연합학습 Non-IID 대응을 위한 잠재 공간 전이 기반 데이터 보강 프레임워크 설계

강태욱, 박지우, 조용윤, 박철영, 신창선*

*순천대학교,

1250070@s.scnu.ac.kr, 1250181@s.scnu.ac.kr, yycho@scnu.ac.kr, cypark@scnu.ac.kr, *csshin@scnu.ac.kr

Design of Latent Space Transfer-based Data Augmentation Framework for Non-IID Federated Learning

Kang Tae Wook, Park Ji Woo, Cho, Young Yoon, Park Chul Young ,Shin Chang Cun*

*Sunchon Univ.

요 약

연합학습(Federated Learning, FL)은 데이터 프라이버시를 보호하며 모델을 공동 학습할 수 있는 유망한 기술이지만, 클라이언트 간 데이터 분포 불균형(Non-IID) 문제로 인해 모델 성능이 저하되는 한계가 있다. 기존 연구에서는 HMM과 Diffusion 모델을 융합하여 시나리오를 생성하는 방안을 제시하였으나, 로컬 클라이언트의 특정 클래스 결여 문제를 직접적으로 보강하기 위한 통신 효율성 및 프라이버시 보전 방안은 여전히 과제로 남아 있다. 본 논문에서는 클래스가 풍부한 클라이언트의 잠재 공간(Latent Space) 정보를 선별적으로 전이하여 데이터를 보강하는 프레임워크를 제안한다. 제안 시스템은 사전 학습된 글로벌 디코더를 활용하여 잠재 벡터로부터 데이터를 복원하며, 참조 빈도 기반의 적응형 노이즈 주입 및 ADR/DSR 보안 지표를 통해 데이터의 유용성과 보안성을 동시에 확보한다.

Ⅰ. 서 론

연합학습(Federated Learning, FL)은 개별 클라이언트의 로컬 데이터를 외부로 노출하지 않고 모델의 가중치만을 공유하여 프라이버시를 보호하는 분산 학습 기법이다. 그러나 실제 환경에서 클라이언트별 데이터 분포가 상이한 Non-IID(Non-Identically and Independently Distributed) 문제는 모델의 안정성을 저하하고 통신 비용의 증대를 초래하는 주요 병목 현상으로 작용한다[1].

저자들은 선행 연구에서 이러한 문제를 완화하기 위해 은닉 마르코프 모델(HMM)과 확산 모델(Diffusion Model)을 결합한 이미지 증강 기법을 제안한 바 있다. 특히 대규모 자율주행 데이터셋인 BDD100K를 기반으로 설계된 계층적 HMM 아키텍처는 시나리오 내 문맥적 요인(날씨, 시간, 장소)의 흐름과 객체 분포를 체계적으로 모델링하여, 현실적인 주행 시나리오 생성의 기반을 마련하였다[2].

본 연구는 선행 연구를 통해 확보된 고품질 시나리오 생성 엔진을 연합학습 시스템 전반으로 확장하여, 클라이언트 간 '잠재 공간(Latent Space)' 정보를 안전하게 공유하고 부족한 데이터 클래스를 보강하는 통합 프레임워크를 제안한다. 이는 개별 클라이언트의 생성 모델을 협력적으로 활용하는 동시에 공유 과정에서의 보안성을 확보하는 데 중점을 둔다.

Ⅱ. 본론

본 절에서는 잠재 공간 전송을 통한 데이터 보강 프로세스와 이를 안전하게 관리하기 위한 보안 공유 알고리즘의 세부 설계를 기술한다.

2.1 잠재 공간 기반 협력형 데이터 보강 프로세스

특정 클래스(예: 야간 주행 등) 데이터가 풍부한 소스 클라이언트는 선행 연구의 HMM-Diffusion 파이프라인을 통해 해당 클래스의 특징을 저차원 잠재 벡터(z)로 추출한다. 잠재 공간 정보는 직접적인 픽셀 데이터를 포함하지 않아 통신 오버헤드가 낮고 일차적인 프라이버시 보호가 가능하다[3]. 본 프레임워크에서는 글로벌 서버가 초기 라운드에서 공공 데이터셋으로 사전 학습된 글로벌 디코더(Global Decoder)를 모든 클라이언트에 배포한다. 타겟 클라이언트는 서버로부터 수신한 타 클라이언트의 잠재 벡터를 이 디코더에 입력하여 로컬 환경에 필요한 가상 데이터를 생성하고, 이를 기존 데이터셋과 병합하여 학습의 다양성을 확보한다.

