사용자별 가중치 표준 편차를 활용한 Federated Learning 성능 향상 기법

김형빈, 김용호, 우기문, 김지하, 박현희*
명지대학교 정보통신공학과

{hbkim, yhkim98, gmwoo, yaki5896, hhpark*}@mju.ac.kr

FedSD : Federated Learning algorithm with Standard Deviation of weights for each user

Hyungbin Kim, Yongho Kim, Gimoon Woo, Jiha Kim, and Hyunhee Park*
Department of Information and Communication Engineering, Myongji University

요 약

본 논문에서는 FL(Federated Learning)에서 데이터 분포의 Non-IID(Independent and Identically Distributed) 문제 개선을 목표로 한다. FL의 경우 edge device들의 데이터 분포가 IID 할 때는 학습에 부정적 영향을 끼치지 않지만, Non-IID 할 때는 클라우드 컴퓨팅 성능에 도달하지 못하는 결과를 보일 수 있다. 본 논문을 통해 우선 기본적인 FL의 개념을 살펴보고, edge device들의 데이터 분포가 Non-IID하여 FL 학습이 적절히 이루어지지 않는 상황을 개선하기 위한 연구를 소개한다. 이를 통해 본 논문에서는 Non-IID 상황에서 사용자별 가중치 표준 편차를 활용하는 기법을 제안하며, 기존 FedAvg(Federated Averaging algorithm)와 제안하는 알고리즘의 accuracy와 loss 비교를 통하여 성능을 검증한다.

Ⅰ. 서 론

하드웨어 장치의 발전에 따라 딥 러닝 성능도 크게 향상되어 다양한 분야에서 활용되고 있다. CNN, DNN 등의 일반적인 딥 러닝 모델을 제작하기 위해서는 사용자들로부터 데이터를 수집하여 이를 활용하여야 한다. 하지만 이 데이터들로 인하여 사생활이 침해되는 것이 문제로 제기될 수 있으며, 따라서 이를 해결하기 위한 여러 딥 러닝 모델이 연구되었다. 이 중에서 구글이 제안한 FL(Federated Learning)은 분산적인 모델 설계로 인하여 사생활 보호와 동시에 높은 성능을 보인다 [1].

FL은 다수의 ED(Edge Device)와 하나의 FL 서버로 구성되어있고, ED에서 학습한 모델을 중앙으로 취합하는 학습 모델이다. 기존 모델들과 달리 데이터를 FL 서버로 수집하지 않기 때문에 사생활 보호라는 특징을 띤다. 하지만 학습 과정에서 각 ED의 데이터를 활용하는 FL의 특성상, 각 ED가 가진 적은 양의 데이터에 의하여 편중된 학습이 이루어지기 쉽다. 특히 데이터 분포가 Non-IID(Independent and Identically Distributed) 한 경우가 있을 수 있다. 때문에, Non-IID 문제로 인하여 클라우드 컴퓨팅 [2] 성능에 도달하지 못하는 문제를 보인다.

따라서 Non-IID 문제를 개선하기 위한 연구가 진행되어 왔다. [3] 에서는 FL 서버가 각 ED에게 공통적인 데이터를 적게 분배하여 Non-IID 문제를 개선한다. 하지만 각 ED에게 주어진 적은 수의 공통 데이터에 의해 과적합된 학습 결과를 보일 수 있기 때문에 논문에서 주장하는 성능을 일반화할 수 없다. [4] 에서는 최근 많은 관심을 받는 메타 러닝을 사용하였다. 메타 러닝은 학습하는 방법 그 자체를 배운다는 목적으로, 적은 양의 데이터를 처리하는 것에 적합한 모델이다. [4] 에서는 FL과 메타 러닝이 합쳐진 FedMeta(Federated Meta-Learning algorithm)를 제안하였으며, 이를 통해 적은 양의 데이터만으로 효과적인 성능을 갖는 모델을 만들어낼 수 있음을 보였다. FedMeta가 정확도 관점에서 FedAvg보다 좋은 성능을 보이지만, 정확한 수치는 86.23%를 나타내어 성능 향상이 필요하다.

