

# 통신 및 네트워크 상황을 고려한 연합학습 참여 기기 선택 기법 연구

이민정, 성민지, 이웅희

한성대학교

{jeong0304, minji179, whlee}@hansung.ac.kr

## A Study on Participation Selection Scheme Considering Communication and Network Situation

Minjeong Lee, Minji Sung, Woonghee Lee

Hansung University

### 요약

기술의 발전에 따라 분산된 개별 기기에서 독립적으로 생산, 수집 및 저장되는 데이터양이 급속도로 증가하였다. 이러한 데이터들을 하나의 서버에 모아 학습을 수행하는 것이 일반적이지만 이러한 방법은 효율성 및 비용 등에서 많은 제약이 따르므로 이를 극복하기 위해 연합학습이 등장하게 되었다. 연합학습은 서버와 다수의 참여 기기들 간 통신들을 반복하여 수행되므로, 참여 기기들의 통신 및 네트워크 상황에 따라 지연이 발생할 수 있고 이는 결국 연합학습 시스템의 성능 저하로 이어지게 된다. 따라서 본 논문에서는 통신과 네트워크 상황에 따라 연합학습 참여 기기를 선택하는 기법을 제안하고 이를 평가하였다.

### I. 서론

기술의 발전에 따라 스마트 기기 및 사물인터넷 장치와 같은 기기들과 여러 기관에서 독립적으로 생산, 수집 및 저장하는 데이터양이 급속도로 증가하고 있다. 이러한 분산형 데이터들을 하나의 서버에 모아 인공지능 모델을 학습시키는 것이 일반적이었다. 그러나 이러한 방법들은 효율성 및 비용 등 많은 제약이 따르게 된다. 따라서 분산된 로컬 데이터를 한곳에 모으지 않고 데이터를 보유한 기기에서 독립적으로 학습을 수행하고 학습 결과를 서버에서 통합하도록 하는 연합학습이 등장하게 되었다[1]. 연합학습을 이용하면 raw 데이터를 송수신할 필요가 없기 때문에 통신자원을 절약하고 개인정보가 보호되는 학습이 가능하다[2]. 하지만 참여 기기 간 통신 및 네트워크의 상황에 차이가 존재하기 때문에 일부 기기의 응답이 늦어질 경우 연합학습 전체가 지연되는 결과를 초래한다. 따라서 본 논문에서는 통신과 네트워크 상황에 따라 연합학습 참여 기기를 선택하는 기법을 제안하고 이를 평가하였다.

### II. 연합학습

연합학습이란 다수의 로컬 클라이언트와 하나의 중앙 서버가 협력하여 데이터가 탈중앙화된 상황에서 글로벌 모델을 학습하는 기술을 말한다[2]. 그림 1은 연합학습의 개념도이다. 먼저, 중앙서버에서 제작한 학습 모델을 참여 기기들에게 배포한다. 참여 기기들은 각 기기들의 로컬 데이터를 활용하여 배포 받은 모델을 학습시킨 후 학습 모델을 중앙서버로 다시 전송한다. 마지막으로, 중앙서버는 학습된 모델을 통합하고, 통합된 모델을 다시 사용자에게 배포한다[3].

서버와 참여 기기는 학습 모델을 여러 번 반복해서 주고받기 때문에 참여 기기의 통신과 네트워크 상황을 고려하는 것이 매우 중요하다. 또한, 서버가 매 라운드마다 참여 기기를 선택하기 때문에 통신과 네트워크를 고려하여 매 라운드마다 적절한 기기를 선택하는 것이 중요하다.

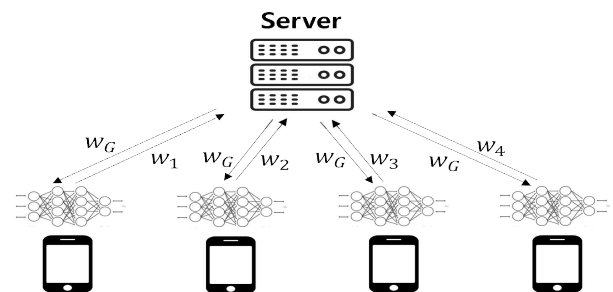


그림 1 연합학습 개념도

### III. 통신 및 네트워크 상황을 고려한 연합학습 참여 기기 선택 기법

#### 1. 구현 알고리즘

서버와 클라이언트 간 통신 및 네트워크 상황을 파악하기 위해 통신과 네트워크 지표들로 나누어 성능 지표들을 측정하고 서버에 저장한다.

