개인화 및 라벨 노이즈를 향하여: 조기 학습 정규화는 연합 학습에서 무엇을 돕는가?[1)]

김동규* 신재우O* 김태현 윤세영
KAIST 김재철AI대학원
{eaststar, yimsungen5, potter32, yunseyoung}@kaist.ac.kr

Toward Personalization and Label-Noise: What does Early Learning Regularization help in Federated Learning?

Donggyu Kim Jaewoo Shin Taehyeon Kim Seyoung Yun
KAIST Kim Jaechul Graduate School of AI

요 약

연합 학습은 각 로컬 교육 데이터를 분산된 상태로 유지하면서 각 클라이언트의 개인 정보를 보호하며 학습하는 협업 기계 학습 설정의 해답으로써 등장했다. 인기에도 불구하고 연합 학습은 개인화 및 라벨 노이즈에 대해 견고하지 않은 문제로 인해 실제 의사 결정 시스템에 배포하는데 여전히 어려움을 겪고 있다. 구체적으로 개인화의 경우에는 연합 학습은 각 사용자에 맞는 모델을 학습하지 않으며, 라벨 노이즈에도 취약한 모습을 보인다. 본 논문에서는 이러한 문제를 목적 함수 최적화의 관점에서 살펴본다. 먼저, 우리는 로컬 모델이 데이터를 over-confident하게 예측하는 문제를 지적한다. 이 문제를 해결하기 위해 간단한 정규화 방법인 Federated Early Learning Regularization (FedELR)을 제안한다. FedELR에서는 각 클라이언트는 강력한 손실함수인 ELR에서 영감을 받아 서버 모델의 출력을 활용하여 로컬 데이터의 출력을 정규화한다. 우리의 방법은 기존 알고리즘과 비교하여 고도로 데이터가 불균형한 상황에 표준 벤치마크 데이터 셋에서 크게 향상된 성능을 보여주며, 라벨 노이즈에도 강건함을 보였다.

1. 서 론

연합 학습(Federated Learning; FL)은 협업 머신 러닝의 가장 인기 있는 패러다임 중 하나이다[2,3]. 일반적으로 연합 학습 프레임워크는 각 클라이언트가 자체적으로 가지는 개인 데이터를 통해 로컬 모델을 업데이트 한 후 모든 로컬 업데이트는 글로벌 모델로 집계시킨다. 특히 연합 평균화(FedAvg)는 클라이언트의 로컬 학습 모델에 대한 집계 방법으로 평균화를 사용한다. 이러한 연합 학습 프레임워크를 사용하면 구조적으로 데이터 수준에서 높은 수준으로 개인 정보 위험을 완화하며 휴대전화 및 태블릿과 같은 엣지 컴퓨팅 장치의 응용 프로그램에 상당한 잠재력을 보인다. 연합 학습의 특성으로 인해 온디바이스(on-device) 학습이 가능하고 클라이언트의 로컬 모델 품질을 지속적으로 향상시킬 수 있다.

최근 FL 알고리즘은보다 정확한 글로벌 모델을 얻기 위해 모델을 집계하는 방법과 클라이언트를 선택하기 위한 새로운 목적 함수를 설계하는 방향으로 발전하고 있다[3]. 다른 한편으로, 실제 의사 결정 시스템에 연합 학습 알고리즘을 적용하는 것에 대해 이목이 집중되고 있지만, 근본적인 문제에 대해서는 다루지 않고 있다. 예로 들어, 의료 영역에서는 일반적인 사용자들의 건강에 대한 일반화된 글로벌 모델보다 사용자별 모델이 더 중요하다. 때때로 상당수의 데이터에 라벨이 잘못 지정되거나 (또는 라벨이 지정되지 않아) 모델이 최신 알고리즘을 실행할 수 없다. 우리는 연합 학습 알고리즘의 실용화를 위해 개인화와 라벨 노이즈 두 가지 문제를 구체적으로 확인했다.

개인화. 데이터 불균형이 존재할 경우 글로벌 모델은 각 클라이언트의 로컬 데이터에 대해 개별적으로 잘 일반화되지 않을 수 있다. 이 문제를 해결하는 것의 핵심은 중앙 서버의 일반화를 위해 로컬 데이터의 특수 지식(개별 클라이언트에 해당하는 지식)과 일반적인 지식(일반적으로 모든 클라이언트들에게 통용되는 지식) 사이에서 균형을 맞추는 것이다.

