Journal of KIIT. Vol. 23, No. 5, pp. 21-28, May 31, 2025. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2025.23.5.21

Boosting 알고리즘을 이용한 네트워크 트래픽 이상 탐지

김인혜*, 전왕수**

Network Traffic Anomaly Detection using Boosting Methodology

Inhye Kim*, Wangsu Jeon**

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임(IITP-2025-RS-2024-00436773)

요 약

최근 네트워크 기술의 발전으로 데이터 트래픽이 급증하고 있으며, 이와 함께 사이버 위협의 규모도 증가하고 있다. 본 논문에서는 IDS/IPS에서 수집된 데이터셋을 활용하여 오버샘플링 기법과 XGBoost, LightGBM, CatBoost 등의 부스팅 알고리즘을 적용한 고급 네트워크 트래픽 이상 탐지 기법을 제안한다. 제안된 접근 방식은 데이터 불균형 문제를 완화하고, 다양한 약한 분류기를 결합하여 강력한 예측 모델을 형성하여 이상 탐지의 정확도와 효율성을 높인다. XGBoost와 LightGBM 모델이 기존 탐지 기법에 비해 0.5~0.9%의 우수한 성능 향상을 보였으며, 실시간 연산 수행 측면에서는 XGBoost가 가장 적합함을 확인하였다. 또한 오버샘플링 기법이 탐지 성능 향상에 기여함을 검증하였다. 본 연구는 동적이고 복잡한 네트워크 환경에서 IDS/IPS 시스템이 효율적으로 이상 징후를 탐지할 수 있도록 돕고, 네트워크 보안의 신뢰성과 안정성을 강화하고자 한다.

Abstract

Recently, with the development of network technology, data traffic has been rapidly increasing, and the scale of cyber threats has also increased. In this paper, we propose an advanced network traffic anomaly detection technique that applies oversampling techniques and boosting algorithms such as XGBoost, LightGBM, and CatBoost using datasets collected from IDS/IPS. The proposed approach addresses the issue of data imbalance and builds a potent prediction model by combining multiple weak classifiers to improve the accuracy and efficiency of anomaly detection. The XGBoost and LightGBM models showed excellent performance improvements of 0.5~0.9%p compared to existing detection techniques, and XGBoost was confirmed to be the most suitable in terms of real-time computation performance. In addition, we proved that the oversampling method improved detection performance. This study aims to help IDS/IPS systems detect anomalies efficiently in dynamic and complex network environments, strengthening the reliability and stability of network security.

Keywords

intrusion detection, anomalous traffic, IDS, machine learning, boosting algorithms

* 경남대학교 창의융합대학 MSC교육부
- ORCID: https://orcid.org/0009-0004-4115-2251
** 경남대학교 AI · SW융합대학 컴퓨터공학부(교신저자)
- ORCID: https://orcid.org/0000-0001-8887-2513

· Received: Feb. 27, 2025, Revised: Apr. 08, 2025, Accepted: Apr. 11, 2025
· Corresponding Author: Wangsu Jeon
Dept. of Computer Engineering, 7 Kyungnamdaehak-ro, Masanhappo-gu, Changwon-si, Gyeongsangnam-do, Republic of Korea
Tel.: +82-55-249-2262, Email: jws2218@naver.com

Ⅰ. 서 론

최근 몇 년 동안 IoT, 클라우드 컴퓨팅, 5G, AI 등의 기술 발전으로 인해 네트워크 환경이 복잡해지고, 인터넷 연결 장치의 확산으로 트래픽이 급격히 증가하였다. 이러한 변화는 네트워크 환경을 위협하는 네트워크 침입과 사이버 범죄의 증가로 이어져 네트워크 보안의 중요성이 더욱 강조되고 있다. 네트워크 보안 문제를 해결하기 위해서는 네트워크 공격의 패턴을 파악하고 네트워크 보안을 보장하기 위한 탄력적인 침입탐지 솔루션이 필요하다.

