Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.27, NO.6, Dec. 2017 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2017.27.6.1385

클래스 불균형 데이터에 적합한 기계 학습 기반 침입 탐지 시스템*

정 윤 경,[1†] 박 기 남,[1] 김 현 주,[2] 김 종 현,[2] 현 상 원[1‡]
[1]성균관대학교, [2]한국전자통신연구원

Machine Learning Based Intrusion Detection Systems for Class Imbalanced Datasets*

Yun-Gyung Cheong,[1†] Kinam Park,[1] Hyunjoo Kim,[2] Jonghyun Kim,[2]
Sangwon Hyun[1‡]
[1]Sungkyunkwan University,
[2]Electronics and Telecommunications Research Institute

요 약

본 논문에서는 정상과 이상 트래픽이 불균형적으로 발생하는 상황에서 기계 학습 기반의 효과적인 침입 탐지 시스템에 관한 연구 결과를 소개한다. 훈련 데이터의 패턴을 학습하여 정상/이상 패킷을 탐지하는 기계 학습 기반의 IDS에서는 훈련 데이터의 클래스 불균형 정도에 따라 탐지 성능이 현저히 차이가 날 수 있으나, IDS 개발 시 이러한 문제에 대한 고려는 부족한 실정이다. 클래스 불균형 데이터가 발생하는 환경에서도 우수한 탐지 성능을 제공하는 기계 학습 알고리즘을 선정하기 위하여, 본 논문에서는 Kyoto 2006+ 데이터셋을 이용하여 정상 대 침입 클래스 비율이 서로 다른 클래스 불균형 훈련 데이터를 구축하고 다양한 기계 학습 알고리즘의 인식 성능을 분석하였다. 실험 결과, 대부분의 지도 학습 알고리즘이 좋은 성능을 보인 가운데, Random Forest 알고리즘이 다양한 실험 환경에서 최고의 성능을 보였다.

ABSTRACT

This paper aims to develop an IDS (Intrusion Detection System) that takes into account class imbalanced datasets. For this, we first built a set of training data sets from the Kyoto 2006+ dataset in which the amounts of normal data and abnormal (intrusion) data are not balanced. Then, we have run a number of tests to evaluate the effectiveness of machine learning techniques for detecting intrusions. Our evaluation results demonstrated that the Random Forest algorithm achieved the best performances.

Keywords: Intrusion Detection System, Machine Learning, Imbalanced Dataset

I. 서 론

침입 탐지 시스템 (IDS, Intrusion Detection System)은 다양한 유형의 네트워크 공격으로부터 시스템과 정보 자산을 보호하는데 사용된다. 일반적으로 정확한 공격 원인 및 패턴 규명을 위해 IDS를

Received(09. 26. 2017), Modified(11. 02. 2017), Accepted(11. 13. 2017)
* 이 논문은 2017년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2016-0-00078, 맞춤형 보안서비스 제공을 위한 클라우드 기반 지능형 보안 기술 개발)
† 주저자, aimecca@skku.edu
‡ 교신저자, swhyun77@skku.edu(Corresponding author)

통해 수집된 트래픽 및 로그에 대한 전문가의 분석 과정을 거치며, 이러한 분석 과정은 위협의 종류에 따라 수일에서 수개월까지 소요된다[13]. 따라서, 즉각적 대응이 어려운 상황이며, 대응이 늦어질수록 복구 비용은 급격히 증가한다.

이를 해결하기 위해, 기계 학습기술을 활용하여 보안 위협 탐지를 자동화하려는 연구가 진행되고 있으며, 산업체에서도 연구 결과를 적극 활용하여 제품화하고 있다. 예를 들어, RSA SA(Security Analytics)는 보안 관련 데이터를 수집하는 빅데이터 플랫폼으로서 특별한 사전 정보나 데이터 분석 전문가 없이도 이상 행동을 스스로 탐지한다[11]. Cyphort의 Intelligent Adaptive Detection Fabric[12]은 공격 정보를 지속적으로 수집하고 학습 및 동작 분석 기술을 사용하여, 다른 보안 도구가 감지하지 못한 새로운 유형의 위협을 발견한다. SK 인포섹이 최근 발표한 보안 플랫폼 Secudium (http://www.skinfosec.com)은 딥러닝 기반 인공지능 엔진을 탑재하여, 기존 전문가가 작성하는 이상행위 판별 규칙을 딥러닝으로 학습한다. 펜타시큐리티시스템의 웹해킹 차단 서비스 클라우드브릭 (https://www.cloudbric.com) 역시 기계 학습 기술을 활용하여 비정상 트래픽을 탐지한다. 이와 같이, 국내외 보안 업체는 최근 기계 학습 엔진을 탑재한 지능형 탐지 기반의 보안 관제 자동화 시스템을 상용화하고 있다.

IDS를 비롯하여 현실 세계에서 수집되는 데이터는 클래스간 비율이 균형적이지 않은 것이 일반적이며, 특히 침입 탐지 문제의 경우 전체 트래픽 데이터 중 침입 데이터 비율이 약 1%로 알려져 있다[2,5]. 기계 학습 기반 IDS 성능 평가에 주로 활용되는 데이터인 KDD Cup 99[7]와 Kyoto 2006+[1] 역시 정상/침입 데이터 클래스간 비율이 불균형적이다. KDD Cup 99의 경우 전체 데이터에서 정상 데이터 비율이 26% 이고, Kyoto 2006+ 데이터의 경우 정상 데이터 비율이 5-6% 수준이다. 이렇게 클래스간 크기가 불균형한 데이터셋을 이용하여 IDS 성능을 실험하는 경우, 큰 클래스 뿐만 아니라 작은 클래스에 대한 인식 성능 또한 평가될 수 있도록 성능 평가 수치를 선정할 필요가 있다.

