클라이언트 자가 균형 연합학습 알고리즘[1)]

배상민[O*] 안수명[*] 윤세영
KAIST 김재철 AI대학원
{bsmn0233, sumyeongahn, yunseyoung}@kaist.ac.kr

Federated Client-side Self-Balancing Algorithm with Entropy Regularization

Sangmin Bae[O*] Sumyeong Ahn[*] Se-Young Yun
KAIST Kim Jaechul Graduate School of AI

요 약

최근 엣지 디바이스의 광범위한 활용성으로 인해, 엣지 디바이스의 한정적인 계산 자원과 각각의 개인 데이터 보안을 유지하며 학습에 사용하는 방법인 연합학습(Federated Learning)에 관한 관심이 많아지고 있다. 연합학습은 크게 개인 데이터를 보유하고 있는 '클라이언트(로컬)'와 모델을 각 클라이언트에게 배포하고 취합하는 '서버(글로벌)'의 통신으로 학습이 진행된다. 이 때, 서버는 클라이언트(엣지 디바이스)의 보안을 침해할 수 없는 상황과 각 클라이언트의 데이터 분포가 불균형한 상황을 가정한다. 일반적으로 최근 연구들은 데이터의 불균형도를 학습된 서버 모델을 통해 간접적으로 측정하여 알고리즘에 사용한다. 하지만 개인 데이터에 대한 보안으로 인해 서버는 실제 데이터에 접근하여 불균형도를 정확하게 측정할 수 없다. 본 논문에서는 클라이언트 측에서 데이터의 불균형도를 고려해, 학습 데이터 클래스의 분포가 균등하도록 배치를 재구성하는 방식을 제안한다. 나아가, 소수 클래스의 데이터에 대한 과적합 문제를 해결하고자 엔트로피 정규화 손실 함수를 제안한다. 제안한 알고리즘은 타 알고리즘과 비교하여 연합학습 상황에서 이미지 분류 성능을 월등히 높여주었다.

1. 서 론

1.1 연구 동기

연합학습(Federated Learning, FL)은 개인 데이터 보안을 침해하지 않고, 클라이언트(또는 엣지 디바이스, 로컬)에서 서버의 글로벌 모델을 학습하기 위한 알고리즘으로 연구되고 있다. [1]에서는 **연합학습의 세 가지 제약 조건**을 다음과 같이 요약하였다. **(1) 데이터 보안**: 개인 정보 보호를 위해, 각 로컬 데이터는 다른 클라이언트 또는 서버가 접근할 수 없다. **(2) 데이터 이질성**: 각 엣지 디바이스는 각기 다른 분포의 훈련 데이터를 가지고 있다. **(3) 통신 대역폭 제한**: 서버와 클라이언트 사이에 통신 대역폭이 제한되어 있다.

[1]은 위의 제약 조건 속에서 단순하고 효율적으로 학습이 가능한 **Fed**erated **Av**era**g**ing (**FedAvg**) 알고리즘을 제안했다. 매 라운드마다 서버는 선택된(또는 사용 가능한) 클라이언트에게 글로벌 모델을 배포해준다. 그리고 각 클라이언트는 개인 데이터를 사용하여 로컬 모델을 업데이트하고, 서버는 각 장치의 데이터 샘플 수에 비례한 가중치 평균을 통해 새로운 글로벌 모델을 만든다.

최신 연구들은 데이터 불균형 상황에서 FedAvg의 성능이 급격하게 떨어짐을 지적하며, 학습된 서버 모델을 통해 간접적으로 현재 모델이 균등한 데이터를 배우도록 알고리즘을 설계했다. 매 라운드에서 학습에 참여할 클라이언트를 선택할 때, 현재 모델에 대한 손실값이 가장 큰 클라이언트 [4], 혹은 학습된 신경망 분류기의 그래디언트 크기가 가장 균등한 클라이언트를 [5] 선택하하는 선행 연구들이 있다. 하지만, 서버 측에서 모델의 파라미터로만 간접적으로 추정했다는 한계점과 각 클라이언트 자체의 데이터 불균형은 여전히 해결할 수는 없다는 문제점이 있다.

