https://doi.org/10.7840/kics.2023.48.8.934

연합학습의 모델 업로드 및 종합을 위한 상향링크 다중 입출력 통신 최적화

유 원 식•, 박 석 환°

Optimizing Uplink MIMO Transmission for Model Uploadand Aggregation in Federated Learning

Wonsik Yoo•, Seok-Hwan Park°

요 약

본 논문은 연합학습 시스템에서 효율적인 모델 업로드를 위한 다중안테나 통신 기술의 최적화를 다룬다. 기존 연구는 연합학습의 로컬 모델 업로드 통신 기술 설계 시 채널 상태 정보가 완벽하다고 가정하였다. 그러나 본 논문은 채널 정보의 부정확성을 고려하여 상향링크 데이터 전송율을 계산하고, 이를 바탕으로 글로벌 종합 모델의 평균 자승 오차를 최소화하는 문제를 정립한다. 이 최적화 문제의 비 컨백스 성질을 극복하고 부 최적해를 효율적으로 얻기 위해 Majorization Minimization 기반의 알고리즘을 제안한다. 또한 모의실험을 통해 제안하는 최적화 알고리즘의 우수성을 검증한다.

키워드 : 연합학습, 상향링크, 강인한 전송, 부정확한 채널 정보, 최적화
Key Words : Federated learning, uplink, robust transmission, imperfect CSI, optimization

ABSTRACT

This paper focuses on optimizing multi-antenna transmission to enhance the efficiency of model upload in digital federated learning systems. Unlike previous research that assumes perfect channel state information (CSI) for local model upload in federated learning, this work takes into consideration the imperfection of CSI and computes the achievable data rates accordingly. The problem of minimizing the mean squared error (MSE) of the global aggregated model is formulated, which is found to be non-convex. To address this non-convexity and obtain an efficient suboptimal solution, we propose an iterative algorithm based on Majorization Minimization. The advantages of the proposed algorithm are validated through numerical results.

Ⅰ. 서 론

연합학습(Federated learning)은 로컬 수집 학습 데이터를 갖춘 모바일 사물인터넷 기기(IDInternet-of-things devices)들이 서버와 메시지를 교환하면서 공통의 기계학습 모델을 학습하는 기술이다. 이 기술은 수집한 로컬 학습 데이터의 보안을 유지하고 무선 자원의 부담을 경감하는 장점으로 큰 관심을 받고

※ This work was supported by the National Research Foundation (NRF) of Korea funded by the MOE under Grant 2019R1A6A1A09031717 and by the MSIT under Grants 2021R1C1C1006557.
◆ First Author : Division of Electronic Engineering, Jeonbuk National University, wonsik0713@jbnu.ac.kr, 학생회원
° Corresponding Author : Division of Electronic Engineering, Jeonbuk National University, seokhwan@jbnu.ac.kr, 정회원
논문번호 : 202304-077-A-RN, Received April 14, 2023; Revised May 25, 2023; Accepted May 29, 2023

있다[1,2]. 연합학습 과정은 여러 라운드로 구성되며, 각 라운드마다 분산된 ID들이 로컬 데이터를 기반으로 학습한 로컬 모델을 서버 노드로 업로드한 후, 서버가 이를 종합하여 글로벌 모델로 갱신하여 ID들에게 알려준다. 따라서 연합학습을 통해 높은 정확도의 글로벌 기계학습 모델을 얻기 위해서는 한정된 무선 자원을 효율적으로 활용하여 로컬 모델을 정확하게 서버 노드에 보고해야 한다.

연합학습 시스템의 로컬 모델 업로드를 위한 상향링크 통신 기술의 최적화에 관한 연구가 수행된 바 있다[3-6]. 기존 문헌들은 채널에서 발생하는 더해지는 잡음 신호가 연합학습의 수렴성에 미치는 영향을 분석하는 과정에서 무선 페이딩 및 간섭 현상을 고려하지 않거나[5], 무선 페이딩을 고려하더라도 채널 정보가 완벽하다는 전제 하에 연합학습 결과 모델의 정확도, 특정 정확도를 획득하는데 소요되는 학습 수행시간 및 에너지 소모 등의 성능 척도를 최적화하는 데 집중하였다[3,4,6].

본 논문은 채널 추정 과정에서 발생하는 채널 추정 오차를 고려하여 데이터 전송율 표현식을 유도하고, 이를 기반으로 서버 노드에서 획득한 글로벌 모델의 평균 자승 오차 (MSE: mean squared error)를 최소화하기 위한 상향링크 전송 기술을 연구한다. 연합학습을 위한 상향링크 통신 방법은 크게 Over-the-Air Computation (AirComp)[6] 등의 아날로그 전송 방식과 로컬 모델의 양자화 및 압축 버전을 전송하는 디지털 전송 방식[4,6]으로 구분되는데, 본 논문은 채널 부정확성에 대한 강인함(robustness)을 얻기 위해 디지털 전송 방식을 고려한다. 수학적으로 정립된 최적화 문제가 비 컨백스 문제임을 고려하여, 효율적인 부 최적해를 얻기 위해 Majorization Minimization (MM)[8] 기반의 반복적인 최적화 알고리즘을 제안한다. 모의실험 결과를 통해 상향링크 전송을 최적화하지 않거나 채널 부정확성을 고려하지 않은 비교 기법들과 비교하여 상당한 성능 향상을 확인한다.

