블록체인 기반 연합학습을 위한 레퍼런스 아키텍처

*고은수 **문종현 *이광기 **손채봉
*이노피아테크 **광운대학교

A Reference Architecture for Blockchain-based Federated Learning

*Goh, Eunsu **Mun, Jong-Hyeon ***Lee, Kwang-Kee ***Sohn, Chae-bong
*Innopiatech **Kwangwoon University

요약

연합학습은, 데이터 샘플을 보유하는 다수의 분산 에지 디바이스 또는 서버들이 원본 데이터를 공유하지 않고 기계학습 문제를 해결하기 위해 협력하는 기술로서, 각 클라이언트는 소유한 원본 데이터를 로컬모델 학습에만 사용함으로써, 데이터 소유자의 프라이버시를 보호하고, 데이터 소유 및 활용의 파편화 문제를 해결할 수 있다. 연합학습을 위해서는 통계적 이질성 및 시스템적 이질성 문제 해결이 필수적이며, 인공지능 모델 정확도와 시스템 성능을 향상하기 위한 다양한 연구가 진행되고 있다.

최근, 중앙서버 의존형 연합학습의 문제점을 극복하고, 데이터 무결성 및 추적성과 데이터 소유자 및 연합학습 참여자에게 보상을 효과적으로 제공하기 위한, 블록체인 융합 연합학습기술이 주목받고 있다. 본 연구에서는 이더리움 기반 블록체인 인프라와 호환되는 연합학습 레퍼런스 아키텍처를 정의 및 구현하고, 해당 아키텍처의 실용성과 확장성을 검증하기 위하여 대표적인 연합학습 알고리즘과 데이터셋에 대한 실험을 수행하였다

1. 서론

대규모 데이터의 활용은 인공지능 기술 발전의 기본 조건이지만 프라이버시 보호, 보안 등의 이유로 데이터의 공유 및 활용이 제한적인 분야가 존재하며, 이에 대한 대안으로 데이터의 직접적인 공유 없이 분산 환경에서 기계학습을 수행할 수 있는 연합학습(Federated Learning) 기술이 활발히 개발되고 있다. 연합학습을 통해 데이터 및 인공지능 기술 활용의 폭발적 확장을 기대할 수 있으며, 파편화 되어있는 데이터 소유 및 활용 주체 간의 협력을 활성화하여 산업 내, 산업 간 융합을 촉진할 수 있다.

연합학습은 다수의 클라이언트에 분산된 학습데이터를 사용하여 심층신경망(DNN: Deep Neural Networks)을 포함한 기계학습 모델을 훈련시킬 수 있는 학습 패러다임으로, 각 로컬 클라이언트가 수집한 원본 데이터를 클라이언트 간의 공유 또는 중앙 서버로 전송하지 않고, 로컬모델 학습에만 사용함으로써 데이터 소유자의 프라이버시를 보호하고, 궁극적으로 데이터 소유 및 활용의 파편화 문제를 해결할 수 있다. 연합학습은 모든 로컬 데이터 세트가 하나의 서버에 공유되는 전통적인 중앙집중식 기계학습 방식 혹은 로컬 데이터 샘플의 동일한 분포(identically distributed)를 가정하는 전통적인 분산접근 방식과는 대비되며, 다수의 다양한 디바이스, 동적 환경 및 시공간으로부터 수집된 데이터가 독립동일분포(iid: independent identically distributed) 조건을 만족하지 못하고 비균일 및 불균형 특성을 지니는 통계적 이질성 문제(Statistical Heterogeneity)와 연합학습에 참여하는 디바이스의 성능과 기능 및 네트워크 환경이 다양하고, 디바이스의 추가, 변동이 지속적으로 발생하는 시스템적 이질성 문제(SystemHeterogeneity)의 해결이 필수적이다.[1]

