Asia-pacific Journal of Convergent Research Interchange
Vol.9, No.12, December 31 (2023), pp.11-19
http://dx.doi.org/10.47116/apjcri.2023.12.02

Efficient and Secure Multi-Server Based Hierarchical Clustering Federated Learning on Non-IID Data

다중 서버 환경에서 Non-IID 데이터를 위한 효율적이며 안전한 계층적 클러스터링 기반 연합학습

Min Seob Lee[1], Ik Rae Jeong[2], Ji Young Chun[3]

이민섭[1], 정익래[2], 천지영[3]

[1] Student, Graduate School of Information Security, Korea University, South Korea, ms_lee@korea.ac.kr

[2] Professor, Graduate School of Information Security, Korea University, South Korea, irjeong@korea.ac.kr

[3] Professor, Department of Big Data and Information Security, Seoul Cyber University, South Korea, jychun@korea.ac.kr

Corresponding author: Ji Young Chun

Abstract: Federated learning emerges as a groundbreaking and innovative distributed learning paradigm, distinctively characterized by its approach that avoids the direct exchange of data between a central server, which plays a pivotal role, and a myriad of client devices scattered across different locations. In this unparalleled methodology, each individual client takes the initiative to train their respective model using their local data. Once this training phase reaches its conclusion, these clients dispatch their model updates to the central server, ensuring a collaborative learning process. This federated approach becomes especially crucial and salient in specific environments, notably the medical sector. In such domains, data frequently deviates from the standard assumption of being independently and identically distributed, presenting unique challenges. Consequently, the academic and industrial worlds are witnessing a burgeoning interest in delving deep into diverse techniques. These techniques range from clustering and personalization to the more avant-garde concept of meta-learning. It's worth noting that a substantial chunk of the existing body of research firmly anchors itself in a framework that places an implicit, and sometimes unwarranted, trust in the central server. Such an over-reliance inadvertently opens the door to potential security vulnerabilities and introduces efficiency bottlenecks, primarily due to the overwhelming server-centric operations. Based on these issues, this study proposes a new federated learning strategy that reduces the load on the central server while leveraging the strengths of clustering methodologies.

Keywords: Federated Learning, Non-IID, Hierarchical Clustering, Proxy Signature

요약: 연합학습은 중앙 서버와 다양한 위치에 분포되어 있는 수많은 클라이언트 장치 간의 데이터를 직접 교환하지 않는 새로운 분산 학습 방법으로 부각되고 있다. 이러한 방법에서 각 개별 클라이언트는 자신의 로컬 데이터를 활용하여 모델을 독립적으로 훈련시킨다. 훈련 단계가 완료되면, 클라이언트들은 모델의 업데이트를 중앙 서버로 전송하여 협력적인 학습

Received: August 03, 2023; 1st Review Result: September 06, 2023; 2nd Review Result: October 11, 2023
Accepted: November 25, 2023

과정을 구현한다. 연합학습 방식은 특히 의료 분야와 같이 데이터가 독립적이고 동일하게 분포되지 않는 환경에서 중요한 역할을 한다. 이러한 분야에서는 데이터의 특성이 표준 가정에서 벗어나기 때문에 해결해야 할 문제가 생긴다. 따라서 학계와 산업계는 클러스터링, 개인화부터 혁신적인 메타 학습에 이르기까지 다양한 기술을 깊게 탐구하는데 큰 관심을 보이고 있다. 주목할 만한 점은, 현재의 연구 대부분이 중앙 서버에 대한 묵시적인 신뢰를 기반으로 하고 있어, 보안 취약점과 서버 중심의 연산으로 인한 효율성 문제가 발생할 수 있다는 것이다. 이러한 문제를 바탕으로, 본 연구는 중앙 서버의 부하를 줄이면서 클러스터링 방법론의 강점을 활용하는 새로운 연합 학습 전략을 제안한다.