본 논문에서는 학습된 글로벌 서버 모델이 아닌, 유일하게 데이터에 접근 가능한 클라이언트 단에서 데이터 이질성을 측

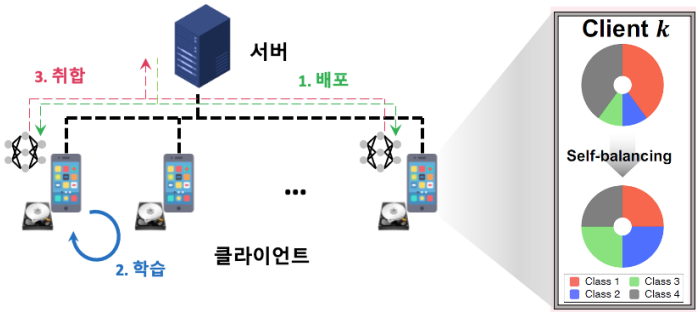

그림 1. FedCSB 알고리즘의 개요도

정하여 활용하는 알고리즘을 제안한다. 특히, 로컬 모델 업데이트 단계에서 학습 데이터의 클래스 분포가 균등해질 수 있도록, 배치를 재구성하는 **클라이언트 자가 균형 방법론**(Federated Client-side Self-Balancing, **FedCSB**)을 제안한다. 전체적인 알고리즘의 개요는 Figure 1에 요약했다.

특정 라운드에서 선택된 클라이언트는 개인 데이터에 대한 불균형도 계산한다. 데이터 불균형도는 클래스의 비율로 설정하였다. 제안된 알고리즘은 학습 데이터 배치를 구성할 때, 측정된 불균형도에 반비례하게 데이터를 샘플링한다. 따라서 소수 클래스에 대해선 오버샘플링을, 다수 클래스는 언더샘플링이 적용되어 클래스 분포를 보다 균등하게 만들어준다. 하지만 이는 적은 수의 클래스 데이터에 대한 과적합을 발생시킬 수 있으므로, 본 논문은 **엔트로피 정규화 손실 함수**(Entropy Regularization, **-ER**)을 통해 문제를 해결했다.

1.2 선행 연구: Federated Averaging

[1]은 개인 데이터 보안을 유지하며 모델을 학습하는 연합학습 프레임워크와 FedAvg 알고리즘을 제안한 논문이다. 하나의

1) 이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임([No.2019-0-00075, 인공지능대학원지원 (한국과학기술원), 10%]과 [No. 2021-0-00907, 능동적 즉시 대응 및 빠른 학습이 가능한 적응형 경량 엣지 연동분석 기술개발, 90%])

글로벌 서버 S와 총 K 명의 클라이언트 C_k 에서 신경망 한 개를 학습하는 문제를 해결하고자 한다. 이 때, 클라이언트 C_k 는 각각 $D_k = \{(x_i, y_i)_{i=1}^{N_k}\}$의 개인 데이터를 가지고 있다. 매 라운드 $r \in \{1,2,\cdots,R\}$에서는 서버가 클라이언트 셋 P를 선택하여, 가장 최신의 글로벌 모델 W_S^r을 배포한다. 이 후, 배포된 모델은 D_k를 학습하여(W_k^r) 다시 서버에게 전송한다. 서버는 데이터 수 N_k에 비례하게 가중 평균된 모델을 생성하고, 이 때 학습에 참여하지 않았던 클라이언트들은 W_S^r로 취급하여 W_S^{r+1}을 얻는다. FedAvg 알고리즘에 대한 슈도 코드는 아래와 같다.

