Journal of KIIT. Vol. 21, No. 12, pp. 47-57, Dec. 31, 2023. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2023.21.12.47

정확도 기반 참여 제한 연합학습 연구

이주원*[1], 전석환*[2], 방준일*[3], 김화종**

Accuracy-based Limited Participation Federated Learning Study

Joowon Lee*[1], Sukhwan Jeon*[2], Junil Bang*[3], and Hwajong Kim**

이 논문은 2023년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과임(2022RIS-005)

요 약

현재 연구되고 있는 많은 인공지능 관련 분야의 경우, 개인정보보호 문제 등으로 인하여 민감한 데이터의 공유가 불가능한 실정이다. 이러한 문제 해결을 위하여 연합학습이 주목받고 있으나, 기존의 연합학습 방법은 낮은 성능을 지닌 일부 클라이언트로 인하여 글로벌 모델 생성 시 성능 저하가 발생할 수 있다. 본 논문에서는 연합학습 과정에서 글로벌 모델의 성능에 악영향을 미치는 파라미터를 제외하고 집계하는 방법인 정확도 기반 참여 제한 연합학습을 제안한다. 제안하는 방법은 정확도를 기반으로 낮은 성능의 클라이언트를 감지하고 글로벌 모델 생성 시 참여를 제한하도록 동작한다. MNIST 데이터를 활용해 구성한 iid 및 non-iid환경에서 성능 평가를 시행하여 기존 연합학습 대비 글로벌 모델이 목표 성능에 도달하는 라운드 횟수를 iid 상황에서는 평균 2라운드, non-iid 상황에서는 평균 7.77라운드 단축하였음을 보인다.

Abstract

In many fields of artificial intelligence research currently underway, the sharing of sensitive data is often impossible due to issues such as privacy protection. Federated learning has been gaining attention as a solution to this problem, however, the conventional federated learning methods can degrade the performance of the global model due to the influence of some low-performance clients. This paper proposes an accuracy-based participation restriction federated learning technique, which excludes parameters that negatively impact the performance of the global model during the federated learning process. The proposed method operates by detecting low-performance clients based on accuracy and limiting their participation in the creation of the global model. Performance evaluation was carried out in iid and non-iid environments using MNIST data, demonstrating that compared to traditional federated learning, the number of rounds required for the global model to reach the target performance is reduced by an average of 2 rounds in the iid situation, and an average of 7.77 rounds in the non-iid situation.

Keywords

federated learning, accuracy, limited participation, non-iid, MNIST

* 강원대학교 컴퓨터정보통신공학과
- ORCID[1]: https://orcid.org/0000-0001-8638-4081
- ORCID[2]: https://orcid.org/0000-0002-0882-6347
- ORCID[3]: https://orcid.org/0000-0003-0582-1572
** 강원대학교 컴퓨터정보통신공학과 교수(교신저자)
- ORCID: https://orcid.org/0000-0002-3822-390X

· Received: Dec. 03, 2023, Revised: Dec. 18, 2023, Accepted: Dec. 21, 2023
· Corresponding Author: Hwajong Kim
Dept. of Computer and Communications Engineering, Kangwon National Univ., Gangwondaehak-gil 1, Chuncheon-si, Gangwon-do, Korea
Tel.: +82-33-250-6323, Email: hjkim3@gmail.com

Ⅰ. 서 론

대다수의 인공지능 분야에서 데이터의 공유와 활용은 큰 숙제이다. 인공지능 학습 시 활용한 데이터의 양과 품질에 따라 모델의 성능이 결정되기 때문이다.

그러나 현재 연구되고 있는 많은 인공지능 관련 도메인들의 경우, 데이터를 쉽게 획득할 수 없는 것이 현실이다. 예를 들어, 정밀의료 또는 신약개발 분야의 경우 전체 분야에서 생산되는 데이터의 양은 매우 많으나, 연구를 원하는 각 기관 또는 대학 등에서 보유하고 있는 데이터의 수는 한정적이며 공유가 실질적으로 불가능한 민감한 데이터이기 때문이다.

정밀의료 분야의 경우 각 대학병원에서 생산되는 데이터의 양은 많으나 모든 데이터가 환자와 연결된 개인정보이다. 따라서 각 대학병원, 연구소 등은 스스로가 생산하고 보유 중인 데이터만을 활용하여 정밀의료 연구를 진행할 수밖에 없으며, 데이터를 공유하고자 하면 개인정보 보호법에 따라 매우 까다로운 데이터 반출 절차를 거쳐야 한다. 이러한 문제점을 데이터의 비식별화(가명화, 익명화 등)처리 및 데이터 결합 전문기관에서의 결합을 통해 해결하고자 하는 노력이 있었으나, 이 또한 데이터심의위원회(DRB) 및 윤리 심의 위원회(IRB)의 심의가 필요하고 데이터 공유 및 결합 절차가 진행된 이후에도 데이터의 유출 방지를 위하여 안심존 서비스 등의 보안대책으로 관리해야 하는 등 비용과 시간의 소요가 매우 크다.

신약개발 분야도 마찬가지이다. 많은 제약사는 인공지능 신약개발을 위하여 일명 Wet-lab(실제 실험 검증 기반 연구실)에서 얻어진 데이터로 데이터 세트를 구축하고 Dry-lab(데이터 분석 및 통계 기반 연구실)에서 활용한다. 생물학적 동등성 시험 및 임상시험의 경우 정밀의료의 사례와 같이 개인정보가 결부되어 있으므로 개인정보 보호법에 따라 보호된다. 또한, 화합물의 실험 결과를 통해 얻은 데이터의 경우 이러한 개인정보 보호법의 영향을 받지는 않으나 데이터의 보유량이 곧 제약사의 경쟁력이 되므로, 데이터를 적극적으로 공개, 공유하지는 못하고 있는 실정이다.

이러한 상황을 효과적으로 해결 가능한 방안은 연합학습이다. 연합학습은 법리적, 정책적, 또는 기술적으로 데이터를 우회하여 공유하는 방법 또는 정밀의료 및 신약개발의 사례처럼 비식별화 기술과 암호화를 거쳐 공유하는 방법이 아니며, 원본 데이터를 공유하지 않으면서도 데이터를 활용한 인공지능 모델 학습이 가능한 학습 방법이다.