핵심어: 연합학습, 독립 동일 분포가 아닌 데이터, 계층적 클러스터링, 프록시 서명

1. 서론

최근 IT 기술의 발전으로 인해 다양한 스마트 기기와 IoT 기기가 보급되고 있다. 이러한 스마트폰, 스마트 워치, 스마트 홈 장치 등 다양한 스마트 기기들은 우리의 행동, 위치, 건강, 소비 패턴 등 다양한 정보를 기록하고 저장한다. 또한, IoT 기기들은 우리 주변의 환경과 사물들을 연결하여 데이터를 수집하고 분석한다. 이렇게 많은 데이터의 생성과 수집은 인공지능 (AI)의 활용이 필수적이게 되었다. AI 기술은 대량의 데이터를 분석하고 패턴을 학습하여 예측, 분류, 추천 등 다양한 작업을 수행할 수 있다. 이러한 점을 통해 인공지능의 활용이 점점 더 중요해지고 있으며, 이를 통해 우리의 생활과 사회가 더욱 편리하고 지능적으로 발전할 수 있다.

중앙 서버에 데이터를 모아 AI 모델을 학습시키는 방법이 주를 이뤘으나 이러한 방식은 많은 통신 비용과 분석 시간, 그리고 개인정보 유출과 같은 여러 문제점을 갖고 있다. 따라서, 최근 데이터를 중앙 서버에 모으지 않고 여러 제약으로부터 비교적 자유롭고 다양하게 데이터의 활용이 가능하도록 해주는 연합학습(Federated Learning)이 등장했다. 연합학습은 분산되어 있는 독립적인 데이터들을 처리하지만, 전체 데이터를 함께 처리하는 것과 유사한 효과를 보인다. 하지만, 독립 동일 분포가 아닌(Non-IID) 데이터로 인한 단점이 존재한다. Non-IID 데이터는 전체적인 학습 성능 저하와 모델의 일반화 능력의 저하가 발생할 수 있다. 따라서, 이러한 문제를 해결하기 위한 기법의 연구가 필요하다.

2. 관련연구

이번 장에서는 연합학습의 주요 개념, Non-IID 데이터 분포 환경에 대한 개념, 제안하는 기법의 근간이 되는 계층적 클러스터링, 프록시 서명에 대해 설명한다.

2.1 Non-IID 환경에서의 연합학습

연합학습의 대표적인 알고리즘 중 하나인 FedAvg는 아래의 알고리즘과 같이 초기화 단계, 참여할 클라이언트를 선택하는 단계, 선택된 클라이언트가 자체 데이터를 사용하여 모델을 로컬로 학습하는 단계, 중앙 서버에 학습된 모델의 가중치를 보내는 단계, 가중치를 평균하여 글로벌 모델을 업데이트 하는 단계로 전체 모델이 수렴할 때까지

위의 과정을 반복한다[1]. 하지만, 위의 FedAvg과 같은 연합학습 기법들을 Non-IID 데이터 환경에 적용할 경우 몇 가지 단점이 발생할 수 있다. 첫번째, 학습 데이터가 제한적이거나 특정한 편향을 보여 성능의 저하가 발생한다. 두번째, 클라이언트 간의 데이터 분포 차이로 인해 중앙 서버가 모델을 통합할 때 일부 클라이언트의 기여가 다른 클라이언트보다 더 크게 반영될 수 있다. 예를 들어, 의료 이미지 분류 모델을 구축하려고 할 때 각 기기나 병원에서 수집된 데이터는 병원의 위치에 따라 하나의 병원 데이터에 특정 지역의 환자 자료만 들어있어 Non-IID한 분포를 갖을 수 있다. Chandran 등의 논문에서는 FedAvg의 단점으로 Non-IID 환경을 제외하고도 분산 설정에서의 최적화, 학습 과정에서의 개인 정보 보호 및 통신 지연과 같은 문제가 있어 이러한 알고리즘 등이 대규모의 제품에 적용되기 어렵다고 설명한다[2]. 이와 같이, 유사성의 차이가 큰 데이터에 적용 가능한 연합학습 기법의 연구가 필요하다.