알고리즘 1. FedAvg [1]

W_S 초기화
for $r = 1, \cdots$ **do**
  클라이언트 셋 $\{C_1, \cdots, C_K\}$에서 P 샘플링
  for $C_k \in P$ **do**
    $W_{C_k}^r \leftarrow ClientUpdate(W_S)$
  end for
  $W_S^{r+1} = \sum_{k=1}^{K} \frac{N_k}{N} W_k^r \ (N = \sum_{k=1}^{K} N_k, W_{C_k \notin P}^r = W_S^r)$
end for

$ClientUpdate(W_S)$:
$W_{C_k} = W_S$
for Local epoch **do**
  $W_{C_k} \leftarrow W_{C_k} - \eta \nabla L_{CE}(D_k)$
end for
return W_{C_k}

2. 본 론

2.1 클라이언트 자가 균형 연합학습 (FedCSB)

본 논문에서는 기존 선행 연구와는 달리, 데이터에 접근이 가능한 클라이언트 단에서 데이터 이질성 문제를 해결하고자 한다. 특히, 데이터 불균형 정도에 근거해 현 모델이 균형 잡힌 데이터를 학습할 수 있도록, 클라이언트의 데이터 D_k를 재구성하는 FedCSB 알고리즘을 제안한다. 특히, 배치 데이터를 새로 샘플링할 때는 **각 클래스 비율에 반비례하게 샘플을 추출한다.** 예를 들어, 1번 클래스를 많이 가진 클라이언트는 1번 클래스에 대해서는 적은 수의 추출을, 다른 소수 클래스에 대해서는 반복 추출을 통해 클래스 비율이 비슷한 새로운 배치 D_k^*를 생성한다. 이는 알고리즘 1의 *ClientUpdate* 함수를 변경했고, 해당 슈도 코드는 알고리즘 2에 정리되어 있다.

FedCSB는 크게 두 가지 부분으로 구성되어 있다. 먼저, 클래스 불균형도를 고려해 각 샘플의 추출 확률을 계산한다. 본 논문에서는 가장 단순한 함수로 **클래스 샘플의 비율에 대한 역수**를 사용하였고, 같은 클래스를 부여받은 샘플들은 모두 같은 값으로 계산된다.

$$u_i = U(x_i, y_i) = \left(\frac{\sum_{j=1}^{N} I[y_j = y_i]}{N} \right)^{-1}$$

그 다음, 각 샘플들의 확률에 따라 **기각 샘플링(rejection sampling) 알고리즘을 사용하여 추출한다.** 매우 간단한 알고리즘이기 때문에, 기존의 연합학습 프레임워크 위에서 설계된 방법론들에 쉽게 적용이 가능하다.

알고리즘 2. FedCSB

$ClientUpdate(W_S)$:
$W_{C_k} = W_S$
$u_i = U(x_i, y_i)$ *where* $(x_i, y_i) \in D_k$
for Local epoch **do**
  for Local iteration **do**
    $D_k^* = \varnothing$
    while True **do**
      if $u_i \leqq p \sim U(0,1)$ **then**
        $D_k^* = D_k^* \cup (x_i, y_i)$
      end if
      if $|D_k^*| = BatchSize$ **then**
        Break
      end if
    end while
    $W_{C_k} \leftarrow W_{C_k} - \eta \nabla L_{CE}(D_k^*)$
  end for
end for
return W_{C_k}

2.2 엔트로피 정규화 손실 함수 (FedCSB-ER)

FedCSB는 기존의 알고리즘과 다르게 클라이언트 단에서 동작함으로써, 효과적으로 데이터 불균형 문제를 해결할 수 있었다. 하지만 **소수 클래스에 대한 복원 추출은 반대로 과적합 문제를 야기한다.** 실제로 FedCSB 알고리즘을 통해 모델을 학습했을 때, 소수 클래스에 대한 Softmax Response (로짓에 대한 소프트맥스의 최댓값, SR [4])가 다른 클래스들보다 매우 높게 학습이 되었다. 적은 수의 샘플에 아무리 다른 데이터 증강 방법(e.g., 회전, 자르기)을 적용해도, 깊은 신경망이 샘플들을 기억하는 양상을 보였다.