라벨 노이즈. 노동집약적인 라벨링 작업 시, 각 클라이언트마다 필연적으로 약간의 라벨 노이즈(라벨이 잘못된 샘플)가 존재하며, 모델은 테스트 데이터에 대한 잘못된 일반화로 수렴된다. 따라서 현대의 심층 네트워크는 이러한 신뢰할 수 없는 데이터에 모델을 완벽하게 맞출 수 있기 때문에 모델 훈련에 강건함이 요구된다. 특히, 노이즈가 많은 클라이언트의 데이터는 모델 집계를 하는 중앙 서버에 부정적인 영향을 줄 수 있다. 단일 모델 훈련의 최근 발전은 노이즈에 강건한 손실 설계, 레이블 수정 활용, 두 개의 신경망을 사용하여 깨끗한 라벨 선택으로 발전했다. 그러나 라벨 노이즈가 주입될 때 연합 학습 환경에서 발생하는 고유한 역학에 대한 연구는 거의 없다.

개인 정보를 보호하면서 동시에 이러한 두 가지 문제를 완화하기 위한 방법으로 본 논문에서 우리는 새롭고 간단한 알고리즘인 FedELR(**Fed**erated **E**arly **L**earning **R**egularization)를 제공한다. FedELR은 조기 학습 정규화(ELR)에서 영감을 얻어 로컬 데이터에서 학습을 진행할 때 이전 반복에서 출력되는 결과에 따라 다음 반복에서 출력되는 예측에 제한을 걸어준다. 최근 연구에 따르면 심층 신경망(Deep Neural Network; DNN)은 초기 학습 단계에서 데이터가 간단하거나 올바른 라벨을 갖는 것을 먼저 학습한 후 결국 자신의 데이터를 암기하여 over-confident하게 예측할 수 있음을 보여준다. 우리는 이러한 작업에서 영감을 받아 데이터가 암기되는 현상을 방지하기 위해 현재 로컬 예측과 이전 서버 예측 사이에 ELR을 적용하는 간단한 접근 방식을 먼저 소개한다. 우리 연구 결과는 연합 학습에 대한 표준 벤치마크 데이터 셋에 대한 방법을 검증하고 개인화 및 레이블 노이즈에 대한 우수성을 보여준다. 특히 FedELR은

1) 이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임([No.2019-0-00075, 인공지능대학원지원(한국과학기술원), 10%]과 [No. 2021-0-00907, 능동적 즉시 대응 및 빠른 학습이 가능한 적응형 경량 엣지 연동분석 기술개발, 90%])

더 많은 불균형 데이터에서 테스트 정확도 곡선이 변동이 적으며 안정적으로 최적화를 이끌어낸다. 우리의 연구는 다음 세 가지로 요약된다.

- 우리는 로컬 데이터를 통해 학습을 하는 과정에서 암기가 극단적으로 발생함을 보여주고 이 현상이 데이터 불균형과 노이즈가 많을수록 각각의 클라이언트들에게 좋지 않은 결과를 도축함을 보인다.
- 이러한 문제를 방지하기 위해 우리는 조기학습 정규화(ELR)에서 영감을 받아 연합 학습 환경에서 로컬 데이터로 학습할 때 제약을 추가하는 프레임워크를 제안한다.
- 데이터 불균형의 설정에 초점을 맞춘 여러 데이터 셋에서 검증을 진행한다. 레이블에 대한 개인화 및 견고성 측면에서 우리의 방법은 추가 통신 비용 및 개인 정보 침해 없이 기존 기준을 크게 능가한다.

2. 본 론

2.1 연합 학습

연합학습의 목적은 다음의 확률적 최적화 문제를 해결하는 것이다.

$$\min_w f(w) \triangleq \sum_{k \in S} p_k F_k(w)$$

여기서 S는 총 클라이언트의 집합이고 p_k는 클라이언트 k의 가중치로 $p_k \geq 0$이며 $\sum_k p_k = 1$이다. 클라이언트 k의 로컬 목적함수는 w의 매개변수로 매개화된 D_k의 데이터 분포를 가지는 로컬 데이터 (x_k, y_k)로 $F_k(w) = E_{x_k \sim D_k}[l_k(x_k, y_k; w)]$를 최소화하는 것이다.

