

경사 압축 알고리즘 기반 통신 효율적인 공중 연합 학습의 연구 동향에 대한 조사

오준석, 김재민, 이동현, 이운성, 송치현, 허동현, 이재환*, 조성래

중앙대학교, 국립공주대학교*

{jsoh, jmkim, dhlee, yslee, chsong, dhur}@uclab.re.kr, jhnlee@kongju.ac.kr*, srcho@cau.ac.kr

A Survey on Research Trends of Gradient Compression Algorithm-Based Communication-Efficient Federated Learning Over-the-Air

Junsuk Oh, Jaemin Kim, Donghyun Lee, Yunseong Lee, Chihyun Song, Donghyeon Hur, Jaehwan Lee*, and Sungrae Cho

Chung-Ang University, Kongju National University*

요약

심층신경망(DNN, Deep Neural Networks)에 기반한 높은 통신 비용으로 인해 연합 학습(FL, Federated Learning)은 다중 액세스 채널의 파형 중첩 특성을 활용하는 공중 계산(OAC, Over-the-Air Computation)을 통합하여 공중 연합 학습(FLOA, FL Over-the-Air)으로 발전했다. 그러나, 병목 현상은 FLOA에서 여전히 발생하고 있다. 이를 위해, 본 논문은 경사 압축 알고리즘 기반 통신 효율적인 FLOA의 최근 연구 동향을 조사한다.

I. 서론

6세대 통신 및 네트워크에서 중요한 주제는 로컬 장치의 계산 잠재력과 네트워크 연결 유연성을 활용하여 빅 데이터에 기반한 기계 학습(ML, Machine Learning) 작업을 중앙 집중형에서 분산형으로 전환하는 것이다. 이에 따라, 분산 학습 패러다임인 FL이 상당한 관심을 얻고 있다 [1]. FL에서 지리적으로 분산된 로컬 장치는 로컬 데이터를 활용하여 로컬 모델을 훈련한 후 집계를 위해 로컬 경사를 매개변수 서버(PS, Parameter Server)로 전송한다. PS는 이후 도출한 글로벌 경사를 각 로컬 장치에게 반환하고 글로벌 모델이 수렴할 때까지 이러한 통신 라운드를 반복한다. 기존 중앙 집중형 학습 패러다임과 다르게 FL은 대규모 원시 데이터를 전송할 필요성이 존재하지 않아 통신 비용을 절감하면서 개인 정보 보호 문제를 상당한 정도로 보장한다.

DNN을 구성하는 수많은 매개변수로 인해 병목 현상이 발생함에 따라, FL은 통신 비용을 줄이기 위해 경사 압축 알고리즘을 고려한다. 이러한 경사 압축 알고리즘에는 양자화, 희소화, 양자화 및 희소화 혼합 방식, 압축 센싱, 모델 가지치기 등이 존재한다. 이러한 경사 압축으로 인해 발생할 수 있는 정보 손실은 글로벌 경사 추정 오류로 이어져 수렴 성능과 학습 성능의 저하를 초래할 수 있음에 따라 최근 경사 압축 알고리즘은 오류 피드백 메커니즘을 통합한다. 그러나, 이러한 경사 압축 알고리즘에도 불구하고, FL은 각 통신 라운드에 참여하는 장치 수에 따라 통신 비용이 선형적으로 증가하기 때문에 병목 현상이 여전히 발생할 수 있다. 이를 위해, 최근 FL은 다중 액세스 채널의 파형 중첩 특성을 활용하는 OAC를 통합한다 [2]. OAC는 각 로컬 장치에 대해 전용 무선 자원을 할당하지 않아 FL의 모델 집계 과정에 적합하다. 따라서, FLOA는 무선 연합 학습(WFL, Wireless Federated Learning)에 대해 혁신적인 방식을 제시한다.

FLOA는 모든 장치가 동일한 시간-주파수 내에서 개별 경사를 전송할 수 있게 보장하고, PS가 전용 무선 자원을 할당할 필요성을 제거한다. 이러한 방식은 참여 장치 수가 증가함에 따라 선형적으로 증가하는 무선 자

원 문제를 방지한다 [3], [4]. 따라서, FLOA는 통신 및 계산을 통합하여 기존 FL보다 높은 대역폭 효율성을 제공하고, 다양한 시나리오 내 데이터 집약적인 실제 응용 프로그램에서 중추적인 역할을 수행하도록 한다. 그러나, DNN을 구성하는 수많은 매개변수에 기반한 높은 통신 비용은 해결되지 않아 FLOA에서 여전히 병목 현상 문제를 초래한다. 이를 해결하기 위해, 최근 FLOA는 경사 압축 알고리즘을 다시 고려한다. 본 논문은 본문에서 이러한 경사 압축 알고리즘 기반 FLOA의 최근 연구 동향을 조사하고, 결론에서 미래의 방향을 제시하여 마무리한다.

II. 본론

[5]는 [3] 및 [4]의 학습 모델 매개변수 전송이 전력 낭비로 귀결된다는 점에 주목하여 분산형 확률적 경사 하강법(DSGD, Distributed Stochastic Gradient Descent)의 각 반복을 분산형 공중 손실 계산 문제로 상정한다. 이를 통해, 협대역 채널 물리 계층 자원제약을 고려하여 아날로그 DSGD (A-DSGD, Analog DSGD)를 제안한다. 해당 소스 코딩 알고리즘은 경사 차원 감소를 위해 경사 희소화, 오류 피드백 메커니즘, 무작위 선형 투영을 사용한다.

