https://doi.org/10.5392/JKCA.2021.21.12.896

능동형 모델 개선 피드백 기술을 활용한 보안관제 시스템 성능 개선 방안

SIEM System Performance Enhancement Mechanism Using Active Model Improvement Feedback Technology

신윤섭, 조인준
배재대학교 대학원 사이버보안학과

Youn-Sup Shin(wildoat1@gmail.com), In-June Jo(injune@pcu.ac.kr)

요약

인공지능 기반 보안관제 시스템은 운영환경에서 발생할 수 있는 학습 데이터 오류, 신규 공격 이벤트 발생으로 인한 오탐 증가 등 문제를 해결하기 위해 피드백 기능이 연구되고 있다. 그러나 한정된 관제 인력의 피드백 수행 방식은 모델 개선에 오랜 시간이 걸리고 숙련되지 않은 관제 인력의 피드백은 오히려 모델 성능 저하의 원인이 될 수 있다. 본 논문에서는 관제 인력 한계 극복, 신규 오탐 개선, 빠른 모델 성능 향상을 위한 능동형 보안관제 모델 개선 프로세스를 제안하였다. 운영 중 예측된 유사 이벤트를 군집화 하고, 피드백이 우선적으로 필요한 군집을 계산하여 운영자에게 대표 이벤트 설명이 가능한 인공지능(eXplainable AI) 기반 시각화도 함께 제시하였다. 수신된 대표 피드백은 동일 군집과 다른 데이터를 계산하여 제외하고 피드백 전파 학습 데이터를 생성한다. 준비된 학습 데이터는 초기 모델과 함께 점진적 학습을 통해 모델을 생성함으로써 성능을 향상시키는 프로세스이다. 제안 프로세스의 실효성 검증을 위해 웹 어플리케이션 방화벽 데이터셋 PKDD2007과 CSIC2012를 선택하여 3개의 시나리오를 통해 실험을 진행하였다. 실험 결과 제안된 프로세스는 피드백을 주지 않았거나 소수 운영자 피드백을 적용한 모델 성능에 비해 모든 지표에서 약 30% 이상의 성능 향상을 확인하였다.

■ 중심어 : | 인공지능 | 머신러닝 | 보안관제 | 피드백 | 점진적 학습 |

Abstract

In the field of SIEM(Security information and event management), many studies try to use a feedback system to solve lack of completeness of training data and false positives of new attack events that occur in the actual operation. However, the current feedback system requires too much human inputs to improve the running model and even so, those feedback from inexperienced analysts can affect the model performance negatively. Therefore, we propose “active model improving feedback technology” to solve the shortage of security analyst manpower, increasing false positive rates and degrading model performance. First, we cluster similar predicted events during the operation, calculate feedback priorities for those clusters and select and provide representative events from those highly prioritized clusters using XAI (eXplainable AI)-based event visualization. Once these events are feedbacked, we exclude less analogous events and then propagate the feedback throughout the clusters. Finally, these events are incrementally trained by an existing model. To verify the effectiveness of our proposal, we compared three distinct scenarios using PKDD2007 and CSIC2012. As a result, our proposal confirmed a 30% higher performance in all indicators compared to that of the model with no feedback and the current feedback system.

■ keyword : | Artificial Intelligence | Machine Learning | Feedback | Incremental Learning |

접수일자 : 2021년 11월 29일
수정일자 : 2021년 12월 16일
심사완료일 : 2021년 12월 16일
교신저자 : 조인준, e-mail : injune@pcu.ac.kr

I. 서 론

보안관제시스템(SIEM, Security Information and Event Management)은 다양한 보안 데이터를 실시간으로 수집하여 빅데이터 기반으로 통계 및 패턴 분석을 수행하고 있었다. 이를 기반으로 보안관제 업무는 보안 위협 상황을 실시간 모니터링 하고 침해사고 예방, 탐지 및 대응하고 있다. 특히 보안관제 대상 장비(방화벽, 웹방화벽, IPS, FW, Web Log 등)에 대한 지속적인 관리 및 보안 시스템에서 발생하는 다양한 이벤트를 분석 및 대응함으로써 탐지된 내용에 대한 명확한 공격 여부를 파악하고 있다. 그러나 보안관제는 엄청난 양의 로그와 경보로 인해 신규 보안 위협 및 보안 관제 업무 증가, 관제 직원의 기술 편차 심화 등 많은 문제들이 도출됨에 따라 이의 해결을 위한 대안으로 인공지능 기술을 도입하고 있다.

인공지능 기반 보안관제 시스템은 보안관제 운영인력의 업무효율 제고를 목표로 지도학습 알고리즘을 적용하여 정오탐 식별하고, 비지도 학습 알고리즘을 적용하여 이상행위를 탐지 하고 있다. 이를 통해 한정된 시간과 가용 자원으로 방대한 데이터를 신속하고 정확하게 분석하여 우선순위가 높은 고위험도 이벤트를 선별함으로써 분석에 소요되는 시간을 최소화 하고 있다. 또한, 운영기간 동안 발생하는 신규 공격 패턴 적용을 위해 신규 이벤트를 주기적으로 재학습 하거나 관제 인력의 피드백(Feedback) 기능을 제공하고 있다. 이와 같이 모델을 개선하기 위한 방안으로 새로운 데이터를 소량의 자원사용으로 신속하게 적용이 가능한 점진적 학습(Incremental Learning)기술이 활발히 연구되고 있다. 이 기술은 전체 데이터를 재학습 하는 방식과 비교하여도 유사한 정확도를 보여주고 있다[1][2].