본 연구의 최종 목표는 클래스 불균형 문제가 있는 경우에도 신뢰성 있는 탐지 성능을 보이는 네트워크 IDS를 개발하는 것이다. 본 논문은 클래스 불균형 데이터에 대해서도 효과적으로 침입을 탐지하는 기계 학습 알고리즘을 선정하기 위해 성능 비교 실험을 수행하였다. 본 논문이 기여한 바는 다음과 같다. 첫째, 실제 환경에서 수집한 네트워크 로그 데이터인 Kyoto 2006+ 데이터셋으로부터 정상과 침입 클래스 간 비율이 다른 실험 데이터셋을 구축하고 공개하여 (https://github.com/inglab/dataset) 서로 다른 연구 결과의 비교가 가능하도록 하였다. 둘째, 기계 학습 연구 분야에서 활용되는 성능 비교 척도를 사용하여 6가지 대표적인 기계 학습 알고리즘의 탐지 성능을 분석하였다. 셋째, 다양한 실험 환경에서 Random Forest 알고리즘이 최고의 성능을 보임을 추가 데이터를 이용한 실험을 통해 확인하였다.

본 논문의 구성은 다음과 같다. 먼저 2절에서 관련 연구를 서술한 후, 3절에서 데이터셋의 통계 분석 결과와 실험 환경, 및 성능 평가 지표를 설명한다. 4절에서 다양한 기계 학습 알고리즘의 성능을 비교 분석한 결과를 설명한다. 마지막으로, 5절에서 결론을 맺는다.

II. 관련 및 배경 연구

2.1 Kyoto 2006+ 및 KDD Cup 99 데이터셋

KDD CUP 99 데이터셋은 1999년에 수집된 데이터로 IDS 성능 평가에 가장 널리 사용되고 있으며, 41개의 feature들을 포함한 약 31만개의 패킷에 대한 로그 정보로 구성된다. KDD CUP 99 데이터셋에는 많은 양의 중복된 레코드가 포함되어 있고, 70% 이상이 테스트 및 훈련 셋에 동시에 존재한다. 따라서, 데이터 중복을 제거한 NSL-KDD 데이터셋이 평가 목적으로 사용되기도 한다. 하지만, 1999년에 수집되어 최근 침입 탐지 패턴을 반영하지 못하고 가상 네트워크 환경에서 수집되어 실제 네트워크 시스템에서 관찰된 패턴과 다르다는 한계가 있다.

Kyoto 2006+ 데이터셋은 교토 대학교에서 2006년부터 2015년까지 다양한 유형의 허니팟 시스템으로부터 수집된 네트워크 로그로 24개의 feature가 포함되어 있다. 이중 처음 14개 feature는 기존 KDD CUP 99 데이터에서 학습에 유의미한 것들을 선정한 것으로, connection duration, service, source bytes, count 등이 있다. 나머지 10개의 feature는 검증 용도로도 사용될 수 있도록 추가되었다. 추가된 feature는 Snort 탐지 여부, 안티 바이러스 탐지 여부, 쉘 코드 탐지 여부, 위협

혹은 정상 세션인지 여부를 나타낸 레이블, 소스 IP 주소, 소스 포트 번호, 목적지 IP 주소, 목적지 포트 번호, 시작 시간, 지속 시간 정보를 가지고 있다.

전체 데이터셋은 다음과 같은 3종류의 클래스로 구성된다. 허니팟에 들어온 패킷은 모두 공격이라는 전제하에 알려진 침입 (-1), 쉘코드 기반의 알려지지 않은 침입 (-2)으로 세분화하고, 정상 서버에 들어오는 패킷은 정상 (2)으로 구분한다.

2.2 불균형 데이터에 대한 인식 성능 측정 배경 연구

기계 학습 기반 인식 알고리즘은 어떤 데이터를 사용했는가에 따라 인식 성능에 영향을 준다. 특히, 훈련 데이터에서 어느 한 클래스가 다른 클래스에 비해 대단히 많거나 적은 경우를 클래스 불균형 (class imbalance) 문제라 하며, 실제 환경에서 수집한 데이터에서 흔히 볼 수 있다. IDS 문제 뿐만 아니라 기계 결함 탐지, 텍스트 분류, 지진이나 폭발 감지, fraud 탐지, 희귀병 탐지 등이 그 예이다.

클래스 불균형 데이터로 기계 학습 알고리즘을 훈련하는 과정에서 발생하는 문제에 대한 해결책은 데이터나 알고리즘 레벨에서 강구되고 있다[14]. 데이터 레벨에서는 다수를 차지하는 클래스에 데이터를 랜덤 필터링하여 크기를 줄이는 under-sampling 과, 소수를 차지하는 클래스에 데이터를 중복하여 데이터를 늘이는 over-sampling 등 re-sampling 하는 방식이 있다. 그러나, under-sampling은 정보 손실을 초래하고 over-sampling은 지나치게 훈련 데이터에 맞추어 학습하는 과적합 문제를 초래할 수 있다. 반면, 알고리즘 레벨에서는 소수 클래스의 에러에 대한 패널티를 높게 설정하거나, 데이터의 크기에 둔감한 SVM, 혹은 퍼지 방식을 사용할 수 있다.

