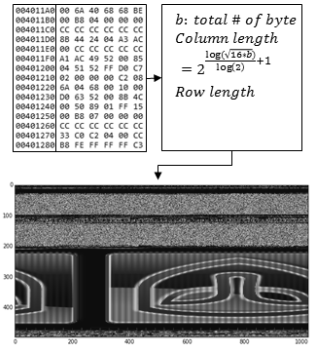


그림 6. 악성코드 바이너리 파일의 시각화
Fig. 6. Visualization of malicious binaries.

그림 7 에서는 여러 개의 악성코드에 대해 같은 알고리즘을 적용하여 이미지로 시각화한 예들을 보여주고 있다. 시각화된 예제들에는 이미지로 보일 수 있는 서명이 나타나고 있다.

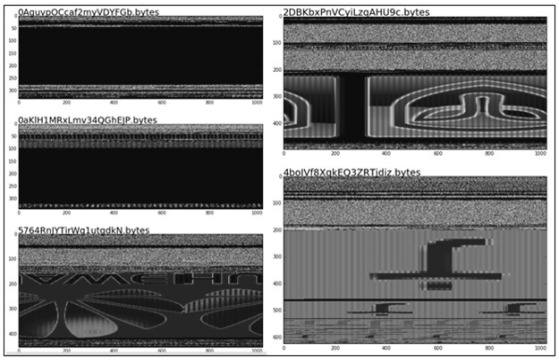
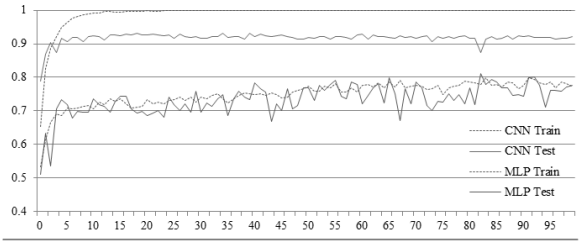


그림 7. 여러 악성코드 바이너리 파일의 시각화
Fig. 7. Visualization of several malicious binary file.

IV. 실험 결과

제안하는 악성코드를 이미지화하고 이미지에 강력한 성능을 보이는 딥러닝 기반의 심층 컨볼루션 인공신경망 방법을 사용한 분류기의 성능 비교를 위해 일반적인 인공신경망 방법을 적용한 분류기와 비교 실험하였다.

학습과정에서는 내부의 파라미터를 미분하여 Loss Function 을 최소화하는 값으로 노드를 갱신하는데, 이때 다른 노드의 값이 반영되어 과적합 문제가 발생할 수 있다[13]. 따라서 그림 8 에서 제안하는 모델의 과적합 여부를 알아보기 위해서 학습과 검증 정확도를 반복회수별로 나타내었다. 반복 학습이 진행될수록 기존의 기계학습 방법 MLP 방법은 학습 및 검증 정확도가 70~80% 사이에서 불안정하였으나 제안하는 CNN 방법은 학습 정확도가 반복회수 15회에 거의 100%에 수렴하였으며 검증 정확도 또한 91~92% 사이에서 안정된 경향을 보였다.



Epochs	10	20	30	40	50	60	70	80	90	100
CNN Train Accuracy	0.9863	0.9967	0.9988	0.9987	0.9989	0.9993	0.9991	0.9995	0.9996	**0.9995**
CNN TestAccuracy	0.9056	0.9306	0.9185	0.9139	0.9139	0.9139	0.9185	0.9213	0.9130	**0.9176**
MLP Train Accuracy	0.7129	0.7110	0.7422	0.7537	0.7610	0.7678	0.7681	0.7765	0.7831	0.7803
MLP Test Accuracy	0.6963	0.6944	0.6972	0.7417	0.7157	0.7861	0.7667	0.7491	0.7472	0.7722

그림 8 CNN, MLP 기반의 악성코드 분류기의 반복회수별 학습 및 정확도 검증
Fig. 8. CNN, for MLP-based learning and accuracy by verifying repetitions.