연합학습은 윤리적, 법리적, 기술적 문제로 인하여 데이터를 공유하기 어렵거나 불가능한 분야에 적용이 가능하다[1]-[3]. 예를 들어, 활용해야 하는 대상 데이터가 환자의 개인정보 및 진단, 치료에 관련된 데이터일 경우, 기존 인공지능 모델 개발 방식과 같이 하나의 중앙 서버에 모든 데이터를 집합하여 활용하는 것이 아니라 일종의 분산학습 방식과 같이 데이터를 학습한 인공지능 모델의 파라미터만 수집하여 결합, 성능을 개선한 공통 모델을 만들어 각 클라이언트에게 재배포하고 공통 모델을 활용할 수 있도록 할 수 있다.

그러나 현재 관련된 연구 분야에서 활용되고 있는 일반적인 연합학습 프레임워크의 경우 특정한 클라이언트가 학습 결과가 좋지 못하여 성능이 나쁠 경우, 이를 결합한 공통 모델 생성 시 악영향을 줄 수 있다는 문제점이 있다.

따라서 본 연구에서는 이러한 문제점을 해결 가능한 정확도 기반 참여 제한 연합학습 방식을 제안하고자 한다. 정확도 기반 참여 제한 연합학습 방식이란, 연합학습 시 개별 클라이언트가 학습한 모델 파라미터를 가져오는 과정에서 각 클라이언트의 자체 데이터로 평가된 모델의 정확도를 함께 수집하고, 이를 토대로 성능이 낮은 모델의 파라미터를 제외하여 공통 모델을 생성하는 방식이다.

본 연구는 5개 클라이언트로 나누어진 환경에서 연합학습을 진행하고, 정확도 기반 참여 제한 연합학습 방식을 적용하여 기존 방식 대비 목표한 정확도 수준에 달성하는 시간 및 최대 정확도 수준을 비교한 연구이다. 이를 위하여 학습에 사용할 데이터 세트는 MNIST 데이터 세트를 사용하였으며, 인공지능 모델은 10개의 클래스에 대하여 다중 분류가 가능한 MLP 모델을 사용하였다. 실험은 각 클라이언트가 보유한 데이터가 모두 균일한 iid 상황과 함께 클라이언트 간 데이터 분포가 일정하지 않아 편향이 발생하는 non-iid 환경으로 두 가지 조건을 가정하여 진행하였다.

본 논문의 구성은 다음과 같다. 2장은 관련 연구 및 방법론으로써, 본 연구에서 활용한 개념인 연합학습, MNIST, iid, MLP 등에 관하여 기술한다. 3장에서는 본 논문에서 제안하는 정확도 기반 참여 제한 연합학습에 대하여 자세하게 기술한다. 4장에서는 구현된 연합학습 방법을 활용하여 실시한 성능 평가에 대해 논한다. 마지막으로 5장에서는 결론에 관해 기술하고 본 연구가 갖는 한계점 및 향후 연구에 대하여 기술한다.

II. 관련 연구

2.1 연합학습(Federated learning)

연합학습이란 구글에서 최초로 제안한 개념[4]으로, 데이터를 한곳에 집중시키지 않고 분산된 환경에서 인공지능 학습을 수행하는 방법이다. 학습 시 참여자는 클라이언트(Client)로 지칭하고, 이는 병원, 대학, 연구소 등의 기관(Institute) 또는 스마트폰, 임베디드 장치 등의 기기(Device)가 될 수 있다. 이러한 클라이언트들은 각자 보유한 데이터를 직접 공유하지 않고 보유한 데이터를 활용하여 인공지능 모델을 학습한다. 학습이 완료된 모델의 파라미터를 서버(Server)로 전송하고 이를 바탕으로 글로벌 모델(Global model)을 만들어 클라이언트에게 재배포한다. 이러한 방식으로 연합학습을 진행하였을 경우, 데이터의 이동 없이도 클라이언트들이 보유한 데이터를 모두 학습할 수 있으며, 데이터를 모아 학습한 것(Centralized learning)과 유사한 효과를 지닌다[5]-[9].

따라서 연합학습은 데이터가 이동하지 않으므로, 데이터가 이동 및 공유되는 상황에서 발생할 수 있는 데이터 유출, 탈취, 저장소에 대한 공격 등을 미리 예방할 수 있다. 또한, 법적으로 데이터를 이동시키는 것이 불가능한 상황에서도 유용하게 적용할 수 있다.

더불어 리소스 관리 부분에서도 이점을 지닌다. 데이터를 수집하여 집적(Integration), 저장하지 않으므로 대규모의 데이터 보관 장치를 준비할 필요가 없으며, 데이터를 전송하는데 필요한 통신 비용이 감소하고 데이터가 각 클라이언트에서 독립적으로 학습되므로 인공지능 학습에 소모되는 시간 또한 더욱 줄어든다.

연합학습의 일반적인 동작 절차는 다음과 같다.

1. 초기화(Initialize) : 서버에서 각 클라이언트가 공통으로 학습할 모델을 정의하고, 모델의 파라미터를 초기화한다.
2. 클라이언트 선택 : 연합학습에 참여할 클라이언트를 선택하고, 통신을 위해 연결한다.
3. 글로벌 모델 배포 (Distribution): 서버에서 정의하였던 공통 학습 모델의 초기 버전을 참여 클라이언트에게 배포한다.
4. 로컬 학습 : 각 클라이언트에서 배포 받은 글로벌 모델을 클라이언트가 보유하고 있는 데이터(로컬 데이터)로 학습하도록 한다.
5. 파라미터 전달 : 각 클라이언트에서 학습된 모델의 파라미터(로컬 모델 파라미터)를 서버로 전송한다.
6. 파라미터 집계(Aggregation) : 서버에서 각 클라이언트로부터 전달받은 로컬 모델 파라미터를 집계하여 글로벌 모델의 파라미터를 갱신한다. 집계에는 다양한 집계 방식을 활용 가능하다.

연합학습 동작 시 위 과정 1~6번을 반복하는 것을 한 번의 round 라고 한다. 연합학습이 처음 시작되었을 때는 round를 1~6까지 모두 수행하며, 두 번째 round부터는 3~6번까지 반복적으로 수행한다[11].

일반적인 연합학습 예시는 그림 1과 같다.

