Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.30, NO.4, Aug. 2020 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2020.30.4.647

자료 전송 데이터 분석을 통한 이상 행위 탐지 모델의 관한 연구

손 인 재,[1†] 김 휘 강[2‡]
[1,2]고려대학교 정보보호대학원(대학원생, 교수)

A Study on the Abnormal Behavior Detection Model through Data Transfer Data Analysis

In Jae Son,[1†] Huy Kang Kim[2‡]
[1,2]Korea University School of Cybersecurity(Graduate student, Professor)

요 약

최근 국가·공공기관 등 중요자료(개인정보, 기술 등)가 외부로 유출되는 사례가 증가하고 있으며, 조사에 따르면 정보유출 사고의 주체로 가장 많은 부분을 차지하고 있는 것이 대부분 권한이 있는 내부자로써 조직의 주요 자산에 비교적 손쉽게 접근할 수 있다는 내부자의 특성으로 외부에서의 공격에 의한 기술유출에 비해 보다 더 큰 피해를 일으킬 수 있다.

이번 연구에서는 업무망과 인터넷망의 분리된 서로 다른 영역(보안영역과 非-보안영역 등)간의 자료를 안전하게 전송해주는 망간 자료전송시스템 전송 로그, 이메일 전송 로그, 인사정보 등 실제 데이터를 이용하여 기계학습 기법 중 지도 학습 알고리즘을 통한 이상 행위 탐지를 위한 최적화된 속성 모델을 제시하고자 한다.

ABSTRACT

Recently, there has been an increasing number of cases in which important data (personal information, technology, etc.) of national and public institutions are leaked to the outside world. Surveys show that the largest cause of such leakage accidents is "insiders." Insiders of organization with the most authority can cause more damage than technology leaks caused by external attacks due to the organization. This is due to the characteristics of insiders who have relatively easy access to the organization's major assets.

This study aims to present an optimized property selection model for detecting such abnormalities through supervised learning algorithms among machine learning techniques using actual data such as CrossNet data transfer system transmission log, e-mail transmission log, and personnel information, which safely transmits data between separate areas (security area and non-security area) of the business network and the Internet network.

Keywords: Insider Threat, Supervised Learning, Multilayer Perceptron, Classification, Weka

I. 서 론

1.1 연구 배경

최근 국가·공공기관 등 중요자료(개인정보, 기술 등)가 외부로 유출되는 사례가 증가하고 있으며, 정보보호 실태조사 보고서에 따르면 침해사고 피해 경험 유형으로는 '랜섬웨어'가 56.3%로 가장 많았고, 다음으로 '악성코드(컴퓨터 바이러스, 웜, 트로이잔

Received(05. 07. 2020), Modified(06. 30. 2020), Accepted(07. 22. 2020)

† 주저자, injae0203@korea.ac.kr
‡ 교신저자, cenda@korea.ac.kr(Corresponding author)

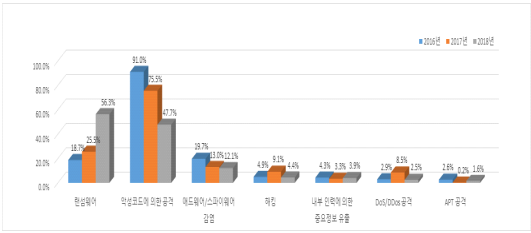

Fig. 1. Types of Intrusion in Accident Damage Experience

등)에 의한 공격(47.7%)', '애드웨어/스파이웨어 감염(12.1%)', '해킹(4.4%)', '내부 인력에 의한 중요정보 유출(3.9%)' 등의 순으로 조사되었다[1].

이러한 내부 직원들이 주체가 된 위협을 '내부자 위협'이라고 하는데, 여기서 말하는 내부자란 조직 내의 시설, 자산, 인원, 시스템에 대해 정상적인 접근권한을 가지고 있는 인원을 말한다. 이들은 외부에 있는 다른 공격자들 보다 더 쉽게 조직 내 중요 자산에 접근할 수 있는 이점이 있기 때문에, 좀 더 작은 노력으로 조직에 중대한 피해를 입힐 수 있다.

1.2 연구 목적

최근 국가·공공기관 등 중요자료가 외부로 유출되는 사례가 증가함에 따라 각 기관은 업무망과 인터넷망을 물리적 또는 논리적으로 분리하여 인터넷을 통해 업무관련 정보에 접근하는 것을 차단함으로써 안전한 업무환경을 구축하고 있다. 그러나, 업무망과 인터넷망을 분리할 경우 인터넷에서 획득한 정보를 업무망으로 전송하거나 업무자료의 외부반출 등 분리된 전산망간에 자료전송이 요구되는 경우 국가·공공기관에서는 분리된 전산망간 자료를 전송하기 위해서는 망간 자료전송시스템을 도입하여 운용 해야만 한다[2].

본 논문에서는 날로 심각해지는 내부자 위협 중에서도 가장 큰 위협인 정보유출을 탐지 및 예측하기 위한 방안을 제안한다. 정보유출 탐지 및 예측을 하기 위해 망간 자료전송시스템 로그, 이메일 전송 로그, 인사정보를 이용하여 내부자의 이상 행위 정보를 탐지하고 분석 하였으며 지도 학습 알고리즘을 통한 이상 행위 탐지를 위한 최적화된 속성 모델을 제시하고자 한다.

