

경사도값의 Count-Sketch 압축방식을 통한 통신 효율적 연합학습

박찬호, 서중하, 강준혁

한국과학기술원

kmapark@kaist.ac.kr, junghaa.seo@kaist.ac.kr, jhkang@kaist.edu

Communication-efficient Federated Learning
via Count-sketch Compression of Gradients

Park Chanhoo, Seo Junghaa, Kang Joonhyuk

Korea Advanced Institute of Science and Technology

요약

본 논문에서는 통신 채널이 적용된 연합학습에서 발생하는 통신지연 문제를 해결하기 위해 선형적 특성을 가지며 아날로그 통신에 적합한 Count-Sketch 압축방식을 소개하고, 이것을 광대역 아날로그 통신에 적용하여 통신 지연이 감소된 연합학습 모델을 제안하였다.

I. 서론

연합학습(Federated learning)[1]은 다수의 로컬 클라이언트가 각각 보유한 개별 데이터로 학습 후 중앙서버를 통해 글로벌 모델을 협력적으로 학습하는 방식으로, 학습에 사용되는 데이터를 상호 간에 공유하지 않는다는 측면에서 정보보호의 장점을 가지며 대용량의 데이터를 직접 전송하기보다는 개별 학습된 모델의 파라미터만을 공유하기에 통신 효율적인 특징을 가진다. 또한 연합학습 방식은 학습에 참여하는 디바이스의 컴퓨팅 능력을 활용하기 때문에 참여하는 클라이언트가 많을수록 계산 능력(속도)이 강화되는 반면, 모델의 파라미터를 공유하기 위해 한정된 통신 자원을 다수의 클라이언트가 사용하기에 통신 부분이 연합학습 속도를 지연시키는 주요 원인이 된다. 이러한 통신지연 문제를 해결하기 위해 개별 학습 횟수를 늘려 최적으로 계산된 파라미터를 전송[1]하거나, 공유되는 파라미터를 수를 압축[2]하고, 무선통신 채널에서 아날로그 신호의 중첩 특성을 통해 주파수 대역폭을 효율적으로 전송하는 방식[3] 등 각 분야에서 다양한 방식으로 심도 있게 연구되고 있다. 본 논문에서는 클라이언트 수와 관계없이 제한된 주파수 대역폭을 효율적으로 재사용하기 위해 광대역 아날로그 방식을 적용하고, [4]에서 제안한 선형적 특성을 가지는 Count-Sketch (CS) 기법으로 전송할 파라미터 수를 압축한 후 채널에서 Air-Computation을 하여 연합학습 간 발생하는 통신 지연 문제를 개선할 수 있는 방법을 제안한다.

II. 본론

2-1 연합학습 기반 무선통신 변조기술 탐지 모델

연합학습 기반 무선통신 변조기술 탐지 수신기는 수식 (1)과 같이 중앙

서버로부터 받은 t 번째 글로벌 모델의 가중치 θ_t 를 기반으로 m 번 수신기에서는 수신 신호 r 과 변조기술 y 를 통해 수신기 모델의 손실함수 $f(\theta_t; \cdot)$ 를 최소화하기 위한, 가중치 경사도값

$$g_m(\theta_t) = \frac{1}{|B_m|} \sum_{r_i \in B_m} \nabla f((\theta_t, r_i), y_i) \quad (1)$$

을 계산하여 중앙서버로 전송한다. 이때, f 는 DNN 기반 수신기 모델의 손실함수, B_m 은 m 번 수신기가 가진 수신 신호 데이터 개수를 의미하며, 중앙서버는 개별 수신기로부터 받은 t 번째 가중치 경사도값들을 연합평균(Federated averaging)을 하여 얻은 글로벌 가중치값

$$\theta_{t+1} = \theta_t - \eta_t \frac{1}{M} \sum_{m=1}^M g_m(\theta_t) \quad (2)$$

을 개별수신기로 보내고 m 번 개별 수신기는 다시 수신기 모델의 손실함수 $f(\theta_{t+1}; \cdot)$ 를 최소화하는 가중치 경사도값 $g_m(\theta_{t+1})$ 을 계산하여 중앙서버로 보내는 식 (1),(2)의 반복적인 과정을 통해 수신기의 변조기술 탐지 성공률을 높이도록 학습한다.

