Flower을 사용한 점진적 연합학습시스템 구성

[1] 강윤희, [2*]강명주

Construction of Incremental Federated Learning System using Flower

[1]Yun-Hee Kang, [2*]Myungju Kang

요 약

인공지능 분야에서 학습모델을 구성하기 위해서는 학습데이터의 수집이 선행되어야 하며, 학습데이터를 학습모델 구성이 이루어지는 중앙 서버로 전달하여야 한다. 연합 학습은 클라이언트 측면의 데이터 이동없이 협력적은 방법으로 전역 학습 모델을 구성하는 기계학습 방법이다. 연합학습은 개인 정보를 보호하기 위해 활용될 수 있으며, 개별 클라이언트에서 로컬 학습모델을 구성한 후 로컬 모델의 매개변수를 중앙에서 집계하여 전역 모델을 업데이트한다. 이 본문에서는 연합학습의 개선을 위해 기존의 학습 결과인 학습 매개변수를 사용한다. 이를 위해 연합학습 프레임워크인 Flower를 사용하여 실험을 수행한 후 알고리즘의 수행시간 및 최적화에 따른 결과를 평가하여 제시한다.

Abstract

To construct a learning model in the field of artificial intelligence, a dataset should be collected and be delivered to the central server where the learning model is constructed. Federated learning is a machine learning method building a global learning model without transmitting data located in a client side in a collaborative manner. It can be used to protect privacy, and after constructing a local trained model on individual clients, the parameters of the local model are aggregated centrally to update the global model. In this paper, we reuse the existing learning parameter to improve federated learning, describe incremental federated learning. For this work, we do experiments using the federated learning framework named Flower, and evaluate the experiment results with regard to elapsed time and precision when executing optimization algorithms.

Keywords: *Federated learning, Distributed learning, Flower, Optimization algorithms, Performance evaluation*

[1] *백석대학교 컴퓨터공학부 부교수(yhkang@bu.ac.kr)*
[2*] *교신저자 넥타르소프트 연구소장(kmjziro@daum.net)*

Received: July. 26, 2023, Revised: Aug. 16, 2023, Accepted: Aug. 21, 2023

I. 서론

스마트 폰, CCTV, 센서가 연결된 초연결 환경의 출현으로 인터넷에 연결된 장치로부터 생성되는 데이터의 종류와 양이 급하게 증가하고 있다. 이를 활용한 응용 요구가 커지고 있으며 데이터의 활용분야와 데이터 시장 또한 확장하고 있다[1]. 국내외적으로 다양한 분야에서 인공지능에 대한 활용은 증가하고 있지만 개인의 의료 데이터, 금융 데이터 등은 개인정보 영역으로 학습 데이터의 프라이버시를 유지하기 위한 다양한 규제로 인해 해당데이터의 외부활용은 어려워지는 추세이다. 이러한 추세의 예로 유럽연합은 EU 회원국가의 국민의 사생활 및 개인정보를 보호하기 위해 GDPR 을 2018 년 제정하여 적용하고 있다[2]. 국내에서도 2020 년 데이터 활용을 위한 법안이 시행되고 있으며, 개인데이터의 활용 범위 및 요구사항을 규정하고 있다[2]. 이러한 변화된 환경에서 개별데이터의 이동없이 학습모델을 구성하기 위한 연구에 대한 요구가 증가되고 있다[3][4].

인공지능 분야에서 학습모델을 구성하기 위해서는 학습데이터의 수집이 선행되어야 하며, 학습데이터를 학습모델 구성이 이루어지는 중앙 서버로 전달하여야 한다. 그러나 외부의 학습데이터를 사용하여 기계 학습모델을 구성하는 과정에 분산 되어있는 데이터들을 한 곳으로 모아 전체 학습데이터를 수집하고 학습시켜야 하며, 이는 데이터 보안 취약성 및 데이터 소유자의 이해관계의 제약이 따른다. 특히 학습데이터의 이동은 사생활 및 개인정보를 보호에 문제점을 갖는다.