1.3 논문 구성

본 연구의 구성은 다음과 같다. 2장에서는 내부자 위협에 관한 선행 연구에 대하여 살펴보고, 3장에서는 망간 자료전송시스템 로그, 이메일 전송 로그, 인사정보 데이터를 가지고 내부자 유형에 따라 로그를 분류하고 데이터 전처리를 수행한다. 4장에서는 지도 학습 알고리즘을 이용하여 매개변수에 따른 분류된 데이터 그룹 간에 영향도를 분석하고 결과 데이터 중 가장 높은 정확도와 유사성을 판단하며, 최적화된 속성 모델에 대한 검증을 실시한다. 5장에서는 결론과 향후 연구 방향에 대하여 서술하였다.

II. 관련 연구

2.1 내부자 위협

2.1.1 내부자 위협 정의

내부자 위협(Insider Threat)은 특정 취약점을 이용할 수 있는 가능성을 의미한다[3]. 반면 내부자 공격(Insider Attack)은 조직의 보안 정책을 위반하는 이벤트 또는 액션의 집합을 의미한다. 내부자 공격은 내부자에 의해 실제 수행되는 오용 자체이고 성공 또는 실패여부와는 무관하다. 내부자에 대한 정의를 기반으로, 내부자 사이버 위협은 합법적인 접근권한을 가진 조직이 신뢰하는 사람이 취약점을 이용하여 조직의 보안 정책을 위반하는 사람으로 정의할 수 있다[4].

2.1.2 내부자 위협 탐지 연구

내부자 정보 유출에 대해 사회적으로 관심이 많아지면서 다양한 분야에서 내부자의 이상 행위 탐지 방안에 대한 연구가 수행되고 있다. 기업의 중요한 지적 자산을 안전하게 보호하기 위해 내부 인증 사용자의 이상 행위를 탐지할 수 있는 방법으로 인증 사용자의 시스템 사용 형태를 확인할 수 있는 주요 변수들을 K-Means 및 SOM 알고리즘을 사용하여 데이터를 군집화하는 데이터 마이닝 기반의 탐지 모델을 제안하였다[5]. 사전에 등록된 유형으로 탐지되지 않아 정상적인 사용으로 분류되는 행위 이벤트들을 분석하여 이상 행위 징후를 탐지하고 CBR을 활용한 이상 탐지 모델 제시를 통한 대응 방법을 제시

하였다[6]. 내부 구성원이 정보시스템을 사용할 때 기록되는 로그를 사용자의 행위를 기반으로 일 단위 인스턴스화하여 특정 기간 동안에 사용자의 여러 행위를 발생 빈도로 요약하고 수치화된 벡터로 표현하는 내부 사용자 이상 행위 모델링 기법을 제시하였다[7]. 비지도학습 알고리즘의 신뢰성 및 정확도를 향상시키기 위해 SMOTE를 사용하여 다양한 데이터 셋을 학습 하고 비지도 학습이 대량의 데이터를 처리하는 예산과 자원을 최소한으로 투자하여 데이터를 가공하는데 효율적이라는 것을 제시하였다[8]. 내부 침입자 이상 탐지를 위해 비지도 학습 앙상블 기반의 알고리즘을 학습하여 내부 침입자의 특정 정보를 가지고 있는 데이터 스트림의 분류 정확도를 향상시키는 방법을 제안하였다[9]. 정상·비정상 소스가 함께 혼합되어 있는 실제 데이터 셋을 기계 학습을 통해서 이상치를 확인하고 추적 하였으며 본인이 속한 그룹과 다른 행동을 보이는 그룹에서 이상 행위에 대한 탐지의 정확성과 유연 적응성을 향상시키는 결과를 제시하였다[10].

2.2 기계 학습

데이터 마이닝은 데이터베이스(Database)에 저장되어 있는 수많은 양의 데이터로부터 각 데이터의 유용한 의미를 가진 상간관계 정보를 추출하여 분석하는 방법이며 기계 학습은 전반적인 데이터 마이닝의 기술적인 환경을 제공하고 있다. 학습에는 지도학습, 비지도학습, 강화학습이 있다[11].

머신 러닝은 학습 데이터(Training Data)에 종속변인(Label)이 있는 경우와 없는 경우로 나누어 학습 방법을 구분한다. 종속변인이 있는 경우 학습 방식을 지도 학습(Supervised Learning)으로 설명하고, 종속변인이 없는 경우 학습 방식을 비지도 학습(Unsupervised Learning)으로 설명한다. 지도 학습 방식은 분류(Classification)와 예측(Prediction) 알고리즘을 활용하여 예측 모형을 개발한다. 비지도 학습 방식은 군집(Clustering) 알고리즘을 활용하여 예측 모형을 개발한다. 알파고(AlphaGo)의 학습 방식으로 유명한 강화 학습(Reinforcement Learning)은 지도 학습 방식에 포함하여 분류하기도 하지만, 대개 독립적인 머신러닝과는 다른 개념으로 구별하기도 한다. 강화 학습은 스스로 생산한 데이터를 피드백하여 예측 모형을 진화시켜 나가는 방식이다. 구글 딥마인드 대표인 데미스하사비스(Denis Hassabis)는 알파고 2.0을 학습시키는 과정에서 바둑의 규칙만을 입력했을 뿐, 인간의 기보는 입력하지 않았다는 인터뷰를 하였다. 즉, 사전에 학습용 데이터를 입력하지 않고 규칙만을 입력하여 가상의 데이터를 스스로 생산하고, 이렇게 생산된 데이터를 활용해 학습하고 진화한다는 의미이다[12].

