Federated Learning 관련 Research Question

Federated Learning에서,

안전하게 데이터 삭제를 할 수 없을까?

Federated Unlearning 관련 논문

Federated Learning의 한계: 데이터 삭제를 안전하게 할 수 없을까?

1. **FedRecovery: Differentially Private Machine Unlearning for Federated Learning Frameworks**
   - 출판: IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 18, 2023

2. **Guaranteeing Data Privacy in Federated Unlearning With Dynamic User Participation**
   - 출판: IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 22, NO. 3, May/June 2025

3. **Model Recovery in Federated Unlearning With Restricted Server Data Resources**
   - 출판 : IEEE INTERNET OF THINGS JOURNAL, VOL. 12, NO. 11, 1 June 2025

Federated Unlearning 개념의 최초 등장

2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS)

FedEraser: Enabling Efficient Client-Level Data Removal from Federated Learning Models

Gaoyang Liu[1], Xiaoqiang Ma[1], Yang Yang[2], Chen Wang[1], Jiangchuan Liu[3]
[1]Huazhong University of Science and Technology, Wuhan, China
[2]Hubei University, Wuhan, China
[3]Simon Fraser University, British Columbia, Canada
[1]{liugaoyang, maxiaoqiang, chenwang}@hust.edu.cn, [2]yangyang@hubu.edu.cn, [3]jcliu@cs.sfu.ca

[Federated Unlearning 개념과 배경]

- 분산 학습 환경에서의 안전한 데이터 삭제

PMLC Lab meeting

2025.08.13 (Wed)

목차

1. Machine Unlearning의 배경

1. Machine Unlearning의 배경

'Right to Be Forgotten'
잊혀질 권리

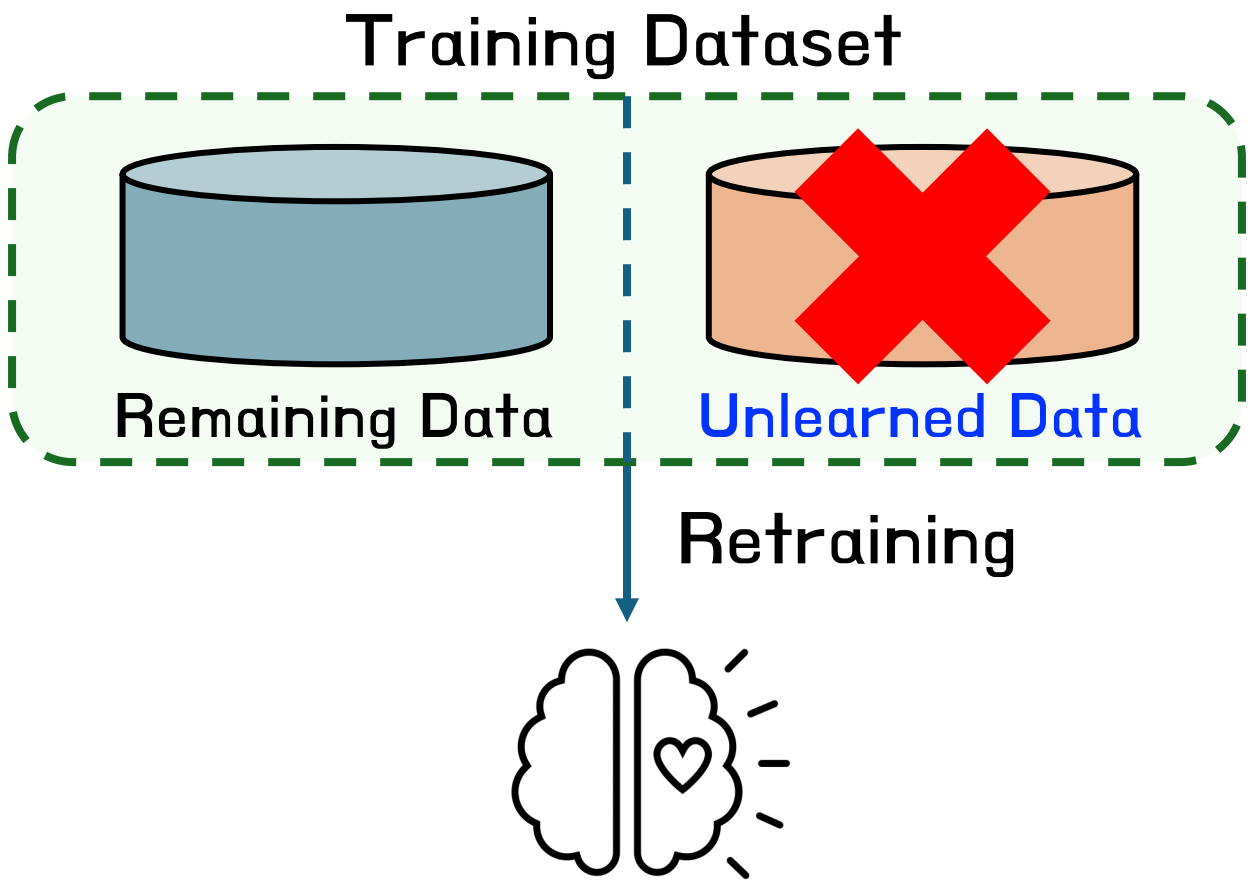
- **최근 10년 이내 전 세계적으로 관련 법률 제정했거나 제정 중**
  - o GDPR (EU), CCPA (USA), APPI (일본), CPPA (캐나다)