Central Server:
Initialize w_0
for each round $t = 1,2,\cdots$ **do**
  $m \leftarrow \max(C \cdot K, 1)$
  $S_t \leftarrow$ (random set of m clients)
  for each client $k \in S_t$ **in parallel do**
    $w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$
  $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$
ClientUpdate(k,w): //*Run on client k*
$\mathcal{B} \leftarrow$ (*split* $\mathcal{P}_k$ into batches of size B)
for each local epoch i from 1 to E do
  for batch $b \in \mathcal{B}$ **do**
    $w \leftarrow w - \eta \nabla \ell(w; b)$
  return w to server

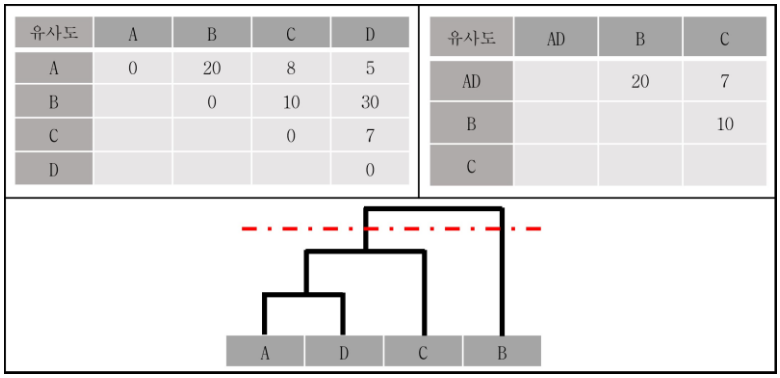
[그림 1] FedAvg 알고리즘

[Fig. 1] FedAvg Algorithm

2.2 계층적 클러스터링(Hierarchical Clustering)

계층적 클러스터링(Hierarchical Clustering)은 비지도 학습 기법 중 하나로, 데이터를 계층적인 구조로 분류하는 알고리즘이다. 데이터 포인트들을 유사도에 따라 그룹화하여 클러스터를 형성하며, 이러한 과정을 통해 클러스터 간의 계층 구조를 만들어낸다. 일반적으로, 계층적 클러스터링은 두가지 유형을 갖는다. 첫번째로, 병합 계층적 클러스터링 기법은 개별 데이터 포인트를 하나의 클러스터를 시작하여 유사한 클러스터를 순차적으로 병합해가는 방식이다[3]. 초기에는 각 데이터 포인트가 독립된 클러스터로 간주되며, 유사도 측정 기준에 따라 가장 유사한 클러스터를 병합하여 계층 구조를 형성한다. 이와 같은 방법의 예시로는, [그림 2]와 같이 먼저 유사도 행렬을

만든다. 표의 숫자는 두 개체 사이의 유사도로서 숫자가 작으면 거리가 가까우며 유사도가 높다는 의미이다. A와 D의 유사도가 높기 때문에 다음의 행렬에서는 AD를 묶어 업데이트를 진행한다. 이러한 과정을 반복하여 아래의 덴드로그램(Dendrogram) 결과로 나타낼 수 있다. 두번째, 분할 계층적 클러스터링은 모든 데이터 포인트를 하나의 클러스터로 시작하여 서로 다른 클러스터로 분할해가는 방식이다. 초기에는 모든 데이터가 하나의 클러스터로 속해 있는 상태에서 분할 과정을 통해 계층 구조를 형성한다. 분할은 가장 큰 클러스터를 분할하여 두 개의 하위 클러스터를 생성하고, 이를 반복한다[4]. Gosh 등은 사용자들의 클러스터 식별을 번갈아 가며 추정하고 경사 하강법을 통해 사용자 클러스터의 모델 매개변수를 최적화하는 알고리즘을 제안하였고[5], Briggs 등은 연합학습 과정에 계층적 클러스터링을 통해 클라이언트 클러스터를 전역 모델과의 유사성에 따라 분리하고 분리된 클러스터는 독립적으로 병렬로 학습되는 방법을 제안하여 더 적은 라운드에서 수렴하게 됨을 보여주었다. 하지만 기존 연합학습과 같이 중앙 서버의 신뢰를 전제로 하고 있어, 안전성이 떨어지며, 학습, 유사성 계산 및 클러스터링 등의 모든 작업이 중앙 서버 하나에서 이루어지기 때문에 효율성이 매우 떨어진다[6].