2.3 지도 학습

지도학습(Supervised Learning)은 훈련용 데이터(Training Data)로부터 하나의 함수를 유추하기 위한 기계 학습(Machine Learning)의 일부 방법이다. 훈련 데이터는 일반적으로 Input 객체에 대한 속성을 벡터 형태로 가지고 있으며 각각의 벡터에 대해 원하는 결과가 무엇인지 명시되어 있다. 이렇게 유추된 함수 중 연속적인 값을 Output 하는 것을 회귀분석(Regression)이라 하고 주어진 Input 벡터가 어떤 종류의 값인지 표식 하는 것을 분류(Classification)라고 한다. 지도 학습기(Supervised Learner)가 하는 작업은 훈련 데이터로부터 주어진 데이터에 대해 예측하고자 하는 값을 올바로 추측해내는 것이다[13].

지도 학습 기법은 일반적으로 예측 및 분류의 대상이 되는 목적변수와 이를 설명하고, 패턴을 찾기 위한 설명변수 그리고 이들을 훈련시킬 훈련용 데이터와 향후 적용 및 정확도 계산 등을 위한 테스트 데이터 등으로 구성이 되어 진다. 지도학습 기법의 가장 핵심적인 부분 중 하나는 바로 훈련용 데이터에서 설명변수를 이용하여, 목적변수를 맞추어 가는 패턴을 인식하는 과정이 알고리즘화 되어있는 예측자(Predictor)이다. 좀 더 자세히 설명하면, 예측자라는 것은 지도학습 기법들을 이용하여 특히 훈련 용(Training) 데이터를 이용하여 생성되어지는 예측 로직(Logic)을 의미한다. 이러한 지도 학습 기법을 이용하여 만들어진 모델 또는 예측자는 성능평가 등을 통해서 향후 모델의 변경 및 조정 작업이 이루어진다. 지도학습 기법의 성능 평가방법은 여러 가지가 있을 수 있다. 가장 대표적인 것이 각종 분류 및 예측의 정확도(Accuracy)가 있고, 이 외에도 수행 속도, 강건성(Robustness), 확장성(Scalability) 그리고 해석력(Interpretability) 등이 있다[14].

2.4 다층 퍼셉트론(Multilayer Perceptron)

퍼셉트론은 딥러닝의 기원이 되는 알고리즘으로서 다수의 신호를 입력으로 받아 하나의 신호를 출력한다. 퍼셉트론이 동작하는 방식은 인간의 뇌 구조와 매우 유사한데, 각 입력 값과 가중치의 곱을 모두 합한 값을 활성화 함수를 통해 판단하여 해당 뉴런을 활성화 할지 하지 않을 지를 결정 한다[15].

다층 퍼셉트론은 입력층과 출력층 사이에 중간층이 존재 하는데 이를 은닉층(Hidden Layer)라고 하며, 모든 입력값은 은닉층의 모든 노드로 전달되고 은닉층의 모든 출력값 역시 전체 출력층으로 전달 된다. 따라서 가중치와 활성화 함수의 개수도 은닉층의 개수에 따라 증가하게 되어 출력을 위해 계산해야 하는 값들이 늘어나게 되는데, 이러한 복잡한 계산식 덕분에 다층 퍼셉트론은 단층 퍼셉트론이 해결 할 수 없는 다양한 문제들을 해결 할 수 있게 된다[16].

본 연구에서는 13개 입력층과 1개의 출력층으로 구성된 다층 퍼셉트론의 은닉층의 개수를 변경하며 성능 및 정확도에 변화가 있는지 실험을 진행 한다.

III. 연구 데이터 및 전처리

3.1 연구 데이터

본 연구에 사용된 데이터는 망간 자료전송시스템의 2년(2018년~2019년) 동안 사용자 PC 자료를 업무망에서 인터넷망으로 전송한 로그이며, 이메일 전송 로그 및 인사정보 데이터도 포함 되어 있다.

자료 전송 로그는 전송 시간, 사용자, 사용자명, 파일명, 파일 확장자, 파일 사이즈로 구성되어 있으며, 이메일 전송 로그는 전송 시간, 발신자, 발신자명, 수신자, 제목, 첨부 파일명, 첨부 파일 사이즈로 구성되어 있다. 인사 정보는 사번, 성명, 직급, 근속년수, 내부평가 점수1, 내부평가 점수2, 직원구분, 퇴직일자로 구성되어 있다.

3.2 데이터 전처리

실제 데이터 분석을 어떠한 방법으로 데이터 전처리를 수행해야 기계학습의 효과를 얻을 수 있을지 연구되어야 하기 때문에 기계학습을 하기 전 데이터 분석이 필요하다. 데이터 전처리는 불필요한 정보를 사전에 제거하고 기계학습에 사용할 수 있는 데이터 형태로 바꾸는 과정이다. 본 연구의 데이터는 기계학습 알고리즘이 이해할 수 있도록 데이터 전처리를 수행하였으며 분류 데이터는 다음과 같다. 파일 확장자는 분류 알고리즘에 의해 Table 1과 같이 데이터가 분류 되었다. 특이사항으로 미분류 파일의 경우 확장자를 위변조한 파일이 확인 되었다.

