

Paper Review

김진수

Federated Learning

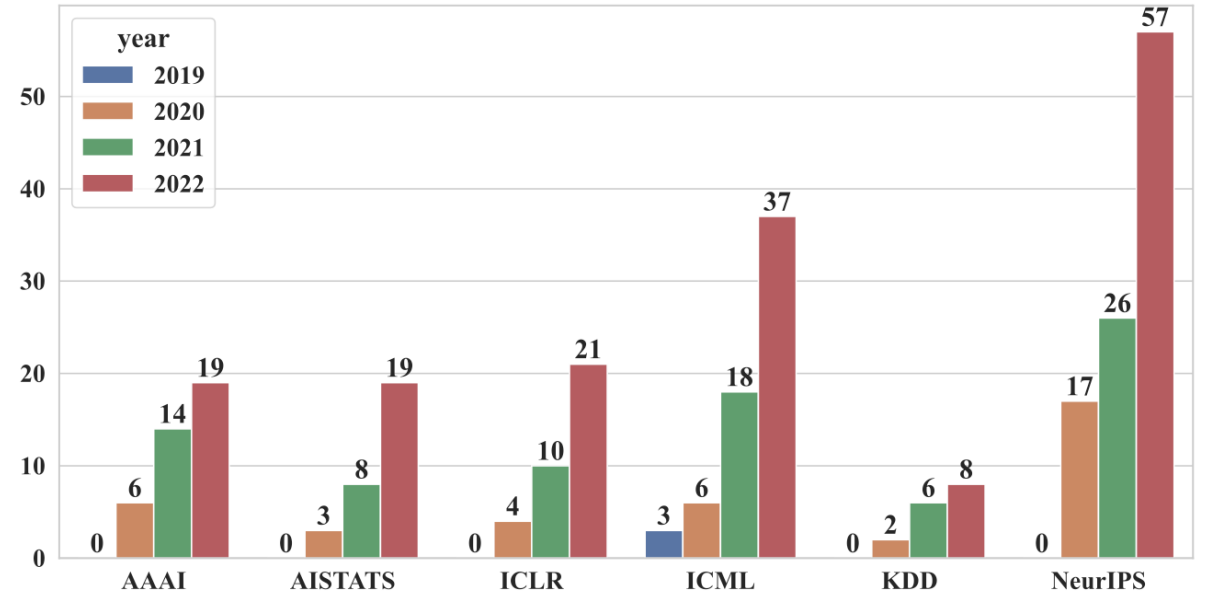
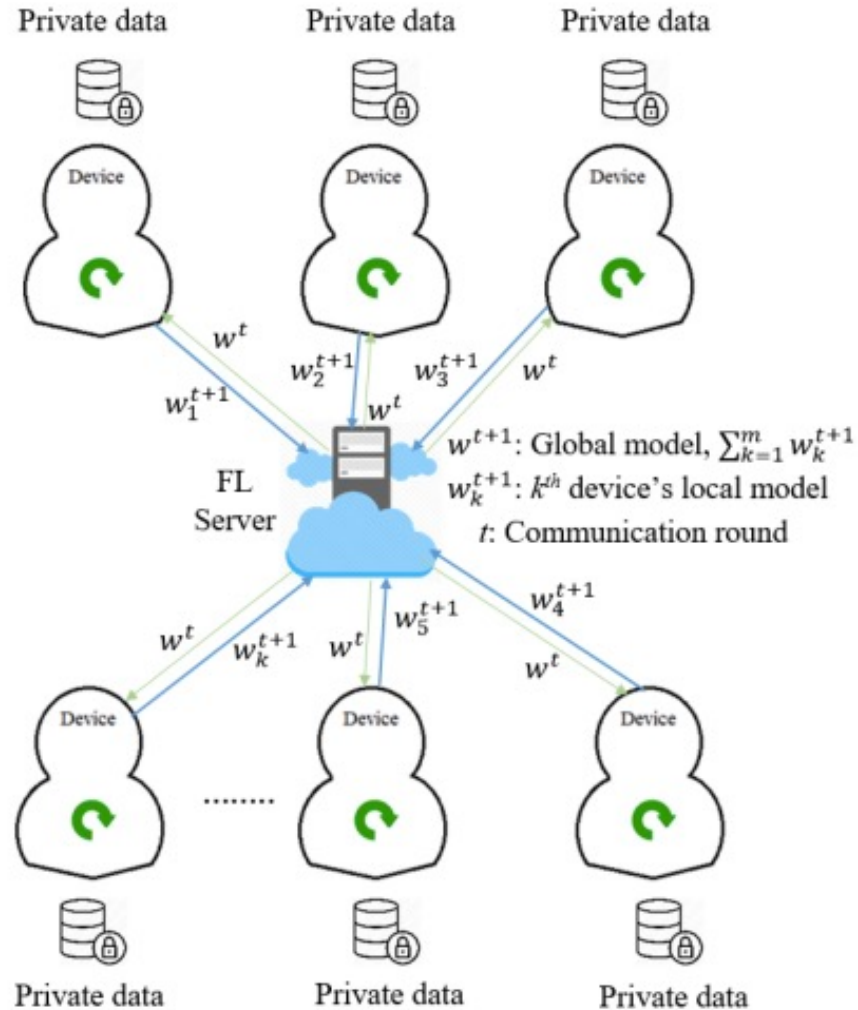
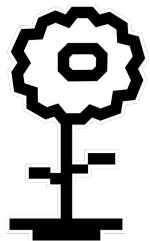