Ⅱ. 시스템 모델

단일 서버가 각각 학습 데이터 집합을 보유한 N_I개의 IoT device (ID)들과 무선 통신을 수행하며 공통의 기계학습 모델을 학습하는 연합학습 시스템을 고려한다. 각 ID k가 보유한 데이터 집합의 샘플 개수를 D_k로 표기하고, 서버 노드와 ID k가 사용하는 안테나 개수를 각각 n_S와 n_I로 표기한다. 연합학습 과정은 여러 회의 라운드로 구성되며, 각 t번째 라운드는 서버가 글로벌 모델을 ID들에게 알려주는 멀티캐스팅으로 시작한다. 각 ID k는 수신한 글로벌 모델을 시작점으로 잡고 자신의 학습 데이터를 이용해 로컬 모델 업데이트를 수행하며, ID k에서 업데이트된 로컬 모델 벡터를 $\boldsymbol{s}_k(t) \in R^{d_R \times 1}$로 표기한다. 여기서 d_R는 모델을 표현하는 실수 파라미터의 개수이다 (예: 심층 신경망에서 가중치와 바이어스 개수). ID들은 자신의 업데이트 모델 벡터들을 상향링크 채널을 통해 서버로 전송하며, 서버는 수신신호 벡터를 이용하여 로컬 모델 벡터들 $\boldsymbol{s}_1(t), \boldsymbol{s}_2(t), \ldots, \boldsymbol{s}_{N_I}(t)$을 회복한 뒤, 다음의 가중치 합을 통해 글로벌 모델을 업데이트한다.

$$\boldsymbol{s}_G(t+1) = \sum_{k=1}^{N_I} (D_k/D_T)\boldsymbol{s}_k(t). \quad (1)$$

(1)에서 전체 학습데이터 샘플개수 $D_T = \sum_{k=1}^{N_I} D_k$을 정의하였다. 위에서 업데이트된 글로벌 모델은 $t+1$번째 라운드 시작 시 서버가 ID들에게 내려주는 글로벌 모델 벡터가 된다. 추후 표기의 편의를 위해 가중치 $\alpha_k = D_k/D_T$을 정의한다.

각 t번째 라운드에서 ID들이 로컬 모델 벡터를 업로드하는 통신 구간 길이를 L 심볼 주기(즉, 채널 사용 횟수)로 가정하면, 각 심볼 주기 내에서 ID k가 전송하는 파라미터 벡터를 $\boldsymbol{\theta}_k \in C^{n_\theta \times 1}$로 표기한다. 표기의 편의를 위해 라운드 및 심볼 인덱스를 삭제했고, n_θ는 각 심볼 구간 동안 보내야 하는 복소수 원소 개수를 의미하고 $n_\theta = (d_R/2)/L$로 계산된다. 파라미터 벡터 $\boldsymbol{\theta}_k$는 평균벡터 $E[\boldsymbol{\theta}_k] = \boldsymbol{0}$, 공분산 행렬 $E[\boldsymbol{\theta}_k \boldsymbol{\theta}_k^H] = \sigma_\theta^2 \boldsymbol{I}_{n_\theta}$을 갖는다고 가정한다.

ID들과 서버 간의 상향링크 통신 채널은 다음의 서버 수신 신호 벡터 $\boldsymbol{y} \in C^{n_S \times 1}$로 모델링한다.

$$\boldsymbol{y} = \sum_{k=1}^{N_I} \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{z}. \quad (2)$$

(2)에서 $\boldsymbol{H}_k \in C^{n_S \times n_I}$는 ID k와 서버 노드 간 채널 행렬, $\boldsymbol{x}_k \in C^{n_I \times 1}$는 ID k의 송신신호 벡터, $\boldsymbol{z} \sim CN(\boldsymbol{0}, \sigma_z^2 \boldsymbol{I})$는 잡음신호 벡터를 의미한다. 각 송신신호 벡터 $\boldsymbol{x}_k$는 전력 제한조건 $E[\| \boldsymbol{x}_k \|^2] \le P$을 만족한다.

본 논문은 채널 정보의 부정확성이 연합학습의 통신

성능에 미치는 영향을 관찰하기 위해, 서버 노드와 ID들이 습득한 추정 채널 행렬 $\hat{\boldsymbol{H}}_k$와 실제 채널 $\boldsymbol{H}_k$ 간 관계를 다음과 같이 모델링한다.

$$\boldsymbol{H}_k = \hat{\boldsymbol{H}}_k + \boldsymbol{E}_k. \tag{3}$$

$\boldsymbol{E}_k$는 채널 추정 오류를 의미하고 모든 원소들은 독립이며, 평균 0, 분산 $\sigma_{e,k}^2$을 갖는다고 가정한다.

