머신러닝을 이용한 악성코드 분류

이길홍*, 김경신**
*서울과학기술대학교 컴퓨터공학과
**청강문화대학교 모바일학과
e-mail:khlee@seoultech.ac.kr*,kskim@ck.ac.kr**

A Malicious Code Classification using Machine Learning

Kilhung Lee*, Kyeong-Sin Kim**
*Dept. of Computer Science, Seoul National University of Science and Technology
**Dept. of Mobile Engineering, Chunggang Munwha Industrial University

요 약

머신러닝 기법을 다양한 분야에 사용되는 연구가 한창이다. 본 논문에서는 악성 코드의 분류 시스템에 머신러닝 기법을 적용하였다. 악성 코드 파일을 적당한 크기로 이미지화하여 텐서 플로우의 인셉션 V3에 적용하였다. 실험 결과, 이미지의 사이즈 조정과 파라미터 조정을 통해 매우 만족할 만한 수준으로 악성 코드를 잘 분류함을 확인할 수 있었다.

1. 서론

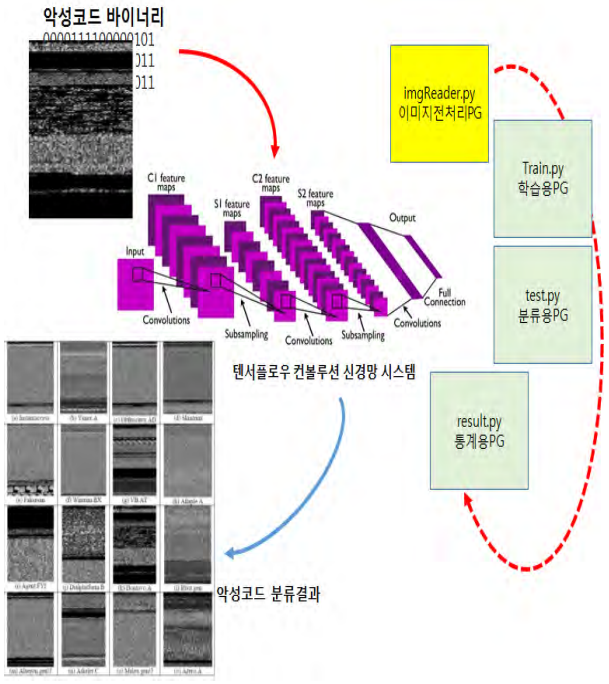
2016년 3월 이세돌과 알파고 대전이후 머신러닝과 딥러닝의 관심은 폭발적으로 상승하고 있다. 가트너는 주목할 10대 기술로 인공지능과 머신러닝을 언급하였다. 로봇, 인공지능은 물론 거의 모든 분야에 이용될 수 있다[1].

머신러닝은 정보보안 분야에 적용 시 매우 효율적인 결과를 낼 수 있는 분야중의 하나이다[2-4]. 네트워크 침입 탐지, 악성코드의 분석, 취약점 분석 등에 머신러닝 방식이 효과적으로 사용될 수 있다[5-8]. 예전에는 지금처럼 악의적인 사이버공격이 그리 다양하지 않았고, 패턴매칭 방식만으로도 침입 탐지나 공격분석이 가능했지만, 현재는 스마트폰, 스마트카, 스마트홈, 스마트팩토리, 스마트그리드 등 사회가 전반적으로 사이버지능화 되어가면서 이에 따른 사이버공격도 지능화되고 다각화 되므로 이를 대응하기 위해서는 머신러닝과 정보보안을 접목한 새로운 기술이 필요하다.

2. 머신러닝 기법을 이용한 악성 코드 분류

본 논문에서는 콘볼루션 신경망 오픈소소인 구글의 텐서플로우를 사용하여 악성코드의 학습 및 분류기를 개발하였다. 기본형 텐서플로우 신경망 라이브러리와 본 연구를 통해 개발한 악성코드 이미지화 프로그램을 포함한 악성코드의 학습과 분류기의 전체 시스템 구조도는 (그림 1)과 같다.

먼저 악성코드 바이트코드를 이미지화하는 작업을 수행한다. 이 프로그램은 설정 과정을 통해 일정 크기의 이비지로 변환하는 과정이다. 이 과정에서 이미지의 크기에 따라 분류 정확도가 영향을 받는다.



(그림 1) 악성코드 분류 시스템 구조도

다음으로 학습 프로그램을 수행한다. MS데이터 셋의 경우 10,868개의 학습용 악성코드 바이트코드와 asm코드가 제공된다. 본 연구에서는 이 10,868개의 악성코드를 통해 학습하였다. 이 과정은 수행하는 컴퓨터 시스템 성능에 매우 영향을 받는다. 이 과정이 끝나면 텐서플로우 콘볼루션 신경망 학습결과 파일이 생성된다.