통신 지표로는 Link Speed, RSSI(Received Signal Strength Indicator)를 측정한다. Link Speed는 장치가 무선 라우터와 통신할 수 있는 최대 속도를 의미하며 고속 인터넷의 경우 100Mbps, 기가비트 인터넷의 경우 1Gbps의 속도를 가진다. RSSI는 수신 신호 세기/강도를 말한다. 일반적인 통신에서는 -99 dBm에서 35 dBm까지의 세기를 가지는 신호를 송출하며 숫자가 높을수록 신호가 강하다는 것을 의미한다.

네트워크 지표로는 Throughput, RTT(Round Trip Time), Jitter, Hop Count를 측정하였다. Throughput은 네트워크 상에서 한 노드로부터 또 다른 노드로 전달되는 단위 시간당 데이터량을 말한다. RTT는 패킷이 목적지에 도달하고 나서 해당 패킷에 대한 응답이 출발지로 다시 돌아오기까지의 시간이다. Jitter는 패킷 지연(Delay)의 변동으로 인해 패킷 간 간격이 일정하지 않는 현상을 말하며, 순간 지연 값과 평균 지연 값과의 차이를 의미한다. 마지막으로 Hop Count는 데이터가 출발지와 목적지 사이에서 통과해야 하는 중간 장치들의 개수를 의미한다.

그림 2는 구현 알고리즘의 순서도를 보여준다. 클라이언트들의 통신 성능 지표 측정이 끝나면 저장된 지표를 서버에서 읽어와 값들을 비교해 통신 성능이 좋은 클라이언트 순으로 나열한다. 이때, 측정값 간에 차이가 커서 변별력이 크고, 데이터 송수신 속도와 관련성이 큰 성능 지표들을 우선적으로 비교하게 하였다. 가장 먼저, Throughput 값들을 비교해 통신 성능이 가장 좋은 순으로 클라이언트를 나열하였다. 만약, 가장 높은 Throughput 값을 가지는 클라이언트가 2개 이상일 경우 추가적으로 RTT 값을 비교하며, 다음으로는 Jitter 값을 비교하도록 알고리즘을 구현하였다. 비교가 끝나면 우선순위가 높은 클라이언트 순으로 서버가 학습을 요청한다.

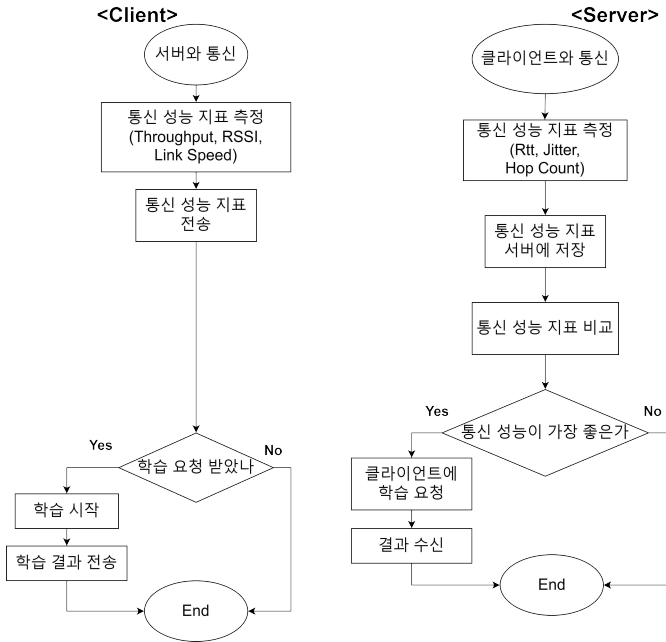


그림 2 구현 알고리즘 순서도

## 2. 실험 및 평가

제안한 기법의 유효성을 평가하기 위한 실험을 진행하였다. 통신 및 네트워크 상황에 따라 적절하게 참여 기기를 선택하는지 평가하기 위해 서로 다른 통신 및 네트워크 상황을 가지는 2개의 클라이언트들과 1대의 서버로 구성된 연합학습 환경을 구성하였으며, 그림 3은 실험 환경을 보여준다. 실험은 Jetson TX2를 클라이언트 기기로 사용하였다. 표 1은 실험에 사용된 Jetson TX2와 서버 PC의 사양을 보여준다.