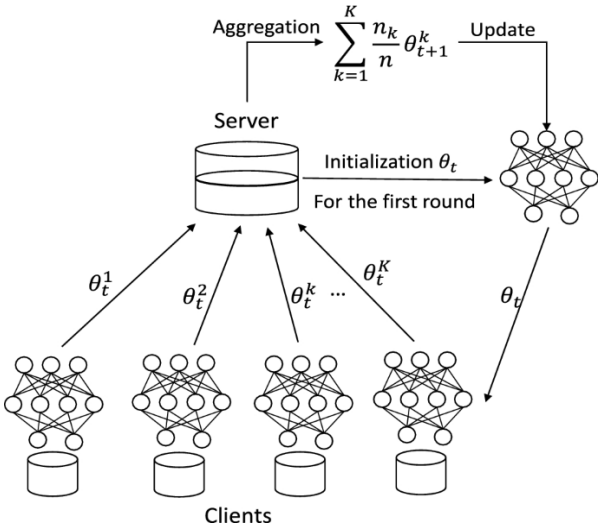

그림 1. 일반적인 연합학습 예시[10]

Fig. 1. Example of federated learning involving multiple clients[10]

2.2 MNIST

MNIST란 Modified National Institute of Standards and Technology database의 약자로써, 손으로 쓴 숫자들로 이루어진 데이터 세트이다. 미국 국립 표준 기술 연구소(NIST)에서 수집한 데이터를 Yann LeCun, Corinna Cortes, Christopher J.C. Burges가 수정하여 만들었다. 이들은 기존 NIST에서 보유한 데이터인 Special Database 1과 Special Database 3을 조합하여 MNIST 데이터 세트를 만들었다[12]-[14].

MNIST는 다음과 같은 특징들을 지니고 있다.

규모 : 60,000개의 학습용 데이터와 10,000개의 테스트 데이터로 구성되어 인공지능 모델의 성능을 테스트하기에 충분한 양을 지닌다.

간결성 : 각 이미지는 28x28픽셀 크기이며, 픽셀이 0~255 사이 정숫값을 지니는 회색조(Grayscale) 이미지로, 처리가 간단하고 복잡하지 않다. 또한, 예측 대상인 레이블이 명확하므로 학습 과정에서 모호성이 적은 편이다.

다양성 : 손글씨 숫자 이미지들은 다양한 사람들에 의해 작성되어 숫자별로 다양한 글쓰기 스타일을 포함하고 있다. 따라서 모델이 다양한 손글씨에 대하여 일반화(Generalization)하는 능력을 얻을 수 있다.

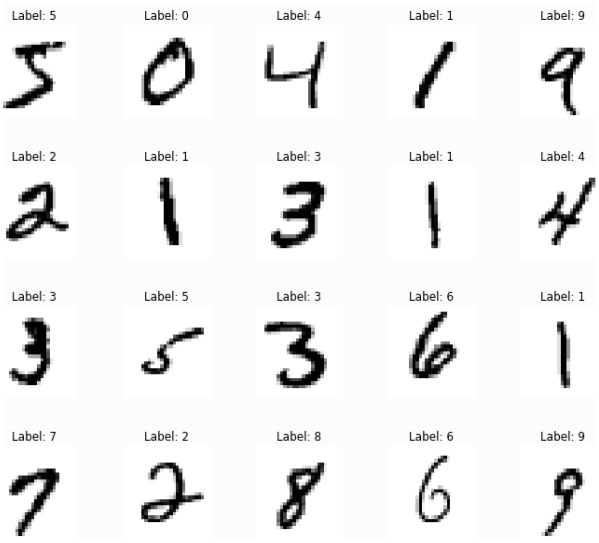

그림 2. MNIST 데이터 세트의 예시
Fig. 2. Example of the MNIST data set

접근성 : 인터넷에서 쉽게 다운로드 받을 수 있으며, 다양한 인공지능 관련 프레임워크에서 쉽게 불러와 사용할 수 있도록 지원하고 있다.

위와 같은 특징들로 인하여 MNIST 데이터 세트는 여러 인공지능 분야에서 다양하게 사용되며 기본적인 벤치마크(Benchmark)의 역할을 하고 있다.

MNIST 데이터 세트의 예시는 그림 2와 같다.

2.3 iid, non-iid

iid란 Independent and identically distributed(독립항등 분포)의 약자로써, 독립적이고 동일하게 분포된 데이터 세트를 의미한다. 데이터의 독립성(Independence)과 동일 분포(Identically distributed)에 대한 설명은 다음과 같다.

독립성 : iid한 데이터 세트는 각각의 데이터가 다른 데이터와 독립적으로 생성되었다는 가정을 기반으로, 하나의 데이터 값 또는 특성이 다른 데이터의 값이나 특성에 영향을 미치지 않는다. 따라서 독립성을 지니는 데이터 세트는 데이터 간에 상호작용과 상관관계가 존재하지 않는다.

동일 분포 : iid한 데이터 세트는 같은 확률 분포에서 독립적으로 추출되어야 한다. 각 데이터는 동일한 분포에서 생성(또는 추출)되고, 데이터에 따라서 분포의 모수 또는 특성이 변화하지 않는다.

만약 데이터 세트가 위 조건을 만족하지 않아 non-iid한 상태일 경우, 인공지능 모델의 학습 및 성능측정은 크게 제한된다[15]-[18].

예를 들어, 데이터가 가진 특성을 바탕으로 0과 1을 분류하는 이진 분류(Binary classification) 모델을 학습한다고 가정한다. 해당 이진 분류 모델을 학습하기 위하여 1이 90%, 0이 10%인 데이터 세트를 사용한다고 가정할 경우, 이진 분류 모델은 모든 예측을 1로 출력하더라도 90%의 정확도를 지닌 것으로 계산된다.

또 다른 경우, 1이 90%, 0이 10%인 데이터 세트를 훈련용으로 학습한 후 1과 0이 각 50%씩 분포한 테스트 데이터 세트로 평가(Evaluation)를 진행한다면 평가 시 해당 모델의 정확도는 크게 낮은 수치로 기록될 것이다.

2.4 MLP

MLP란 Multi Layer Perception(다층 퍼셉트론)의 약자로, 지도학습에 사용되는 인공 신경망의 가장 기본적인 형태이다. 여러 층의 뉴런(또는 노드)으로 구성된 피드 포워드 네트워크이다[19].