따라서 본 논문에서는 FL 서버에서의 데이터 처리 형태를 개선하기 위한 FedSD(Federated Learning with Standard Deviation algorithm)를 제안하며, 다음 장에서 FedSD의 구성을 소개한다.

Ⅱ. FedSD

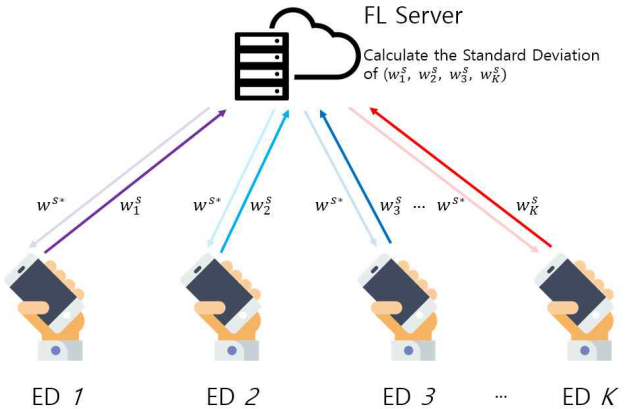

그림 1. FedSD

본 논문에서는 그림 1 형태의 FedSD 알고리즘을 제안한다. 그림 1 에서 모델 파라미터 w_1^s, w_2^s, w_3^s, w_K^s는 각 ED에서 학습되어 FL 서버로 전송되는 가중치를 의미한다. 기존 FL 기법에서는 FL 서버에서 이 데이터들의 평균화가 이루어지고, 평균화를 거쳐 최적화된 모델 파라미터 w^{s*}는 다운 링크를 통해 모든 ED에게 전달된다. 각 ED에서 학습하는 과정에서 데이터 분포가 IID 상태이면 기존 기법을 계속 사용할 수 있겠으나 Non-IID 상태에서는 모델 성능이 저하될 수 있다. 이를 개선하기 위하여

FedSD를 제안하며 아래 알고리즘 1 에 FedSD를 의사코드로 작성하였다.

알고리즘 1 FedSD

1: $n = 0$
2: **for** $k = 1,\cdots,K$ **do**
3: 각 ED 별 학습을 진행
4: FL 서버에 학습 결과 가중치를 전송
5: **end for**
6: // 아래는 FL 서버에서의 작동 원리를 의미
7: **input:** 각 ED 별 학습 결과 가중치 $(w_1^s,\cdots,w_K^s)$
8: $S(w^s) = \sum_{k=1}^{K} w_k^s$
9: $A(w^s) = S(w^s) \ / \ K$
10: $SD(w^s) = \sqrt{\frac{\sum_{k=1}^{K}(w_k^s - A(w^s))^2}{K}}$
11: **for** $k = 1,\cdots,K$ **do**
12: **if** $A(w^s) - SD(w^s) <= w_k^s <= A(w^s) + SD(w^s)$
13: $w_n^{SD} = w_k^s$
14: $n = n + 1$
15: **end for**
16: $SDA(w^{SD}) = \sum_{n=1}^{N} w_n^{SD}$
17: **output:** 최적화된 모델 파라미터 $SDA(w^{SD})$는 각 ED에게 initialization

기존 FL과 FedSD의 가장 큰 차이는 사용자별 가중치 표준 편차를 활용하는 것이다(알고리즘 1. Line 10-17). 서버에서 ED에 데이터를 전송하기 이전에 $w_k^s(k = 1,\dots,K)$의 평균에서 표준 편차 범위 내의 $w_n^{SD}(n = 1,\dots,N)$만을 평균화하여 ED에 전송한다. 이를 통해 데이터 분포가 Non-IID 할 때 covariate shift를 유발하는 feature들을 제외할 수 있다.

Ⅲ. 실험

제안하는 기법의 성능을 확인하기 위하여 LeNet-5 [5] 모델을 사용하였다. 데이터로는 MNIST 데이터셋 [6] 을 사용하였으며, 10 개의 ED에 대해서 Non-IID 상태가 되도록 데이터를 분포하였다. 5 개의 ED에는 무작위로 데이터를 배분하였고, 다른 5 개의 ED에는 각각의 ED가 1~3 개의 클래스를 갖도록 배분하였다. 성능 확인을 위한 환경 설정은 아래 표 1 과 같다.