그림 3 실험 환경

이러한 환경에서, 제안한 알고리즘을 적용했을 때와 랜덤하게 클라이언트들을 선택하는 기존 기법을 적용했을 때의 실험 결과를 비교하였다. 서버가 클라이언트에게 학습을 요청한 시점부터 학습 결과를 전송받을 때까지의 시간의 평균값을 측정하였으며 그림 4는 본 실험의 결과를 보여준다. 제안한 알고리즘을 적용하였을 경우, 클라이언트들의 통신 및 네트워크 성능 지표 값들을 비교하여 통신 성능이 좋은 클라이언트에게 학습을 요

청한다. 반면 기존 기법은 랜덤하게 클라이언트를 선택하여 학습을 요청하였다. 따라서 제안한 알고리즘을 적용한 경우, 랜덤으로 클라이언트를 선택하는 기존 기법을 적용하였을 때 보다 연합학습이 적게는 20초 많게는 50초 정도 빠르게 수행된 것을 볼 수 있다. 이 결과를 통해 제안한 기법을 사용하면 통신 및 네트워크 상황에 따라 참여 기기를 적절하게 선택할 수 있고 이는 연합학습 시스템의 성능 향상으로 이어짐을 알 수 있다.

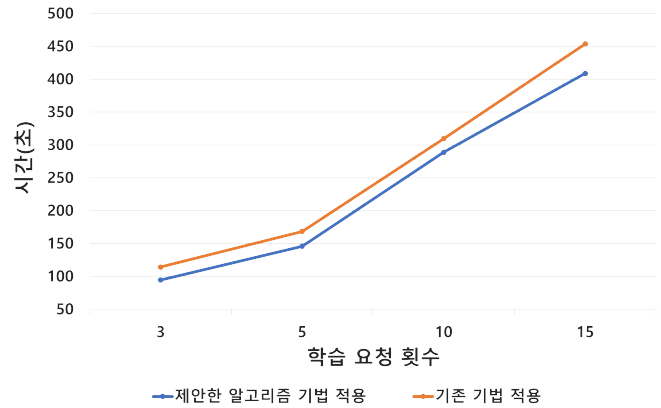


그림 4 학습 요청 횟수에 따른 연합학습 수행 시간

구성요소	사양	
	Jeston TX2	PC
Processor	Dual-core Denver + quad-core ARM A57	AMD Ryzen 7 6800 @3.20GH
Memory	8GB 128-bit LPDDR4 Memory 1866 MHz - 59.7 GB/s	32GB
OS	Ubuntu 18.04	Windows 11

표 1. 실험에 사용된 기기들의 사양

## IV. 결론

본 논문에서는 통신과 네트워크 상황에 따라 연합학습 참여 기기를 선택하는 기법을 제안하고 이를 평가하였다. 참여 기기들의 통신 및 네트워크 상황을 측정하고 이를 바탕으로 적절하게 참여 기기를 선택하는 알고리즘을 제안하였으며 실제 기기로 구성된 연합학습 환경을 구성하여 실험을 수행하였다. 실험 결과를 통해 제안한 기법을 사용할 경우 통신 및 네트워크 상황에 따라 참여 기기를 적절하게 선택할 수 있음을 보였으며, 이를 통해 연합학습 시스템의 성능 향상이 가능하다는 것을 보였다.

## ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2021R1G1A1092939).

## 참 고 문 헌

- [1] 이승민. "연합학습 기술 동향 및 산업적 시사점". 한국전자통신연구원, 2020.
- [2] McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." Artificial intelligence and statistics. PMLR, 2017.
- [3] Li, Qibin, Zeyi Wen, and Bingsheng He. "Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection." (2019).