https://doi.org/10.7840/kics.2024.49.7.911

불균형 네트워크 데이터의 이상유형 분류를 위한 미세 조정 연구

조 무 곤◆, 김 미 르*, 권 민 혜°

Fine-Tuning Anomaly Classifier for Unbalanced Network Data

Mugon Joe◆, Miru Kim*, Minhae Kwon°

요 약

최근 네트워크 침입 시도가 증가함에 따라, 신속하고 적절한 대응의 중요성이 강조되고 있다. 각각의 침입 방식에 따라 대응 방안이 다르기 때문에, 효과적인 대응을 위해서는 네트워크 이상유형을 정확하게 파악해야 한다. 이에 오토인코더 기반 이상탐지 기술에 분류 모델을 추가적으로 활용하여 네트워크 이상유형을 분류하는 연구가 주목받고 있다. 그러나 네트워크 데이터는 수집이 어려운 비정상 데이터에 대해 불균형 문제를 가지고 있으며, 오토인코더 기반 이상탐지로부터 탐지된 데이터와 분류 모델의 학습 데이터 간의 이상유형 분포 차이로 인해 성능의 한계를 보인다. 이러한 한계를 극복하기 위해 본 논문에서는 분류 모델에 미세 조정 기법을 적용하는 방안을 제안한다. 모의 실험 결과, 제안하는 시스템이 기존 연구 결과 대비 모든 평가 항목에서 우수한 성능을 보였다.

Key Words : Anomaly Detection, Anomaly Classifier, Autoencoder, Fine-tuning

ABSTRACT

With the recent surge in network intrusion attempts, there is a growing emphasis on prompt and accurate responses. Given that each intrusion type necessitates a distinct response approach, accurately identifying network anomalies is crucial for an effective response. Research on classifying network anomalies using classification models with autoencoder-based anomaly detection has garnered significant attention. However, network data presents an unbalance problem for abnormal data, which is challenging to collect, leading to limited performance. This limitation arises from the disparity in distribution between the detected data of autoencoder-based anomaly detection and the training data of classification models. To address this issue, this study proposes a solution that employs fine-tuning techniques with the classification model. Simulation results demonstrate that the proposed system surpasses previous research results across all evaluation metrics.

Ⅰ. 서 론

네트워크 기술의 발전에 따라 온라인에서 공유되는 중요한 개인 정보의 양이 급증하면서, 네트워크 상에서 통신 되는 정보를 대상으로 한 네트워크 침입 시도가 빠르게 증가하고 있다[1,2]. 정확한 침입 유형을 파악하

※ 본 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단(RS-2023-00278812), 그리고 정보통신기획평가원(IITP-2021-0-00739)의 지원을 받아 수행된 연구임.
◆ First Author : Soongsil University School Department of Intelligent Semiconductors, whanrhs115@soongsil.ac.kr, 학생회원
° Corresponding Author : Soongsil University School of Electronic Engineering, minhae@ssu.ac.kr, 종신회원
* Soongsil University School Department of Intelligent Semiconductors, mirukim00@soongsil.ac.kr, 학생회원
논문번호 : 202402-037-A-RN, Received February 25, 2024; Revised April 3, 2024; Accepted April 9, 2024

는 것은 신속하고 적절한 대응을 하기 위해 중요하기 때문에, 이를 위한 네트워크 이상유형 분류에 관한 연구가 크게 주목받고 있다. 특히 오토인코더(autoencoder; AE) 기반 이상탐지 기술을 활용하여 네트워크 이상유형을 분류하는 연구가 활발히 이루어지고 있다[3-8].

AE 모델은 입력 데이터의 축소와 복원을 수행하는 생성형 모델로, 이를 활용한 이상탐지는 입력 데이터와 복원 데이터 간의 오차를 바탕으로 비정상 데이터를 탐지한다[3-5]. 구체적으로 해당 방식은 AE 모델을 정상 데이터로 학습한 후, 학습에 경험하지 않은 비정상 데이터의 높은 복원 오차를 통해 이상탐지를 수행한다. 하지만 기존 이상탐지 기술은 입력층과 출력층의 정보만 활용하여 탐지 성능의 한계를 보이기에[6-8], 이를 보완하기 위한 방안으로 계층적 오토인코더(hierarchical autoencoder; HAE) 기반 이상탐지 방식이 제안되었다[6-8]. 해당 방식에서는 은닉층의 정보를 활용하여 다수의 탐지 구간을 통해 보다 정밀한 이상탐지를 진행한다.

HAE 기반 이상탐지를 비롯한 AE 계열의 이상탐지 기술은 정상 및 비정상에 대한 판단만 수행하므로, 탐지된 데이터의 구체적인 이상유형 분류를 위해서는 별도의 분류 모델이 필요하다[9,10]. 하지만 별도의 분류 모델을 구성하더라도 탐지된 데이터의 정확한 이상유형 분류에는 다음과 같은 문제점을 가진다.

우선, 네트워크 데이터는 일반적으로 불균형 데이터로 구성되어 있다[11-13]. 이는 이상유형 간의 샘플 수 차이에서 비롯되는 문제로, 대부분의 비정상 데이터는 정상 데이터에 비해 양이 적고 수집이 어렵다. 이와 같은 불균형 상황에서 데이터 샘플이 부족한 이상유형의 경우 탐지 및 분류 성능에 한계를 가진다. 또한, HAE 기반 이상탐지를 통해 탐지된 데이터와 전체 네트워크 데이터는 이상유형 분포가 일치하지 않는다[8]. 즉, 분류 모델이 실제로 분류해야 하는 데이터와 학습 데이터 간의 이상유형 분포가 서로 상이하다. 이러한 상황에서 단순히 전체 데이터를 사용하여 분류 모델[14-19]을 학습하게 되면, 학습 데이터와 탐지된 데이터 간의 분포 차이로 인해 성능 저하가 발생할 수 있다.

이에 본 논문에서는 미세 조정 기법[20-22]을 통해 위 두 가지 문제 상황에서도 높은 성능을 내는 이상유형 분류 시스템을 제안한다. 제안하는 시스템은 HAE 모델을 통해 네트워크 데이터의 이상탐지를 수행한 후, 탐지된 데이터의 이상유형을 분류하는 구조이다. 구체적으로 이상탐지 단계에서는 다수의 탐지 구간으로 비정상 데이터를 계층적으로 탐지하도록 하며, 이상유형 분류 단계에서는 각 구간별로 분류 모델을 배치하여 탐지된 데이터의 이상유형을 분류하도록 한다. 이때 각 구간별 분류 모델은 미세 조정을 통해 탐지된 데이터의 분포에 최적화되도록 한다. 즉, 각 탐지 구간별 분류 모델이 최대의 성능을 내도록 시스템을 설계한다.

본 논문은 다음과 같이 구성되어 있다. Ⅱ장에서는 본 연구와 관련된 선행 연구에 대하여 살펴본다. Ⅲ장에서는 제안하는 이상유형 분류 시스템에 대하여 서술한다. 이후 Ⅳ장에서는 제안하는 시스템의 성능을 평가하고 분석한다. Ⅴ장에서는 본 연구에 대한 결론을 맺는다.