네트워크 트래픽은 정상과 악성으로 나뉘며, 악성 트래픽은 Dos(Denial of Service), R2L(Root to Local), U2R(User to Root), Probe로 분류된다[1]. 네트워크 트래픽 이상 탐지는 크게 시그니처 기반 탐지와 이상 기반 탐지로 나눌 수 있으며 제로데이 공격, 무단 액세스, 기존 보안 시스템을 회피하는 기타 악성 활동과 잠재적인 위협을 식별하는데 중요한 역할을 한다. 시그니처 기반 탐지는 트래픽을 공격 시그니처 데이터베이스와 비교하여 알려진 위협을 식별하는 데 효과적이지만 시그니처가 없는 새로운 유형의 공격(제로데이 공격)에는 적절하지 않다[2]. 이상 기반 탐지는 네트워크 동작을 모니터링하고 정상적인 동작에서 벗어난 위협을 식별하여 이전에 알려지지 않은 공격을 탐지하지만 종종 새로운 정상적인 활동을 악성으로 오탐하기도 한다[3]-[5]. 기존 침입 탐지 시스템(IDS, Intrusion Detection System)과 침입 방지 시스템(IPS, Intrusion Prevention System)은 네트워크 트래픽을 모니터링하고 분석하는 데 널리 사용되며, 이러한 시스템은 비정상적인 패턴을 탐지하는 데 도움이 되지만 끊임없이 진화하는 사이버 공격의 특성으로 인해 실시간으로 이상 탐지가 가능한 고급 탐지 기법이 필요하다.

최근에는 머신러닝 활용으로 트래픽 패턴 자동과 분류해 지속적으로 변화하는 사이버 공격 위협에 대응하는 기법이 연구되고 있다. 특히 여러 약한 분류 알고리즘을 결합하여 강력한 예측 모델을 생성하는 XGBoost과 LightGBM은 이상 탐지를 위한 강력한 도구로 등장했다[6][7]. 이러한 고급 알고리즘은 탐지 정확도와 효율성을 향상시켜 네트워크 트래픽 이상을 더욱 정확하게 식별할 수 있도록 한다.

본 연구에서는 IDS/IPS 장비를 통해 수집된 데이터 세트에 Boosting 알고리즘을 적용하여 네트워크 트래픽 이상 탐지 성능을 향상시키는 것을 목표로 한다. 이를 위해 데이터 불균형 문제를 완화하고 다양한 약 분류기를 결합하여 강력한 예측 모델을 구축하는 기법을 제안한다. 이를 통해 이상 탐지의 정확성을 높이고 오탐률을 줄여 더 나은 네트워크 보안과 안정성을 보장하고자 한다.

서론에 이어 2장에서는 네트워크 침입탐지에 인공 지능 모델을 적용한 연구 사례를 소개하며, 3장에서는 이상 침입탐지 분류에 적용한 모델을 제시하고, 4장에서는 모델별 탐지 정확도를 실험한 결과를 분석하고 5장에서 논문의 결론을 맺는다.

Ⅱ. 관련 연구

최근 침입탐지 시스템 분야에서는 인공지능 및 기계학습과 같은 최신 기술들을 적용하려는 연구가 활발히 이루어지고 있다. 특히 기존에 알려지지 않은 공격을 탐지할 수 있는 이상 징후 탐지 분야에서 인공지능 기술의 활용이 증가하고 있으며, 이러한 방법은 높은 탐지 정확도와 효과적인 보안 대응 성능을 보이고 있다.

Boosting 모델을 적용한 연구는 다음과 같다. S. Bhattacharya et al.[8]은 오픈 소스 Kaggle 데이터 세트를 사용하여 침입 탐지를 위한 PCA-Firefly 기반 XGBoost 분류 모델을 제안하였다. 이 모델은 침입 탐지에 있어 높은 정확도와 낮은 오탐률을 보여준다. 그러나 Kaggle이라는 한 데이터셋 만을 사용하여 다양한 네트워크 환경과 공격 유형을 반영하는 데이터셋에 대한 검증과 실시간 탐지에 대한 평가가 부족하다. A. Alqahtani et al.[9]은 IoT 네트워크 침입탐지를 위한 6가지 최신 Boosting 기반 머신러닝 알고리즘의 성능을 비교 평가하여 가장 효과적인 Boosting 기반 알고리즘을 제시하였다. 제시한 모델은 고정된 IoT 장치의 데이터셋을 사용하였으며 대규모 IoT 네트워크에서 테스트되지 않아 산업, 의료, 자동차 등과 같은 다양한 IoT 장치가 포함된 실제 IoT 네트워크 환경에서는 제한적 요소가 있다.