연합 학습의 표준 알고리즘인 FedAvg는 학습률 η로 로컬 모델 w_k^t를 학습하며 E 단계마다 w^t와 동기화를 진행한다.

$$w_k^t \triangleq \begin{cases} w_k^{t-1} - \eta \nabla F_k(w_k^{t-1}) & \text{if } t \bmod E \neq 0 \\ w^t & \text{if } t \bmod E = 0 \end{cases}$$

이후 임의로 선택된 클라이언트들의 집합인 S^t의 클라이언트들 $k \in S^t$에 대한 w_k^t들을 평균 내어 글로벌 모델인 w^t를 학습한다.

$$w^t \triangleq \sum_{k \notin S^t} p_k w^t + \sum_{k \in S^t} p_k w_k^t$$

2.2 조기 학습 정규화

조기 학습 정규화(ELR)[4]는 현재 예측 출력과 이전 예측 출력 간의 차이를 정규화하여 모델을 훈련시키는 최신 알고리즘이다. ELR은 일반적으로 레이블 노이즈가 있는 기존 딥 러닝 작업에 적용되었으며 노이즈 데이터에 대한 견고성을 향상시키는 것으로 나타났다. ELR의 주된 목적은 잘못된 라벨이 암기가 되는 것을 방지하는 것이다. 조기 학습 정규화의 목적 함수는 다음과 같다.

$$L_{ELR} \triangleq L_{CE} + \frac{\lambda}{n} \sum_{i=1}^{n} \log(1 - < p >)$$

라벨 스무딩 라벨을 스무딩 시키는 것은 예측의 over-confident 문제를 방지하여 심층 신경망의 성능을 향상시키는 일반적인 방법이다[5]. 라벨 스무딩은 기저 사실인 원-핫 벡터 y를 하드 타겟 y^{LS}의 가중치 혼합으로 대체하여 일반화를 쉽게 하는 기술이다.

$$y_k^{LS} = \begin{cases} 1 - \beta & \text{if } y_k = 1 \\ \dfrac{\beta}{K-1} & otherwise \end{cases}$$

2.3 실험 설정

우리는 이미지 데이터에 대한 훈련된 분류기의 보정에 대한 DSD의 영향을 분석하기 위해 수많은 실험을 진행한다. 우리는 다양한 심층 아키텍처와 표준 데이터 셋을 사용하여 실험한다.

실험 설정 실험에서 CIFAR-10 데이터 셋을 사용한다. 본 실험은 VGG-11 및 ResNet-8 모델에서 진행되었다.

CIFAR 10 데이터 셋 32 x 32 해상도의 이미지 6만 장으로 이루어져 있으며 10개의 클래스로 나뉘어져 각 클래스마다 6000장의 이미지가 있는 dataset이다. 이미지 6만장 중 5만장이 training 이미지이며, 1만장은 test 이미지이다.

Dirichlet 불균형 데이터를 임의로 만들기 위해 모든 클라이언트의 훈련 데이터가 벡터 q로 매개변수화된 N개의 클래스에 대한 독립적인 범주형 분포에 따라 클래스 라벨을 사용한다고 가정한다($q_i \geq 0,\ i \in [1, N]$ and $\| q \|_1 = 1$). 이때 Dirichlet 분포, $q \sim Dir(\alpha p)$,를 사용한다. 여기서 p는 N개의 클래스에 대한 사전 분포를 나타내며, $\alpha > 0$는 클라이언트들 간의 동일성을 제어하는 농도 변수이다.

2.4 실험 결과

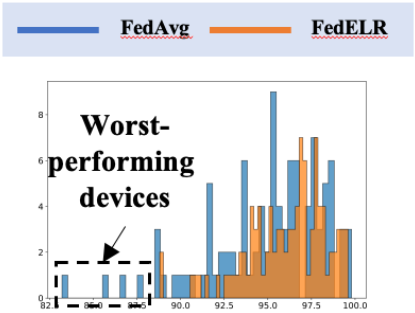
noise	FedAvg	FedELR
0.2	36.04	34.21
0.4	35.12	36.22
0.6	29.19	31.12
0.8	21.89	24.45
1.0	14.42	18.64

표 1 ResNet-8을 사용하여 Dirichlet CIFAR-10 데이터 셋에서 noise를 달리하였을 때 FedAvg와 FedELR의 global server의 정확도 결과

Dirichlet CIFAR-10 데이터 셋을 ResNet-8 모델을 사용하여 FedAvg와 FedELR의 중앙서버의 정확도를 비교해보았다. 클라이언트 20명으로 총 100라운드를 매 라운드마다 4명의 클라이언트를 uniformly random으로 선택하여, 각 클라이언트마다 5번의 학습을 진행하였다. 이때, Dirichlet 변수인 $\alpha = 0.01$이다.