Ⅱ. 선행 연구

2.1 오토인코더 모델을 활용한 이상탐지 연구

AE 모델은 학습에서 경험한 데이터를 낮은 복원 오차로 복원하는 특성을 가진다. 따라서 정상 데이터를 이용해 AE 모델을 학습하면, 학습에 경험했던 정상 데이터와 유사한 특징을 가지는 데이터는 낮은 복원 오차를 가진다. 반대로 정상 데이터와 상이한 특징을 가지는 데이터는 높은 복원 오차를 가진다. 따라서 AE 기반 이상탐지는 이러한 복원 오차 특성을 바탕으로 이상탐지를 수행한다[3-5]. 하지만 기존의 AE 기반 이상탐지 연구는 AE 모델의 은닉층 정보를 활용하지 않고 입력층과 출력층의 정보만 사용한다. 이러한 제한적인 정보의 활용은 탐지 성능의 한계를 가져올 수 있다. 따라서 더욱 정밀하고 높은 성능의 이상탐지를 위해서는 은닉층 정보를 고려해야 한다[6-8].

이러한 한계를 극복하기 위해 AE 모델의 은닉층 정보를 활용하여 보다 정밀한 탐지를 수행하는 HAE 기반 이상탐지 방식이 제안되었다[6,7]. HAE 기반 이상탐지는 복원 데이터를 모델에 재입력하여, 입력 데이터와 복원 데이터에 대한 각 은닉층 출력을 비교하여 이상여부를 계층적으로 탐지한다. 또한 HAE 기반 이상탐지 방식은 이상정도에 따라 각 구간마다 이상유형을 분담하여 비정상 데이터를 탐지하기 때문에, 구간별로 탐지되는 데이터의 이상유형의 분포가 서로 상이하다는 결과를 확인한 연구가 진행되었다[8]. 이에 본 논문에서는 HAE 모델의 각 탐지 구간에 분류 모델을 구성하여, 구간별로 분포가 상이한 데이터의 이상유형을 분류하는 시스템을 제안한다.

2.2 이상유형 분류를 고려하는 연구

네트워크 공격에는 다양한 이상유형들이 존재하며, 각 이상유형마다 대응 방식이 상이하기 때문에 정확한 이상유형 분류가 필요하다[9,10]. 이러한 네트워크 이상유형을 분류하기 위해 서포트 벡터 머신[14]과 결정 트리[15] 등 다양한 분류 모델을 활용한 연구들이 진행되었

다. 특히 서포트 벡터 머신은 분류 문제에서 일반적으로 사용되는 방식으로, 대표적인 기계 학습 알고리즘이다[14]. 해당 방식은 이상유형 간의 최적의 결정 경계를 찾고, 이를 통해 새로운 데이터의 결정 위치에 따라 이상유형을 분류하는 방식이다. 그러나 네트워크 데이터는 방대한 양과 고차원의 데이터로 구성되어 있어, 단순한 방식의 기계 학습 알고리즘은 분류 성능의 한계를 가진다[16].

이러한 한계를 극복하기 위해 신경망 모델을 활용한 연구가 주목받고 있다[16-19]. 신경망 기법을 활용하여 네트워크 데이터에 최적화된 분류 모델을 설계한 연구에서는 일반적으로 크로스 엔트로피 손실함수를 사용하여, 모델의 예측 확률값과 실제값의 차이를 최소화하도록 모델을 학습하였다[16-18]. 이후 해당 모델이 방대한 양의 데이터 처리와 고차원 데이터를 비선형적으로 처리하는 데 용이함을 보였다[19]. 하지만 네트워크 데이터는 정상 데이터가 대부분을 차지하므로, 분류 대상이 되는 데이터가 정상 데이터에 편향된다[9,10]. 이로 인하여 모델이 비정상 데이터를 탐지 및 분류하는 데 민감하지 않게 된다.

이와 같은 문제를 해결하고자 정상 데이터를 기반으로 학습한 AE 모델을 통해 이상탐지를 진행하고, 전체 데이터를 기반으로 학습한 분류 모델을 통해 탐지된 데이터의 이상유형을 분류하는 다단계 이상유형 분류 방식이 제안되었다[9,10]. 해당 방식은 정상 데이터와 비정상 데이터를 분리함으로써, 분류 모델에 입력되는 데이터가 정상 데이터에 편향되는 문제를 완화하였다. 하지만 분류 모델의 학습에 사용한 전체 데이터 또한 불균형 문제를 가지고 있기에, 기존 연구[9,10]에서는 학습 데이터 샘플이 부족한 비정상 데이터의 분류에 어려움이 있다. 이에 본 논문에서는 이상탐지로부터 탐지된 데이터를 활용하여 학습 데이터가 부족한 이상유형도 효과적으로 분류할 수 있는 시스템을 고려한다.

2.3 미세 조정을 활용한 모델 최적화 연구

사전 학습된 모델의 파라미터는 학습 데이터의 분포에 맞춰 최적화되기 때문에 새로운 분포의 데이터를 입력하였을 때, 높은 성능을 기대하기 어렵다. 이에 따라 모델을 새로운 분포의 데이터에 최적화하기 위한 미세 조정에 관한 연구가 진행되고 있다[20-22].

대표적으로 사용되는 미세 조정 기술 중 부분 조정 방식[20-22]은 사전 학습된 모델의 파라미터 중 일부를 선택하여 미세 조정하는 방식이다. 이는 선택된 파라미터에 대해서만 새로운 분포의 데이터로 학습하여, 계산량을 줄이면서도 모델의 성능을 효과적으로 향상시킬 수 있다. 이외에도 모델의 마지막 은닉층에 새로운 파라미터를 추가하여 특정 문제를 다루기에 유용한 추가 방식[23,24]과, 불필요한 파라미터를 제거하거나 매우 작은 값으로 변경하여 모델 경량화에 사용되는 대체 방식[25,26]이 존재한다.

본 논문에서 고려하는 HAE 기반 이상탐지로부터 탐지된 데이터는 전체 네트워크 데이터와 이상유형 분포가 서로 상이하므로, 미세 조정 기술 중 부분 조정 방식을 활용하여 탐지 구간별 최적화된 분류 모델을 설계하도록 한다. 즉, HAE 기반 이상탐지에 최적화된 이상유형 분류 시스템을 제안한다.

Ⅲ. 제안하는 이상유형 분류 시스템

본 논문에서 제안하는 네트워크 이상유형 분류 시스템은 비정상 데이터를 탐지하는 이상탐지 단계와 탐지된 비정상 데이터를 분류하는 이상유형 분류 단계로 구성된다. 본 장에서는 HAE 기반 이상탐지를 바탕으로 제안하는 시스템의 이상탐지 단계에 대하여 서술한 후, 이상유형 분류 단계에 대하여 서술한다.

3.1 계층적 오토인코더 기반 이상탐지 단계

네트워크 이상탐지 대상 데이터는 실수 벡터 공간 $\mathbb{R}$에 대해 d개의 feature로 구성된 샘플 $\mathbf{x} \in \mathbb{R}^d$를 원소로 가지는 집합 $\mathcal{X}$로 정의하며, 네트워크 데이터 $\mathcal{X}$의 이상유형은 $\mathcal{Y}$로 정의한다. 자연수 집합 $\mathbb{N}$에 대해 이상유형 $\mathcal{Y}$의 원소 $y \in \mathbb{N}$는 네트워크 데이터의 이상유형 종류 J개에 대하여 $1 \le y \le J$를 만족한다. $\mathcal{Y}$는 이상유형 분포 Q를 따르며, 이는 $\mathcal{Y} \sim Q$와 같이 정의한다. 제안하는 시스템의 이상탐지 단계에서는 전체 데이터 $\mathcal{X}$에 대한 정상과 비정상에 대한 탐지만 진행하고, 이후 이상유형 분류 단계에서 다수의 비정상 유형에 대한 분류를 진행한다.

