https://doi.org/10.7840/kics.2025.50.12.1811

실시간 레이블 분포 변화 적응을 위한 개인화 연합 사후훈련 연구

임 경 진•, 박 희 원*, 김 미 르*, 조 무 곤*, 권 민 혜°

Personalized Federated Post-Training for Real-time Adaptation to Label Distribution Shifts

Kyungjin Im•, Heewon Park*, Miru Kim*, Mugon Joe*, Minhae Kwon°

요 약

인공지능 모델은 학습 완료 후 개별 디바이스에 배포되며, 수집한 데이터를 대상으로 예측을 수행한다. 이때 수집된 데이터의 레이블 분포는 실시간으로 변화하기 때문에, 학습 시 경험한 분포와 실제 환경의 분포 간 차이로 인해 성능이 저하될 수 있다. 이를 해결하기 위해 사후훈련이 필요하지만, 디바이스별 데이터 양이 제한적이고 로컬 데이터에 과도하게 적응하여 사전훈련 모델이 보유한 정보가 소실될 수 있다. 본 논문에서는 Fisher Information Matrix(FIM)을 활용해 사전훈련 모델의 중요한 정보를 유지하면서 다수 디바이스가 협력하여 개인화된 모델을 생성하는 Fed-AFIR(**Fed**erated-**A**daptation with **FI**M **R**egularization)을 제안하며, 실험을 통해 기존 방식 대비 우수한 성능을 보임을 확인하였다.

키워드 : 연합학습, 사후훈련, 개인화, 레이블 분포 이동
Key Words : Federated Learning, Post-training, Personalization, Label Shift

ABSTRACT

Artificial intelligence systems deployed to individual devices are exposed to shifting label distributions over time, degrading model performance. While post-training has been studied to address this, limited local data and overfitting can cause loss of critical knowledge in the pre-trained model. We propose Fed-AFIR(**Fed**erated **A**daptation with **FI**M **R**egularization), leveraging the Fisher Information Matrix (FIM) to preserve critical parameters while enabling collaboration across devices, consistently outperforming existing approaches under dynamic label distribution and heterogeneous device data conditions.

※ 본 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2023-00278812, RS-2025-02214082).
• First Author : Soongsil University School of AI Convergence, dlarudwls1473@soongsil.ac.kr, 학생회원
° Corresponding Author : Soongsil University School of Electronic Engineering and Department of Intelligent Semiconductors, minhae@ssu.ac.kr, 종신회원
* Soongsil University Department of Intelligent Semiconductors, heewon012@soongsil.ac.kr, 학생회원; mirukim00@soongsil.ac.kr, 학생회원; mugon@soongsil.ac.kr
논문번호 : 202507-153-A-RN, Received July 4, 2025; Revised August 6, 2025; Accepted August 7, 2025

Ⅰ. 서 론

기술의 발전과 함께 인공지능 시스템은 다양한 분야에서 활용되고 있으며, 시간에 따라 변화하는 데이터 분포에 지속적으로 노출된다. 특히, 분류 모델의 경우, 데이터가 속한 레이블의 분포가 실시간으로 변화하는 환경에서 성능이 크게 저하될 수 있다[1-5]. 일반적으로 실제 환경의 데이터는 일부 레이블에 샘플이 집중된 롱테일 분포를 따르며[6-8], 이로 인해 모델은 자주 등장하는 레이블에 편향되기 쉽고, 드문 레이블에 대한 성능은 떨어지게 된다[9,10]. 이러한 모델을 디바이스에 배포할 경우, 모델은 시간에 따라 변화하는 레이블 분포에 적절히 대응하지 못해 예측에 대한 추론 성능이 감소하게 된다[11,12]. 예를 들어, 네트워크 이상 탐지 시스템에서는 초기에 정상 데이터가 대부분이지만, 시간이 지나며 다양한 공격 유형이 등장하면서 모델의 정확도가 저하될 수 있다.

이러한 문제를 해결하기 위해 사후훈련(post-training)에 대한 연구가 활발히 진행되고 있다[1-5]. 기존의 사후훈련 기술은 각 디바이스가 개별적으로 수집한 로컬 데이터만을 이용하여 독립적으로 모델을 업데이트하는 방식이지만, 다음과 같은 두 가지 한계가 존재한다. 첫째, 각 디바이스는 실제 환경에서 실시간으로 데이터를 수집하지만, 수집 가능한 데이터의 양이 제한적이기 때문에 충분한 학습이 이루어지기 어렵다. 이로 인해 디바이스의 독립적인 학습만으로는 성능 향상에 한계가 있다[13,14]. 이에 따라, 여러 디바이스들이 협력하여 공통된 모델을 학습하면서도 각 디바이스가 수집한 데이터 특성을 반영할 수 있는 개인화 연합학습(personalized federated learning) 기반의 사후훈련이 요구된다. 둘째, 디바이스의 로컬 모델이 새롭게 수집한 데이터에 과도하게 적응하면서, 사전훈련(pre-trained) 모델이 보유한 중요한 정보가 소실될 수 있다[15]. 그 결과 모델의 일반화 성능이 저하되어 변화하는 데이터 레이블 분포에 효과적으로 적응하지 못하는 문제가 발생할 수 있다[16]. 이러한 문제를 완화하기 위해, 사후훈련 과정에서 사전훈련 모델의 중요한 정보를 보존하는 기술이 필요하다.

본 논문에서는 변화하는 데이터 레이블 분포에 적응하는 동시에, 다수의 디바이스가 협력하면서 개인화된 모델을 생성할 수 있는 Fisher Information Matrix (FIM) 기반 개인화 연합 사후훈련 알고리즘인 Fed-AFIR(**Fed**erated-**A**daptation with **FI**M **R**egularization)을 제안한다. 그림 1은 Fed-AFIR의 전체 구조를 보여준다. 제안하는 알고리즘은 사전훈련이 완료된 모델로부터 각 파라미터의 중요도를 FIM을 통해 추정하고, 중요도가 높은 파라미터가 보존되는 방향으로 사후훈련을 수행한다. 또한, 각 디바이스의 로컬 모델을 공유 층(shared layers)과 개인화 층(personalized layers)으로 나누어, 개인화 연합 사후훈련을 수행함으로써 각 디바이스의 고유한 데이터 특성에 맞춘 모델을 효과적으로 학습할 수 있도록 한다.

본 논문은 다음과 같이 구성되어 있다. II장에서는 본 연구와 관련된 선행 연구에 대하여 살펴본다. III장에서는 시스템 설정 및 문제 정의를 서술하고, IV장에서는 제안하는 FIM 기반 개인화 연합 사후훈련 시스템에 대하여 서술한다. 이후 V장에서는 제안하는 시스템의 우수성을 입증하기 위한 실험 결과에 대해 서술하고 VI장에서 본 연구에 대한 결론을 맺는다.

