

초소형, 저전력 네트워크 기반 Federated Learning의 효율적인 운용을 위한 Model Parameter 전송 방법

최진혁, 이창민, 김성륜*

연세대학교

jh.choi@ramo.yonsei.ac.kr, cmlee@ramo.yonsei.ac.kr, *slkim@ramo.yonsei.ac.kr

Communication - Efficient Federated Learning Model Update Method

Choi Jin Hyuk, Changmin Lee, Seong-Lyun Kim*

Yonsei Univ.

요약

본 논문은 초소형, 저전력 단말에서 인공지능 적용을 위한 무선 통신 기반의 Machine Learning 기법인 연합 학습(Federated Learning)의 적용성을 확인하고 Physical-layer Network Coding (PNC) 기반의 다중 ACK 동시 회신 기법을 활용하여 저전력 네트워크의 효율적 운용 방법에 대해 제안한다. 추가로 ML 학습 측면과 통신 비용 측면을 고려한 Server 모델의 재전송 Threshold의 최적값을 도출하여 Communication-Efficient FL 방식을 제안한다.

I. 서론

오늘날 인공지능 분야에서 효율적인 학습 방식은 다양한 형태로 연구되고 있다. 기존의 인공지능을 활용하기 어려운 초소형, 저전력 하드웨어에서 인공지능을 활용하기 위하여 통신을 기반으로 학습하는 모델이 소개되고 있다. 대표적인 예로는 구글에서 제안한 Federated Learning(FL)[1]이 있다. FL은 계산 유닛을 가진 다수의 단말(device)이 무선 통신을 이용하여 하나의 Global Model Update에 참여하고, Update된 Model을 전송받는 반복적인 학습 과정을 통해 Model을 학습하는 방식이다. FL 학습 방식은 학습을 위해 사용되는 데이터의 Privacy를 보존하면서 동시에 저사양 단말들의 학습 성능을 올려주는 장점이 있다. 그러나 FL 방식은 단말과 서버의 잦은 통신으로 인해 무선 통신 채널에 부하가 크기 때문에, FL의 통신 방식에 적용될 수 있는 시스템 모델 관련 연구와[3] FL 적합 통신 프로토콜 연구가 진행되고 있다[4]. 본 논문은 초소형, 저전력 하드웨어에서의 인공지능 활용성을 확대하고 다중 ACK 동시 회신을 통해 네트워크의 효율성을 증대하기 위한 연구이다.

II. 본론

본문에서는 첫째로 FL과 PNC를 응용한 ACK 수신 기법에 대해 간략히 설명한다. 둘째로 PNC 응용 ACK 수신 기법이 FL에서 동작하는 과정에 대해 설명한다. 마지막으로 PNC 응용 ACK 수신 기법이 적용된 FL의 상황에서 optimal한 parameter 범위에 대해 분석한다.

- Federated Learning(FL)

FL 방식은 반복해서 랜덤한 수의 단말들이 Global Model Update에 참여하고, 참여한 단말들은 서버로부터 이전 시기의 단말들이 Update한 weight의 평균을 전달받는다. 각 단말은 서버로부터 전송받은 weight를 자신의 모델에 초기화한 후, 가지고 있는 데이터를 통해 모델을 학습하는 과정을 반복한다. FL 구성도는 그림 1과 같다.

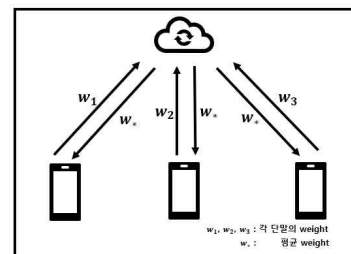


그림 1. Federated Learning 예시

- Physical-layer Network Coding 응용 ACK

Physical Network Coding[5]은 물리 계층에서 서로 중복된 신호를 수신하여 Mapping 작업과 각 수신 노드에 저장된 정보를 통해 분리해내는 기법이다. 이러한 네트워크 코딩 방식은 통신 과정에 필요한 time slot을 효과적으로 감소시키는 이점을 가지고 있지만, 하나의 신호를 알고 있는 네트워크 구조에만 적용될 수 있다는 한계점을 가지고 있다. ACK 신호의 경우 미리 정보를 알고 있을 수 있다는 가정하에 이러한 PNC의 특징을 통해 분리해낼 수 있다. 또한, ACK 수신을 중첩하여 받아 ACK 신호의 개수를 분리해 낼 수 있는 PNCed ACK 기법을 제안한다.