HAE 기반 이상탐지는 전체 K개의 탐지 구간으로 구성되어 있으며 $k(1 \le k \le K)$번째 탐지 구간은 d_k개의 출력 노드를 가지고 있다. HAE 모델은 $\mathcal{X}$에 대하여 k 번째 탐지 구간에 잠재 벡터 집합 $\mathcal{Z}_k$를 출력하며, $\mathcal{Z}_k$는 잠재 벡터 $\mathbf{z}_k \in \mathbb{R}^{d_k}$를 원소로 가진다. 각 탐지 구간에서 비정상으로 분류된 잠재 벡터 데이터는 집합 $\mathcal{Z}_k' \subseteq \mathcal{Z}_k$로 표현하고, 이상점수 함수 $\epsilon_k(\mathbf{z}_k)$[6]와 임계치 δ_k를 사용하여 수식 (1)과 같이 정의한다.

$$\mathcal{Z}_k' = \{\mathbf{z}_k | \epsilon_k(\mathbf{z}_k) \ge \delta_k, \mathbf{z}_k \in \mathcal{Z}_k\} \quad (1)$$

이상점수 함수 $\epsilon_k(\mathbf{z}_k)$는 k 번째 탐지 구간의 출력

z_k에 대한 이상점수를 연산하고, 해당 이상점수 $\epsilon_k(\boldsymbol{z}_k)$가 임계치 δ_k보다 크거나 같은 경우 비정상 데이터로 탐지한다. 이때 탐지된 데이터 $\mathcal{Z}_k'$의 이상유형은 $\mathcal{Y}_k' \subseteq \mathcal{Y}$로 나타낸다. $\mathcal{Y}_k'$는 분포 $\boldsymbol{Q}_k'$를 따르며, 이는 $\mathcal{Y}_k' \sim \boldsymbol{Q}_k'$와 같이 정의한다.

k 번째 탐지 구간에서 비정상 데이터로 판단되지 않은 데이터는 $k+1$번째 탐지 구간에서 이상탐지를 진행하며, 최종적으로 K 번째 탐지 구간에서 $\epsilon(\boldsymbol{z}_K) < \delta_K$를 만족하는 데이터는 정상 데이터로 판단한다. 앞서 기술한 이상탐지 단계의 구조는 그림 1에서 확인 가능하다.

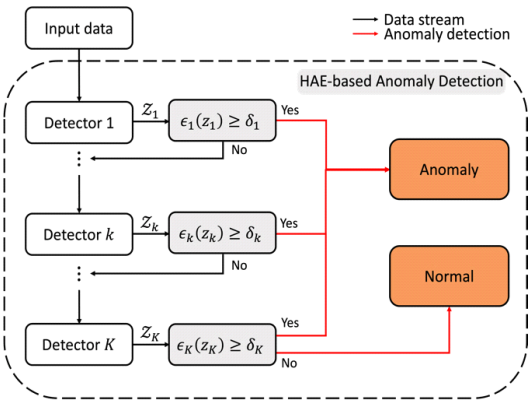


그림 1. HAE 기반 이상탐지 구조
Fig. 1. Structure of HAE-based anomaly detection

3.2 미세 조정 기반 이상유형 분류 단계

제안하는 시스템의 이상유형 분류 단계에서는 이상탐지 단계의 k 번째 탐지 구간에서 비정상으로 탐지된 데이터의 유형 분류를 진행한다. 이를 위해 각 탐지 구간마다 유형 분류 모델이 존재한다. 유형 분류 모델의 학습 과정은 잠재 벡터 집합 $\mathcal{Z}_k$를 통한 분류 모델의 사전 학습 과정과 탐지된 데이터 $\mathcal{Z}_k'$을 사용하여 사전 학습된 분류 모델을 탐지 구간별 분포에 맞춰 미세 조정하는 과정으로 구성된다. 제안하는 시스템의 전체적인 구조는 그림 2에서 확인 가능하다.

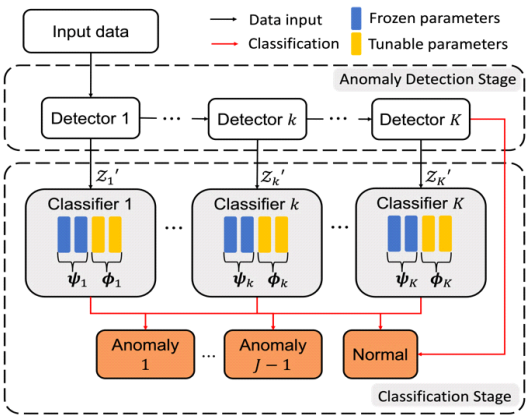


그림 2. 제안하는 시스템 구조
Fig. 2. Structure of proposed system

3.2.1 탐지 구간별 분류 모델 사전 학습

k 번째 탐지 구간의 분류 모델은 $\{\boldsymbol{\psi}_k, \boldsymbol{\phi}_k\}$으로 표현되며, $\boldsymbol{\psi}_k$는 고정 파라미터로 사전 학습에서만 업데이트 되며, $\boldsymbol{\phi}_k$는 조정 파라미터로 사전 학습과 미세 조정 모두에서 업데이트 된다. 이러한 분류 모델 $\{\boldsymbol{\psi}_k, \boldsymbol{\phi}_k\}$에 입력 데이터 $\mathcal{Z}_k$와 이상유형 $\mathcal{Y}$를 입력하였을 때, 크로스 엔트로피 손실함수 $\mathcal{L}(\mathcal{Z}_k, \mathcal{Y}; \{\boldsymbol{\psi}_k, \boldsymbol{\phi}_k\})$는 입력 데이터 $\mathcal{Z}_k$에 대한 예측 확률값와 이상유형 $\mathcal{Y}$에 대한 실제값의 차이를 나타낸다. 이에 사전 학습 과정의 목적함수는 수식 (2)와 같이 표현할 수 있다.

$$\{\boldsymbol{\psi}_k^+, \boldsymbol{\phi}_k^+\} = \underset{\{\boldsymbol{\psi}_k, \boldsymbol{\phi}_k\}}{\operatorname{argmin}} \mathbb{E}_{\mathcal{Y} \sim \boldsymbol{Q}} \mathcal{L}(\mathcal{Z}_k, \mathcal{Y}; \{\boldsymbol{\psi}_k, \boldsymbol{\phi}_k\}) \quad (2)$$

수식 (2)는 전체 데이터 $\mathcal{X}$에 대한 각 탐지 구간의 잠재 벡터 집합 $\mathcal{Z}_k$와 $\mathcal{Y} \sim \boldsymbol{Q}$을 사용하여, 손실 함수 $\mathcal{L}$이 최소가 되도록 하는 사전 학습 모델 $\{\boldsymbol{\psi}_k^+, \boldsymbol{\phi}_k^+\}$을 구하는 것을 목적으로 한다. 이는 각 탐지 구간의 분류 모델 $\{\boldsymbol{\psi}_k, \boldsymbol{\phi}_k\}$이 전체 이상유형 분포 $\mathcal{Y} \sim \boldsymbol{Q}$에 대하여 전반적인 특징을 학습하는 것이 목적임을 의미한다.