Ⅱ. 선행 연구

2.1 분포 변화 적응 학습

개별 디바이스는 실시간으로 수집되는 데이터의 레이블 분포가 변화하는 환경에서 초기에 배포된 모델을 활용하여 추론을 수행한다[1-5]. 이때 추론 시점의 레이블 분포가 모델 학습 시점의 분포와 다를 경우 모델의 성능이 저하되는 문제가 발생할 수 있다[11,12]. 이를 해결하기 위해, 데이터 분포 변화에 적응하기 위한 사후훈련 기법들이 활발히 연구되고 있다[1-5]. 예를 들어, Unbiased Online Gradient Descent (UOGD)는 사전훈련 데이터셋의 균일한 레이블 분포를 활용하여 변화하는 레이블 분포에 모델이 편향되지 않도록 지속적으로 사후훈련을 수행한다[1]. Risk-sensitive Online Gradient Descent (ROGD)는 변화하는 환경에서의 예측 리스크를 최소화하는 방향으로 모델을 점진적으로 학습한다[2]. 또한, Adapting To LAbel Shift (ATLAS)는 시점별 데이터 레이블 분포 간 관계를 반영하여 파라미터를 조정함으로써 분포 변화에 적응하도록 학습한다[1]. 이러한 사후훈련 방식들은 단일 디바이스 환경을 가정하며, 해당 가정의 환경에서 실시간으로 수집되는 로컬 데이터의 양이 제한적인 상황일 경우 성능 향상에 한계가 있다[13,14]. 본 논문은 이러한 한계를 극복하기 위해, 다

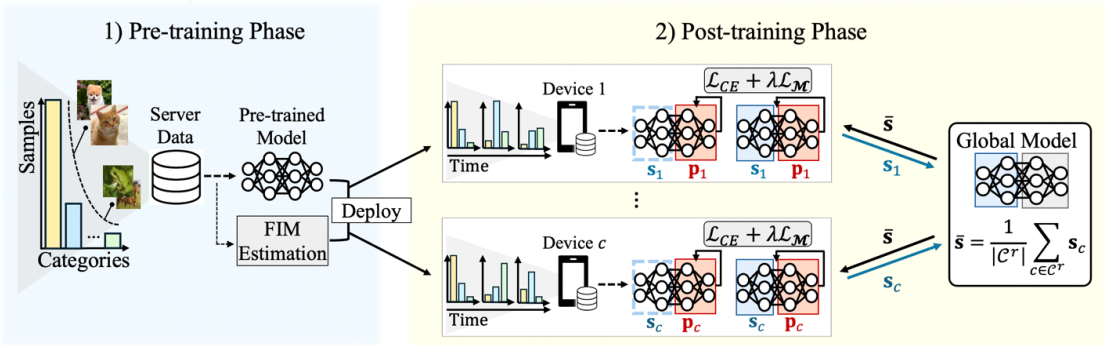


그림 1. 제안하는 Fed-AFIR의 시스템 구조
Fig. 1. Structure of the proposed Fed-AFIR system

수의 디바이스가 협력하는 연합학습 기반 사후훈련 방식을 도입하여 데이터 부족 문제를 완화하고, 변화하는 레이블 분포에 효과적으로 적응할 수 있도록 한다.

2.2 연합학습

연합학습은 분산된 데이터 환경에서 다수의 디바이스가 데이터를 직접 공유하지 않고 협력하여 인공지능 모델을 학습하는 기술이다[17-26]. 각 디바이스는 독립적으로 모델을 학습한 후, 학습된 모델 파라미터를 서버와 공유함으로써 전체 데이터의 직접적인 노출 없이 서로의 정보를 간접적으로 활용할 수 있다. 이러한 방식은 데이터의 유출 위험을 최소화하면서, 각 디바이스의 데이터 부족 문제를 완화할 수 있다[14].

대표적인 연합학습 알고리즘인 FedAvg는 각 디바이스가 보유한 로컬 데이터를 이용하여 모델을 학습한 후, 모델의 모든 파라미터를 서버로 전송하고, 서버는 이를 취합하여 글로벌 모델을 생성한 후 다시 각 디바이스에 배포한다[19]. 이와 같이 모든 디바이스가 동일한 글로벌 모델을 공유하는 구조는 디바이스 간 데이터 레이블 분포가 상이한 경우 각 디바이스의 고유한 데이터 특성을 반영하지 못하는 한계가 있다[20-26].

이를 보완하기 위해 여러 개인화 연합학습 기법들이 제안되어 왔다[22-26]. 대표적으로, 로컬 모델을 공유 층과 개인화 층으로 분리하여 학습하는 부분 공유 기법이 있다. 이 기법은 공유 층을 통한 디바이스 간 협력 학습과 개인화 층을 통한 개별 디바이스 특성에 맞춘 개인화 학습을 가능하게 한다. 예를 들어, FedBABU는 로컬 학습 단계에서 공유 층만을 학습하고, 연합학습을 완료한 이후 개인화 층만을 로컬에서 별도로 학습한다[23]. 그러나 이러한 연합학습 방식은 공유 층만이 연합학습을 수행하고 개인화 층은 연합학습을 수행하지 않기 때문에, 공유 층과 개인화 층 간의 호환성 문제가 발생할 수 있다. 또한 실제 환경에서는 디바이스가 수집하는 데이터의 레이블 분포가 시간에 따라 변화하지만, 대부분의 기존 개인화 연합학습 연구는 이러한 분포 변화 환경을 가정하지 않고 있다. 본 논문은 이러한 호환성 문제와 분포 변화에 대한 적응력 부족을 해결할 수 있는 알고리즘을 제안한다.

2.3 Fisher Information Matrix (FIM)

FIM은 주어진 데이터가 특정 파라미터에 대해 얼마나 많은 정보를 담고 있는지를 정량적으로 나타내는 지표로, 파라미터의 확률 분포를 기반으로 정보의 양을 측정하는 데 사용된다. 일반적으로 FIM의 값이 클수록 해당 파라미터는 모델의 예측에 큰 영향을 미치는 중요한 파라미터임을 의미한다. 이는 해당 파라미터가 모델의 출력 확률 분포를 형성하는 데 높은 기여도를 가진다는 것을 의미하며, 미세한 파라미터의 변화만으로 모델의 출력이 크게 달라질 수 있음을 나타낸다. 따라서 이러한 파라미터를 보존하는 것은 사전훈련 모델이 학습한 주요 정보를 유지하는 데 중요한 역할을 한다. Elastic Weight Consolidation은 FIM을 활용하여 새로운 데이터를 학습할 때 기존에 학습된 지식을 잊지 않도록 규제하지만[15], 단일 디바이스의 사후훈련 환경에만 적용되므로 데이터가 제한된 다수의 디바이스 환경에서는 성능 향상에 한계가 있다[13-14].

FedAS는 FIM을 도입한 개인화 연합학습 알고리즘으로, FIM의 대각합을 활용하여 각 디바이스

의 연합학습 기여도를 평가하고 공유 층의 파라미터를 가중 집계한다[24]. 이러한 개인화 연합학습 방식은 데이터 부족 문제를 완화할 수 있으나, 개별 디바이스의 로컬 모델 학습 과정에서는 FIM을 활용하지 않기 때문에 모델이 사전훈련을 통해 확보한 중요 파라미터들을 소실할 수 있다[15,16]. 이에 본 논문에서는 사전훈련 단계에서 계산한 FIM 값을 개인화 연합 사후훈련 과정에서 손실 함수의 정규화 항으로 활용하여, 사전훈련 모델의 중요 파라미터를 보존할 수 있도록 한다. 이 접근은 각 디바이스의 고유한 데이터 특성을 반영하면서도, 파라미터 수준에서 안정성을 유지하는 개인화 연합 사후훈련을 가능하게 한다.

Ⅲ. 시스템 설정

본 장에서는 사전훈련 과정과 시간에 따라 레이블 분포가 변화하는 사후훈련 과정의 시스템 설정을 서술하고, 이를 바탕으로 문제를 정의한다.