클래스 불균형 데이터셋을 훈련 데이터로 사용하는 경우, 객관적 성능 평가가 가능한 수치 선정이 중요하다. 그러나, IDS 성능 평가에서 흔히 사용되는 정확도(accuracy)나 ROC curve 만으로는 다음과 같은 이유로 성능을 신뢰하기 어렵다. 정확도는 전체 테스트 데이터에서 정상과 이상 데이터를 제대로 예측한 수의 비율로서, 다수 클래스에 대한 인식 성능이 높으면 소수 클래스를 완전히 무시해도 좋은 결과를 얻을 수 있다. 예를 들어, 100개의 테스트 사례 중 A 클래스가 99 개이고, B 클래스가 1 개이면, 입력된 데이터를 모두 A로 판단하면 정확도는 99%가 된다. ROC curve는 클래스를 분류하는 판별 기준치를 변화시키면서 민감도와 특이도의 변화를 시각화 한 것으로, 3 클래스 이상의 사례에는 적용하기 어렵다. 또한, 클래스 판별 기준치의 변화에 따른 민감도와 특이도 모두 필요하기 때문에, 여러 연구간 비교가 어렵다.

따라서, 본 연구에서는 불균형 데이터에 대한 기계 학습 알고리즘 기반 침입 탐지 성능 평가 지표로 정확도(accuracy), 정밀도(precision), 재현력(recall), F-score 등을 사용한다. 아래에서 각 지표의 의미를 식과 함께 설명한다.

TP(True Positive)가 정탐의 수, FP(False Positive)가 오탐의 수를 의미할 때, 정밀도(precision)는 식 (1)로 계산된다. 즉, 알고리즘이 검출한 positive의 개수 중 실제로 positive인 데이터의 비율을 의미한다.

$$\mathrm{Pr}ecision = \frac{TP}{TP+FP} \tag{1}$$

반면, 재현력(recall)은 탐지하고자 하는 데이터를 얼마나 잘 찾아냈는지를 의미한다. 식 (2)를 보면 FN(False Negative)가 실제 침입의 사례였으나 정상 사례로 잘못 판단한 미탐의 수를 의미할 때, 재현력은 침입으로 판단한 갯수를 전체 침입의 수로 나눈 비율이다. 민감도(sensitivity), 탐지율(detection rate)이라고도 불린다.

$$Recall = \frac{TP}{FN+TP} \tag{2}$$

정밀도를 높이려면, 해당 클래스에 속할 확률이 매우 높은 경우만 예측하면 된다. 이 때, 재현력의 성능은 낮아지는데, 실제 해당 클래스에 속하지만 확률이 낮아서 제외된 데이터들이 다수 존재하기 때문이다.

따라서, 정밀도와 재현력을 동시에 고려하는 harmonic mean인 F_β score가 유용하다. P가 precision, R이 recall을 의미할 때, F_β 수치는 다음과 같이 계산된다.

$$F_\beta = \frac{(1+\beta^2)(P\times R)}{(\beta^2 P+R)} \tag{3}$$

따라서, β를 조정함으로써 중요하게 생각하는 수

치를 반영할 수 있다. F_1은 정밀도와 재현력은 동등하게 고려하는 수치이고, F_2는 재현력을 정밀도보다 2배 더 중요하게 고려하는 반면, $F_{0.5}$는 정밀도를 재현력보다 2배 더 중요하게 고려하는 수치이다.

2.3 기계 학습 기반 침입 탐지 연구

기계 학습 (Machine Learning)은 인공지능 기술의 한 분야로서, 사람의 명시적인 프로그래밍 없이 기계가 주어진 훈련 데이터로부터 규칙성이나 패턴을 발견하여 클래스 분류하는 기능을 수행한다. 분류하려는 클래스 값을 활용하여 학습하면 지도 학습(supervised learning), 그렇지 않으면 비지도 학습(unsupervised learning)으로 분류한다.

기계 학습을 이용한 IDS 기술은 크게 signature 기반 방식과 anomaly 기반 방식 기법으로 나뉜다. Signature-based는 침입의 패턴을 학습하여 탐지하는 방식이며 지도 학습 알고리즘(SVM, Decision Tree, Neural Network, KNN 등)을 활용하고, anomaly detection은 다수의 밀집된 데이터가 정상이라 가정하고 소수의 아웃라이어를 비정상으로 탐지하는 방식으로 비지도 학습 알고리즘 (one-class SVM, K-means clustering 등)을 사용한다. 본 논문은 signature-based 방식을 채택하고 있다.