3.2.2 사전 학습된 분류 모델 미세 조정

각 이상탐지 구간에서 탐지된 데이터의 이상유형의 분포 $\boldsymbol{Q}_k'$는 전체 이상유형 분포 $\boldsymbol{Q}$와 상이하기 때문에, 고성능의 이상유형 분류를 위해 사전 학습된 이상유형 분류 모델 $\{\boldsymbol{\psi}_k^+, \boldsymbol{\phi}_k^+\}$을 분포 $\boldsymbol{Q}_k'$에 최적화되도록 미세 조정하는 과정이 필요하다. 사전 학습된 모델 $\{\boldsymbol{\psi}_k^+, \boldsymbol{\phi}_k^+\}$의 미세 조정 과정의 목적함수는 수식 (3)과 같이 표현할 수 있다.

$$\boldsymbol{\phi}_k^* = \underset{\boldsymbol{\phi}_k^+}{\operatorname{argmin}} \mathbb{E}_{\mathcal{Y}_k' \sim \boldsymbol{Q}_k'} \mathcal{L}(\mathcal{Z}_k', \mathcal{Y}_k'; \{\boldsymbol{\psi}_k^+, \boldsymbol{\phi}_k^+\}) \quad (3)$$

수식 (3)은 탐지된 잠재 벡터 집합 $\mathcal{Z}_k'$과 해당하는 이상유형 $\mathcal{Y}_k' \sim \boldsymbol{Q}_k'$을 사용하여 손실 함수 $\mathcal{L}$이 최소가 되도록 하는 조정 파라미터 $\boldsymbol{\phi}_k^*$를 구하는 것을 목적으로 한다. 이를 통해 각 탐지 구간별 분류 모델 $\{\boldsymbol{\psi}_k^+, \boldsymbol{\phi}_k^*\}$은 분포 $\boldsymbol{Q}_k'$에 최적화되어, 고성능의 분류를 수행한다.

결과적으로 제안하는 이상유형 분류 시스템은 이상탐지 단계와 이상유형 분류 단계로 구성된다. 이상탐지 단계에서는 HAE 기반 이상탐지를 통해 비정상 데이터를 탐지하고, 이상유형 분류 단계에서는 각 탐지 구간별 분류 모델 $\{\psi_k^+, \phi_k^*\}$로 탐지된 데이터의 이상유형을 분류한다.

Ⅳ. 이상유형 분류 성능 평가

제안된 이상유형 분류 시스템의 성능을 평가하기 위한 모의 실험을 진행한다. 이에 3가지의 이상유형 분류 방식과 성능을 비교하여 제안하는 시스템의 유효성과 일반화된 우수성을 평가한다.

4.1 실험 설정

본 절에서는 성능 평가를 진행하기 위한 실험 설정에 관하여 서술한다. 먼저, 실험에 사용할 네트워크 데이터 셋을 설명한 뒤, 성능 평가를 수행할 대상에 대하여 서술한다.

4.1.1 불균형 네트워크 데이터 셋

본 논문에서는 두가지 대표적인 불균형 네트워크 데이터 셋을 이용하여 실험을 진행한다. NSL-KDD 데이터 셋은 총 148,417개의 샘플과 정상 데이터를 포함한 5개의 이상유형으로 구성되어 있으며 각 이상유형 별 샘플 수는 표 1을 통해 확인할 수 있다[27]. 해당 표를 통해 NSL-KDD 데이터 셋은 전체 데이터 샘플 중 상위 2개의 이상유형은 88%, 하위 2개의 이상유형은 3%를 차지하고 있는 것을 확인할 수 있다.

CSE-CIC-IDS 2018 데이터 셋은 총 15,450,706개의 샘플과 15개의 이상유형으로 구성되어 있으며 각 이상유형 별 샘플 수는 표 2를 통해 확인할 수 있다[28]. 전체 데이터 샘플 중 상위 4개의 이상유형은 83%, 하위 4개의 이상유형은 1% 미만으로 구성된다.

본 논문에서는 이와 같은 불균형 네트워크 데이터 셋을 이용한 실험을 진행하였으며, 일반화된 결과를 위

표 1. 데이터 샘플 수 (NSL-KDD)
Table 1. Number of data sample (NSL-KDD)

Category	Class	Train data	Test data
Normal	1	67,343	9,711
Dos	2	45,927	7,458
Probe	3	11,656	2,421
R2L	4	995	2,754
U2R	5	52	200

표 2. 데이터 샘플 수 (CSE-CIC-IDS 2018)
Table 2. Number of data sample (CSE-CIC-IDS 2018)

Category	Class	Train data	Test data
Normal	1	10,159,055	1,269,882
DDoS-HOIC	2	548,810	68,601
DDoS-LOIC-HTTP	3	460,953	57,619
DoS-Hulk	4	373,331	46,666
Bot	5	228,953	28,619
Bruteforce-Web	6	154,688	19,336
Bruteforce-XSS	7	150,071	18,759
Infilteration	8	129,547	16,193
DoS-Goldeneye	9	33,206	4,151
DoS-SlowHTTP	10	11,112	1,389
DoS-Slowloris	11	8,729	1,099
DDoS-LOIC-UDP	12	1,384	173
FTP-Bruteforce	13	489	61
Bruteforce	14	184	23
SQL-Injection	15	70	9

해 5개의 랜덤시드에 대한 평균과 표준편차를 기준으로 이상유형 분류 모델의 성능 평가를 진행하였다. 이를 바탕으로 제안된 시스템의 최저 성능과 비교 방식의 최고 성능을 비교하여, 제안된 시스템의 일반화된 우수성을 확인하였다.

4.1.2 성능 평가 대상

본 논문에서 제안하는 이상유형 분류 시스템의 우수성을 입증하기 위한 성능 비교의 대상은 다음과 같다.

- **제안하는 시스템**: 제안하는 시스템의 학습 과정은 이상탐지 단계의 HAE 모델 학습과 이상유형 분류 단계의 분류 모델 사전 학습 및 미세 조정으로 구성된다. 이상탐지 단계의 HAE 모델 학습은 학습 데이터 내 정상 데이터만을 사용하여 진행한다. 이상유형 분류 단계의 분류 모델 사전 학습은 학습 데이터 내 모든 정상 및 비정상 데이터를 사용하여 학습한다. 이후 미세 조정 과정은 학습이 완료된 HAE 모델에 학습 데이터 내 모든 정상 및 비정상 데이터를 통과시켜 각 탐지 구간별 비정상 데이터로 분류된 데이터를 사용하여 학습한다.
- **다단계 탐지(Multi-stage)[9]**: 제안하는 시스템과 동일하게 이상탐지 단계와 이상유형 분류 단계로 구성되는 다단계 탐지 방식으로, HAE 모델 학습과 분류 모델 사전 학습 과정을 거친다. 해당 방식은 제안하는 시스템과 동일한 구조이나 분류 모델의 미세 조정 과정을 거치지 않는다. 본 비교 모델을 통해 미세

조정 과정의 유효성을 확인한다.