X. Zhang et al.[10]은 시계열 정보의 중요성을 강조하여 Self-attention 메커니즘과 Gradient Boosting 의사결정 트리를 결합한 네트워크 이상 탐지 모델을 제안하였다. 검증을 위해 실제 네트워크 트래픽과 공격 유형을 혼합하여 시뮬레이션한 데이터 셋인 UNSW-NB15을 사용하였다. 실제 네트워크 데이터에 대한 검증이 필요하다.

여러 모델을 결합한 하이브리드 기법과 딥러닝을 적용한 연구는 다음과 같다. Z. Ding et al.[11]은 불균형 데이터 세트를 대상으로 이상 탐지를 위해 KNN(K-Nearest Neighbors)과 GAN(Generative Adversarial Networks)을 결합한 하이브리드 기법을 제안하였다. 또한, 데이터 불균형과 과적합 문제를 해결하기 위해 V. Balyan et al.[12]은 EGA(Enhanced Genetic Algorithm), PSO(Particle Swarm Optimization) 및 IRF(Improved Random Forest) 방법을 결합하는 하이브리드 침입 탐지 시스템을 제안하였다. NSL-KDD 데이터 세트에서 평가되었으며 이상 탐지에 있어 높은 정확도를 달성하였다. 두 연구 모두 특징 선택의 어려움과 새로운 공격 유형에 대한 고려 부족, 그리고 계산의 복잡성 및 실제 데이터가 아닌 훈련 데이터로 인한 오탐율 발생 문제가 있다.

최근 네트워크 트래픽 데이터의 시간적, 공간적 특징을 활용한 딥러닝 모델 적용 연구는 다음과 같다. Y. Xiuye와 C. Liyong[13]은 다중 스케일 CNN(Convolutional Neural Network)을 사용하여 IoT 환경에서의 침입 탐지 분석 네트워크 모델을 제안하였다. J. Kim et al.[14]은 계층적 LSTM(Long Short-Term Memory) 구조를 사용하여 전체 패킷 정보를 효과적으로 처리하였다. 특히 다양한 길이의 세션을 처리할 수 있어 유연한 LSTM 기반 모델을 제시하였다. 이러한 딥러닝 모델을 적용하는 경우 복잡한 특징 추출을 위한 패턴 인식 능력으로 계산의 복잡성과 학습과 추론 시간이 길어져 실시간 처리에는 적합하지 않을 수 있으며, 과적합의 단점이 있다.

본 논문에서는 데이터의 불균형을 처리하고 더 빠른 추론 속도와 과적합에 강한 부스팅 알고리즘을 실제 IDS/IPS 시스템을 통해 수집된 네트워크 트래픽 데이터에 적용하여 이상 침입탐지를 분류하고자 한다. 본 연구는 탐지 정확도를 높이고 오탐률을 최소화하는 데 중점을 둠으로써 보다 효과적인 네트워크 이상 탐지 모델을 개발하기 위한 지속적인 노력에 기여하는 것을 목표로 한다.

III. 네트워크 침입탐지 시스템

본 논문은 장비에 들어오는 패킷 정보를 이용하여 네트워크 침입탐지의 정상과 비정상을 분류한다. 이 방법은 그림 1과 같이 Raw data에서 데이터를 전처리하고, 부스팅 계열의 모델인 XGBoost, CatBoost, LightGBM, 그리고 DNN 모델을 각각 성능을 비교한다.

3.1 데이터 전처리

데이터 전처리 과정은 상관관계 분석을 통해 불필요한 특징과 결측치를 제거한다.

