무선 네트워크에서의 베이지안 방법론 기반 적응형 연학학습

장기현, 홍준표
동국대학교, 홍익대학교
lightjang2001@naver.com, *jp_hong@hongik.ac.kr

Adaptive Bayesian Federated Learning for Wireless Networks

Gi hyeon Jang, Jun-Pyo Hong
Dongguk University, Hongik University

요 약

기존 빈도주의 연합학습은 블랙박스적 특성을 가져 학습 과정 및 모델에 대한 해석이 어려워 연산력 및 통신 상황이 비대칭적인 현실적인 무선 엣지 네트워크 환경에서 적응적인 동작이 제한되어 비효율성이 발생하였다. 이와 같은 기존 연합학습의 비효율성을 극복하기 위해 베이지안 접근을 연합학습에 도입해 학습 모델에 대한 해석을 가능케 하며, 이를 기반으로 단말의 통신 및 연산 상황에 따라 적응적으로 로컬학습, 통신을 수행하는 기법을 제안하였다. 시뮬레이션을 통해 제안기법이 지역 모델에 대한 불확실성 정보를 기반으로 효과적인 자원 효율적 지역 학습과 전송을 수행하고, 기존 빈도주의 접근의 연합학습보다 향상된 학습성능을 달성함을 확인하였다.

Ⅰ. 서 론

기존의 중앙집중식 학습을 통한 인공지능 구축에서의 데이터 프라이버시, 높은 서비스 지연과 통신비용 등의 문제를 해결하기 위한 대안으로 연합 학습(Federated Learning, FL) 이 부상하게 되었다[1]. FL은 각 에지 디바이스(edge device)가 본인의 데이터를 서버로 전송하지 않고 분산적으로 모델을 학습하도록 함으로써, 개인정보 유출 위험과 통신 병목 현상을 효과적으로 완화할 수 있다.

분산 학습의 특성상 FL에 참여하는 에지 디바이스들 간의 서로 다른 스펙 및 가용 연산력을 가져 지역 학습에 소요되는 시간이 크게 차이가 날 수 있다. 서버는 모든 에지 디바이스로부터의 지역 학습 결과를 모두 수신한 후 집계를 수행하게 되므로, 일부 가용 연산력이 낮은 에지 디바이스에 의해 연합학습이 크게 지연되는 낙오자 효과(straggler effect)가 발생할 수 있다. 이와 더불어 무선 에지 네트워크 환경에서는 각 에지 디바이스들이 서로 이질적인 통신 채널을 겪게되어 앞서 설명한 낙오자 효과가 심화될 수 있다. 이와 같은 연산 및 통신 자원의 불균형에 따른 학습지연을 완화하기 위해 모델 프루닝(pruning), 업데이트 양자화, 업데이트 압축 등이 연구되었으나 빈도주의 접근 기반의 기존 연합학습은 블랙박스적인 성격을 가져, 학습 결과 및 과정을 고려한 효과적인 연합학습 구축이 어려운 문제를 갖는다.

뿐만 아니라 에지 디바이스들이 상이한 데이터를 보유하는 현실적인 무선 네트워크 환경에서는 에지 디바이스마다 지역 학습의 결과가 크게 달라질 수 있어 서버에서의 집계를 통해 일반화된 전역모델의 수렴이 어려워지는 문제가 발생할 수 있다. 이와 같은 gradient divergence 문제 완화를 위해 FedAvg 알고리즘을 기반으로 FedProx, SCAFFOLD 등 다양한 FL 알고리즘에 개발되었으나 데이터가 희소한 환경에서 큰 성능 개선을 보이지 못하는 문제를 갖는다.

앞서 설명한 연산 및 통신 자원의 비대칭성, 데이터의 희소성과 이질성으로 인한 기존 FL의 한계점을 근원적으로 해소하기 위해, 본 연구에서는 베이지안 접근을 FL에 도입하였다,

베이지안 학습에서는 먼저 사전 확률(prior distribution)을 정의하여 모델 파라미터에 대한 초기의 믿음 또는 가정을 확률적 형태로 나타낸다. 데이터가 수집될 때마다 베이즈 정리(Bayes' theorem)를 적용하여 이 사전 확률을 데이터로부터 관측한 정보를 통해 갱신하며, 이를 통해 사후 확률을 얻는다. 이러한 사후 확률은 모델의 파라미터 값에 대한 모든 가능한 정보를 담고 있으며, 불확실성을 명확하게 표현할 수 있다.