Table 1. File Extension Classification

Group	Description
Document	hwp, ppt, xls, ppt, doc, etc.
Image	bmp, jpg, gif, png, tif, etc.
Media	avi, mpeg, swf, wma, etc.
Compression	zip, tar, war, gz, egg, etc.
Program	tmp, sql, bak, iso, dwg, etc.
System	bat, conf, dat, dll, exe, etc.
Etc	der, pfx, p12
Unclassified	exe1, jsp1, htm_, bbb, etc.

파일 확장자 분류 비율은 Table 2와 같이 문서(86.4%), 압축(8.5%), 이미지(4.2%), 미디어(0.5%),프로그램(0.1%), 미분류(0.1%), 시스템(0.1%), 기타(0.0%) 순으로 나타났으며, 파일 확장자 분류 비율 중 두 번째에 해당하는 압축 파일의 경우 파일 사이즈 크기에 따른 분류 비율은 Table 3과 같다.

Table 2. File Extension Classification Rate

Group	Count	Rate
Document	111,779	86.4%
Compression	10,996	8.5%
Image	5,405	4.2%
Media	672	0.5%
Program	163	0.1%
Unclassified	144	0.1%
System	112	0.1%
Etc	36	0.0%

Table 3. Compression File Size Classification Rate

Less than 100M	Over 100M	Over 500M	Over 1G
9,036 (82.18%)	1,960 (17.83%)	769 (7.00%)	403 (3.67%)

파일명 분류는 Python의 여러 형태소 분석 라이브러리 중에서 Twitter 클래스를 이용하여 파일명의 단어를 추출 하였다. KoNLPy는 다음과 같은 다양한 형태소 분석, 태깅 라이브러리를 제공하여 여러 프로그래밍 환경에서 손쉽게 개발할 수 있도록 제공하고 있다[17]. Fig. 2.는 형태소 분석을 위해 사용된 알고리즘으로 파일명을 2개 단어를 기준으로 분석하는 알고리즘의 코드 구조의 일부이다.

이를 활용하여 특정 단어의 정보를 분석하고 추출한 결과 숫자, 영문, 부정확한 단어 등 그룹별로 분류하기 힘든 문제가 확인 되었으나, 최종적으로 대외비, 업무, 개인, 시스템, 기타, 미분류 6가지로 분류하였다. 파일명 분류 비율은 Table 4와 같이 업무(91.6%), 미분류(5.8%), 개인(1.2%), 대외비(0.8%), 기타(0.5%), 시스템(0.1%) 순으로 나타났다.

```
f = open(file, encoding="euc-kr", errors='ignore')
lines = f.read()

from konlpy.tag import Twitter
nlpy = Twitter()
nouns = nlpy.nouns(lines)

from collections import Counter
count = Counter(nouns)

tag_count = []
tags = []

for n, c in count.most_common(20000):

    dics = {'tag': n, 'count': c}

    if len(dics['tag']) >= 2 and len(tags) <= 10000:
        tag_count.append(dics)
        tags.append(dics['tag'])

for tag in tag_count:
    print("{}".format(tag['tag']), end=',')
    print("{}".format(tag['count']))


    print(" {:<14}".format(tag['tag']), end='₩t')
    print("{}".format(tag['count']))
    print("₩n-----------------------------------")
    print(" 명사 총  {}개".format(len(tags)))
    print("-----------------------------------₩n₩n")

return tags
```

Fig. 2. Morphological Analysis Technique Algorithm

Table 4. File Name Classification Rate

Group	Count	Rate
Document	118,483	91.6%
Unclassified	7,526	5.8%
Individual	1,507	1.2%
Restricted Document	1,023	0.8%
Etc	644	0.5%
System	124	0.1%

IV 이상 행위자 탐지 모델

4.1 탐지 모델 구성

본 연구에서 최적화된 모델을 생성하기 위해 제3장에서 데이터 전처리를 통한 자료 전송 로그, 이메일 전송 로그, 인사 정보 데이터를 통해 총 13개의 속성을 추출하였다. 자료 전송 로그에서는 전송 시간(평일 9시~18시), 파일명 분류(대외비, 문서, 개인, 시스템, 기타, 미분류), 확장자 분류(문서, 이미지, 미디어, 압축, 프로그램, 시스템, 기타, 미분류), 파일 크기, 대용량 파일(전송 파일 100M 초과), 파일명 본인 이름(파일명의 본인이름 포함), 이메일 전송 로그에서는 메일 수신자(수신자의 본인 회사 또는 상용 메일 계정) 여부, 인사 정보에는 직급, 년차, 직원 구분(정규직, 계약직), 내부평가1(정보보안 의식평가), 내부평가2(해킹메일 대응 훈련), 퇴직자를 선정 하였다.