MLP는 입력층, 하나 이상의 은닉층, 그리고 출력층으로 이루어져 있으며, 각 층은 여러 뉴런으로 구성되어 있다. MLP의 핵심은 은닉층 개념을 도입함으로써 단순 퍼셉트론의 한계를 극복하는 것이다. 단순 퍼셉트론은 선형 분리 가능한 문제만을 해결할 수 있었으나, MLP는 은닉층을 이용하여 비선형 문제 또한 해결할 수 있다.

MLP를 구성하는 각 뉴런은 가중치(Weight), 편향(Bias), 활성화 함수(Activation Function)를 지닌다. 가중치는 입력 신호가 출력에 미치는 영향을 조절하며, 편향은 뉴런이 얼마나 쉽게 활성화될지를 조절한다. 활성화 함수는 뉴런의 출력값을 결정하는데 사용되며, Sigmoid, Tanh, ReLU등의 비선형 함수가 주로 사용된다.

MLP는 지도 학습(Supervised learning) 알고리즘으로, 신경망의 가중치를 학습하기 위해 역전파(Back-propagation) 알고리즘이 사용된다. 역전파 알고리즘이란 학습 데이터를 통해 실제 출력과 예측 출력 간의 오차를 계산하고, 이 오차를 최소화하는 방향으로 가중치를 조정하는 방법이다.

일반적인 MLP 모델의 예시는 그림 3과 같다.

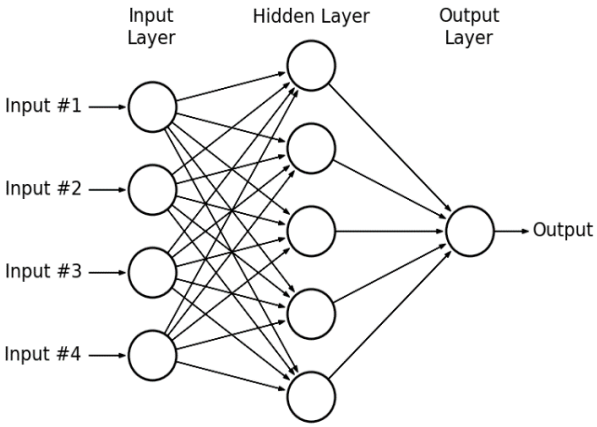

그림 3. 일반적인 MLP 모델의 형태[19]
Fig. 3. Form of a general MLP model[19]

III. 정확도 기반 참여 제한 연합학습

일반적으로 널리 사용되는 연합학습 방법 및 모델의 경우 모든 클라이언트의 파라미터를 수집하고 집계하는 방식으로 동작한다. 그러나 모든 클라이언트의 파라미터를 제한 없이 집계한다면, 성능이 좋지 않은 클라이언트로부터 전달된 파라미터로 인하여 서버가 생성하는 글로벌 모델의 성능에 악영향을 미칠 수 있다. 따라서 본 논문에서는 클라이언트별 정확도 점수를 기반으로 하는 참여 제한 연합학습을 제안한다.

본 논문에서 제안하고자 하는 정확도 기반 참여 제한 연합학습은 연합학습 진행 시 파라미터 집계 단계에서 각 클라이언트의 로컬 evaluation 정확도를 기반으로 집계 포함 여부를 결정함으로써 가장 낮은 정확도를 지닌 클라이언트의 파라미터, 즉 글로벌 모델의 성능에 악영향을 미칠 수 있는 파라미터를 제외하고 집계하도록 구현하였다.

정확도 기반 참여 제한 연합학습을 구현하기 위하여, 다음 과정을 진행하도록 설계하였다.

초기 모델 파라미터 배포 : 모든 클라이언트가 매 실험 시 동일한 파라미터를 지닌 상태로 시작할 수 있도록 서버에서 저장해 두었던 초기 형태의 MLP 모델 파라미터를 배포한다.

로컬 학습 : 각 클라이언트는 로컬에서 보유한 학습 데이터를 활용하여 학습을 진행한다.

로컬 평가 : 각 클라이언트는 로컬에서 보유한 테스트 데이터를 활용하여 evaluation을 진행한다.

파라미터 및 정확도 전달 : 기존 연합학습의 로컬 모델 파라미터 전달과 더불어, 로컬 평가 단계에서 산출한 정확도(Accuracy)또한 함께 전달한다.

파라미터 집계 : 서버에서 각 클라이언트로부터 전달받은 로컬 모델 파라미터를 FedAVG 알고리즘을 활용하여 집계한다. 이때, 전달받은 정확도를 확인하고 정확도가 가장 낮은 클라이언트의 파라미터는 집계에 포함하지 않는다.

글로벌 모델 배포 : 서버에서 파라미터 집계과정을 거쳐 생성한 글로벌 모델을 각 클라이언트에게 배포한다. 이 과정에서 각 클라이언트의 모델 파라미터는 글로벌 모델의 파라미터로 업데이트된다.