사전 훈련 단계 ($t=0$): 서버 G는 사전훈련에 사용되는 데이터셋 $\mathbf{X}_G^0$과 이에 대한 레이블 $\mathbf{Y}_G^0$을 보유하고 있으며, 사전훈련 데이터셋($\mathbf{X}_G^0$, $\mathbf{Y}_G^0$)은 롱테일 레이블 분포 $\mathbf{Y}_G^0 \sim \mathbf{Q}_G^0$를 따른다. 서버는 cross-entropy 손실 함수 $\mathcal{L}_{CE}\left(\boldsymbol{\theta}_G; (\mathbf{X}_G^0, \mathbf{Y}_G^0)\right)$를 기반으로 사전훈련 모델 $\boldsymbol{\theta}_G$를 학습한다.

사후 훈련 단계 ($1 \le t \le T$): 사전훈련 모델 $\boldsymbol{\theta}_G$는 디바이스 집합 $\mathcal{C}$의 각 디바이스 $c \in \mathcal{C}$에 배포되며, 디바이스 c의 로컬 모델 $\boldsymbol{\theta}_c$는 사전훈련 모델과 동일한 구조를 가진다. 이후 각 디바이스는 전체 시간 T동안, 시점 t ($1 \le t \le T$)마다 변화하는 레이블 분포 $\mathbf{Y}_c^t \sim \mathbf{Q}_c^t$를 따르는 데이터 ($\mathbf{X}_c^t$, $\mathbf{Y}_c^t$)를 수집한다. 디바이스 c의 로컬 데이터셋 $\mathbf{X}_c^t \subset \mathbb{R}^{M_c^t \times N}$은 총 M_c^t개의 샘플로 구성되며, 각 샘플 $\mathbf{x}_{c,m}^t \in \mathbb{R}^{1\times N}$은 N차원의 feature를 포함한다. 데이터셋 $\mathbf{X}_c^t$에 대응하는 레이블은 $\mathbf{Y}_c^t$로 표현되며, 로컬 데이터셋 $\mathbf{X}_c^t$는 다음과 같이 표현할 수 있다.

$$\mathbf{X}_c^t = \left[\mathbf{x}_{c,1}^t, \mathbf{x}_{c,2}^t, \ldots, \mathbf{x}_{c,m}^t, \ldots, \mathbf{x}_{c,M_c^t}^t\right]^{\mathrm{T}} \tag{1}$$

이때 각 디바이스 c가 시점 t에서 수집한 사후훈련 데이터는 독립적인 레이블 분포 $\mathbf{Q}_c^t$를 따른다고 가정한다. 레이블 분포 $\mathbf{Q}_c^t$는 수식 (2)와 같이 정의한다.

$$\mathbf{Q}_c^t = (1-\alpha^t)\mathbf{Q}_G^0 + \alpha^t \mathbf{Q}_c^T \tag{2}$$

위 수식에서 α^t는 t시점에서의 분포 변화 계수이며, $\mathbf{Q}_c^T$는 디바이스 c의 최종 시점 T에서의 레이블 분포를 나타낸다. 디바이스 c는 서버에서 배포 받은 사전훈련 모델을 기반으로, $\mathbf{Y}_c^t \sim \mathbf{Q}_c^t$를 따르는 사후훈련 데이터셋 ($\mathbf{X}_c^t$, $\mathbf{Y}_c^t$)을 활용하여 사후훈련을 수행한다.

개인화 연합학습 기반 사후훈련 과정은 각 시점 t 마다 수행되며, 각 통신 라운드 r ($1 \le r \le R$) 마다 일부 디바이스 집합 $\mathcal{C}^r \subset \mathcal{C}$이 학습에 참여한다. 연합학습에 참여한 디바이스는 사후훈련 데이터셋 ($\mathbf{X}_c^t$, $\mathbf{Y}_c^t$)를 활용하여 자신의 로컬 모델 $\boldsymbol{\theta}_c = \{\mathbf{s}_c, \mathbf{p}_c\}$를 학습한다. 이때, $\mathbf{s}_c$는 연합학습을 위해 서버와 파라미터를 공유하는 공유 층, $\mathbf{p}_c$는 서버와 파라미터를 공유하지 않고 로컬에서만 학습되는 개인화 층을 의미한다. 디바이스 c는 로컬 모델의 공유 층 $\mathbf{s}_c$를 서버에 전송하고, 서버는 개별 디바이스로부터 $\mathbf{s}_c$를 취합하여 글로벌 모델 $\bar{\mathbf{s}}$을 수식 (3)과 같이 생성한다.

$$\bar{\mathbf{s}} = \frac{1}{|\mathcal{C}^r|}\sum_{c\in\mathcal{C}^r} \mathbf{s}_c \tag{3}$$

이후 서버는 글로벌 모델 $\bar{\mathbf{s}}$를 모든 디바이스에 전송하고, 각 디바이스는 로컬 모델의 공유 층 $\mathbf{s}_c$을 글로벌 모델 $\bar{\mathbf{s}}$로 업데이트한다.

이러한 개인화 연합학습 기반 사후훈련의 목적은 각 디바이스에 개인화 모델을 생성하면서도 공통된 정보를 공유할 수 있는 모델 집합 $\{\bar{\mathbf{s}}^*, \mathbf{p}_1{}^*, \ldots, \mathbf{p}_{|\mathcal{C}|}{}^*\}$을 찾아 전체 손실 함수를 최소화하는 것이다.

$$\{\bar{\mathbf{s}}^*, \mathbf{p}_1{}^*, \ldots, \mathbf{p}_{|\mathcal{C}|}{}^*\} = \arg\min_{\{\bar{\mathbf{s}}, \mathbf{p}_1, \ldots, \mathbf{p}_{|\mathcal{C}|}\}} \sum_{c=1}^{|\mathcal{C}|} \mathcal{F}_c(\bar{\mathbf{s}}, \mathbf{p}_c)$$
$$\mathcal{F}_c(\bar{\mathbf{s}}, \mathbf{p}_c) \coloneqq \mathbb{E}_{\mathbf{Y}_c^t \sim \mathbf{Q}_c^t} \mathcal{L}_{CE}\left(\{\bar{\mathbf{s}}, \mathbf{p}_c\}; (\mathbf{X}_c^t, \mathbf{Y}_c^t)\right) \tag{4}$$

수식 (4)의 $\mathcal{F}_c(\bar{\mathbf{s}}, \mathbf{p}_c)$은 디바이스 c의 로컬 모델 $\{\bar{\mathbf{s}}, \mathbf{p}_c\}$과 시점 t에 수집된 로컬 데이터 ($\mathbf{X}_c^t$, $\mathbf{Y}_c^t$)를 이용하여 계산된 손실 함수의 기댓값으로 정의된다. 실제 환경에서는 많은 수의 디바이스가 존재하기 때문에 수식 (4)를 달성하는 데 어려움이 있다. 따라서 각 디바이스의 학습 목적은 로컬에서 수행하는 방식으로 다음과 같이 재정의할 수 있다.

$$\{\mathbf{s}_c^*, \mathbf{p}_c^*\} = \arg\min_{\{\mathbf{s}_c, \mathbf{p}_c\}} \mathbb{E}_{\mathbf{Y}_c^t \sim \mathbf{Q}_c^t} \mathcal{L}_{CE}\left(\{\mathbf{s}_c, \mathbf{p}_c\}; (\mathbf{X}_c^t, \mathbf{Y}_c^t)\right) \quad (5)$$

디바이스 c는 수식 (5)에 대한 로컬 학습을 수행하고, 서버는 지속적으로 공유 층을 집계하는 방식을 수행함으로써 정의된 문제를 해결한다.