그러나 상기에서 언급한 피드백 프로세스는 소수의 관제 요원이 발생하는 이벤트에 대해 일일이 틀린 예측 이벤트를 분석하고, 이를 토대로 예측에 대한 수정이 필요할 경우 단일 이벤트를 대상으로 피드백을 부여하는 방식이다. 이러한 소수의 관제 인력에 의한 피드백 수행 방식은 초기 학습 데이터 대비 모델에 영향이 미비하여 많은 시간이 걸리는 문제가 발생됨과 동시에 숙련되지 않은 운영자의 피드백은 오히려 모델 성능 저하의 원인이 될 수 있다.

본 논문에서는 상기와 같은 보안관제센터 운영 관점에서 빠르게 변화하는 사이버 위협 최소화, 인력 한계를 극복, 관제 업무 효율성 향상시킬 수 있는 새로운 능동형 모델 개선 피드백 프로세스를 제안하였다. 제안방안은 인공지능 기반 보안관제시스템 초기 모델에서 운영 중에 예측되는 이벤트를 군집화 하여 피드백 우선순위를 계산한다. 계산된 우선순위에 따라 대표 이벤트와 모델 설명을 운영자에게 제시함으로써 효율적인 피드백을 수집할 수 있도록 하였다. 수집된 피드백은 군집별 피드백 적용 이벤트를 분류하여 전파함으로써 소수의 피드백으로 다수의 학습데이터를 생성이 가능하도록 하였다. 이러한 점진적 학습기술을 활용하여 모델을 개선한다.

제안기술의 검증은 웹 어플리케이션 방화벽 데이터셋을 사용하여 3가지 시나리오별 모델 평가 기준의 비교 분석을 통해 효과성 및 성능 향상을 검증하였다.

II. 관련 연구 및 동향

1. 머신러닝 기반의 사이버 보안 연구 현황

디지털로 전환은 온라인 기반 경제 활동을 급성장시키고 있다. 사회전반에서 재택근무, 원격수업 등의 디지털로 전환이 일어나면서 보안 사각지대를 공격하는 사이버 위협이 증가하고 있다. 글로벌 컨설팅 기업 PWC(PWC Digital Trust Insight Snapshot)에서 2021년 1월 실시한 CEO 설문조사에서도 기업들 사이에서도 사이버 위협을 가장 우려하고 있다. 또한, CISO, CIO 대상으로 수행한 설문 조사에서도 사이버 보안과 관련된 투자액이 지속적으로 증가할 것으로 전망하고 있다[3].

주목할 부분은 사이버 보안 시장에서의 인공지능 기술이 2019년 88억 달러에서 2026년 384억 달러로 연평균 23.3%성장을 예상하고 있다. Accenture Research는 사이버 보안 리더들이 성공적인 공격 감소, 정확한 사고 탐지, 대응 품질, 비용절감을 위해 인공지능 및 자동화 기술에 우선 투자해야 한다고 권고하고 있다[4].

머신러닝 기반 지도 학습 정오탐 식별에 대한 연구는 네트워크 트래픽 정보(payload)와 다양한 알고리즘을 통해 지속적으로 성능향상이 이루어지고 있다. G. Betarte은 보안전문가에 의해 페이로드에서 문자 기반의 특징을 추출하였으며 이를 기반으로 Random Forest, KNN, SVM 등 분류 알고리즘을 통해 좋은 성능을 제안하였다[5]. Erxue Min은 성능을 높이기 위해 패킷 헤더와 페이로드에서 특징 추출을 위해 네트워크 데이터에서 중요한 통계적인 방법과 워드 벡터 후 Text-CNN을 통한 특징 추출 후 Random Forest 알고리즘을 이용한 방법을 제안하였다[6]. 국내 인공지능 기반 플랫폼 적용 기관 결과 보고서에 따르면 공격 유형별 시그니처 모델을 개발하고 Auto Encoder 및 합성곱신경망 알고리즘을 사용하여 정확도 99.67%를 달성하였다[7].

보안관제에서의 인공지능 기술 도입은 한정된 시간과 가용 자원으로 기하급수적으로 증가하고 있는 방대한 데이터를 정확하고 신속하게 분석하고 우선순위가 높은 고위험 이벤트를 선별함으로써, 방대한 보안 데이터 분석에 소요되는 인간의 한계를 극복함과 동시에 고도화된 보안 위협에 능동적으로 대처하기 위해 발전하고 있다.

2. 보안관제 피드백 연구 현황

인공지능 기반 보안관제시스템은 오탐과 미탐을 줄이고 보안관제 인력의 업무 효율성 향상을 목적으로 구축된다. 하지만, 구축 후 운영 단계에서 발생할 수 있는 초기 학습 데이터 오류 문제, 동일 공격 발생에 대한 차단 및 대응 지연, 신규 이벤트 발생으로 인한 오탐 증가 등이 요인이 되어 지속적인 모델 개선 프로세스가 필요하다. 이를 위해 다양한 개선 기술을 적용하고 있다. 대표적으로 정확도 및 정합성을 올리기 위해 예측한 결과를 사용자가 정정하여 재학습시키는 기법을 많이 사용한다. 운영자는 만족스러운 예측 결과와 불만족스러운 예측 결과에 대해 각각 다른 레이블을 부여하여 재학습함으로써 머신 러닝 모델의 성능 향상시킬 수 있었다.