[그림 2] 계층적 클러스터링 동작 과정

[Fig. 2] Hierarchical Clustering Process

2.3 프록시 서명(Proxy Signature)

프록시 서명은 전자 서명 기술 중 하나로, 한 사람이 다른 사람의 대리로 서명을 생성할 수 있는 기술이다. 프록시 서명을 통해 서명을 생성하면, 서명자의 개인 정보와 신원이 노출되지 않는다. 이는 개인 정보 보호 및 익명성을 강화할 수 있는 중요한 기법이다. 또한, 서명자와 프록시 간의 역할 분담을 허용하여, 작업을 분산시킬 수 있고 효율성을 향상시킬 수도 있다. 따라서, 본 논문에서는 아래와 같은 프록시 서명을 알고리즘을 적용한다[7].

- $(\boldsymbol{Para}, \boldsymbol{s}) \leftarrow \boldsymbol{ParaGen}(\boldsymbol{\ell})$: 파라미터 값을 생성한다.
- $\boldsymbol{sk_{ID_i}} \leftarrow \boldsymbol{KeyExtract}(\boldsymbol{Para}, \boldsymbol{ID_i}, \boldsymbol{s})$: 생성된 파라미터 값과 본 서명자와 위임 서명자의 ID값을 넣어 비밀키를 생성한다.

- $\boldsymbol{\sigma_S \leftarrow StandardSign(Para, M, sk_{ID})}$: 파라미터 값, 서명자의 비밀키 $\boldsymbol{sk_{ID}}$, 메시지 M을 입력으로 받아 메시지에 대한 서명 $\boldsymbol{\sigma_S}$을 생성한다.

- $\boldsymbol{\{True, False\} \leftarrow StandardVer(Para, ID, M, \sigma_S)}$: 파라미터 값, 서명자의 ID, 서명 $\boldsymbol{\sigma_S}$를 입력으로 받아, 메시지 M 이 유효한 서명인 경우 $\boldsymbol{True}$를 출력, 그렇지 않은 경우 $\boldsymbol{False}$를 출력하여 서명을 검증한다.

- $\boldsymbol{\sigma_w \leftarrow DelegationGen(Para, W, sk_{ID_a})}$: 파라미터 값, 원 서명자의 비밀키 $\boldsymbol{sk_{ID_a}}$, 대리 서명자와의 계약을 뜻하는 W 값을 입력으로 받아 대리 서명 값인 $\boldsymbol{\sigma_w}$를 생성한다.

- $\boldsymbol{\sigma \leftarrow ProxySign(Para, W, \sigma_w, sk_{ID_b}, M)}$: 파라미터 값, 대리 서명자와의 계약을 뜻하는 W 값, 대리 서명 값인 $\boldsymbol{\sigma_w}$, 프록시 서명자의 비밀키 $\boldsymbol{sk_{ID_b}}$, 메시지 M을 입력으로 받아, 프록시 서명 값인 $\boldsymbol{\sigma}$을 생성한다.

- $\boldsymbol{\{True, False\} \leftarrow ProxyVer(Para, ID_a, ID_b, W, M, \sigma)}$: 파라미터 값, 원 서명자의 ID값인 $\boldsymbol{ID_a}$, 프록시 서명자의 ID값인 $\boldsymbol{ID_b}$, W 값, 서명된 메시지 M, 프록시 서명 값인 $\boldsymbol{\sigma}$를 입력으로 받아 유효한 프록시 서명인지 검증한다.