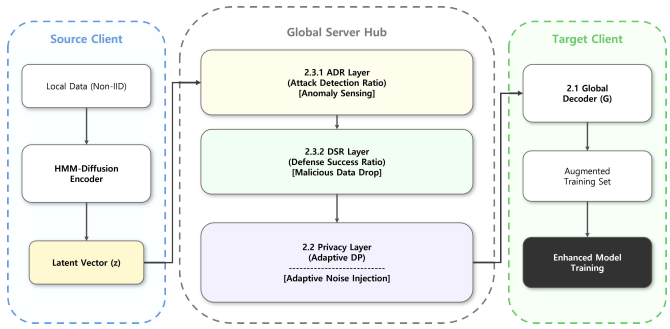

그림 1.제안하는 잠재 공간 전이 기반 연합학습 데이터 보강 프레임워크의 전체 구성도

2.2 참조 기반 적응형 노이즈 주입 알고리즘

데이터 공유 과정에서 특정 클라이언트의 민감 정보가 역추적되는 것을 방지하기 위해 차등 개인정보 보호(Differential Privacy) 기법을 적용한다. 특히, 기존의 중요도 및 상태 기반 적응형 DP 메커니즘을 확장하여 글로벌 서버가 개별 잠재 공간 정보의 활용 빈도를 실시간으로 모니터링하여 노이즈의 강도를 동적으로 제어하는 적응형 노이즈 주입 메커니즘을 제안한다[4-5].

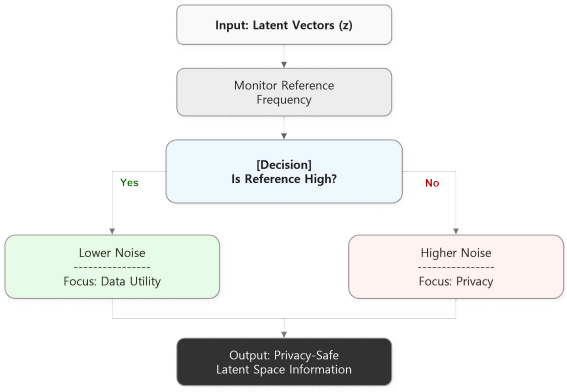

그림 2. 참조 기반 적응형 노이즈 주입 (Adaptive DP)

많은 클라이언트가 공통으로 필요로 하는 범주적 특징(예: 일반적인 도로 형태 등)을 지닌 고빈도 참조 데이터의 경우, 생성되는 보강 데이터의 유용성(Utility)을 보존하기 위해 노이즈 강도를 상대적으로 하향 조정하여 생성 이미지의 실사성과 품질을 유지한다. 반면, 특정 클라이언트에만 편중되어 나타나는 특이 데이터는 역추적을 통한 프라이버시 노출 위험이 높으므로 노이즈 강도를 상향 조정하여 보안성을 강화한다. 이러한 접근은 보안성 제고뿐만 아니라 잠재 공간 내에 의도적인 변동성을 부여함으로써 데이터의 다양성(Diversity)을 확보하고, 특정 데이터군에 대한 글로벌 모델의 편향성을 완화하는 보수적인 효과를 동시에 거둔다.

2.3 ADR 및 DSR 지표 기반의 보안성 검증 체계

잠재 공간 정보를 공유하는 과정에서 발생할 수 있는 보안 위협에 대응하고 시스템의 신뢰성을 보장하기 위해 본 연구에서는 ADR(Attack Detection Ratio) 및 DSR(Defense Success Ratio) 지표를 활용한 보안 검증 체계를 통합한다.