재학습 기법은 보안 분야와 같은 실시간으로 이벤트 및 경보가 발생하는 환경에서 학습된 패턴을 유지하면서 지속적으로 모델을 학습할 수 있는 온라인 학습(Online Learning) 즉, 점진적 학습이 활발히 연구되고 있다[8].

초기 YuanTong[9]의 연구에서는 침입탐지 데이터에 대해 점진적 학습과 FSVM(fuzzy support vector machine) 기반의 일반학습방법을 비교하였다. 점진적 학습 방법에 대해 빠른 학습 속도에 초점을 맞춰 적용해본 결과 두 학습 방법이 유사한 수준의 성능결과를 보였다.

사이버 감시 정찰[10]에서는 C&C 서버에 수집한 데이터를 대상으로 소량의 자원으로 짧은 시간에 빠른 학습을 시킨 점진적 학습과 일괄학습을 비교 실험하였다. 그 결과 점진적 학습이 500MB 이하 메모리 환경에서 학습 소요시간을 10배 이상 단축시키는 결과를 보였다. 또한 유지훈[11], Liang 연구에서도 일괄학습과 점진적 학습을 비교 실험하였다. 그 결과 점진적 학습에 소요되는 시간이 현저하게 감소함에도 불구하고 두 학습 기법이 유사한 정확도를 보여주었다[12].

보안 분야에서 피드백 프로세스의 대표적인 연구 사례로는 미국의 MIT에서 2016년 발표한 연구를 들 수 있다. 이 연구에서는 웹로그와 방화벽로그를 대상으로 [그림 1]과 같은 피드백 운영 프로세스를 제안하였다. 관제인력의 노하우 학습과 피드백에 의한 예측 및 분석 모델 업데이트 운영은 86.8% 향상된 탐지율을 보였다[13][14].

그림 1. AI security analytics operations

국내 인공지능기반 보안관제 솔루션도 보안 전문가와 인공지능의 결합을 통해 오탐을 줄이고 정확도를 향상시킬 수 있는 선순환 구조의 피드백 프로세스를 개발하여 제공하고 있다[15].

3. 모델 개선 피드백 프로세스 고찰

보안운영센터의 31%는 2~5명, 36%는 6~25명 정도

의 인력으로 구성되고, 보안관제시스템 49%가 일일 5,000여개 경보를 발생시킨다. 또한, 보안운영센터 운영인력의 70%는 매일 10개 이상의 경보를 조사하고, 단일 경보를 조사하는데 10분 이상 걸린다고 응답하였다. 이들 조사 경보 중 50% 이상이 오탐으로 보고되고 있다. 이와 같은 결과는 현재의 모델 개선의 필요성을 나타냄과 동시에 일일 100여건의 소수 피드백 데이터가 생성될 것으로 추정하고 있다[16-18].

이와 같이 보안운영센터의 관제 인력은 이미 압도적으로 많은 수의 이벤트 및 경보를 지속적으로 조사해야 하는 상황에 직면해 있다. 기존 보안관제 피드백 연구는 점진적 학습기법을 사용하여 학습시간과 시스템 자원 절감 효과에 초점이 맞추어 연구되었다. 그러나 대부분의 보안운영센터에서 문제가 되고 있는 관제 인력 부족과 분석 건수 한계는 고려되지 않았다. 소수의 피드백은 모델 개선을 위해 너무 많은 시간이 소모되는 문제와 대용량 이벤트 및 경보 발생 환경에 적용할 경우 모델 개선 효과가 반감되는 현상이 예상된다.

본 논문에서는 위 피드백 프로세스와 점진적 학습기술을 대용량 이벤트 및 경보 발생 환경에 적용할 경우, 관제 인력의 분석 건수 한계를 고려하여 초기 학습 데이터 오류 문제, 신규 이벤트 발생으로 인한 오탐 등을 빠르고 효율적으로 해결할 수 있는 최적의 피드백 기법 적용 방안을 제안하였다.

Ⅲ. 머신러닝 기반 능동형 보안관제 모델 개선 프로세스 제안

본 논문에서 제안하는 머신러닝 기반 능동형 보안관제 모델 개선 프로세스는 [그림 2]와 같다. 인공지능 보안관제 시스템 초기 모델에 의해 예측된 이벤트를 수집하는 단계로 시작하여 피드백 이벤트 선정, 피드백 제안 및 수행, 피드백 전파 및 적용 단계로 구성된다. 프로세스 설명과 초기 모델 생성을 위해 공개된 웹 어플리케이션 방화벽 데이터셋을 사용하였다.