Ⅲ. 모델 양자화, 상향링크 전송 및 모델 종합

본 논문은 각 ID k가 파라미터 벡터 $\boldsymbol{\theta}_k$를 디지털 정보로 양자화/압축 후 채널 코딩을 수행하여 상향링크 전송하는 디지털 연합학습을 가정한다[3,4,6]. 이에 ID k는 벡터 $\boldsymbol{\theta}_k$에 가장 가까운 양자화 부호어(codeword) 벡터를 선택하여 bit 열로 변환 후 채널 코딩을 수행한다. 선택된 양자화 벡터를 $\hat{\boldsymbol{\theta}}_k = \boldsymbol{\theta}_k + \boldsymbol{q}_k$로 표기 및 모델링하고, 채널 코딩을 통해 얻는 송신 신호 벡터를 $\boldsymbol{x}_k \sim CN(\boldsymbol{0}, \boldsymbol{V}_k)$로 표기한다. 공분산 행렬 $\boldsymbol{V}_k$는 전력제한 조건 $tr(\boldsymbol{V}_k) \leq P$을 만족한다. 충분한 길이의 소스 부호화 및 가우시안 소스 코드북[3,6]을 가정하면, 양자화 손실벡터 $\boldsymbol{q}_k$는 $\boldsymbol{\theta}_k$와 연관이 없고 $\boldsymbol{q}_k \sim CN(\boldsymbol{0}, \boldsymbol{Q}_k)$의 분포를 따른다.

서버노드는 (2)의 수신신호 벡터 $\boldsymbol{y}$로부터 ID들의 송신신호들을 $\boldsymbol{x}_{\pi(1)} \rightarrow \boldsymbol{x}_{\pi(2)} \rightarrow \ldots \rightarrow \boldsymbol{x}_{\pi(N_I)}$의 순서로 순차적 간섭제거 (SIC: successive interference cancellation) 기법을 통해 복호화한다고 가정한다. 채널 오류 항에 대한 데이터 전송율 평균의 closed-form 계산이 어려우므로, log-det 함수의 concave 성질과 Jensen 부등식을 이용하여 획득 가능한 데이터 전송율 $\boldsymbol{R}_{\pi(k)}$을 다음과 같이 유도한다.

$$\begin{aligned} R_{\pi(k)} = f_{\pi(k)}(\boldsymbol{V}) = &-\log_2\det(\boldsymbol{\Omega}_k) \\ &+ \log_2\det\left(\boldsymbol{\Omega}_k + \hat{\boldsymbol{H}}_{\pi(k)} \boldsymbol{V}_{\pi(k)} \hat{\boldsymbol{H}}_{\pi(k)}^H\right). \end{aligned} \tag{4}$$

수식 (4)에서 표기 $\boldsymbol{V} = \{\boldsymbol{V}_k\}_{k=1}^{N_I}$을 정의했고, $\boldsymbol{\Omega}_k$는 채널 부정확성 및 SIC 복호화 순서를 고려하여 제거되지 않은 모든 간섭 신호 및 잡음 신호의 공분산 행렬을 정의하며, 다음과 같이 계산된다.

$$\begin{aligned} \boldsymbol{\Omega}_k = &\sum_{m=k+1}^{N_I} \hat{\boldsymbol{H}}_{\pi(m)} \boldsymbol{V}_{\pi(m)} \hat{\boldsymbol{H}}_{\pi(m)}^H \\ &+ \left(\sum_{m=1}^{N_I} \sigma_{e,m}^2 tr(\boldsymbol{V}_m) + \sigma_z^2\right) \boldsymbol{I}_{n_S}. \end{aligned} \tag{5}$$

ID k의 데이터 전송율 $\boldsymbol{R}_k$은 양자화된 파라미터 벡터 $\hat{\boldsymbol{\theta}}_k$을 업로드하는데 사용되므로, 부호율-왜곡 이론[9]에 따르면 양자화 잡음 공분산 행렬 $\boldsymbol{Q}_k$는 다음 조건을 만족하도록 설계된다.

$$\begin{aligned} I(\boldsymbol{\theta}_k; \hat{\boldsymbol{\theta}}_k) = g_k(\boldsymbol{Q}_k) = &\log_2\det\left(\sigma_\theta^2 \boldsymbol{I}_{n_\theta} + \boldsymbol{Q}_k\right) \\ &- \log_2\det(\boldsymbol{Q}_k) \leq R_k. \end{aligned} \tag{6}$$

(6)에서 파라미터 벡터 $\boldsymbol{\theta}_k$는 기 정의한 공분산 행렬 $\sigma_\theta^2 \boldsymbol{I}_{n_\theta}$의 복소 가우시안 분포를 따른다고 가정하였다.