- **심층 신경망(DNN)**[19]: 신경망 방식의 분류 모델을 학습 데이터 내 모든 정상 및 비정상 데이터로 학습한다. 해당 방식은 제안하는 시스템의 이상유형 분류 모델과 같은 구조를 가지지만 이상탐지 단계를 적용하지 않는다는 차이점을 가진다. 이를 통해 제안하는 시스템에서 사용된 이상탐지 단계의 유효성을 확인한다.
- **서포트 벡터 머신(SVM)**[14]: 분류 문제에서 일반적으로 사용되는 서포트 벡터 머신 방식이다. 해당 방식은 제안하는 시스템에서 사용된 신경망 방식과 상이한 기계 학습 방식으로, 단순한 알고리즘으로 인하여 분류 성능에 한계를 가진다. 이를 통해 제안하는 시스템에서 사용된 신경망 기반 분류 모델의 우수성을 확인한다.

제안된 시스템과 다단계 탐지에서 사용한 HAE 모델은 각각 2개의 은닉층을 가진 인코더와 디코더로 구성되며, 총 4개의 탐지 구간을 가진다. 제안된 시스템과 다단계 탐지, 심층 신경망에서 사용한 탐지 구간별 분류 모델은 2개의 은닉층을 가지며, 은닉층의 노드 수는 입력층의 노드 수와 동일하게 설정한다. 본 실험에서 사용한 성능 평가 대상들의 자세한 하이퍼파라미터 설정은 표 Appendix A.2와 표 Appendix A.3을 통해 확인할 수 있다.

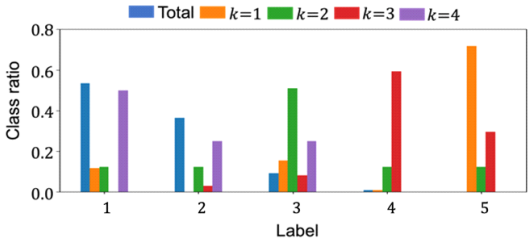


그림 3. k 번째 탐지 구간의 이상유형 분포
Fig. 3. Class distribution of k-th detection stage

4.2 탐지 구간별 이상유형 분포 분석

제안하는 시스템의 분류 모델에는 전체 데이터를 사용한 사전 학습과 탐지된 데이터를 활용한 미세 조정이 적용된다. 이에 본 논문에서는 NSL-KDD와 CSE-CIC-IDS 2018 데이터 셋을 사용한 모의 실험을 통해, 각 탐지 구간의 이상유형 분포를 확인한다. 이를 통해 제안하는 시스템의 사전 학습과 미세 조정에 사용되는 데이터가 서로 상이함을 실증한다. 즉, 전체 데이터와 탐지된 데이터의 이상유형 분포가 일치하지 않음

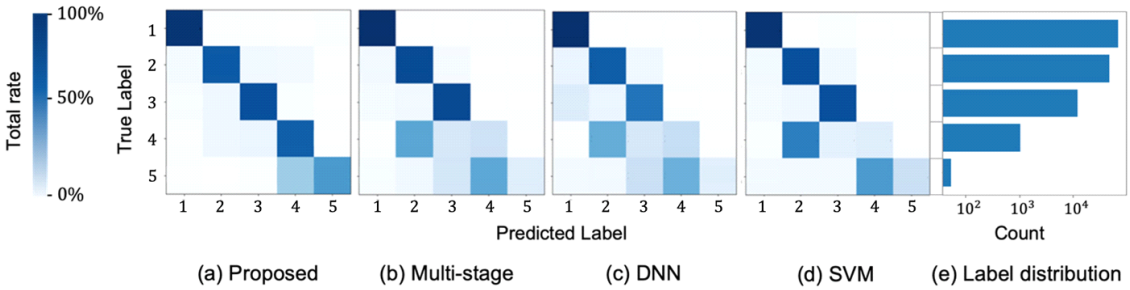


그림 4. 성능 결과 분석 (NSL-KDD 데이터 셋)
Fig. 4. Performance analysis (NSL-KDD dataset)

표 3. 성능 비교 결과 (NSL-KDD)
Table 3. Performance comparisons (NSL-KDD)

NSL-KDD dataset[27]: mean (±1 standard deviation)							
	Accuracy	Recall	Precision	F1-score	Fall-out	Specificity	MCC
Proposed	**0.9036** (±0.0161)	**0.9062** (±0.0123)	**0.9085** (±0.0176)	**0.9039** (±0.0072)	**0.0396** (±0.0098)	**0.9633** (±0.0088)	**0.8505** (±0.0259)
Multi-stage [9]	0.8338 (±0.0044)	0.8432 (±0.0035)	0.8329 (±0.0108)	0.8398 (±0.0053)	0.0439 (±0.0024)	0.9338 (±0.0023)	0.7644 (±0.0072)
DNN [19]	0.8595 (±0.0053)	0.8543 (±0.0074)	0.8451 (±0.0129)	0.8483 (±0.0093)	0.0599 (±0.0014)	0.9401 (±0.0049)	0.7834 (±0.0066)
SVM [14]	0.8214 (±0.0032)	0.8297 (±0.0044)	0.8508 (±0.0071)	0.8096 (±0.0027)	0.0714 (±0.0047)	0.9286 (±0.0046)	0.7407 (±0.0053)

을 보인다.

그림 3은 NSL-KDD 데이터 셋에 대해 전체 데이터와 구간별로 탐지된 데이터의 이상유형 분포를 나타낸 그래프이다. 해당 그래프에서 가로축은 이상유형을 의미하며, 세로축은 각 이상유형의 비율을 의미한다. 먼저 그림 3에서 Total로 표현한 전체 데이터의 경우 이상유형 1, 2가 대부분의 비율을 차지하는 것을 확인할 수 있다. 반면 탐지된 데이터의 경우 전체 데이터의 이상유형 분포와 상이한 결과를 보였다. 특히 k=1 구간에서는 이상유형 5, k=2 구간에서는 이상유형 3, k=3 구간에서는 이상유형 4, 5, k=4 구간에서는 이상유형 1, 2, 3이 주로 탐지되었다. 주 탐지 이상유형은 데이터 불균형 상황을 효과적으로 다룰 수 있음을 의미하며, 해당 분포 차이는 전체 데이터를 HAE 모델을 통해 계층적으로 탐지하였기 때문이다. 이와 같은 주 탐지 이상유형과, 전체 데이터와 탐지된 데이터 간의 이상유형 분포 차이는 미세 조정의 필요성을 입증한다. CSE-CIC-IDS 2018 데이터 셋에 대한 결과는 그림 Appendix A.1에서 확인 가능하다.

4.3 비교 성능 평가

표 3과 그림 4는 각각 NSL-KDD 데이터 셋에 대한 제안하는 시스템, 다단계 탐지, 서포트 벡터 머신, 그리고 심층 신경망 모델의 이상유형 분류 실험 결과이다.

표 3에서 확인할 수 있듯이, 제안하는 시스템은 Accuracy, Recall, 그리고 Precision의 세 가지 항목에서 다단계 탐지 대비 약 8%의 성능 향상을 보였다. 이 결과는 미세 조정 과정이 성능 향상에 큰 역할을 하는 것을 의미한다. 또한, 제안하는 시스템은 심층 신경망과 서포트 벡터 머신 대비 약 7.5% 향상된 Accuracy를 보였다. 이는 제안하는 시스템의 이상탐지 단계와 미세 조정 과정의 유효성을 입증한다. 추가적으로 각 비교 방식의 표준편차를 고려하였을 때, 제안하는 시스템의 최저 성능이 비교 대상 중 가장 높은 성능을 보인 심층 신경망의 최고 성능 대비 약 3% 성능 향상을 보였다. 이는 제안하는 시스템의 일반화된 우수성을 나타낸다.