이러한 접근 방식의 주요 이점은 데이터가 제한적이거나 불균형한 환경에서도 효과적으로 일반화가 가능하며, 모델의 예측이나 결정에 대한 신뢰도 평가가 가능하다는 점이다. 특히 베이지안 접근법은 각 파라미터의 불확실성을 수치적으로 표현할 수 있기 때문에, 중요한 파라미터를 효율적으로 선정하거나 자원을 최적화하는 데 활용될 수 있다. 또한 베이지안 접근은 모델의 가중치를 확률분포로 가정하고, 관측된 데이터에 기반한 사후분포를 추론함으로써 불확실성을 효과적으로 반영할 수 있다. 이를 통해 소량의 데이터나 이질적인 환경에서도 높은 일반화 성능을 달성할 수 있으며, 매개변수의 신뢰도를 기반으로 연산 및 통신 자원에 적응적인 학습 전략을 설계할 수 있다는 장점도 지닌다.

즉, 제안기법은 베이지안 방법론을 통해, 모델 매개변수에 대한 신뢰도 정보에 기반하여 학습모델을 해석하고 가용 연산력/통신 상황에 적응적으로 동작할 수 있게 함으로써, 기존 연합학습 기법의 한계를 극복해 학습을 가속화하고 학습성능을 최대화하는 것을 목표로 한다.

Ⅱ. 본론

본 논문에서는 하나의 기지국과 여러개의 에지 디바이스로 이루어진 무선 에지 네트워크 환경에서 (Federated Learning)을 수행하는 상황을 다룬다. 에지 디바이스들은 non-i.i.d. 데이터셋 D_k를 가지고 있으며, D_k의 크기는 평균 $\bar{\lambda}$인 포아송(poisson) 분포를 따른다. 서버는 각 라운드 t의 시작 시점마다 파라미터 $\boldsymbol{\theta}_t = (\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) \in \mathbb{R}^{2d}$를 기반으로 불확실성을

$$u_{t,i} = \frac{\sigma_{t,i}}{|\mu_{t,i}|}. \quad (1)$$

다음과 같이 정량화하여 각 에지 디바이스에게 전달한다. 여기서 $\boldsymbol{\mu}_t = [\mu_{t,1}, \mu_{t,2}, ..., \mu_{t,d}]$와 $\Sigma_t = \mathrm{diag}(\sigma_{t,1}^2, \sigma_{t,2}^2, ..., \sigma_{t,d}^2)$는 각각 분포의 평균과 대각화된 공분산 행렬을 나타낸다. 각 에지 디바이스는 전달받은 불확실성 정보를 바탕으로 자신이 보유한 연산 및 통신 자원의 제한 안에서 불확실성이 높은 일부 파라미터에 대해서만 집중적으로 로컬 학습을 진행하고 결과를 서버로 보낸다. 반면 불확실성이 낮은 파라미터는 그대로 유지하여 자원을 효율적으로 활용하도록 한다.

연산 부하와 무선통신환경을 바탕으로 최대화된 학습 파라미터의 수 $d_{t,k}$와 송신 파라미터 수 $\bar{d}_{t,k}$는 다음과 같이 정의될 수 있다

$$d_{t,k} = \min\left\{d, \lfloor \frac{\bar{\tau}_{\mathrm{train,t}} C_k}{2D_k(c_{\mathrm{fwd}}M + c_{\mathrm{bwd}})} \rfloor - d\right\}, \quad (2)$$

$$\bar{d}_{t,k} = \min\left\{d_{t,k}, \lfloor \frac{(\bar{\tau}_{\mathrm{round}} - \bar{\tau}_{\mathrm{train,t}})\mathrm{R}_{\mathrm{comm,t,k}}}{2q} \rfloor\right\}. \quad (3)$$

