

연합학습에서의 Heterogeneous 로컬 모델들을 고려한

Batch Statistics Tracking 방법 분석

김민재[○] 문수목

서울대학교 전기정보공학부

mjkim@snu.ac.kr, smoon@snu.ac.kr

Analysis of Batch Statistics Tracking Methods Considering

Heterogeneous Local Models in Federated Learning

Minjae Kim[○] Soo-Mook Moon

Dept. of Electrical and Computer Engineering, Seoul National University

요 약

연합학습은 데이터를 한곳에 모으지 않고, 각기 다른 데이터를 가지고 있는 로컬 장치들이 서로의 데이터로 학습을 진행하고 그 결과물을 서버와 주고받으며 단일 모델을 학습시키는 방법론이다. 연합학습은 그 특성상 클라이언트들이 직접 학습을 진행하게 되므로 클라이언트의 연산 및 통신 능력을 고려하여 각 클라이언트가 서로 다른 크기의 로컬 모델을 가지는 학습 방법들이 연구되고 있다. 하지만 각 클라이언트가 학습하는 로컬 모델의 크기가 다른 경우엔, Batch Normalization 방법을 적용함에 있어 평균과 분산과 같은 통계값들을 어떤 식으로 트래킹 해야 할지에 대한 명확한 해결책이 없는 상황이다. 본 연구에서는 이러한 상황에서 효과적인 배치 트래킹을 위한 다양한 방법들을 실험을 통해 비교 분석해보았고, 그 결과 적은 수의 클라이언트 및 적은 양의 데이터로도 효과적인 Batch Statistics Tracking이 가능함을 확인하였다.

1. 서 론

연합학습은 분산학습의 일종으로, 학습에 사용되는 데이터들을 한곳에 모아 학습을 진행하는 일반적인 학습 방법들과는 달리 자신만의 데이터를 가지고 있는 클라이언트들이, 학습한 로컬 모델들을 서버와 주고받으며 단일한 글로벌 모델을 학습한다. 클라이언트들은 이를 위해, 로컬에서 모델을 학습할 수 있는 연산 능력뿐만 아니라 서버와 모델을 주고받을 수 있는 통신 능력 또한 요구된다. 최근에 소개된 연합학습 알고리즘은 서버와 클라이언트 간의 통신 비용을 크게 줄였지만, 여전히 로컬 모델을 학습하기 위한 연산 능력은 클라이언트에게 요구된다. 만약 클라이언트의 리소스가 부족할 경우 연합학습에 참여할 수 없는 문제를 해결하기 위해 클라이언트의 리소스를 고려하여 로컬 모델의 크기를 유연하게 조절하는 연구들이 진행되고 있다. 대표적으로 HeteroFL [1]에서는 로컬 모델들이 글로벌 모델의 커널의 부분집합을 가지도록 하였다.

여러 가지 정규화 방법 중 Batch Normalization은 대부분의 모델이 채택하고 있는 가장 표준적인 정규화 방법이다. Batch Normalization은 학습 과정에서 모든 레이어의 입력값들을 미니 배치의 평균과 분산을 이용해 정규화하는 동시에 inference 시에 사용할 평균과 분산 통계를 트래킹한다. 하지만 로컬 모델의 크기가 서로 다르고 데이터가 분산된 연합학습의 경우에는 어떠한 방식으로 통계를 트래킹해야 할지가 명확하지 않다. HeteroFL에서는 학습 중간에는 각 레이어의 입력값들의 평균과 분산을 트래킹 하고 있지 않다가, 학습이 종료된 후에 클라이언트들에게 query 하는 방식의 sBN(Static batch normalization)을 제시하였다.

하지만 로컬 모델의 크기가 서로 다른 이유가 리소스가 부족한 클라이언트들을 위해서라는 점을 고려하면, 리소스가 부족한 클라이언트들이 배치 정규화에 쓰일 평균과 분산을 측정하기 위해 본인의 로컬 모델보다 큰 글로벌 모델을 inference 하지 못할 수도 있을뿐더러 단순히 각 레이어의 입력들에 대한 통계치를 알기 위해 모든 클라이언트가 모든 데이터에 대한 inference를 진행하는 방식은 비효율적이다. 따라서 본 연구에서는 이러한 경우에 배치 정규화에 사용될 통계치를 얻을 수 있는 효율적인 방법들을 고안해보았다. 우선 기존의 방식대로 학습 중간에 꾸준히 트래킹하면서, 단순히 이를 평균 내는

* 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2020R1A2B5B02001845).