[6]은 압축 아날로그 DSGD (CA-DSGD, Compressed A-DSGD)를 제안하고, [5]의 AWGN (Additive White Gaussian Noise) MAC을 Fading MAC으로 확장한다.

[5] 및 [6]에서 공중 집계에 사용되는 인코딩되지 않은 아날로그 변조는 디지털 변조 기반 시스템에 배포하기 어렵다는 점에 주목하여 [7]은 1비트 양자화, 디지털 변조 및 PS의 과반수 투표를 기반으로 하는 1비트 광대역 디지털 집계(OBDA, One-Bit Broadband Digital Aggregation)를 제안한다. 특히, FL에서 로컬 경사 통계를 알 수 없다는 점에 주목하여 FLOA에서 로컬 경사 정규화 가정을 제거하기 위해 1비트 양자화를 채택한다.

[8]은 [7]의 1비트 양자화 및 절단된 채널 반전이 수렴 속도 및 학습 성능을 저하한다는 점에 주목하여 수렴 분석 기반 전력 제어를 포함하는 오

류 피드백 OBDA (EFOBDA, Error-Feedback OBDA)를 제안한다.

[9]는 [5] 및 [6]의 무작위 선형 투영과 [7]의 1비트 양자화를 기반으로 전력 제어와 장치 스케줄링을 공동 최적화하는 1비트 압축 감지 아날로그 집계(OBCSAA, One-Bit Compressive Sensing Analog Aggregation)를 제안한다.

[10]은 [5] 및 [6]의 작업에서 무작위 선형 투영이 수신기의 계산 복잡성을 증가시킨다는 점에 주목하여 전력 제어 및 장치 스케줄링을 공동 최적화하는 시간 상관 희소화 및 하이브리드 집계(TCS-H, Time-Correlated Sparsification with Hybrid aggregation)를 제안하고, 장치 간 같은 글로벌 희소 경사와 장치 간 다른 로컬 희소 경사를 각각 OAC와 디지털 통신을 통해 집계한다.

[11]는 [5] 및 [6]의 모델 희소화 및 무작위 압축 기반 전송기와 PS에 대한 수정 터보 압축 감지(Modified Turbo Compressive Sensing)라는 수신기를 설계하여 공동 연합 다중 작업 학습(OA-FMTL, Over-the-Air Federated Multi-Task Learning)을 제안한다.

III. 결론

본 논문은 경사 압축 알고리즘 기반 통신 효율적인 FLOA의 최근 연구 동향에 대한 조사를 제시한다. 각 연구는 수렴 성능과 통신 비용 간 절충점을 체계적으로 분석하여 학습 성능을 저하하지 않으면서 통신 효율적임을 증명한다. 따라서 통신 효율적인 FLOA를 위한 새로운 경사 압축 알고리즘은 앞으로 수렴 성능 대 통신 비용의 관점에서 다른 어떤 경사 압축 알고리즘보다 더 효율적임을 증명해야 한다.

ACKNOWLEDGMENT

“본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음” (IITP-2024-RS-2022-00156353, IITP-2024-RS-2023-00258639)

참고 문헌

- [1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, “Communication-efficient learning of deep networks from decentralized data,” *Proceedings of Artificial intelligence and statistics conference*, vol. 54. PMLR, 2017, pp. 1273-1282.
- [2] B. Nazer and M. Gastpar, “Computation over multiple-access channels,” *IEEE Transactions on information theory*, vol. 53, no. 10, pp. 3498-3516, 2007.
- [3] G. Zhu, Y. Wang, and K. Huang, “Broadband analog aggregation for low-latency federated edge learning,” *IEEE Transactions on wireless communications*, vol. 19, no. 1, pp. 491-506, 2020.
- [4] K. Yang, T. Jiang, Y. Shi, and Z. Ding, “Federated learning via over-the-air computation,” *IEEE Transactions on wireless communications*, vol. 19, no. 3, pp. 2022-2035, 2020.
- [5] M. M. Amiri and D. Gündüz, “Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air,” *IEEE Transactions on signal processing*, vol. 68, pp. 2155-2169, 2020.
- [6] M. M. Amiri and D. Gündüz, “Federated learning over wireless fading channels,” *IEEE Transactions on wireless communications*,

vol. 19, no. 5, pp. 3546-3557, 2020.

- [7] G. Zhu, Y. Du, D. Gündüz and K. Huang, “One-bit over-the-air aggregation for communication-efficient federated edge learning: Design and convergence analysis,” *IEEE Transactions on wireless communications*, vol. 20, no. 3, pp. 2120-2135, 2021.
- [8] Y. Liu, D. Liu, G. Zhu, Q. Shi, and C. Zhong, “Over-the-air federated edge learning with error-feedback one-bit quantization and power control,” 2023, arXiv:2303.11319.
- [9] X. Fan, Y. Wang, Y. Huo and Z. Tian, “1-bit compressive sensing for efficient federated learning over the air,” *IEEE Transactions on wireless communications*, vol. 22, no. 3, pp. 2139-2155, 2023.
- [10] Y. Sun, S. Zhou, Z. Niu and D. Gündüz, “Time-correlated sparsification for efficient over-the-air model aggregation in wireless federated learning,” *Proceedings of IEEE International conference on communications*, Seoul, Korea, Republic of, 2022, pp. 3388-3393.
- [11] H. Ma, X. Yuan, Z. Ding, D. Fan and J. Fang, “Over-the-air federated multi-task learning via model sparsification, random compression, and turbo compressed sensing,” *IEEE Transactions on wireless communications*, vol. 22, no. 7, pp. 4974-4988, 2023.