서버 노드는 ID 송신신호 벡터들을 복호화함으로써 얻은 파라미터 벡터들 $\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \ldots, \hat{\boldsymbol{\theta}}_{N_I}$을 이용하여 글로벌 모델 업데이트 벡터 $\boldsymbol{\theta}_G = \sum_{k=1}^{N_I} \alpha_k \boldsymbol{\theta}_k$에 근접한 벡터를 얻고자 다음의 선형 처리를 수행한다.

$$\hat{\boldsymbol{\theta}}_G = \sum_{k=1}^{N_I} u_k \hat{\boldsymbol{\theta}}_k. \tag{7}$$

상기 모델 종합 출력의 정확성을 정량화하기 위해 다음 MSE 손실 함수를 고려한다.

$$\begin{aligned} &E\left[\| \hat{\boldsymbol{\theta}}_G - \boldsymbol{\theta}_G \|^2\right] = f_{MSE}(\boldsymbol{Q}, \boldsymbol{u}) \\ &= n_\theta \sigma_\theta^2 \sum_{k=1}^{N_I} |u_k - \alpha_k|^2 + \sum_{k=1}^{N_I} |u_k|^2 tr(\boldsymbol{Q}_k). \end{aligned} \tag{8}$$

(8)에서 표기 $\boldsymbol{Q} = \{\boldsymbol{Q}_k\}_{k=1}^{N_I}$와 $\boldsymbol{u} = \{u_k\}_{k=1}^{N_I}$을 정의했다.

Ⅳ. 문제 정의 및 최적화

본 절은 연합학습의 정확도에 직접적인 영향을 미치는 (8)의 모델 종합 MSE 손실 함수를 최소화하기 위해 ID들의 전송 공분산 행렬 $\boldsymbol{V}$, 양자화 잡음 공분산 행렬 $\boldsymbol{Q}$, 서버 노드의 선형 처리 계수 $\boldsymbol{u}$의 합동 최적화를 목표로 한다. 이 문제를 다음과 같이 수학적으로 정의할 수 있다.

$$\begin{aligned}\min._{\boldsymbol{V},\boldsymbol{Q},\boldsymbol{u},\boldsymbol{R}}\ & f_{MSE}(\boldsymbol{Q},\boldsymbol{u})\\ s.t.\quad & R_{\pi(k)} \le f_{\pi(k)}(\boldsymbol{V}),\ \forall k,\\ & g_k(\boldsymbol{Q}_k) \le R_k,\ \forall k,\\ & tr(\boldsymbol{V}_k) \le P,\ \forall k.\end{aligned} \tag{9}$$

최적화 문제 (9)는 비 컨백스 문제이므로, 변수 집합을 $\{\boldsymbol{V},\boldsymbol{Q},\boldsymbol{R}\}$와 $\boldsymbol{u}$로 분할하여 반복적인 업데이트를 통해 부 최적해를 획득하는 알고리즘을 제안한다. $\{\boldsymbol{V},\boldsymbol{Q},\boldsymbol{R}\}$이 주어졌을 때 최적의 $\boldsymbol{u}$는 $\partial f_{MSE}(\boldsymbol{Q},\boldsymbol{u})/\partial \boldsymbol{u}^* = \boldsymbol{0}$을 만족하는 다음 값들로 주어진다.

$$u_k = \frac{\alpha_k}{1 + tr(\boldsymbol{Q}_k)/\left(n_\theta \sigma_\theta^2\right)}. \tag{10}$$

(9)에서 $\boldsymbol{u}$이 고정 상수로 취급될 경우, 모든 목적함수, 제한조건 함수들이 컨백스 함수들의 차로 표현되는 difference-of-convex (DC) 문제로 분류된다. 따라서 비 컨백스 성질을 유발하는 항을 적절한 기준점과 선형화하여 얻은 컨백스 문제의 해를 반복적으로 찾아가는 MM 기법[8]을 적용하여 부 최적해를 획득할 수 있고, 이는 비 컨백스 최적화 문제를 다루는 데 활발히 적용되고 있는 SCA 기법의 한 예로도 볼 수 있다[10]. 따라서 글로벌 수렴 조건을 만족할 때까지 MM 기법과 수식 (10)을 이용하여 $\{\boldsymbol{V},\boldsymbol{Q},\boldsymbol{R}\}$와 $\boldsymbol{u}$을 반복 업데이트하는 알고리즘을 적용할 수 있으며, 해당 알고리즘을 표 'Algorithm 1'에 기술한다. 표에서 문제 (11)은 다음과 같다.

$$\begin{aligned}\min._{\boldsymbol{V},\boldsymbol{Q},\boldsymbol{R}}\ & f_{MSE}(\boldsymbol{Q},\boldsymbol{u}')\\ s.t.\quad & R_{\pi(k)} \le \tilde{f}_{\pi(k)}(\boldsymbol{V},\boldsymbol{V}'),\ \forall k,\\ & \tilde{g}_k(\boldsymbol{Q}_k,\boldsymbol{Q}_k') \le R_k,\ \forall k.\end{aligned} \tag{11}$$

함수 $\tilde{f}_{\pi(k)}(\boldsymbol{V},\boldsymbol{V}')$와 $\tilde{g}_k(\boldsymbol{Q}_k,\boldsymbol{Q}_k')$는 각각 $f_{\pi(k)}(\boldsymbol{V})$와 $g_k(\boldsymbol{Q}_k)$의 비 컨백스 항을 기준점 $\boldsymbol{V}'$와 $\boldsymbol{Q}_k'$와 함께 1차 Taylor 급수로 근사화하여 얻은 함수를 의미하며, 다음과 같이 정의된다.

