딥러닝 및 이미지화를 통한 악성코드 분류 시스템

조현진*, 홍윤영*, 김호원*
*부산대학교
hyunjin@islab.re.kr, hyy0238@pusan.ac.kr, howonkim@gmail.com

Malware Classification System Using Deep Learning and Imaging

Jo Hyun Jin*, Hong Yoon Yung*, Kim Ho Won*
*Pusan National Univ.

요 약

최근 분산서비스거부(DDoS) 공격, 랜섬웨어, 개인정보 유출, 해킹 등 악성코드로 인한 피해가 급증하고 있다. 특히 기존 악성코드를 분석하던 기법으로는 은닉기술, 자가변형 등 발달한 분석 방해 기법으로 지능화된 악성코드들의 분석이 어려운 실정이며 악성코드의 종류가 다양해지고 있어 새로운 형태의 악성코드가 많이 등장하고 있다. 본 논문에서는 딥러닝 모델과 이미지화된 악성코드를 이용하여 악성코드를 분류하는 시스템을 제안한다. 시스템에서 사용된 딥러닝 모델인 샴 네트워크는 유사도를 기반으로 분류를 하여 새로운 형태의 악성코드에 대해서도 비슷한 악성코드 패밀리로 분류가 가능하다. 또한 적은 데이터로도 학습이 가능한 One-shot learning이 가능하여 적은 양의 데이터를 가진 악성코드 패밀리에 대해서도 분류가 가능하다. 제안하는 시스템은 평균 97.5%의 정확도로 악성코드를 분류하며 데이터가 적은 악성코드 패밀리에 대해서 약 90%의 정확도로 분류하며 일반적인 CNN(Convolutional Neural Network)과 비교하여 좋은 성능을 보인다.

Ⅰ. 서 론

최근 연구에서는 이미지 인식에 뛰어난 성능을 보이는 딥러닝을 이용하여 이미지화된 악성코드를 분석하는 기법들이 제안되고 있다 [1] [2]. 이미지화된 악성코드들은 같은 패밀리 간 유사성이 높게 나오기 때문에 이미지 인식을 하는 딥러닝으로도 높은 분류 성능을 확인할 수 있다. 하지만 기존 이미지 인식 모델들은 소량의 데이터로 훈련된 악성코드에 대해서 유사한 악성코드 패밀리가 있더라도 제대로 분류하지 못 하는 결과를 보인다. 이에 본 논문에서는 유사도 기반의 분류를 하는 샴 네트워크를 악성코드 분류에 적용하여 기존 패밀리와 유사한 형태의 소량의 악성코드 패밀리에 대해서도 악성코드로 분류되도록 하는 기법을 제안한다.

Ⅱ. 본 론

2.1 샴 네트워크

Gregory 등[3]은 하나의 데이터로만 학습하여도 정확한 추론이 가능한 One-shot learning 기반의 샴 네트워크를 제안하였다. 샴 네트워크는 동일한 파라미터를 공유하는 트윈 네트워크(Twin-network)를 포함하는 딥러닝 모델이다. 그림 1과 같이 트윈 네트워크 중 상위 네트워크는 입력으로 검증받고자 하는 이미지가 주어지고, 하위 네트워크는 입력으로 검증용 참조 이미지가 주어진다. 트윈 네트워크는 두 이미지로부터 특징 벡터(Feature vector)를 추출하고 특징 벡터간의 L_1거리를 계산하여 벡터 간의 유사도를 계산하는 Prediction layer로 전달한다. Prediction layer에서 특징 벡터 간의 유사도를 구하는 식은 식 1과 같다.

$$\sum_j \left(\alpha_j \left| h_1^{(j)} - h_2^{(j)} \right|\right) \quad (1)$$

h_1과 h_2는 상위 네트워크와 하위 네트워크에서 출력된 각각의 특징 벡터, α_j는 특징 벡터의 j번째 특징 간 L_1거리에 곱해지는 Prediction layer의 파라미터이다. Prediction layer는 특징 벡터 간의 유사도에 시그모이드 함수를 적용하여 0과 1사이의 확률을 출력한다. 그림 1과 같은 구조로 입력 이미지가 참조 이미지와 동일한 클래스이면 특징 벡터 간의 유사도가 커지도록, 다른 클래스면 유사도가 작아지도록 학습을 하여 새로운 데이터에 대해서 재학습 없이 높은 성능의 분류가 가능하다.

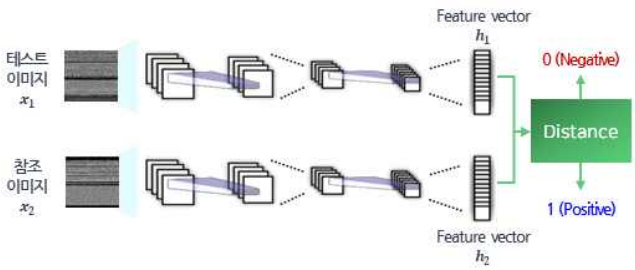

그림 1. 시스템 구조

2.2 시스템 구조

본 논문에서 제안하는 시스템의 구조는 그림 1과 같다. 데이터세트는 Microsoft Malware Dataset [4]를 사용하였다. 해당 데이터세트는 총 9가지의 악성코드 패밀리로 구성이 되어있다. 각 데이터는 Hex값으로 구성된 바이너리 덤프 파일을 가지고 있으며 이를 0은 검정색, 255는 흰색으로 변경하는 Gray scale 이미지화를 적용한다. 샴 네트워크의 트윈 네트워크는 MobileNet V2 [5]로 구성하였다.

