

# UE-BS 네트워크 지능 증강을 위한 연합학습

진민규<sup>1</sup>, 최진혁<sup>1</sup>, 이지훈<sup>1</sup>, 오승은<sup>1</sup>, 김성륜<sup>1</sup>

연세대학교 전기전자공학과<sup>1</sup>

{mgjin, jh.choi, jhlee, seoh}@ramo.yonsei.ac.kr, slkim@yonsei.ac.kr

## Enhancing UE-BS Network Knowledge through Federated Learning

Min-Gyu Jin<sup>1</sup>, Jin-Hyuk Choi<sup>1</sup>, Ji-Hoon Lee<sup>1</sup>, Seung-Eun Oh<sup>1</sup>, Seong-Lyun Kim<sup>1</sup>

School of Electrical & Electronic Engineering, Yonsei Univ.<sup>1</sup>

### 요약

머신러닝과 6G 기술이 발전함에 따라, User devices는 Deep Neural Network (DNN) task를 수행하기 시작한다. 해당 task를 수행하기 위해서 각 user는 자신이 가지고 있는 데이터셋으로 로컬 모델을 학습한다. 하지만, 컴퓨팅 자원의 부족 등으로 각 user는 방대한 데이터셋을 가지고 학습하기에는 한계가 있다. 따라서 다른 user의 knowledge를 공유하기 위해 하나의 Base Station (BS)를 두고, User들이 로컬 모델의 파라미터를 공유하는 Federated Learning (FL)이 많이 연구되고 있다. 이때 각 로컬 데이터셋의 양만을 고려하여 weight를 결정하는 Federated Averaging (FedAvg) 기법은 일반적으로 성능이 뛰어나지만, non-i.i.d 데이터셋에서는 취약함을 드러낸다. 따라서 본 논문에서는 non-i.i.d 데이터셋에서도 User-Basestation (UE-BS) 네트워크의 지능 증강을 효과적으로 달성하기 위해 로컬 데이터셋의 엔트로피와, 지니불순도를 고려한 FL 기법을 제안하고, 해당 기법이 FedAvg 기법보다 지식 증강을 효과적으로 달성할 수 있음을 입증한다.

### I. 서론

머신러닝과 6G 기술이 발전하면서, User devices는 Deep Neural Network (DNN) task를 수행한다. 이를 위해 각 유저는 자신이 수집한 로컬 데이터셋으로 모델을 학습하고 task를 처리한다. 하지만, 일반적으로 각 user device의 컴퓨팅 자원은 부족한 경우가 많기에 방대한 데이터셋으로 모델을 학습하는 것이 어렵고, 이에 따라 성능이 나빠질 수 있다. 이를 해결하기 위해 방대한 데이터셋을 직접 학습하지 않고도, 다른 유저의 knowledge를 sharing 할 수 있는 기법인 Federated Learning (FL)이 활발히 연구되고 있다 [1]. FL은 유저 각각의 데이터셋으로 각각 로컬 모델을 학습한 후 학습의 결과인 로컬 모델 파라미터를 BS로 전송하고, BS에서 다시 글로벌 파라미터를 계산하여 모든 유저에게 다시 공유하는 것이다. 이때, BS에서는 각 로컬 데이터셋의 양을 weight로 주어서 로컬 파라미터들의 평균을 내는 기법인 FedAvg 기법이 자주 이용되고 있다 [2]. 하지만 해당 기법은 로컬 데이터셋이 non-i.i.d인 경우 글로벌 모델이 잘 수렴하지 않는, 즉 UE-BS 네트워크의 지능 증강이 활발히 일어나지 않는 경우가 많다 [3]. 특히, 일반적으로 유저의 데이터셋은 유저만의 특성을 반영한 경우가 많으므로 non-i.i.d 분포를 따르는 경우가 많다.

이에 본 논문에서는 다수의 non-i.i.d 데이터셋을 가진 유저들과 BS가 FL을 진행할 때, 유저 데이터셋 분포의 엔트로피와 지니불순도를 고려한 FL 기법을 제안하고, 시뮬레이션을 통해 제안 기법이 네트워크의 지능증강과 높은 업무효율을 효과적으로 달성할 수 있음을 입증한다.