개인 및 조직은 자신의 데이터를 삭제 요청할 권리가 있다 !

→ **Privacy 관점으론** 😊, **ML 모델 관점으론** 😭
  - o 데이터셋에서 타겟 데이터만 삭제
  - o 이미 학습된 모델에서 타겟 데이터의 영향 제거

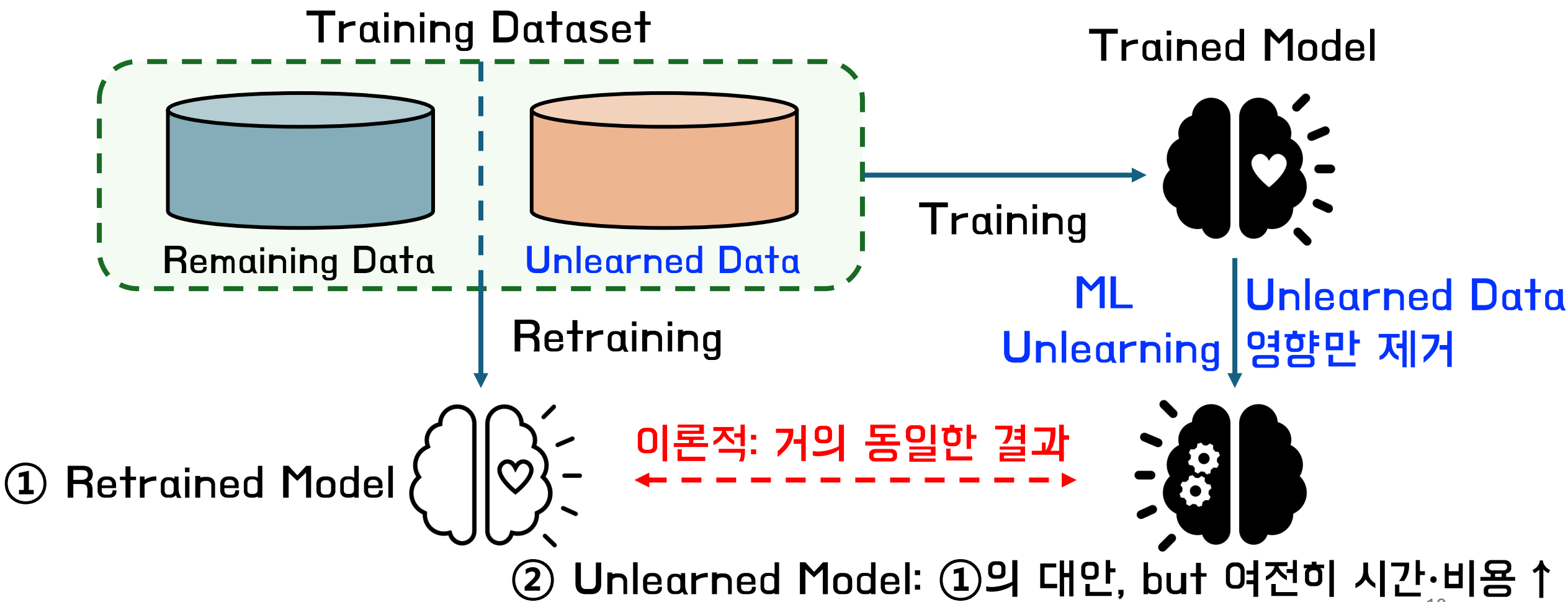
2. Machine Unlearning의 분류

2. Machine Unlearning의 분류 - Naïve Model (1/3)



① Retrained Model: 가장 확실, but 시간·비용 ↑

2. Machine Unlearning 분류의 - Unlearned Model (2/3)



2. Machine Unlearning의 분류 (3/3)

- Strong Unlearning : ① Retrained Model ② Unlearned Model처럼
  삭제 요청된 데이터가 학습에 전혀 사용되지 않은 것과 동일한 상태를 목표로 함
  * 단, 실제 적용 난이도가 높음

[Approximate Machine Unlearning]

- Weak Unlearning : 모델 출력 (예: Accuracy)이 유사하도록 조정하는 것을 목표
  ① 특정 데이터의 영향을 역으로 적용