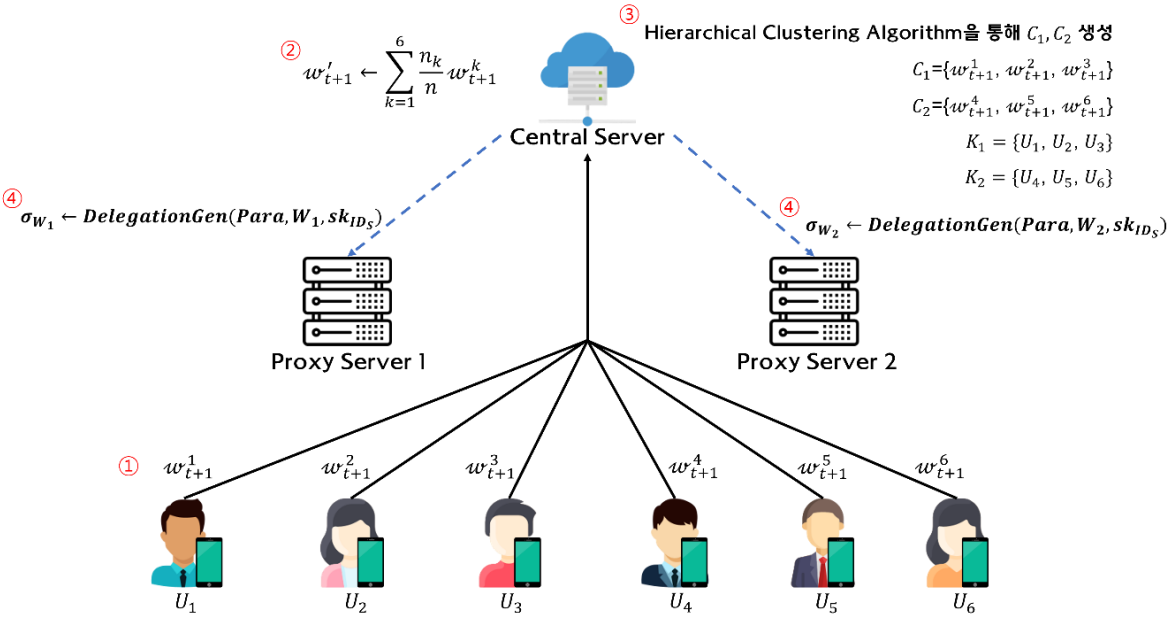
3. 제안 기법

본 논문에서 제안하는 기법은 기존 연합학습 기법에서의 단점 두가지를 해결하기 위해 설계되었다. 첫번째, 연합학습은 일반적으로 중앙 서버가 신뢰할 수 있는 엔티티로 가정되는 환경에서 진행된다는 점, 두번째, 중앙 서버에서 모든 계산과 통신을 처리하므로 비효율적일 수 있다는 점이다. 제안하는 기법은 프록시 서버들을 활용한 학습과정을 통해 역할을 분산시켜 보안 및 신뢰성 완화하고, 전체적인 효율성을 향상시킬 수 있다.

[표 1] 기호 설명

[Table 1] Notation

Notation	
$\boldsymbol{U_k}$	Client(User)
$\boldsymbol{C_x}$	Cluster
$\boldsymbol{w_{t+1}^k}$	$\boldsymbol{k}$'th client's weight
$\boldsymbol{K_x}$	Clustered group of users
$\boldsymbol{\sigma_{w_x}}$	Proxy signature value of $\boldsymbol{x}$'th proxy server
$\boldsymbol{Para}$	Parameter
sk_{ID}	Signer's private key
σ_S	Digital signature