$$\begin{aligned}\tilde{f}_{\pi(k)}(\boldsymbol{V},\boldsymbol{V}') =& -\log_2\det(\boldsymbol{\Omega}_k')\\ & - \frac{1}{\ln 2} tr\left((\boldsymbol{\Omega}_k')^{-1}(\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_k')\right)\\ & + \log_2\det\left(\boldsymbol{\Omega}_k + \hat{\boldsymbol{H}}_{\pi(k)}\boldsymbol{V}_{\pi(k)}\hat{\boldsymbol{H}}_{\pi(k)}^H\right),\end{aligned} \tag{12}$$

Algorithm 1. 제안하는 최적화 알고리즘

Initialize $\{\boldsymbol{V}',\boldsymbol{Q}',\boldsymbol{R}',\boldsymbol{u}'\}$.
repeat
  Set $\{\boldsymbol{V},\boldsymbol{Q},\boldsymbol{R}\}$ as solution of (11).
  Update $\boldsymbol{u}$ according to (10).
  Update $\{\boldsymbol{V}',\boldsymbol{Q}',\boldsymbol{R}',\boldsymbol{u}'\} \leftarrow \{\boldsymbol{V},\boldsymbol{Q},\boldsymbol{R},\boldsymbol{u}\}$.
until convergence

$$\begin{aligned}\tilde{g}_k(\boldsymbol{Q}_k,\boldsymbol{Q}_k') =& -\log_2\det(\boldsymbol{Q}_k)\\ & + \log_2\det\left(\sigma_\theta^2 \boldsymbol{I}_{n_\theta} + \boldsymbol{Q}_k'\right)\\ & + \frac{1}{\ln 2} tr\left(\left(\sigma_\theta^2 \boldsymbol{I}_{n_\theta} + \boldsymbol{Q}_k'\right)^{-1}(\boldsymbol{Q}_k - \boldsymbol{Q}_k')\right).\end{aligned} \tag{13}$$

결과적으로 (11)의 최적화 문제는 컨백스 문제로 분류되고, CVX[11] 등 잘 알려진 툴박스를 활용하여 해를 구할 수 있다.

V. 공분산 행렬 전달을 위한 오버헤드 분석

제안하는 연합학습을 위한 상향링크 전송 기술을 동작시키기 위해서는 서버 노드가 전체 상향링크 네트워크의 채널 행렬 정보 $\hat{\boldsymbol{H}}_1, \hat{\boldsymbol{H}}_2, \ldots, \hat{\boldsymbol{H}}_{N_I}$을 획득한 후, Algorithm 1을 수행하여 송신 공분산 행렬 $\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_{N_I}$을 포함하는 최적화 변수들을 결정한다. 이후 계산된 각 공분산 행렬 $\boldsymbol{V}_k$을 ID k에 알려주는 절차가 필요하다. 이 때 서버가 ID k에 보내는 정보의 양을 B_k로 표기하면 (단위: bit), B_k는 다음과 같이 정량화할 수 있다.

$$B_k = \zeta n_I^2. \tag{14}$$

상기 수식에서 ζ는 각 실수 성분 (I 혹은 Q 성분)을 양자화하는 데 사용하는 bit 개수(즉, 해상도)를 나타내며, ζ가 클수록 정확한 공분산 행렬 정보를 ID에 알려주는 반면 하향링크 채널의 통신 부담이 증가하기 때문에, 이 두 측면을 모두 고려한 적절한 해상도 값 ζ을 결정할 필요가 있다. 공분산 행렬 $\boldsymbol{V}_k$를 표현하는 데 필요한 실수 성분 개수는 다음 근거를 통해 n_I^2로 계산하였다. $\boldsymbol{V}_k$는 총 n_I^2개의 복소수로 구성되나, Hermitian 대칭 성질을 이용하여 대각 원소들 기준 한 방향(예: 우-상)을 배제할 수 있고, 대각 원소들은 모두

양의 실수임을 감안하여 허수부를 배제할 수 있음을 고려하여 다음과 같이 계산하였다.

$$\underbrace{\frac{n_I(n_I+1)}{2}}_{\substack{\text{대각 원소 기준}\\ \text{한 방향 제거}}} \times 2 - n_I = n_I^2. \tag{15}$$

서버가 ID k로 내려주는 하향링크 통신의 데이터 전송율을 R_k^{dl}로 표기하면, $\boldsymbol{V}_k$를 ID k에 알리는 데 소요되는 지연시간은 B_k/R_k^{dl}로 계산되며, 여기서 전송율 R_k^{dl}은 하향링크 통신 전략(예: 다중접속 방식, 빔포밍 방식 등)에 의해 결정된다. 본 논문은 모델 업로드를 위한 상향링크 통신 최적화를 집중적으로 다루고 있기 때문에, 하향링크의 통신 오버헤드와 지연시간, 그리고 연합학습의 성능에 미치는 영향 등의 분석은 향후 연구 주제로 남겨 놓는다.