연합 학습(FL, Federated Learning)은 데이터 이동없이 개별 클라이언트의 데이터로 구성된 로컬 학습모델을 서버에서 집계하여 중앙의 전역 모델을 로컬 학습모델에 반영 업데이트하는 기계학습 기법이다[3]. 연합학습은 개인 정보를 보호하기 위해 제안된 기계 학습 패러다임으로 개별 클라이언트에서 로컬 학습모델을 구성한 후 로컬 모델의 매개변수를 중앙에서 집계하여 전역 모델을 업데이트하는 과정을 수행한다. 효율적인 연합 학습을 위해서는 통신비용, 시스템 이질성, 통계적 이질성 및 프라이버시 문제 해결을 요구한다[5][6][7][8].

기계학습은 수집된 데이터셋에 대한 분석 후 학습 모델을 구성 후 데이터셋을 입력으로 받아 학습과정에서 모델 파라메터의 최적화를 수행한다. 연합학습의 성능은 클라이언트의 상이한 네트워크 대역폭 및 계산처리능력 등 제한된 자원으로 인해 대규모 로컬연산 보다 느릴 수 있다. 연합학습이 수행되기 위해서는 학습에 참여하는 클라이언트 간 통신이 유지되어야 하므로 네트워크 환경은 연합학습에 주요한 영향을 미친다. 중앙에서 반복적으로 데이터를 학습에 입력하여 학습모델을 최적화하는 것과는 달리 연합학습에서는 통신 라운드 수와 통신 비용에 의해 종속된다. 각 클라이언트의 데이터 집합이 상대적으로 작음에 따라 추가적인 연산을 사용하여 모델을 훈련하는데 요구되는 통신 라운드의 수 또는 통신 비용을 낮추는 것이 요구된다.

본 논문에서는 연합학습의 최적화(federated optimization)를 개선하기 위한 기존의 학습결과를 활용하며 학습에 필요한 횟수를 줄이기 위한 점진적 연합학습 시스템 구성을 제시한다. 이를 위한 점진적 연합학습을 기술하고, 연합학습 프레임워크인 Flower[9]를 사용하여 연합학습 알고리즘을 적용하여 실험을 수행한 후 알고리즘의 수행시간 및 최적화에 따른 결과를 평가하여 제시한다.

II. 관련연구

본 절에서는 기존의 중앙 서버에서의 학습모델 구성과의 차이점을 연합학습의 개요에서 설명하고, 연합학습 수행의 단계 및 단계의 수행 내용을 연합학습 과정에서 기술한다. 마지막으로 FedSGD, FedAvg, FedProx 등의 연합학습 알고리즘 특징을 기술한다.

2.1 연합학습 개요

중앙집중형 학습모델(centralized training model) 구성은 데이터 전달과정에서 중앙서버로부터 개인정보 유출의 잠재적 문제점을 갖는다. 연합학습은 다수의 클라이언트를 갖는 중앙처리

기반의 기계학습 모델을 분산학습(distributed learning)으로 구성하는 방법이다[3]. 연합학습은 클라이언트가 소유 및 제어하고 있는 데이터가 분산되어 있는 환경에서 데이터를 한곳에 모으지 않고 각 클라이언트에서 학습한 뒤 이를 활용한다는 점에서 horovod 기반의 분산학습과 차별점을 갖는다[10].

일반적으로 스마트폰에 저장된 데이터는 개인의 사진, 위치정보 및 검색단어 등의 민감한 개인정보를 포함한다. 2016 년 구글은 차세대 기계학습 방법으로 구글이 개발한 안드로이드와 iOS 용 가상 키보드 앱 스마트폰인 지보드에 연합학습을 적용한다. 사용자의 단어 추천을 위한 지보드 서비스의 학습모델을 구성하기 위해 연합학습을 제안한다. 제안된 연합학습 기반 기계학습 방법은 개인 맞춤형 학습모델 개발을 위해 다수의 클라이언트가 참여하여 학습모델을 구성하는 방법을 활용한다[4]. 그림 1 은 전체적인 지보드 서비스의 동작을 보인 것이다. 각 모바일 디바이스는 개별 디바이스의 자료만으로 학습을 수행하여 학습결과인 지역학습모델을 구성한 후 학습을 통해 구성된 지역학습모델을 서버에 전달한다. 서버에서는 개별 디바이스로 부터 전달된 학습모델을 사용하여 새로운 전역모델을 생성한 후 이를 다시 개별 디바이스에 전달한다. 연합학습을 사용한 지보드 서비스 구성은 개인 자료의 유출없이 사용자의 입력 특성을 기반으로 단어추천을 수행할 수 있으며, 이 과정에서 다른 디바이스에서의 학습 결과를 반영한 전역모델을 사용한다.