Ⅳ. FIM 기반 개인화 연합 사후훈련

본 장에서는 제안하는 Fed-AFIR의 전체 구조를 설명한다. Fed-AFIR는 FIM을 활용하여 사후훈련 손실함수를 정의하고, 제안하는 개인화 연합학습 알고리즘으로 사후훈련을 수행한다. Fed-AFIR의 동작 과정은 Algorithm 1에 명시되어 있다.

4.1 FIM 기반 사후훈련 손실함수 정의

서버는 사전훈련 데이터셋 ($\mathbf{X}_G^0$, $\mathbf{Y}_G^0$)을 활용하여 모델 $\boldsymbol{\theta}_G$을 학습한 이후, 사전훈련 모델 $\boldsymbol{\theta}_G$의 중요 파라미터를 평가하기 위해 FIM을 활용한다. FIM을 활용하여 각 파라미터가 모델의 예측 결과에 미치는 변화량 $\boldsymbol{\mathcal{M}}$은 수식 (6)과 같이 계산한다.

$$\boldsymbol{\mathcal{M}} = \left[\nabla_{\boldsymbol{\theta}_G} \ell(\boldsymbol{\theta}_G; \mathbf{X}_G^0)\right]\left[\nabla_{\boldsymbol{\theta}_G} \ell(\boldsymbol{\theta}_G; \mathbf{X}_G^0)\right]^{\mathrm{T}} \quad (6)$$

위 수식의 $\ell(\boldsymbol{\theta}_G; \mathbf{X}_G^0)$는 입력 데이터 $\mathbf{X}_G^0$에 대해 사전훈련 모델 $\boldsymbol{\theta}_G$이 예측한 로그 확률을 의미한다. $\boldsymbol{\mathcal{M}}$의 값이 클수록 해당 파라미터가 모델의 예측 결과에 미치는 영향이 크다는 것을 의미하며, 사전훈련 모델이 학습한 중요 파라미터임을 나타낸다. 따라서, $\boldsymbol{\mathcal{M}}$의 값이 큰 파라미터의 변화를 최소화하도록 규제함으로써, 사전훈련 모델이 가진 중요한 정보를 사후훈련 과정에서도 유지할 수 있도록 한다.

사후훈련 단계에서는 사전훈련 단계에서 계산된 $\boldsymbol{\mathcal{M}}$을 기반으로 손실 함수를 정의하고, 이를 활용하여 개인화 연합 사후훈련을 수행한다. 우선 각 디바이스 c의 로컬 모델 $\boldsymbol{\theta}_c$이 사전훈련 모델 $\boldsymbol{\theta}_G$로부터 변화한 정도를 기준으로, $\boldsymbol{\mathcal{M}}$의 값이 큰 파라미터는 사전훈련 모델 $\boldsymbol{\theta}_G$의 파라미터로부터 크게 변화하지 않도록 규제하는 정규화 항 $\mathcal{L}_{\boldsymbol{\mathcal{M}}}(\boldsymbol{\theta}_c, \boldsymbol{\theta}_G)$을 수식 (7)과 같이 정의한다.

$$\mathcal{L}_{\boldsymbol{\mathcal{M}}}(\boldsymbol{\theta}_c, \boldsymbol{\theta}_G) = \boldsymbol{\mathcal{M}} \times (\boldsymbol{\theta}_c - \boldsymbol{\theta}_G)^2 \quad (7)$$

Algorithm 1 Fed-AFIR

Input: Pre-trained model $\boldsymbol{\theta}_G$, FIM $\boldsymbol{\mathcal{M}}$, communication rounds R, local epochs E
Server distributes $\boldsymbol{\theta}_G$ and $\boldsymbol{\mathcal{M}}$ to $|\mathcal{C}|$ devices
Each device $c \in \mathcal{C}$:
Initializes local model $\boldsymbol{\theta}_c = \{\mathbf{s}_c, \mathbf{p}_c\} \leftarrow \boldsymbol{\theta}_G$
for communication round r: 1 to R **do**
  for each device $c \in \mathcal{C}^r$ **do**
    for local epoch e: 1 to E **do**
      Update shared layers $\mathbf{s}_c \leftarrow \bar{\mathbf{s}}$
      Freeze shared layers $\mathbf{s}_c$
      Update personalized layer $\mathbf{p}_c$ as (9)
    end for
    for local epoch e: 1 to E **do**
      Unfreeze shared layers $\mathbf{s}_c$
      Update local model $\{\mathbf{s}_c, \mathbf{p}_c\}$ as (10)
    end for
    Send $\mathbf{s}_c$ to server
  end for
  Server calculates $\bar{\mathbf{s}} = \frac{1}{|\mathcal{C}^r|}\sum_{c \in \mathcal{C}^r} \mathbf{s}_c$
end for

이후, 각 디바이스 c는 수식 (8)과 같이 정규화 항 $\mathcal{L}_{\boldsymbol{\mathcal{M}}}(\boldsymbol{\theta}_c, \boldsymbol{\theta}_G)$과 cross-entropy 손실함수 $\mathcal{L}_{CE}\left(\boldsymbol{\theta}_c; (\mathbf{X}_c^t, \mathbf{Y}_c^t)\right)$를 포함한 사후훈련 손실 함수 $\mathcal{L}_{post}\left(\boldsymbol{\theta}_c, \boldsymbol{\theta}_G; (\mathbf{X}_c^t, \mathbf{Y}_c^t)\right)$를 최소화하도록 학습을 수행한다.

$$\mathcal{L}_{post}\left(\boldsymbol{\theta}_c, \boldsymbol{\theta}_G; (\mathbf{X}_c^t, \mathbf{Y}_c^t)\right) = \mathcal{L}_{CE}\left(\boldsymbol{\theta}_c; (\mathbf{X}_c^t, \mathbf{Y}_c^t)\right) + \lambda \mathcal{L}_{\boldsymbol{\mathcal{M}}}(\boldsymbol{\theta}_c, \boldsymbol{\theta}_G) \quad (8)$$

위 수식에서 λ는 정규화 항의 정도를 조절하는 계수로, 사후훈련 과정에서 사전훈련 모델 $\boldsymbol{\theta}_G$의 중요 파라미터가 지나치게 변경되지 않도록 조정한다.

4.2 개인화 연합학습 알고리즘

제안하는 Fed-AFIR는 모델을 공유 층과 개인화 층으로 나누며, 수식 (8)에서 정의한 사후훈련 손실함수 $\mathcal{L}_{post}\left(\boldsymbol{\theta}_c, \boldsymbol{\theta}_G; (\mathbf{X}_c^t, \mathbf{Y}_c^t)\right)$를 이용하여 개인화 연합 사후훈련을 수행한다. 이때, 연합학습 과정에서 로컬 모델의 공유 층 $\mathbf{s}_c$을 글로벌 모델 $\bar{\mathbf{s}}$로 갱신한 뒤, 총 두 단계에 걸쳐 로컬 모델 학습을 수행한다.