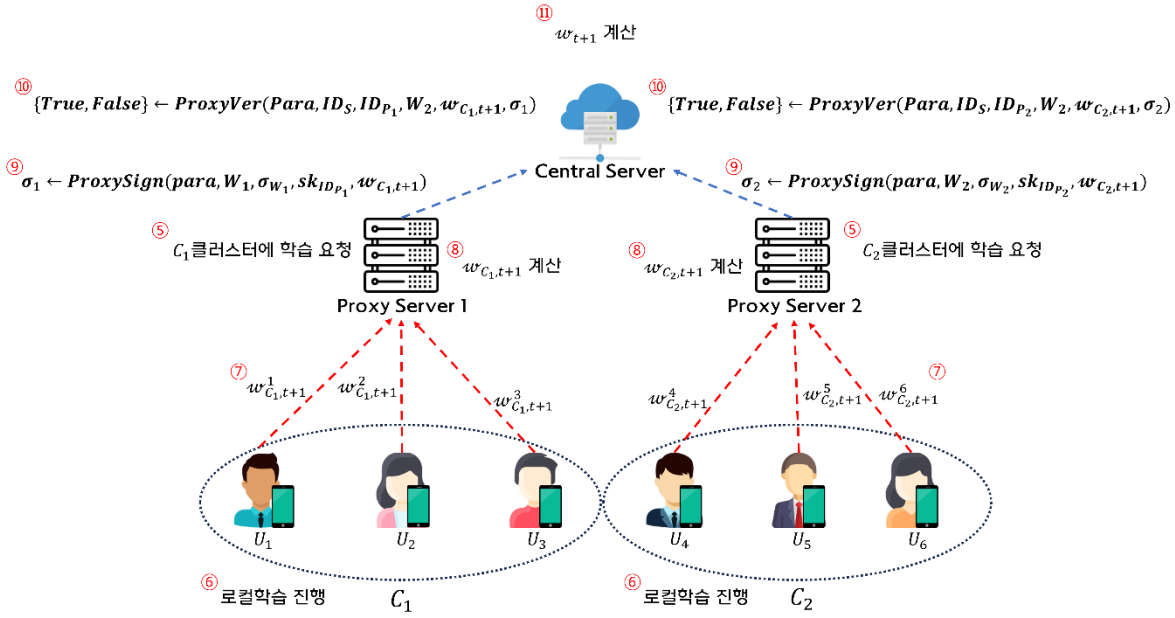


[그림 3] 제안 기법의 동작 과정 1

[Fig. 3] Operational Process of the Proposed Method 1

제안하는 알고리즘의 구성요소로는 기존 연합학습에서 사용되는 중앙 서버, 클라이언트들, 본 논문에서 새롭게 추가하는 프록시 서버 2대가 있다. 먼저, 중앙 서버는 학습을 진행하기전 가중치를 초기화하고 시작한다. [그림 3]과 같이 $\boldsymbol{U_1 \sim U_6}$ 의 클라이언트들은 로컬학습을 통해 클라이언트들의 가중치 값인 $\boldsymbol{w^k_{t+1}}$ 를 계산하여 중앙 서버에게 보낸다. 클라이언트들의 가중치 값을 받은 중앙 서버는 연합학습에서 사용되는 FedAvg(연합평균)의 알고리즘을 사용하여, 클라이언트들의 가중치를 통해 글로벌 모델을 업데이트 한다. 일반적인 연합학습은 이러한 단계를 반복하지만, 본 논문에서는 계층적 클러스터링 알고리즘을 통해 클라이언트 가중치 값을 유사도가 높은 값끼리 분류하는 과정을 거친다. $\boldsymbol{U_1 \sim U_6}$ 의 가중치 값들을 비교하여 클러스터링을 진행한 후 $\boldsymbol{C_1 = \{w^1_{t+1}, w^2_{t+1}, w^3_{t+1}\}, C_2 = \{w^4_{t+1}, w^5_{t+1}, w^6_{t+1}\}}$으로 나뉘었을 때, 중앙 서버는 프록시 서버 2대에게 권한을 위임하며 중앙 서버의 서명 값을 생성한다.

이어 [그림 4]에서는 중앙 서버로부터 $\boldsymbol{C_1, C_2}$ 각각의 클러스터의 클라이언트 정보를 받은 프록시 서버1, 프록시 서버2는 해당 클러스터에 속하는 클라이언트들에게 로컬학습을 요청한다. 예를 들어 $\boldsymbol{C_1}$ 의 경우, $\boldsymbol{U_1 \sim U_3}$ 의 클라이언트들은 로컬학습을 진행한 결과를 다시 프록시 서버1에게 보내준다. 중앙 서버가 FedAvg 알고리즘을 진행했던 것과 마찬가지로 프록시 서버1도 클라이언트 3명의 가중치 값으로 $\boldsymbol{C_1}$ 가중치 값을 계산한다. 중앙 서버로 보내기전 모든 과정이 끝나면, 프록시 서버1은 중앙 서버에게 계산 결과를 메시지로 넣어 생성한 프록시 서명 값을 보낸다. 이러한 프록시 서명 값을 받은 중앙 서버는 프록시 서명 검증 단계를 통해, 위임했던 프록시 서버가 맞는지 확인후, $\boldsymbol{C_1, C_2}$ 들의 가중치 데이터를 다시 한번 계산을 통해 한번의 라운드를 마친다.



[그림 4] 제안 기법의 동작 과정 2

[Fig. 4] Operational Process of the Proposed Method 2

4. 효율성 및 안전성 분석

4.1 효율성 분석

본 논문에서 제안한 알고리즘과 Briggs 등이 제안한 클러스터링 기반 연합학습 기법과의 효율성을 아래의 [표 2]과 같은 식을 통해서 비교한다.

