열악한 환경에서도 효과적인 블록체인 기반의 탈중앙화 연합학습 플랫폼

김민재 박상현○ 김재윤 문수묵

서울대학교 전기·정보공학부

alswo901@snu.ac.kr, {lukepark, jaeykim}@altair.snu.ac.kr, smoon@snu.ac.kr

Robust Decentralized Federated Learning Platform based on Blockchain

Min-Jae Kim Sanghyeon Park○ Jae-Yun Kim Soo-Mook Moon

Department of Electirical and Computer Engineering, Seoul National University

요 약

본 연구에서는 중앙 서버 없이 각 로컬 장치들이 서로를 평가할 수 있는 블록체인 기반 탈중앙화 연합학습 플랫폼을 제안한다. MNIST 데이터셋을 파레토 분포에 따라 각 로컬 장치들에게 Non-IID한 방식으로 분배함으로써 서로 다른 분포의 데이터를 가지는 평가 환경을 구성했다. 실험 결과 제안하는 플랫폼을 통해 참여자들이 상호 평가 및 투표할 수 있었으며, 각자에게 이득이 되는 독립된 모델로 수렴함을 확인할 수 있었다.

1. 서 론

연합학습(Federated learning)은 학습에 필요한 데이터들을 로컬 장치(local device)에 그대로 놓아둔 채, 학습된 결과물을 서버와 로컬 장치 간에 주고받으며, 정확도가 높은 글로벌 모델을 만드는 것을 목표로 한다. 연합학습은 데이터의 프라이버시(privacy)의 중요성과 로컬 장비의 성능이 향상됨에 따라 최근 주목받고 있지만, 각 로컬 장치들이 가지고 있는 데이터들의 분포가 편향될 수 있는 등 해결해야 할 문제들이 여전히 남아있다[1][3].

본 연구는 노드간 상호 평가 및 투표를 통해 로컬 장치들이 가지고 있는 데이터들이 서로 다른 분포를 가짐으로써 발생하는 문제를 해결하고자 한다. 로컬 장치가 가지고 있는 데이터들의 다양한 분포를 시뮬레이션하기 위해, 노드들에게 0~9까지의 손글씨 데이터셋인 MNIST을 파레토 분포(Pareto distribution)에 따라 나누어주었다. 또한, 로컬 장치들이 가지고 있는 연합학습을 방해하고자 하는 비잔틴 노드(Byzantine node)의 존재를 포함하기 위해, 본 연구에서는 레이블링(labeling)이 잘못된 데이터를 가지고 있는 비잔틴 노드들의 비율을 30%로 상정하였다. 종래의 서버를 이용한 연합학습의 일종인 Federated Averaging [2] 방법을 위 환경에 적용할 경우 그림 1에서

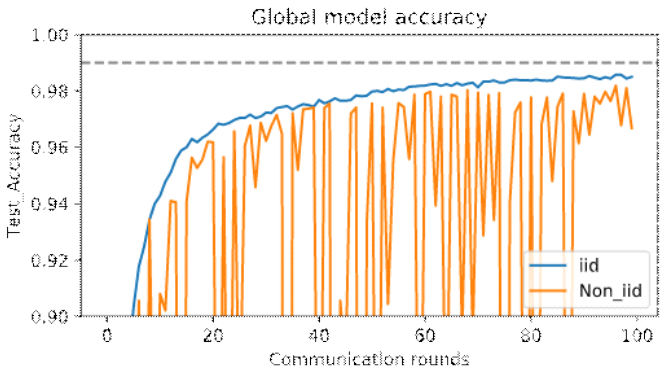

그림 1 Federated Averaging의 결과

* 이 논문은 BK21플러스 사업 및 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No.2020R1A2B5B02001845)