첫 번째 단계로, 글로벌 모델로 갱신된 로컬 모델의 공유 층 $\mathbf{s}_c$을 고정하고 개인화 층 $\mathbf{p}_c$만을 수식 (9)와 같이 학습한다.

$$\mathbf{p}_c \leftarrow \mathbf{p}_c - \eta \nabla \mathcal{L}_{post}\left(\boldsymbol{\theta}_c, \boldsymbol{\theta}_G; (\mathbf{X}_c^t, \mathbf{Y}_c^t)\right) \quad (9)$$

위 수식의 η는 학습률을 의미한다. 첫 번째 단계에서 연합학습에 직접적으로 참여하지 않은 개인화 층만을 학습함으로써 개인화 층과 공유 층 사이의 정보 격차를 줄여 공유 층과 개인화 층 간 호환성을 확보할 수 있다.

두 번째 단계에서는 공유 층 $\mathbf{s}_c$과 개인화 층 $\mathbf{p}_c$을 수식 (10)과 같이 업데이트한다.

$$\{\mathbf{s}_c, \mathbf{p}_c\} \leftarrow \{\mathbf{s}_c, \mathbf{p}_c\} - \eta \nabla \mathcal{L}_{post}\left(\boldsymbol{\theta}_c, \boldsymbol{\theta}_G; (\mathbf{X}_c^t, \mathbf{Y}_c^t)\right) \quad (10)$$

첫 번째 단계에서 확보한 공유 층과 개인화 층 간의 호환성을 바탕으로, 두 번째 단계에서는 로컬 모델 전체를 안정적으로 학습할 수 있으며 이를 통해 각 디바이스의 데이터 레이블 분포에 최적화된 모델을 생성할 수 있다.

이와 같은 방법으로, 본 논문에서 제안하는 Fed-AFIR는 사전훈련 단계에서 계산된 FIM 정보를 활용하여 중요 파라미터의 변화를 최소화하도록 한다. 또한 디바이스의 로컬 모델을 공유 층과 개인화 층으로 나누어 사후훈련을 수행함으로써, 각 디바이스의 데이터 레이블 분포를 반영한 개인화 모델을 학습할 수 있도록 한다.

V. 실험 결과

본 장에서는 제안하는 알고리즘의 우수성을 다양한 데이터 레이블 분포 변화 환경과 비교 알고리즘을 통해 실험적으로 입증한다.

5.1 실험 설정

데이터셋: 제안하는 Fed-AFIR의 성능을 평가하기 위해, 네트워크 데이터셋인 NSL-KDD[27]와 CSE-CIC-IDS 2018[28], 그리고 이미지 데이터셋인 CIFAR-10[29], CIFAR-100[29]을 대상으로 실험을 진행하였다. NSL-KDD는 총 148,417개의 샘플을 가지는 대표적인 불균형 네트워크 침입 탐지 데이터셋으로, 네 개의 공격 레이블과 한 개의 정상 레이블로 구성되어 있다. CSE-CIC-IDS 2018 데이터셋은 총 15,450,706개의 샘플로 구성되며, 열 네 개의 공격 레이블과 한 개의 정상 레이블로 이루어진 불균형 네트워크 침입 탐지 데이터셋이다. CIFAR-10은 32×32 feature의 RGB 컬러 이미지 60,000개로 구성되어 있으며, 열 개의 이미지 레이블이 균일한 비율로 분포되어 있다. CIFAR-100은 CIFAR-10과 동일한 크기와 이미지 수로 구성되어 있으며, 백 개의 이미지 레이블이 균일한 비율로 분포되어 있는 데이터셋이다.

데이터 설정: NSL-KDD, CIFAR-10, CIFAR-100 데이터셋은 기본적으로 학습 데이터와 테스트 데이터로 분리되어 제공되며, CSE-CIC-IDS 2018 데이터셋은 전체 데이터를 5:1의 비율로 분할하여 학습 데이터와 테스트 데이터를 구성하였다. 이후 모든 학습 데이터는 사전훈련 데이터셋과 사후훈련 데이터셋으로 나누기 위해 4:1의 비율로 분할하였다. 이때 사전훈련 데이터셋은 데이터의 레이블 분포가 불균형한 실제 환경을 고려하여 구성하였으며, 네트워크 데이터셋은 본래 불균형 레이블 분포를 따르기 때문에 해당 분포를 유지한 상태로 사전훈련 데이터셋을 구성하여 실험을 진행하였다. 이미지 데이터셋은 본래 균일한 레이블 분포를 따르기 때문에 사전훈련 데이터셋의 레이블 인덱스가 증가할수록 샘플 수가 지수적으로 감소하는 롱테일 레이블 분포를 따르도록 설정하였다. 이때, 데이터의 레이블 분포는 불균형 계수 Imbalance Factor (IF)를 사용하여 나타내며, 불균형 계수 IF는 가장 많은 샘플을 가진 레이블의 샘플 수 N_{max}를 가장 적은 샘플을 가진 레이블의 샘플 수 N_{min}로 나눈 값 $IF = \frac{N_{max}}{N_{min}}$으로 정의된다. 실제적인 롱테일 레이블 환경을 고려하여 불균형 계수 IF는 50으로 설정하였다[30-32]. 사후훈련 데이터셋은 디바이스 간 데이터 레이블 분포가 상이한 환경을 구성하기 위해, 총 100개의 디바이스에 디리클레 분포를 기반으로 데이터셋을 분배하였다[21-24,26].

레이블 분포 변화 환경: 본 실험은 데이터의 레이블 분포가 시간에 따라 변화하는 환경을 반영하기 위해, 각 디바이스에 대해 초기 레이블 분포 $\mathbf{Q}_c^0$와 최종 레이블 분포 $\mathbf{Q}_c^T$를 설정하고, 수식 (2)에 따라 시점 $t \in [1, 100]$ 에서의 레이블 분포 $\mathbf{Q}_c^T$를 설정하였다[1,2,5]. 초기 레이블 분포 $\mathbf{Q}_c^0$는 사전훈련 단계에서 사용된 분포와 동일하게 설정하였으며, 최종 레이블 분포 $\mathbf{Q}_c^T$는 디리클레 분포를 활용하여 각 디바이스별로 상이한 레이블 분포를 따르도록 설정하였다. 수식 (2)에 정의된 시점 t에서의 레이블 분포 변화 $\mathbf{Q}_c^T$는 분포 변화 계수 α'를 기반으로 설정되며, $\alpha' = 0$일 경우 초기 분포 $\mathbf{Q}_c^0$를 따르고, $\alpha' = 1$일 경우 최종 분포 $\mathbf{Q}_c^T$를 완전히 따르도록 구성하였다. 이는 시간에 따라 사전훈련 모델이 기반한 초기 분포에서 디바이스마다 서로 다른 레이블

분포로 변화하도록 하기 위함이다. 본 실험에서는 t가 증가함에 따라 데이터의 레이블 분포가 점진적으로 또는 급격하게 이동하는 실제 환경을 반영하기 위해, 분포 변화 계수 α^t를 그림 2와 같이 두 가지 형태의 패턴으로 정의하였다. 사인(Sin.) 형태는 α^t가 시간에 따라 사인 함수 곡선을 따르며 점진적인 분포 이동을 나타낸다. 사각(Squ.) 형태는 α^t가 일정 간격마다 0과 1로 급격히 전환되는 비연속적인 분포 변화를 나타낸다.

비교 알고리즘: 제안하는 FIM 기반 개인화 연합 사후훈련 알고리즘의 우수성을 입증하기 위한 성능 비교 대상은 다음과 같다.