Kyoto 2006+ 데이터셋[1]에 기계 학습 알고리즘을 적용하여 침입 패턴을 인식하는 연구는 다음과 같다. Song et al.[2]은 one-class SVM 모델에 기반한 침입 탐지 접근법을 제안했다. Sallay et al.[3]은 온라인 자기 훈련 기능의 SVM 시스템을 기반으로 하는 실시간 침입 탐지 경보 classifier를 제안했다. Chitrakar et al.[4]은 점진적 SVM 알고리즘을 기반으로 침입 탐지 시스템을 개발했다. Ishida et al.[5]는 OptiGrid 클러스터링과 Grid 기반 클러스터 레이블링 알고리즘을 결합하여 공격 트래픽을 식별하는 침입 탐지 방식을 제안했다. Ambusaidi et al.[6]는 상호 정보 기반 알고리즘으로 최적의 특징을 선택하고 최소 제곱 지원 벡터(Least Square SVM) 기반으로 개발한 IDS 성능을 기존 SVM을 적용한 성능과 비교 분석하였다. Sahu et al.[10]은 J48 Decision Tree 알고리즘 기반의 침입 탐지 시스템의 성능을 분석했다. Kishimoto et al.[8]은 네트워크 트래픽의 트렌드가 동적으로 변하는 상황에 대응하기 위해 다양한 트래픽 데이터셋들을 학습시켜 얻어진 여러 개의 classifier들을 복합적으로 활용하는 침입 탐지 시스템을 제안했다. Beaver et al.[9]는 [8]과 유사하게 여러 개의 침입 탐지용 classifier들이 존재하는 상황에서 adaptive boosting 방식 기반의 침입 탐지 시스템을 제안했다.

Table 1.은 대표적인 연구의 실험 환경 및 성능 분석 결과를 보여준다. Kyoto 2006+ 데이터셋을 이용한 IDS 탐지 연구가 존재하지만, 표에서 보듯이 실험에 사용된 데이터가 동일하지 않고, 성능 분석 수치 또한 accuracy, precision, recall, F_1-score, ROC curve graph, false discovery rate 등으로 다양하여 연구 결과간의 객관적인 수치 비교가 어렵다.

III. 실험 방법

본 절에서는 클래스 불균형 데이터에 적합한 기계 학습 기반 IDS를 찾기 위하여, 실험용 데이터를 선정하는 과정과 실험 절차를 기술한다.

3.1 실험 데이터 구축

Kyoto 2006+ 데이터셋은 2006년 11월부터 2015년 12월까지 9년 2개월간 수집된 총 19.78GB에 달하는 빅 데이터이다. 실험에 사용되는 데이터는 다음과 같은 기준으로 선정하였다. 첫째, 데이터 크기가 커지면 노이즈가 많이 발생되어 오탐지 확률이 커지므로 3-7일 이내의 기간 동안 연속적으로 수집된 데이터를 선정한다.

둘째, 수집된 데이터 량이 많고 다양한 유형(즉, 정상/알려진 침입/쉘 코드 사용 알려지지 않은 침입)이 고르게 분포된 기간의 데이터를 선정한다. 이러한 기준을 적용하여, 2013년 1월 3-9일 동안 수집된 데이터를 테스트 데이터로, 2013년 2월 19-25일 사이에 수집된 데이터를 훈련 데이터로 선정하였다. 추가적으로, 극단적인 클래스 불균형 데이터에 대한 성능 평가를 위하여, 훈련 셋에서 침입 데이터가 전체의 1%에 해당하도록 랜덤하게 under-sampling 하였다.

Fig.1.은 실험 데이터 구축 과정을 도식화 하고 각 과정에서 해당 클래스의 크기 정보를 보여준다.

Table 1. The results of previous research on intrusion detection using Machine Learning algorithms

Study	*Method*	*Data & Result*
Sallay [3]	Online self-trained SVM	Accuracy : 0.98 precision : 0.99 Recall : 0.98 F1-score : 0.98
Chitrakar [4]	CSV-ISVM	2007.11.1.~3 Precision : 0.90 False discovery rate : 0.02
Ishida [5]	OptiGrid Clustering & Grid-based Labelling	Training set : 2007.11.1.~3 (attack rate 1%) Test set : 2007.12.1.,8,15,22, 2008.12.1.,9,15,22, 2009.7.1.,8,15,22 Performance analysis using ROC Curve
Ambusaidi [6]	Mutual Information Feature Selection	Training set: randomly extract 152,460 data from 2009.8.27.~31 Test set: 2007.11.1.~3 Accuracy : 0.998 Precision : 0.996 False discovery rate : 0.001

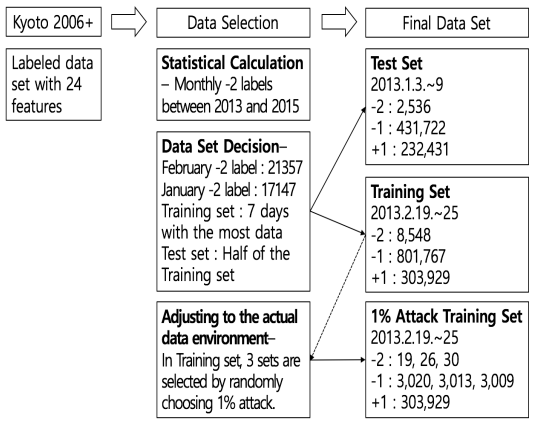

Fig. 1. The process of building training and test datasets for evaluation

Table 2.를 보면, 기본적으로 Kyoto 2006+ 데이터셋에는 침입 데이터(-1과 -2 레이블)가 다수를 차지함을 알 수 있다. 이는 허니팟 서버를 이용하여 공격 트래픽을 수집한 특수한 상황에서 기인하는 것으로, 실제 정상적인 네트워크 환경에서는 관찰되는 공격 패킷의 비율이 약 1% 정도가 일반적이다[2]. 이를 반영하여 침입 데이터의 비율이 전체 셋의 1%가 되도록 랜덤하게 샘플링한 3개의 훈련 데이터를 만든 통계 결과를 Table 3.에서 보이고 있다. 정상 클래스의 데이터는 기존 훈련 데이터의 정상 클래스 데이터를 동일하게 사용하였다.

