

# 근사 복호 기반 연합학습의 시스템 지연 및 연산 오버헤드 분석

권정민, 박형곤\*

강원대학교, \*이화여자대학교

jungmin.kwon@kangwon.ac.kr, \*hyunggon.park@ewha.ac.kr

## Latency and Computational Overhead Analysis of Approximate Decoding - based Federated Learning

Jungmin Kwon, \*Hyunggon Park

Kangwon National University, \*Ewha Womans University

### 요약

근사 복호 기반 연합학습은 UDP 기반 전송 환경에서 오류 복원뿐 아니라, 모델 파라미터 중 주요 정보만을 선택적으로 전송 및 복호하여 전송 효율을 극대화하는 특징이 있다. 이러한 방식은 전체 파라미터를 전송하는 기존 UDP 기반 FL 대비 전송 데이터 양을 크게 줄이며, 통신 대역폭 및 에너지 사용을 절감한다. 다만, 복호화 연산 복잡도에 따라 학습 지연이 증가할 수 있다. 따라서 본 논문에서는 근사 복호 기반 FL의 전송 효율성과 시스템 오버헤드 간의 균형 관계를 분석한다.

### I. 서론

연합학습(Federated Learning, FL)은 다수의 단말이 데이터를 분산적으로 보유한 상태에서 로컬 학습을 수행하고, 그 결과를 서버로 전달하여 전역 모델을 업데이트하는 분산 학습 구조로 주목받고 있다. 그러나 무선 네트워크 환경에서 FL은 패킷 손실이나 비트 오류가 누적되어 FL 성능이 저하되는 문제가 발생한다. 이러한 문제 상황을 보완하기 위해 최근에는 오류 복원 기능을 포함한 내결함(fault-tolerant) 연합학습 기법이 활발히 연구되고 있으며, 특히 User Datagram Protocol(UDP) 기반 환경에서 동작하는 FL 시스템의 설계 및 최적화에 대한 연구가 집중적으로 이루어지고 있다 [1]–[3]. 근사 복호 기반 FL(Approximate Decoding - based FL, AD-FL)은 전송된 모델 파라미터 중 주요 정보만을 선택적으로 복호하여 전달함으로써 전송 효율을 크게 향상시키는 접근법으로 주목받고 있다 [4]–[5]. 이러한 방법은 모든 파라미터를 완전 복원하거나 재전송하지 않고, 학습에 기여도가 높은 중요 파라미터만을 복호하여 활용함으로써 통신 부하를 크게 줄인다.

한편, AD-FL은 통신 효율 측면에서 유리하지만, 선택적 복호화로 인하여 복호 연산 복잡도 증가라는 잠재적 한계를 내포한다. 즉, 복호화 연산을 위한 추가 계산이 시스템 지연(latency)을 유발하여 학습 효율을 저하시킬 가능성이 있다. 따라서, 본 연구에서는 AD-FL의 시스템 지연 및 연산 오버헤드를 정량적으로 평가하고, 전송 효율성과 계산 복잡도 간의 균형적 관계를 분석함으로써 근사 복호 기반 효율적 FL 시스템의 구현 가능성과 실효성을 검토하고자 한다.

### II. 근사 복호 기반 연합학습

본 논문에서 고려하는 AD-FL 시스템의 전송 구조와 복호 메커니즘을 중심으로, UDP 기반 환경에서의 모델 파라미터 전달 효율을 분석한다. 본 연구에서는 다수의 로컬 디바이스와 하나의 글로벌 서버로 구성된 중앙집중형 연합학습 구조를 고려하며, 전송 방향에 따라 송신 주체가 달라질 수

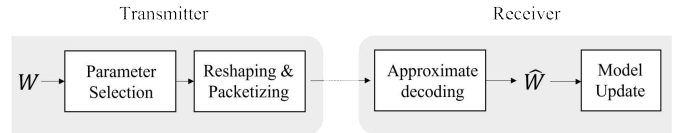


그림 1. 근사 복호 기반 FL 시스템의 모델 파라미터 전달 및 학습 과정

있음을 가정한다. 즉, 모델 업데이트가 상향 전송(uplink)될 때는 로컬 디바이스가 전송 측이 되며, 반대로 전역 모델이 하향 전송(downlink)될 때는 서버가 전송 측의 역할을 수행한다.