학습을 완료하면 생성된 콘볼루션 신경망을 활용하여, 테스트 파일에 대해서 분류를 시험한다. 시험 결과에 따라

악성 코드를 분류하고, 정확히 분류하였는가에 대한 통계를 내서 시스템의 정확도를 계산한다.

3. 실험 및 결과

개별 악성코드 파일별 혹은 파일단위별 분류의 정확도와 각종 수행간의 통계사항을 정리하여 정리한 결과를 (표 1)에 보였다.

실험은 크게 두 가지 방식으로 진행하였다. 스크래치(Scratch) 방식은 이미지를 학습하고 그 결과에 새로운 이미지를 평가하는 기본적인 CNN 방식의 학습 방식인 반면, 이러한 신경망을 이용한 이미지 학습의 정확도를 높이기 위해 최근 나온 방식이 파인튜(Fine tune) 방식이다. Fine tune 방식은 신경망으로 학습결과를 남기면서 체크포인트를 남겨서 이전에 학습한 내용을 또 다시 학습하는 방식을 말한다. 후자가 당연히 정확도가 높으리라 예상하겠지만 이미지의 형태에 따라서 또는 복잡성에 따라서 스크래치 방식이 더 정확도가 높은 경우도 나온다. 따라서 본 연구에서는 Scratch 방식과 Fine tune방식을 번갈아가면 실험하여 정확도를 측정하였다.

<표 1> Scratch 방식과 Fine tune 방식의 악성코드 분류 결과

구 분	실험 방식 및 결과	
Method	Scratch	Fine tune
Validation	600	600
Steps	1000 / 500	1000 / 500
Batch size	64	64
Learning rate	0.01 / 0.001	0.01 / 0.001
Weight decay	0.00004	0.00004
Model	InceptionV3	InceptionV3
Result	62%	85%

4. 결론 및 연구과제

본 연구에서는 머신러닝 기법을 이용하는 악성코드 분류 시스템을 구현하였다. 구글의 텐서플로우 인셉션 라이브러리를 활용하였고, 학습 방식은 Scratch 방식과 Fine tune 방식을 서로 비교하였다. MS 캐글 데이터 셋 10,868개를 통해 Scratch 방식은 62%, Fine tune 방식은 85%의 정확도를 기록하였다. 이러한 결과를 바탕으로, 이미지 변환 및 처리 과정과 파라미터 튜닝을 통해서 정확도를 좀 더 높인다면, 머신러닝을 활용한 악성 코드 분류 시스템은 그 결과의 정확도와 유용성 면에서 요구조건을 만족시킬 수 있는 효과적인 방안이 될 수 있음을 확인할 수 있었다.

참고문헌

[1] Z. Bazrafshan, H. Hashemi, S. M. H. Fard, A. Hamzeh, "A survey on heuristic malware detection techniques", Proc. 5th Conf. Inf. Knowl. Technol. (IKT), pp. 113-120, 2013.

[2] Tian, R. Batten, L.M. and Versteeg. S.C., "Function length as a tool for malware classification. 3rd International Conference on Malicious and Unwanted Software (MALWARE), 2008.

[3] Tian, R. Batten, L. Islam, R. and Versteeg, S., "An automated classification system based on the strings of trojan and virus families", 4th International Conference on Malicious and Unwanted Software: MALWARE 2009, pp. 23-30.

[4] Park, Y. Reeves, D. Mulukutla, V. Sundaravel, B., " Fast malware classification by automated behavioral graph matching", Proc. Of Sixth Annual Workshop on Cyber Security and Information Intelligent Research (CSIIRW'10), 2010.

[5] Islam, R., Tian R., Batten, L., Versteeg, S., " Classification of Malware Based on String and Function Feature Selection", 2nd Cybercrime and Trustworthy Computing Workshop, 2010.

[6] 민승욱. 조형진. 신진섭. 류재철, "머신러닝 기법을 이용한 안드로이드 악성코드 탐지 기법", 한국정보과학회, 한국정보과학회 학술발표논문집 39(1C), 2012.6, pp. 280-282.

[7] 석선희. 김호원, "Convolutional Neural Network 기반의 악성코드 이미지화를 통한 패밀리 분류", 정보보호학회논문지 제26권 제1호, 2016.2, pp. 197-208.

[8] 김혜정. 윤은준, "악성코드로부터 빅데이터를 보호하기 위한 이미지 기반의 인공지능 딥러닝 기법", 전자공학회논문지 제54권 제2호, 2017.2, pp. 76-82.