기본적인 방식을 고려하였다. 또한, 학습이 끝나고 클라이언트들에게 query를 할 때, HeteroFL처럼 모든 클라이언트의 모든 데이터가 아니라, 특정 클라이언트 또는 클라이언트가 가진 데이터 일부분만 활용하는 방식을 실험해 보며 효율적인 트레이킹이 가능한지 분석해보았다.

2. 배경 지식

가장 잘 알려진 연합학습 방법은 FedAvg[2] 으로, 연합학습에 참여하는 클라이언트들이 글로벌 모델을 내려받은 후, 본인이 가지고 있는 데이터로 로컬 모델을 학습하고, 서버가 업로드된 로컬 모델들의 파라미터를 평균 내는 방식으로 글로벌 모델의 학습이 진행된다. 이러한 연합학습은 의료계 데이터나, 스마트폰에서의 Next word prediction과 같이 주로 Privacy가 중요한 학습 환경에서 활용될 수 있다. 연합학습에 참여하는 클라이언트들은 서로 다른 환경에서 데이터를 얻기 때문에 데이터들이 서로 heterogeneous 하며, 매우 많은 클라이언트가 학습에 참여하고 한 라운드에 전체 클라이언트의 일부분만 학습에 참여하는 등의 여러 특성이 있다. 특히 클라이언트가 가지고 있는 데이터의 양뿐만 아니라 데이터들이 속해있는 class 들의 분포도 heterogeneous하며, 이는 글로벌 모델을 방해하는 매우 중요한 요소이다.

Batch Normalization [3] 은 internal covariate shift 현상을 해결하기 위해 내부 레이어들의 입력값들을 정규화하는 방법론이다. 모델을 학습할 때는, 학습에 사용되는 미니 배치의 평균과 분산을 이용하여 정규화를 진행하며, inference 시에는 학습 도중에 트레이킹한 전체 데이터의 평균과 분산 값을 이용하여 deterministic 하게 레이어들의 입력값을 정규화한다.

3. Naive 한 Tracking 방법 분석

기존 배치 정규화 방법과 같이, 클라이언트들이 서로 다른 크기의 로컬 모델을 학습할 때 통계값들을 트레이킹하고 서버에서 단순히 이들을 평균 내어 이를 inference 시에 사용하는 단순한 방식을 먼저 고려해 보았다. HeteroFL 과 마찬가지로, 로컬 모델들이 글로벌 모델의 커널의 부분 집합을 가지도록 설정하였다. 실험은 CIFAR100 데이터셋에 각각 (64, 128, 256, 512) 의 channel을 가진 3x3 Convolutional layer 들과 하나의 Fully connected layer로 이루어진 모델을 글로벌 모델로 하여 진행되었다. 이때 로컬 모델의 크기를 4단계(a,b,c,d) 나누어 실험을 진행하였으며 (a,b,c,d) 에 해당하는 로컬 모델은 각각 글로벌 모델의 커널을 $\left(\frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{4}{4}\right)$ 만큼 Pruning 한 모델이다. 이때, 각 크기의 로컬 모델을 학습하는 클라이언트의 비율은 서로 같다. 로컬 모델의 크기가 서로 다르므로 학습 과정에서 각 레이어의 입력값들의 통계를 트레이킹하더라도, 크기가 작은 로컬 모델을 학습하는 클라이언트의 데이터는 전체 채널 중 일부분만 트레이킹 하게 된다. 이러한 점은 각 클라이언트가 트레이킹하는 통계를

을 서로 일치하지 않게 만들어 성능을 제한하는 요소가 된다.

실험은 데이터의 분포가 IID(Independent and identically distributed) 한 경우와 Non-IID 한 경우, 두 가지 경우로 나누어 진행되었으며, Non-IID 한 경우는 각 client가 특정 class에 속하는 데이터를 가지는 분포가 Dirichlet(0.5) 분포를 따르도록 하였다. 실험 결과는 다음과 같다.