그림 1과 같이 전송 측에서는 로컬 학습을 통해 갱신된 모델 파라미터  $W$  중 중요도가 높은 요소를 선택적으로 추출하고, 이를 벡터 형태로 구조화하거나 압축하여 전송 패킷을 구성한다. 수신 측은 채널 오류로 패킷 손실이 발생한 경우, 수신된 주요 파라미터를 이용해 근사 복호를 수행하고 누락된 값을 복원한다. 복호된 모델 파라미터는 이후 전역 모델 집계에 사용되며, 이를 통해 통신 효율과 학습 정확도 간의 균형을 유지한다.

### III. 실험 결과

본 실험에서는 각 학습 라운드마다 100개의 로컬 디바이스 중 20%의 디바이스가 랜덤하게 참여하는 시스템을 고려한다. 이때, 네트워크는 UDP를 기반으로 동작하며, 따라서 재전송은 발생하지 않는다. 본 실험에서는 MNIST 데이터셋을 이용하여 이미지 분류 작업을 수행하며 이때 각 데이터셋은 non-IID(Independent and Identically Distribution)에 따라 분포되어있다고 가정한다. 본 실험에서 각 로컬 모델은 두 개의 hidden layer로 구성된 DNN 구조를 사용하며, 각 layer는 100개의 hidden node로 구성한다.

본 실험에서는 제안 기법의 성능을 비교하기 위해 1) TCP 기반의 모델 파라미터 전송 방식 [6], 2) 손실된 모델 파라미터를 0으로 대체하는

**표 1. 각 알고리즘의 Uplink 지연, Downlink, 코딩, 전체 지연(msec) 비교 결과**

	Uplink 지연 (A)	Downlink 지연 (B)	코딩 지연 (C)	전체 지연 (A+B+C)
TCP	162.00	25.30	2.71	190.01
Zero-filling	44.70	8.37	<b>2.70</b>	61.77
EDEN	71.00	12.70	21.61	105.31
SysNC	46.31	8.09	26.44	74.92
AD-FL	<b>23.19</b>	<b>3.99</b>	27.19	<b>54.37</b>

zero-filling 기법 [7], 3) 모델 파라미터를 양자화 하여 압축한 후 전송하는 EDEN 기법 [8], 4) 원 데이터 이외에 추가적으로 패킷을 전달하여 신뢰성을 향상시키는 systematic 네트워크 코딩(SysNC) 기법[9]을 고려한다. 본 실험에서 평가하는 AD-FL은 중요 모델 파라미터의 추출과 복호 과정에서 특이값 분해(Singular Value Decomposition, SVD) 기반의 모델 파라미터 분해 및 저랭크 근사 과정을 적용한다 [5].

각 알고리즘의 전송 효율을 평가하기 위해 UDP 기반 연합학습 환경에서 네 가지 지연 성분, 즉 uplink 지연, downlink 지연, 코딩 지연, 그리고 전체 지연을 측정하였다. uplink 지연은 로컬 디바이스에서 서버로 모델 파라미터  $W$  전체가 전송되는 데 소요되는 시간이며, downlink 지연은 서버에서 갱신된 전역 파라미터  $W$ 를 로컬 디바이스로 전달하는 데 걸리는 시간을 의미한다. 코딩 지연은 부호화 과정과 복호화 과정을 포함하며 근사 복호나 오류 정정 연산을 수행하는 데 필요한 전체 계산 시간을 나타낸다. 전체 지연은 세 지연의 합으로 정의한다.