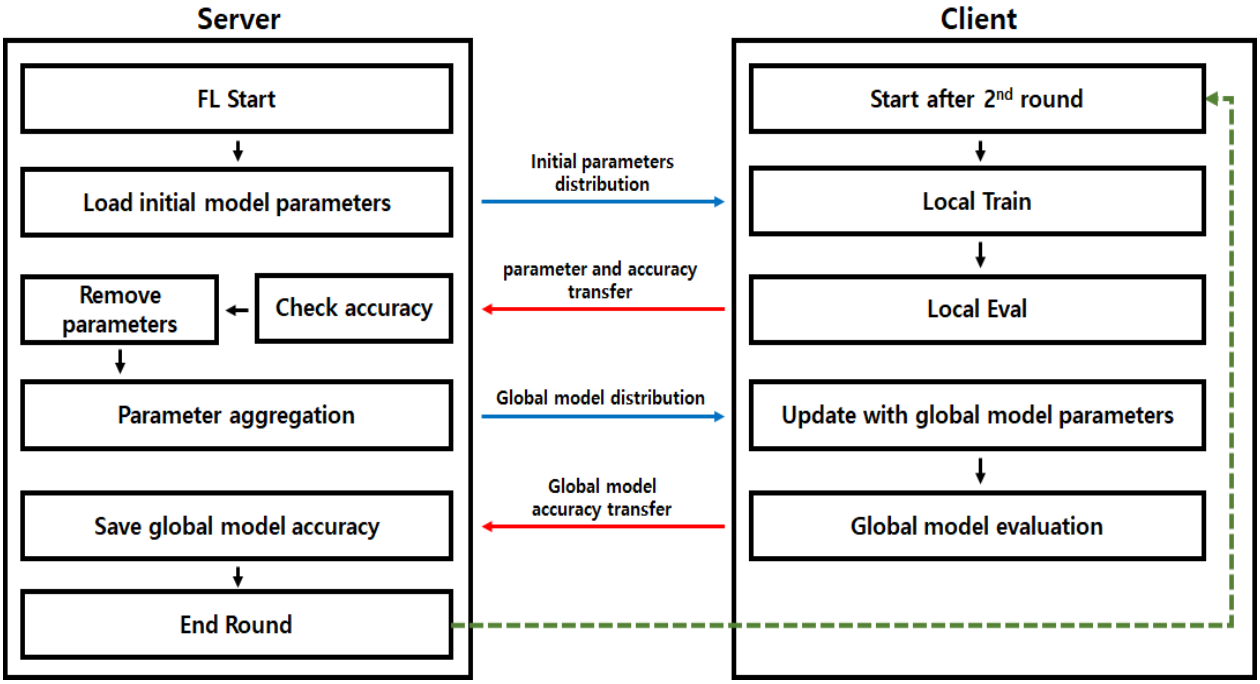

그림 4. 정확도 기반 참여 제한 연합학습 구조
Fig. 4. Accuracy-based limited participation federated learning architecture

글로벌 모델 평가 : 서버에서 배포한 글로벌 모델을 각 클라이언트에서 보유한 테스트 데이터로 evaluation을 진행하고, 정확도를 서버로 전달한다.

글로벌 모델 평가 결과 저장 : 서버는 전달받은 클라이언트별 글로벌 모델의 평가 결과인 정확도를 저장한다.

연합학습을 처음 시작하였을 경우 1~8단계를 진행하며, 이후 2번 round부터는 1번 초기 모델 파라미터 배포 단계를 제외한 2~8단계를 연합학습이 종료될 때까지 반복한다. 제안하는 정확도 기반 참여 제한 연합학습의 구조는 그림 4와 같다.

IV. 실험 설계 및 성능 평가

4.1 데이터 분할

데이터 세트는 MNIST 데이터 세트를 활용하며, 각기 두 가지의 상황을 가정하여 5개의 클라이언트에 분배할 데이터를 iid 상황과 non-iid 상황별로 각기 나누었다.

iid 상황의 경우 분할된 데이터의 형태는 아래 표 1과 같다.

iid 상황에서는 모든 클라이언트가 12,000개의 훈련 데이터와 2,000개의 테스트 데이터를 가질 수 있도록 분할하였으며, 각 클라이언트가 보유한 데이터의 레이블별 분포 비율은 유사하다.

non-iid 상황의 경우 분할된 데이터의 형태는 아래 표 2와 같다.

non-iid 상황에서는 모든 클라이언트가 6,000개의 훈련 데이터와 2,000개의 테스트 데이터를 가지도록 분할되었으나, 클라이언트 5의 경우 이상 분포를 지닌 클라이언트가 되도록 설정하여 학습 데이터 중 1~7까지의 레이블은 존재하지 않되, 8과 9레이블이 각 3000여 개씩 존재하도록 분할하였다. 클라이언트가 이러한 데이터 분포를 지닐 경우, 인공지능 모델의 로컬 학습 시 8과 9에 대해서만 학습하게 된다. 이후 테스트 데이터를 활용한 로컬 evaluation 과정에서 모든 레이블에 대하여 모델 예측 검증과정을 거치게 된다면 evaluation 정확도가 극도로 낮게 평가된다.

표 1. iid 상황에서의 클라이언트별 데이터 분포
Table 1. Data distribution by client in iid situation

Client	Data set	0	1	2	3	4	5	6	7	8	9	Total
Client 1	train set	1206	1351	1176	1228	1184	1048	1208	1279	1127	1193	12000
	test set	175	234	219	207	217	179	178	205	192	194	2000
Client 2	train set	1175	1381	1171	1259	1163	1096	1162	1210	1165	1218	12000
	test set	195	216	199	201	201	193	200	206	192	197	2000
Client 3	train set	1153	1381	1195	1203	1160	1086	1197	1232	1180	1213	12000
	test set	198	236	207	187	181	176	184	198	203	230	2000
Client 4	train set	1195	1357	1220	1199	1148	1094	1181	1247	1181	1178	12000
	test set	205	219	209	208	189	175	194	204	200	197	2000
Client 5	train set	1194	1272	1196	1242	1187	1097	1170	1297	1198	1147	12000
	test set	207	230	198	207	194	169	202	215	187	191	2000
Total		6903	7877	6990	7141	6824	6313	6876	7293	6825	6958	70000

표 2. non-iid 상황에서의 클라이언트별 데이터 분포
Table 2. Data distribution by client in non-iid situation

Client	Data set	0	1	2	3	4	5	6	7	8	9	Total
Client 1	train set	592	671	581	608	623	514	608	651	551	601	6000
	test set	175	234	219	207	217	179	178	205	192	194	2000
Client 2	train set	614	680	595	620	561	534	600	628	576	592	6000
	test set	195	216	199	201	201	193	200	206	192	197	2000
Client 3	train set	577	700	564	655	563	539	563	621	605	613	6000
	test set	198	236	207	187	181	176	184	198	203	230	2000
Client 4	train set	598	681	607	604	600	557	599	589	560	605	6000
	test set	205	219	209	208	189	175	194	204	200	197	2000
Client 5 (Abnomal)	train set	0	0	0	0	0	0	0	0	2998	3002	6000
	test set	207	230	198	207	194	169	202	215	187	191	2000
Total		3361	3867	3379	3497	3329	3036	3328	3517	6264	6422	40000

4.2 학습 환경 정의

인공지능 모델은 MNIST 데이터를 학습하고, 데이터가 지닌 10개의 레이블을 분류할 수 있는 MLP 모델을 사용하였다. 또한 모델 학습 시 모델의 성능 수렴 속도를 조절하여 정확도 기반 참여 제한 방법을 적용하였을 때와 적용하지 않았을 때의 차이가 가시성이 높을 수 있도록 모델의 복잡도를 제한하였으며, 집계 이전에 각 로컬 모델의 학습이 과다하게 진행되지 않도록 1 round 당 1 Epoch의 학습을 진행하도록 하였다.