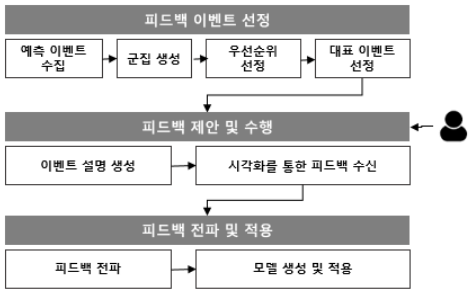


그림 2. 능동형 보안관제 모델 개선 프로세스

1. 데이터 전처리 및 초기 모델 생성

본 논문에서 사용된 데이터셋은 웹 어플리케이션 방화벽 이벤트인 PKDD2007[19]과 CSIC2012[20]를 선정하였다. 데이터 전처리는 이벤트의 정확한 판단을 위해 901개의 범주형(Category), 숫자형(Numeric), 텍스트형(Text) 피처를 추출하였다[21]. [표 1]과 같이 PKDD2007 데이터셋은 정상(Normal)과 7가지 유형의 공격(Attack) 레이블을 포함하고 있으며 CSIC2012 데이터셋은 정상(Normal)과 9가지 유형의 공격(Attack) 레이블을 포함한다.

표 1. PKDD2007, CSIC2012 데이터셋 레이블 비교

구분	PKDD2007	CSIC2012
레이블 수	7개	9개
공통 레이블	Valid	Valid
	SqlInjection	SqlInjection
	XSS	XSS
	XPathInjection	XPathInjection
	LdapInjection	LdapInjection
	SSI	SSI
데이터별 레이블	OSCommanding	CRLFi
	-	FormatString
	-	BufferOverflow
데이터 수	75,728	57,676

제안하는 프로세스 설명 및 실험을 위해 초기 모델은 Catboost 알고리즘과 PKDD2007 데이터셋 100%를 사용하였다. 운영 시 신규로 발생하는 공격 이벤트는 상이한 4개의 레이블이 포함된 CSIC2012 데이터셋

70%를 사용하여 피드백을 부여하였다. 이를 통해 능동형 모델 개선 기술을 설명하면 다음과 같다.

2. 능동형 모델 개선 프로세스 단계

2.1 피드백 이벤트 선정

초기 모델에서 발생한 신규 이벤트 중 피드백이 우선적으로 수행되어야할 이벤트를 선출한다. 이를 위해 유사한 이벤트를 군집화하고 군집별 우선순위 및 대표 이벤트를 계산하여 운영자에게 제시한다.

신규 이벤트를 대상으로 가장 최적의 군집을 생성하기 위해 유사성(Similarity)을 확인하는 유클리디언 거리(Euclidean Distance) 기반의 K-means 알고리즘을 사용하였다. K-means 알고리즘은 K개의 군집 별 중심(centroid)을 생성하여 유클리디언 거리 기반으로 각 중심에서 가까운 그룹을 분류 한다. CSIC2012 데이터셋을 K-means 알고리즘을 사용, [그림 3]과 같이 64개의 군집을 생성하였다.

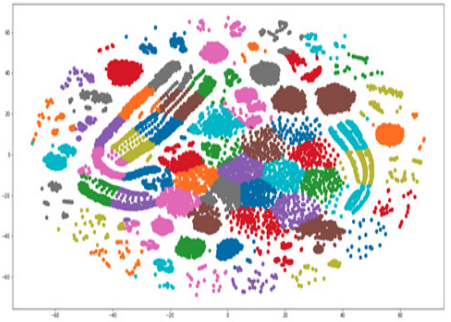


그림 3. CSIC2012 데이터셋 군집 시각화

64개의 이벤트 군집 중 우선적으로 피드백 되어야 하는 이벤트를 선정하기 위해 군집의 신뢰도 점수(CSS, Cluster Confidence Score)를 측정한다. 군집의 신뢰도 점수는 초기 모델이 예측한 이벤트별 신뢰도(ICS, Individual Confidence Score)와 군집의 중심에 대한 거리를 기반 지수(B, distance to cluster center)를 사용했다.

$$Cluster\ Confidence\ Score,\ CSS = \frac{\sum_{1}^{n} B_i ICS_i}{n-1} \quad (1)$$

위 두 가지 지수를 동시에 고려하기 위해 곱한 후 군집별 평균을 계산한다. 식 (1)에서 군집의 신뢰도 점수는 0부터 1까지 나타난다. 이는 상대적 지수로써, 낮을수록 더 우선적으로 피드백이 필요한 군집으로 판단한다.

피드백이 필요한 우선순위 군집별 군집의 데이터 성질을 가장 잘 표현하는 대표 이벤트를 선정하기 위해 코사인 유사도(Cosine Similarity) 점수를 측정한다.

$$\text{similarity} = = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \quad (2)$$

이벤트의 유사성(similarity)은 식 (2) 에서와 같이 코사인 유사도 함수를 사용하여 군집 원소 A(i) 와 군집 내 이벤트의 데이터 평균값 B(i) 을 계산한다.(2), 유사성 지수는 군집의 원소와 군집 내 데이터 평균이 완전히 같을 경우 양수, 최대 점수 1, 90°의 각을 이룰 경우 0, 180°로 완전히 반대 방향인 경우 음수, 최대 점수 -1의 값을 갖는다. 점수가 가장 높은 대표 이벤트를 군집의 대표로 운영자에게 피드백을 제안한다.

2.2 피드백 제안 및 수행

피드백 이벤트 선정을 통해 최우선 피드백 군집과 대표 이벤트가 선정되었다. 운영자는 피드백 레이블을 지정하기 위해 대표 이벤트를 분석해야 한다. 본 프로세스에서는 원활한 분석과 빠른 피드백 수행을 위해 설명이 가능한 인공지능(XAI, eXplainable AI) 기술을 통해 대표 이벤트의 설명 근거를 생성했다.