Ⅵ. 모의실험 결과

본 절은 모의실험 결과를 통해 Ⅳ절에서 제안한 연합학습 손실 최소화 알고리즘의 우수성을 검증한다. 반경 100 m의 원형 영역 내에 서버 노드는 중앙에 위치하고 ID들의 위치는 독립적으로 랜덤하게 생성한다. ID k의 채널 행렬 $\boldsymbol{H}_k$의 모든 원소는 독립이며 $CN(0,\rho_k)$의 분포를 따른다. ID k와 서버 노드 간 거리를 γ_k라 할 때, 경로감쇠 ρ_k는 $\rho_k = \rho_0(\gamma_k/\gamma_0)^{-\eta}$로 주어지며, 경로감쇠 지수는 $\eta = 3$, 기준 경로감쇠 및 거리는 $\rho_0 = 10$와 $\gamma_0 = 30$ m로 설정한다. 추정 채널행렬 $\hat{\boldsymbol{H}}_k$도 독립적인 가우시안 분포를 따르며, 각 원소의 분산은 $\rho_k(1-\beta)$로 설정한다. 채널 추정 오류 행렬 $\boldsymbol{E}_k$의 각 원소의 분산 σ_e^2는 $\sigma_e^2 = \rho_k\beta$로 가정한다. 즉, $\beta \in [0,1]$는 채널 전력 중 추정 오류의 비중을 의미하며, $\beta \to 0$일수록 정확한 채널 정보, $\beta \to 1$일수록 부정확한 채널 정보임을 의미한다. 다른 기법 간의 연합학습 성능 비교를 위해 다음과 같이 정의되는 정규화된 MSE (NMSE: normalized MSE) 성능을 고려한다.

$$E\left[\|\hat{\boldsymbol{\theta}}_G - \boldsymbol{\theta}_G\|^2\right] / E\left[\|\boldsymbol{\theta}_G\|^2\right] = E\left[\|\hat{\boldsymbol{\theta}}_G - \boldsymbol{\theta}_G\|^2\right] / \sum_{k=1}^{N_I} \alpha_k^2 n_\theta \sigma_\theta^2. \tag{16}$$

식 (4)와 같이 상향링크 데이터 전송율에 영향을 미치는 SIC 복호화 순서 $\pi(\cdot)$는 임의로 고정되었다고 가정한다.

Ⅳ절에서 제안한 알고리즘의 우수성 확인을 위해 다음 3개 기법의 NMSE 성능을 비교한다.

(1) Fixed $\boldsymbol{V}$: $\boldsymbol{V}_k = (P/n_I)\boldsymbol{I}_{\boldsymbol{n_I}}$을 고정하고, 남은 변수만 최적화한 기법

(2) Non-robust opt.: $\hat{\boldsymbol{H}}_k = \boldsymbol{H}_k$, 즉, 추정 채널이 완벽하다는 가정 하에 최적화한 기법

(3) Proposed robust opt.: Ⅳ절에서 제안한 최적화 기법 (Algorithm 1)

그림 1은 $N_I = 4, n_S \in \{4,8\}, n_I = 2, \beta = 0.1$의 연합학습 환경에서 상향링크 signal-to-noise ratio (SNR) P/σ_z^2 대비 평균 NMSE 성능 비교 그래프를 보여준다. SNR이 증가함에 따라 'Fixed $\boldsymbol{V}$' 및 'Non-robust opt.'의 비교 기법들은 지속적으로 감소하지 않고 일정 수준으로 수렴하는 NMSE 성능을 보이는 반면, 'Proposed robust opt.' 기법은 0에 가까워질 때까지 지속적으로 감소하는 NMSE 성능을 획득하는 것을 볼 수 있다. 서버 노드의 안테나 개수가 $n_S = 4$에서 $n_S = 8$로 증가할 경우, $N_I = 4$개의 ID들이 보낸 상향링크 신호의 간섭 제어 및 복호화 능력이 강화되고, 결과적으로 연합학습을 위한 NMSE 성능이 향상되는 것을 관찰할 수 있다.

그림 2는 $N_I = 4, n_S \in \{4,8\}, n_I = 2$, 15 dB SNR의 환경에서 채널 추정 오류 비중인 β의 증가에 따른 평균 NMSE 성능 비교 그래프를 보여준다. 채널 정보에 오차가 있을 경우 (즉, $\beta > 0$), 'Proposed