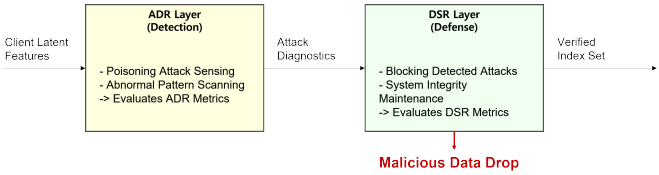

그림 3. ADR 및 DSR 기반 보안 검증 프레임워크

먼저 ADR 지표 산출을 위해 글로벌 서버는 수신된 잠재 벡터의 통계적 분포를 정밀하게 분석하며, 기존의 정상적인 글로벌 분포 범위를 벗어난 이상치(Outlier)를 실시간으로 탐지하여 악의적인 조작 시도를 식별한다. 이와 병행하여 DSR 지표를 통해 감지된 공격 벡터를 학습 프로세스에서 완전히 배제하고, 무결성이 엄격히 검증된 클라이언트의 인덱스 집합만을 최종 업데이트에 반영함으로써 글로벌 모델의 강건성을 정량적으로 평가하고 유지한다. 이러한 다중 방어 기제는 악의적인 의도를 가진 클라이언트가 왜곡된 잠재 정보를 주입하여 전체 학습 모델을 오염시키려는 시도를 효과적으로 차단하는 핵심적인 보안 기틀을 제공한다.

Ⅲ. 결론

본 논문에서는 연합학습 환경의 Non-IID 문제를 해결하기 위해 HMM-Diffusion 기반의 잠재 공간 전이 및 보안 공유 프레임워크를 설계하고 그 이론적 타당성을 고찰하였다. 제안된 시스템은 저차원 잠재 벡터 (z) 공유를 통해 통신 효율성을 확보하는 동시에, 참조 빈도 기반의 적응형 노이즈 주입 알고리즘을 설계함으로써 데이터의 유용성과 프라이버시 보호를 동시에 달성할 수 있는 기틀을 마련하였다.본 연구는 구현 및 실험적 검증에 앞서 보안성과 효율성을 극대화한 시스템 아키텍처를 제안하는 데 중점을 두었으며, 특히 ADR 및 DSR 지표를 통한 다중 보안 검증 체계의 설계는 데이터 오염 공격에 대한 시스템의 강건성을 이론적으로 보장한다. 향후 연구에서는 BDD100K와 같은 대규모 주행 데이터셋을 활용하여 본 설계안을 실제 연합학습 시뮬레이션 환경에서 구현하고, 제안한 알고리즘이 기존의 고정형 노이즈 방식 및 데이터 증강 기법 대비 모델의 정확도와 보안 지표 측면에서 거두는 정량적 성과를 입증함으로써 연구의 완결성을 높일 계획이다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2024-00407739).

참 고 문 헌

[1] H. Kim, H. Jeon, E. Choi, W. Jang, S. Jeong, and C. Shim, "Applying Generative Models to Improve Non-IID Problems in Federated Learning," KICS Conf. Proc., pp. 76-77, 2025.

[2] T. Kang, C. Park, C. Shim, S. Jeong, and C. Shin, "A Study on HMM-Diffusion based Image Augmentation for Solving Non-IID Problems in Federated Learning," Proc. KIIT Fall Conf., p. 1, 2025.

[3] D. M. S. Bhatti and B. J. Choi, "Enhancing IoT Healthcare with Federated Learning and Variational Autoencoder," Sensors, vol. 24, no. 11, p. 3632, Jun. 2024.

[4] W. Zheng, Q. . Zhao, and H. . Xie, "Research on Adaptive Noise Mechanism for Differential Privacy Optimization in Federated Learning", J. Knowl. Learn. Sci. Technol., vol. 3, no. 4, pp. 383 - 392, Dec. 2024.

[5] M. Gong, K. Pan, Y. Xie, A. K. Qin, and Z. Tang, "Preserving differential privacy in deep neural networks with relevance-based adaptive noise imposition," Neural Networks, vol. 125, pp. 131 - 141, May 2020.