Figure 1. Google G-board service based on federated learning
그림 1. 구글 지보드 서비스 연합학습적용

2.2 연합학습 학습과정

학습모델은 데이터의 관계를 표현하는 함수로 이해할 수 있으며, 학습모델 파라메터 값은 학습에 사용되는 훈련과정에서 모델에 의해 조정된다. 모델 파라메터는 학습모델, 데이터셋 특징 및 데이터 운영에 따라 달라진다. 학습모델의 개선을 위해서는 학습의 오차를 최소화하는 최적화 과정에서의 파라메터의 조정 과정을 요구한다.

연합학습은 다수의 클라이언트에 저장된 데이터로부터 최적화된 전역 학습모델을 생성하는 것을 목표로 한다. 이를 위해 연합학습은 개별 클라이언트에서 생성된 데이터가 로컬로 저장되고 처리된다는 제약조건을 보장하고 각 클라이언트와 중앙서버의 주기적인 통신을 통해 전역학습모델을 업데이트한다.

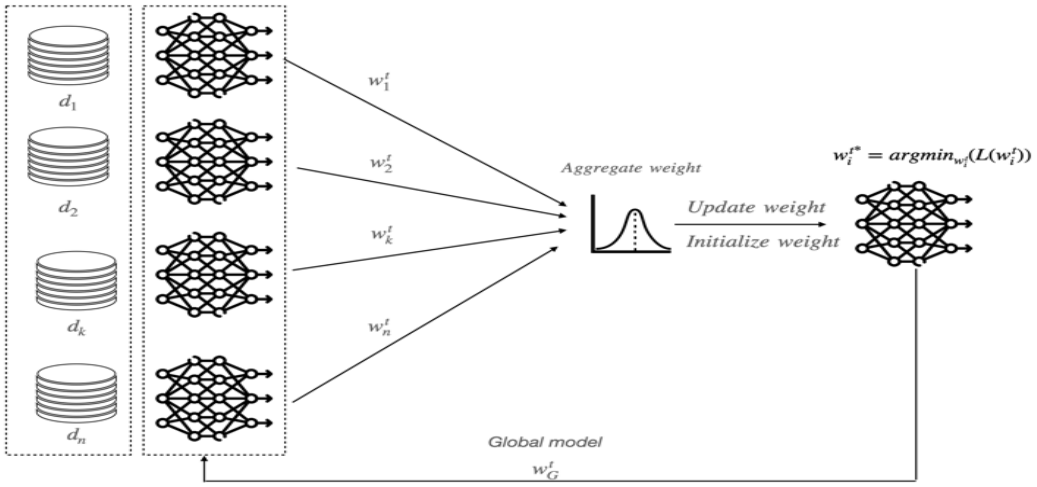

Figure 2. The overall process of federated learning
그림 2. 전체 연합학습 과정

그림 2 는 연합학습과정을 보인 것이다. 그림 2 에서 먼저 서버는 $t = 0$ 에서의 초기 매개변수인 가중치 ω_G^0의 학습모델을 클라이언트에게 배포한다. 각 클라이언트는 로컬 데이터를 활용해 배포받은 모델을 기반으로 학습된 학습모델을 구성하는 매개변수 $1 \le i \le n$ 인 ω_i^t를 서버로 전달한다. 서버는 학습된 매개변수를 통합하여 ω_G^t 를 갖는 전역 모델을 구성한 후 다시 클라이언트에 배포한다. 연합학습는 학습데이터를 외부 데이터 이동 없이 클라언트에서 유지하며 개별적인 지역 학습모델을 구성하며, 학습과정에서 얻은 매개변수인 가중치를 서버로 전달한다. 서버는 지역 학습모델의 매개변수의 통합한 후 통합된 결과를 기반으로 최적화를 수행한 후 전역모델을 생성한다. 이후 서버는 클라이언트의 지역 학습모델을 갱신한다[3][4][5].