표 1. 테스트 환경 설정

테스트 환경	
edge device	10
local epoch	1
batch size	100
optimizer	SGD
train/test ratio	4:1

FedSD를 사용하였을 때 accuracy 및 loss 측면에서 성능이 향상됨을 그림 2 와 그림 3 에서 확인할 수 있다. Communication round 40 이후 FedSD의 accuracy와 loss가 FedAvg보다 향상된 성능을 보인다. 특히 communication round 100 일 때, FedAvg의 accuracy와 loss는 0.9281, 0.2580 을 보였고 FedSD는 0.9381, 0.2217 을 보여 각 측면에서 0.01, 0.0363 의 성능 향상이 이루어짐을 확인할 수 있다.

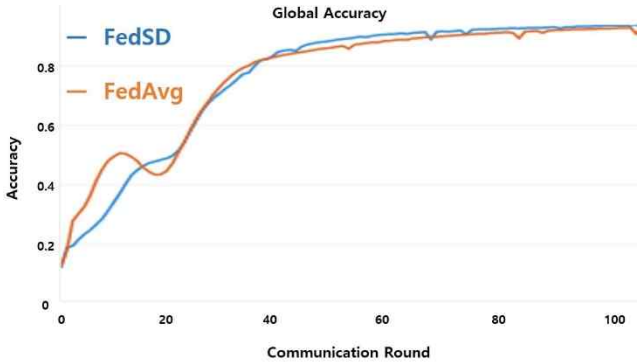

그림 2. Accuracy 비교

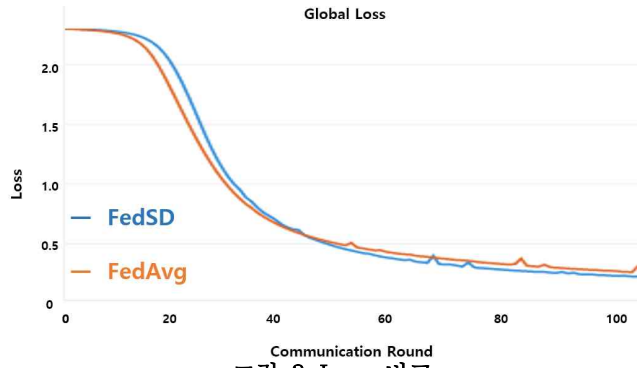

그림 3. Loss 비교

Ⅳ. 결론

본 논문에서는 FL에서 데이터 분포의 Non-IID 문제를 개선하기 위하여 FedSD를 제안하였다. 제안하는 기법을 사용하여 데이터 분포가 Non-IID 할 때 covariate shift를 유발하는 feature들을 제외할 수 있으며, 같은 실험 환경에서 이전 기법과의 accuracy, loss 비교 실험을 통해 제안하는 기법의 accuracy, loss 성능이 이전 기법에 비하여 향상됨을 확인하였다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-00368, 6G 서비스를 위한 인공지능/머신러닝 기반 자율형 MAC 개발)

참 고 문 헌

[1] B.McMahan, E. Moore, D. Ramage, S.Hampson, and B. A. y Acras, "Communication-efficient learning of deep networks from decentralized data," in Artificial Intelligence and Statistics. PMLR, 2017, pp. 1273- 1282.

[2] Armbrust, Michael, et al. "A view of cloud computing," Communications of the ACM 53.4 (2010): 50-58

[3] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, V. Chandra, "Federated learning with Non-IID data," arXiv preprint arXiv:1806.00582, 2018.

[4] F. Chen, M. Luo, Z. Dong, Z. Li, and X. He, "Federated meta-learning with fast convergence and efficient communication," arXiv preprint arXiv:1802.07876, 2018.

[5] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278- 2324

[6] LeCun, Yann. "The MNIST database of handwritten digits." http://yann.lecun.com/exdb/mnist/ (1998).