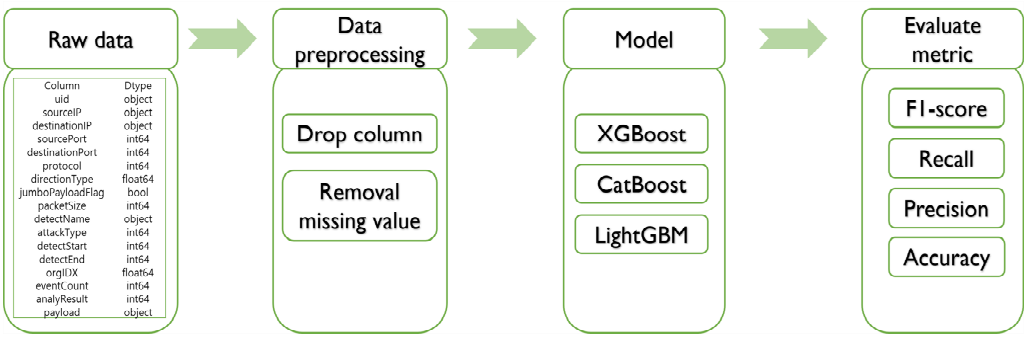

그림 1. 데이터 전처리부터 학습 및 평가까지 제안된 학습 프로세스
Fig. 1. Proposed learning process from data pre-processing to train and evaluate

특징 제거를 위하여 식별자 역할을 하는 'uid', 네트워크 정보를 포함하는 'sourceIP', 'destinationIP', 그리고 'payload', 'jumboPayloadFlag', 'detectName' 데이터는 상관관계가 가장 낮아 삭제하며, 이 데이터는 정탐, 오탐을 구분할 때 불필요하다. 그 후 결측치가 있는 행을 제거한다.

최종적으로 선별된 특징은 네트워크 트래픽의 출발지 포트(sourcePort), 목적지 포트(destinationPort), 통신 프로토콜(Protocol), 트래픽의 방향 유형(Direction type), 패킷의 크기(Packet size), 공격 유형(Attack type), 탐지 시작 시간(detectStart)과 종료 시간(detectEnd), 소속 조직(orgIDX)과 이벤트 발생횟수(eventCount)를 학습에 사용하도록 하며, 이 특징들은 정탐과 오탐을 구분하는 데 필요하다.

이때, 네트워크 데이터는 표 1과 같이 정상과 비정상의 클래스 불균형하여 학습에 어려움이 있거나 인근 특징들의 개수가 부족하여 일반화가 되지 않는 문제점이 있다. 이러한 문제점을 해결하기 위해 SMOTE(Sythetic Minority Over-Sampling Technique)[15]를 사용한다. 이 방법은 데이터의 개수가 불균형이 있을 때, 적은 클래스의 표본을 임의의 값으로 추가하여 새로운 데이터를 생성하는 방법이다.

표 1. SMOTE를 이용한 데이터 오버샘플링 결과
Table 1. Results of data oversampling using SMOTE

	without SMOTE	with SMOTE
Normal	5,224,683	20,211,263
Abnormal	20,211,263	20,211,263

SMOTE는 그림 2의 같이 임의의 소수 클래스 데이터(초록색)로부터 인근 소수 클래스 사이에 거리를 계산하여 새로운 데이터(빨간색)를 생성한다. SMOTE는 식 (1)과같이 먼저 임의의 소수 클래스에 해당하는 관측치 X를 잡고, 그 X로부터 가장 가까운 K개의 이웃 X_{nn}를 찾는다. 그리고 이 K개의 X_{nn}과 X 사이에 임의의 새로운 데이터 X'를 생성하며 식 (1)과 나타낼 수 있다.

$$Sythetic X' = X + u(X_{nn} - X) \quad (1)$$

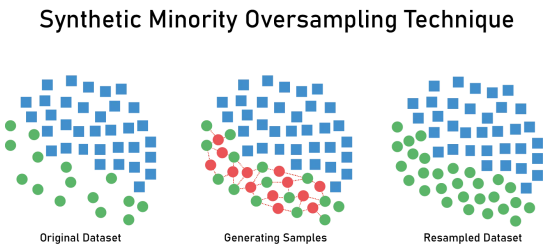

그림 2. SMOTE를 이용한 데이터 오버샘플링 프로세스
Fig. 2. Data oversampling process using SMOTE

3.2 XGBoost

XGBoost는 Gradient Boosting을 확장한 모델로, 적은 연산량으로 높은 성능을 보이는 것이 특징이다. 기본적으로 여러 개의 결정 트리를 순차적으로 학습하여 각 트리는 이전 트리의 오류를 Gradient Boosting으로 최소화한다. 그리고 각 트리의 오버피팅을 방지하기 위해 정규화(Regularization)를 사용한다. XGBoost는 빠른 학습 속도와 낮은 메모리 사용량, 그리고 과적합을 방지하는 강력한 일반화 성능으로 인해 다양한 분야에서 사용된다[16].