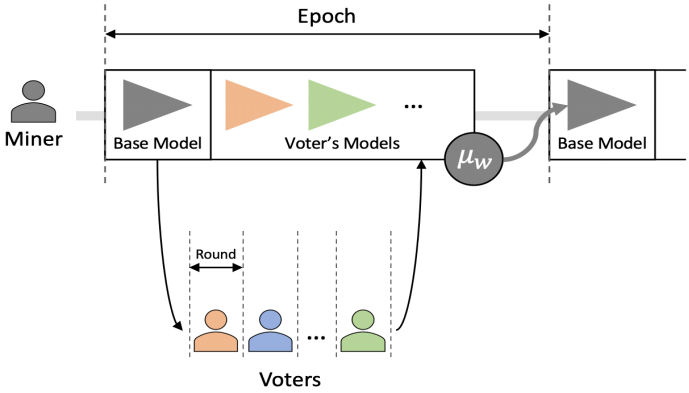

그림 2 블록체인 기반의 연합학습 플랫폼

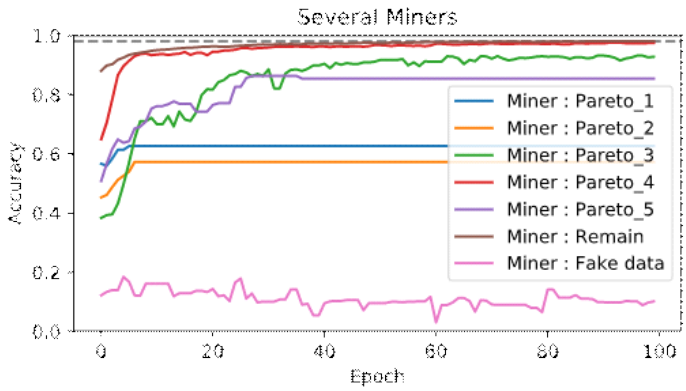

그림 3 매 에폭마다 생성된 모델의 정확도

처럼 글로벌 모델의 수렴성이 나빠진다. 본 연구에서는 불균일한 데이터 분포 환경을 다룰 수 있도록 블록체인 기반의 탈중앙화 연합학습 플랫폼을 제안함으로써 다음 두 목표를 이루고자 한다. 첫째, 각 노드가 가지고 있는 데이터들이 불균일하고 심지어 거짓된 데이터가 섞여 있어도 글로벌 모델의 수렴성을 유지한다. 둘째, 비슷한 분포의 데이터를 가지고 있는 노드들이 자신들의 데이터에 적합한 독자적인 글로벌 모델을 가진다.

2. 블록체인 기반 연합학습 플랫폼

본 연구에서는 알고리즘 1과 같은 블록체인 기반의 연합학습 플랫폼(Blockchain-based Federated Learning Platform)을 제안한다.

알고리즘 1: 블록체인 기반의 연합학습 알고리즘

Input: Data X_t from $t = 1, \ldots, n$ nodes and selected h Miners
for Epoch $e = 0, 1, \ldots$ **do**
  for m **in parallel over** h Miners **do**

$$w_m^{(e)} = \frac{1}{\sum Reputation} \sum_{tx \in transactions} Reputation_{sender} * w_{tx}$$

    Miners evaluate every transaction with their own data and make new Block with Base Model $w_m^{(e)}$
  end
  for round $r = 0, 1, \ldots$ **do**
    i) Randomly choose a node t
    ii) Node t evaluate and vote k miners with X_t
    iii) Download base model of miners node t voted for
    iv) Train base models and upload results as a transaction
  end
end

1) 라운드마다 n개의 노드 중 하나가 랜덤하게 선택된다.

2) 선택된 노드는 각 마이너가 제시한 최신 블록들의 베이스 모델(base model)을 받아, 자신이 가지고 있는 데이터에 대해 높은 정확도를 보이는 k개의 블록을 선택(Voting)한다.