대용량 파일은 Table 3과 같이 100M 이상 파일이 전체 압축 파일 대비 17.825%의 많은 비중을 차지하여 속성으로 추가 하였으며, 파일명 본인 이름 속성은 파일명 분류 시 미분류로 분류된 데이터를 통해 나온 특징이다. 그 밖에 파일명 분류, 확장자 분류 속성은 범주형 데이터를 이진 특성의 수치형 데이터로 변환 하였으며, 전송 시간, 대용량 파일, 직원 구분, 메일 수신자, 퇴직자 속성은 해당하는 데이터는 1로 변경해 주고, 나머지는 0으로 채워주었다.

4.2 탐지 모델 분석 방법

본 연구는 탐지 모델 개발을 위해 활용한 머신러닝 분석 알고리즘은 다층 퍼셉트론(Multilayer Perceptron)을 이용하였다. 다층 퍼셉트론은 활성화 함수의 개수가 은닉층의 개수에 따라 증가하게 되어 출력을 위해 계산해야 하는 값들이 늘어나게 되는데, 이러한 복잡한 계산식 덕분에 단층 퍼셉트론이 해결 할 수 없는 다양한 문제들을 해결 할 수 있게 된다[16].

탐지 모델의 독립변수로 선정된 업무 시간, 파일 크기, 파일명 분류, 확장자 분류, 대용량 파일, 파일명 본인 이름, 메일 수신자, 직급, 년차, 직원 구분, 내부평가1, 내부평가2, 퇴직자의 13개 입력층과 1개의 출력층으로 구성된 다층 퍼셉트론의 은닉층의 개수를 변경하며 성능 및 정확도에 변화가 있는지 분

석 한다.

분석 도구로는 뉴질랜드 와이카토 대학교에서 개발한 기계 학습 소프트웨어 제품으로 Weka 3.8 버전을 사용하여 데이터 마이닝 작업을 위한 기계 학습 알고리즘 분석을 진행하였다.

4.3 탐지 모델 학습 데이터 분석

4.3.1 모델 분석 지표

본 연구에서 자료 전송 이상 행위 탐지 모델의 분류정확도는 TP Rate(True Positive Rate : 진양성율), FP Rate (False Positive Rate : 위양성률), Accuracy(정확도), Recall(재현율), Precision(정밀도), F-measure 지표를 활용하여 예측모형별 성능을 측정 분석하고, 예측유효성은 TP Rate, FP Rate, Accuracy, Recall, Precision, F-measure, ROC(Receiver Operating Characteristic) Area 지표를 활용하여 예측모형별 성능을 측정 분석하였다. 예측모형 성능평가 결과 중 최종 판정을 위해 최우선으로 고려하는 분석지표는 예측유효성의 경우 ROC Area 지표이다[18].

4.3.2 다층 퍼셉트론 파라미터

다층 퍼셉트론 분석 알고리즘에 사용하는 파리미터는 Table 5와 같다. 본 연구에서는 Weka에서 기본으로 제공하는 파라미터 디폴트는 -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a 와 같다.

Table 5. Multilayer Perceptron Parameter

Parameter	Definition	Default
L	LearningRate	0.3
M	Momentum	0.2
N	TrainingTime	500
V	ValidationSetSize	0
S	Seed	0
E	ValidationThreshold	20
H	HiddenLayers	a

4.3.3 학습 데이터 모델 분류

학습 데이터 탐지 모델은 선택된 속성 13개의 영향도와 정확도를 파악하기 위해 8개의 모델로 분류하였으며 해당 모델은 Table 6과 같다. Model 1은 모든 속성을 포함하고 있으며, Model 2는 내부평가 제외, Model 3은 퇴직자 제외, Model 4는 직원 구분 제외, Model 5는 대용량 파일 제외, Model 6은 직급 제외, Model 7은 년차 제외, Model 8은 파일명 업무 시간, 파일명 본인 이름 속성을 제외 하였다.

Table 6. Learning Data Model Definition

No.	M1	M2	M3	M4	M5	M6	M7	M8
1	O	O	O	O	O	O	O	X
2	O	O	O	O	O	O	O	O
3	O	O	O	O	O	O	O	O
4	O	O	O	O	O	O	O	O
5	O	O	O	O	X	O	O	O
6	O	O	O	O	O	O	O	X
7	O	O	O	O	O	O	O	O
8	O	O	O	O	O	X	O	O
9	O	O	O	O	O	O	X	O
10	O	O	O	X	O	O	O	O
11	O	X	X	X	X	X	X	X
12	O	X	X	X	X	X	X	X
13	O	O	X	O	O	O	O	O

4.3.4 학습 데이터 모델 수행 결과

본 연구에서는 각 모델별로 다층 퍼셉트론 파리미터 은닉층(Hidden Layer)의 개수(a, 3, 5, 7, 10, 15, 20, 25, 30)에 따라 성능 및 정확도에 변화가 있는지 분석 한다. 학습 데이터 모델은 은닉층의 개수 변경에 따라 성능 및 정확도가 다르게 측정되었으며 전체 모델의 수행 결과는 Table 7과 같다.

Model 1은 Hidden Layer가 5개 일 때 정확도(TP Rate : 0.598)가 제일 높고, Model 2은 3개 일 때 정확도(TP Rate : 0.695)가 제일 높고, Model 3은 5개 일 때 정확도(TP Rate : 0.443)가 제일 높고, Model 4은 5개 일 때 정확도(TP Rate : 0.690)가 제일 높고, Model 5은 a개 일 때 정확도(TP Rate : 0.586)가 제일 높고, Model 6은 20개 일 때 정확도(TP Rate : 0.684)가 제일 높고, Model 7은 15개 일 때 정확도(TP Rate : 0.701)가 제일 높고, Model 8은 20개 일

때 정확도(TP Rate : 0.701)가 제일 높았다.