3.3 CatBoost

CatBoost는 Categorical Boosting의 약자로, 범주형 데이터를 효과적으로 처리하도록 만들어진 Gradient Boosting 모델이다. 일반적인 Gradient Boosting과 달리, CatBoost는 범주형 변수를 별도의 전처리 과정 없이 사용할 수 있는 장점이 있다. 또한 CatBoost는 Ordered boosting을 통해 데이터의 불균형에도 좋은 성능을 보인다[17].

3.4 LightGBM

LightGBM은 Gradient Boosting 알고리즘의 변형으로, 대용량 데이터를 빠르게 학습할 수 있는 장점이 있다. LightGBM은 리프 중심 트리 분할(Leaf-wise Splitting) 방식을 사용하여 기존의 수평 분할 방식보다 적은 연산량으로 성능 개선을 하였다. 또한 LightGBM은 메모리 사용량이 적고 학습 속도가 빠르다[18].

3.5 DNN

DNN(Deep Neural Network)은 은닉층의 개수와 깊이를 크게 설정한 인공 신경망으로 비선형성을 효율적으로 학습할 수 있는 장점이 있다. 실험에 사용한 DNN 모델은 8층으로 이루어진 모델로써, 은닉층(Hidden Layer)의 개수는 512로 설정하고, 층마다 Layer Normalization을 추가한 구조를 사용한다.

Ⅳ. 실험 환경 및 결과

본 논문의 실험 환경은 다음과 같다. 운영체제는 Ubuntu 18.04, CPU는 AMD Ryzen Threadripper PRO 5995WX, 메모리는 512GB이다. 그리고 병렬처리 보드는 RTX 3090Ti 24GB를 사용한다.

실험에 사용한 데이터는 KISTI에서 제공하는 네트워크 침입탐지 데이터를 사용한다[19]. 이 데이터는 총 데이터의 개수는 25,435,946이며, 학습 데이터는 20,348,756개, 테스트 데이터 5,087,190개로 구성된다.

학습에 사용된 파라미터는 XGBoost, LightGBM, CatBoost는 기본 설정값을 사용하고, DNN은 최적화 함수(Optimizer)는 RAdam, 배치사이즈(batch size)는 512, 에포크(Epoch)는 100, 학습률(Learning rate)은 0.001로 설정하여 실험했다.

실험결과는 모델별로 SMOTE를 적용하기 전, 후의 성능을 비교하고, 성능평가를 위해 정밀도(Precision), 재현율(Recall), 정확도(Accuracy) 3가지를 사용한다. Precision은 실제 음성인 대상 중 예측과 실제 값이 음성으로 일치한 데이터의 비율로 식 (2)와 같이 계산하고, Recall은 실제 양성인 대상 중 예측과 실제 값이 양성으로 일치한 데이터의 비율로 식 (3)과 같이 계산한다. 그리고 정확도는 전체 데이터 수 중 예측 결과와 실제 값이 같은 데이터의 비율로 식 (4)와 같이 계산한다. 그리고 Precision과 Recall의 비율을 계산할 수 있는 F1-score를 식 (5)와 같이 계산한다.

$$\mathrm{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$Accuracy = \frac{TN+TP}{TN+FN+FP+TP} \tag{4}$$

$$F1score = \frac{\mathrm{Precision}+Recall}{2 \bullet Precision \bullet Recall} \tag{5}$$