$\bar{\tau}_{train,t}$동안, 각 장치 k는 각자의 데이터셋 D_k을 사용해 글로벌 분포 $q_{\tilde{\theta}_{t,k}}(\mathrm{w})$를 사전분포로 간주하여 변분추론(variational inference)을 수행한다. 구체적으로는 장치 k의 손실함수는

$$L_{k(\tilde{\theta})} = \lambda KL(q_{\tilde{\theta}}(\mathrm{w}) \| q_{\tilde{\theta}_{t,k}}(\mathrm{w})) - \mathrm{E}_{\mathrm{q}_{\tilde{\theta}}(\mathrm{w})}(\log\ p(D_k|\mathrm{w})), \quad (4)$$

로 주어진다. 여기서 λ는 로컬 학습동안 글로벌 사후분포로부터의 편차 정도를 조절하는 양의 양수를 의미한다. 두 가우시안 변분 분포간의 KL 발산(Kullback-Leibler Divergence)은

$$\mathrm{KL}\left[\mathrm{q}_{\tilde{\theta}}(\mathbf{w}) \| \mathrm{q}_{\tilde{\theta}_{\mathrm{t,k}}}(\mathbf{w})\right] = \frac{1}{2}(\log\frac{|\tilde{\Sigma}_{\mathrm{t,k}}|}{|\check{\Sigma}|} - \mathrm{d} + \mathrm{tr}\left[\tilde{\Sigma}_{\mathrm{t,k}}^{-1}\check{\Sigma}\right] + (\tilde{\mu}_{t,k} - \check{\mu})^T \tilde{\Sigma}_{t,k}^{-1}(\tilde{\mu}_{t,k} - \check{\mu})), \quad (5)$$

으로 주어진다.식 (4)에서 기대값 항은 로컬 학습 중 변분 파라미터가 로컬데이터셋에 보다 잘 들어맞도록 유도하는 역할을 한다. 하지만 현실적으로 기대값 항을 계산하는 것은 어렵기 때문에, MC 샘플링을 통해

$$\mathbb{E}_{q_{\tilde{\theta}}(\mathbf{w})}[\log\ p(D_k|\mathrm{w})] \approx \frac{1}{\mathrm{M}}\sum_{m=1}^{M}\log\ p(D_k|\mathrm{w}^{(m)}), \quad (6)$$

로 나타낼 수 있다. 여기서 $\mathbf{w}^{(\mathbf{i})}$는 변분 사후 분포 $q_{\tilde{\theta}}(\mathrm{w})$로부터 추출한 i-번째 MC 샘플을 의미한다. 결과적으로 (5)와 (6) 식을 사용하여, 장치 k에서의 로컬 학습을 마친 후의 변분 파라미터 $\tilde{\theta}_{t,k}^{*} = (\tilde{\mu}_{t,k}^{*}, \tilde{\Sigma}_{\mathrm{t,k}}^{*}) \in \mathrm{R}^{2d_{t,k}}$는, 확률적 경사 하강법을 통해서 구해진다.

이 최대 파라미터 수를 토대로 디바이스는 주어진 자원 내에서 변분 파라미터 $\check{\theta}_{t,k}^{*} \in \mathbb{R}^{2\bar{d}_{t,k}}$를 전송하며, 만약 특정 디바이스가 학습 결과를 완전히 전송하지 못한다면, 서버는 수신되지 않은 부분을 글로벌 모델의 기존 값으로 보완하여 전역 모델을 갱신한다.

서버는 수신한 지역 모델들의 정보를 바탕으로 글로벌 사후 분포 $\boldsymbol{\theta}_{t+1} = (\mu_{t+1}, \Sigma_{t+1})$를 다음과 같이 갱신한다

$$\Sigma_{t+1}^{-1} = \sum_{k\in K}\pi_k \Sigma_{t,k}^{-1}, \quad 4()$$

$$\mu_{t+1} = \Sigma_{t+1}\sum_{k\in K}\pi_k \Sigma_{t,k}^{-1}\mu_{t,k}, \quad (7).$$

이후로는 $\boldsymbol{\theta}_{t+1}$를 다음 라운드에 모든 디바이스에게 다시 전달하는 과정을 반복한다. 이때 서버는 각 디바이스의 연산력과 통신 자원을 고려하여 전체 네트워크의 학습 성능이 최대화될 수 있도록 지역 학습 시간과 통신 자원을 최적화하여 할당한다.

