연합학습에 대한 최신 연구동향

정택준§, 임동건§, 김주령§, 방인규†■, 김태훈§■

§국립한밭대학교 컴퓨터공학과, †국립한밭대학교 지능미디어공학과

{20211929, 20211893, 20222021}@edu.hanbat.ac.kr, {ikbang, thkim}@hanbat.ac.kr

Recent Research Trends in Federated Learning

Taekjun Jeong§, Donggeon Im§, Juryeong Kim§, Inkyu Bang†■, Taehoon Kim§■

§Department of Computer Engineering, Hanbat National University

†Department of Intelligence Media Engineering, Hanbat National University

요 약

연합학습은 분산 환경에서 프라이버시를 보호하며 모델을 학습하는 기술이다. 본 논문은 FedAvg 알고리즘의 동작 원리와 Non-IID 데이터 및 보안 취약점이라는 두 가지 핵심 문제를 분석한다. CGAN 기반 데이터 증강, 차등 프라이버시, 안전한 집계 등 최신 해결 기법들을 소개하고, 각 기법의 트레이드오프를 고려한 선택이 중요함을 제시한다.

Ⅰ. 서 론

최근 데이터 볼륨이 크게 증가하면서 중앙 집중식 데이터 저장 및 분석이 비실용적이며, 민감한 개인 정보 보호 문제도 발생하고 있다. 이에 따라 연합학습(Federated Learning, FL)이 주목받고 있다. 연합학습은 프라이버시 민감하거나 대용량의 분산된 데이터로부터 딥러닝 모델을 효율적으로 학습시키는 접근방식이다. 모델 훈련을 원시 훈련 데이터에 대한 직접적인 접근의 필요성으로부터 분리함으로써 상당한 프라이버시 이점을 제공한다.

연합학습의 대표적인 알고리즘인 FedAvg(Federated Averaging)는 각 클라이언트에서 로컬 확률적 경사 하강법을 수행하고, 서버에서 이들의 모델을 평균화하는 방식으로 동작한다. 그러나 실제 환경에서는 클라이언트 간 데이터 분포가 다른 Non-IID 문제와 그래디언트 유출 공격과 같은 보안 취약점이 존재한다. 본 논문에서는 연합학습의 동작 원리를 설명하고, 주요 취약점과 이를 해결하기 위한 최신 연구 동향을 살펴본다.

Ⅱ. 연합학습 동작 원리 및 문제

연합학습의 대표적인 알고리즘으로 FedAvg(Federated Averaging) 알고리즘이 있다[1]. 그림 1은 FedAvg의 기본 동작 원리를 보여준다. 중앙 서버는 초기 글로벌 모델을 초기화한 후, 각 라운드마다 선택된 클라이언트들에게 현재 글로벌 모델을 전송한다(①). 각 클라이언트는 로컬 데이터를 mini-batch로 분할하여 E 에포크 동안 SGD를 반복 수행한다(②). 학습이 완료된 로컬 모델 파라미터는 서버로 업로드되고(③), 서버는 각 클라이언트의 데이터 수에 비례한 가중 평균으로 이를 집계하여(④) 새로운 글로벌 모델을 생성한다. 이 과정을 반복하며 모델을 개선한다.

그러나 연합학습에도 취약점들이 있다. 첫째, Non-IID 데이터 문제이다. 실제 환경에서 각 클라이언트의 데이터는 서로 다른 분포를 가지는 Non-IID(Non-Independent and Identically Distributed) 특성을 보인다. 예를 들어 sEMG 신호 분석, 의료 영상, 모바일 키보드 입력 등의 응용 분야에서 사용자마다 데이터 패턴이 상이하며, 이는 모델 수렴 속도 저하와 정확도 감소를 야기한다. 둘째, 보안 및 프라이버시 위협이다. 연합학습은 원시 데이터를 공유하지 않아 프라이버시를 보호한다는 장점이 있으나, 클라이언트가 전송하는 그래디언트 정보로부터 원본 훈련 데이터를 복원하는 그래디언트 유출 공격(Gradient Leakeage Attack)이 가능하다[4]. 이는 연합학습의 근본적인 목적인 프라이버시 보장을 위협한다.