그림 3 은 연합학습 수행과정 프로토콜에 따른 처리흐름을 보인 것으로 C_1, C_2, C_3 및 C_4 로 구성된 4 개의 클라이언트와 1 개의 서버 S 로 구성된다. 연합학습 과정은 선택, 형상관리, 리포트의 3 단계로 이루어진다.

- 선택 단계는 초기 학습에 참여하는 클라이언트의 등록 과정을 수행한다. 등록 후 형상관리 단계에서 서버는 학습에 참여하는 클라이언트는 선정 후 학습에 참여하는 클라이언트에는 작업 수행을 위한 초기학습 매개변수 및 학습을 위한 하이퍼매개변수 정보를 전달한다.
- 형상관리단계에서 클라이언트는 자신의 데이터셋을 사용하여 학습을 진행하며 라운드 별로 로컬모델을 생성한다.
- 리포트 단계에서 클라이언트 자신의 학습 결과를 서버에 전달하고 서버는 FedAvg 와 같은 연합학습 알고리즘을 사용하여 전역학습 모델을 구성한다. 서버는 전역학습 모델을 체크포인트 하여 라운드 별 학습모델을 저장한다.

그림 3 에서와 같이 초기 3 개의 클라이언트인 C_1, C_2, 및 C_3 부터 연합학습 참여에 대한 응답을 전달받게 된다. 선택단계에서는 C_1, C_2, 및 C_3 가 연합학습에 참여하도록 등록되며, 형상관리과정에서 정의된 정책에 따라 초기 학습 매개변수를 전달한 후 지역학습을 진행한다. 이과정에서 이전에 학습결과의 학습매개변수는 초기 매개변수로서 활용할 수 있다. 리포트 단계에서는 네트워크 단절로 인해 C_1 와 C_2, 클라이언트로부터 전달받은 두개의 학습결과를 집계한 후 전역 학습모델을 구성한다. 이후 다음 라운드에서 사용하기 위해 체크포인트를 구성한다.