Model 1과 Model 2를 비교 하였을 때 직원 내부평가(정보보안 의식 평가, 해킹메일 대응 훈련)와는 영향이 없었다. Model 2와 Model 3을 비교 하였을 때 퇴직자 속성이 탐지율의 영향을 크게 주었으며, Model 5의 파일 사이즈 100M 초과 속성의 경우 탐지율을 높이는 효과 보여주었다. Model 4의 직원 구분, Model 6의 직급 속성과는 관련이 있는 걸로 보였으며, Model 7의 년차 속성과 Model 8의 업무 시간, 파일명 본인 이름 속성은 관련이 없는 걸로 확인 되었다.

최종적으로 이상 행위 탐지 최적화된 모델을 선정하기 위해 판단 지표로 TP Rate와 ROC Area 지표이다. TP Rate는 실제로 이상 행위 징후로 판단되는 데이터 셋이 연구 모델에서 이상 행위 징후로 판단되어진 데이터 셋으로 예측한 비율이다. ROC Area 지표는 예측모형 성능평가 결과 중 최종 판정을 위해 최우선으로 고려하는 예측유효성 분석 지표이다[15]. 따라서 Model 7번의 학습 데이터 결과 TP Rate(0.701%), ROC Area(1.000)가 다른 모델보다 정확성이 높다고 평가 할 수 있다.

Table 7. Learning Data Model Performance Result

No.	Layer	Time (sec)	TP Rate	FP Rate	ROC Area
M1	a	109.27	0.523	0.477	0.999
	3	56.45	0.494	0.506	0.999
	5	102.63	0.598	0.402	1.000
	7	134.56	0.523	0.477	0.999
	10	186.44	0.575	0.425	1.000
	15	217.53	0.586	0.414	0.999
	20	280.14	0.569	0.431	1.000
	25	344.44	0.592	0.408	0.999
	30	415.63	0.569	0.431	1.000
M2	a	91.84	0.586	0.414	0.999
	3	51.85	0.695	0.305	0.999
	5	75.48	0.615	0.385	0.999
	7	99.57	0.678	0.322	1.000
	10	136.28	0.603	0.397	1.000
	15	198.98	0.609	0.391	1.000
	20	255.86	0.563	0.437	1.000
	25	336.44	0.598	0.402	1.000
	30	370.8	0.615	0.385	0.999
M3	a	86.56	0.437	0.563	0.926
	3	48.92	0.425	0.575	0.964
	5	71.88	0.443	0.557	0.971
	7	97.31	0.425	0.575	0.925
	10	140.2	0.431	0.569	0.942
	15	197.13	0.431	0.569	0.961
	20	239.51	0.437	0.563	0.914
	25	298.34	0.443	0.557	0.959
	30	353.96	0.437	0.563	0.963
M4	a	86.93	0.626	0.374	0.999
	3	49.48	0.598	0.402	0.999
	5	71.81	0.690	0.310	0.999
	7	98.59	0.592	0.408	0.999
	10	138.93	0.586	0.414	0.999
	15	196.05	0.684	0.316	0.999
	20	245.08	0.684	0.316	0.999
	25	297.68	0.667	0.333	0.999
	30	352.35	0.667	0.333	0.999
M5	a	85.58	0.586	0.414	0.999
	3	50.75	0.569	0.431	0.998
	5	70.58	0.563	0.437	0.999
	7	97.17	0.546	0.454	0.999
	10	143.02	0.471	0.529	0.999
	15	185.77	0.500	0.500	0.999
	20	185.77	0.500	0.500	0.999
	25	302.35	0.500	0.500	0.999
	30	352.34	0.506	0.494	0.999
M6	a	85.36	0.569	0.431	0.999
	3	51.43	0.557	0.443	0.999
	5	77.43	0.603	0.397	0.999
	7	100	0.615	0.385	0.999
	10	130.05	0.615	0.385	0.999
	15	188.68	0.609	0.391	0.999
	20	258.9	0.684	0.316	1.000
	25	294.79	0.684	0.316	0.999
	30	357.87	0.621	0.379	0.999
M7	a	86.67	0.690	0.310	0.999
	3	53.04	0.603	0.397	1.000
	5	72.16	0.690	0.310	0.999
	7	97.64	0.695	0.305	0.999
	10	129.24	0.632	0.368	0.999
	15	275.83	0.701	0.299	1.000
	20	268.27	0.632	0.368	1.000
	25	303.61	0.626	0.374	1.000
	30	363.4	0.603	0.397	1.000
M8	a	69.28	0.667	0.333	0.999
	3	49.63	0.644	0.356	0.999
	5	72.14	0.667	0.333	0.999
	7	92.65	0.701	0.299	0.999
	10	123.36	0.672	0.328	0.999
	15	175.72	0.690	0.310	0.999
	20	242.36	0.701	0.299	0.999
	25	291.06	0.672	0.328	0.999
	30	406.4	0.609	0.391	0.999

※ 실험환경 : Intel Core i3-6100(3.7GHz), 4.0GB RAM, Windows 10

4.3.5 학습 모델별 시각화

학습 모델별 학습 데이터 수행 결과 최적화된 모델의 속성별 분포도는 Fig. 3~7과 같다. 주요 특징으로는 파일명 속성의 기타와 미분류, 확장자 속성의 압축과 미분류, 직급 속성에서는 중간 직급, 년차 속성은 3년 미만과 20년 이상 근속, 직원 구분 속성은 계약직에서 이상 징후가 더 많이 분포 되었다.