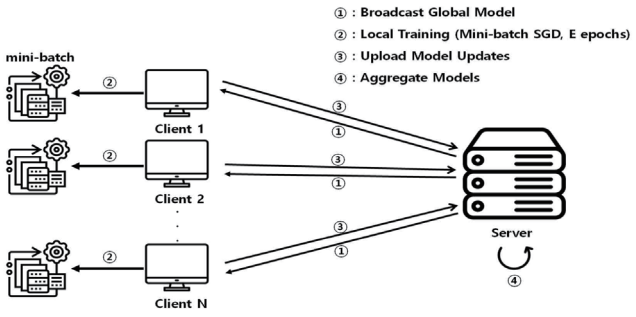

그림 1 FedAvg 알고리즘의 동작 구조

Ⅲ. 취약점 문제 해결 동향

Non-IID 문제를 해결하기 위한 대표적인 접근법으로 데이터 증강 기법이 있다. CGAN(Conditional Generative Adversarial Network) 기반 방법은 각 클라이언트에서 부족한 클래스의 합성 데이터를 생성하여 데이터 분포의 불균형을 완화한다[2]. 이 기법은 레이블 정보를 조건으로 활용하여 특정 클래스의 데이터를 선택

■ Corresponding Authors: Inkyu Bang (ikbang@hanbat.ac.kr), Taehoon Kim (thkim@hanbat.ac.kr)

적으로 생성할 수 있어, 클라이언트 간 데이터 분포 차이를 효과적으로 줄일 수 있다. 실제 응용 분야에서도 Non-IID 문제 해결 연구가 활발하다. FedAssist는 AI 기반 의수족을 위한 sEMG 신호 디코딩에 연합학습을 적용한 프레임워크로[3], 데이터 소유권을 보호하면서도 분산된 협력적 모델링을 가능하게 한다. 특히 클라이언트 선택 전략과 지역적 정규화 기법을 통해 sEMG 데이터의 Non-IID 특성을 완화하고 모델 정확도를 향상시켰다. 그림 2는 FedAssist의 성능을 다른 연합학습 알고리즘들과 비교한 결과를 보여주고 있다. 세 가지 Non-IID 시나리오(이질적 레이블 분포, 이질적 특징 분포, 이질적 데이터 볼륨)에서 FedAssist(빨간 점선)는 FedAvg, FedProx, FedMix 등 기존 연합학습 기법들보다 일관되게 높은 정확도와 빠른 수렴 속도를 달성했다. 특히 Non-IID Case 1과 Case 3에서는 다른 모델들이 학습 정체(plateau) 현상을 보이는 반면, FedAssist는 글로벌 라운드가 증가함에 따라 지속적으로 성능이 향상되었다. 이는 로컬 레벨 데이터 보강과 글로벌 레벨 모델 미세조정 전략이 Non-IID 환경에서 효과적임을 입증한다.

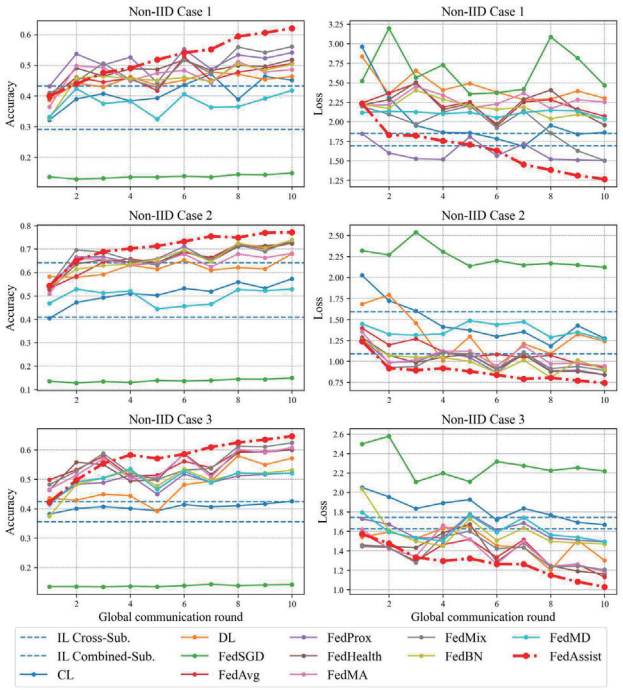

그림 2 세 가지 Non-IID 시나리오에서 FedAssist와 기존 연합학습 알고리즘들의 성능 비교 [3].