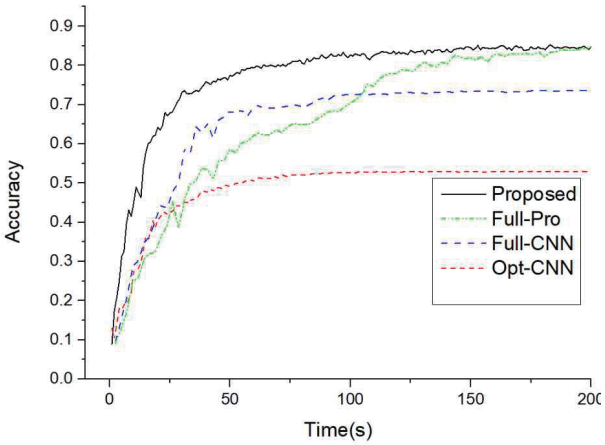

그림 1 연합학습 기법 성능 비교

그림 1의 Opt-CNN은 기존 Frequentist CNN[4]의 논문의 우선도 지표를 바탕으로 본 논문에서의 최적화 기법을 적용하였고, Full-CNN은 최적화를 적용하지 않은 일반 CNN과 비교하였다. Full-Pro는 최적화를 진행하지 않고 모든 파라미터를 전부 보낸 베이지안 접근을 활용한 BNN이다. 그림 1의 내용을 분석한다면, 기존의 BNN은 보내는 파라미터의 양이 CNN과 비교하였을때 2배의 양을 보내기 때문에 모든 모델이 학습속도가 빠른 초기에는 비교적으로 낮은 성능을 갖지만, 빈도주의 기반 딥러닝이 가지는 성능적 한계 때문에 결국 성능의 역전이 일어나게 된다. 한편 본 논문의 제안방안은 모든 비교방안을 통틀어서 수렴 속도와 성능 두가지 방면에서 모두 좋은 성능을 보여주며, 기존의 연합학습에서 추구하고자 하는 목표를 성공적으로 달성하였다고 볼 수 있다.

Ⅲ. 결론

본 논문에서는 가용 연산력과 통신상황을 반영하는 베이지안 방법론을 활용한 적응형 연합학습을 구현하였다. 가중치와 무선통신 환경에 대한 수학적인 분석을 바탕으로 데이터 희소성, 이질적 환경등 기존 연합학습의 한계점으로 꼽히던 문제들을 해결하였고 시뮬레이션을 통해, 적응형 빈도주의 딥러닝, 베이지안 연합학습 등 다양한 비교방안과의 수렴 속도와 성능 면에서의 차이를 보였으며, 이를 통해 효과적으로 기존 연합학습 기법의 한계를 극복하였다는 것을 보였다.

참 고 문 헌

[1] Yeo Ung Gi, Lee Joon Woo, “A Survey on Trends and Analysis of Robust Federated Learning Techniques,”Proceedings of Symposium of the Korean Institute of communications and Information Sciences, Jeju, 2024.

[2] J. Nguyen, et al. "Federated Learning with Buffered Asynchronous Aggregation," in Proc. 25th Int. Conf. Artificial Intelligence and Statistics (AISTATS 2022), pp. 3581 - 3607, Virtual Event, Mar. 2022.

[3] L. Liu et al., "A Bayesian Federated Learning Framework With Online Laplace Approximation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 46, no. 1, pp. 1 - 16, Jan. 2024.

[4] Y. Jiang et al.,"Model Pruning Enables Efficient Federated Learning on Edge Devices,"IEEE Trans. Neural Netw. Learn. Syst., vol. 34, no. 12, pp. 10374 - 10386, Dec. 2023.