클라이언트 데이터 이질성에 강인한 생성적 적대 신경망 기반 연합학습 알고리즘

장원준, 박현서, 이시현
한국과학기술원

wonjun_jang@kaist.ac.kr, phseo2000@kaist.ac.kr, sihyeon@kaist.ac.kr

Generative Adversarial Network-Based Federated Learning Algorithm Robust to Client Data Heterogeneity

Won-Jun Jang, Hyeon-Seo Park, Si-Hyeon Lee
KAIST

요 약

본 논문은 연합학습 상황에서 서버의 라벨이 없는 데이터를 이용한 앙상블 증류를 할 때에 생성적 적대 신경망을 이용하여 클라이언트 별 데이터 분포가 크게 상이할 때에도 강인하게 학습할 수 있는 알고리즘을 제안한다. 또한 제안하는 앙상블 증류 알고리즘이 클라이언트 데이터를 모두 합친 데이터 분포에서 최적의 성능을 가짐을 이론적으로 증명하고, CIFAR-10 데이터셋에 대해 성능의 우월성을 검증한다.

Ⅰ. 서 론

연합학습 분야는 클라이언트들이 각각 갖고 있는 데이터로 학습한 모델을 서버에게 보내면 서버가 그를 모아서 서버 모델을 학습하고 다시 클라이언트에게 보내는 과정을 반복함으로써 학습하는 알고리즘에 대해 연구하는 분야이다 [1]. 연합학습 과정에서 서버는 클라이언트의 데이터를 모두 받아오지 않으므로 프라이버시와 통신 효율 면에서 장점이 있다. Google 이 첫 연합학습 알고리즘인 FedAVG 알고리즘을 제시한 이후로, 연합학습 과정에서 통신 효율 [2], 프라이버시 [3], [4], 그리고 클라이언트 데이터 분포의 상이성 극복 [5]-[7]과 같은 분야에서 많은 연구가 진행되고 있다. 특히 서버는 클라이언트 대비 데이터 저장 및 계산 용량이 더 크므로 서버에 추가적인 데이터가 있을 수 있고, 이를 이용한 앙상블 증류를 통해 클라이언트 데이터 분포의 상이를 극복하는 알고리즘들이 제시되었다. 본 논문은 생성적 적대 신경망을 이용하여 더 좋은 성능의 분류 모델을 학습할 수 있는 앙상블 증류 알고리즘을 제시한다.

Ⅱ. 본 론

분류 모델을 학습시키기 위한 본 논문에서 고려하는 연합학습 상황은 다음과 같다. 하나의 서버와 C 명의 클라이언트가 있고, 클라이언트는 라벨이 있는 작은 데이터셋, 서버는 라벨이 없는 큰 데이터셋이 있다고 가정한다. 서버는 저장 용량이 클라이언트보다 클 것이므로 큰 데이터셋을 가지고 있을 수 있고, 데이터 라벨링은 비용이 많이 드는 작업이므로 서버에 라벨이 없는 큰 데이터셋이 있다는 가정은 충분히 현실적이다.

앙상블 증류 과정은 다음과 같다. t 라운드에 자신의 데이터셋으로 학습한 클라이언트 c 의 모델이 f_c^t 라고 하자. 서버는 자신의 데이터셋 U 안의 데이터 x 에 대해 다음과 같이 의사 라벨을 만든다.

$$\tilde{y}(x) = \sum_{c=1}^{C} w_c(x)\mathrm{f}_c^t(x)$$

이 때 $w_c(x)$ 는 데이터 x 에 대해 클라이언트 c 의 출력이 얼마나 반영되는지를 결정하는 가중치이고, $\sum_{c=1}^{C} w_c(x) = 1$이다.

이렇게 각 라벨이 없는 데이터 x에 대해 의사 라벨 $\tilde{y}(x)$ 를 계산하고, 의사 라벨이 있는 데이터셋 $\tilde{U} = \{(x_i, \tilde{y}_i)\}_{i=1}^{N}$ 을 구성한 후에, 서버는 자신의 모델 f_s^t 의 파라미터를 클라이언트 파라미터의 평균으로 초기화한 후에 $\tilde{U}$에 대해 추가로 학습한다. 이를 통해 클라이언트 각각의 데이터 분포가 크게 상이하더라도 서버 데이터셋으로의 추가 학습을 통해 더 좋은 성능의 서버 모델을 학습할 수 있게 된다.

FedDF[6]를 제시한 논문에서는 이러한 앙상블 증류를 제시하고, 몇 가지 가정 하에서 서버 모델 성능의 상한을 제시했다. FedDF 알고리즘은 모든 $c = 1, \ldots, C$ 와 모든 서버 데이터 x에 대해서, $w_{c_{DF}}(x) = \frac{1}{C}$로 클라이언트 별 동일한 균등 가중치를 이용하여 의사 라벨을 만든다.

Fed-ET 알고리즘은 서버 데이터 x에 대해 보다 잘 아는 클라이언트 모델의 출력이 더 많이 반영되도록 가중치를 조정한다 [7]. $\mathrm{f}_c^t(x)$ 가 데이터 x 에 대한 클라이언트 c 의 출력 logit 이라고 하자. 이 때 클라이언트 c의 가중치 $w_{c_{ET}}$는 다음과 같다.

$$w_{c_{ET}}(x) = \frac{\mathrm{Var}\left(f_c^t(x)\right)}{\sum_{c'=1}^{C} \mathrm{Var}\left(f_{c'}^t(x)\right)}$$

이는 logit 이 크게 차이나는 클라이언트 모델이 데이터에 대해 더 확신하는 모델이고, 그만큼 해당 데이터에 대해 더 잘 아는 모델일 것이라는 가설에서 기인한다.