- Local: 각 디바이스가 로컬 데이터를 활용하여 단독으로 모델을 학습하는 방식으로, 디바이스가 가진 데이터의 양이 제한적이기 때문에 성능 향상에 한계를 가진다.
- ATLAS[1]: 레이블 분포 변화에 대응하기 위해 제안된 로컬 학습 알고리즘으로, 시점에 따라 학습률을 조정함으로써 분포 변화에 적응한다.
- UOGD[1]: 레이블 분포 변화에 대응하기 위해 제안된 로컬 학습 알고리즘으로, 레이블이 없는 상태에서도 모델을 업데이트할 수 있도록 한다.
- ROGD[2]: 레이블 분포 변화에 대응하기 위해 제안된 로컬 학습 알고리즘으로, 예측 리스크를 최소화하며 모델을 점진적으로 업데이트한다.
- FedAvg[19]: 대표적인 연합학습 알고리즘으로, 디바이스가 로컬 모델 전체를 서버와 공유하며 연합학습을 수행한다. 이는 연합학습 과정에서 개별 디바이스의 데이터 특성을 반영할 수 없기 때문에 해당 알고리즘과 비교하여 개인화 연합학습의 유효성을 확인한다.
- FedBABU[23]: 개인화 연합학습 알고리즘으로, 연합학습 단계에서 공유 층만을 학습하고, 연합학습이 끝난 후 개인화 층만을 로컬 데이터로 재학습한다. 연합학습과 개인화 학습이 분리되어 수행되기 때문에 두 층 간 호환성 문제가 더욱 심화될 수 있다. 해당 알고리즘과의 비교를 통해 두 층 간 호환성 확보의 이점을 확인한다.
- FedAS[24]: 개인화 연합학습 알고리즘으로, FIM을 활용하여 공유 층 집계 시 가중 평균을 적용한다. 개인화 연합 사후훈련 과정에서 제안하는 Fed-AFIR와 동일하게 FIM을 활용하지만, 사전훈련 모델의 정보 소실 문제는 고려하지 않는다. 해당 알고리즘과의 비교를 통해 사후훈련 과정에서 사전훈련 모델의 중요한 정보를 유지하는 방식의 이점을 확인한다.

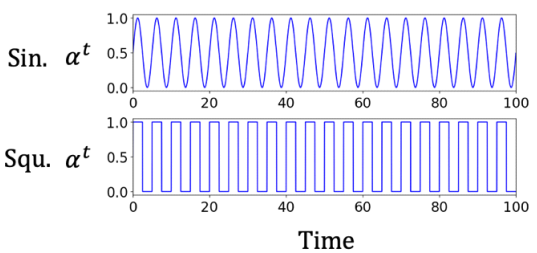


그림 2. 시점 t에 따른 분포 유형별 α^t의 변화
Fig. 2. Temporal variation of α^t by distribution type

5.2 실험 결과

제안하는 알고리즘의 성능을 평가하기 위해, 본 실험에서는 훈련 데이터셋과 동일한 분포를 갖는 테스트셋을 사용하여 100개의 디바이스에 대해 $t \in [1, 100]$ 구간의 평균 성능을 측정하였다. 실험은 5개의 랜덤시드에 대해 수행되었으며, 랜덤시드별 성능의 평균과 표준편차를 함께 나타내었다.

5.2.1 네트워크 데이터셋에 대한 성능 분석

표 1은 네트워크 데이터셋 NSL-KDD, CSE-CIC-IDS 2018에 대해 두 가지 분포 변화 유형(Sin., Squ.) 상황에서의 성능 비교 결과를 포함하고 있다. 먼저, NSL-KDD에서 두 가지 분포 변화에 대해 평균적으로 Local은 61.6%, ATLAS는 63.2%, UOGD는 60.4%, ROGD는 61.1%의 정확도를 나타내었으며, 분포 변화에 적응하기 위한 알고리즘들이 Local과 유사한 수준의 성능을 보였다. 이와 동일하게 CSE-CIC-IDS 2018에서도 ATLAS, UOGD, ROGD는 모두 Local과 유사한 정확도를 보였다. 이는 분산된 데이터 환경에서 제한된 양의 로컬 데이터만으로는 시간에 따른 네트워크 공격 유형 변화에 효과적으로 대응하기 어렵다는 것을 의미한다. 반면에 제안하는 Fed-AFIR는 전체 로컬 기반 사후훈련 알고리즘(Local, ATLAS, UOGD, ROGD)의 평균 성능 대비 NSL-KDD에서 13.6%, CSE-CIC-IDS 2018에서 8.3% 성능 향상을 보였으며, 단일 디바이스 기반 사후훈련 방식의 한계를 보완할 수 있음을 확인할 수 있다.

한편, 개인화 연합학습 알고리즘인 FedAS는 모든 분포 변화에 대해 평균적으로 FedAvg 대비 NSL-KDD에서 0.7%, CSE-CIC-IDS 2018에서 1.5% 더 높은 성능을 보였다. 이는 각 디바이스의 고유한 데이터 특성을 반영한 개인화 학습이 성능 향상

표 1. 레이블 분포 변화에 따른 실험 결과
Table 1. Performance Comparison under Label Distribution Shifts

Method	NSL-KDD		CSE-CIC-IDS 2018		CIFAR-10		CIFAR-100	
	Sin.	Squ.	Sin.	Squ.	Sin.	Squ.	Sin.	Squ.
Local	$61.8_{\pm 2.5}$	$61.3_{\pm 2.9}$	$78.3_{\pm 3.3}$	$78.3_{\pm 2.9}$	$49.6_{\pm 3.9}$	$50.4_{\pm 4.1}$	$21.7_{\pm 1.1}$	$22.1_{\pm 1.2}$
ATLAS [1]	$62.9_{\pm 0.6}$	$63.4_{\pm 0.6}$	$76.8_{\pm 2.1}$	$77.2_{\pm 2.1}$	$51.8_{\pm 0.7}$	$51.0_{\pm 0.7}$	$24.5_{\pm 0.5}$	$24.8_{\pm 0.5}$
UOGD [1]	$59.0_{\pm 0.7}$	$61.9_{\pm 0.6}$	$78.4_{\pm 1.9}$	$78.6_{\pm 1.8}$	$59.8_{\pm 0.5}$	$58.1_{\pm 0.6}$	$26.0_{\pm 0.5}$	$26.1_{\pm 0.5}$
ROGD [2]	$59.7_{\pm 0.8}$	$62.5_{\pm 0.7}$	$78.3_{\pm 1.9}$	$78.5_{\pm 1.8}$	$60.0_{\pm 0.5}$	$58.4_{\pm 0.6}$	$26.1_{\pm 0.4}$	$26.2_{\pm 0.5}$
FedAvg [19]	$67.3_{\pm 1.7}$	$68.6_{\pm 1.2}$	$81.3_{\pm 1.4}$	$81.7_{\pm 1.4}$	$61.5_{\pm 0.8}$	$60.6_{\pm 1.5}$	$28.2_{\pm 0.5}$	$28.6_{\pm 0.5}$
FedBABU [23]	$64.5_{\pm 1.5}$	$65.5_{\pm 1.4}$	$81.6_{\pm 1.5}$	$81.6_{\pm 1.3}$	$56.4_{\pm 0.9}$	$56.2_{\pm 1.0}$	$26.2_{\pm 0.6}$	$26.3_{\pm 0.6}$
FedAS [24]	$68.4_{\pm 0.7}$	$68.5_{\pm 0.7}$	$82.7_{\pm 2.0}$	$82.7_{\pm 1.9}$	$64.6_{\pm 0.6}$	$63.6_{\pm 0.5}$	$32.6_{\pm 0.8}$	$32.2_{\pm 0.7}$
Fed-AFIR (Ours)	**$70.6_{\pm 1.1}$**	**$69.4_{\pm 1.7}$**	**$84.6_{\pm 2.0}$**	**$84.7_{\pm 1.9}$**	**$65.7_{\pm 0.7}$**	**$66.0_{\pm 0.5}$**	**$33.1_{\pm 0.8}$**	**$33.5_{\pm 0.7}$**

에 기여했음을 보여준다.