표 1에 제시된 결과에서 확인할 수 있듯이, 제안한 AD-FL은 다른 기법에 비해 uplink 및 downlink 지연이 현저히 감소하였다. 이러한 지연 감소는 중요 모델 파라미터만을 선택적으로 전송함으로써, 각 통신 라운드에서 교환되는 데이터 양이 크게 줄어든 결과이다. 반면, 코딩 지연은 파라미터 인코딩 및 재구성 과정에서 수행되는 행렬 분해와 저랭크 근사 연산으로 인해 계산 복잡도가 증가함에 따라 비교적 큰 값을 보이는 것을 확인할 수 있다. 그럼에도 불구하고, 전체 지연(total delay)은 모든 비교 기법 중 가장 낮게 측정되었으며, 이는 선택적 전송으로 인한 통신 효율 향상이 코딩 오버헤드 증가분을 상쇄함을 의미한다.

### III. 결론

본 논문에서는 AD-FL 시스템을 대상으로 UDP 기반 전송 환경에서의 모델 파라미터 효율화와 지연 특성을 분석하였다. AD-FL은 중요 모델 파라미터만을 선택적으로 전송함으로써 uplink 및 downlink 지연을 크게 감소시켰으며, 이는 전체 통신 효율을 향상시키는 핵심 요인으로 작용하였다. 반면, 행렬 분해와 저랭크 근사 과정에서 발생하는 연산 복잡도로 인해 코딩 지연이 다소 증가하는 경향을 보였으나, 전체 지연을 기준으로 할 때 여전히 가장 낮은 값을 나타냈다. 이러한 결과는 AD-FL 기법이 통신 지연과 연산 복잡도 간의 균형을 효과적으로 달성함을 보여주며, 제한된 대역폭이나 비신뢰성 전송 환경에서도 실질적인 적용 가능성이 있음을 시사한다.

### ACKNOWLEDGMENT

이 논문은 이정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No.RS-2025-16066913, No. RS-2025-00556192)을 받아 수행된 연구임.

### 참 고 문 헌

- [1] K. Wei, J. Li, M. Ding, C. Ma, H. Su, B. Zhang, and H. V. Poor, "User-level privacy-preserving federated learning: Analysis and performance optimization," *IEEE Transactions on Mobile Computing*, vol. 21, no. 9, pp. 3388 - 3401, 2021.
- [2] R. Chandrasekaran, K. Ergun, J. Lee, D. Nanjunda, J. Kang, and T. Rosing, "FHDnn: Communication efficient and robust federated learning for AIoT networks," in *Proceedings of the 59th ACM/IEEE Design Automation Conference (DAC)*, 2022, pp. 37 - 42.
- [3] H. Ye, L. Liang, and G. Y. Li, "Decentralized federated learning with unreliable communications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 3, pp. 487 - 500, 2022.
- [4] H. Zhou, J. Cheng, X. Wang, and B. Jin, "Low rank communication for federated learning," in *Database Systems for Advanced applications (DASFAA) 2020 International Workshops*, Springer, 2020, pp. 1 - 16.
- [5] 권정민, 박형근, "부분적 모델 파라미터 손실에 따른 연합학습 성능 저하를 완화하기 위한 근사 데이터 전송 전략" 2024년도 한국통신학회 하계종합학술발표회 논문집, 2024, pp. 926-927.
- [6] S. Ha, I. Rhee, and L. Xu, "CUBIC: A new TCP-friendly high-speed TCP variant," *ACM Special Interest Group on Operating Systems Review*, vol. 42, no. 5, pp. 64 - 74, 2008.
- [7] C. Shiranthika, Z. H. Kafshgari, P. Saeedi, and I. V. Bajić, "SplitFed resilience to packet loss: Where to split, that is the question," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2023, pp. 367 - 377.
- [8] S. Vargaftik, R. B. Basat, A. Portnoy, G. Mendelson, Y. B. Itzhak, and M. Mitzenmacher, "EDEN: Communication-efficient and robust distributed mean estimation for federated learning," in *International Conference on Machine Learning (ICML)*. PMLR, 2022, pp. 21 984 - 22 014.
- [9] D. E. Lucani, M. Médard, and M. Stojanovic, "Systematic network coding for time-division duplexing," in *IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2010, pp. 2403 - 2407.