Table 2. Statistic of training and test data. The number in parentheses denotes the percentage of the label in the set.

Label	Training (2013.2.19-25)	Test (2013.1.3-9)
1 (normal)	303,929 (27.3%)	232,431 (34.9%)
-1 (abnormal)	801,767 (72.0%)	431,722 (64.8%)
-2 (shellcode)	8,548 (0.8%)	2,536 (0.4%)
Total	**1,114,244**	**666,689**

Table 3. Statistic of training data when the abnormal data are under-sampled. The number in parentheses denotes the percentage of the label in the set.

Label	Attack 1% Training A	Attack 1% Training B	Attack 1% Training C
1 (normal)	303,929 (99%)		
-1 (abnormal)	3,020 (0.99%)	3,013 (0.99%)	3,009 (0.99%)
-2 (shellcode)	19 (0.01%)	26 (0.01%)	30 (0.01%)
Total	**306,968**		

3.2 성능 평가 수치 및 기계 학습 알고리즘 선정

네트워크 침입 탐지 시스템에서는 침입인 경우를 잘 탐지하는 것이 중요 하므로, 클래스의 재현력이 정밀도보다 중요하다. 따라서, 본 연구에서는 F_1-score와 F_2-score 수치를 모두 제시한다.

지도 학습 알고리즘 중 대표적인 알고리즘인 Support Vector Machine (SVM), Decision

Tree, Random Forest, K-nearest Neighbors (KNN), Naive Bayes, Neural Network 알고리즘들을 Python 3.6 버전에서 Scikit-learn 0.19.0 버전을 이용하여 구현하였으며, 실험 환경으로는 Intel Core i5 3.1GHz, 8GB 메모리, Mac OS 환경의 PC를 사용하였다. 알고리즘의 파라미터 값은 대부분 기본 설정치를 이용하였고, 몇 가지 값은 여러 번의 실험을 통해 최적치를 활용하였다. 예를 들면, Decision Tree의 max deep은 4로, KNN 알고리즘의 nearest instance 수는 3으로 설정하였다.

실험에 사용된 feature는 KDD Cup 99에서 침입 탐지에 중요한 것으로 선별된 14 항목과 Kyoto 2006+에서 새로 만든 10개 중 4개 (IDS_detection, Malware_detection, Ashula (shellcode) detection, duration) 등 총 18개의 속성이다. 데이터는 원래의 값을 사용한 경우와 정규화 처리된 경우로 테스트 하였으며, 본 논문에서는 정규화된 결과를 보인다.

IV. 실험 결과

본 절에서는 3.1절에서 선정한 훈련 및 테스트 데이터를 3.2절에서 선정한 지도방식 기계 학습 알고리즘에 적용하여 예측한 결과를 기술한다. 우선, 데이터셋의 원 레이블 별 (1, -1, -2)의 세 가지 클래스를 예측하는 경우를 실험했다. 다음으로, -1과 -2 레이블이 침입 데이터를 의미하는 점을 반영하여, 하나의 침입 클래스로 병합하여 정상과 침입의 두 가지 클래스로 예측하는 경우로 나누어 실험을 수행했다.

4.1 3 class 예측 결과

이미 기술한 바와 같이, Kyoto 2006+ 데이터셋에는 1(정상), -1(알려진 방식의 침입), -2(쉘 코드를 이용한 침입) 세 가지 레이블의 데이터들이 존재한다. Table 4.는 각각의 알고리즘을 통해 이 세 가지 클래스를 예측한 성능을 정확도, 정밀도, 재현력, F1-score, F2-score 순서로 보여준다. 이 수치는 각 클래스별로 산출된 수치에 해당 클래스 크기를 가중치로 평균한 결과이다.

Naive Bayes 알고리즘을 제외한 모든 지도 학습 알고리즘들이 모든 지표에서 0.95 이상의 우수한 성능을 보였다. 특히, 결정 트리 계열의 알고리즘인 Decision Tree, Random Forest가 모든 지표에서 비슷한 수준의 우수한 성능을 보였으며, 최고 accuracy와 F_1, F_2-score가 0.99에 달하였다.

Table 4. Performance comparison of supervised machine learning algorithms using 3 classes

Classifier	*Accuracy*	*Precision*	*Recall*	*F_1-Score*	*F_2-Score*
SVM	0.95	0.95	0.95	0.95	0.95
Decision Tree	0.96	0.96	0.96	0.96	0.96
Random Forest	0.99	0.99	0.99	0.99	0.99
Naive Bayes	0.86	0.89	0.86	0.86	0.87
Neural Network	0.98	0.98	0.98	0.98	0.98
KNN	0.98	0.98	0.98	0.98	0.98

불균형 데이터의 특성상, 가중치 평균의 경우 극히 적은 수를 차지하는 -2 레이블에 대한 예측 성능이 낮더라도 전체 평균 성능은 높을 가능성이 있으므로, 각 클래스별 탐지 성능 또한 확인하였다. Table 5.에서 보듯이, Random Forest 알고리즘은 각 클래스에 대해서도 고르게 좋은 성능을 보였다.