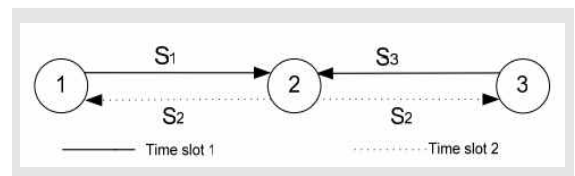


그림 2. Two way relay channel (TWRC)에서 Physical-Layer Network Coding 적용 예시

- FL & PNCed ACK

Federated Learning의 서버가 단말에 weight값을 전달하는 단계는 전송 방식에 따라 통신 자원의 효율성이 달라질 수 있다. 또한, 단말이 수신하는 weight 정보는 신뢰성을 요구하므로, ACK 수신이 요구된다. 이러한 조건에 따라 Unicast 방식을 사용할 경우, FL의 weight update 1회에 들

알고리즘 1: Federated Learning(FL) with PNCed ACK

```

1: Server executes:
2:  $\zeta \leftarrow$  재전송 기준 설정
3: for 각 round  $t = 1, 2, \dots$  do
4:    $p \leftarrow \max(C \cdot |\text{전체 단말}|, 1) \triangleright C$  학습 참여 비율
5:    $S \leftarrow \{\text{전체 } p \text{ 단말}\}$ 
6:    $S_T \leftarrow \{\text{통신 성공 } p \text{ 단말}\}$ 
7:    $S_F \leftarrow \{\text{통신 실패 } p \text{ 단말}\}$ 
8:    $\gamma = \frac{S_T}{S_T + S_F}$ 
9:   if  $\gamma > \zeta$ :
10:    for 각 device  $k \in S$  in parallel do
11:       $w_{t+1}^k \rightarrow \text{FL}(w_{\text{global}}, k_t^{S_T})$ 
12:       $w_{t+1}^k \rightarrow \text{FL}(w_{\text{device}}, k_t^{S_F})$ 
13:       $w_{t+1} \leftarrow \sum_{k=1}^S \frac{1}{p} w_{t+1}^k$ 
14:    else:
15:       $w_{\text{global}}$  재전송
16:
17: FL( $w, k$ ):
18:   weight 값  $w$  로 초기화
19:   for device local data do
20:      $w \leftarrow w - \nabla \text{Loss}(w)$ 
21:   return  $w$  to server

```

여가는 Delay가 너무 길어지게 된다. 이를 해결하기 위해, PNC를 활용하여 다수의 ACK를 동시에 데이터를 전송하고 충돌하여 수신된 ACK 신호를 분리하는 FL 시스템 모델을 가정한다. FL 시스템 모델에서 서버는 PNCed ACK을 통해 중첩된 ACK의 개수를 확인하여 참여 단말에 weight가 올바르게 전달되었는지를 인지한다. 기존의 방식의 경우 weight값을 수신하지 못한 단말이 존재할 경우, 재전송으로 인한 통신 자원이 낭비된다. 이러한 통신 자원의 효율을 위해 FL의 학습에 따른 모델의 정확도와 weight 초기화의 관계에 대한 분석을 통해 모델의 학습에 영향을 미치는 단말 그룹의 weight 초기화 비율에 대해 분석하고자 한다. 추가로 PNC 기반 ACK 수신 FL에서 서버가 허용하는 ACK 수신 비율의 범위에 대해 분석하여, Communication-Efficient한 FL parameter를 찾는다. 위의 설명된 방식은 아래 알고리즘1과 같다.