이벤트 설명을 위해 해석하고자 하는 예측 데이터 값의 근방에서 모델이 어떻게 작동하였는지 설명하는 LIME(Local Interpretable Model-agnostic Explanations) 라이브러리를 사용하였다. LIME 은 예측 이벤트 설명을 위해 블랙박스 모형의 개별 예측 이벤트를 설명하기 위해 국소적 대리 모델(local surrogate model)을 생성하고, 개별 예측 이벤트를 입력으로 예측 설명 데이터를 시각화하여 표시할 수 있다. 아를 위해 이벤트 군집 생성시 만들어진 데이터와 군집 명을 레이블로 학습하여 국소적 대리 모델을 생성하고 대표이벤트를 개별 예측 이벤트로 입력함으로써 예측에 대한 설명 근거를 생성하였다.

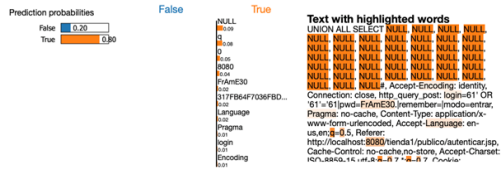


그림 4. XAI 기반 예측 근거 시각화

[그림 4]는 우선순위 군집 (Cluster 1) 대표 이벤트를 개별 예측 이벤트로 입력하여 모델의 근거 데이터를 시각화한 예시이다. 우선 군집의 대표이벤트 설명은 양의 상관관계가 있는 특징은 True (빨간색) 으로 표시되고 그렇지 않으면 False (파란색) 으로 표시 되어 긍정적, 부정적인 상관관계를 표시한다. 예시 이벤트는 “NULL”, “q”, “sysusers” 등의 문자열이 긍정적 상관관계로 시각화하여 설명함으로써 운영자가 해당 군집이 “SQL Injection” 공격에 유사함을 빠르게 판단, 피드백을 부여할 수 있다.

2.3 피드백 전파 및 적용

운영자가 부여한 대표 이벤트의 피드백 레이블을 기준으로 동일 군집에 대해 레이블을 전파한다. 소수의 피드백으로 다수의 학습 데이터를 만드는 과정이다.

우선 동일 군집 내에 대표 이벤트와의 상이한 데이터를 제외하고 전파한다. 이를 위해 피드백 적용 이벤트와 비적용 이벤트를 분리하기 위해 집단 내 표본 중 다른 데이터(Outlier)를 분류할 수 있는 OCSVM (One-class Support Vector Machine) 알고리즘을 사용했다. OCSVM 은 피드백 받은 대표 이벤트의 특징 벡터와 위치 특징 값을 사용하여 학습하고 대표 이벤트와 유사한 이벤트를 구분 짓는 경계를 설정한다. 해당 경계를 기준으로 경계 내에 포함될 경우 피드백을 전파할 적용 이벤트로, 벗어난 경우 비적용 이벤트로 판단한다.

예로 우선순위 군집 (Cluster 1) 데이터에 대표 이벤트를 기준으로 Inlier, Outlier를 판단하였으며 Inlier 적용 이벤트만 피드백 학습 데이터로 사용한다. 적용 이벤트가 선정되면 피드백 레이블을 전파한다.

군집별, 대표 이벤트를 통해 수신된 피드백 레이블(class)을 적용 이벤트(Inlier) 에 전파하여 추가 학습 데이터가 준비되었다. 새로운 모델을 만들기 위해 온라인 학습 기법인 점진적 학습을 수행하여 모델을 생성한다. CatBoost 알고리즘은 초기 생성된 모델을 로드하여 피드백 학습 데이터를 추가하는 점진적 학습 수행했다.

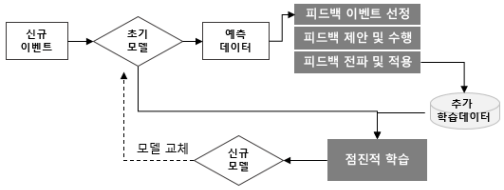


그림 5. 점진적 학습에 의한 신규모델 교체

[그림 5]와 같이 점진적 학습에 의한 신규모델 교체 프로세스는 신규 모델 학습과 초기 모델 교체 프로세스가 분리됨으로써 운영 시 업무의 연속성을 보장할 수 있다.

Ⅳ. 실험 및 평가

본 장에서는 3장에서 제안한 방법과 데이터를 이용하여 소수의 피드백을 통한 능동형 모델 개선 프로세스가 모델 성능 향상에 실효성이 있는지 실험을 수행하였다.

실험을 위해서 아래 [표 2]와 같이 3가지 시나리오를 통해 실험을 수행하고 평가한다.