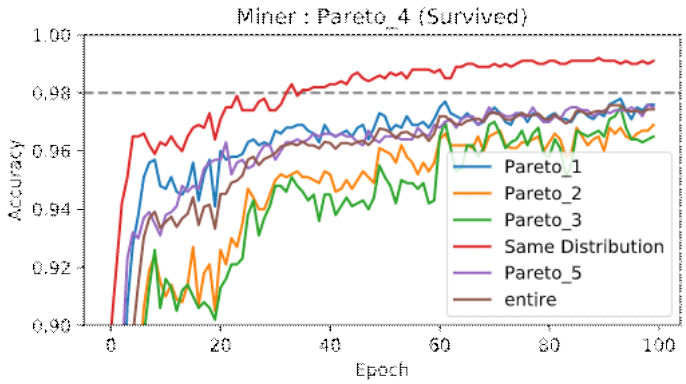

그림 4 특정 마이너가 생성한 모델의 정확도

3) 투표한 블록의 모델을 자신의 데이터로 학습한 결과물을 트랜잭션(transaction)의 형태로 해당 블록에 업로드 한다.

4) 에폭마다 마이너들은. 블록의 트랜잭션들에 기록된 모델들을 자신이 가지고 있는 데이터로 평가하여, 해당 트랜잭션을 송신한 노드의 평점(Reputation)을 조절한다.

5) 마이너들은 4)에서 누적된 평점에 따라 블록의 트랜잭션들에 기록된 모델들의 가중치(weight)를 가중평균 내어 새로운 베이스 모델을 만들고, 이를 기반으로 하는 새로운 블록을 생성한다. 그러나 새로운 베이스 모델의 정확도가 기존의 베이스 모델에 비해 현저히 낮을 경우 기존 모델을 유지한다.

3. 실험 및 평가

편향된 데이터 분포와 비잔틴 노드가 존재하는 상황을 실험하기 위해 다음과 같은 상황을 상정하였다. 총 100개의 노드는 7가지로 분류되는데, 7가지 범주별로 하나씩 마이너를 가진다. 7가지 종류 중 5가지에 해당하는 노드들은 각기 서로 다른 파레토 분포를 따르는 MNIST 데이터들을 가진다. 다른 한 종류의 노드들은 거짓된 데이터들을 가지고 있는 비잔틴 노드들이며, 나머지 한 종

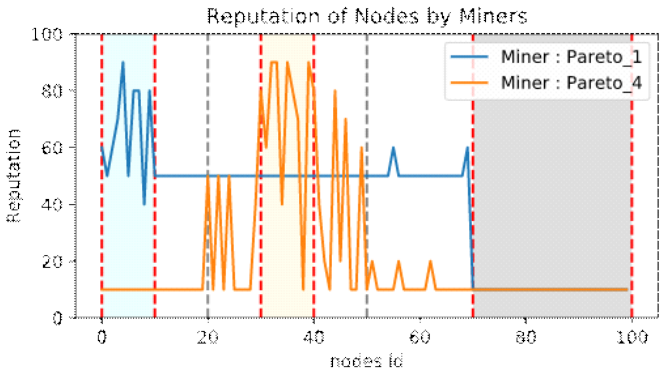

그림 5 마이너가 노드들을 평가한 평점

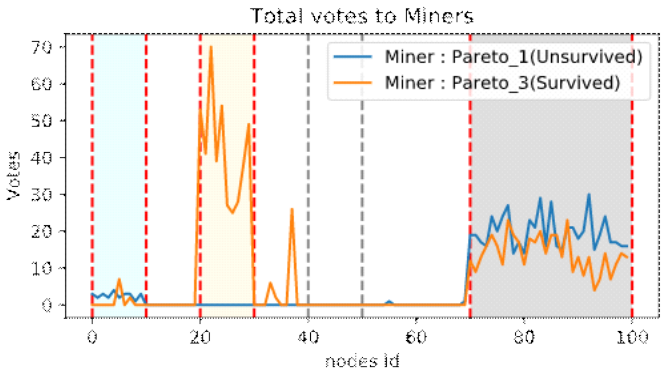

그림 6 마이너가 노드들로 받은 투표수

류는 상대적으로 균일하게 분포된 '나머지 (remain) 데이터'를 가진다. 본 실험에서는 2개의 합성곱 레이어(Convolution layer)와 2개의 전연결 레이어(Fully Connected layer)로 구성된 CNN(Convolutional Neural Network) 모델을 사용했다.