모델별 성능을 비교한 결과는 표 2와 같다. SMOTE를 전 · 후를 비교하면, XGBoost가 기존 모델 대비 0.7% 향상되었고, CatBoost와 LightGBM 모델도 0.5%~0.9% 정도 올랐다. 그러나 DNN 모델은 SMOTE 적용 여부에 상관없이 정확도 79.44%로 성능 향상이 없었다. DNN 모델의 경우 사용한 데이터의 특성 정보가 오히려 오버피팅을 유발하여 학습이 잘 안되어 정확도가 낮은 문제가 있었다. 그리고 신경망 구조는 기존의 부스팅 모델보다 연산량이 많은 문제점이 있다. 실시간성을 고려하기 위해서는 부스팅 계열을 사용하는 것이 옳다고 판단된다. 신경망 알고리즘은 기존의 부스팅 알고리즘보다 연산량이 280배 차이 나는 것을 확인할 수 있었고, 이에 따라 본 연구에서는 DNN과 같은 신경망 모델보다는 XGBoost, CatBoost, LightGBM를 사용하는 것이 효과적이며, 이 중에서 가장 속도가 빠르고 성능이 좋은 XGBoost가 최적의 모델이다.

그림 3은 네 가지 분류 모델의 ROC 커브 분석 결과, XGBoost, CatBoost, LightGBM 모델은 1에 가까운 결과를 보였고, DNN 모델은 낮은 0.50의 차트를 볼 수 있다.

표 2. 정량적 평가 결과 비교
Table 2. Comparison of quantitative evaluation results

	Model (Time)	Precision	Recall	F1Score	Accuracy
with out SMOTE	XGBoost (0.24s)	98%	99%	99%	99.23%
	CatBoost (0.66s)	97%	98%	97%	98.34%
	LightGBM (0.64s)	98%	99%	99%	99.05%
	DNN (185s)	40%	50%	44%	79.44%
with SMOTE	XGBoost (0.53s)	99%	99%	99%	99.9%
	CatBoost (1.39s)	98%	98%	98%	98.87%
	LightGBM (1.37s)	99%	99%	98%	99.99%
	DNN (349s)	40%	50%	44%	79.44%