그래디언트 유출 공격은 연합학습의 핵심 위협으로, 전송되는 그래디언트 정보로부터 민감한 훈련 데이터를 복원할 수 있다. 이에 대응하기 위해 다양한 방어 기법이 제안되었다[4]. 차등 프라이버시(Differential Privacy)는 그래디언트에 통계적 노이즈를 추가하여 개별 데이터 포인트의 정보를 마스킹한다. 노이즈 수준을 조절하여 프라이버시 보호 강도를 제어할 수 있으나, 과도한 노이즈는 모델 성능 저하를 야기한다. 안전한 집계(Secure Aggregation)는 암호화 프로토콜을 활용하여 서버가 개별 클라이언트의 그래디언트를 직접 볼 수 없도록 한다. 서버는 오직 집계된 결과만 복호화할 수 있어 강력한 프라이버시 보호를 제공하지만, 암호화 연산으로 인한 계산 오버헤드가 발생한다. 그래디언트 압축은 전송되는 그래디언트의 정보량을 줄여 공격 표면을 감소시킨다. Top-k 선택이나 양자화 기법을 통해 그래디언트를 희소화하거나 저정밀도로 표현하며, 통신 효율성 향상과 보안 강화를 동시에 달성할 수 있다. 이러한 기법들은 보안성, 모델 성능, 계산 비용 간의 트레이드오프를 고려하여 응용 환경에 따라 적절히 선택되어야 한다.

Ⅳ. 결론

본 논문에서는 연합학습의 핵심 알고리즘인 FedAvg의 동작 원리와 함께 Non-IID 데이터 문제 및 보안 취약점이라는 두 가지 주요 과제를 분석하였다. Non-IID 문제 해결을 위해 CGAN 기반 데이터 증강과 FedAssist와 같은 응용 프레임워크가 제안되었으며, 보안 강화를 위해서는 차등 프라이버시, 안전한 집계, 그래디언트 압축 등의 방어 기법이 연구되고 있다.

각 해결 기법은 프라이버시 보호, 모델 성능, 계산 비용 간의 트레이드오프를 수반한다. 향후 연구는 이러한 기법들을 통합하여 Non-IID 환경에서도 높은 보안성과 성능을 동시에 달성하는 방향으로 진행될 필요가 있다. 또한, 의료, IoT, 모바일 기기 등 다양한 실제 응용 분야에서의 검증과 최적화가 요구된다. 연합학습은 프라이버시 보호와 효율적인 분산 학습을 동시에 달성할 수 있는 유망한 기술이며, 지속적인 연구를 통해 더욱 실용적인 시스템으로 발전할 것으로 기대된다.

ACKNOWLEDGMENT

본 연구는 2025년 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구결과 (2022-0-01068) 및 2025년도 교육부 및 대전광역시의 재원으로 대전RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)의 결과임 .(2025-RISE-06-002)

참 고 문 헌

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Aguera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artif. Intell. Statist. (AISTATS), 2017.

[2] 장현진, 김요한, "연합학습의 Non-IID 문제 해결을 위한 CGAN 기반 데이터 증강 기법," 한국전자통신학회 논문지(Journal of the Korea Institute of Electronic Communication Sciences), vol. 20, no. 1, pp. 105-112, 2025.

[3] H. Lee, M. Jiang, and Q. Zhao, "FedAssist: Federated learning in AI-powered prosthetics for sustainable and collaborative learning," in 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2024, doi: 10.1109/EMBC53108.2024.10781961.

[4] W. Wei and L. Liu, "Gradient leakage attack resilient deep learning," IEEE Transactions on Information Forensics and Security, vol. 17, pp. 303-316, 2022.