표 2. 능동형 모델 개선 프로세스 실험 시나리오

	시나리오 1	시나리오 2	시나리오 3
목표	초기 모델에 피드백 없이 신규 이벤트를 정확히 판단하는가?	초기 모델에 운영자 판단에 의한 소수 피드백을 적용하여 신규이벤트를 정확히 판단하는가?	초기 모델에 운영자 판단에 의한 소수 피드백을 능동형 모델 개선에 의해 적용하고 신규이벤트를 정확히 판단하는가?
실험 방법	A 데이터 (100%) 초기 모델 생성, B데이터(30%)로 예측 성능 측정	A 데이터 (100%) 초기 모델 생성, B 데이터 피드백 (100건) 점진적 학습 신규 모델 생성, B데이터(30%)로 예측 성능 측정	A 데이터 (100%) 초기 모델 생성, B 데이터 피드백 (100건), 능동형 모델 개선 프로세스 수행으로 B 데이터(70%)에 전파, 점진적 학습 신규 모델 생성, B데이터(30%)로 예측 성능 측정
초기 모델	PKDD2007 데이터 (100%)	PKDD2007 데이터 (100%)	PKDD2007 데이터 (100%)
피드백 수행	-	CSIC2012 데이터 피드백 (100건)	CSIC2012 데이터 피드백 (100건)

데이터 전파	-	-	능동형 모델 개선 프로세스 수행, CSIC2012 데이터 (70%) 전파
피드백 학습 데이터 건수	-	100건	40,373건 중 13,764건 추출
검증	CSIC2012 데이터 (30%)	CSIC2012 데이터 (30%)	CSIC2012 데이터 (30%)

시나리오 1 (No-FB) : 초기 모델에 피드백이 없이 신규 이벤트를 정확히 판단하는지 실험

첫 번째 시나리오는 초기 모델이 피드백 없이 운영 중 발생되는 새로운 공격 유형이 포함된 데이터를 잘 예측할 수 있는지에 대한 관점이다. PKDD2007 데이터 100%를 활용하여 초기 모델을 생성하였으며 6개의 공통 레이블, 3개의 다른 레이블이 포함된 CSIC2012 데이터 30%를 검증 데이터로 사용하였다.

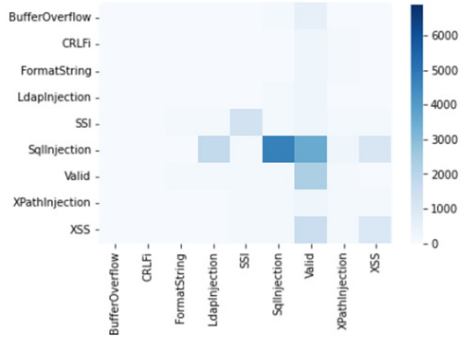


그림 6. 시나리오 1 (No-FB) Confusion Matrix

표 3. 시나리오 1 (No-FB) 성능

	precision	Recall	Accuracy	F1-Score
시나리오1 (NO-FB)	55.07	53.08	58.71	54.06

Confusion Matrix [그림 6]을 보면 6개의 공통 레이블 중에서도 SqlInjection 공격명은 다른 공격명으로 분류 되고 있으며, CRLFi, FormatString, BufferOverflow 와 같은 새로운 공격 유형에 대해서는 대부분 정상(Valid) 등 제대로 분류되지 않고 있는 점을 확인했다. [표 3] 예측 성능은 Precision 55.07%, Recall 53.08%, Accuracy 58.71%, F1-Score 54.06% 로 낮은 성능 결과를 보여준다.

시나리오 2 (Human-FB) : 초기 모델에 운영자 판단에 의한 소수 피드백을 적용하여 신규 이벤트를 정확히 판단하는지 실험

두 번째 시나리오는 초기 모델이 운영자 판단에 따라 소수의 피드백을 받아 모델에 적용하고 운영 중 발생되는 새로운 공격 유형이 포함된 데이터를 잘 예측할 수 있는지에 대한 관점이다. PKDD2007 데이터 100%를 활용하여 초기 모델을 생성하였으며 CSIC2012 데이터 중 운영자 판단에 의해 소수 이벤트 100건을 피드백 데이터로 점진적 학습에 적용하여 모델을 생성하였다. 생성된 모델을 사용하여 시나리오 1번과 동일하게 6개의 공통 레이블, 3개의 다른 레이블이 포함된 CSIC2012 데이터 30%를 검증 데이터로 사용하였다.

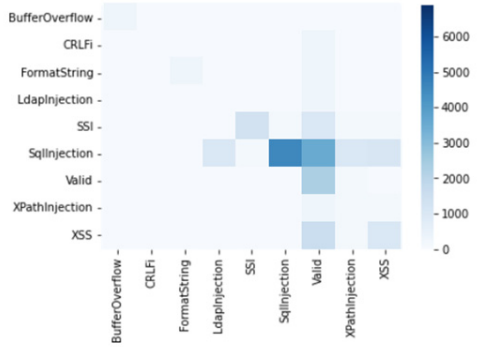


그림 7. 시나리오 2 (Human-FB) Confusion Matrix

표 4. 시나리오 2 (Human-FB) 성능

	precision	Recall	Accuracy	F1-Score
시나리오2 (Human-FB)	59.01	57.06	59.81	58.02

Confusion Matrix [그림 7]을 보면 일부 정상(Valid) 으로 분류되었던 CRLFi, FormatString, BufferOverflow 이벤트가 공격으로 분류되고 있으나 시나리오 1과 비교하여 거의 변동된 부분을 찾기 힘들다. [표 4] 예측 성능은 Precision 59.01%, Recall 57.06%, Accuracy 59.81%, F1-Score 58.02% 로 성능에서도 큰 변화가 없다.