최근, 연합학습 클라이언트 및 데이터 샘플의 평가 및 선정(Client/Sample Evaluation andSelection)을 통한 모델 정확도와 시스템 성능 향상[2][3], 악의적 참여자 및 가짜 데이터(Malicious clients and false data)를 걸러 내기 위한 다양한 연구가 진행되고 있으며, 특히 중앙서버 의존형 연합학습의 문제점인 단일 장애점 문제(single point of failure problem)를 극복하고, 데이터 무결성 및 추적성(integrity, traceability)과 연합학습 데이터 소유자 및 참여자에게 보상을 제공하기 위하여 IoT, 헬스케어 등의 응용분야를 중심으로 블록체인 융합기술이 활발히 개발 및 적용되고 있다. [4]

본 연구에서는 이더리움 기반 블록체인 인프라와 호환되는 연합학습 레퍼런스 아키텍처를 정의 및 구현하고, 해당 아키텍처의 실용성과 확장성을 검증하기 위하여 대표적인 연합학습 알고리즘과 데이터셋에 대한 실험을 수행하였다. 블록체인 기반 연합학습의 구성 요소로, 스마트 컨트랙트와 모델 평가자(Evaluator) 및 학습 참여자(Trainer)를 정의 하였으며, 연합학습 참여자 간의 학습모델 공유를 위하여 IPFS를 활용하였다. 기존 중앙 집중형 연합학습에서는, 중앙 서버의 제어 하에 로컬 노드의 선정과 각 라운드(round) 마다의 학습결과가 취합되었지만, 블록체인 기반 연합학습에서는 스마트 컨트랙트를 통해 연합학습의 평가와 학습이 관리된다. [5]

본 연구에서는 디바이스-교차(Cross-device) 연합학습 시나리오와 사일로-교차(Cross-silo) 시나리오를 토대로 블록체인 기반 연합학습 레퍼런스 아키텍처를 구현하였으며, 디바이스-교차 시나리오에 대해 우선 검증을 시행하였다.

2. 레퍼런스 아키텍처

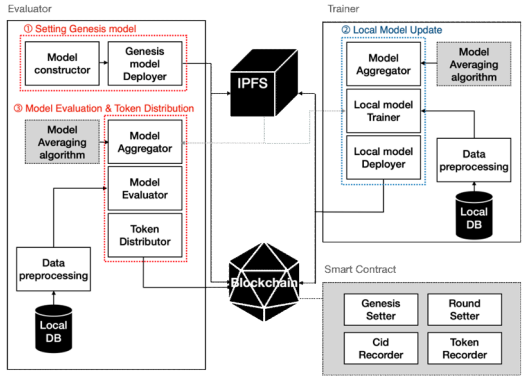

그림 1 연합학습 Cross-Device 시나리오를 위한 블록체인 기반의 전체 레퍼런스 아키텍처

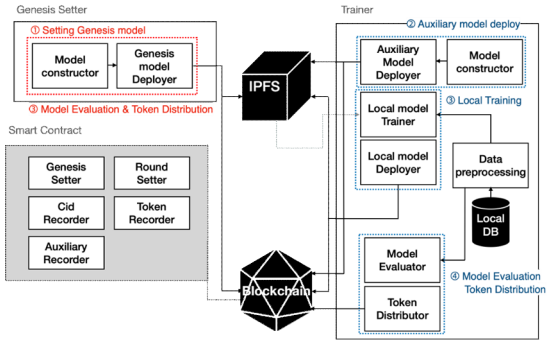

그림 2 연합학습 Cross-Silo 시나리오를 위한 블록체인 기반의 전체 레퍼런스 아키텍처

그림 1은 디바이스-교차 시나리오의 전체 아키텍처이다. 평가자(Evaluator)는 제네시스 모델 생성과 라운드 별로 업데이트된 로컬 모델들을 집계하여 글로벌 모델을 만들고, 로컬 모델들의 라운드 별 평가를 담당한다. 평가가 끝난 뒤 Evaluator는 각 클라이언트들의 기여도를 바탕으로 토큰을 배부한다. 로컬 학습 참여자(Trainer)는 자신의 로컬 저장 공간에 있는 개인적인 데이터셋을 활용해 로컬 모델 학습을 진행한다. 이 때 Evaluator와 Trainer는 절대 서로 직접 소통하지 않으며, 각각 스마트 컨트랙트의 함수를 호출해서 데이터를 등록한다.