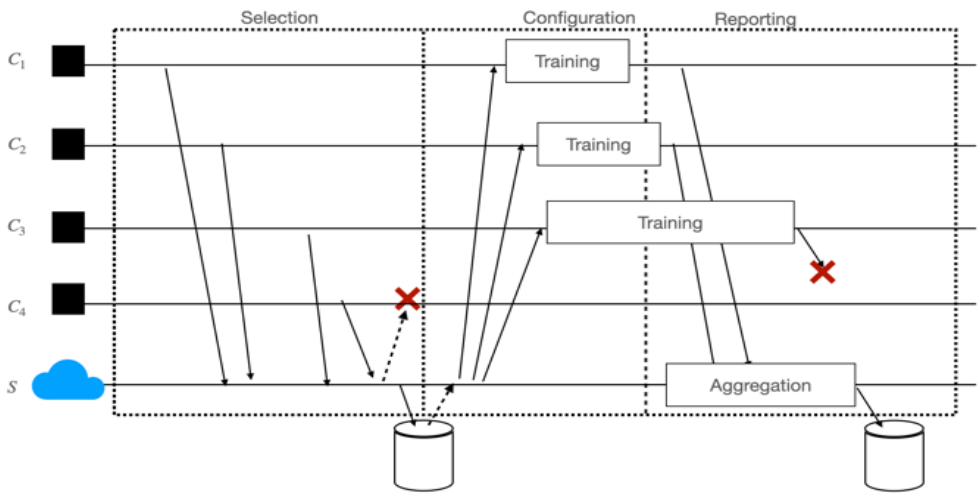

Figure 3. The protocol of federated learning implementation
그림 3. 연합학습 수행 프로토콜

2.3 연합학습 알고리즘

확률적 경사 하강 (SGD, Stochastic Gradient Descent)는 기계 학습에서 가장 기본적으로 사용되는 최적화 알고리즘 중 하나로, 경사 하강법(Gradient Descent)의 확률적인 버전이다. 확률적 경사 하강에서 기계학습 모델 생성은 학습데이터를 섞은 후 선택한 데이터 점을 사용하여 학습의 오차를 최소화하는 가중치를 반복적으로 수정하는 최적화 진행에 의해 결정된다. 확률적 경사 하강에서 함수의 오차를 최소화하기 위해 경사와 학습률을 조정한다. 경사는 학습율의 방향을 정하기 위한 함수 변화율(the rate of change of a function)을 의미하며, 학습률은 경사하강 수행 중 가중치를 수정하는 이동폭을 의미한다. SGD 에서 경사는 전체 데이터 집합의 임의 하위 집합에서 계산된 이후 경사 하강의 한 단계를 생성하기 위해서 사용된다. 수식 (1)에서 가중치 ω 는 현재의 가중치에서 학습률 η 와 ω 의 가중치를 오차를 줄이기 위한 오차 미분값을 구하기 위한 손실함수 F 를 반영하여 가중치를 수정한다. 손실함수 F 는 예측값과 실제값의 차이를 비교하기 위한 함수로서 반복학습을 통해 이들 두개 값을 일치하기 위해 차이값을 최소화하여 최적화한다.

$$w = w + \eta \nabla F(\omega) \text{ where } \omega \in \mathbb{R}^d \quad \text{- (1)}$$

연합학습에 적용되는 주요알고리즘은 FedSGD(Federated Stochastic Gradient Descent), FedAvg(Federated Averaging), FedAdam, FedAdagrad 등이 있다.

FedSGD 알고리즘[3]은 각 통신 라운드마다 정의된 비율의 클라이언트를 선택하고, 해당 장치의 데이터 집합을 사용한다. 이에 따라 FedSGD 에서 경사는 각 장치의 훈련 데이터 갯수와 비례하여 중앙에서 평균화되고, SGD 와 같이 경사 하강의 한 단계를 생성하기 위해서 사용된다. 매 라운드마다 전체 클라이어트 중 m 개의 client 가 가진 모든 데이터를 한 번에 학습한다. 즉, 학습에 참여하는 클랑이언트들이 가진 데이터의 전체 배치의 경사하강를 1 회 계산하여 1 번의 지역 갱신(local update)만 반영하는 특징을 갖는다. 수식 (2) 에서 m 은 전체 클라이언트의 수, 클라이언트 i 의 손실함수 F_i 는 ω 을 매개변수로 갖는 i 번째 클라이언트, p_i 는 i 번째 클라이언트가 처리하는 데이터의 가중치이다. FedSGD 목표는 목적함수 f(ϖ) 를 최소화하기 위해 학습과정을 통해 모든 클라이언트 데이터의 평균모델을 찾는다.

$$\min_{\omega \in \mathbb{R}^d} f(\omega) = \frac{1}{m} \sum_{i=1}^{m} p_i F_i(\omega) \quad \text{- (2)}$$

수식 (3)는 새로운 가중치 w_{t+1} 는 t 시점에서 학습 매개변수는 ω_t 에 클라이언트의 학습결과를 반영하여 갱신하여 얻는다. $\nabla F_i(\omega)$ 는 클라이언트 i 가 다루는 데이터의 손실을 구하기 위한 경사값이다.