본 논문이 제시하는 연합학습 알고리즘은 세 단계로 이루어져 있다.

1. 서버는 서버 데이터셋 U 를 이용하여 생성 모델 G를 훈련한 후 클라이언트들에게 배포한다.
2. 클라이언트 c는 서버가 제공한 G를 이용하여 가짜 데이터셋을 만들고, 자신이 가지고 있는 진짜 데이터셋을 이용하여 생성적 적대 신경망[8]의 판별자 D_c 를 학습한다. 특히, 진짜 데이터에 대해서는 1, 가짜 데이터에 대해서는 0 을 출력하도록 학습한다. 그 후 D_c를 서버에 보낸다.
3. 서버와 클라이언트는 앙상블 증류를 이용하여 연합학습을 진행한다.

앙상블 증류 단계에서 서버는 클라이언트 c의 판별자 D_c 와 모델 f_c^t 를 이용하여 의사 라벨을 만든다. 서버 데이터 x에 대해, "odd" Φ_c를 다음과 같이 정의한다.

$$\Phi_c(x) := \frac{D_c(x)}{1 - D_c(x)}$$

본 논문이 제시하는 가중치 w_c^*는 다음과 같다.

$$w_c^*(x) = \frac{\Phi_c(x)}{\sum_{c'=1}^{C} \Phi_{c'}(x)}$$

또한, 다음 정리가 알려져있다.

정리 1 [8]. 클라이언트 c의 데이터 분포가 p_c, 생성 모델 G가 만드는 데이터 분포가 p_g라고 하면, 최적으로 학습된 D_c는 다음과 같다.

$$D_c(x) = \frac{p_c(x)}{p_c(x) + p_g(x)}$$

이를 통해 판별자 D_c를 통해 만드는 odd 가 $\Phi_c(x) = \frac{p_c(x)}{p_g(x)}$ 를 근사함을 알 수 있고, 데이터 x에 대해 $\Phi_c(x)$는 해당 데이터가 c 의 데이터 분포에서 나왔을 확률 $p_c(x)$ 에 비례한다. 또한 $p_c(x)$가 다른 클라이언트에 비해 클수록 c의 모델 f_c^t의 출력의 성능이 다른 클라이언트에 비해 좋을 것이다(f_c^t 는 $p_c(x)$ 에서 loss 를 최소화하도록 학습하므로). 따라서, 제시하는 가중치 w_c 는 앙상블 과정에서 해당 데이터를 더 잘 학습했을 클라이언트 모델에 더 큰 가중치를 주는 알고리즘이라고 할 수 있다.

위 가중치를 이용하여 다음 정리를 증명할 수 있다.

정리 2. Loss 가 convex 하고, D는 모든 클라이언트 데이터 분포의 평균 분포이고, $L_D(f)$는 모델 f의 D에서의 평균 loss, f_c는 클라이언트 c의 데이터 분포에서 평균 loss 가 가장 작은 모델이라 하자. 이 때 모든 모델 f에 대해,

$$L_D\left(\sum_c w_c^* f_c\right) \le L_D(f).$$

이는 제시하는 앙상블로 생성한 의사 라벨이 어떠한 단일 모델로 생성한 의사 라벨의 성능보다 좋음을 의미한다.

다음은 CIFAR-10 에서 C=20 이고 클라이언트 데이터 분포가 Dirichlet(0.1) 분포를 따를 때(충분히 클라이언트 데이터 분포가 불균등할 때), FedAVG, 그리고 균등 가중치 $w_{c_{DF}}$, 분산 기반 가중치 $w_{c_{ET}}$, 마지막으로 본 논문이 제시하는 생성적 적대 신경망 기방 가중치 w_c^* 들을 이용한 앙상블로 50 라운드 학습한 모델의 최종성능을 나타낸 표이다. 표의 값은 3 회 서로 다른 랜덤 시드에 대해 반복하여 평균을 취한 값이다.

알고리즘	최종 성능(정확도, %)
FedAVG	47.95
$w_{c_{DF}}$ 기반 앙상블 증류	72.32
$w_{c_{ET}}$ 기반 앙상블 증류	71.60
w_c^* 기반 앙상블 증류	**80.96**

이는 본 논문이 제시하는 가중치 기반 앙상블의 성능이 이미지 데이터셋에서 잘 동작함을 보인다.

Ⅲ. 결 론

본 논문에서는 생성적 적대 신경망을 이용하여 최적의 앙상블을 수행할 수 있는 가중치 모델링을 제시했고, 이를 통해 높은 성능의 이미지 분류 모델을 학습하였다.

ACKNOWLEDGMENT

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No, 2022R1A2C2092151).

참 고 문 헌

[1] McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." Artificial intelligence and statistics. PMLR, 2017.

[2] Bernstein, Jeremy, et al. "signSGD: Compressed optimisation for non-convex problems." International Conference on Machine Learning. PMLR, 2018.

[3] Geyer, Robin C., Tassilo Klein, and Moin Nabi. "Differentially private federated learning: A client level perspective." arXiv preprint arXiv:1712.07557 (2017).

[4] Truex, Stacey, et al. "LDP-Fed: Federated learning with local differential privacy." Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking. 2020.

[5] Li, Tian, et al. "Federated optimization in heterogeneous networks." Proceedings of Machine learning and systems 2 (2020): 429-450.

[6] Lin, Tao, et al. "Ensemble distillation for robust model fusion in federated learning." Advances in Neural Information Processing Systems 33 (2020): 2351-2363.

[7] Cho, Yae Jee, et al. "Heterogeneous ensemble knowledge transfer for training large models in federated learning." arXiv preprint arXiv:2204.12703 (2022).

[8] Goodfellow, Ian, et al. "Generative adversarial networks." Communications of the ACM 63.11 (2020): 139-144.