① 은 Evaluator 디바이스에서 이루어지는 Genesis model setting 과정을 의미한다. 제네시스 모델은 Model Constructor에 에 의해 제공되며 Genesis model deployer에 의해 IPFS에 먼저 업로드 된다. 이 때 제네시스 모델의 cid(content identifier)를 반환 받게 되며, 이것을 이더리움 스마트 컨트랙트에 등록한다.

② 는 Trainer에서 이루어지는 로컬 학습 진행 단계이다. 본 논문에서 Model averaging algorithm은 Google에서 제안된 FedAvg[6]가 사용되었다. Trainer들은 본인의 학습용 데이터를 Data preprocessing 모듈을 사용해 전처리를 진행한다. 전처리된 데이터는 Local model Trainer 모듈에서 학습용으로 사용된다. Local model Trainer는 지정된 하이퍼 파라미터에 따라 학습을 진행한다. 로컬 학습이 끝나면 Genesis model setting 과정과 마찬가지로 IPFS에 학습된 모델을 업로드하고, 스마트 컨트랙트에 이를 기록한다.

③ 은 Model Evaluation 및 Token distribution 과정이다. 먼저 Model Aggregator는 FedAvg알고리즘을 수행하여 해당 글로벌 라운드의 글로벌 모델을 생성한다. Trainer들이 수행한 로컬 학습을 평가하기 위해 Evaluator는 검증된 평가용 데이터를 가지고 있다고 가정한다.[7] 이러한 평가용 데이터는 Data preprocessing을 통해 전처리한 뒤, Model Evaluator에 의해 평가 데이터로서 사용된다. 평가는 글로벌 모델의 손실(loss)과 로컬 모델의 손실의 차로써 기여도 점수로 계산되고, 이를 기반으로 Token Distributor에 의해 학습에 참가한 Trainer들에게 토큰을 배부한다. Token Distributor는 각각의 클라이언트의 기여도 점수에 따라 비례하게 토큰을 분배하며, 이 결과를 스마트 컨트랙트에 기록한다. 이 때 기여도가혹은 음수로 측정된다면, 해당 클라이언트는 토큰을 수여할 수 없다.

그림 2는 사일로-교차 시나리오의 전체 아키텍처이다. 사일로-교차 시나리오에서 각각의 조직은 디바이스-교차 시나리오를 따르고 있다.

① 은 SettingGenesis model 과정을 의미한다. 디바이스-교차 시나리오에서 Evaluator가 Genesis model을 배포했던 점과는 달리, 사일로-교차 시나리오는 따로 Evaluator가 정해져 있지 않기 때문에 Genesis Setter는 genesis model을 배포하는 것 그 이상의 작업은 하지 않는다.

②는 Trainer들이 각각 진행하는 Auxiliary model deploy단계이다. 사일로-교차 시나리오에서는 각각의 Trainer들이 보조 모델(Auxiliary model)을 배포하고, 이 보조 모델들에 대해서 자신이 배포한 보조 모델의 평가자 역할을 맡게 된다.

③은 Trainer들의 로컬 학습 단계이다. Trainer들은 본인이 평가자로서 참가한 보조 모델이 아닐 경우에만 학습에 참여하며, 디바이스-교차 시나리오와 마찬가지의 방법으로 학습을 진행한다.

④는 모델 평가 및 토큰 배부 단계이다. Trainer들은 본인이 평가자로서 참가한 보조 모델에 대해서 평가를 진행한다. 이 때 사용되는 평가용 데이터셋은 ③에서 사용한 데이터와 동일하다.