$$w_{t+1} = \omega_t + \eta \nabla F_i(\omega) - (3)$$

FedAvg 는 FedSGD 알고리즘의 확장으로 중앙 서버와 경사를 공유하는 방법 대신 클라이언트의 로컬 모델을 공유하는 방식을 제공한다. FedAvg 는 FedSGD 가 연합학습 과정에서 1 라운드의 local update 만 반영한다는 것은 제약점을 해결하기 위해 연합학습에 참여하는 클라이언트가 일정한 배치수 만큼 학습을 수행한 후의 매개변수를 서버로 전달하는 과정을 수행한다. 클라이언트는 학습데이터를 나누어 학습하는 미니배치 효과를 얻게 되어 전역 파라메터가 수렴하는 시간을 단축할 수 있다. FedAvg 는 FedSGD 의 일반화방식으로서 각 클라이언트에게 FedSGD 에 비해 지역 갱신(local update)을 반복시켜 수식 (4)의 평균화 단계에 도달하기 전, 더 많은 계산을 수행할 수 있다. FedAvg 알고리즘은 파라미터 누적을 동기적으로 진행하기 때문에 각 학습 라운드는 가장 느린 디바이스의 속도와 같다. 게다가 모델은 학습 라운드가 이미 진행 중일 때 중간에 참가할 수 있는 참가자를 고려하지 않는다.

$$g_k = \nabla F_k(\omega_t)$$
$$w_{t+1} = w_t + \eta g_k \text{ - (4)}$$

연합학습 최적화을 위해 적응형 알고리즘인 FedAdagrad 에서는 동일 기준으로 갱신되는 클라이언트의 가중치 매개변수에 개별 기준을 적용하도록 한다. 이를 위해 가중치 갱신시 학습률에 반비례로 적용되어 높은 값을 가지는 매개변수에서는 상대적으로 적은 변화를 주고 반대로 적게 이동한 매 매개변수에서는 큰 변화를 반영하도록 한다. FedAdam 은 클라이언트의 가중치의 변화율값과 변화율의 제곱값의 지수이동평균을 활용하여 학습단계의 변화량을 조절한다.

III. 점진적 연합학습

3.1 Flower 개요

Flower 는 연합학습을 위한 라이브러리 수준의 플랫폼을 제공하는 연합학습 응용 프레임워크로서, 10,000 개 이상의 다수 클라이언트를 구성하여 학습을 진행할 수 있는 확장성을 가지고 있다[8]. Flower 는 클라이언트로서 모바일 장치와 같은 다양한 장치를 지원하며 기존의 인공지능학습 프레임워크인 사이킷런(Scikit-learn) 파이토치(PyTorch)와 텐서플로(TensorFlow)를 사용하여 응용을 구성할 수 있도록 지원한다.

3.2 FL 응용 구성

점진적 연합학습을 위해 Flower 응용는 서버와 클라이언트 구성되며, MNIST 의 분류를 진행하기 위해 응용에 점진적 FL 모델 학습을 수행하도록 작성한다. 이들 간의 통신은 G-RPC 를 사용하여 이루어진다. 서버와 클라이언트 간 통신을 위한 TCP 포트를 구성하다.

서버는 클라이언트의 학습결과를 집계하는 역할을 수행하며 정해진 수행 횟수의 집계를 마치고 종료한다. 서버는 집계된 학습결과의 연합학습을 위한 알고리즘을 전략(strategy)에 정의하여 FL 수행 환경을 구성한다. 연합학습을 위한 알고리즘은 클래스로 구성된 FedAvg, FedAdam, FedYogi 를 선택하여 FL 의 최적화를 수행하도록 서버에 적용한다.

- on_fit_config_fn 는 선택과정에서 필요한 하이퍼파라매터를 클라이언트에 전달
- initial_parameters 는 클라이언트 전달을 위한 초기학습 매개변수를 설정하며, 점진적 FL 을 위해 사전에 저장된 학습 매개변수가 있는 경우 초기합습 매개변수로 선정함
- evaluate_fn 는 서버에서의 전역 학습모델의 성능을 평가하기 위한 함수를 설정

클라이언트는 서버로 부터 전달된 학습 매개변수를 입력으로 기계학습의 매개변수를 설정한 후

학습을 수행한다. 학습 수행을 위해서는 fit() 함수를 사용하며, 생성된 로컬 학습모델의 파라메터와 성과지표인 손실값(loss) 과 정확률(accuracy)를 서버에 전달한다.

