머신러닝 기반의 키보드 트리거 감지에 관한 연구*

박연균, 윤종희
영남대학교 컴퓨터공학과
pyk2181@gmail.com, youn@yu.ac.kr

A Study on Detection of a Keyboard Trigger Based on Machine Learning

Yeon-Kyun Park, Jonghee Youn
Dept. of Computer Engineering, Yeungnam University

요 약

대부분의 컴퓨터 사용자는 키보드를 사용하여 정보를 입력한다. 이때 키보드와 같은 입력장치에 관련된 트리거가 발생할 수 있는데, 키보드 보안과 관련해 키보드 트리거에 대해 정의하고 머신러닝의 분류 모델을 통해 이를 감지해봄으로써 사용된 두 모델 간의 성능을 비교해 보고자 한다.

1. 서론

현대사회에서 우리는 컴퓨터에 많은 것을 의존하고 있다. 간단한 문서 작업부터 시작해 금융업무와 같은 보안 중요도가 높은 일도 컴퓨터를 통해 해결한다. 이렇듯이 컴퓨터와 관련된 업무에서 보통의 사용자는 입력장치를 사용하여 정보를 입력하는데, 이때 트리거가 존재할 수 있다.

트리거는 정해진 특수한 조건을 만족할 경우 지정된 시점에 작동하도록 설계되어 있다. 일반적으로 가상 머신의 존재 유무, 아키텍처, 시간 등을 기준으로 삼고 있으며 사람과의 상호작용에 관련된 트리거의 경우 키보드, 마우스와 같은 입력장치를 통해 작동된다. 본 논문에서는 키보드 트리거에 대해 정의하고, 머신러닝의 분류 모델을 통해 이를 감지하여 모델 간의 성능을 비교하고자 한다.

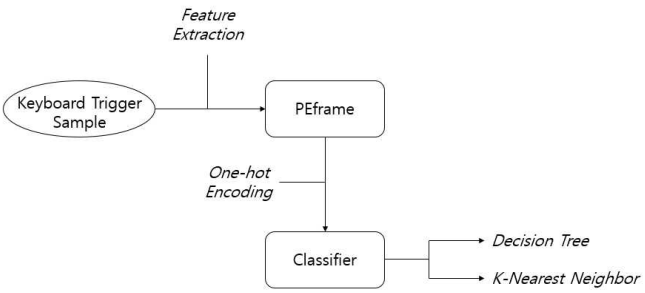
2. 본론

2.1 키보드 트리거

키보드 트리거는 무엇을 기준으로 하는가에 따라 적용될 수 있는 범위가 매우 넓다. 좁게는 키로거(Keylogger)와 같은 악성코드에 포함되어 사용자의 아이디나 비밀번호 등의 개인 정보를 탈취하기 위해 사용되는 경우로 한정하기도 하지만, 키보드에 의한 입력이 발생하는 경우 자체를 트리거로 간주할 수 있기 때문이다. 본 연구에서는 키로거가 아니라고 해서 트리거 역시 존재하지 않는다고 확신할 수 없다는 점을 고려하여 키보드 트리거를 사용자의 입출력 존재 여부에 따라 넓은 범위에서 판별하기로 하고, 입출력이 감지되는 경우 키보드 트리거가 존재한다고 정의하였다.

2.2 연구 개요

키보드 트리거 감지를 위한 전체적인 연구 진행 과정은 그림 1과 같다.



(그림 1) 키보드 트리거 감지 연구 과정

머신러닝은 목적에 따라 지도 학습(Supervised Learning), 비지도 학습(Unsupervised Learning), 강화 학습(Reinforcement Learning) 등으로 분류된다[1]. 이 중에서 키보드 트리거를 감지하기 위해 지도 학습의 분류 모델을 사용하여 두 모델의 결과를 비교 분석하고자 했으며. 이를 위해 100개의 샘플을

* 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2018R1D1A1B07050647)

수집해 Feature 추출을 진행하였다.

2.3 PEframe

Feature는 머신러닝의 성능에 더없이 큰 영향을 미친다[2]. 똑같은 샘플에서 특징을 추출했다고 해도 어떤 특징을 사용하는지, 서로 다른 특징을 어떻게 조합하는지에 따라 그 결과가 달라지기 때문이다.