제안하는 시스템의 성능 향상에 대한 원인은 그림 4를 통해 분석할 수 있다. 그림 4의 (a), (b), (c), 그리고 (d)는 각각의 이상유형 분류 방식의 예측 결과를 나타내는 그래프로 가로축은 모델이 예측한 레이블, 세로축은 실제 레이블을 의미한다. 따라서 예측 레이블과 실제 레이블이 동일한, 대각선의 색상이 진할수록 모델의 예측 성능이 높음을 의미한다. 그림 4의 (e)는 각 이상유형의 샘플 수를 표현한 그래프로, 가로축은 각 이상유형

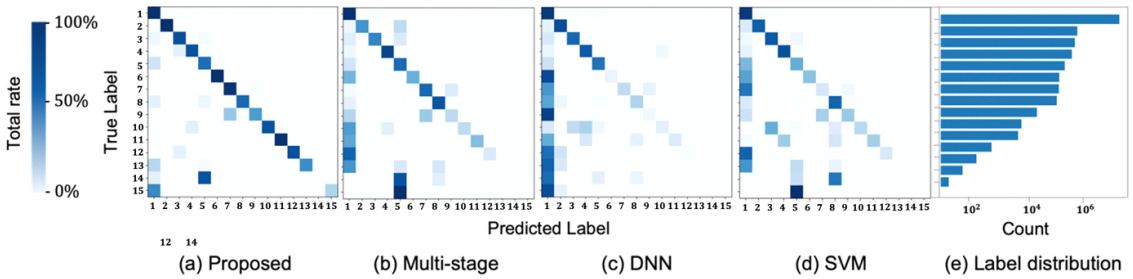


그림 5. 성능 결과 분석 (CSE-CIC-IDS 2018 데이터 셋)
Fig. 5. Performance analysis (CSE-CIC-IDS 2018 dataset)

표 4. 성능 비교 결과 (CSE-CIC-IDS 2018)
Table 4. Performance comparisons (CSE-CIC-IDS 2018)

CSE-CIC-IDS 2018 dataset[28]: mean (±1 standard deviation)							
	Accuracy	Recall	Precision	F1-score	Fall-out	Specificity	MCC
Proposed	**0.9500** (±0.0032)	**0.9342** (±0.0078)	**0.9442** (±0.0055)	**0.9392** (±0.0021)	**0.0222** (±0.0093)	**0.9778** (±0.0064)	**0.9293** (±0.0019)
Muti-stage [9]	0.8547 (±0.0045)	0.8348 (±0.0071)	0.8333 (±0.0023)	0.8341 (±0.0092)	0.1667 (±0.0056)	0.7139 (±0.0037)	0.8233 (±0.0088)
DNN [19]	0.8272 (±0.0024)	0.8236 (±0.0029)	0.8182 (±0.0103)	0.8209 (±0.0049)	0.0196 (±0.0028)	0.7804 (±0.0028)	0.7809 (±0.0068)
SVM [14]	0.8405 (±0.0037)	0.8195 (±0.0068)	0.8212 (±0.0035)	0.8204 (±0.0078)	0.1805 (±0.0058)	0.6995 (±0.0035)	0.8195 (±0.0074)

에 해당하는 데이터의 샘플 수, 세로축은 이상유형을 나타낸다.

그림 4를 통해서 데이터의 양이 충분한 이상유형 1, 2, 3에 대하여 (a), (b), (c), (d) 모두 좋은 성능을 내는 것을 확인 가능하다. 반면 데이터 샘플 수가 충분하지 않은 이상유형 4, 5에 대하여 제안하는 시스템 (a)만이 다른 비교 방식 (b), (c), (d) 대비 높은 성능을 보인다. 이는 제안하는 방식이 미세 조정을 통해 샘플 수가 적은 이상유형에 대한 성능을 개선하여 전체적인 성능 향상을 이끌었음을 의미한다. 해당 결과를 통해 제안하는 방식이 데이터 불균형으로 인한 성능 저하 문제를 성공적으로 해결 할 수 있는 방안임을 입증할 수 있다.

표 4와 그림 5는 각각 CSE-CIC-IDS 2018 데이터 셋에 대한 제안하는 시스템, 다단계 탐지, 서포트 벡터 머신, 그리고 심층 신경망 모델의 이상유형 분류 실험 결과이다.

표 4에서 제안하는 시스템은 비교 대상 중 가장 높은 성능을 보인 다단계 탐지 대비 Accuracy, Recall, 그리고 Precision 항목이 약 12% 성능 향상되었다. 마찬가지로 제안하는 시스템과 다단계 탐지의 표준편차를 고려하였을 때, 제안하는 시스템의 최저 성능이 다단계 탐지의 최고 성능 대비 약 11% 성능 향상을 보였다. 또한 심층 신경망과 서포트 벡터 머신 대비 제안하는 시스템의 Accuracy가 약 14% 향상되었다. 이와 같이 실험에 적용한 두 데이터 셋 모두에서 동일한 경향성의 성능 향상 결과를 통해 제안하는 방식의 일반화된 우수성을 확인할 수 있다.

CSE-CIC-IDS 2018 데이터 셋의 실험 결과 또한 NSL-KDD 데이터 셋의 실험 결과와 동일하게 그림 5를 통해 성능 향상의 원인 분석이 가능하다. 그림 5의 (a)에 해당하는 제안하는 시스템은 (b), (c), (d)에 해당하는 비교 방식이 제대로 분류하지 못하는 샘플 수가 적은 이상유형에 대하여 올바른 분류를 수행하여 전체적인 성능이 개선되었음을 확인할 수 있다. 이를 통해 제안하는 방식의 미세 조정 과정이 데이터 불균형에 따른 성능 저하를 개선할 수 있음을 확인할 수 있다.

V. 결 론

본 논문에서는 네트워크 데이터의 불균형 문제와, HAE 기반 이상탐지로부터 탐지된 데이터와 분류 모델의 학습 데이터 간의 이상유형 분포 차이를 해결하기 위해, 미세 조정을 활용한 이상유형 분류 시스템을 제안하였다. 제안하는 시스템은 이상탐지 단계의 각 탐지 구간마다 다수의 유형 분류 모델을 배치하여 네트워크 이상유형 분류를 수행한다. 분류 모델은 전체 데이터의 전반적인 특성을 사전 학습하며, 이후 분류 모델을 탐지 구간별 이상유형 분포에 최적화되도록 미세 조정한다. 네트워크 데이터를 사용한 모의 실험 결과, 제안하는 방식을 적용하였을 때 기존의 비교 모델 대비 분류 성능이 향상된 것을 확인하였다. 이러한 결과를 통해 미세 조정 과정이 네트워크 데이터의 불균형 문제 해결에 유의미한 기여를 하였음을 실험적으로 입증하였다.

References

[1] S. Lan, Y. Ma, W. Huang, W. Wang, H. Yang, and P. Li, "DSTAGNN: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting," in *ICML*, Baltimore, USA, Jul. 2022.