3.3 실험 결과

본 실험에서는 Flower 1.3 을 기반으로 MNIST 데이터셋을 사용하여 FL 응용을 구성한다. 응용은 1 개의 서버와 2 개의 클라이언트가 수행하도록 분산컴퓨팅 환경과 단일 서버환경의 2 종류 환경을 구성한다. 분산 컴퓨팅환경의 클라이언트는 파이썬 3.8 으로 Raspberry PI 4 와 Jetson Nano 에 각각 클라이언트로 구성한다. 서버는 Intel(R) Xeon(R) 서버에 구성한다. 단일서버환경 은 클라이언트와 서버를 Intel(R) Xeon(R) 서버에 구성한다. 표 1 은 FL 의 FedAvg 을 사용하여 20 라운드의 수행에 대한 라운드 별 평균응답시간과 표준편차를 실험결과를 보인 것이다. 실험결과의 분석을 통해 라운드별 응답시간은 분산컴퓨팅 환경이 단일서버환경 보다 응답시간은 2.67 배 및 표준편차는 3.4 배임을 관찰되었다. 이를 통해 FL 의 주요한 수행시간의 네트워크 상의 메시지 전송시간에 의해 결정됨을 확인할 수 있다.

Table 1. Comparison of response time between distributed computing environment and single server
표 1. 분산컴퓨팅 환경과 단일서버 환경 응답시간

	Distributed computing environment	Single server environment
Average response time of rounds (second)	1.21	0.46
Standard deviation	0.33	0.11

그림 4 는 점진적 FL 의 최적화 알고리즘을 FedAvg, FedAdam 및 FedAdagrad 적용한 성능지표인 분류 정확률을 보인 것이다. 학습결과인 모델 파라메터는 numpy 의 저장 형식으로 저장한다. 클라이언트의 학습 파라메터 설정 시 체크포인트를 통해 저장된 마지막 학습매개변수를 반영하여 클라이언트의 학습이 이루어짐에 따라 단속적인 클라이언트 환경에서도 전체 학습시간을 줄일 수 있는 잇점을 얻을 수 있음을 확인할 수 있다. 지역학습 매개변수의 산술평균을 적용하는 FedAvg 는 정확률인 점진적으로 개선되는 것을 확인할 수 있으며, 모멘텀이 적용되는 FedAdam 과 FedAdagrad 의 경우 정확률의 변화가 큰것을 확인할 수 있다.

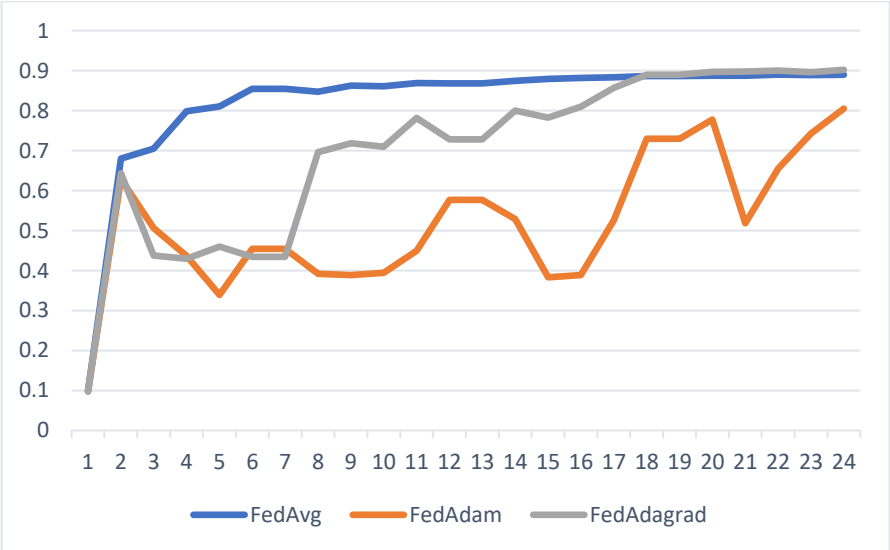

Figure 4. Comparison of optimization algorithm of incremental federated learning
그림 4. 점진적 FL 의 최적화 알고리즘 비교

IV. 결론 및 향후 연구 방향

다양한 장치를 통해 프라이버시에 민감한 데이터생성이 빠르게 증가하고 있다. 서비스 공급자는 개인화된 서비스를 공급하기 위해 민감한 데이터에 접근이 필요하지만 이러한 정보의 접근은 용이하지 않다. FL 은 학습모델을 생성하는 클라이언트와의 협력을 통해 전역적인 학습모델을 생성함으로써 민감한 데이터 이동 없이 전체 데이터를 대상으로 학습모델을 생성하는 이점을 제공한다.