2.3 실험 결과

본 장에서는 샴 네트워크의 성능을 기존 이미지 분류 딥러닝 모델 [6]과 비교하여 성능을 검증하였다. 동일한 데이터세트로 학습한 2개의 모델

의 정확도를 비교하였으며 각 클래스 별 Confusion Matrix로 성능을 비교하였다. 사용된 데이터세트의 문제점은 특정 클래스의 데이터가 다른 클래스에 비해 부족한 점이다. 해당 클래스는 Simda 패밀리 클래스이며 기존 이미지 분류 모델에서는 해당 클래스의 분류 성능이 낮게 나온다.

```
Class Name : [1], a number of Class : 1541,
Class Name : [2], a number of Class : 2148,
Class Name : [3], a number of Class : 2937,
Class Name : [4], a number of Class : 455,
Class Name : [5], a number of Class : 42,
Class Name : [6], a number of Class : 751,
Class Name : [7], a number of Class : 389,
Class Name : [8], a number of Class : 1222,
Class Name : [9], a number of Class : 1013,
```

그림 2 . 데이터세트의 각 클래스 분포

샴 네트워크는 One-shot learning이 가능하여 작은 클래스의 데이터에 대해서도 좋은 성능의 분류가 가능하다. 그림 3과 그림 4는 기존 이미지 분류 모델과 샴 네트워크의 Confusion matrix이다. 5번째 패밀리에 대해서 일반 분류 모델은 58%의 정확도를 보이는 반면 샴 네트워크는 90%를 보이는 점을 확인할 수 있다.

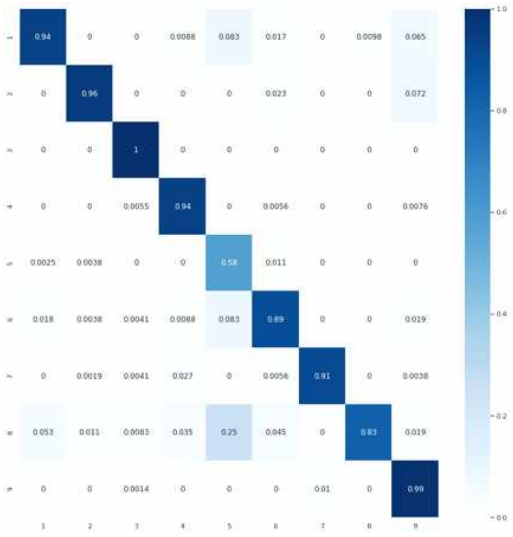

그림 3 . 기존 이미지 분류 모델의 Confusion matrix

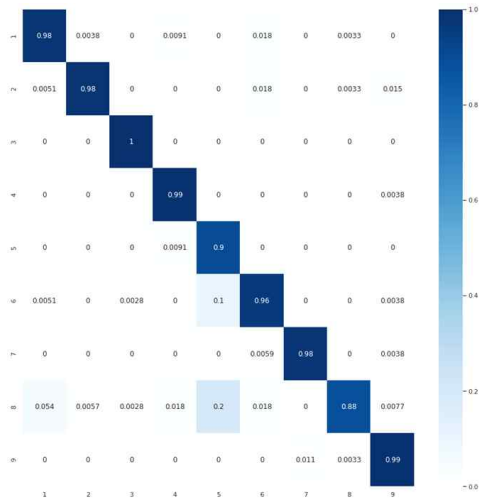

그림 4 . 샴 네트워크의 Confusion matrix

Ⅲ. 결론

본 논문에서는 샴 네트워크를 사용하여 소량의 악성코드 패밀리에 대해서도 높은 성능으로 분류가 가능한 기법을 제안하였다. 새로운 형태의 악성코드들이 많이 생기고 있으며 이에 소량의 데이터만 확보되어도 정확한 분류가 가능한 샴 네트워크를 악성코드 분류에 적용하여 높은 성능을 보이는 점을 검증하였다. 추후 악성코드의 이미지화 방법을 다양하게 하여 각 성능을 검증할 예정이다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2019-0 00097, 스마트공장 네트워크 인프라용 보안 칩 및 실시간 제어프로토콜 보안 기술 개발)

참 고 문 헌

[1] 조영복. "딥러닝 기반의 R-CNN 을 이용한 악성코드 탐지 기법." 한국디지털콘텐츠학회 논문지 19.6 (2018): 1177-1183.

[2] 석선희, and 김호원. "Convolutional Neural Network 기반의 악성코드 이미지화를 통한 패밀리 분류." 정보보호학회논문지 26.1 (2016): 197-208.

[3] Koch, Gregory, Richard Zemel, and Ruslan Salakhutdinov. "Siamese neural networks for one-shot image recognition." ICML deep learning workshop. Vol. 2. 2015.

[4] Ronen, Royi, et al. "Microsoft malware classification challenge." arXiv preprint arXiv:1802.10135 (2018).

[5] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[6] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 (2012): 1097-1105.