

연합학습을 위한 액티브 러닝 기법

안진현, 마예은, 박서윤, 유철우
명지대학교

wlsgus3396@mju.ac.kr, k41031614@mju.ac.kr, dhn04100@mju.ac.kr, cwyou@mju.ac.kr

A Study on Active Learning for Federated Learning

Jin-Hyun Ahn, Yeeun Ma, Seoyun Park, and Cheolwoo You
Myongji University

요약

연합학습은 통신 효율성, 개인 정보 보호 및 공정성 측면에서 많은 연구가 이뤄졌으나 각 기기에서 이뤄지는 레이블링과 관련된 연구는 상대적으로 덜 진행되어왔다. 이 논문에서는 연합학습 프레임워크에 액티브러닝을 적용하는 연구를 제안한다. 각 기기에서 개별적으로 액티브러닝을 적용하는 기법보다 더 나은 방법으로써 연합학습을 위한 액티브러닝 (FAL) 을 제안한다. 이 방법에서 클라이언트는 분산 최적화 방식으로 연합학습에 더 많은 정보량을 제공하는 인스턴스를 획득한다. 전통적인 무작위 샘플링 전략, 클라이언트 수준 별도의 액티브러닝 (SAL) 및 제안된 FAL 을 비교하여 연합학습 모델의 테스트 정확도를 비교한다. 실험적으로 제안된 FAL 이 이미지 분류 작업에서 다른 기법보다 나은 성능을 갖음을 보인다.

I. 서론

레이블링은 연합학습에서 많이 연구되고 통신 효율, 공정성, 강인함, 개인화 등의 이슈 [1-2]이외에도 실제 구현에서 간과되거나 무시되어서는 안 되며, 여러 클라이언트가 고려되기 때문에 레이블링 비용이 일반적인 중앙집중형 머신러닝보다 훨씬 높을 수 있다. 그 이유로 연합학습의 경우에 레이블링 비용을 반드시 고려해야 하며 본 논문에서는 이 문제를 고려하여 연합학습 프레임워크에서 레이블링 전략을 연구한다.

기존 연합학습 연구들과 다르게, 실제적으로 연합학습에 참여하는 클라이언트는 연합학습 실행 이전에 학습 데이터셋에 레이블을 지정해야 한다. 본 논문에서는 각 클라이언트가 레이블링 전략을 적용하여 효과적인 인스턴스에 레이블링을 진행하는 스텝을 고려한다.

본 논문에서는 연합학습을 위한 액티브 러닝 전략을 제안하는데, 이는 연합학습에 참여하는 클라이언트들이 각각 국소적인 학습이 목표가 아니라, 연합학습이 목표가 되기 때문에 연합학습 측면에서 정보량이 높은 인스턴스를 우선적으로 샘플링하는 방법이다. 이를 위해서 각 클라이언트는 액티브 러닝을 위해 학습하는 모델을 국소적으로 학습시키지 않고 연합학습을 통해 학습시켜서, 연합학습 측면에서 효율적인 인스턴스를 고를 수 있게 한다.

제안된 방법을 검증하기 위해 레이블링 단계가 있는 연합학습 프레임워크를 설정하였으며, 이 프레임워크 하에서 다양한 레이블링 전략을 비교한다: 1) 무작위 샘플링을 사용한 전통적인 연합학습, 2) 클라이언트 수준에서 별도의 액티브러닝을 적용 (SAL), 그리고 3) 제안하는 기법인 연합학습을 위한 액티브러닝 (FAL).

액티브러닝은 분명히 전통적인 중앙 집중식 학습에서 무작위 샘플링보다 우수한 성과를 보였다. 하지만

액티브러닝을 연합학습 프레임워크에서 고려하고 액티브러닝이 연합학습의 성능에 미치는 영향을 조사한 연구는 거의 없었다. Ahmed 및 Aussel의 연구 [3-4]는 SAL 이 무작위 선택 (전통적인 FL)보다 우수하며 액티브러닝 연합학습 및 중앙 집중식 학습 환경 모두에 동일하게 유익하다는 것을 실험적으로 보였다. 본 연구는 연합학습 환경에서 액티브러닝이 레이블링 비용을 줄일 수 있으며 제안한 FAL 이 연합학습 환경에서 액티브러닝의 성능을 크게 향상시킨다는 것을 보인다.

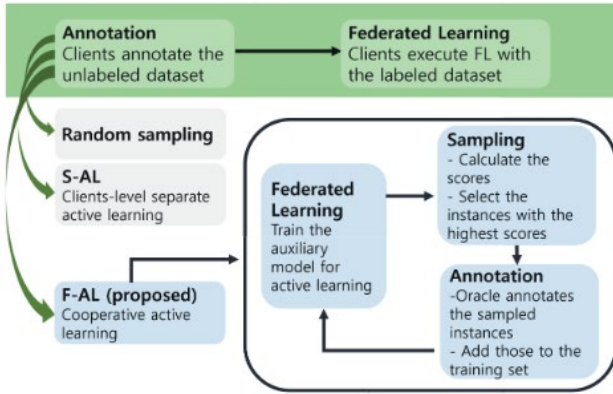
다양한 액티브러닝 알고리즘과 데이터셋을 사용한 다양한 실험을 통해 액티브러닝이 연합학습 환경에서 효과적임을 실험적으로 입증합니다. 수치 결과는 액티브러닝 방법이 연합학습 모델의 테스트 정확도 측면에서 무작위 샘플링을 능가한다는 것을 나타낸다.

또한 FAL 이 다른 방법들보다 우수함을 보여서, 제안한 방법이 향후 연합학습의 실제적인 적용 시에 우수한 기술로서 적용될 수 있음을 보인다.