[2] S. Han, X. Hu, H. Huang, M. Jiang, and Y. Zhao, "ADBench: Anomaly detection benchmark," in *NeurIPS*, New Orleans, USA, Nov. 2022.
(https://doi.org/10.1109/TGRS.2022.3207165)

[3] S. Tu, M. Waqas, A. Badshah, M. Yin, and G. Abbas, "Network intrusion detection system (NIDS) based on pseudo-siamese stacked autoencoders in fog computing," *IEEE Trans. Serv. Comput.*, vol. 16, no. 6, pp. 4317-4327, 2023.
(https://doi.org/10.1109/TSC.2023.3319953)

[4] D. Aguilar, M. Pérez, O. González, K. Choo, and E. Susarrey, "Towards an interpretable autoencoder: A decision-tree-based autoencoder and its application in anomaly detection," *IEEE Trans. Dependable Secur. Comput.*, vol. 20, no. 2, pp. 1048-1059, 2022.
(https://doi.org/10.1109/TDSC.2022.3148331)

[5] Y. Moon, M. Kwon, B. Lee, and J. Noh, "Performance enhancement of autoencoder-based audio data anomaly detection via weakly supervised learning," *J. KICS*, vol. 48, no. 3, pp. 382-390, 2023.
(https://doi.org/10.7840/kics.2023.48.3.382)

[6] M. Kim, H. Kye, and M. Kwon, "Network anomaly detection system using hidden layer information of autoencoder," *J. KICS*, vol. 47, no. 9, pp. 1310-1321, 2022.

(https://doi.org/10.7840/kics.2022.47.9.1310)

[7] H. Kye, M. Kim, and M. Kwon, “Hierarchical autoencoder for network intrusion detection,” in *IEEE ICC*, Seoul, Korea, May 2022.

[8] M. Kim, H. Kye, and M. Kwon, “A study on detection role of hierarchical stage in autoencoder based network intrusion detection systems,” in *KICS Winter Conf.*, Pyeongchang, Korea, Feb. 2022.

[9] M. Verkerken, L. D’hooge, D. Sudyana, Y. Lin, T. Wauters, B. Volckaert, and F. Turck, “A novel multi-stage approach for hierarchical intrusion detection,” *IEEE Trans. Netw. Serv. Manag.*, vol. 20, no. 3, pp. 3915-3929, 2023. (https://doi.org/10.1109/TNSM.2023.3259474)

[10] Z. Tauscher, Y. Jiang, K. Zhang, J. Wang, and H. Song, “Learning to detect: A data-driven approach for network intrusion detection,” in *IEEE IPCCC*, Austin, USA, Jan. 2021.

[11] J. Lee, S. Shin, S. Yoon, and T. Kim, “Survey on artificial intelligence & machine learning models and datasets for network intelligence,” *J. KICS*, vol. 47, no. 4, pp. 625-643, 2022. (https://doi.org/10.7840/kics.2022.47.4.625)

[12] Y. Choe and K. Oh, “A study on the introduction of CTGAN oversampling algorithm to improve imbalance problem in intrusion detection data,” *J. KICS*, vol. 45, no. 12, pp. 2114-2122, 2020. (https://doi.org/10.7840/kics.2020.45.12.2114)

[13] A. Jumabek, S. Yang, and Y. Noh, “CatBoost-based network intrusion detection on imbalanced CIC-IDS-2018 dataset,” *J. KICS*, vol. 46, no. 12, pp. 2191-2197, 2021. (https://doi.org/10.7840/kics.2021.46.12.2191)

[14] M. Shen, K. Ye, X. Liu, L. Zhu, J. Kang, S. Yu, Q. Li, and K. Xu, “Machine learning-powered encrypted network traffic analysis: a comprehensive survey,” *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 791-824, 2022. (https://doi.org/10.1109/COMST.2022.3208196)

[15] O, Almomani, M. Almaiah, A. Alsaaidah, S. Smadi, A. Mohammad, and A. Althun, “Machine learning classifiers for network intrusion detection system: Comparative study,” in *IEEE ICIT*, Amman, Jordan, Jul. 2021.

[16] B. Borisenko, S. Erokhin, and I. Martishin, “Analysis of machine learning methods for detecting computer attacks,” in *IEEE SYNCHROINFO*, Pskov, Russian Federation, Jul. 2023.

[17] A. Kamali, K. Chougdali, and K. Abdellatif, “A new intrusion detection system based on convolutional neural network,” in *IEEE ICC*, Rome, Italy, Oct. 2023.

[18] M. Masum and H. Shahriar, “Tl-NID: Deep neural network with transfer learning for network intrusion detection,” in *IEEE ICITST*, London, UK, Dec. 2020.

[19] A. Sadeghzadeh, S. Shiravi, and R. Jalili, “Adversarial network traffic: Towards evaluating the robustness of deep-learning-based network traffic classification,” *IEEE Trans. Netw. Serv. Manag.*, vol. 18, no. 2, pp. 1962-1976, 2021. (https://doi.org/10.1109/TNSM.2021.3052888)

[20] J. Liang, C. Zhao, M. Wang, X. Qiu, and L. Li, “Finding sparse structures for domain specific neural machine translation,” in *AAAI*, Vancouver, Canada, Feb. 2021.

[21] G. Yuan, Y. Li, S. Li, Z. Kong, S. Tulyakov, X. Tang, Y. Wang, and J. Ren, “Layer freezing & data sieving: Missing pieces of a generic framework for sparse training,” in *NeurIPS*, New Orleans, USA, Nov. 2022.

[22] Y. Sung, J. Cho, and M. Bansal, “VL-adapter: Parameter-efficient transfer learning for vision-and-language tasks,” in *IEEE CVPR*, New Orleans, USA, Jun. 2022.

[23] A. Stickland and I. Murray, “BERT and PALs: Projected attention layers for efficient adaptation in multi-task learning,” in *ICML*, Long Beach, USA, Jun. 2019.

[24] R. Mahabadi, J. Henderson, and S. Ruder, “Compacter: Efficient low-rank hypercomplex adapter layers,” in *NeurIPS*, Virtual, Dec. 2021.

[25] A. Potapczynski, G. Loaiza-Ganem, and J.

Cunningham, "Invertible gaussian reparameterization: Revisiting the gumbel-softmax," in *NeurIPS*, Virtual, Dec. 2020.

[26] S. Ravi, A. Venkatesh, G. Fung, and V. Singh, "Optimizing nondecomposable data dependent regularizers via lagrangian reparameterization offers significant performance and efficiency gains," in *AAAI*, New York, USA, Feb. 2020.

[27] G. Meena and R. Choudhary, "A review paper on IDS classification using KDD 99 and NSL KDD dataset in WEKA," in *IEEE Comptelix*, Jaipur, India, Jul. 2017.

[28] L. Liu, G. Engelen, T. Lynar, D. Essam, and W. Joosen, "Error prevalence in NIDS datasets: A case study on CIC-IDS-2017 and CSE-CIC-IDS-2018," in *IEEE CNS*, Austin, USA, Oct. 2022.