표 1 Tracking 여부에 따른 글로벌 모델의 성능

Distribution	Tracking	Accuracy
IID	No	47.14
	Naive	47.58
	sBN	51.4
Non-IID	No	45.27
	Naive	46.95
	sBN	48.57

위 표에서 볼 수 있듯, 단순한(Naive) 트레이킹 방식은 전혀 트레이킹을 하지 않고 inference 시에서도 mini batch의 통계를 이용하는 것보단 좋은 성능을 보이긴 하지만 sBN 방식에 비해서는 나쁜 성능을 보였다. 따라서 로컬 모델의 크기가 서로 다른 경우에는, 기존의 Naive 한 방식 대신 학습이 끝난 후 하나의 글로벌 모델에 대한 트레이킹 방식이 필요함을 알 수 있다.

4. Tracking 비율 분석

이번에는 HeteroFL에서 제시한 sBN(static batch normalization) 방법을 사용하되, 모든 클라이언트가 모든 데이터를 이용하여 통계를 구하는 것이 아니라, 특정 클라이언트만, 혹은 클라이언트가 가진 데이터 일부분만을 활용하여 트레이킹할때의 글로벌 모델의 성능 변화를 분석해보았다. 우선, 전체 클라이언트 중 트레이킹에 참가하는 비율을 2%에서 10%까지 2% 간격으로 늘려가며 글로벌 모델의 정확도를 비교한 결과는 그림1, 2와 같다. 데이터의 분포가 IID 한 경우인 그림1과, Non-IID 한 그림 2 두 가지 경우 모두 대략 전체 클라이언트 중 6%이상 트레이킹에 참가할 때부터, 모든 클라이언트가 트레이킹에 참여하는 Full sBN 과의 성능 차이가 거의 없음을 확인할 수 있었다. 따라서, 연합학습에서 전체 데이터의 통계를 알고 싶은 경우에 대략 6% 정도의 클라이언트만 트레이킹에 참가하더라도 전체 데이터에 의한 통계값에 근사한 값을 얻을 수 있음을 확인할 수 있다.

다음은, 위 실험에서 조금 더 나아가서, 트레이킹에 참가하는 클라이언트들이 본인이 가지고 있는 데이터의 10%만 트레이킹할 때, 어느 정도의 클라이언트가 트레이킹에 참여해야 하는지에 관한 실험을 진행해보았다. 데이터의 분포가 IID 한 그림 3의 경우와 Non-IID 한 그림 4의 두 가지 경우 모두 대략 전체 클라이언트의 50%~60% 정도

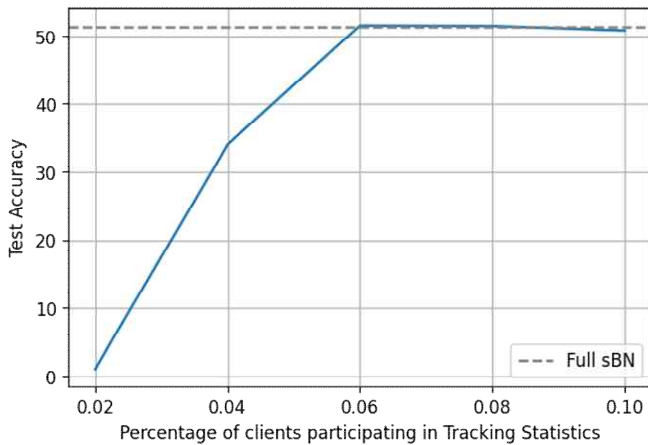


그림 1 . 데이터 분포가 IID 할 때, 트래킹에 참여하는 클라이언트의 수에 따른 글로벌 모델의 정확도 분석

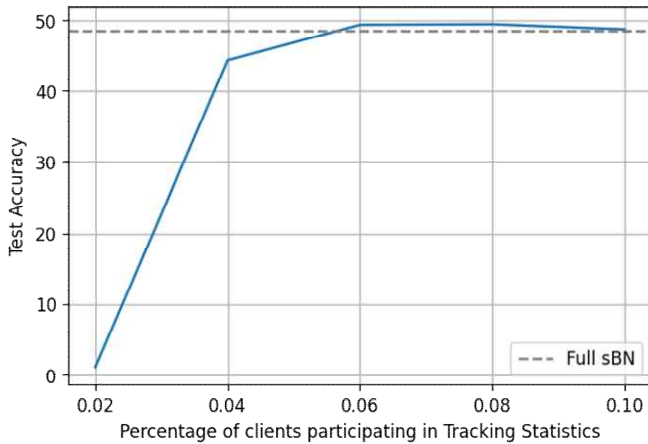


그림 2 . 데이터 분포가 Non-IID 할 때, 트래킹에 참여하는 클라이언트의 수에 따른 글로벌 모델의 정확도 분석

수준에서, 모든 클라이언트가 본인이 가진 모든 데이터를 가지고 Tracking 하는 Full sBN에 근접하였다.