3.1 마이너들에 의해 생성된 다양한 모델

그림 3의 결과는 7명의 마이너가 매 에폭(50 라운드)마다 생성하는 새 모델의 정확도를 전체 테스트 데이터셋으로 측정한 결과이다. 실험 결과, 초기 라운드에서 높은 성능을 가진 모델을 가진 마이너들은 노드들로부터 꾸준히 선택받은 데 비해, 초기 모델의 성능이 좋지 않은 마이너들은 이어지는 라운드들에서 선택받지 못해 성능이 개선되지 못했다. 그림 4는 7개의 마이너 중 비교적 많은 투표를 받은 'Miner : Pareto_4'가 매 에폭마다 만드는 새 모델의 정확도를 서로 다른 파레토 분포를 따르는 테스트 데이터셋에 대해 측정한 결과이다. 마이너가 가지고 있는 데이터의 분포와 같은 분포의 데이터셋에 대한 정확도가 다른 분포의 데이터셋에 비해 높음을 확인할 수 있다.

3.2 노드와 마이너의 상호 평가

각 노드와 마이너가 서로를 평가한 결과가 이를 뒷받침한다. 그림 5에서 마이너는 본인이 가지고 있는 데이터와 같은 분포를 가지고 있는 노드들을 높게 평가하는 동시에 비잔틴 노드들(ID 70부터 100까지)은 낮게 평가한다. 이를 통해 마이너들은 비잔틴 노드들을 제외하면서도, 본인과 같은 분포를 가진 노드들은 적극적으로 학습에 사용하고 있음을 알 수 있다. 그림 6에서는 노드들이 마이너들을 평가한 결과를 비교하였는데, 꾸준히 선택을 받은 마이너가 다른 마이너들에 비해 많은 투표를 받았음을 관찰할 수 있다. 특히, 마이너들은 같은 분포를 가지는 노드들로부터 비교적 많은 투표를 받았음을 알 수 있다. 반면 비잔틴 노드들은 모든 마이너의 학습을 골고루 방해하려는 양상을 띤다.

4. 결론 및 향후 연구

본 연구에서는 블록체인 기반의 탈중앙화 연합학습 플랫폼을 고안하여 로컬 장치들이 본인이 가지고 있는 데이터를 이용해 서로를 평가하는 시스템을 도입하였다. 연합학습 과정에서 각 로컬 장치들은 서로에게 투표하고 평점을 매김으로써 비잔틴 노드를 최대한 제외하고 자신에게 유리한 데이터를 가진 노드들을 적극적으로 학습에 활용했다.

MNIST 데이터셋 분류의 경우, 하나의 적절한 글로벌 모델이 존재할 수 있다. 하지만, NLP(Natural Language Processing)와 같은 보다 다양한 과제의 경우, 모두를 위한 하나의 글로벌 모델보단, 사람들의 특성에 맞는 다양한 모델이 필요할 수 있다. 따라서 이미지 분류에서 더 나아가 MNIST 이외의 데이터셋에 대해서도 이러한 탈중앙화 방식의 연합학습 플랫폼이 유효한지 검증해보는 것이 앞으로의 연구 과제이다.

참 고 문 헌

[1] Li, Tian, et al. "Federated learning: Challenges, methods, and future directions." arXiv preprint arXiv:1908.07873 (2019).

[2] McMahan, H. Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." arXiv preprint arXiv:1602.05629 (2016).

[3] Zhao, Yue, et al. "Federated learning with non-iid data." arXiv preprint arXiv:1806.00582 (2018).