3. 실험 결과

본 장에서는 디바이스-교차 시나리오에 대해서 우선 단독으로 검증을 시행했으며, 먼저 독립동일분포(iid:independent identically distributed) 데이터셋에 대하여 기초적인 딥 러닝 태스크를 제안된 프레임워크가 수행한 결과를 확인하고, 실제 연합학습 환경과 가까운 Non-iid 데이터셋에서도 마찬가지로 모델 학습 성능 결과를 확인한다.

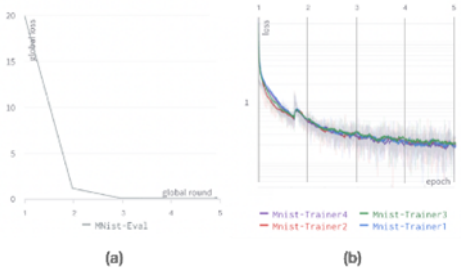

그림 3 MNIST 데이터셋에 대한 (a)global loss 그래프와 (b)local loss 그래프

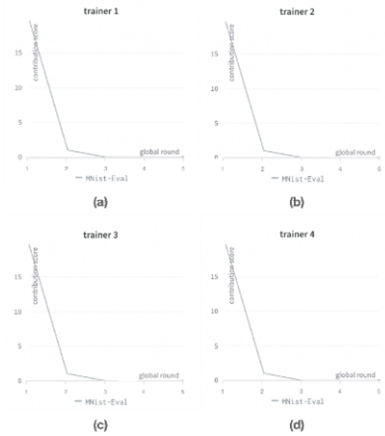

그림 4 MNIST 데이터셋 학습에 참여한Trainer들의 기여도 그래프 (a)Trainer1 (b)Trainer2 (c)Trainer3 (d)Trainer 4

그림3 은 iid 데이터 분포를 가지는 MNIST 데이터셋을 사용한

CNN 분류 네트워크의 (a)global loss 및 (b)local loss 그래프이다. 연합학습은5개의 global round, 2개의 local epoch로 진행되었다. (a)에서 알 수 있듯이, global loss의 경우 global round가 진행됨에 따라 점차 감소하여 대략0.1에 수렴하는 양상을 보였다. (b)에서는 누적 에포크에 따른local loss를 확인할 수 있다. 큰 세로 구분선은 global round를 의미한다. 네 Trainer들 모두 20.0 이상의 loss로부터 5개의 global round 이후 local loss가 대략 0.2로 수렴하는 모습을 보였다.

그림 4는 MNIST 데이터셋 학습에 참여한 4인의 Trainer들의 기여도 점수 그래프이다. 기여도의 경우 global loss 대비 참가자들의 라운드별 local loss의 차로 계산된다. 4인 모두 초반 2라운드에서 양의 기여도를 기록하였지만, 3라운드부터는 학습 모델이 수렴함에 따라 0에 가까운 기여도를 보였다.

