데이터 이질성에 강인한 무선 연합학습을 위한 Over-the-Air 기반 모델 사후분포 집계 기법

홍준표

홍익대학교

jp_hong@hongik.ac.kr

Over-the-Air Posterior Aggregation for Wireless Federated Learning Robust to Data Heterogeneity

Jun-Pyo Hong

Hongik Univ.

요 약

본 논문은 각 단말이 보유한 데이터가 희소하고 분포가 극히 상이한 무선 환경에서 기존 연합학습 기법들이 갖는 학습 장애를 극복하기 위해, 일반적으로 활용되는 빈도주의(frequentist) 학습에서 벗어나 베이지안(Bayesian) 학습을 활용하는 새로운 무선 연합학습 프레임워크를 제안하였다. 즉, 베이지안 접근법을 도입함에 따른 새로운 지역 업데이트 전송 및 집계(aggregation) 기법을 개발하였으며, 학습 데이터가 적고 non-i.i.d한 환경에서 기존 무선 연합학습 기법 대비 큰 학습성능 이득을 가짐을 시뮬레이션을 통해 확인하였다.

Ⅰ. 서 론

연합학습은 단말이 보유한 데이터를 중앙 서버에 전달하지 않고도 인공지능 모델을 학습시킬 수 있다는 특징을 가져, 인공지능 구축에서 데이터 프라이버시 보장 및 지역 맞춤형 인공지능 구축을 위한 목적으로 많은 각광을 받고 있다. 하지만 사용자 간의 취향 및 단말 이용 패턴, 단말의 연산 및 통신 자원이 다를 경우, 지역 학습의 결과가 크게 상이해 이를 종합시 다양한 지역모델들의 중간지점이 도출되어 학습 성능이 저하될 수 있는 문제를 갖는다 [1]. 특히 독립적인 다수의 모바일 단말이 연합학습에 참여하는 현실적인 무선 네트워크 환경에서 이와 같은 특징이 두드러지게 나타난다.

이와 같이 상이한 지역 학습결과에 따라 학습 성능 저하를 나타내는 기존 빈도주의 접근 연합학습의 한계점을 근원적으로 해결하기 위해, 본 논문에서는 베이지안 접근을 도입한 새로운 무선 연합학습을 제안한다.

Ⅱ. 본론

기존의 빈도주의 접근 기반의 연합학습에서 지역학습은 주어진 학습 데이터셋 D_k를 가장 잘 표현할 수 있는 모델의 단일 매개변수 θ_k도출을 목표로 한다. 하지만 본 연구에서 고려하는 베이지안 학습은 데이터셋 D_k를 기반으로 모델 매개변수의 사후분포(posterior distribution) $p(\theta|D_k)$ 도출을 목표로 한다. 도출된 확률 분포는 다양한 매개변수에 대한 신뢰도 정보를 제공함으로써 분산 학습 결과들의 효과적인 집계를 가능케 한다. 즉, 지역 사후분포들 $\{p(\theta|D_1), p(\theta|D_2), ..., p(\theta|D_K)\}$을 대표할 수 있는 분포 $p(\theta|D)$를 도출해 이를 글로벌 사후분포로 업데이트 한다. 이때, 지역 사후분포로부터의 정보손실을 최소화할 수 있도록 글로벌 사후분포는 다음과 같이 지역 사후분포의 곱 연산을 통해 도출된다 [2]

$$p(\theta|D) = \frac{1}{C}\prod_{k=1}^{K} p(\theta|D_k) \quad (1)$$

여기서 C는 정규화 상수이다.