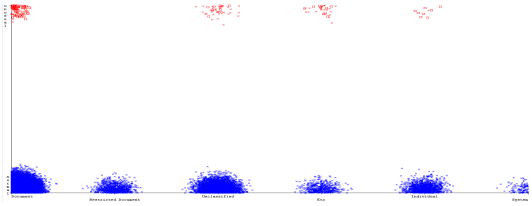

Fig. 3. File Name Classification Attribute Distribution Plot

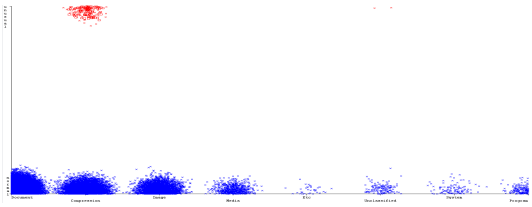

Fig. 4. Extension Classification Attribute Distribution Plot

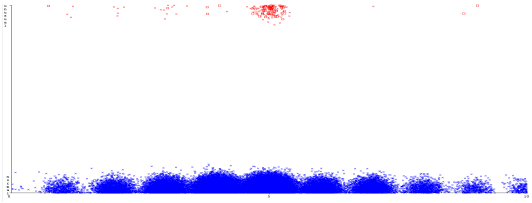

Fig. 5. Employee Rank Attribute Distribution Plot

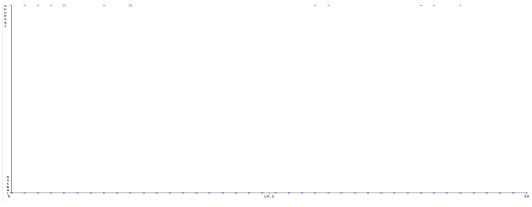

Fig. 6. Employee Years Attribute Distribution Plot

Fig. 7. Employee Type Attribute Distribution Plot

4.4 탐지 모델 검증

제안한 이상 행위 탐지모델을 검증하기 위해서 전체 학습 데이터 24개월 중 임의로 선정한 12개월 데이터를 각각 분리하여 검증 데이터로 사용 하였다. 검증 방법은 Table 6과 동일한 속성으로 학습 데이터 모델 수행 결과에서 각 학습 모델별 TP Rate 비율이 제일 높았던 다층 퍼셉트론 파리미터 은닉층(Hidden Layer) 개수를 동일하게 설정하고 TP Rate를 측정하였다.

탐지 모델 검증 결과 Table 8과 같이 2018년 데이터는 Model 8이 TP Rate(0.699), 2019년 데이터는 Model 2과 7이 TP Rate(0.738)로 이상 행위 탐지를 위한 최적화된 속성 모델 검증에서 일정한 탐지 결과를 확인할 수 있다.

Table 8. True Positive Rate

No.	Hidden Layer	2018	2019
Model 1	5	0.566	0.656
Model 2	3	0.673	0.738
Model 3	5	0.434	0.459
Model 4	5	0.681	0.705
Model 5	a	0.549	0.656
Model 6	20	0.664	0.721
Model 7	15	0.681	0.738
Model 8	20	0.699	0.705

V 결 론

5.1 결론

본 연구에서는 내부자 이상 행위가 늘어나고 있고, 국가, 공공기관의 경우 전자문서 혹은 각종 기밀이 포함된 자료들이 일단 유출되면 매우 다양한 경로를 통해 급속도로 전파되어 막대한 피해가 발생하기

때문에 이러한 사고를 미연에 방지하고자 기계학습 알고리즘을 이용한 내부자 이상 행위 탐지 방법을 제안하고 검증하였다.

다층 퍼셉트론의 파라미터 은닉층 개수가 증가할수록 소요시간이 증가 하고 정확도를 높여주는 효과가 있었지만, 특정 모델에서는 은닉층 개수가 증가에 따른 정확도를 높여 주진 않았다. 또한, 은닉층의 개수가 증가하게 되면 대용량 데이터의 경우 처리 속도가 증가함으로 탐지 모델에서 빠른 처리를 위해서는 고사양의 PC 또는 서버가 필요할 수 도 있다.

본 연구의 탐지 모델 주요 특징으로는 파일명 속성의 기타와 미분류, 확장자 속성의 압축과 미분류, 직급 속성에서는 중간 직급, 년차 속성은 3년 미만과 20년 이상 근속, 직원 구분 속성은 계약직에서 이상 징후가 더 많이 분포 되었다. 또한, 권한을 가진 내부자들이 자료 유출을 시도하기 위해 주로 퇴직자의 경우 파일 확장자를 변조시키거나 대용량 파일로 압축하는 등 이상 행위를 하여 회사 메일 또는 상용 메일로 전송하였으며, 퇴직일자가 가까워질수록 정상 사용자의 비해 자료 전송 빈도수가 높아지는 것을 확인할 수 있었다.