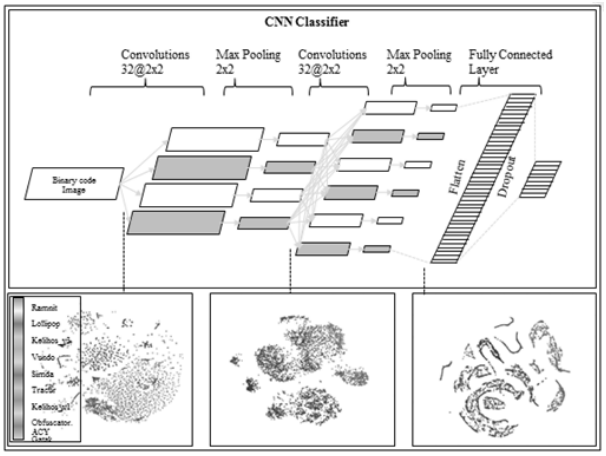


그림 9. 시각화한 모델 내부에서의 데이터 분포
Fig. 9. Model data distribution in the interior.

그림 9 에서 제안하는 모델 내부에서의 데이터의 분포를 t-SNE 알고리즘을 사용하여 시각화하였다. 데이터가 인공신경망에 입력되면 심층으로 갈수록 각 노드의 파라미터와 곱해지는데, 이를 시각화하여 군집화하는 경향을 통해 모델 내부에서의 학습 정도를 파악할 수 있다.

입력 초기의 여러 악성코드 데이터가 서로 뒤섞인 것에 비해 모델 심층으로 갈수록 악성코드의 종류가 다름

에도 불구하고 비슷한 이미지로 군집화하여 나타난다. 따라서 분류기의 학습 정도가 잘 진행되었음를 직관적으로 판단할 수 있으며 실험에 의해 증명되었다.

아래 그림 10 에서 분류기 출력층에서 내부 파라미터들과 곱해진 악성코드 데이터의 분포를 시각화하였다. 중앙에 Kelihos_v3 악성코드와 kelihos_v1 악성코드가 비슷한 위치에 맵핑된 것을 제외하고 종류별로 군집화하는 경향을 보였다.

점점 진화하고 있는 새로운 악성코드의 탐지에 대해서도 인공 신경망 학습으로 신속하게 분류하는 것이 가능하며 새롭게 나타나는 악성코드에 대해 빠른 탐색이 가능함을 보이고 있다.

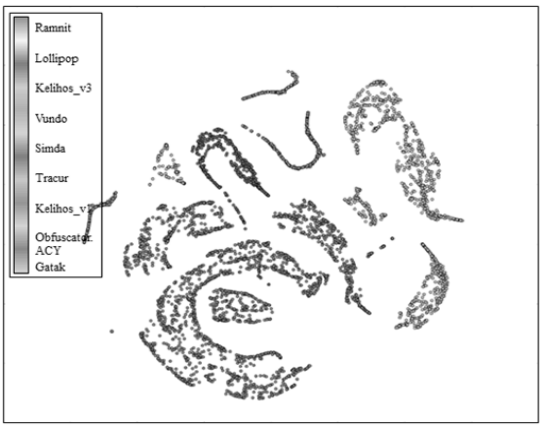


그림 10. 분류기 출력층에서의 악성코드 데이터의 군집화
Fig. 10. Clustering of malicious code data from the output layer.

Ⅳ. 결론 및 향후 연구 방향

기존에 사용되는 메일에 포함되거나 첨부된 파일의 악성코드를 탐지하는 방식은 단순히 데이터베이스에서 파일의 변화와 레지스트리를 검색하는 방식이다. 해싱값을 활용하는 하나의 알고리즘으로서, 입력된 값의 해싱값 변화에 민감하여 정확한 값을 찾기가 힘들다. 따라서 백신을 사용한다고 하더라도 많은 양의 메일이 첨부파일에 의해 랜섬웨어와 같은 악성바이러스 감염이 이루어지며 진짜 메일과 구별하기 힘들어 속기가 쉽다. 따라서 인공지능의 딥러닝 학습을 이용하여 악성코드의 패턴을 인식하고 데이터를 스스로 학습한다. 또한 작은 값의 변화에 민감하지 않고 정확하게 악성코드를 찾아내도록 이미지화하여 보여줌으로써 증가하고 있는 보안에 대처할 수 있다. 또한 악성코드로부터 빅데이터 보호를 할 수 있는 방법이 될 수 있으리라 짐작한다. 그러나 딥러닝 연구를 하기 위한 많은 양의 라벨화된 악성코드 데이터가 이미 존재하므로 이것을 활용하여 연구를 진행할 수 있지만 진화된 연구를 위해서는 우리나라에서 빅데이터 사용의 규제가 좀 더 자유로워져야 할 것이다.