II. 본론

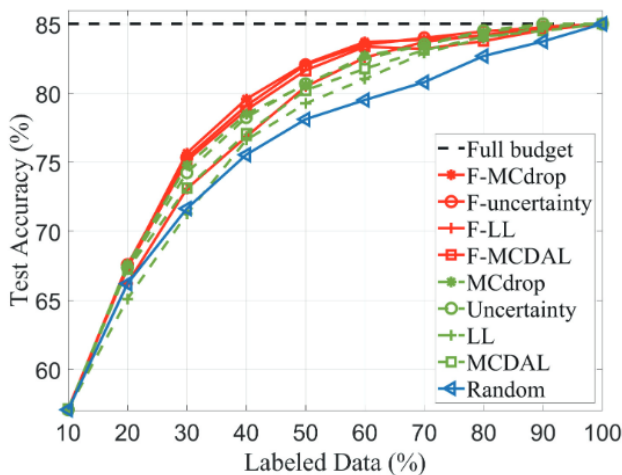
본논문에서는 연합학습에 참여하는 클라이언트에 초기 레이블링된 샘플이 주어졌을 때, 주어진 레이블링 한도 내에서 샘플링하는 상황을 고려한다. 이 때, 동일한 한도 하에서 더 높은 성능을 보이는 전략이 높은 레이블링 전략이라고 여겨질 수 있다. 아래 그림과 같이 세가지 방법을 고려하는데, 첫 번째로 Random sampling 은 각 클라이언트 별로 레이블링을 적용할 시에 어떠한 액티브러닝 전략도 활용하지 않고 임의로 골라서 레이블링을 하는 방법을 뜻한다. 두 번째로 SAL 같은 경우에는 클라이언트 별로 독자적으로 액티브러닝을 적용하는 전략을 의미하고, 마지막으로 FAL 이 본 논문에서 제안하는 전략인데, FAL 하에서는

클라이언트들이 액티브 러닝 단계에서 연합학습을 수행하여 연합학습 측면에서 효과적인 샘플을 골라서 레이블링을 수행하는 방법이다.



위 세 가지 전략의 성능을 실험하기 위해 아래와 같은 실험환경을 고려하였다. CIFAR10 이미지 분류 태스크를 고려하였으며 5 개의 클라이언트가 각각 서로 겹치지 않는 10000 개의 이미지를 갖고 있다고 가정하였다. 초기에는 10%에 해당하는 1000 개의 이미지가 라벨링되어있는 환경을 고려하였으며 각 라운드당 10%가 추가적으로 라벨링되는 실험환경을 가정하였다. 학습 모델은 Resnet-18 을 고려하였고, 각 클라이언트의 국소적 학습 수는 1, 글로벌 학습 수는 로스가 특정 임계치 이하로 떨어질 때까지 반복하는 환경을 가정하였다. 그래프의 각 데이터는 3 번의 실험의 평균 테스트 정확도를 바탕으로 나타나진다.

액티브러닝 방법으로는 최대 엔트로피, MCdrop [5], Learning loss [6], MCDAL [7] 방법을 고려하였다. 아래 그래프에서 “F-“ 가 붙은 성능은 각 액티브러닝 기법을 FAL 전략에 적용한 기법의 성능을 나타내고, 초록선들은 SAL 전략의 성능을 의미한다. 아래 그림에서 볼 수 있듯이, 전체적으로 다양한 액티브러닝 전략에 대해서 FAL 전략이 훨씬 우수한 성능을 보임을 확인할 수 있으며, 또한 FAL 과 SAL 둘다 Random sampling 보다 확연하게 우수한 성능을 보임을 확인할 수 있다.



III. 결론

본 논문에서는 연합학습 환경에서 레이블링 전략으로서 FAL 을 제안하였다. 비교 전략으로서

Random sampling 과 SAL 을 비교하여 FAL 이 우수한 성능을 나타냄을 확인하였다. FAL 은 각 클라이언트가 액티브러닝을 적용 할 때 협동적으로 연합학습을 수행하여 연합학습 시에 효율적인 샘플을 선택하는 방법으로서 레이블링이 높은 비용을 요구하는 현 기계학습 산업계에 유용하게 적용될 수 있을 것이라고 기대된다.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No.1711195089, RS-2023-00212836).

참 고 문 헌

- [1] V. Smith, C.-K. Chiang, M. Sanjabi, and A. Talwalkar, “Federated multi-task learning,” arXiv preprint arXiv:1705.10467, 2017.
- [2] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks,” Proceedings of Machine Learning and Systems, vol. 2, pp. 429– 450, 2020.
- [3] L. Ahmed, K. Ahmad, N. Said, B. Qolomany, J. Qadir, and A. Al-Fuqaha, “Active learning based federated learning for waste and natural disaster image classification,” IEEE Access, vol. 8, pp. 208 518– 208 531, 2020.
- [4] N. Aussel, S. Chabridon, and Y. Petetin, “Combining federated and active learning for communication-efficient distributed failure prediction in aeronautics,” arXiv preprint arXiv:2001.07504, 2020
- [5] Y. Gal and Z. Ghahramani, “Dropout as a bayesian approximation: Representing model uncertainty in deep learning,” in international conference on machine learning. PMLR, 2016, pp. 1050– 1059.
- [6] D. Yoo and I. S. Kweon, “Learning loss for active learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 93– 102.
- [7] J. W. Cho, D.-J. Kim, Y. Jung, and I. S. Kweon, “Mcdal: Maximum classifier discrepancy for active learning,” arXiv preprint arXiv:2107.11049, 2021.