네트워크 침입탐지 시스템에서 CNN 기반 특징 추출 연구 동향

양서연[1], 오현영[2*]
[1]가천대학교 인공지능학과 학부생
[2]가천대학교 인공지능학과 교수
{gkfhfhz123, hyoh}@gachon.ac.kr

Research Trends of CNN-based Feature Extraction in Network Intrusion Detection Systems

Seo-Yeon Yang, Hyunyoung Oh
Dept. of AI, Gachon University

요 약

네트워크 침입탐지시스템(NIDS)은 시그니처 기반 탐지를 넘어 이상 탐지로 확장되었고, 이 과정에서 머신러닝과 딥러닝 기법의 활용이 보편화되었다. 특히 합성곱신경망(CNN)은 데이터로부터 계층적 특징을 자동 학습하는 능력으로 주목받지만, 네트워크 트래픽의 고차원성과 비정형성으로 인해 학습 전 및 학습 중 단계에서 특징 재표현(feature re-encoding) 이나 특징 최적화(feature selection)를 병행할 필요가 여전히 크다. 본 논문은 CNN 기반 NIDS에서 제안된 특징 추출 연구의 흐름을 정리하고, 의미적 재부호화와 유전 알고리즘 기반 CNN 구조, 입력 동시 최적화, 다중 특징 융합을 중심으로 비교 분석한다. 이를 통해 CNN을 NIDS에 적용할 때 데이터 표현의 보존과 강조, 구조 탐색의 자동화가 성능 및 일반화에 미치는 영향을 논의한다.

1. 서론

머신러닝은 일찍부터 이상 탐지에 적용되어 왔으며, 네트워크 IDS도 통계적 탐지에서 SVM과 랜덤 포레스트, 나아가 딥러닝 모델로 발전해 왔다. CNN은 이미지 처리 분야에서 탁월한 성능을 보여 왔고, 계층적 특징을 자동으로 학습하는 능력 때문에 NIDS로의 전용이 시도되고 있다. 그러나 네트워크 트래픽은 이미지와 달리 고정 격자 구조가 없는 비정형 시계열, 토큰 데이터이기 때문에, 단순한 1차원 배열 입력만으로는 프로토콜 필드 경계나 토큰 순서 같은 의미 정보가 손실되어 학습 효율이 저하될 수 있다. 따라서 CNN의 자동 특징 학습 능력에만 의존하기보다, 학습 전 또는 학습 과정에 표현 보존과 의미 강조를 위한 명시적 특징 추출, 변환이 병행되어야 한다. 본 논문은 이러한 문제의식 아래, CNN 기반 IDS에서 제안된 의미적 재부호화[1], 유전 알고리즘 기반 CNN 최적화[2], 특징 융합 기반 딥러닝[3]을 중심으로 기술적 흐름을 고찰한다.

2. CNN 기반 IDS와 특징 추출 기법

2.1 의미적 재부호화(Semantic Re-encoding)

Wu 등[1]은 네트워크 트래픽의 비정형성으로 인한 의미 정보 손실을 보완하기 위해 의미적 재부호화를 제안하였다. 핵심 아이디어는 정상과 공격 트래픽을 구분하는 데 유의미한 토큰(단어) 신호를 강조하고, 기능적으로 유사한 토큰을 병합하여 차원을 축소하면서도 의미를 보존하는 것이다. 구체적으로 Word Table Reordering에서 정상/공격 데이터의 단어 빈도를 비교해 종합 단어 빈도(CWF) 차이를 기준으로 어휘를 재배열하고, Word Re-mapping을 통해 모든 샘플을 등길이 벡터로 사상한다. 이렇게 재표현된 입력을 CNN에 제공하면, 각 벡터가 정상/공격뿐 아니라 공격 유형 간 의미 차이를 반영하게 되어 학습이 구조화된 표현 공간에서 이루어진다. NSL-KDD와 UNSW-NB15 등 데이터셋에서 기존 CNN 대비 정확도와 탐지율 향상이 확인되었다[1]. 요컨대, 의미적 재부호화는 사전 특징 추출, 강조 단계를 통해 CNN의 표현 학습을 보조하는 접근이다.