Table 5. Performance comparison of the Random Forest algorithm for each class

Label	*Precision*	*Recall*	*F_1-Score*	*F_2-Score*	*number of instances*
-2	1.00	1.00	1.00	1.00	2536
-1	0.99	1.00	0.99	1.00	431722
1	0.99	0.98	0.99	0.98	232431

4.2 2 class 예측 결과

-1, -2 레이블이 모두 침입 클래스를 나타낸다는 점을 고려하여, 기존 3 클래스 데이터셋을 정상과 침입의 2 클래스로 분류하는 실험을 수행하였다 (Table 6.). 대부분의 알고리즘이 3 클래스 분류 결과와 유사하나 약간 낮은 성능을 보였고, Random Forest는 3 클래스 분류와 동일한 성능을 보였다.

Table 6. Performance comparison of supervised machine learning algorithms using 2 classes

Classifier	*Accuracy*	*Precision*	*Recall*	F_1-*Score*	F_2-*Score*
SVM	0.95	0.95	0.95	0.95	0.95
Decision Tree	0.97	0.97	0.97	0.97	0.97
Random Forest	0.99	0.99	0.99	0.99	0.99
Naive Bayes	0.85	0.89	0.85	0.86	0.86
Neural Network	0.97	0.97	0.97	0.97	0.97
KNN	0.98	0.98	0.98	0.98	0.98

4.3 Attack 1% 훈련셋에서 2 class 예측 결과

훈련 데이터에서 침입 클래스의 크기가 전체 데이터의 1%에 해당하는 환경에서 실험한 결과는 Table 7,8,9.에 제시되어 있다. 전반적으로 대부분의 알고리즘에서 인식 성능이 현저히 낮아진 것을 알 수 있다. Naive Bayes를 제외하고 성능이 크게 감소하였다 (F_1, F_2 수치 0.1~0.2 감소). SVM 알고리즘은 클래스 불균형 데이터에 대해서도 강건하다는 기존 연구 결과[14]와 일치하지 않았다.

Table 7. Performance comparison of supervised machine learning algorithms using 1% attack Training A

Classifier	*Accuracy*	*Precision*	*Recall*	F_1-*Score*	F_2-*Score*
SVM	0.76	0.86	0.76	0.76	0.89
Decision Tree	0.81	0.87	0.81	0.81	0.82
Random Forest	0.92	0.90	0.90	0.90	0.90
Naive Bayes	0.87	0.90	0.87	0.87	0.88
Neural Network	0.78	0.86	0.78	0.78	0.79
KNN	0.82	0.88	0.82	0.82	0.83

Table 8. Performance comparison of supervised machine learning algorithms using 1% attack Training B

Classifier	*Accuracy*	*Precision*	*Recall*	F_1-*Score*	F_2-*Score*
SVM	0.76	0.85	0.76	0.76	0.78
Decision Tree	0.81	0.88	0.81	0.81	0.82
Random Forest	0.88	0.91	0.88	0.89	0.89
Naive Bayes	0.86	0.89	0.86	0.86	0.87
Neural Network	0.80	0.87	0.80	0.81	0.81
KNN	0.82	0.88	0.82	0.82	0.83

Table 9. Performance comparison of supervised machine learning algorithms using 1% attack Training C

Classifier	*Accuracy*	*Precision*	*Recall*	F_1-*Score*	F_2-*Score*
SVM	0.76	0.85	0.76	0.76	0.78
Decision Tree	0.85	0.89	0.85	0.85	0.86
Random Forest	0.90	0.92	0.90	0.90	0.90
Naive Bayes	0.86	0.89	0.86	0.86	0.87
Neural Network	0.81	0.88	0.81	0.82	0.82
KNN	0.81	0.87	0.81	0.81	0.81

4.4 추가 데이터셋과의 성능 비교

본 실험 결과의 일반성을 확인하기 위하여 2014년도의 데이터에 대하여 추가적인 실험을 수행하였다. 2014년 12월 8~14일 데이터를 트레이닝 셋으로 2014년 8월 25~31일 데이터를 테스트 셋으로 동일한 실험을 반복한 결과 (Table 10.의 통계치 참조), 역시 Random Forest가 가장 좋은 성능을 보였다 (Table 11.). 앞선 실험과 동일하게 공격 데이터 비율을 1%로 랜덤하게 under-sampling 하여 구축한 훈련 데이터셋으로 실험한 경우에도 F_1-*score* 및 F_2-*score*가 0.95로 다른 알고리즘에 비하여 좋은 결과를 보였다.

Table 10. Statistic of the second dataset

Label	Training (2014.12.8-14)	Test (2014.8.25-31)
1 (normal)	260,329 (10.9%)	113,298 (9.4%)
-1 (abnormal)	2,128,046 (89.0%)	1,087,923 (90.6%)
-2 (shellcode)	174 (0.0%)	166 (0.0%)
Total	2,388,549	1,201,387

Table 11. Performance comparison of supervised machine learning algorithms using 3 classes

Classifier	*Accuracy*	*Precision*	*Recall*	*F_1-Score*	*F_2-Score*
SVM	0.86	0.94	0.86	0.88	0.87
Decision Tree	0.97	0.97	0.97	0.97	0.97
Random Forest	0.99	0.99	0.99	0.99	0.99
Naive Bayes	0.10	0.74	0.10	0.04	0.12
Neural Network	0.97	0.97	0.97	0.97	0.97
KNN	0.97	0.97	0.97	0.97	0.97

4.5 논의

실험 결과 Random Forest 알고리즘이 다양한 환경에서 일관적으로 최고의 성능을 보이며, 클래스의 크기가 불균형한 경우에도 각 클래스에 대해 균일하게 좋은 성능을 보였다.