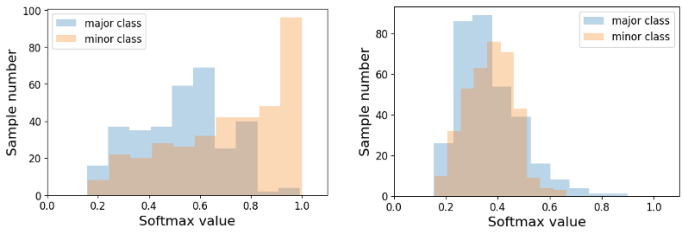

그림 2. FedCSB(왼쪽)와 FedCSB-ER(오른쪽)의 SR 히스토그램

본 논문에서는 과적합 문제 해결을 위해 **로짓의 엔트로피 값을 정규화 손실 함수로 제안한다.** 엔트로피는 균등 벡터에서 최댓값을 갖고, 반대로 원-핫 벡터에서 최솟값을 갖는다. 따라서, 과적합으로 원-핫 벡터에 가까워진 소수 클래스 샘플들에 대해 엔트로피 값을 키워주는 방향으로 같이 학습할 수 있도록 설계했다. 기존의 교차 엔트로피 함수에 정규화 항을 추가한 꼴로, 정확한 수식은 다음과 같다.

$$L = L_{CE} - \lambda H(q) = L_{CE} + \lambda \sum_{c=1}^{C} q_c \log q_c$$

이 때, q는 임의의 샘플에 대한 신경망의 출력값(로짓)의 소프트맥스 함수값이다. λ는 하이퍼파라미터로, $\lambda = 1$을 사용했다. 본 논문에서는 최종적으로 FedCSB와 Entropy Regularization (ER)을 결합한 FedCSB-ER 알고리즘을 제안한다.

2.3 실험 환경

본 논문에서는 VGG-16-BN [6] 신경망 모델에서 기존의 알고리즘들과 불균형 상황에서의 연합학습 성능을 비교했다. 특히, α 파라미터로 조절되는 ***Dirichlet* 분포를 사용하여 불균형된 CIFAR-10 [7] 데이터셋을 인위적으로 생성했다.** 생성된 데이터셋의 불균형도는 α 파라미터의 크기와 서로 반비례한다.(그림 3 참고)

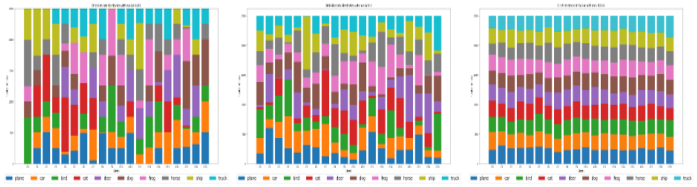

그림 3. $\alpha = 0.001$(왼쪽), $\alpha = 0.1$(중간), $\alpha = 10.0$(오른쪽)에서의 CIFAR-10 데이터셋 분포도

총 100번의 라운드(R)동안 $K = 20$ 클라이언트들은 128개의 샘플로 구성된 배치로 한 번의 에폭만큼 신경망을 학습하였다. 매 라운드마다 학습에 참여하는 클라이언트가 2명과 4명인 상황($|P| = 2$ or 4)에서 실험하였다. 학습을 위한 최적화 방식은 확률적 경사 하강법(SGD)을 사용하였다. 0.1의 학습률과 0.9 크기의 모멘텀 값을 설정했다. 학습률은 각각 50번째와 75번째 라운드에서 0.1배만큼 감소시키며 학습을 진행했다. 실험 결과의 타당성을 위해, 3번의 무작위 실험을 통해 성능을 분석했다.

2.4 비교군

본 논문에서는 총 네가지의 비교군을 설정했다. **(1) 중앙 학습**: 서버가 모든 데이터에 접근 가능한 상황으로, 연합학습이 아닌 일반적인 상황을 의미한다. **(2) FedAvg [1]**: 처음으로 제시된 연합학습 알고리즘이다. **(3) FedProx [8]**: 글로벌 모델과 각 로컬 모델 사이의 거리를 정규화 손실 함수로 제안했다. **(4) FedAvg_pdp [5]**: 각 라운드에서 선택된 클라이언트만을 취합하는 알고리즘이다.