* 교신저자

2.2 유전 알고리즘 기반 CNN 최적화

CNN은 계층적 구조를 통해 입력으로부터 자동으로 특징을 추출할 수 있지만, 입력 데이터 차원이 크고 불필요한 속성이 섞여 있는 경우 학습 효율이 떨어질 수 있다. 이를 보완하기 위해 Nguyen 등[2]은 유전 알고리즘(Genetic Algorithm, GA)을 적용한 CNN 최적화 기법을 제안하였다. GA는 개체 집단을 진화시키는 과정에서 적합도가 높은 해를 선택하는 탐색 기법으로, 이 논문에서는 CNN의 네트워크 블록 연결 방식과 입력 특징 하위집합을 동시에 탐색하는 데 활용되었다. GA는 초기 개체 집단을 무작위 비트열(염색체)로 생성하고, 각 개체는 선택된 특징의 하위집합과 CNN 블록 연결 구조를 동시에 부호화한다. 이후 교차와 돌연변이를 반복하여 새로운 세대를 형성하고, 각 개체는 CNN을 학습한 뒤 탐지 정확도, 오탐률, 연산 비용 등을 종합한 적합도(fitness) 값으로 평가된다. 이를 통해 불필요한 입력 특징은 자연스럽게 제거되고, 중요한 특징이 강조된 상태에서 CNN이 학습할 수 있다. 동시에 블록 연결 방식을 탐색하여 과도하게 깊거나 복잡한 구조는 배제하고, 최적의 모델이 선택된다. 이와 같은 기법은 CNN을 진화적 탐색을 통해 CNN이 학습하는 특징 표현 공간 자체를 최적화했다는 점을 주목할 수 있다. 실험 결과 GA 기반 CNN은 기존 CNN보다 학습 속도가 개선되었고, NSL-KDD, UNSW-NB15 등 주요 데이터셋에서 정확도 및 탐지율 향상뿐 아니라 FPR을 낮추는 데에도 기여하였다.

2.3 특징 융합 기반 딥러닝 접근

Ayantayo 등[3]은 CNN의 특징 추출 능력을 보완하기 위해 여러 특징 공간을 결합(feature fusion)하는 방식을 제안하였다. 네트워크 트래픽은 단일 표현만으로 설명하기 어려운 다양한 속성을 가지는데, 연구진은 이를 통계적 특징(패킷 길이, 바이트 수 등), 프로토콜 기반 특징(플래그, 서비스 타입 등), 시계열적 패턴으로 나누어 각각 추출하였다. 이후 이러한 이질적 특징들을 CNN과 함께 학습하도록 결합하는데, CNN은 지역적 패턴과 공간적 상관관계를 잘 포착하는 반면, 다른 특징 집합은 전역적 구조나 시간적 연속성을 반영하여 학습 과정에서 상호 보완적 역할을 한다. 특히 연구에서는 특징 추출 기법을 여러 특징을 결합한 후 모델에 입력하는 early-fusion과 각 특징으로 개별 학습 후 결과를 병합하는 late-fusion 의 두 가지로 나누어 조합 방식에 따른 성능 차이를 비교하였다. 실험 결과, 단일 CNN 기반 학습보다 다층적 특징 융합을 적용했을 때 탐지 정확도와 재현율이 향상되었으며, 공격 유형 간 구분 능력이 개선되었다. 이는 CNN이 추출하는 저차원 지역 패턴에, 명시적으로 정의된 전역적, 시계열적 특징을 더해 IDS의 탐지 성능을 강화한 사례라 할 수 있다. 따라서 특징 융합 접근은 CNN을 중심에 두되, 다양한 특징 추출 과정을 병행해 전체 표현 공간을 풍부하게 만드는 방법으로 해석할 수 있다.

3. 결론 및 향후 연구

CNN은 계층적 특징을 자동 학습하는 강점을 가지지만, 네트워크 트래픽의 비정형성, 고차원성으로 인해 단독 사용만으로는 한계가 존재한다. 최근 연구는 CNN을 특징 추출기로 활용하면서 명시적 재표현(의미적 재부호화)[1], 진화적 최적화(GA 기반 구조, 입력 동시 탐색)[2], 다중 특징 융합[3]을 결합해 성능과 일반화를 개선하는 경향을 보였다. 나아가 CNN을 전통 ML 또는 RNN/트랜스포머 계열과 결합하는 하이브리드 구조도 활발하다.

향후 과제로는 (1) 신뢰성 향상을 위한 예측 설명가능성(XAI) 및 불확실성 정량화, (2) 임베디드, 엣지 환경을 고려한 경량화와 연산 예산 친화적 설계, (3) 암호화 트래픽과 개인정보 보호 상황에서도 유효한 특징 재표현과 도메인 일반화 기법, (4) 데이터셋 편향과 분포 이동에 강건한 특징 학습, (5) 공격자 모델을 반영한 적대적 학습과 오탐/미탐 비용 기반 목적함수 설계가 요구된다. 이러한 방향은 CNN 기반 NIDS의 탐지 성능, 일반화, 안정성을 동시에 끌어올리는 실질적 경로가 될 것이다.

사사문구

이 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. RS-2024-00337414, SW공급망 운영환경에서 역공학 한계를 넘어서는 자동화된 마이크로 보안 패치 기술 개발)과 한국산업기술기획평가원의 지원(No. RS-2024-00406121, 자동차보안취약점기반위협분석시스템개발(R&D))을 받아 수행된 연구 결과임.

참고문헌

[1] Zhendong Wu et al, “A network intrusion detection method based on semantic Re-encoding and deep learning,” Journal of Network and Computer Applications, vol. 168, pp. 102688, 2020.

[2] Minh Tuan Nguyen et al, “Genetic Convolutional Neural Network for Intrusion Detection Systems,” Future Generation Computer Systems, vol. 110, pp. 418-427, 2020.

[3] Ayantayo et al, “Network intrusion detection using feature fusion with deep learning,” Journal of Big Data, vol. 10, no. 1, pp. 167, 2023.