기존 존재하는 MNIST 데이터 세트를 4.1에서 기술한 바와 같이 클라이언트별로 학습할 수 있도록 DataLoader를 재정의하였으며, MLP 모델 학습이 진행된 후 학습용 데이터가 아닌 테스트 데이터를 활용하여 정확도를 산출할 수 있도록 Train 동작 부를 재정의하였다.

또한, Pytorch 라이브러리를 사용한 MLP 모델 정의 시 학습이 진행되지 않은 초기 모델은 파라미터를 임의로 지정한다. 매 실험 시 모델의 파라미터가 무작위로 초기화될 경우, 특정 클라이언트의 초기 성능이 높게 또는 낮게 산출될 수 있으므로, 초기 MLP 모델을 저장하여 모든 클라이언트가 매 실험 시 같은 파라미터로 시작하도록 하였다.

연합학습 환경은 Adap사에서 공개한 오픈소스 연합학습 프레임워크인 Flower FL을 활용하여 구성하였으며, Pytorch 라이브러리의 MLP 모델 파라미터를 기반으로 FedAVG를 활용한 parameter aggregation을 진행하도록 하였다. Flower FL은 gRPC를 사용하여 서버 - 클라이언트 간 통신을 구현한다[20].

연합학습에 활용된 서버 및 클라이언트의 환경은 다음과 같다.

운영체제 : Ubuntu 20.04 LTS

가상환경 : Anaconda3

CPU : Intel(R) Core(TM) i7-6700K CPU @ 4.00GHz 4-Core

GPU : GeForce GTX TITAN X 12GB

RAM : Samsung DDR4 2133Mhz 16GB * 4

4.3 성능 평가

제안한 정확도 기반 참여 제한 연합학습의 성능 평가를 위하여 본 연구에서는 정확도 기반 참여 제한 연합학습과 일반적인 연합학습을 각기 구현하였으며, 이들 각각의 성능 평가를 위하여 목표한 정확도인 90%까지의 성능에 도달하기까지 몇 회의 round가 소요되는지 측정하였다.

성능 평가 실험은 보유 데이터가 iid한 상황과 non-iid한 상황 두 가지를 모두 진행하였으며, 한 번의 실험(Test)당 정확도 기반 참여 제한 연합학습 30 round, 일반 연합학습 30 round를 진행하고 결과를 기록하여 비교하였다. 일 회 실험의 비교는 우연성에 따라 성능의 우열이 결정될 수 있으므로, 각 상황에서의 실험은 30회씩 진행하여 평균값을 비교하였다.

따라서, 진행한 실험의 목록은 다음과 같다.

iid) 정확도 기반 참여 제한 30 round 30회

iid) 일반 연합학습 30 round 30회

non-iid) 정확도 기반 참여 제한 30 round 30회

non-iid) 일반 연합학습 30 round 30회

1번과 2번 실험 성능을 비교한 내용은 아래 표 3번과 같으며, 3번과 4번 실험 성능을 비교한 내용은 아래 표 4번과 같다.

성능 평가 결과의 내용은 다음과 같다.

test : 실험 회차

acclimit : 정확도 기반 참여 제한 연합학습 실험

nomal : 일반 연합학습 실험

difference : acclimit - nomal 의 차이

round : 글로벌 모델의 성능 목표치인 90% 정확도에 도달하는데 걸린 round 횟수

acc avg before 90% : 글로벌 모델의 성능 목표치인 90% 정확도에 도달하기 이전 round 들의 글로벌 모델 evaluation 정확도 평균. 연합학습이 얼마나 안정적으로 진행되었는지를 나타낸다.

max acc : 해당 실험 회차의 연합학습이 진행되는 30 round 간 가장 높은 정확도 수치

avg : 30회 test 간의 평균

표 3. iid 상황에서의 성능 평가 결과

Table 3. Performance evaluation results in iid situation

	acclimit			nomal			difference		
test	round	acc. avg. before 90%	max acc.	round	acc. avg. before 90%	max acc.	round	acc. avg. before 90%	max acc.
1	4	71.71	95.70	6	72.46	95.78	-2	-0.75	-0.08
2	9	73.78	95.55	5	73.13	96.07	4	0.65	-0.52
3	4	75.39	95.88	5	73.86	96.03	-1	1.52	-0.15
4	6	77.29	95.75	7	73.53	95.66	-1	3.77	0.09
5	11	66.96	84.18	16	71.56	87.43	-5	-4.60	-3.25
6	5	71.45	95.73	4	78.05	95.93	1	-6.61	-0.20
7	5	73.59	95.51	5	68.19	84.82	0	5.40	10.69
8	4	77.32	95.57	4	80.59	96.02	0	-3.27	-0.45
9	4	69.97	95.74	5	72.34	95.79	-1	-2.37	-0.05
10	4	68.92	84.62	6	64.49	84.70	-2	4.43	-0.08
11	6	74.15	95.44	21	70.13	86.49	-15	4.02	8.95
12	5	67.21	95.58	4	69.44	95.92	1	-2.23	-0.34
13	6	76.96	95.75	30	84.03	93.49	-24	-7.07	2.26
14	7	75.16	95.63	3	78.80	95.86	4	-3.64	-0.23
15	5	75.20	95.62	5	75.46	95.89	0	-0.26	-0.27
16	5	64.28	84.74	4	75.46	95.87	1	-11.18	-11.13
17	4	77.34	95.73	29	83.07	95.14	-25	-5.72	0.59
18	9	76.14	95.88	6	73.01	95.75	3	3.13	0.13
19	4	73.68	95.90	4	69.46	84.81	0	4.22	11.09
20	4	77.08	95.67	6	67.43	84.82	-2	9.64	10.85
21	4	74.64	95.83	5	66.67	84.72	-1	7.97	11.11
22	5	74.58	95.78	4	64.97	84.75	1	9.61	11.03
23	12	64.52	84.61	4	73.67	96.07	8	-9.15	-11.46
24	6	69.52	84.66	5	67.01	84.54	1	2.51	0.12
25	6	79.17	95.83	5	65.54	84.72	1	13.63	11.11
26	4	76.27	95.95	6	66.96	87.62	-2	9.31	8.33
27	6	75.69	95.61	4	72.36	96.11	2	3.33	-0.50
28	8	75.04	95.58	7	69.27	84.74	1	5.77	10.84
29	11	80.72	95.69	20	79.46	95.49	-9	1.26	0.20
30	6	72.40	95.58	4	67.90	84.57	2	4.50	11.01
avg.	5.97	73.54	93.84	7.97	72.28	91.19	-2	1.26	2.66