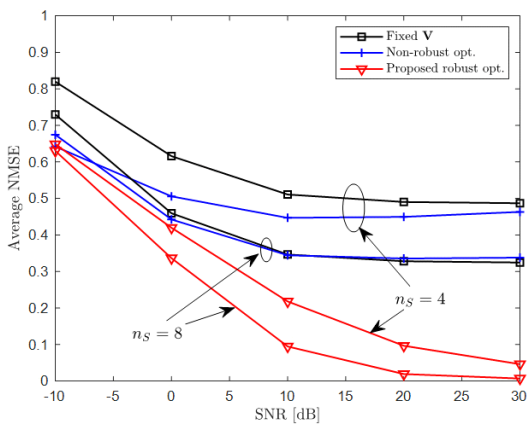

그림 1. SNR P/σ_z^2 대비 평균 NMSE 성능
Fig. 1. Average NMSE versus the SNR

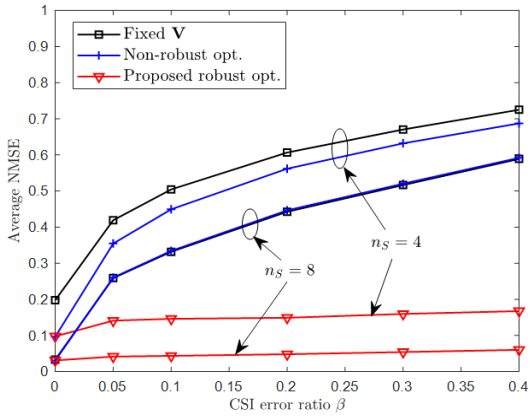

그림 2. 채널 추정 오류 비중 β 대비 평균 NMSE 성능
Fig. 2. Average NMSE versus the CSI error ratio β

robust opt.' 기법이 채널 오차를 고려하지 않는 'Non-robust opt.' 기법 대비 상당히 낮은 NMSE을 획득하는 것을 알 수 있다. 두 기법 간의 성능 차이는 추정 오차 β가 클수록 더 현저해진다. 'Proposed robust opt.' 기법은 채널 오차를 고려하여 $\{\boldsymbol{V}, \boldsymbol{Q}, \boldsymbol{R}, \boldsymbol{u}\}$을 설계하기 때문에, NMSE 성능이 β에 대해 매우 느리게 증가하기 때문으로 해석된다.

마지막으로 그림 3은 그림 1과 같은 환경에서 본 논문에서 다룬 상향링크 전송 기술로 인해 발생하는 글로벌 모델의 NMSE 오차가 실제 기계학습의 성능에 미치는 영향을 관찰한다. 이를 위해 파라미터 벡터 $\boldsymbol{s} = [a \;\; b \;\; c]$로 표현되는 비선형 회귀 모델 $v(x;\boldsymbol{s}) = ax^2 + bx + c$의 fitting 문제를 고려한다. 주어진 ground-truth 파라미터 $\boldsymbol{s}^* = [a^* \;\; b^* \;\; c^*]$에 대해 ID k의 데이터 집합을

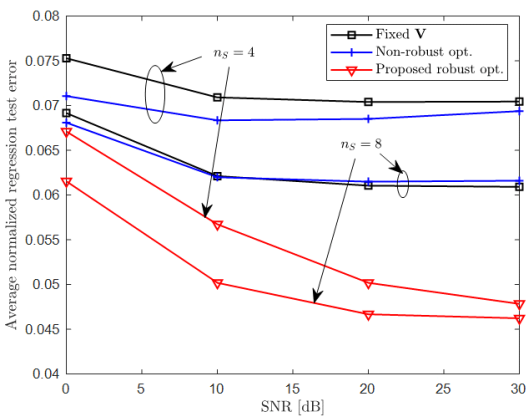

그림 3. SNR 대비 정규화된 평균 회귀 테스트 오차
Fig. 3. Average normalized regression error versus the the SNR

$\overline{D}_k = \{(x_{k,t}, y_{k,t}) | t = 1,2,\ldots,D_k\}$로 표기하고, 각 샘플 $(x_{k,t}, y_{k,t})$은 $x_{k,t} \sim N(0,10)$와 $y_{k,t} = v(x_{k,t};\boldsymbol{s}^*) + z_{k,t}$의 분포를 통해 독립적으로 생성한다. 여기서 $z_{k,t}$는 $z_{k,t} \sim N(0,1)$의 분포를 따른다.

$\overline{D}_{test} = \{(x_{test,t}, y_{test,t}) | t = 1,2,\ldots,D_{test}\}$는 테스트 데이터 집합을 나타내며, 위와 동일한 모델로 생성한다. $f_L(\boldsymbol{s}) = (1/D_T)\sum_{k=1}^{N_I} D_k f_{L,k}(\boldsymbol{s})$로 정의되는 글로벌 손실 함수를 최소화하는 벡터 $\boldsymbol{s} = [a \;\; b \;\; c]$을 찾는 것이 연합학습의 목적이며, $f_{L,k}(\boldsymbol{s})$는 ID k에서의 로컬 평균자승오류(MSE: mean squared error) 함수를 나타내며 $f_{L,k}(\boldsymbol{s}) = (1/D_k)\sum_{t=1}^{D_k}(y_{k,t} - v(x_{k,t};\boldsymbol{s}))^2$와 같이 정의된다. 이를 위해, 매 라운드에서 각 ID는 로컬 MSE 함수에 대한 경사하강법을 통해 $n_L = 20$번의 로컬 업데이트를 수행하고, 서버는 ID들로부터 수집한 파라미터 벡터들을 (1)의 수식을 통해 종합하여 다음 라운드를 위해 ID들에게 전송하며, ID들과 서버가 모델을 주고받는 글로벌 반복 횟수는 $n_G = 100$로 가정한다. 매 실험마다 Ground-truth 파라미터 벡터 $\boldsymbol{s}^* = [a^* \;\; b^* \;\; c^*]$의 원소들은 독립적인 $N(0,1)$의 분포를 통해 생성하며, 총 100개의 랜덤 생성 파라미터 벡터들에 대해 상기 기술된 연합학습을 각각 수행한 후, 테스트 데이터 집합을 통해 다음과 같이 측정한 평균 회귀 테스트 오차 성능을 관찰한다.