[표 2] 시간 효율성 비교 수식

[Table 2] Time Efficiency Comparison Formula

Briggs[6]	$2n * T_{tx} + T_{FedAvg_n}$ $(T_{FedAvg} = \sum_{i=1}^{n} w_i)$
Ours	$2n * T_{tx} + 2c + \frac{1}{c} T_{FedAvg_n} + T_{FedAvg_c}$

n : $number\ of\ clients$, c : $number\ of\ clients$
T_{tx} : $weight\ Transmission\ Time$
T_{FedAvg_n} : $FedAvg\ value\ computed\ from\ n\ weights$

실행 시간 측면에서, Briggs 등이 제안한 기법은 매 라운드마다 하나의 서버가 FedAvg 알고리즘 수행, 유사성 계산, 클러스터 분류 등을 전부 처리해야 되기 때문에 서버의 처리량에 따라 차이가 크다. 또한, 위 [표 2]의 식을 통해, 클라이언트의 수인 n 이 증가할수록 서버에서 FedAvg의 알고리즘을 통해 계산해야 하는 과정이 매우 증가한다. 반면에, 우리가 제안한 기법은 중앙 서버가 하는 일들을 프록시 서버가 위임을 받아 일부를 처리하기 때문에, 중앙 서버는 FedAvg 과정을 프록시 서버 개수만큼의 연산만

진행하고, 프록시 서버는 Briggs 등의 기법과 달리, 프록시 서버가 많을수록 그 수만큼 반비례하여 동시에 처리가 가능하다. 따라서 결과적으로 봤을 때, 클라이언트의 수가 늘어날수록 기존의 기법보다 더 빠른 처리가 가능하므로, 여러 논문에서 언급했던 대규모 처리의 어려움을 해결할 수 있다.

수렴 속도 측면에서, Briggs 등의 기법은 앞서 얘기한 것과 같이 매 라운드가 끝나기 위해서는 클러스터링 알고리즘을 순차적으로 진행해야 한다. 하지만, 제안한 기법은 동시에 여러 개의 클러스터에 대한 학습을 진행하기 때문에 더 적은 라운드에서 수렴할 수 있다. 확장성으로 볼 때, 대용량 데이터셋에서 효과적인지를 판단하기 위해서는 분산 시스템 환경, 계산 리소스가 중요하다. 본 기법은 프록시 서버를 활용하여 더 분산된 시스템을 제안하였고, 병렬로 학습을 진행하므로 클라이언트 수에 따라 중앙 서버의 과부하가 높아지는 Briggs 등의 기법과 달리 클라이언트의 수가 많아져도 프록시 서버의 수에 따라 확장성이 높일 수 있다는 장점이 있다.

[표 3] 수렴 속도 및 확장성 비교

[Table 3] Comparison of Convergence Speed and Scalability

Scheme	수렴 속도 (Convergence Speed)	확장성 (Scalability)
[6]	$\boldsymbol{R \propto \frac{1}{C}}$	$\boldsymbol{S \propto \frac{1}{n}}$
Ours	$\boldsymbol{R \propto K}$	$\boldsymbol{R \propto C}$

4.2 안전성 분석

기존 제안되고 있는 연합학습의 기법은 학습 및 연산을 모두 서버에서 진행한다. 중앙 서버가 클라이언트의 개인 데이터를 받지 않고도 학습이 가능해서 보안 및 프라이버시 이점을 갖지만, 중앙 서버가 해킹이나 악의적인 공격에 취약하다면 클라이언트의 데이터 또는 모델에 대한 위협이 발생할 수 있어 동형 암호와 재암호화를 활용하여 암호화된 데이터 처리와 접근 제어를 개선하는 연구도 진행되고 있다[8]. 또한, 데이터 전송 과정에서 중간자 공격이 발생할 수 있다[9]. 하지만, 본 논문에서 제안하는 알고리즘은 하나의 중앙 서버만 두지 않고, 프록시 서버라는 보조 개체를 활용해서 위험성을 줄이는 역할을 한다. 따라서, 기존에서 제안된 기법과 비교했을 때, 중앙 서버의 데이터 보안을 높이고 프록시 서버와의 서명의 전달 과정을 통해 더 안정성을 높였다.