시나리오 3 (AL-FB) : 초기 모델에 운영자 판단에 의한 소수 피드백을 능동형 모델 개선에 의해 적용하여

신규 이벤트를 정확히 판단하는지 실험

세 번째 시나리오는 초기 모델이 운영자 판단에 따라 소수의 피드백을 받아 능동형 모델 개선에 의해 모델을 개선하고 운영 중 발생되는 새로운 공격 유형이 포함된 데이터를 잘 예측할 수 있는지에 대한 관점이다. PKDD2007 데이터 100%를 활용하여 초기 모델을 생성하였으며 CSIC2012 데이터 중 운영자 판단에 의해 소수 이벤트 100건 피드백 데이터 기반으로 능동형 모델 개선 프로세스에 적용하였다. 운영자의 피드백 100건으로 CSIC2012 70% 데이터 40,373건에 능동형 모델 개선 프로세스를 적용 하였을 때 13,764건의 피드백 학습데이터가 생성되었으며 생성된 데이터를 점진적 학습에 적용하여 모델을 생성하였다. 생성된 모델을 사용하여 시나리오 1,2번과 동일하게 6개의 공통 레이블, 3개의 다른 레이블이 포함된 CSIC2012 데이터 30%를 검증 데이터로 사용하였다.

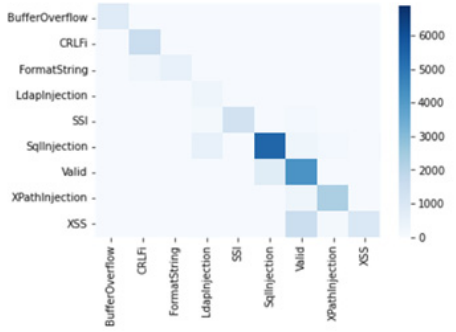


그림 8. 시나리오 3 (AL-FB) Confusion Matrix

표 5. 시나리오 3 (AL-FB) 성능

	precision	Recall	Accuracy	F1-Score
시나리오2 (Human-FB)	59.01	57.06	59.81	58.02

Confusion Matrix [그림 8]을 보면 SqlInjection 공격명, 정상(Valid) 등 6개의 공통 레이블뿐만 아니라 CRLFi, FormatString, BufferOverflow 와 같은 신규 공격명 또한 정상적으로 분류함을 확인 할 수 있다. [표 5] 예측 성능 Precision 88.61%, Recall 90.53%, Accuracy 89.42%, F1-Score 89.56% 로 높은 성능 향상을 확인하였다.

V. 결 론

보안관제시스템은 다양한 보안 데이터를 빅데이터 기반으로 통계 및 패턴 분석을 수행하였다. 신규 보안 위협 증가, 대용량 이벤트 분석의 용이성, 보안관제 업무 증가로 인해 발생하는 다양한 문제를 해결하고자 인공지능 기술이 활발히 연구 및 적용 되었다. 그러나 운영환경에서 발생할 수 있는 초기 학습 데이터 오류, 신규공격 이벤트 발생으로 인한 오탐 증가 등의 문제는 추가적인 피드백 기술 연구에 의해 보완되고 있다.

하지만 지금까지의 피드백 기술은 관제요원이 이벤트에 대해 일일이 예측 이벤트를 분석하여 수정이 필요한 이벤트에 피드백을 부여하는 방식이다. 또한, 소수의 관제 인력에 의한 피드백 수행 방식은 초기 학습 모델에 미비한 영향으로 오랜 시간이 걸리고 숙련되지 않은 관제 인력의 피드백은 오히려 모델 성능 저하의 원인이 될 수 있다.

본 논문에서는 이러한 한계를 극복하기 위한 능동형 보안관제 모델 개선 프로세스를 새롭게 제안하였다. 제안하는 방법은 이벤트를 수집하는 단계로 시작하여 피드백 이벤트 선정, 피드백 제안 및 수행, 피드백 전파 및 적용 단계로 구성되며 이를 통해 효율적인 피드백 레이블 수집과 점진적 학습을 통한 성능향상을 확인하였다.

또한, 소수의 피드백을 통해 제안된 능동형 모델 개선 프로세스가 모델의 성능을 향상에 실효성이 있는지 실험을 수행하였다. 최대한 유사성과 추가 피드백 데이터를 반영하기 위해 웹 어플리케이션 방화벽 데이터셋으로 PKDD2007과 CSIC2012를 선택하였다. 이를 대상으로 3개의 시나리오를 설정하여 실험을 진행하였다. 실험 결과 제안된 프로세스는 피드백을 주지 않았거나 소수 운영자 피드백을 적용한 모델 성능에 비해 모든 지표에서 [그림 9]와 같이 약 30% 이상의 성능 향상을 보였다.

능동형 모델 개선 프로세스는 보안관제 업무에서 소수의 관제인력을 통해 생성되는 피드백으로 모델을 신속하게 개선할 수 있다. 시간 및 인력 활용 측면에서 개선된 프로세스를 적용함으로써 신규 공격 변화에 신속하게 대응하고, 시스템 및 인적 자원에 대한 비용 절감

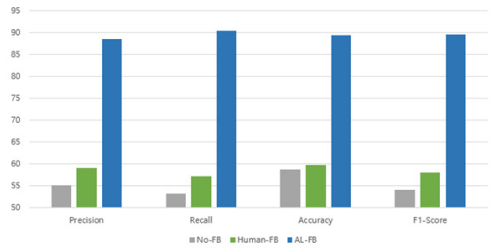


그림 9. 시나리오 성능 지표 비교

효과에 따라 보안관제 업무 개선에 기여할 수 있을 것으로 보인다. 이를 통해 인공지능 도입 목적에 따라 보안관제 시스템 성능을 효율적으로 개선하고 관제 인력은 좀 더 창조적이고 전문적인 업무에 집중할 수 있도록 도움이 될 수 있다.