이와 같은 베이지안 접근은 학습에 대한 추가적인 정보를 제공하지만 정확한 지역 사후분포 도출과 전송에는 방대한 연산력과 통신자원을 필요로 한다. 따라서 본 연구에서는 효율성을 위해 사후분포를 정규분포로 제한하고 지역 학습에서 mean-field approximation과 variation inference를 적용해 지역 사후분포를 근사하는 variational 사후분포 $q_{\mathbf{w}}(\theta|D_k)$를 도출한다. 즉, 단말 k의 지역 variational 사후분포는 평균 μ과 표준편차 σ로 구성된 매개변수 $\mathbf{w}$로 결정되며, 각 단말은 $\mathbf{w}$에 대한 업데이트 정보를 전송함으로써 지역 학습결과를 서버로 전달한다. 이때, 모델의 각 매개변수마다의 평균, 표준편차 정보를 전송해야 하기 때문에, 기존의 FedAvg와 비교해 전송해야 할 데이터의 양이 2배로 증가하여 채널 용량이 제한적인 무선 환경에서 병목현상이 악화되는 문제가 발생할 수 있다. 이에 variational 매개변수 전송 및 종합에 Over-the-Air Computation (AirComp)을 고려한다. AirComp는 무선 채널의 중첩현상을 활용한 전송 방법으로써, 모든 단말들이 동일한 채널로 동시에 전송해 수신단이 중첩된 신호를 수신하도록 하여 마치 평균 연산이 전송 중에 수행되도록 하는 방법이다. 기존의 빈도주의 연합학습의 경우, 모델 종합이 지역 모델들의 평균으로 정의되기 때문에 AirComp의 단순 적용으로도 통신을 크게 절감할 수 있었으나 베이지안 연합학습의 경우 (1)과 같이 분포의 곱으로 사후분포 종합이 이루어지기 때문에 AirComp 적용을 위해서는 전송기법의 변형을 필요로 한다. 이에 베이지안 연합학습에서 AirComp를 통해 통신자원을 절감하기 위해서는 송신단과 수신단에서 추가적인 전송 신호처리가 필요하다. AirComp에 대한 구체적인 원리와 방법은 지면 관계상 생략하고, 이에 대한 의사코드는 Algorithm 1과 같다.

제안 기법의 성능 이득을 확인하기 위해 표 1과 같은 제한적인 환경에서 대표적인 연합학습 알고리즘과의 시뮬레이션을 진행하였다. 그림 1을 통해 학습 데이터가 소수이며 극도로 치우져진 극단적인 환경에서도 제안 기법이 상당히 높은 정확도를 달성할 수 있음을 확인할 수 있었다. FedProx[4]의 경우, proximal term 도입을 통해 FedAvg[3]보다는 향상

Algorithm 2 Bayesian FL with Over-the-Air Posterior Aggregation

Initialize:
  $t \leftarrow 0$
  $\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t \leftarrow$ Random initialization
BS distributes $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ to all devices
while convergence criterion is not met **do**
// Phase 1: Update global covariance matrix
  for each device $k \in \mathcal{K}$ **in parallel do**
    $\boldsymbol{\rho}_{t,k} \leftarrow \mathrm{diag}\left(\boldsymbol{\Sigma}_t^{-1}\right)$
    for $i = 1$ to E **do**
      $\boldsymbol{\rho}_{t,k} \leftarrow \boldsymbol{\rho}_{t,k} - \eta \nabla_{\rho} L_k(\boldsymbol{\mu}_t, \boldsymbol{\rho}_{t,k})$
    end for
    Compute $\boldsymbol{\Delta}_{\rho_{t,k}}$
    Optimize power allocation $\{\mathbf{p}_{t,k}^{(n_1)}\}_{n_1=1}^{N_1}$
    Form $\{\mathbf{x}_{t,k}^{(n_1)}\}_{n_1=1}^{N_1}$ with (30) and transmit them
  end for
  BS reconstructs $\hat{\boldsymbol{\Delta}}_{\rho_t}$ from $\{\mathbf{y}_t^{(n_1)}\}_{n_1=1}^{N_1}$
  $\boldsymbol{\rho}_{t+1} = \boldsymbol{\rho}_t + \bar{\delta}_{\rho_t} \hat{\boldsymbol{\Delta}}_{\rho_t}$
  BS distributes $\boldsymbol{\Sigma}_{t+1} = \mathrm{diag}\left(\boldsymbol{\rho}_{t+1}\right)^{-1}$ to all devices
// Phase 2: Update global mean vector
  for each device $k \in \mathcal{K}$ **in parallel do**
    $\boldsymbol{\nu}_{t,k} \leftarrow \boldsymbol{\Sigma}_{t,k}^{-1} \boldsymbol{\Sigma}_{t+1} \boldsymbol{\mu}_t$
    for $i = 1$ to E **do**
      $\boldsymbol{\nu}_{t,k} \leftarrow \boldsymbol{\nu}_{t,k} - \eta \nabla_{\nu} L_k\left(\boldsymbol{\nu}_{t,k}, \mathrm{diag}\left(\boldsymbol{\Sigma}_{t+1}^{-1}\right)\right)$
    end for
    Compute the normalized update vector $\boldsymbol{\Delta}_{\nu_{t,k}}$
    Optimize power allocation $\{\mathbf{p}_{t,k}^{(n_2)}\}_{n_2=N_1+1}^{N_1+N_2}$
    Form $\{\mathbf{x}_{t,k}^{(n_2)}\}_{n_2=1}^{N_1+N_2}$ and transmit them
  end for
  BS reconstructs $\hat{\boldsymbol{\Delta}}_{\nu_t}$ from $\{\mathbf{y}_t^{(n_2)}\}_{n_2=N_1+1}^{N_1+N_2}$
  $\boldsymbol{\nu}_{t+1} = \boldsymbol{\nu}_t + \bar{\delta}_{\nu_t} \hat{\boldsymbol{\Delta}}_{\nu_t}$
  BS distributes $\boldsymbol{\mu}_{t+1} = \boldsymbol{\nu}_{t+1}$ to all devices
  $t \leftarrow t + 1$
end while