2-2 Count-Sketch (CS) Compression

연합학습간 공유되는 가중치 경사도 값의 개수를 줄이기 위해 사용할 CS 압축은 그림 1과 같이 중앙서버와 개별 수신기 간에 사전 공유된 사인 해시값 h_j^s 과 버킷해시값 h_j^b 을 통해 개별수신기가 가진 d 개의 가중치 경사도 값을 $(r \times c)$ 개로 압축하여 전송할 수 있다. 중앙서버에서는 식 (3)과 같이 가중치 경사도값 g_i 를 i 번 해시값에 의해 매트릭스의 각 행에 저장된 값들의 중앙값을 통해 추정치를 얻을 수 있다.

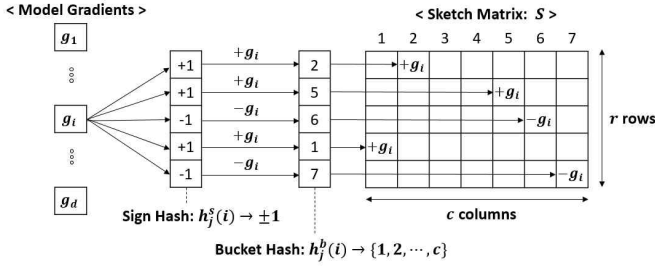


그림 1

$$\text{Estimates}[g_i] = \text{Median}\{h_1^s \cdot S[1, h_1^b], \dots, h_r^s \cdot S[r, h_r^b]\} \quad (3)$$

특히, CS압축은 식 (4)와 같이 선형적 특성을 가지며,

$$\sum_m S(g_m(\theta_t)) = S(\sum_m (g_m(\theta_t))) \quad (4)$$

식 (5)와 같이 중앙서버는 각 개별 수신기로부터 받은 압축된 매트릭스의 합으로부터 가중치 경사도 연합평균값의 추정치 $\hat{g}(\theta_t)$ 를 얻을 수 있다.

$$\hat{g}(\theta_t) = \frac{1}{M} \cdot U(S(\sum_m (g_m(\theta_t)))) \quad (5)$$

이때, $S(\cdot)$ 는 압축 함수, $U(\cdot)$ 는 복원함수를 의미한다.

2-3 광대역 아날로그 전송을 통한 주파수 대역폭 재사용

디지털 통신방식과 비교하여 광대역 아날로그 통신의 장점은 그림 2와 같이 주파수 대역폭을 효율적으로 사용하여 통신지연을 감소시킬 수 있다는 것이다. 아날로그 전송은 신호의 중첩(Superposition) 성질을 활용하여 주어진 대역폭을 온전하게 재사용 가능한 반면에 디지털 전송방식은 클라이언트 간 상호 간섭을 회피하기 위해 대역폭을 개별 할당하게 된다. 따라서 학습에 참여하는 클라이언트 수가 늘어날수록 디지털 방식은 개인에게 할당되는 주파수 대역폭이 줄어들기 때문에 아날로그와 디지털 방식의 통신 지연 격차는 증가하게 된다.

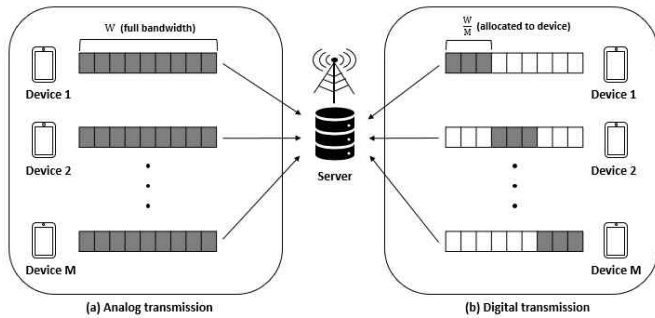


그림 2

한편, 중앙서버에서는 개별 클라이언트의 파라미터들이 합해진 값을 채널을 통해 얻을 수 있는데, 이러한 특성은 앞서 소개한 선형성을 가지는 CS 압축방식을 적용하기에 적합하다.