키보드 트리거 감지를 위한 Feature는 각 샘플에서 호출되는 전체 API 목록으로 정하고, 이를 추출하기 위해 PEframe[3]을 사용하였다. PEframe은 PE 파일 정적 분석을 위한 파이썬 기반의 오픈 소스 도구로, API 호출을 비롯해 PE 파일에 대한 다양한 정보를 확인할 수 있다. 이렇게 추출된 API 호출 데이터는 One-hot Encoding을 거쳤으며, 결과적으로 총 2017개의 Feature가 생성되었다.

	trigger_index	CertCloseStore	CertFindCertificateInStore	CertFreeCertificateChain	CertFreeCertificateContext	CertGetNameStringW
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	1	0	0	0	0	0
3	0	0	0	0	0	0
4	1	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0
9	0	0	0	0	0	0
10	1	0	0	0	0	0
11	1	0	0	0	0	0

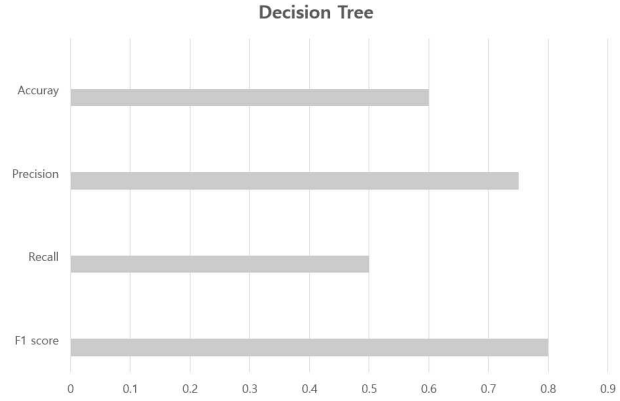
(그림 2) PEframe을 통한 API 호출 목록

2.4 머신러닝 모델 구축

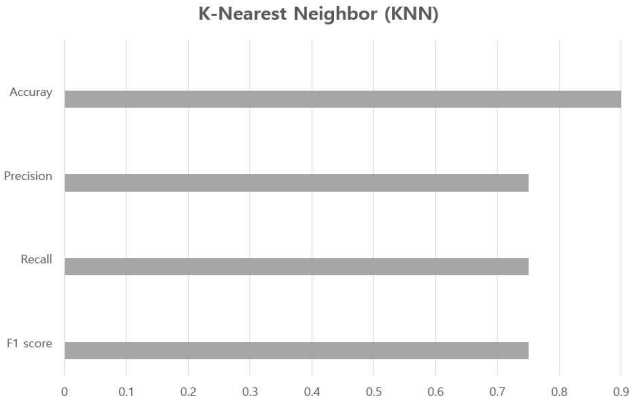
앞서 추출한 Feature로 구성된 데이터셋에서 훈련 데이터와 학습 데이터를 나누었으며, Decision Tree와 K-Nearest Neighbor 두 모델은 머신러닝에서 사용빈도가 높아 이 데이터를 학습시키기 적절하다고 판단하여 선정되었다. Decision Tree는 목적에 맞게 분류 기준을 설정하여 이에 따라 데이터를 구분하며, K-Nearest Neighbor는 유사한 특성을 가진 데이터의 경우 유사한 범주에 속할 가능성이 크다는 점을 이용하는 거리기반 분류 모델이다.

2.5 실험

키보드 트리거를 감지하는 각 모델의 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 score 값을 통해 성능을 평가하여 그래프로 나타내었다. Decision Tree의 성능이 전반적으로 좋지 않은 수치를 보이는 것과는 반대로 K-Nearest Neighbor의 경우 90%의 정확도를 보였다. 즉, API 호출에 따른 Feature를 사용해 만든 데이터셋에서 Decision Tree보다 K-Nearest Neighbor 모델이 더 우수한 것으로 나타났다.



(그림 3) Decision Tree 모델 성능 평가



(그림 4) K-Nearest Neighbor 모델 성능 평가

3. 결론

본 논문에서는 키보드 트리거와 관련된 샘플을 수집하고 One-hot Encoding을 거친 API 호출 목록에 대한 Feature로 데이터셋을 구성하였으며, 이 데이터를 이용해 Decision Tree와 K-Nearest Neighbor 지도 학습 모델을 통해 학습시켜 키보드 트리거의 존재 여부에 따라 분류 및 비교해 보았다.

향후에는 본 연구에서 사용되지 않은 다양한 머신러닝 모델을 적용하고, API 호출 이외의 여러 Feature를 조합해 봄으로써 더 효과적인 데이터셋을 구성해 정확도를 높이는 연구로 확장해 나갈 것이다.

참고문헌

[1] 이동근. “기계학습 기반의 악성코드 탐지기법 분석.” 국내석사학위논문 순천향대학교 대학원, 2018.
[2] 변지윤, 김대호, 김희철, 최상용. “머신러닝 기반 악성코드 분류를 위한 재귀적 속성 추출 알고리즘.” 한국컴퓨터정보학회논문지, 26(2), 61-68. 2021.
[3] “PEframe”, https://github.com/guelfoweb/peframe