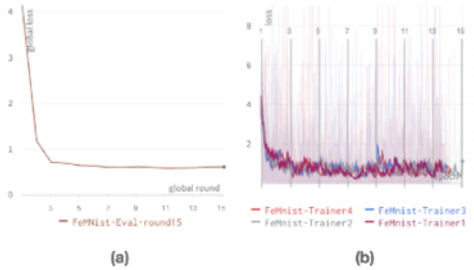

그림 5 FEMNIST 데이터셋에 대한
(a)global loss 그래프와 (b)local loss 그래프

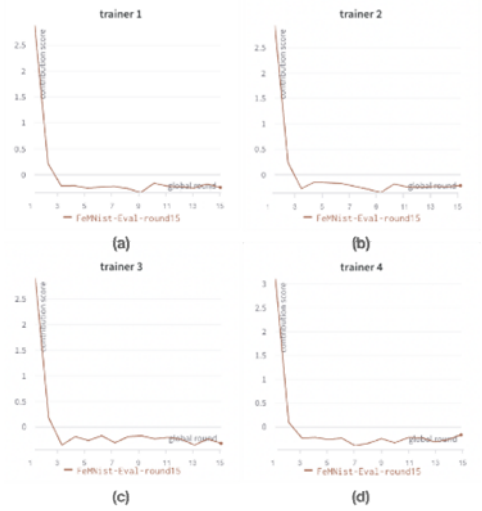

그림 6 FEMNIST 데이터셋 학습에 참여한Trainer들의 기여도 그래프
(a)Trainer1 (b) Trainer 2 (c) Trainer 3 (d)Trainer 4

그림 5는 non-iiid 데이터셋에 해당하는 FEMNIST 데이터셋을 사용하여 진행된 학습 결과이다. 네트워크는 그림4, 그림5에서 사용된 CNN 네트워크와 동일한 모델이 사용되었다. 15개의 global round, 2개의 local epoch로 진행되었다. (a)는 global loss로서, 1라운드에서 4.15를 기록하였으며 마지막 라운드에서의 global loss는 대략 0.61로 수렴하였다. (b)는 누적 에포크에 따른 local loss 그래프이다. 마찬가지로 4인의 Trainer들에 대해 수행되었으며, 15개의 global round가 진행됨에 따라 네 개의 loss값도 수렴하는 양상을 보였다.

그림 6은 FEMNIST 데이터셋 학습에 참여한 4인의 Trainer들의 기여도 그래프이다. MNIST 데이터 실험과 마찬가지로, 초반2개의 global round 까지 양의 기여도를 기록하다가, global loss가 수렴해감에 따라 음의 기여도, 즉 global model보다 개선되지 못한 모습을 보였다.

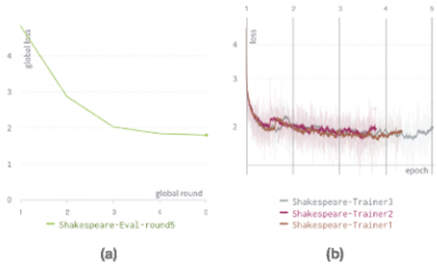

그림 7 Shakespeare 데이터셋에 대한
(a) global loss 그래프와 (b) local loss 그래프

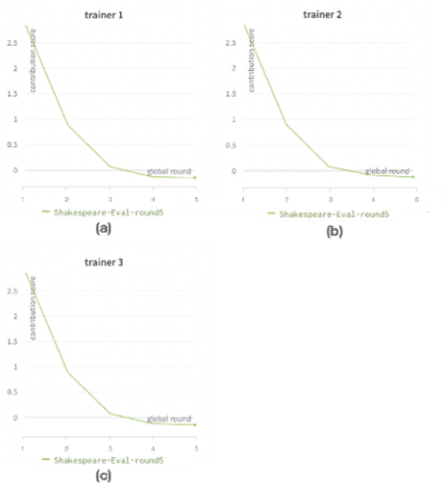

그림 8 Shakespeare 데이터셋 학습에 참여한 Trainer들의 기여도 그래프
(a)Trainer1 (b)Trainer 2 (c)Trainer 3

그림 7은 non-iid 데이터셋에 해당하는 Shakespeare 데이터셋과 LSTM(Long Short-Term Memory) 모델을 사용하여 진행된 학습 결과이다. 5개의 global round, 5개의 local epoch로 진행되었다. (a)는 global loss로서, global round가 진행됨에 따라 수렴하는 양상을 보였다. (b)는 누적 에포크에 따른 local loss 그래프로서, 마찬가지로 global round가 진행됨에 따라 수렴하는 모습을 보이고 있다.