$$E_{\boldsymbol{s}^*}\left[\frac{\frac{1}{D_{test}}\sum_{t=1}^{D_{test}}\left(y_{test,t} - v(x_{test,t};\boldsymbol{s}_G)\right)^2}{var\left(\{y_{test,t}\}_{t=1}^{D_{test}}\right)}\right]. \quad (17)$$

상기 수식에서 $\boldsymbol{s}_G$는 각 실행마다 n_G번의 글로벌 반복을 통해 획득한 글로벌 모델 벡터를 의미한다. 그림 3에서 보듯이, 비교 기법 대비 안정된 NMSE 성능을 획득하는 제안 전송 기법이 실제 비선형 회귀모델의 fitting 문제에서도 우수한 학습 성능을 달성하는 것을 볼 수 있다.

Ⅶ. 결 론

연합학습의 모델 종합 정확도를 높이기 위한 상향링크 전송 기술의 최적화 문제를 다루었다. 기존 문헌과 달리 상향링크 채널 정보의 부정확성을 고려하여 상향링크 데이터 전송율 수식을 계산 및 활용하여 최적화 문제를 정립 후, 해당 문제가 비 컨백스 문제임을 고려하여 MM 기반의 반복적 최적화 알고리즘을 제안하였다. 모의실험 결과를 통해 채널 부정확성을 무시하고 설계한 기존 기법 대비 성능 이득을 검증하였다. 서버의 모델 종합 정확도를 높이기 위해 상향링크 무선 채널의 중첩성을 활용하는 아날로그 연합학습 및 ID들 간의 효율적인 스케줄링 기법[7]들과의 결합 연구를 향후 중요한 연구 방향으로 고려한다. 또한 ID들의 데이터 집합들의 분포 간 이형성, 로컬 및 글로벌 반복 횟수 등 학습 모델의 업데이트량, 그리고 연합학습 알고리즘으로 인한 요소 등을 반영한 상향링크 전송 기술의 최적화도 흥미로운 연구 주제가 될 것이다.

References

[1] H. Kim, Y. Kim, C. You, and H. Park, "Efficient distributed clustering algorithm for large-scale federated learning," *J. KICS*, vol. 47, no. 1, pp. 198-205, Jan. 2022. (https://doi.org/10.7840/kics.2022.47.1.198)

[2] D.-Y. Lee and H. Lee, "Autoencoder-based model compression schemes for federated learning," *J. KICS*, vol. 48, no. 3, pp. 295-305, 2023. (https://doi.org/10.7840/kics.2023.48.3.295)

[3] M. Huh, D. Yu, and S.-H. Park, "Signal processing optimization for federated learning over multi-user mimo uplink channel," *2021 ICOIN*, pp. 495-498, Jeju Island, Korea, 2021. (https://doi.org/10.1109/ICOIN50884.2021.9333891)

[4] Z. Yang, M. Chen, C. S. Hong, and M. Shikh-Bhaei, "Energy efficient federated learning over wireless communication networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1935-1949, Mar. 2021. (https://doi.org/10.1109/TWC.2020.3037554)

[5] X. Wei and C. Shen, "Federated learning over noisy channels: Convergence analysis and design examples," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 2, pp. 1253-1268, Jun. 2022. (https://doi.org/10.1109/TCCN.2022.3140788)

[6] S.-H. Park and H. Lee, "Completion time minimization of fog-RAN-assisted federated learning with rate-splitting transmission," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 10209-10214, Sep. 2022. (https://doi.org/10.1109/TVT.2022.3180747)

[7] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2022-2035, Mar. 2020. (https://doi.org/10.1109/TWC.2019.2961673)

[8] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 76-88, Jan. 2016. (https://doi.org/10.1109/TSP.2015.2480042)

[9] A. E. Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge, U.K.: Cambridge Univ. Press, 2011.

[10] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794-816, Feb. 2017. (https://doi.org/10.1109/TSP.2016.2601299)

[11] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," *Second Edition, Third Printing*, pp. 1-786, Ver. 2.2, Jan. 2020. [Online]. Available: http://cvxr.com/cvx

유 원 식 (Wonsik Yoo)

2018년 8월 : 전북대학교 전자공학부 졸업
2022년 9월~현재 : 전북대학교 전자정보공학부 석사과정
<관심분야> 통신신호처리, 최적화, 기계학습

박 석 환 (Seok-Hwan Park)

2005년 : 고려대학교 전기전자전파공학부 학사
2011년 : 고려대학교 전자전기공학과 박사
2012년~2014년 : NJIT in USA, 박사후연구원
2014년~2015년 : 삼성전자 책임연구원
2015년~현재 : 전북대학교 조교수/부교수
2022년~현재 : IEEE Trans. Wireless Commun. 편집위원
<관심분야> 통신신호처리, 최적화, 기계학습