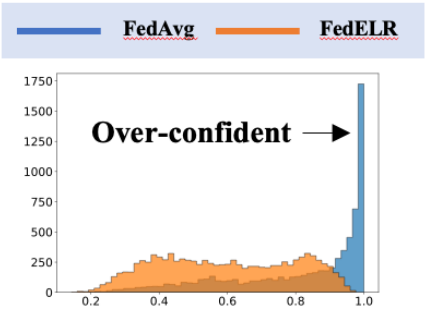
FedELR은 FedAvg보다 noise가 강해질수록 더욱 좋은 결과를 보인다. noise가 0.2일 때는 FedELR이 1.83%p 떨어지지만, noise가 0.4일 때 1.10%p, 0.6일 때 1.93%p, 0.8일 때 2.56%p 그리고 1.0일 때 4.22%p 높다. 이는 FedELR이 기존의 모델인 FedAvg보다 노이즈가 강한 환경에서 global server가 더욱 강건함을 확인할 수 있게 해준다.

그림 1은 VGG-11을 사용하여 40%의 노이즈가 존재하는 Dirichlet CIFAR-10 데이터 셋에 대해 FedAvg와 FedELR의 global server의 정확도 결과를 비교한 그림이다.



Personalization

그림 1 VGG-11을 사용하여 Dirichlet CIFAR-10 데이터 셋을 학습 한 후 각 클라이언트에서 미세조정을 한 결과



Calibration

그림 2 VGG-11을 사용하여 Dirichlet CIFAR-10 데이터 셋을 학습 한 후 테스트 데이터 셋에서 softmax 값이 제일 큰 데이터의 히스토그램

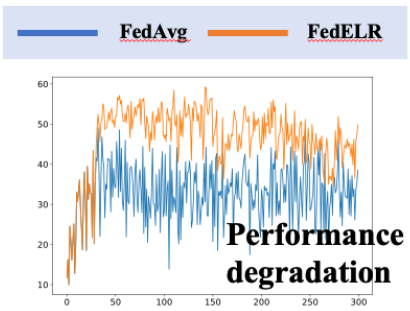
FedAvg의 정확도가 FedELR보다 현저히 떨어지는 것을 확인할 수 있다.

그림 2는 VGG-11을 사용하여 40%의 노이즈가 존재하는 Dirichlet CIFAR-10을 사용하여 학습한 global model을 각 클라이언트에 보내준 뒤 미세 조정을 한 결과이다. 결과를 보게 되면, FedELR의 global model을 사용하여 미세조정을 하는 것이 FedAvg의 global model을 사용하여 미세조정을 하는 것 보다 좋은 성능을 보인다. 또한, 평균 성능과 현저히 낮은 성능을 보이는 클라이언트들 또한 모두 FedAvg를 통해 학습한 것들이다.

그림 3은 VGG-11을 사용하여 40%의 노이즈가 존재하는 Dirichlet CIFAR-10을 사용하여 학습한 뒤 테스트 데이터 셋에서 softmax 값이 제일 큰 데이터의 히스토그램을 그린 것이다. FedAvg가 FedELR보다 현저히 높은 confident를 가지는 over-confident 상태임을 확인 할 수 있으며, 우리가 제안한 FedELR은 over-confident 문제를 방지했다.

3. 결 론

본 논문에서는 각 클라이언트의 개인 정보보호하며 학습하는 연합학습 환경에서 조기 학습 정규화를 결합하는 프레임워크를 제안하고 이미지 분류 과제에 대해서 정확도 향상, 노이즈에 대한 강건함 그리고 over-confident 방지를 실험적으로 확인했다. 본 논문에서 제안한 프레임워크를 Federated



Robustness to Label Noise

그림 3 VGG-11을 사용하여 Dirichlet CIFAR-10 데이터 셋에 노이즈를 주었을 때 global server의 정확도 비교

Early Learning Regularization (FedELR)이라고 칭하며, 기존의 FedAvg보다 좋은 성능을 보였으나 다른 개인화 연합학습 모델과의 비교를 하지 않은 한계점을 가지고 있다. 따라서 개인화 모델들과의 비교 및 적용하는 연구가 진행되어야 할 것이다.

[참고문헌]

[1] Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach.Advances in Neural Information Processing Systems, 33, 2020.

[2] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. InArtificial Intelligence and Statistics, pages 1273-1282, 2017.

[3] Durmus Alp Emre Acar, Yue Zhao, Ramon Matas Navarro, Matthew Mattina, Paul N What-mough, and Venkatesh Saligrama. Federated learning based on dynamic regularization. In Interfelix 2020federatednational Conference on Learning Representations, 2021.

[4] Sheng Liu, Jonathan Niles-Weed, Narges Razavian, and Carlos Fernandez-Granda. Early-learning regularization prevents memorization of noisy labels.CoRR, abs/2007.00151, 2020

[5] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Re-thinking the inception architecture for computer vision. InProceedings of the IEEE conference on computer vision and pattern recognition, pages 2818-2826, 2016.