5. 결론

본 논문에서는 Non-IID 데이터 문제를 효율적이고 안전하게 해결하기 위한 계층적 클러스터링 기반 연합학습을 제안하였다. 연합학습은 분산된 데이터를 처리하고 전역 모델을 학습하는데 있어서 중앙 서버의 신뢰를 전제로 하고 있으며, 중앙 서버에서 모든 계산과 통신을 처리하는 비효율성도 존재한다. 이에 대한 대안으로 기존에 제안되었던 계층적 클러스터링 기반 연합학습에 프록시 서버를 도입하여 학습과정을 분산시키고

보안성 및 신뢰성을 강화하도록 설계하였다. 알고리즘의 연산식을 통해서 효율성을 비교하여 본 논문이 효과적이라는 것을 보였다. 또한 안정성을 비교하였지만, 추후 연구에서는 실험을 진행하여 제안된 알고리즘의 성능과 효과를 더욱 확장할 예정이다. 우선적으로, 다양한 데이터셋과 실제 응용 분야에서의 실험을 통해 제안된 알고리즘의 일반화 능력과 확장성을 평가할 것이다. 더 나아가, 보안 및 프라이버시 측면에서의 추가적인 보호기법 및 프록시 서버의 역할을 최적화하는 방법에 대한 연구도 진행할 것이다.

6. 감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1F1A1063992).

References

[1] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, Blaise Aguera y Arcas, Communication-Efficient Learning of Deep Networks from Decentralized Data, Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR, (2017)
DOI: http://dx.doi.org/10.48550/arXiv.1602.05629

[2] Pravin Chandran, Raghavendra Bhat, Avinash Chakravarthy and Srikanth Chandar, Divide-and-Conquer Federated Learning Under Data Heterogeneity, CS & IT Conference Proceedings, (2021), Vol.11. No.13, pp.21-33.
DOI: http://dx.doi.org/10.5121/csit.2021.111302

[3] William H. E. Day, Herbert Edelsbrunner, Efficient algorithms for agglomerative hierarchical clustering methods, Journal of classification, (1984), Vol.1, pp.7-24.
DOI: http://dx.doi.org/10.1007/BF01890115

[4] Maurice Roux, A Comparative Study of Divisive and Agglomerative Hierarchical Clustering Algorithms, Journal of Classification, (2018), Vol.35, pp.345-366.
DOI: http://dx.doi.org/10.1007/s00357-018-9259-9

[5] A. Ghosh, J. Chung, D. Yin, K. Ramchandran, An efficient framework for clustered federated learning, IEEE Transactions on Information Theory, (2022), Vol.68, No.12, pp. 8076-8091.
DOI: http://dx.doi.org/10.1109/TIT.2022.3192506

[6] C. Briggs, Z. Fan, P. Andras, Federated learning with hierarchical clustering of local updates to improve training on non-IID data, 2020 International Joint Conference on Neural Networks (IJCNN), (2020)
DOI: http://dx.doi.org/10.1109/IJCNN48605.2020.9207469

[7] W. Wu, Y. Mu, W. Susilo, J. Seberry, X. Huang, Identity-based proxy signature from pairings, International Conference on Autonomic and Trusted Computing, ATC, (2007)
DOI: http://dx.doi.org/10.1007/978-3-540-73547-2_5

[8] Chun-I Fan, Ya-Wen Hsu, Cheng-Han Shie, Yi-Fan Tseng, ID-Based Multireceiver Homomorphic Proxy Re-Encryption in Federated Learning, ACM Transactions on Sensor Networks, (2022), Vol.18, No.4, pp.1-25.
DOI: http://dx.doi.org/10.1145/3540199

[9] Xiang Ma, Haijian Sun, Rose Qingyang Hu, Yi Qian, A new implementation of federated learning for privacy and security enhancement, GLOBECOM 2022-2022 IEEE Global Communications Conference, IEEE, (2022)
DOI: http://dx.doi.org/10.1109/GLOBECOM48099.2022.10001614