- Weight 초기화 비율과 FL 학습 정확도 관계 실험

PNCed ACK의 재전송과 관련한 단점은 FL에서 장점으로 사용될 수 있다. Update에 참여한 단말 중 서버로부터 weight값을 수신하지 못하고 이전 weight로 새로 학습하는 비율이 p 라고 했을 때, 전체 모델의 수렴에 영향을 미치지 않는 비율 p 에 대해 분석하고자 한다. 비율 p 는 서버가 용인할 수 있는 ACK 수신 수 / Update 참여 단말 수라고 볼 수 있다. 비율 p 를 찾기 위해서 FL Simulation을 다음과 같은 조건에서 진행하였다. 기본적인 FL 알고리즘은 python으로 구성되었고, 전체 단말의 개수는 100대이고, 매번 학습에 참여하는 단말의 비율은 0.3으로 설정하여 총 30대가 참여할 수 있도록 구성하였다. 사용되는 CNN 학습 모델은 Pytorch에서 제공하는 image classification tutorial 모델을 사용하였다. 데이터는 CIFAR10을 사용하였고 기본적인 cropping을 추가하였다. Non-IID 상황을 가정하였기에 CIFAR10 데이터를 각 단말이 불균형한 상태로 데이터를 갖도록 하였다. 모델의 수렴을 판단하는 정확도 기준은 현재 최신 기술의 경우 96% 이상을 가지고 있지만, 이 실험의 의미는 모델 수렴을 위한 global weight update 필수 비율과 통신 횟수의 관계를 보여주기 위함이기 때문에 50%로 설정하였다.

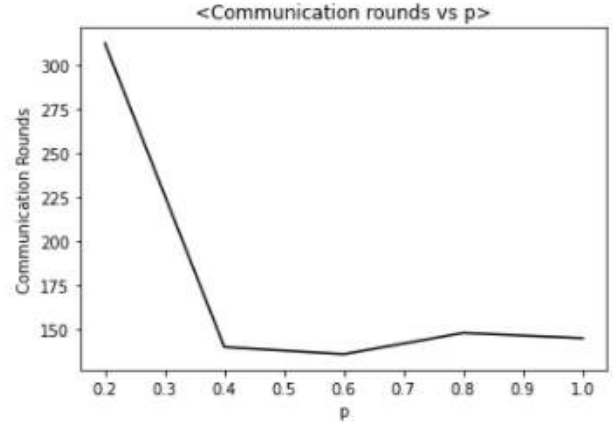


그림 3. p 에 따른 Communication Round

III. 결론

본 논문에서는 통신 기반의 Machine Learning 알고리즘인 Federated Learning에서 발생할 수 있는 통신 지연 상황에 대해서 분석하였다. 먼저 PNCed ACK이라는 PNC기반 ACK 분석 기법의 특성을 활용하여, 단말의 수신 비율에 대해 계산한다. 이후 FL을 통한 training에 영향을 미치지 않는 수신 비율을 실험적으로 분석하여 FL에서 발생하는 통신 지연을 최소화하는 방안에 대해 제시하였다. 추후 다양한 기법의 통신 기반 ML 학습 알고리즘을 보다 효율적으로 운용하기 위한 다양한 통신 프로토콜의 개발과 그에 따른 응용 학습 알고리즘 연구를 진행할 예정이다.

ACKNOWLEDGMENT

본 연구는 국방생체모방 자율로봇 특화연구센터를 통한 방위사업청과 국방과학연구소 연구비 지원으로 수행되었습니다 (UD190018ID).

참 고 문 헌

- [1] Jakub Konecny, H. Brendan McMahan, Felix X. Yu Peter Richtarik, Ananda Theertha Suresh, and Dave Bacon. "Federated learning: Strategies for improving communication efficiency." In NIPS Workshop on Private Multi-Party Machine Learning, 2016.
- [2] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. "Communication-efficient learning of deep networks from decentralized data." In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017.
- [3] K. Bonawitz et al., "Towards federated learning at scale: System design," arXiv preprint arXiv:1902.01046, 2019.
- [4] Mohammed Aledhari, Rehman Razzak, Reza M Parizi, and Fahad Saeed. 2020. Federated learning: A survey on enabling technologies, protocols, and applications. IEEE Access 8 (2020), 140699 - 140725.
- [5] S. Zhang, S. C. Liew, P. P. Lam, "Hot Topic: Physical-layer Network Coding," ACM MobiCom '06, pp. 358-365, Sept. 2006.