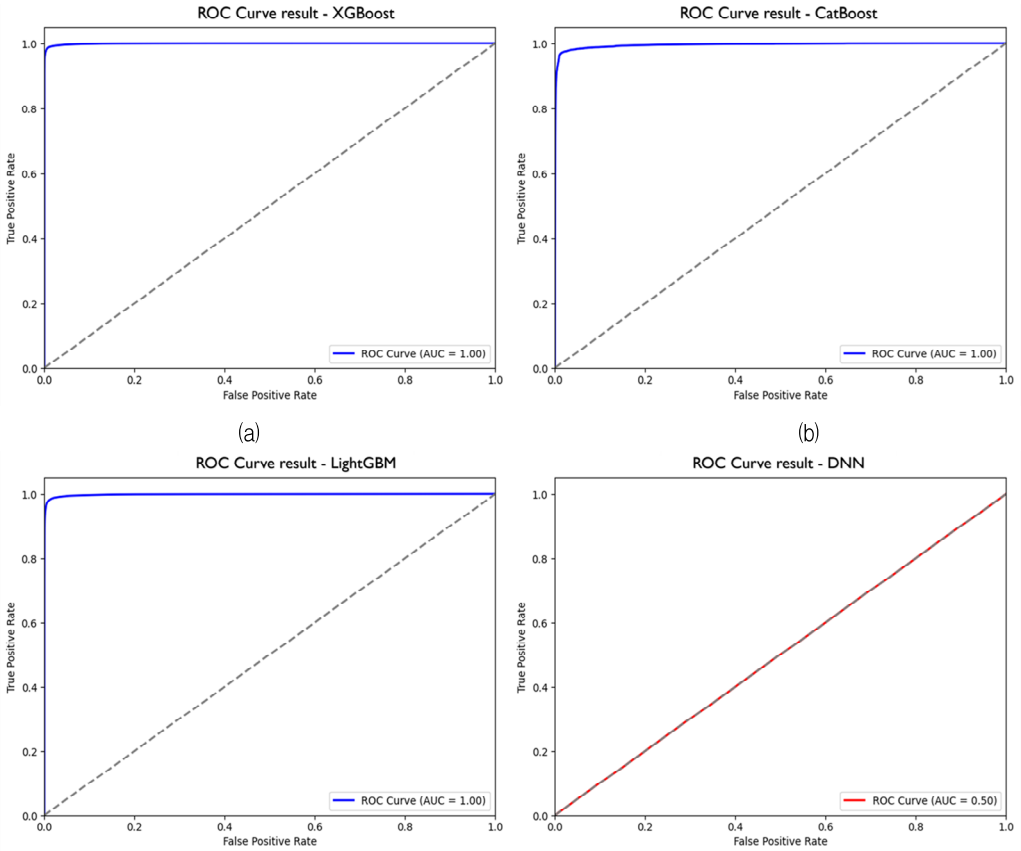

그림 3. ROC Curve를 이용한 정확도 분석결과 (a) XGboost, (b) CatBoost, (c) LightGBM, (d) DNN
Fig. 3. Visualization of accuracy analysis using ROC curves is presented as follows (a) XGBoost, (b) CatBoost, (c) LightGBM, and (d) DNN

간단한 DNN 모델의 구조는 성능이 낮아 추가적인 테스트가 필요하다. 그리고 네트워크는 여러 유형의 공격이 존재하는데 본 연구에서 사용한 데이터는 이상 여부만 판단하므로 특정 공격을 탐지하기에는 부족하다. 향후 연구에서 데이터를 추가로 사용하여 실험이 필요하다.

Ⅴ. 결론 및 향후 과제

본 연구는 네트워크 환경에서의 데이터 불균형과 실시간 연산을 수행하기 위하여 오버샘플링과 부스팅 알고리즘을 사용한 네트워크 트래픽 이상 탐지 기법을 제안하였다. 연구에 사용된 모델은 XGBoost, LightGBM, CatBoost, DNN 모델로 각각 분석한 결과 XGBoost와 LightGBM이 0.5~0.9%의 성능 향상을 보였다. 그리고 실시간 연산 수행을 위해서는 XGBoost가 0.53초로 가장 빠른 모델임을 확인할 수 있었고, 오버샘플링 방법 또한 성능 향상에 도움이 됨을 확인하였다. 또한, Boosting 알고리즘은 IDS/IPS 환경에서 높은 정확도와 실시간 이상 탐지에 적합함을 보였다. 특히 XGBoost는 비교적 짧은 연산 시간과 높은 정확도로 인해 실시간성이 요구되는 실제 보안 시스템에서의 활용 가능성이 높을 것으로 판단된다.

향후 연구는 오버샘플링의 범위를 넓히기 위하여 새로운 방법론을 사용하여 일반화가 잘되도록 개선하려고 한다. 그리고 실시간성을 고려한 가볍고, 최적화된 모델을 개발할 계획이다.

References

[1] Y. Xiao, C. Xing, and T. Zhang, "An Intrusion Detection Model Based on Feature Reduction and Convolutional Neural Networks", IEEE Access, Vol. 7, pp. 42210-42219, Mar. 2019. https://doi.org/10.1109/ACCESS.2019.2904620.

[2] Y. Lee and C. Lee, "Security System Load Reduction in SDN", The Journal of Korean Institute of Communications and Information Sciences, Vol. 46, No. 12, pp. 2251-2261, Dec. 2021. https://doi.org/10.7840/kics.2021.46.12.2251.

[3] A. Khraisat, I. Gondal, and P. Vamplew, "Survey of intrusion detection systems: techniques, datasets and challenges", Cybersecurity, Vol. 2, No. 1, pp. 1-22, Jul. 2019. https://doi.org/10.1186/s42400-019-0038-7.

[4] M. Ring, S. Wunderlich, and D. Scheuring, "A survey of network-based intrusion detection data sets", Computers & Security, Vol. 86, pp. 147-167, Sep. 2019. https://doi.org/10.1016/j.cose.2019.06.005.

[5] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)", Military Communications and Information Systems Conference (MilCIS), Canberra, ACT, Australia, pp. 1-6, Nov. 2015. https://doi.org/10.1109/MilCIS.2015.7348942.

[6] N. Moustafa, B. Turnbull, and K. K. R. Choo, "An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things", IEEE Internet of Things Journal, Vol. 6, No. 3, pp. 4815-4830, Jun. 2019. https://doi.org/10.1109/JIOT.2018.2871719.