본 논문에서는 악성코드를 이미지화하고 크기를 재조정하여 딥러닝 기반의 심층 컨볼루션 인공신경망을 통해 학습하고 분류하였다. 바이너리 파일을 이미지화하여 1/10크기로 재조정한 입력을 사용하여 91.7% 정확도로 악성코드를 구별해 내었다. 또한, 다른 방법과 비교하기 위해 MLP 방법의 분류기와 비교 실험하고, 학습 및 검증 정확도를 반복 회수별로 그래프화하여 2차원 공간에 시각화하였다. 따라서 악성코드로 부터의 탐색에 대한 정확도를 향상 시켜 보안을 강화할 수 있다. 추가적인 연구로 인공지능을 활용한 악성코드 탐색을 스마트폰 등 SNS를 통해 일어날 수 있는 악성코드 감염 예방에 적용 시켜 빅데이터를 보호 하고자 한다.

감사의 글

본 연구는 문화체육관광부 및 한국저작권위원회의 2016년도 저작권기술개발사업(No. 2016-CCP-9500)의 연구결과와 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구(No. NRF-2015R1A2A2A01006824)로 인한 결과물임을 밝힙니다.

REFERENCES

[1] Luo, Xin, and Qinyu Liao, “Awareness Education as the key to Ransomware Prevention,” Information Systems Security 16.4 pp. 195-202, 2007.
[2] P. Vinod, R. Jaipur, V. Laxmi and M. Gaur, “Survey on malware detection methods,” Proceedings of the 3rd hackers′ workshop on computer and internet security,pp. 74-79, March 2009.
[3] https://www.kaggle.com/c/malware-classification
[4] A. Kumar, N. Sharma, A. Khanna and S. Gandhi, “Analysis of machine learning techniques used in malware classification in cloud computing environment,” International journal of computer applications, Vol. 133, pp. 15-18, 2016.
[5] M. Ahmadi, D. Ulyanov, S. Semenov, M. Trofimov and G. Giacinto, “Novel feature extraction, selection and fusion for effective malware family classification,” Proceedings of the 6th ACM conference on data and application security and privacy, pp. 183-194, 2016.

[6] L. Nataraj, S. Karthikeyan, G. Jacob and B. S. Manjunath, “Malware images: visualization and automatic classification,” Proceedings of the 8th international symposium on visualization for cyber security, pp. 4, 2011.
[7] Guyon and A. Elisseeff, “An introduction to variable and feature selection,” Journal of machine learning research, vol. 3, pp. 1157-1182, 2003.
[8] J. G. Dy and C. E. Brodley, “Feature selection for unsupervised learning,” Journal of machine learning research, vol. 5, pp. 845-889
[9] A. Krizhevsky, I. Sutskever and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Advances in neural information processing systems, pp. 1097-1105, 2012.
[10] T. N. Sainath, A. R. Mohamed, B. Kingsbury and B. Ramabhadran, “Deep convolutional neural networks for LVCSR,” 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 8614-8618, 2013.
[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov and A. Rabinovich, “Going deeper with convolutions,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-9, 2015.
[12] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” Journal of Machine Learning Research, vol. 15, pp. 1929-1958, 2014.
[13] D. R. Wilson and T. R. Martinez, “The general inefficiency of batch training for gradient descent learning,” Neural Networks, vol. 16, pp. 1429-1451, 2013.

저 자 소 개

김 혜 정(정회원)
1987년 경북대학교 졸업(이학사)
1989년 경북대학교 전자공학과 석사 졸업(공학석사)
2005년 경북대학교 컴퓨터공학과 박사 졸업(공학박사)

<주관심분야 : 데이터베이스보안, 인공지능, 보안>

윤 은 준(정회원)
1995년 경일대학교 졸업(공학사)
2003년 경일대학교 컴퓨터공학과 석사 졸업(공학석사)
2007년 경북대학교 컴퓨터공학과 박사 졸업(공학박사)

<주관심분야 : 정보보호, 컨텐츠보안, 저작권보호, 암호학, 네트워크보안, 인증>