Random Forest는 정규화되지 않은 데이터 원래 값을 그대로 사용한 실험의 경우에도 3 클래스 인식 및 2 클래스 인식 F_1, F_2 수치가 0.99 였고, 1% attack 경우에는 Training A, B, C 모든 경우 0.9 였다. 즉, 정규화 과정을 거치지 않아도 유사한 성능을 보였다. 이것은, Random Forest가 decision tree 기반의 앙상블 모델이기 때문에, 다양한 조건에서도 좋은 결과를 보인 것으로 판단된다.

반면, 정규화 여부에 따라 성능이 크게 차이가 난 알고리즘은 SVM과 Naive Bayes 였다. SVM은 정규화하지 않은 경우 3 클래스 인식에서 F_1, F_2 수치가 0.90, 2 클래스 인식에서 F_1, F_2 수치가 0.77, 0.79로 크게 낮아졌다. 특히, Naive Bayes는 데이터 비정규화시 대단히 낮은 성능을 보였다. 3 클래스 인식에서 F_1, F_2 수치가 0.15, 0.31, 2 클래스 인식에서 F_1, F_2 수치가 0.18, 0.37로 매우 낮았다.

또한, 1% attack에서는 SVM과 Naive Bayes 알고리즘 모두 훈련셋에 따라 성능 차이가 컸다. SVM은 Training A에서 F_1, F_2 수치가 0.74, 0.76, Training B에서 0.78, 0.79 였으나, Training C에서는 0.23, 0.39로 크게 낮았다. Naive Bayes는 Training A에서 F_1, F_2 수치가 0.79, 0.81, Training B에서 0.93, 0.93, Training C에서 0.18, 0.39로 성능의 편차가 매우 컸다. 따라서, SVM과 Naive Bayes의 경우 데이터의 정규화가 반드시 필요하다.

Random Forest 알고리즘은 속도 면에서도 장점을 보였다. 본 실험의 3 클래스 인식 문제에서 Random Forest는 38초가 걸려, 26초 걸린 Naive Bayes 알고리즘과 29초 걸린 Decision Tree 다음으로 속도가 빨랐다. 반면, Neural Network는 6분, SVM은 10분, KNN은 2시간 이상 소요 되었다. 따라서, 빅데이터를 처리하거나 실시간 처리가 필요한 경우에도 Random Forest가 적합한 알고리즘임을 알 수 있다.

V. 결 론

본 연구에서는 기계 학습 기반 네트워크 침입 탐지 시스템 개발을 위하여 지도 학습 방식의 다양한 기계 학습 알고리즘들의 성능을 비교 분석하였다. 본 논문이 네트워크 침임 탐지 연구 분야에 기여한 바는 다음과 같다.

IDS에서 수집되고 훈련용으로 사용되는 데이터가 클래스 불균형 특성을 보인다는 점을 고려하여, 최근 수집된 Kyoto 2006+ 데이터셋을 통계적으로 분석하여 다양한 정상 대 침입 클래스 비율의 훈련 및 테스트 셋을 구축하였다. 다음으로, 기계 학습의 대표적인 패턴 인식 성능 분석 지표인 accuracy, precision, recall, *F_1-score* 및 *F_2-score*를 침입 탐지 알고리즘의 성능 분석 지표로 기계 학습 알고리즘의 성능을 비교 분석하였다. 클래스 불균형 데이터를 다루는 경우, *F_1-score*는 정밀도와 재현력을 동시에 고려하기 위해 필수적으로 사용하는 수치이다. 또한, IDS 문제는 침입 데이터에 대한 재현력이 중

요하기 때문에 F_2-score를 사용했다.

실험 결과, Random Forest 등 결정 트리 계열의 알고리즘이 네트워크 침입 탐지에 있어서 일관적으로 우수한 성능을 보였다. 또한, 데이터를 정규화하지 않은 경우에도 좋은 성능을 보였다. 이 결과에 따라 본 연구에서는 Random Forest 알고리즘을 사용한 IDS 시스템을 개발 중이며, 향후 수집되는 데이터에 대해서도 성능을 분석할 예정이다.

침입 데이터가 전체 훈련 셋에서 1%를 차지하는 극단적 불균형 데이터의 경우에는 성능이 현저히 하락하였다. 따라서, 추가적인 학습이 필요한 경우 클래스간의 크기 비율을 1:3 혹은 1:4 수준으로 훈련 데이터를 구축하는 것이 중요하다.

본 연구에서는 지도 학습 기술 기반 네트워크 침입 탐지 성능을 분석하였다. 이를 위하여 선정한 실험용 훈련 데이터에서 침입 클래스가 70% 정도를 차지하였고, 레이블 정보도 가지고 있어 지도 학습 방식이 가능하다. 하지만 실제 네트워크 환경에서 수집된 로그에서 침입 클래스가 약 1% 정도로 알려져 있으며, 훈련에서 학습하지 못한 새로운 유형의 침입 패턴도 발생할 경우 미탐지 할 수 있다. 이러한 한계를 극복하기 위하여, 향후 연구에서는 본 연구에서 구축한 실험 데이터에 대하여 비지도 학습 기반 알고리즘과 딥러닝 기술을 비교 분석할 계획이다.

References

〔1〕 Song, Jungsuk, Takakura, Hiroki, Okabe, Yasuo, Eto, Masahi, Inoue, Daisuke, and Nakao, Koji, "Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation," Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, pp. 29-36, Apr. 2011.