조 무 곤 (Mugon Joe)

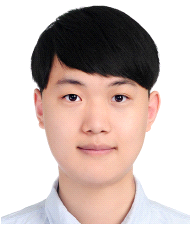
2024년 2월 : 숭실대학교 전자정보공학부 IT융합전공 학사

2024년 3월~현재 : 숭실대학교 지능형반도체학과 석사과정

<관심분야> 모바일네트워크, 이상탐지 기술, 인공지능

[ORCID:0009-0003-4111-491X]

김 미 르 (Miru Kim)



2019년 2월~2022년 8월 : 숭실대학교 전자정보공학부 IT융합전공 학사

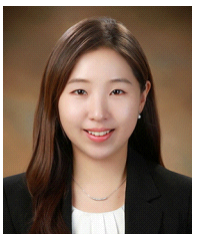
2022년 9월~2024년 2월 : 숭실대학교 지능형반도체학과 석사

2024년 3월~현재 : 숭실대학교 지능형반도체학과 박사과정

<관심분야> 이상탐지 기술, 인공지능, 연합학습

[ORCID:0000-0002-5394-4780]

권 민 혜 (Minhae Kwon)



2011년 8월 : 이화여자대학교 전자통신공학과 학사

2013년 8월 : 이화여자대학교 전자통신공학과 석사

2017년 8월 : 이화여자대학교 전자전기공학과 박사

2017년 9월~2018년 8월 : 이화여자대학교 전자전기공학과 박사 후 연구원

2018년 9월~202년 2월 : 미국 Rice University, Electrical and Computer Engineering, Postdoctoral Researcher

2020년 3월~현재 : 숭실대학교 전자정보공학부 IT융합전공 조교수

<관심분야> 모바일네트워크, 이상탐지기술, 인공지능, 강화학습, 자율주행

[ORCID:0000-0002-8807-3719]

Appendix

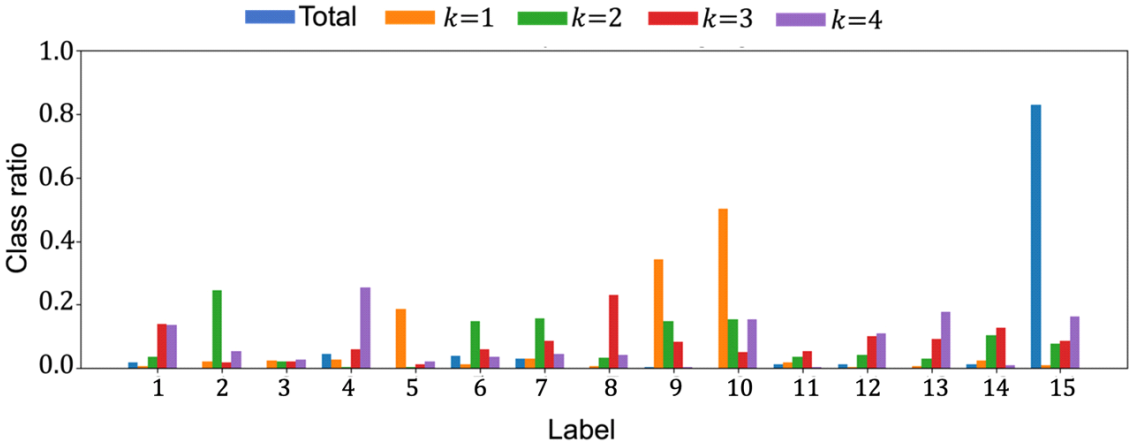


그림 A1. k 번째 탐지 구간의 이상유형 분포
Fig. A1. Class distribution of k-th detection stage

표 A1. 기호 정리 표
Table A1. Table of notation

Notation	Meaning	Notation	Meaning
$\mathcal{X}$	Total data set	x	Sample of $\mathcal{X}$
$\mathcal{Y}$	Label of $\mathcal{X}$	y	Sample of $\mathcal{Y}$
d	Number of features	J	Number of anomaly types
$\boldsymbol{Q}$	Class distribution of $\mathcal{Y}$	$\boldsymbol{Q}_k{}'$	Class distribution of $\mathcal{Y}_k'$
K	Number of total detection stage	k	Detection stage
d_k	Output node of k-th detection stage	$\mathcal{Z}_k$	Output of k-th detection stage
z_K	Sample of $\mathcal{Z}_k$	$\mathcal{Z}_k{}'$	Detected data of k-th detection stage
$\epsilon_k(\boldsymbol{z}_k)$	Anomaly score function	δ_k	Threshold
$\mathcal{Y}_k'$	Label of $\mathcal{Z}_k{}'$	$\boldsymbol{\psi}_k$	Frozen parameters
$\boldsymbol{\phi}_k$	Tunable parameters	$\{\boldsymbol{\psi}_k, \boldsymbol{\phi}_k\}$	Classification model
η	Learning rate	$\mathcal{L}$	Cross-entropy loss

표 A2. NSL-KDD 데이터 셋 모델 하이퍼파라미터 설정
Table A2. Hyperparameter settings at NSL-KDD dataset

Hyperparameter	Value	Hyperparameter	Value
Autoencoder		Classification model	
Encoder hidden	[27, 21, 15]	$\{\psi_1, \phi_1\}$	[15, 15, 5]
Decoder hidden	[15, 21, 27]	$\{\psi_2, \phi_2\}$	[27, 27, 5]
Number of detection stage	4	$\{\psi_3, \phi_3\}$	[21, 21, 5]
Number of hidden layer	2	$\{\psi_4, \phi_4\}$	[15, 15, 5]
Epochs	200	ϕ_k	Final layer
Batch size	128	Optimizer	Adam
Optimizer	Adam	Activation function	ELU
Learning rate	1e-3	Epochs(Pre-train/Fine-tune)	100/50
Activation function	ELU	Batch size(Pre-train/Fine-tune)	128/64
SVM		DNN	
C	1	Epochs	100
Kernel	rbf	Batch size	128
Degree	3	Optimizer	Adam
Gamma	scale	Learning rate	1e-3
Tol	1e-3	Hidden layer	2
Max_iter	-1	Hidden node	[27, 27, 5]
Class_weight	None	Activation function	ReLU

표 A3. CSE-CIC-IDS 2018 데이터 셋 모델 하이퍼파라미터 설정
Table A3. Hyperparameter settings at CSE-CIC-IDS 2018 dataset

Hyperparameter	Value	Hyperparameter	Value
Autoencoder		Classification model	
Encoder hidden	[68, 51, 34]	$\{\psi_1, \phi_1\}$	[34, 34, 15]
Decoder hidden	[34, 51, 68]	$\{\psi_2, \phi_2\}$	[68, 68, 15]
Number of detection stage	4	$\{\psi_3, \phi_3\}$	[51, 51, 15]
Number of hidden layer	2	$\{\psi_4, \phi_4\}$	[34, 34, 15]
Epochs	100	ϕ_k	Final layer
Batch size	256	Optimizer	Adam
Optimizer	Adam	Activation function	ELU
Learning rate	1e-4	Epochs(Pre-train/Fine-tune)	50/30
Activation function	ELU	Batch size(Pre-train/Fine-tune)	128/64
SVM		DNN	
C	1	Epochs	50
Kernel	rbf	Batch size	256
Degree	3	Optimizer	Adam
Gamma	scale	Learning rate	1e-4
Tol	1e-3	Hidden layer	2
Max_iter	-1	Hidden node	[68, 68, 15]
Class_weight	None	Activation function	ReLU