표 4. non-iid 상황에서의 성능 평가 결과
Table 4. Performance evaluation results in non-iid situation

	acclimit			nomal			difference		
test	round	acc. avg. before 90%	max acc.	round	acc. avg. before 90%	max acc.	round	acc. avg. before 90%	max acc.
1	6	75.12	95.39	12	68.77	83.20	-6	6.35	12.19
2	8	68.95	84.10	22	64.20	82.33	-14	4.75	1.77
3	6	69.10	86.21	10	74.24	94.08	-4	-5.14	-7.87
4	7	69.84	83.74	11	72.82	94.38	-4	-2.97	-10.64
5	7	67.21	83.86	11	73.34	94.26	-4	-6.13	-10.40
6	10	66.22	95.05	11	69.75	94.07	-1	-3.53	0.98
7	8	67.63	94.94	18	68.47	82.27	-10	-0.84	12.67
8	8	70.21	83.97	24	76.88	94.25	-16	-6.68	-10.28
9	10	65.04	83.74	17	75.52	94.12	-7	-10.49	-10.38
10	9	67.30	83.84	21	70.41	94.06	-12	-3.11	-10.22
11	13	74.20	94.78	22	72.28	93.77	-9	1.92	1.01
12	14	72.90	94.94	8	72.90	94.59	6	0.00	0.35
13	9	67.79	83.68	23	69.39	93.85	-14	-1.59	-10.17
14	7	74.99	95.04	29	77.58	93.88	-22	-2.59	1.16
15	13	61.35	83.45	20	66.09	83.02	-7	-4.74	0.43
16	8	64.64	83.85	24	73.48	93.03	-16	-8.84	-9.18
17	6	69.92	83.87	27	76.26	94.56	-21	-6.35	-10.69
18	14	73.84	94.58	14	60.84	83.34	0	13.00	11.24
19	11	64.36	86.67	13	65.13	85.32	-2	-0.77	1.35
20	8	72.84	94.91	9	64.54	82.81	-1	8.30	12.10
21	7	67.15	83.95	20	74.80	94.33	-13	-7.65	-10.38
22	10	68.44	94.93	12	72.42	94.32	-2	-3.98	0.61
23	9	66.81	84.02	11	74.41	94.25	-2	-7.60	-10.23
24	7	67.60	94.96	11	71.71	83.74	-4	-4.11	11.22
25	13	62.17	83.32	11	74.14	94.27	2	-11.97	-10.95
26	8	67.75	83.80	15	71.30	93.72	-7	-3.55	-9.92
27	6	69.65	83.86	11	67.78	83.17	-5	1.87	0.69
28	13	76.38	95.21	15	72.35	94.26	-2	4.03	0.95
29	9	67.53	83.90	29	76.48	93.41	-20	-8.95	-9.51
30	9	64.24	94.76	25	74.44	93.80	-16	-10.19	0.96
avg	9.10	68.71	88.44	16.87	71.42	90.82	-7.77	-2.72	-2.37

iid 상황에서의 성능 평가 결과, 정확도 기반 참여 제한 연합학습이 일반 연합학습보다 평균 2 round 빠르게 목표 성능에 도달하였다. 또한, 목표 성능에 도달하기 이전까지 일반 연합학습보다 평균적으로 1.26%p 높은 글로벌 모델 정확도를 보여 학습이 안정적으로 진행되었음을 확인할 수 있었으며, 30 round 간의 최대 정확도에서도 2.66%p 앞선 것을 확인 가능하였다.

non-iid 상황에서의 성능 평가 결과, 정확도 기반 참여 제한 연합학습이 일반 연합학습보다 평균 7.77 round 빠르게 목표 성능에 도달하여 non-iid 상황의 경우 학습 소요시간을 기존 대비 53% 수준으로 크게 감소시킬 수 있음을 보였다. 그러나, 목표 성능에 도달하기 이전까지 일반 연합학습보다 평균 2.72%p의 글로벌 모델 정확도가 감소하였으며, 30 round 간의 최대 정확도에서도 2.37%p 감소한 모습을 보였다.

V. 결론 및 향후 과제

본 연구에서는 연합학습 시 글로벌 모델의 성능에 악영향을 미칠 수 있는 파라미터를 제외하고 집계할 수 있도록 새로운 정확도 기반 참여 제한 연합학습을 제안하였다. 이를 위하여 클라이언트가 학습한 로컬 모델의 정확도를 확인하고, 참여 클라이언트 중 가장 정확도가 낮은 클라이언트가 전달한 모델 파라미터를 집계에서 제외하는 방법을 구현하였다. 제안한 연합학습 방법을 비교 성능 평가를 위하여 정확도 기반 참여 제한 연합학습과 일반 연합학습을 각각 구현하고, iid와 non-iid로 각기 다른 데이터 분포 상황에서 연합학습 진행 시 목표 성능에 도달하는 round 횟수와 글로벌 모델의 정확도를 측정하여 비교하였다. 이로써 기존 방식의 일반적인 연합학습보다 정확도 기반 참여 제한 연합학습이 효과적으로 학습 소요시간을 감소시킬 수 있음을 보였다.

본 연구에서는 효율적인 연합학습을 위하여 정확도 기반 참여 제한 연합학습 방법을 제안 및 구현하였고, 이를 통해 다양한 상황에서 학습 소요시간을 절감시킬 수 있음을 보였다는 점에서 연구에 의의를 찾을 수 있다. 그러나, 제안한 방법에 대하여 MNIST와 같은 벤치마크 데이터가 아닌 실세계(Real-World)의 데이터를 활용한 복잡하고 심도 있는 실험이 이루어지지 않았다는 점, 클라이언트의 데이터 분포가 단순한 비율적 불균형이 아닌 특정

레이블 부재 상황일 경우 정확도 점수를 척도로써 이용할 수 없다는 점, 특정 클라이언트의 파라미터를 집계에서 완전히 배제할 경우 해당 클라이언트가 지닌 데이터 특성을 글로벌 모델에 반영하지 못하여 분류 성능이 하락할 가능성이 존재한다는 점 등에서 본 연구의 한계점을 찾을 수 있다. 따라서 이러한 한계점 극복을 위한 연구를 본 논문의 향후 과제로 하고자 한다.