위 실험을 통해 많은 데이터를 트래킹하는 소수의 클라이언트, 혹은 적은 데이터를 트래킹하는 많은 클라이언트 두 가지 경우 모두 효과적으로 전체 데이터의 통계를 추정해낼 수 있음을 확인할 수 있었다. 다만, 트래킹하는 데이터의 양이 극단적으로 적어질 경우, 배치의 크기가 너무 작아져 통계를 추정하는데 어려움이 있을 수 있으므로, 배치 사이즈가 극단적으로 작아지는 것은 경계해야 한다.

5. 결론 및 향후 연구

Batch Normalization은 거의 모든 딥러닝 네트워크가 사용할 정도로 필수적인 요소가 되었지만 로컬 모델의 크기가 서로 다른 경우에, inference 시에 사용할 데이터의 통계를 구하는 데 어려움이 있다. 이에, 본 연구에서는 적은 수의 클라이언트, 혹은 클라이언트가 가지고 있는 데이터 중 일부분만을 이용하더라도 효과적인 트래킹이 가능함을 확인하였다. 다만, 추가적인 트래킹 과정

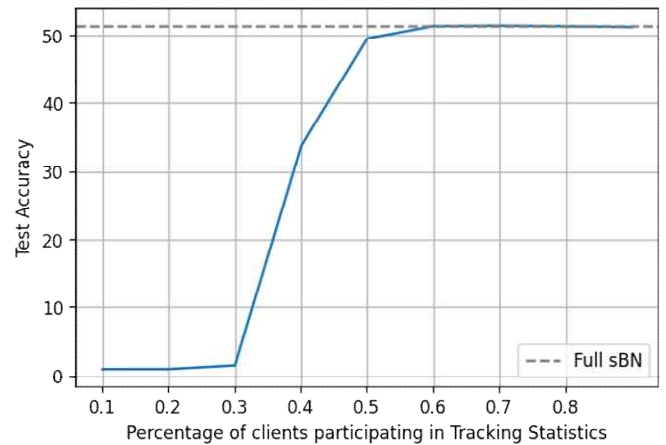


그림 3 . 데이터 분포가 IID 할 때, 가진 데이터의 10%를 트래킹하는 클라이언트의 수에 따른 글로벌 모델의 정확도 분석.

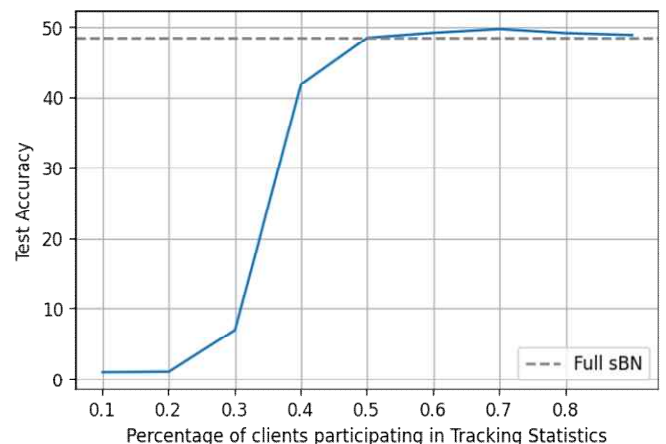


그림 4 . 데이터 분포가 Non-IID 할 때, 가진 데이터의 10%를 트래킹하는 클라이언트의 수에 따른 글로벌 모델의 정확도 분석.

자체가 불필요한 오버헤드일 수 있으므로, 향후에는 학습 과정에서 자연스럽게 학습 데이터들의 통계를 트래킹할 수 있는 방법들이 연구될 것으로 기대된다.

참 고 문 헌

- [1] Enmao Diao, Jie Ding, & Vahid Tarokh (2021). HeteroFL: Computation and Communication Efficient Federated Learning for Heterogeneous Clients. In International Conference on Learning Representations.
- [2] McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." Artificial intelligence and statistics. PMLR, 2017.
- [3] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." International conference on machine learning. PMLR, 2015.