2-4 아날로그 통신과 CS압축 방식을 적용한 연합학습

각 클라이언트는 그림 3과 같이 학습된 가중치 값을 CS기법으로 압축하고 전력 $\alpha_m(t)$ 을 할당하여 중앙서버로 전송한다.

$$\alpha_m(t) = \frac{\gamma_m(t)}{h_m(t)} \quad (6)$$

이때, 가중치 값들은 무선 채널을 통과하면서 합해지고 중앙서버에서 받게 될 수신 신호는 식 (7)과 같다.

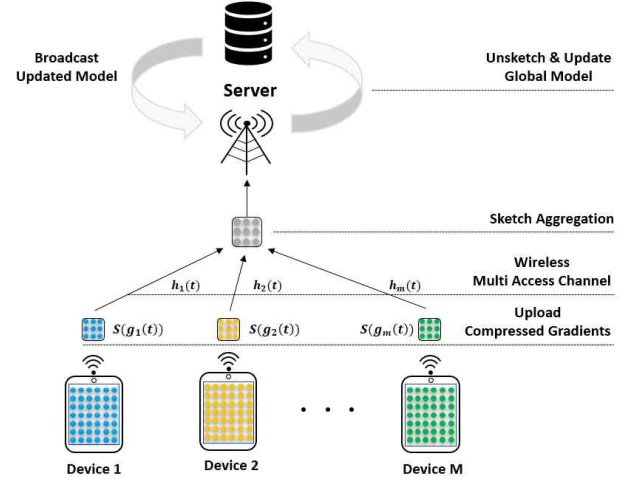


그림 3

$$y(t) = \sum_m \{\gamma_m(t) \cdot S(g_m(\theta_t))\} + z(t) \quad (7)$$

중앙서버에서는 식 (8)과 같이 노이즈를 포함한 압축된 경사도 값의 추정치를 복원할 수 있다.

$$\hat{S}(g(\theta_t)) = \frac{y(t)}{\bar{\gamma}} \quad (8)$$

$$s.t. \bar{\gamma} = \frac{1}{M} \sum_m \gamma_m$$

마지막으로 CS 복원함수를 이용하여 경사도값의 연합평균 추정값을 복원하고 추정값을 통하여 글로벌 모델을 업데이트 할 수 있다.

$$\hat{g}(\theta_t) = U\{\hat{S}(g(\theta_t))\} \quad (9)$$

III. 결론

본 논문에서는 통신 채널이 적용된 연합학습에서 발생하는 통신지연 문제를 해결하기 위해 선형적 특성을 가지는 아날로그 통신에 적합한 Count-Sketch 압축방식을 소개하고, 이것을 광대역 아날로그 통신에 적용하여 통신 지연이 감소된 연합학습 모델을 제안하였으며 현재, 추가적인 성능 향상을 위해 CS 매트릭스의 크기와 Top-K 값 등 하이퍼 파라미터를 조정하여 최적화하기 위한 연구를 지속 중이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음. (IITP-2020-2020-0-01787)

참고 문헌

- [1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Artificial Intelligence and Statistics. PMLR, 2017, pp. 1273–1282.
- [2] Lin, Yujun, et al. "Deep gradient compression: Reducing the communication bandwidth for distributed training." arXiv preprint arXiv:1712.01887 (2017).
- [3] Zhu, Guangxu, Yong Wang, and Kaibin Huang. "Broadband analog aggregation for low-latency federated edge learning." IEEE Transactions on Wireless Communications 19.1 (2019): 491–506.
- [4] Rothchild, Daniel, et al. "Fetchsgd: Communication-efficient federated learning with sketching." International Conference on Machine Learning. PMLR, 2020.