### II. 시스템 모델

본 논문에서는  $N$  개의 유저들이 각각의 데이터 셋을 가지고 로컬 모델

을 학습한 후 로컬 모델의 파라미터를 BS로 보내는 환경을 고려하였다. 이때,  $i$  번째 유저가 가지고 있는 데이터셋의 클래스의 개수는  $|D_i|$ 다. 또한 각 유저들은 BS에게 자신의 데이터셋 분포의 엔트로피와 지니불순도를 전송한다. 여기서 데이터셋 분포의 엔트로피는 데이터셋에 속한 데이터 클래스의 다양성을 의미한다. 즉 데이터셋 분포의 엔트로피 값이 클수록 더 다양한 데이터 종류를 가지고 있음을 의미하고, 다음과 같이 구할 수 있다.

$$H(D_i) = - \sum_{k=1}^{|D_i|} p(x_k) \log p(x_k) \quad (1)$$

여기서  $p(x_k)$ 는  $k$ 번째 클래스에 속한 데이터가 뽑힐 확률이다. 두 번째로 고려한 지니불순도는 데이터 클래스 분포가 얼마나 균일하게 분포되어 있는지를 측정하는 데 사용된다. 즉 지니 불순도 값이 클수록 데이터 클래스 분포가 균형잡혀있다는 것을 의미하고, 아래와 같이 구할 수 있다.

$$G(D_i) = 1 - \sum_{k=1}^{|D_i|} p(x_k)^2 \quad (2)$$

각 유저들이 모두 로컬 파라미터, 엔트로피, 지니불순도를 전송한 이후에 BS는 해당 정보를 가지고 글로벌 모델을 다음과 같이 업데이트하고 글로벌 파라미터를  $N$ 개의 유저들에게 다시 전송해준다.

$$w(t+1) = \sum_{i=1}^N c_i * w_i(t) \quad (3)$$

여기서  $w_i(t)$ 는  $i$ 번째 유저의  $t \in T$  번째 iteration의 로컬 파라미터이고  $w(t+1)$ 은  $t$ 번째 iteration이 끝난 후 업데이트된 글로벌 파라미터,

$\eta$ 는 학습률이다. 마지막으로  $c_i$ 는 weighted coefficient로 정규화된 엔트로피와 지니불순도의 weighted sum으로 다음과 같이 구하고 해당 식에서  $0 \leq \alpha \leq 1$ 을 만족한다.

$$c_i = \alpha * \frac{H(D_i)}{\sum_{k=1}^N H(D_k)} + (1 - \alpha) * \frac{G(D_i)}{\sum_{k=1}^N G(D_k)} \quad (4)$$

다양한 종류의 데이터를 가지고 로컬 모델을 학습한 유저가 적은 종류의 데이터를 가지고 학습한 유저보다 task를 위한 성능이 좋은 knowledge를 가지고 있을 가능성이 높다. 따라서 엔트로피가 높은 데이터셋을 가진 유저에게 더 높은 weight를 부여한다. 또한 데이터는 다양하더라도 하나의 종류의 데이터를 다른 종류의 데이터에 비해 엄청나게 많이 가지고 있다면 task를 처리하기 위한 성능이 좋지 않은 knowledge를 가질 수 없기에, 지니불순도 값이 높은 데이터셋을 가진 유저에게 더 높은 weight를 부여하여 글로벌 모델의 성능을 높여서 전체 네트워크의 지능 증강을 달성하고자 한다.

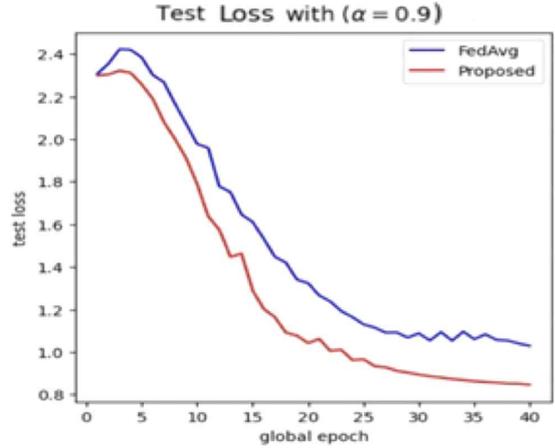
### III. 실험 및 성능평가

본 논문에서 제안한 기법의 성능을 평가하기 위해 7 개의 유저가 각자의 데이터셋을 가지고 하나의 BS과 함께 FL을 하는 환경을 구성하였다. 또한 데이터셋은 MNIST를 이용하였고 유저들의 non-i.i.d 환경을 위해 유저1 - 유저5는 한 종류의 데이터만 가지도록 하였고, 유저6과 유저7은 일곱 종류 이상의 데이터를 가지도록 구성하였다. 데이터셋의 양은 각 유저별로 20개에서 40개 사이로 랜덤하게 발생시켰다. FL에 참여하고 있는 유저와 BS는 ResNet 모델을 학습한다. 각 유저는 1 번의 global epoch을 하기 전 3번의 local epoch을 진행하고 총 40번의 global epoch을 진행하였다. 마지막으로 여러번의 실험을 통해  $\alpha = 0.9$ 로 설정하였다. 성능을 평가하기 위해 본 논문에서 제안한 기법을 이용했을 때와 FedAvg 기법을 이용했을 때의 모델의 Loss와 Accuracy를 비교한다.