  ② 파라미터 일부 조정 및 성능 저하 최소화를 위한 튜닝
  * 단, 완전한 데이터 삭제 보장은 불가

3. Unlearning 알고리즘 설계 시 필수 고려사항

3. Unlearning 알고리즘 설계 시 필수 고려사항

- **Consistency**: Unlearned Model과 Retrained Model의 유사성↑
- **Accuracy**: 삭제 이후에도 Test Data에서 높은 정확도 유지
- **Verifiability**: Unlearning이 제대로 수행됐는지 검증 가능 필요
- **Efficiency**: 계산·통신 비용 최소화
- **Provable Guarantees**: 수학·이론적으로 삭제의 완전성 보장 필요
- **Privacy Preservation: 새로운 정보 유출 야기 방지**

4. Federated Unlearning시 고려해야 할 주요 과제

4.1 Federated Learning의 특징

[FL 기본 구조 (분산 학습)]

- 중앙서버: 모델 관리 및 최신성 유지
- 클라이언트: 로컬 데이터 관리,
  정기적/비정기적인 모델 업데이트 참여

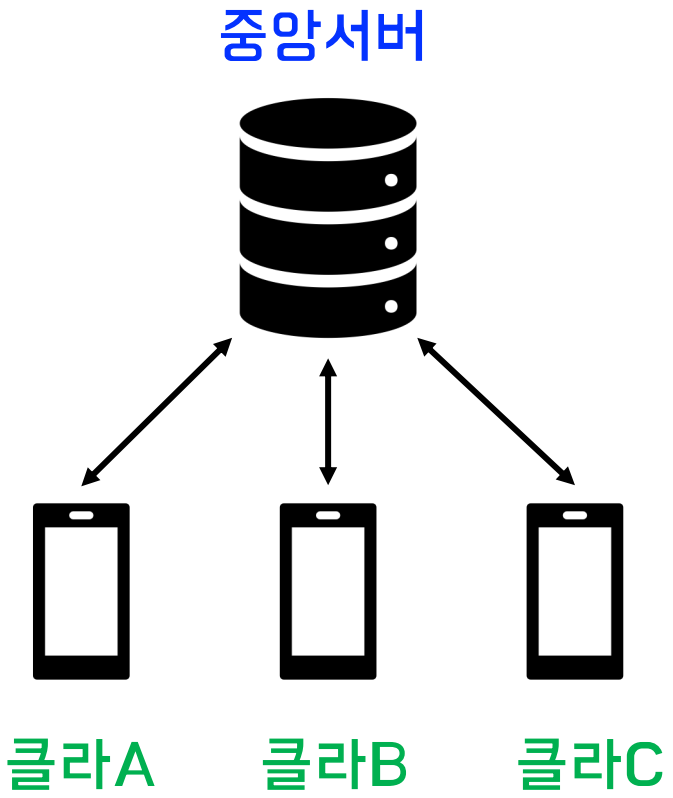
[시나리오]