제안하는 알고리즘인 Fed-AFIR는 모든 분포 변화 유형에 대해 NSL-KDD에서 FedAS 대비 2.3% 더 높은 성능을 보였으며, CSE-CIC-IDS 2018에서도 2.3% 성능 향상을 보였다. 이는 제안하는 Fed-AFIR가 FIM 기반 정규화를 통해 사전훈련 모델의 중요 파라미터를 유지함으로써 변화하는 레이블 분포 환경에 적응하였음을 의미한다.

5.2.2 이미지 데이터셋에 대한 성능 분석

표 1에 포함된 CIFAR-10, CIFAR-100의 결과는 이미지 데이터셋에서의 성능 비교를 보여준다. 이미지 데이터셋에서도 네트워크 데이터셋과 유사한 성능 경향성을 보였다. 모든 분포 환경에 대해 전체 로컬 기반 사후훈련 방식은 CIFAR-10과 CIFAR-100에서 각각 평균적으로 54.9%, 24.7%의 낮은 성능을 보인 반면, 연합학습 기반 사후훈련 방식인 FedAvg는 로컬 기반 사후훈련 방식 대비 각각 11.1%, 15.0% 성능 향상을 보였다. 이는 이미지 데이터셋에서도 연합학습 기반 접근 방식이 단일 디바이스 환경에서의 성능 저하를 완화하는 데 효과적임을 보여준다.

개인화 연합학습 알고리즘인 FedAS는 CIFAR-10과 CIFAR-100에서 각각 64.1%, 32.4%의 정확도를 기록하였으며, FedAvg 대비 각각 5.1%, 14.1% 성능 향상을 보였다. 이러한 결과는 이미지 데이터셋에서도 개별 디바이스의 데이터 분포를 반영한 개인화 연합학습이 성능 향상에 기여했음을 나타낸다. 이와 대조적으로, FedBABU는 모든 네트워크 데이터셋과 이미지 데이터셋에 대해 FedAvg보다 낮은 성능을 기록하였다. 이는 연합학습 단계에서 공유 층만을 학습한 후, 연합학습 이후에 개인화 층을 로컬에서 학습하는 방식이 공유 층과 개인화 층 간 호환성 문제를 심화시켰기 때문에 성능이 저하되었음을 알 수 있다.

제안하는 Fed-AFIR는 모든 분포 변화에 대해 평균적으로 CIFAR-10에서 65.8%, CIFAR-100에서 33.3%로 가장 높은 성능을 보였다. 특히 레이블 수가 많은 CIFAR-100에서 제안하는 Fed-AFIR는 FedAS 대비 2.8%의 성능 향상을 보였다. 이는 복잡한 이미지 분류 환경에서도 제안하는 Fed-AFIR가 사전훈련 모델의 중요 파라미터를 FIM 기반 정규화를 통해 보존함과 동시에 공유 층과 개인화 층 간의 호환성을 효과적으로 유지함을 알 수 있다. 이를 통해 Fed-AFIR는 변화하는 레이블 분포 환경에 적응하고, 각 디바이스에 개인화된 모델을 생성할 수 있음을 확인할 수 있다.

Ⅵ. 결 론

본 논문에서는 변화하는 데이터의 레이블 분포에 적응하면서 개인화된 모델을 생성할 수 있는 FIM 기반 개인화 연합 사후훈련 알고리즘 Fed-AFIR를 제안하였다. 제안하는 Fed-AFIR는 사후훈련 과정에서 디바이스에 배포된 모델이 학습한 정보가 소실되는 문제를 완화하고자, 사전훈련 단계에서 FIM을 활용하여 파라미터 중요도를 추정하고 사후훈련 과정에서 중요 파라미터를 보존할 수 있도록 한다. 또한, 로컬 모델을 공유 층과 개인화 층으로 분리하여 학습함으로써, 디바이스별 고유한 데이터 특성을 효과적으로 반영할 수 있도록 한다. 다양한 레이블 분포 변화 환경에서

실험을 진행한 결과, 기존 사후훈련 및 개인화 연합학습 연구 결과 대비 높은 개인화 성능과 분포 변화 적응력을 확인할 수 있었다. 이러한 결과를 통해 제안하는 Fed-AFIR가 데이터의 레이블 분포가 변화하는 환경에 적응하면서도 디바이스의 고유한 데이터 특성을 반영할 수 있음을 입증하였다.

References

[1] Y. Bai, Y. Zhang, P. Zhao, M. Sugiyama, and Z. Zhou, “Adapting to online label shift with provable guarantees,” in *NeurIPS*, Los Angeles, USA, Nov. 2022.

[2] R. Wu, C. Guo, Y. Su, and K. Q. Weinberger, “Online adaptation to label distribution shift,” in *NeurIPS*, New York, USA, Dec. 2021.

[3] Y. Y. Qian, Y. Bai, Z. Y. Zhang, P. Zhao, and Z. H. Zhou, “Handling new class in online label shift,” in *IEEE ICDM*, Shanghai, China, Sep. 2023. (https://doi.org/10.1109/ICDM58522.2023.00162)

[4] S. Park, S. Yang, J. Choo, and S. Yun, “Label shift adapter for test-time adaptation under covariate and label shifts,” in *IEEE/CVF ICCV*, Paris, France, Oct. 2023. (https://doi.org/10.48550/arXiv.2308.08810)

[5] S. Garg, S. Balakrishnan, and Z. Lipton, “Domain adaptation under open set label shift,” in *NeurIPS*, New York, USA, Nov. 2022.

[6] J. Leevy, T. Khoshgoftaar, R. Bauder, and N. Seliya, “A survey on addressing high-class imbalance in big data,” *J. Big Data*, vol. 5, no. 42, pp. 1-30, Nov. 2018. (https://doi.org/10.1186/s40537-018-0151-6)

[7] D. Samuel, Y. Atzmon, and G. Chechik, “From generalized zero-shot learning to long-tail with class descriptors,” in *IEEE/CVF WACV*, Hawaii, USA, Jan. 2021. (https://doi.org/10.1109/WACV48630.2021.00033)

[8] M. Kim, M. Joe, and M. Kwon, “Improving network attack classification on imbalanced real-world intrusion incident datasets,” in *ACM MobiSys*, California, USA, Jun. 2025. (https://doi.org/10.1145/3711875.3734549)

[9] A. Fernandez, S. Garcia, F. Herrera, and N. Chawla, “SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary,” *J. Artificial Intell. Res.*, vol. 61, no. 1, pp. 863-905, Apr. 2018. (https://doi.org/10.1613/jair.1.11192)

[10] Z. Zhong, J. Cui, S. Liu, and J. Jia, “Improving calibration for long-tailed recognition,” in *IEEE/CVF CVPR*, Tennessee, USA, Jun. 2021. (https://doi.org/10.1109/CVPR46437.2021.01622)

[11] T. Wei, Z. Mao, Z. H. Zhou, Y. Wan, and M.-L. Zhang, “Learning label shift correction for test-agnostic long-tailed recognition,” in *ICML*, Vienna, Austria, Jul. 2024.