[7] M. A. Talukder, M. M. Islam, and M. A. Uddin, "Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction", Journal of Big Data, Vol. 11, No. 1, pp. 1-33, Feb. 2024. https://doi.org/10.1186/s40537-024-00886-w.

[8] S. Bhattacharya, S. R. Krishnan, and P. K. R. Maddikunta, "A Novel PCA-Firefly Based XGBoost Classification Model for Intrusion Detection in Networks Using GPU", Electronics, Vol. 9, No. 2, pp. 219, Jan. 2020. https://doi.org/10.3390/electronics9020219.

[9] A. Alqahtani, A. Alqahtani, and M. Alqahtani, "A Comparative Study of Using Boosting-Based Machine Learning Algorithms for IoT Network Intrusion Detection", Journal of Network and Systems Management, Vol. 31, No. 4, pp. 97, Nov. 2023. https://doi.org/10.1007/s44196-023-00355-x.

[10] X. Zhang, Y. Zhao, and J. Zhao, "Research on Anomaly Network Detection Based on Self-Attention Mechanism", Sensors, Vol. 23, No. 10, pp. 4758, May 2023. https://doi.org/10.3390/s23104758.

[11] Z. Ding, Y. Zhao, and X. Chen, "Imbalanced data classification: A KNN and generative adversarial networks-based hybrid approach for intrusion detection", Future Generation Computer Systems, Vol. 131, pp. 240-254, Jun. 2022. https://doi.org/10.1016/j.future.2022.01.026.

[12] V. Balyan, S. Kumar, and S. Gupta, "A Hybrid Intrusion Detection Model Using EGA-PSO and Improved Random Forest Method", Sensors, Vol. 22, No. 16, p. 5986, Aug. 2022. https://doi.org/10.3390/s22165986.

[13] Y. Xiuye and C. Liyong, "Network Intrusion Detection Method Based on Multi-Scale CNN in Internet of Things", Wireless Communications and Mobile Computing, Vol. 2022, Oct. 2022. https://doi.org/10.1155/2022/8124831.

[14] J. Kim, H. Kim, and I. Kim, "Hierarchical LSTM-Based Network Intrusion Detection System Using Hybrid Classification", Applied Sciences, Vol. 13, No. 5, pp. 3089, Feb. 2023. https://doi.org/10.3390/app13053089.

[15] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. “SMOTE: Synthetic Minority

Over-sampling Technique", *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.

[16] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, pp. 785-794, Aug. 2016. https://doi.org/10.1145/2939672.2939785.

[17] L. Prokhorenkova, G. Gusev, and A. Vorobev, "CatBoost: unbiased boosting with categorical features", Advances in Neural Information Processing Systems 31 (NeurIPS 2018), Montréal, Canada, pp. 6639-6649, Dec. 2018.

[18] G. Ke, Q. Meng, and T. Finley, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree", Advances in Neural Information Processing Systems 30 (NIPS 2017), California, USA, pp. 3149-3157, Dec. 2017.

[19] Korea Institute of Science and Technology Information, "Network Intrusion Detection Dataset(2022)", https://aida.kisti.re.kr/data, https://doi.org/10.23057/55. [accessed: Apr. 01, 2025]

저자소개

김 인 혜 (Inhye Kim)

2009년 2월 : 평생교육진흥원 컴퓨터공학(공학사)
2011년 8월 : 경남대학교 교육대학원 컴퓨터교육(교육학석사)
2017년 8월 : 경남대학교 첨단공학과(공학박사)
2020년 3월 ~ 현재 : 경남대학교 창의융합대학 조교수
관심분야 : 미래인터넷, 무선인터넷 프로토콜, DTN, IoT, 인공지능

전 왕 수 (Wangsu Jeon)

2016년 2월 : 경남대학교 컴퓨터공학과(학사)
2018년 2월 : 경남대학교 대학원 융합IT공학과(공학석사)
2022년 2월 : 경남대학교 대학원 융합IT공학과(공학박사)
2022년 3월 ~ 현재 : 경남대학교 컴퓨터공학부 박사후과정
관심분야 : 컴퓨터 비전