〔2〕 Song, Jungsuk, Takakura, Hiroki, Okabe, Yasuo, and Kwon, Yongjin, "Correlation analysis between honeypot data and IDS alerts using one-class SVM," Intrusion Detection Systems, InTech, pp. 173-192, Mar. 2011.

〔3〕 Sallay, Hassen and Sami Bourouis, "Intrusion detection alert management for high speed networks: current researches and applications," Security and Communication Networks, vol. 8, no. 18, pp. 4362-4372, Dec. 2015.

〔4〕 Chitrakar, Roshan, and Chuanhe Huang, "Selection of Candidate Support Vectors in incremental SVM for network intrusion detection," Computers & Security, vol. 45, no. 16, pp. 231-241, Sep. 2014.

〔5〕 Ishida, Moriteru, Hiroki Takakura, and Yasuo Okabe, "High-performance intrusion detection using optigrid clustering and grid-based labelling," Proceedings of IEEE/IPSJ 11th International Symposium on Applications and the Internet, pp. 11-19, Jul. 2011.

〔6〕 Ambusaidi, Mohammed A., He, Xiangjian, Nanda, Priyadarsi, and Tan, Zhiyuan, "Building an intrusion detection system using a filter-based feature selection algorithm," IEEE Transactions on Computers, vol. 65, no.10, pp. 2986-2998, Jan. 2016.

〔7〕 KDD Cup 1999. Available on: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, Oct. 2007.

〔8〕 Kishimoto, Kazuya, Hirofumi Yamaki, and Hiroki Takakura, "Improving performance of anomaly-based ids by combining multiple classifiers," Proceedings of IEEE/IPSJ 11th International Symposium on Applications and the Internet, pp. 366-371, Jul. 2011.

〔9〕 Beaver, Justin M., Christopher T. Symons, and Robert E. Gillen, "A learning system for discriminating variants of malicious network traffic," Proceedings of the Eighth Annual Cyber Security and Information Intelligence Research Workshop, pp. 23-26, Jan. 2013.

〔10〕 Sahu, Shailendra and Babu M. Mehtre, "Network intrusion detection system using J48 Decision Tree," Proceedings of IEEE International Conference on

Advances in Computing, Communications and Informatics, pp. 2023-2026, Aug. 2015.

[11] RSA Security Analytics Data Sheet, Available on: https://www.rsa.com/content/dam/rsa/PDF/h13414-ds-pdf-sa-overview.pdf

[12] Cyphort Adaptive Detection Fabric Data Sheet, Available on: http://go.rt.com/rs/181-NTN-682/images/CYPHORT_DataSheet.pdf

[13] 2016 Cost of Cyber Crime Study & the Risk of Business Innovation. Ponemon Institute. Available on: https://www.ponemon.org/local/upload/file/2016%20HPE%20CCC%20GLOBAL%20REPORT%20FINAL%203.pdf

[14] Visa, Sofia and Anca Ralescu, "Issues in Mining Imbalanced Data Sets - A Review Paper," Proceedings of the Sixteen Midwest Artificial Intelligence and Cognitive Science Conference, pp. 67-73, Apr. 2005.

〈저자소개〉

정 윤 경 (Yun-Gyung Cheong) 정회원
1996년 2월: 성균관대학교 정보공학과 졸업
1998년 2월: 성균관대학교 정보공학과 석사
2007년 8월: 노스캐롤라이나주립대학교 전산학과 박사
2008년~2010년: 삼성전자 종합기술원 전문연구원
2010년 10월~2014년 8월: 덴마크 IT University of Copenhagen post-doc
2015년 3월~현재: 성균관대학교 소프트웨어대학 조교수
〈관심분야〉 인공지능, 지능적 스토리텔링 및 게임 AI

박 기 남 (Kinam Park) 학생회원
2016년 8월: 충남대학교 항공우주공학과 졸업
2017년 3월~현재: 성균관대학교 소프트웨어플랫폼학과 석사과정
〈관심분야〉 네트워크 및 시스템 보안

김 현 주 (Hyunjoo Kim) 정회원
2000년 2월: 성균관대학교 정보공학과 공학사
2002년 2월: 성균관대학교 컴퓨터공학과 공학석사
2016년 8월: 성균관대학교 컴퓨터공학과 공학박사
2002년 1월~현재: 한국전자통신연구원 정보보호연구본부 선임연구원
〈관심분야〉 네트워크 보안, 악성코드 탐지, 빅데이터 분석, 클라우드 보안

김 종 현 (Jonghyun Kim) 정회원
2000년 5월: 오클라호마주립대 컴퓨터과학과 공학석사
2005년 5월: 오클라호마주립대 컴퓨터과학과 공학박사
1995년~1998년 삼성전자 SW연구개발 연구원
2005년~현재 한국전자통신연구원 책임연구원
〈관심분야〉 정보보호, 네트워크보안, 클라우드 보안, 네트워크 포렌식

현 상 원 (Sangwon Hyun) 정회원
2002년 2월: 성균관대학교 전기전자컴퓨터공학부 졸업
2004년 2월: 서울대학교 컴퓨터공학과 석사
2011년 12월: 노스캐롤라이나주립대학교 전산학과 박사
2016년 3월~현재: 성균관대학교 소프트웨어대학
〈관심분야〉 네트워크 및 시스템 보안