된 학습성능을 달성하였지만 빈도주의 학습의 한계로 그 이득이 제한적임을 확인할 수 있다. SCAFOLD[5]의 경우 control variate을 통해 지역 학습의 발산을 보정함으로써 non-i.i.d. 데이터에 따른 성능 저하를 완화라는데, control variate이 잡음에 상당히 취약한 특성을 가져, 본 연구에서 고려하는 무선 채널 환경에서는 학습 도중 발산하는 현상을 보였다.

데이터셋	MNIST
지역 데이터셋 사이즈	10샘플/단말
데이터 분포	1 label/단말
단말 수	40개
단말 배치	서버를 중심으로 한 2차원 원형 평면 내 균등 분포
네트워크 반경	200m
채널 이득	경로감쇄 + Rayleigh 페이딩
인공지능 모델	6-layer CNN

표 1. 시뮬레이션 환경

Ⅲ. 결론

본 논문에서는 베이지안 접근 기반의 무선 연학학습 기법을 제안하여, 현실적인 무선 네트워크 환경에서 기존의 빈도주의 접근 기반 연합학습이 갖는 한계점을 극복하였다. 후속 연구를 통해 제안 기법의 수렴도 분석과 보다 다양한 환경에서의 성능을 분석할 계획이다.

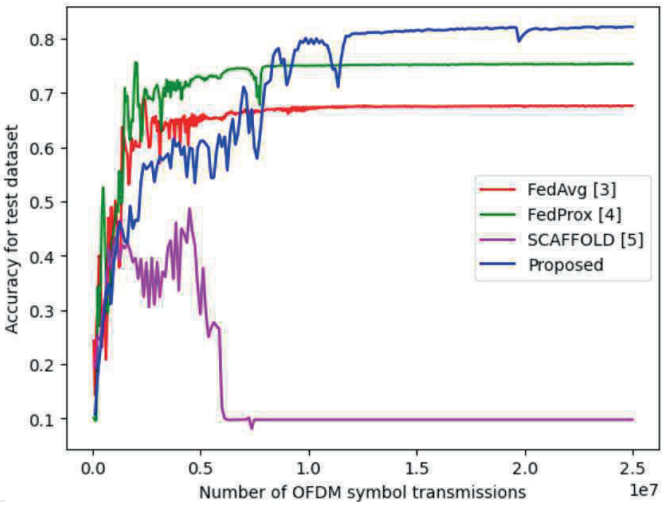

그림 1. 학습성능 비교

ACKNOWLEDGMENT

This work was supported by the National ResearchFoundation of Korea (NRF) Grant funded by the KoreaGovernment (MSIT) under Grant RS-2024-00464570..

참 고 문 헌

[1] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, Federated learning with non-iid data, 2018. arXiv: 1806.00582v2 [cs.LG]

[2] T. P. Hill, “Conflation of probability distributions,” Trans. Amer. Math. Soc., vol. 363, no. 6, pp. 3351-3372, Jun. 2011.

[3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, “Communication-efficient learning of deep networks from decentralized data,” in Proc. Int. Conf. Artif. Intelli. Stat., Apr. 2017.

[4] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwarlkar, and V. Smith, “Federated optimization in heterogeneous networks,” in Proc. Mach. Learn. Syst. Mar. 2020.

[5] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, “SCAFFOLD: Stochastic controlled averaging for federated learning,” in Proc. Int. Conf. Mach. Learn., Jul. 2020.[1] Davies R. W.″ The Data Encryption standard in perspective,″Computer Security and the Data Encryption Standard, pp. 129-132.

[2] Miles E. Smid, ″From DES to AES,″ 2000, (http//www.nist.gov/aes).

[3] Shamir, A. ″On the security of DES,″ Advances in Cryptology, Proc.Crypto ′85, pp. 280-285, Aug. 1985.

[4] NIST, ″Announcing the Advanced Encryption Standard(AES),″FIPS PUB ZZZ, 2001, (http://www.nist.gov/aes).

[5] Daemen, J., and Rijmen, V. “AES Proposal: Rijndael, Version2.,” Submission to NIST, March 1999.