클라이언트 A

: *제 데이터 삭제 부탁드립니다.*

→ 서버, 다른 클라이언트들: ??? 😭

→ Federated Unlearning 필요성



4.2 Federated Unlearning시 고려해야 할 주요 과제

- **Knowledge Permeation**: 업데이트 영향이 모든 모델에 확산
- **Data Isolation**: 데이터는 클라에서 직접 관리하여 접근 혹은 영향 측정 어려움
- **Who Unlearn**: 어떤 주체(클라·서버)가 실행·관리 할 것인지 선택의 어려움
- **Unlearn What**: 삭제 대상 범위 정하기 어려움 (범위에 따라 난이도·비용 달라짐)
- **Constrained Resources**: 연산량 많은 Unlearning 알고리즘 적용의 한계
- **Participants Heterogeneity**: 클라마다 하드웨어 스펙, 품질 등이 다름
- **Client Dynamics**: 클라가 학습에 참여했다가 빠졌다 하는 동적 환경의 한계

4.2 Federated Unlearning시 고려해야 할 주요 과제

- **Knowledge Permeation**: 업데이트 영향이 모든 모델에 확산
- **Data Isolation**: 데이터는 클라에서 직접 관리하여 접근 혹은 영향 측정 어려움
- **Who Unlearn**: 어떤 주체(클라·서버)가 실행·관리 할 것인지 선택의 어려움
- **Unlearn What**: 삭제 대상 범위 정하기 어려움 (범위에 따라 난이도·비용 달라짐)
- **Constrained Resources**: 연산량 많은 Unlearning 알고리즘 적용의 한계
- **Participants Heterogeneity**: 클라마다 하드웨어 스펙, 품질 등이 다름
- **Client Dynamics**: 클라가 학습에 참여했다가 빠졌다 하는 동적 환경의 한계

4.2 Federated Unlearning시 고려해야 할 주요 과제

- **Knowledge Permeation**: 업데이트 영향이 모든 모델에 확산
- **Data Isolation**: 데이터는 클라에서 직접 관리하여 접근 혹은 영향 측정 어려움
- **Who Unlearn**: 어떤 주체(클라·서버)가 실행·관리 할 것인지 선택의 어려움
- **Unlearn What**: 삭제 대상 범위 정하기 어려움 (범위에 따라 난이도·비용 달라짐)
- **Constrained Resources**: 연산량 많은 Unlearning 알고리즘 적용의 한계
- **Participants Heterogeneity**: 클라마다 하드웨어 스펙, 품질 등이 다름
- **Client Dynamics**: 클라가 학습에 참여했다가 빠졌다 하는 동적 환경의 한계

5. Federated Unlearning 접근 방식

5.1 Federated Unlearning 접근 방식

1. Passive Unlearning : 서버만 Unlearning을 수행

- Server-standalone Unlearning

: 저장된 과거 데이터(gradient, 기여도 정보 등)에 의존

* 단, 서버의 저장 공간 요구량 큼

- Client-aided Unlearning

: 각 라운드에서 수집한 과거 gradient만으로 재구성

2. Active Unlearning : 서버·클라이언트 함께 Unlearning 수행

- 데이터 일부 혹은 클라이언트 자체 삭제

* 단, 클라이언트 참여 부담과 통신량 증가

5.2 Federated Unlearning 지금까지 진행된 연구 한계

1. Proof of Unlearning
   - 특정 데이터 영향이 모델에서 제거 됐는지 검증 및 증명 방법 부족
2. 클라이언트 동적 참여 모델링
   - FL 환경에서 클라이언트가 지속적으로 참여·이탈하는 데 관련 설계 필요
3. Fairness and Explainability
   - Unlearning 모델의 공정성에 미치는 영향 분석
4. **Privacy Threat**
   - Unlearning 과정에서 데이터 및 모델 업데이트 정보가 새로 노출될 가능성↑
     예) 삭제 요청자의 데이터 패턴이 역으로 유추되는 위험 등

6. 다음주 리뷰할 논문

6. 다음주 리뷰할 논문

FedRecovery: Differentially Private Machine Unlearning for Federated Learning Frameworks

Lefeng Zhang, Tianqing Zhu, Haibin Zhang, Ping Xiong, and Wanlei Zhou, *Senior Member, IEEE*

7. 참고 자료

7. 참고 자료

PEPR '24 – Learning and Unlearning Your Data in Federated Settings (USENIX)