이 논문에서는 Flower 프레임워크를 통해 네트워크 통신을 통해 MNIST 의 분류를 진행하기 위해 응용에 점진적 FL 모델 학습을 수행하도록 설계한 후 성능평가를 수행하였다. 점진적 학습을 통해 라운드별 응답시간 분석을 통해 네트워크 대역폭의 영향을 관찰하였으며, 점진적인 FL 에서의 최적화 알고리즘의 수행결과를 제시하였다.

연합학습은 급증하는 데이터의 소유권과 활용의 문제를 해결하는 대안으로서 이종 산업간 데이터 공유과정에서의 데이터 이질성과 신뢰성을 높이기 위해 전이학습, 암호화 알고리즘 및 블록체인 등의 접목이 필요하다.

V. Acknowledgement

이 논문은 2022 년도 정부(산업통상자원부)의 재원으로 한국산업기술평가관리원의 지원을 받아 수행된 연구임(No.20022177, 고위험 현장 안전관리를 위한 에지 단말간 AI 협업 기술 기반의 개방형 안전관리 서비스 개발)

VI. 참고문헌

[1] Alfred Zimmermann, Rainer Schmidt, Lakhmi C. Jain, Architecting the Digital Transformation - Digital Business, Technology, Decision Support, Management. Intelligent Systems Reference Library 188, ISBN 978-3-030-49639-5, Springer, 2021.
[2] Voigt, P.; Bussche, A.v.d. The EU General Data Protection Regulation (GDPR): A Practical Guide; Springer Publishing Company, Incorporated: 2017.
[3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. of the 20th International Conference on Artificial Intelligence and Statistics, 2017, pp. 1273–1282.
[4] Konecˇný, J.; McMahan, H.B.; Ramage, D.; Richtárik, P. Federated Optimization: Distributed Machine Learning for On-Device Intelligence. arXiv 2016, arXiv:1610.02527
[5] Kairouz, Peter, et al. (58 others), “Advances and open problems in federated learning”, Foundations and Trends in Machine Learning 14.1–2, 1-210, 2021. https://ml-ops.org/content/references.html
[6] R. Shokri, and V. Shmatikov, "Privacy-Preserving Deep Learning," in Proc. of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS), 2015, pp. 1310–1321.
[7] Rasheed, M.A.; Uddin, S.; Tanweer, H.A.; Rasheed, M.A.; Ahmed, M.; Murtaza, H. Data privacy issue in Federated Learning Resolution using Block Chain. VFAST Trans. Softw. Eng. 2021, 9, 51–61.
[8] T. Li, A. K. Sahu, A. Talwalkar and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," in IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50-60, May 2020, doi: 10.1109/MSP.2020.2975749.
[9] Flower: A Friendly Federated Learning Framework (https://flower.dev/)
[10] Sergeev, A., & Balso, M.D. (2018). Horovod: fast and easy distributed deep learning in TensorFlow. *ArXiv, abs/1802.05799*.

저자소개

강윤희(*Yunhee Kang*)

1993 년 8 월 동국대학교 대학원 컴퓨터공학과 석사
2002 년 8 월 고려대학교 대학원 컴퓨터과학과 박사
2000 년 3 월~현재 백석대학교 컴퓨터공학부 교수

관심분야 :Distributed System, Artificial Intelligence , Cloud Computing

강명주 (*Myungju Kang*)

2009 년 ~ 2012 년 : 자바정보기술(주) SI 사업본부장
2013 년 ~ 2017 년 : 트리니티(주) 빅데이터사업본부장
2018 년 ~ 현 재 : ㈜넥타르소프트 부사장

관심분야 : BigData, Natural Language Process, Artificial Intelligence