2.5 실험 결과

다양한 불균형 정도와 학습에 참여하는 클라이언트 수 상황에서의 알고리즘 성능은 표 1에 정리되어있다. 먼저, 중앙 학습의 성능이 연합학습 상황에서보다 항상 높은 것을 볼 수 있다. 다음으론 불균형도가 적을수록, 더 많은 수의 클라이언트가 학습에 참여할수록 성능이 일관되게 증가하였다.

비교군 알고리즘 중에서는 데이터 불균형 문제를 해결하기 위해 제안된 최신 알고리즘(FedProx, FedAvg_pdp)이 더 높은 정확도를 보여주었다. 하지만, 본 논문에서 제안한 **FedCSB 알고리즘이 최신 알고리즘 대비 최대 12.9%p를, FedAvg 대비 22.6%p 향상을 보였다.** 이를 통해, 각 클라이언트가 자신이 보유한 데이터의 균형을 맞추는 방식이 실제로 연합학습의 성능 향상을 야기할 수 있음을 확인하였다.

또한, FedCSB와 FedCSB-ER 알고리즘 성능 비교를 통해, 소수 데이터에 대한 과적합 문제를 엔트로피 정규화 손실 함수가 해결해주는 지를 검증하였다. **FedCSB 대비 FedCSB-ER 알고리즘은 5.6%p만큼 성능을 향상시켰고, 모든 실험 환경에서 가장 높은 정확도를 보였다.**

3. 결 론

본 논문에서는 기존 선행 연구와는 달리, 데이터에 접근이 가능한 클라이언트에서 데이터 이질성 문제를 해결하는 알고리즘을 제안했다. **클래스 샘플 비율에 반비례하게 새로운 배치를 생성하는 FedCSB**과 **복원 추출에 의한 과적합 문제를 방지하기 위해 엔트로피 정규화를 추가한 FedCSB-ER 알고리즘**을 제시하였다. 최종적으로, 제안한 알고리즘을 선행 연구와의 성능 비교를 통해 그 효과를 입증하였다. 본 논문처럼 클라이언트만이 데이터에 직접 접근이 가능함을 효율적으로 활용하는 알고리즘에 대한 지속적인 연구가 필요하다는 당위성을 제시하고자 한다.

표 1. 각 알고리즘들의 이미지 분류 정확도

알고리즘	$\|P\| = 2$			$\|P\| = 4$		
	$\alpha = 0.001$	$\alpha = 0.1$	$\alpha = 10.0$	$\alpha = 0.001$	$\alpha = 0.1$	$\alpha = 10.0$
Centralized	**91.1**					
FedAvg	51.3	54.9	57.1	65.7	67.5	69.0
FedProx	58.5	74.1	78.8	72.9	80.0	82.0
FedAvg_pdp	53.2	74.6	78.8	74.0	80.2	82.2
FedCSB	66.1	77.5	78.2	76.9	81.7	81.9
FedCSB-SR	**71.7**	**79.0**	**79.9**	**79.9**	**82.9**	**83.5**

참고 문헌

[1] Brendan McMahan, Eider Moore, Daniel Ramage, SethHampson, and Blaise Agueray Arcas. Communication-efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics, pages 1273-1282.PMLR, 2017.

[2] Yae Jee Cho, Jianyu Wang, and Gauri Joshi. Client selection in federated learning: Convergence analysis and power-of-choice selection strategies. arXiv preprint arXiv:2010.01243,2020.

[3] Yang, M., Wong, A., Zhu, H., Wang, H., & Qian, H. Federated learning with class imbalance reduction. arXiv preprint arXiv:2011.11266, 2020.

[4] Geifman, Y., & El-Yaniv, R. Selective classification for deep neural networks. arXiv preprint arXiv:1705.08500, 2017.

[5] Li, X., Huang, K., Yang, W., Wang, S., & Zhang, Z. (2019). On the convergence of fedavg on non-iid data. International Conference on Learning Representations, ICLR 2020

[6] Simonyan, Karen, and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

[7] Krizhevsky, Alex, and Geoffrey Hinton. Learning multiple layers of features from tiny images. (2009)

[8] Li, Tian, et al. Federated optimization in heterogeneous networks. arXiv preprint arXiv:1812.06127 (2018)