References

[1] W. Li, et al., "Privacy-preserving federated brain tumour segmentation", International Workshop on Machine Learning in Medical Imaging MLMI 2019: Machine Learning in Medical Imaging, Vol. 11861, pp. 133-141, Oct. 2019. https://doi.org/10.1007/978-3-030-32692-0_16.

[2] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-Institutional Deep Learning Modeling Without Sharing Patient Data: A Feasibility Study on Brain Tumor Segmentation", International MICCAI Brainlesion Workshop BrainLes 2018: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Vol. 11383, pp. 92-104, Jan. 2019. https://doi.org/10.1007/978-3-030-11723-8_9.

[3] M. J. Sheller, et al., "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data", Nature Scientific Reports, Vol. 10, No. 12598, Jul. 2020. https://doi.org/10.1038/s41598-020-69250-1.

[4] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data", Artificial intelligence and statistics, PMLR, Vol. pp. 1273-1282, Apr. 2017.

[5] Q. Dou, et al., "Federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study", NPJ digital medicine, Vol. 4, No. 1, pp. 1-11, Mar. 2021. https://doi.org/10.1038/s41746-021-00431-6.

[6] J. O. Terrail, et al., "Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer", Nature Medicine, Vol. 29, pp. 135-146, Jan. 2023. https://doi.org/10.1038/s41591-022-02155-w.

[7] S. Pati, et al., "Federated learning enables big data for rare cancer boundary detection", Nature Communications, Vol. 13, No. 7346, Dec. 2022. https://doi.org/10.1038/s41467-022-33407-5.

[8] C. I. Bercea, et al., "Federated disentangled representation learning for unsupervised brain anomaly detection", Nature Machine Intelligence, Vol. 4, pp. 685-695, Aug. 2022. https://doi.org/10.1038/s42256-022-00515-2.

[9] I. Dayan, et al., "Federated learning for predicting clinical outcomes in patients with COVID-19", Nature Medicine, Vol. 27, pp. 1735-1743, Sep. 2021. https://doi.org/10.1038/s41591-021-01506-3.

[10] H. Zhu, H. Zhang, and Y. Jin, "From federated learning to federated neural architecture search: a survey", Complex & Intelligent Systems, Vol. 7, pp. 639-657, Jan. 2021. https://doi.org/10.1007/s40747-020-00247-z.

[11] J. Bang, S. Hong, S. Jeon, J. Lee, and H. Kim, "Federated Learning Study using Motion Recognition Model based on Joint Data", Journal of KIIT, Vol. 21, No. 3, pp. 39-48, Mar. 2023. http://dx.doi.org/10.14801/jkiit.2023.21.3.39.

[12] L. Bottou, et al., "Comparison of classifier methods: A case study in handwritten digit recognition", Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol 3 - Conference C: Signal Processing (Cat No 94CH3440-5), Jerusalem, Israel, Oct. 1994. https://doi.org/10.1109/ICPR.1994.576879.

[13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition", Proceedings of the IEEE, Vol. 86,

No. 11, pp. 2278-2324, Nov. 1998. https://doi.org/10.1109/5.726791.

[14] Yann LeCun, Corinna Cortes, Chris Burges. "MNIST handwritten digit database", 1998. http://yann.lecun.com/exdb/mnist/index.html [accessed : Nov. 29, 2023]

[15] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-iid data", arXiv:1806.00582, Jun. 2018. https://doi.org/10.48550/arxiv.1806.00582.

[16] L. Cao, "Non-IIDness Learning in Behavioral and Social Data", The Computer Journal, Vol. 57, No. 9, pp. 1358-1370, Sep. 2014. https://doi.org/10.1093/comjnl/bxt084.

[17] S. Kim, H. Lee, J. Bang, S. Hong, and H. Kim, "A Study on Probabilistic Sampling Techniques Based on Data Distribution Estimation to Solve Federated Learning Statistical Heterogeneity Problems", Proceedings of Symposium of the Korean Institute of communications and Information Sciences, Vol. 46, No. 11, pp. 1941-1949, Nov. 2021. http://dx.doi.org/10.7840/kics.2021.46.11.1941.

[18] H. Lee, S. Hong, J. Bang, and H. Kim, "Study of Optimization Techniques to Apply Federated Learning on Class Imbalance Problems", The Journal of Korean Institute of Information Technology, Vol. 19, No. 1, pp. 43-54, Jan. 2021. http://dx.doi.org/10.14801/jkiit.2021.19.1.43.

[19] H. Mohamed, "Assessment of Artificial Neural Network for bathymetry estimation using High Resolution Satellite imagery in Shallow Lakes: Case Study El Burullus LakeAssessment of Artificial Neural Network for bathymetry estimation using High Resolution Satellite imagery in Shallow Lakes", International water technology conference, pp. 12-14, Mar. 2015.

[20] D. J. Beutel, et al., "Flower: A Friendly Federated Learning Research Framework", arXiv:2007.14390, Jul. 2020. https://doi.org/10.48550/arXiv.2007.14390

저자소개

이 주 원 (Joowon Lee)

2017년 ~ 2023년 : 강원대학교 정보통신공학전공(공학사)
2023년 ~ 현재 : 강원대학교 컴퓨터정보통신공학과 석사과정
관심분야 : 빅데이터, 기계학습, 인공지능

전 석 환 (Sukhwan Jeon)

2022년 : 강원대학교 기계의용공학과(공학사)
2022년 ~ 현재 : 강원대학교 컴퓨터정보통신공학과 석사과정
관심분야 : 기계학습, 딥러닝, 빅데이터

방 준 일 (Junil Bang)

2018년 9월 ~ 2020년 8월 : 강원대학교 컴퓨터정보통신공학과(공학석사)
2020년 9월 ~ 현재 : 강원대학교 컴퓨터정보통신공학과 박사과정
관심분야 : 데이터분석, 머신러닝

김 화 종 (Hwajong Kim)

1984년 3월 : KAIST 전기및전자공학과(공학석사)
1988년 3월 : KAIST 전기및전자공학과(공학박사)
1988년 3월 ~ 현재 : 강원대학교 컴퓨터정보통신공학과 정교수
관심분야 : 데이터 통신, 컴퓨터네트워크, 네트워크 프로그래밍, 빅데이터