그림 8은 Shakespeare 데이터셋 학습에 참여한 3인의 Trainer들의 기여도 그래프이다. Shakespeare 데이터셋 또한 이전 실험들과 마찬가지로 global loss가 수렴함에 따라 점차 Trainer들의 기여도가 감소하는 양상을 보였다.

4. 결론

본 연구에서는 블록체인 기반 연합학습을 위한레퍼런스 아키텍처를 설계, 구현하였으며, 디바이스-교차 시나리오에 대해서 iid 데이터셋 및 non-iid 데이터셋에 대한 실험을 통하여 해당 아키텍처가 실용성, 확장성 면에 있어서 유효함을 입증하였다. 실험 결과 iid 데이터셋에 대해서 안정적으로 학습이 진행되었음을 미루어보아 기초적인 딥러닝 작업에 대하여 적용이 가능함을 알 수 있었고, 이를 토대로 제안한 아키텍처가 실제 연합학습에 더 적합한 데이터인 non-iid 데이터셋에서도 충분히 활용이 가능함을 입증하였다. 모델 평가자(Evaluator)가 검증용/평가용 데이터를 보유하고 있다는 가정 하에, 디바이스-교차(Cross-device) 연합학습 시나리오 및 사일로-교차(Cross-silo) 연합학습 시나리오의 레퍼런스 아키텍처를 제안하였다.

로컬 학습 참여자(Trainer)들은 각 라운드의 학습 기여도에 따라 보상을 받게 되며, 학습에 고의적으로 악영향을 미치는 참가자에 대한 처벌 체계도 추가로 고려할 수 있다. 또한 클라이언트 및 데이터의 평가, 선정 그리고 보상 메커니즘 등의 기능을 추가로 개발하여 사용자 개인 기기의 데이터를 활용한 방송 컨텐츠 추천 시스템 등에 활용할 수 있도

록 연구를 확장할 계획이며, 연합학습 수명주기관리 MLOps[8]와의 연계를 통해 활용도를 넓혀갈 예정이다.

ACKNOWLEDGEMENTS

이 논문은2022년도 정부의 재원으로 한국과학기술일자리진흥원의 지원을 받아 "디지털 헬스케어를 위한 블록체인 융합 원격임상시험 서비스 개발" 과제로서 수행된 연구임

참 고 문 헌(References)

[1] Kwangkee, Lee et al."동적인 디바이스 환경에서적응적 연합학습 기술"IT지식포털 주간기술동향2052호. (2022)
[2] Lai, F., Zu, X.,Madhyastha, H. V., and Chowdhury, M., "Oort: Efficient federated learning viaguided participant selection.", 15th USENIX Symposium on Operating SystemsDesign and Implementation, pp. 19-35, July 2021
[3] Jaemin Shin,Yuanchun Li, Yunxin Liu, and Sung-Ju Lee., "FedBalancer: data and pace controlfor efficient federated learning on heterogeneous clients.", The 20th AnnualInternational Conference on Mobile Systems, Applications and Services., pp.436-449, June 2022
[4] Wang, Zhilin, and QinHu. "Blockchain-based federated learning: A comprehensive survey.", 2021,arXiv preprint arXiv:2110.02182.
[5] S. K. Lo et al.,"Towards Trustworthy AI: Blockchain-based Architecture Design forAccountability and Fairness of Federated Learning Systems,", IEEE Internetof Things Journal, 2022.
[6] McMahan, Brendan, etal. "Communication-efficient learning of deep networks from decentralizeddata." Artificial intelligence and statistics. PMLR, 2017.
[7] Cai, Harry, DanielRueckert, and Jonathan Passerat-Palmbach. "2cp: Decentralized protocols totransparently evaluate contributivity in blockchain federated learningenvironments." arXiv preprint arXiv:2011.07516 (2020).
[8] Seung-hoo Hong andKang-yoon Lee "The Study on the Implementation Approach of MLOps on FederatedLearning System", Journal of Internet Computing and Services, Vol. 23, No. 3, pp. 97-110, Jun. 2022