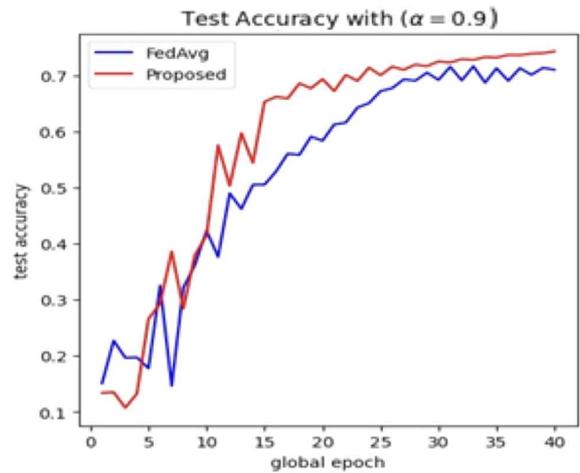
그림 1과 그림 2는 각각 모델의 Loss와 Accuracy를 비교한 그림이다. 글로벌 모델의 Loss는 본 논문에서 제안한 기법을 이용했을 때가 FedAvg 기법을 이용했을 때보다 항상 더 낮은 값을 가지는 것을 확인할 수 있다. 그림 2에서는 모델의 Accuracy를 나타낸 것이다. 우선 모델의 Accuracy가 수렴하기 전까지는 두 기법에서의 Accuracy가 더 높았다가 낮았다가를 반복한다. 하지만, 본 논문에서 제안한 모델의 Accuracy가 수렴하고 난 뒤 부터는 항상 FedAvg 기법보다 성능이 좋음을 확인할 수 있다. 수렴 또한 본 논문에서 제안한 기법에서는 약 Epoch = 15 일 때부터 하는 것을 확인할 수 있고, FedAvg 기법에서는 약 Epoch = 25 일 때부터 하는 것을 확인할 수 있다. 즉, 수렴 속도 또한 본 논문에서 제안한 기법에서 훨씬 빠른 것을 확인할 수 있다.

### IV. 결론

본 논문에서는 여러 개의 유저와 하나의 BS가 FL을 할 때, BS에서 글로벌 파라미터를 업데이트하는 기법을 제안했다. 데이터셋 분포의 엔트로피를 통해 데이터 종류의 다양성과, 데이터셋 분포의 지니불순도를 통해 데이터가 얼마나 균등하게 분포되어 있는지를 동시에 고려함으로써 non-i.i.d 데이터셋에서의 기존 FedAvg 방식보다 Accuracy가 좋아졌음을 보였다. 즉, 전체 네트워크의 지능 증강이 효과적으로 이루어졌다. 또한 Accuracy가 높은 곳으로 수렴하는데까지 걸리는 시간도 짧아졌는데, 이를 통해 향후 새로운 유저의 업무 효율을 증가시킬 수 있다.



(그림 1) FedAvg와 제안한 기법의 Test Loss 비교



(그림 2) 제안한 기법과 FedAvg 사이의 Test Accuracy 비교

### ACKNOWLEDGMENT

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2022R1A5A1027646 & No. 2023-11-1836).

### 참고 문헌

- [1] Z. Du, C. Wu, T. Yoshinaga, K. -L. A. Yau, Y. Ji and J. Li, "Federated Learning for Vehicular Internet of Things: Recent Advances and Open Issues," in IEEE Open Journal of the Computer Society, vol. 1, pp. 45-61, 2020, doi: 10.1109/OJCS.2020.2992630.
- [2] T. Sun, D. Li and B. Wang, "Decentralized Federated Averaging," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 4, pp. 4289-4301, 1 April 2023, doi: 10.1109/TPAMI.2022.3196503.
- [3] F. Sattler, S. Wiedemann, K. -R. Müller and W. Samek, "Robust and Communication-Efficient Federated Learning From Non-i.i.d. Data," in IEEE Transactions on Neural Networks and Learning

Systems, vol. 31, no. 9, pp. 3400-3413, Sept. 2020, doi:  
10.1109/TNNLS.2019.2944481.