[12] M. Joe, M. Kim, and M. Kwon, “Fine-tuning anomaly classifier for unbalanced network data,” *J. KICS*, vol. 49, no. 7, pp. 911-922, Jul. 2024. (https://doi.org/10.7840/kics.2024.49.7.911)

[13] L. Alzubaidi, et al., “A survey on deep learning tools dealing with data scarcity: Definitions, challenges, solutions, tips, and applications,” *J. Big Data*, vol. 10, no. 46, Apr. 2023. (https://doi.org/10.1186/s40537-023-00727-2)

[14] Z. Ma, Z. Li, Y. Shi, and J. Chen, “FedSum: Data-efficient federated learning under data scarcity scenario for text summarization,” in *AAAI*, Washington, D.C., USA, Apr. 2025. (https://doi.org/10.1609/aaai.v39i18.34129)

[15] J. Kirkpatrick, et al., “Overcoming catastrophic forgetting in neural networks,” *National Academy Sci.*, vol. 114, no. 13, pp. 3521-3526, Mar. 2017. (https://doi.org/10.1073/pnas.1611835114)

[16] R. A. Bafghi, N. Harilal, C. Monteleoni, and M. Raissi, “Parameter efficient fine-tuning of self-supervised ViTs without catastrophic forgetting,” in *IEEE/CVF CVPR Wkshps.*, Washington State, USA, Jun. 2024. (https://doi.org/10.1109/CVPRW63382.2024.00371)

[17] G. Legate, N. Bernier, L. Page-Caccia, E. Oyallon, and E. Belilovsky, "Guiding the last layer in federated learning with pre-trained models," in *NeurIPS*, Los Angeles, USA, Dec. 2023.

[18] G. Sun, U. Khalid, M. Mendieta, P. Wang, and C. Chen, "Exploring parameter-efficient fine-tuning to enable foundation models in federated learning," in *IEEE Big Data*, Washington, D.C., USA, Dec. 2024. (https://doi.org/10.1109/BigData62323.2024.10825712)

[19] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *AISTATS*, Florida, USA, Apr. 2017.

[20] S. Luo, Y. Xiao, X. Zhang, Y. Liu, W. Ding, and L. Song, "PerFedRec++: Enhancing personalized federated recommendation with self-supervised pre-training," *ACM Trans. Intell. Syst. and Technol.*, vol. 15, no. 5, Nov. 2024. (https://doi.org/10.1145/3664927)

[21] Y. Tan, G. Long, J. Ma, L. Liu, T. Zhou, and J. Jiang, "Federated learning from pre-trained models: A contrastive learning approach," in *NeurIPS*, Los Angeles, USA, Nov. 2022.

[22] H. Park, M. Kim, and M. Kwon, "Robust partial share federated learning algorithm against model poisoning attack," *J. KICS*, vol. 48, no. 11, pp. 1387-1398, Nov. 2023. (https://doi.org/10.7840/kics.2023.48.11.1387)

[23] J. Oh, S. Kim, and S.-Y. Yun, "FedBABU: Towards enhanced representation for federated image classification," in *ICLR*, Virtual, Jan. 2022.

[24] X. Yang, Y. Wang, L. Zhang, and S. Zhou, "FedAS: Bridging inconsistency in personalized federated learning," in *IEEE/CVF CVPR*, Washington State, USA, Jun. 2024. (https://doi.org/10.1109/CVPR52733.2024.01139)

[25] H. Kye and M. Kwon, "Partial federated learning based network intrusion system for mobile devices," in *ACM MobiHoc*, New York, USA, Oct. 2022. (https://doi.org/10.1145/3492866.3561257)

[26] H. Park, M. Kim, and M. Kwon, "Personalized federated sensing for heterogeneous environment," in *IEEE Sensors Lett.*, vol. 9, no. 4, pp. 1-4, Apr. 2025. (https://doi.org/10.1109/LSENS.2024.3464518)

[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. (https://doi.org/10.1109/COMPTELIX.2017.8004032)

[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*, Texas, USA, Oct. 2022. (https://doi.org/10.1109/CNS56114.2022.9947235)

[29] A. Krizhevsky, "Learning multiple layers of features from tiny images," University of Toronto, Vancouver, Canada, Dec. 2024.

[30] J. Luo, F. Hong, J. Yao, B. Han, Y. Zhang, and Y. Wang, "Revive re-weighting in imbalanced learning by density ratio estimation," in *NeurIPS*, Vancouver, Canada, Dec. 2024.

[31] F. Du, P. Yang, Q. Jia, F. Nan, X. Chen, and Y. Yang, "Global and local mixture consistency cumulative learning for long-tailed visual recognitions," in *IEEE/CVF CVPR*, British Columbia, Canada, Mar. 2023. (https://doi.org/10.1109/CVPR52729.2023.01518)

[32] Y. Cui, M. Jia, T. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *IEEE/CVF CVPR*, California, USA, Jun. 2019. (https://doi.org/10.1109/CVPR.2019.00949)

임 경 진 (Kyungjin Im)

2022년 3월~현재 : 숭실대학교 AI융합학부

<관심분야> 인공지능, 연합학습, 모바일 네트워크

[ORCID:0009-0007-0557-5371]

박 희 원 (Heewon Park)

2024년 2월 : 숭실대학교 전자정보공학부 IT융합전공

2024년 3월~현재 : 숭실대학교 지능형반도체학과 석사과정

<관심분야> 인공지능, 연합학습, 모바일 네트워크

[ORCID:0009-0006-0446-3151]

김 미 르 (Miru Kim)

2022년 8월 : 숭실대학교 전자정보공학부 IT융합전공

2022년 9월 : 숭실대학교 지능형반도체학과 석사

2024년 3월~현재 : 숭실대학교 지능형반도체학과 박사

<관심분야> 인공지능, 연합학습, 이상탐지기술

[ORCID:0000-0002-5394-4780]

조 무 곤 (Mugon Joe)

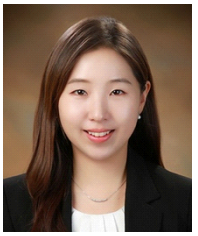
2024년 2월 : 숭실대학교 전자정보공학부 IT융합전공

2024년 3월~현재 : 숭실대학교 지능형반도체학과 석사과정

<관심분야> 인공지능, 모바일 네트워크, 이상탐지 기술

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

권 민 혜 (Minhae Kwon)



2011년 8월 : 이화여자대학교 전자정보통신공학과 학사

2013년 8월 : 이화여자대학교 전자공학과 석사

2017년 8월 : 이화여자대학교 전자전기공학과 박사

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

2018년 9월~2020년 2월 : 미국 Rice University, Electrical and Computer Engineering, Postdoctoral Researcher

2020년 3월~현재 : 숭실대학교 전자정보공학부 및 지능형반도체학과 부교수

<관심분야> 모바일 네트워크, 이상탐지기술, 인공지능, 강화학습, 자율주행

[ORCID:0000-0002-8807-3719]