본 연구의 한계점은 다수의 관제 인력을 운영하고 있는 환경에서 개개인의 지식, 경력 등에 따라 피드백 레이블이 다를 수 있고 모델 성능에 영향을 줄 수 있다. 이는 향후 관제 인력별 피드백 데이터를 통해 각각의 모델을 생성하고 평가 및 검증함으로써 모델을 선택적으로 운영하는 방식으로 개선할 수 있다.

앞으로는 본 연구를 기반으로 XAI 기반 피드백 대상 데이터를 선정하며 자동화된 머신러닝(AutoML, Auto Machine Learning) 기술로 최적의 알고리즘으로 스스로 재학습함으로써 분석가의 개입을 최소화할 수 있는 기술을 연구를 하고자 한다.

참 고 문 헌

[1] M. Maloof and R. Michalski, "A method for partial-memory incremental learning and its application to computer intrusion detection," IEEE International Conference on Tools with Artificial Intelligence, 1995(11).

[2] Xiaoming Yuan, "A Concept Drift Based Ensemble Incremental Learning Approach for Intrusion Detection," IEEE International Conference on Internet of Things, 2018(8).

[3] Markets and Markets, "Artificial Intelligence in Cybersecurity Market," MarketsandMarkets Research Private Ltd. All rights reserved, 2020.

[4] Accenture Research, "State of Cyber Resilience Report," 2020.

[5] Gustavo Betarte, "Web Application Attacks Detection Using Machine Learning Techniques," IEEE International Conference on Machine Learning and Applications (ICMLA), 2018(12).

[6] Erxue Min, "Anomaly-Based Intrusion Detection through Text-Convolutional Neural Network and Random Forest," Hindawi Security and Communication Networks, 2018(7).

[7] 한국지능정보사회진흥원, *2020년도 인공지능기반 적응형 보안시스템 3차 구축 사업추진결과보고서*, 2021.4, https://egov.nia.or.kr/

[8] Shuai Zheng, "Effective Information Extraction Framework for Heterogeneous Clinical Reports Using Online Machine Learning and Controlled Vocabularies," JMIR Publications, 2017(5).

[9] YuanTong Dong, "Research of Intrusion Detection Method Based on IL-FSVM," IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2019.

[10] 신경일, "사이버 감시정찰의 정보 분석에 적용되는 점진적 학습 방법과 일괄 학습 방법의 성능 비교," 정보처리학회논문지, KIPS transactions on software and data engineering, 소프트웨어 및 데이터 공학, Vol.7, No.3, pp.99-106, 2018(7).

[11] 유지훈, "지속적인 사이버 공간의 위협 정보 학습을 위한 특징 선택과 점진적 학습," 국방보안연구소, 국방과보안, 제1권, 제2호, pp.203-225, 2019.

[12] Liang, Yan, et al. "Automatic Security Classification Based on Incremental Learning and Similarity Comparison," IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2019.

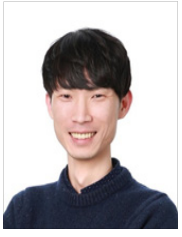
[13] MIT, "Training a big data machine to defend," IEEE International Conference on Intelligent Data and Security (IDS), 2016.

[14] PatternEX, "The Holy Grail of: Teaming

humans and machine learning for detecting cyber threats," ACM SIGKDD Explorations Newsletter, 2019.
[15] 이글루시큐리티, "SPiDER TM AI Edition 제품 설명," http://www.igloosec.co.kr/
[16] SANS, 2018 Security Operations Center Survey, A SANS 2018 Survey, 2018.
[17] CISCO, "Anticipating the Unknowns," CISCO CYBERSECURITY SERIES 2019 ASIA PACIFIC CISO BENCHMARK STUDY, 2019.
[18] CRITICALSTART Research Report, *THE IMPACT OF SECURITY ALERT OVERLOAD,* 2019.
[19] http://www.lirmm.fr/pkdd2007-challenge/
[20] https://www.tic.itefi.csic.es/torpeda/datasets.html
[21] 정일옥, *전이학습과 불균형 데이터 처리를 통한 침입 탐지 성능향상에 관한 연구,* 고려대학교, 박사학위논문, 2021(8).

저 자 소 개

신 윤 섭(Youn-Sup Shin) 정회원



■ 2003년 2월 : 배재대학교 컴퓨터공학과(공학사)
■ 2011년 6월 : 동국대학교 정보보호학과(공학석사)
■ 2012년 11월 ~ 현재 : 이글루시큐리티 책임연구원

〈관심분야〉 : 정보보호, 인공지능, 머신러닝

조 인 준(In-June Jo) 정회원

■ 1982년 2월 : 전남대학교 계산통계학과 졸업
■ 1985년 2월 : 전남대학교 전자계산학과 석사
■ 1999년 2월 : 아주대학교 컴퓨터공학과 박사
■ 1983년 ~ 1993년 : 한국전자통신연구원 선임연구원
■ 1994년 ~ 현재 : 배재대학교 사이버보안학과 교수

〈관심분야〉 : 정보보호, 컴퓨터네트워크보안, 컴퓨터시스템응용