마지막으로 업무망에서 인터넷망으로 자료 전송 시 기관의 특성에 따라 보안정책이 상이할 수 있지만 대용량 파일, 압축 파일, 퇴직 예정자, 미분류 등 본 논문에서 주요 특징으로 분류된 속성들에 대한 보안 정책을 재확인 할 필요가 있으며, 추가적으로 파일 용량 및 확장자 제한 등 보안을 강화할 필요가 있다.

5.2 향후 연구

본 논문에서 탐지 모델의 정확도를 높이기 위해서는 기존에 이용한 탐지 지표 보다 더욱 다양한 탐지 지표(파일명과 본문 내용의 일치성, 확장자 세분화, 파일 사이즈 그룹화, 추가 인사 정보 등)가 필요해 보이며, 더 나아가 매체제어 솔루션, 단말 이상행위 탐지 및 대응 솔루션(EDR), 출력물 시스템 로그 등 다양한 정보보호시스템의 로그 데이터를 이용하여 연구해 본다면 좀 더 정확한 탐지율을 높힐 수 있을 것이다.

그 외에도 기존에 알려진 다른 기계학습 알고리즘을 이용하여 다양한 테스트를 진행 해보거나, 두 개 이상의 기계학습 알고리즘을 결합하여 사용하는 앙상블 방법의 연구를 통해 내부자의 이상행위의 적합한 알고리즘을 제안하고 조직의 중요한 내부 정보를 좀 더 안전하게 지킬 수 있는 연구를 할 필요가 있다.

References

[1] Korea Internet & Security Agency, “2018 Information Security Survey Report,” Apr. 2019.

[2] National Intelligence Service, “National Public Institutions Security Conformity Validation Guide,” Jun. 2017.

[3] CERT Insider Threat, http://www.cert.org/insider_threat/

[4] Jang-hyuk Ko, “A Study on the Analysis of Insider Behavior Based on Machine Learning for Information Leak Detection,” Aug. 2018.

[5] Hyun-Song Jang, “Data-mining Based Anomaly Detection in Document Management System,” Oct. 2015.

[6] Young-baek Kwon, In-seok Kim, “A Study on Anomaly Signal Detection and Management Model Uing Big Data,” JIIBC, vol. 16, no. 6, pp. 287-294, Dec. 2016.

[7] Hae-dong Kim, “Insider Threat Detection based on User Behavior Model and Novelty Detection Algorithms,” Korea University, Aug. 2017.

[8] Ho-Jin Lee “Feature Selection Practice for Unsupervised Learning of Credit Card Fraud Detection,” Korea University, Feb. 2017.

[9] Pallabi Parveen, Nate McDanial, Varun S. Hariharan, “Unsupervised Ensemble based Learning for Insider Threat Detection,” Sep. 2012.

[10] Eldardiry, H., Sricharn,k.,Liu, j., Hanley,J., Price,B., Brdiczka, O., & Bart,E., “Multi-source fusion for anomaly detection: using across-domain and across-time peer-group consistency checks,“ Jun.

2014.
〔11〕 Tae-ho Kim, “Feature Selection Optimization in Unsupervised Learning for Insider threat Detection,” Korea University, Jun. 2018.
〔12〕 Mi-ae Oh, “A Study on Social security Big Data Analysis and Prediction Model based on Machine Learning,” Korea Institute for Health and Social Affairs, Dec. 2017.
〔13〕 Wikipedia, “machine learning”, https://en.wikipedia.org/wiki/Weka_(machine_learning)
〔14〕 Turban, E., J. E. Aronson, and T. P. Liang, “Decsion Support Systems and Intelligent Systems, (7th Edition),” Prentice Hall Inc., Apr. 2004.
〔15〕 Jason Roell, “From Fiction to Reality: A Beginner’s Guide to Artificial Neural Networks,” Jun. 2017.
〔16〕 Jayesh Bapu Ahire, “The XOR Problem in Neural Networks,” Dec. 2017.
〔17〕 KoNLPy Library, https://konlpy.org/ko/latest/
〔18〕 Jong-hyun Lee, “Exploring the Prediction Model of Underachieving Ratio in Middle School Mathematics Using Machine Learning,” Feb. 2020.

〈저자소개〉

손 인 재 (In Jae Son) 정회원
2009년 2월: 조선대학교 컴퓨터통계학과 학사
2016년 6월~현재: 한국가스안전공사 사이버보안부 근무
2018년 9월: 고려대학교 정보보호대학원 사이버보안학과 석사과정
〈관심분야〉 정보보호 정책, 정보시스템 보안, 데이터 마이닝, 블록체인

김 휘 강 (Huy Kang Kim) 종신회원
1998년 2월: KAIST 산업경영학과 학사
2000년 2월: KAIST 산업공학과 석사
2009년 2월: KAIST 산업및시스템공학과 박사
2004년 5월~2010년 2월: 엔씨소프트 정보보안실장, Technical Director
2010년 3월~현재: 고려대학교 정보호호대학원 교수
〈관심분야〉 온라인게임 보안, 네트워크 보안, 네트워크 포렌직, 침입탐지시스템, 봇넷탐지