Figure 2: The number of pulished FL papers in top-tie conference from 2019-2022.

Federated Learning

Framework

“손쉽게 적용할 수 있도록 도와줌”



Flower



Benchmark

“검증/비교를 도와줌”



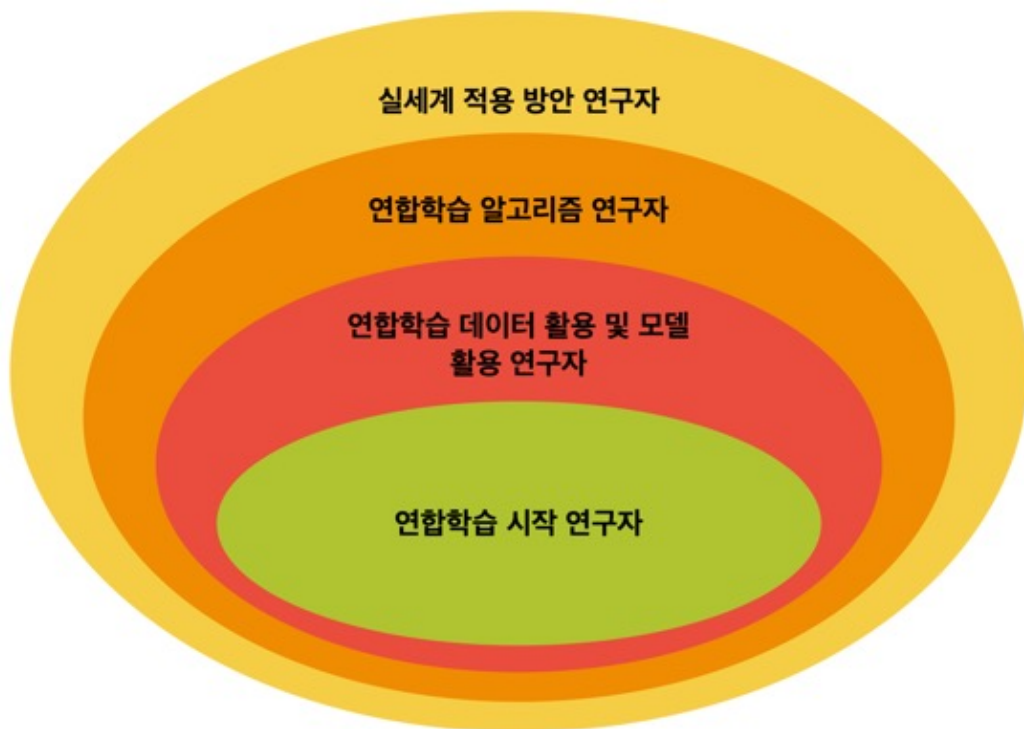
microsoft/**msrflute**

Federated Learning Utilities and Tools for
Experimentation



연합학습 개방형 플랫폼의 발전과 문제점에 대한 체계적 비교 분석

$$\min_{\mathcal{W}} \mathcal{F}(\mathcal{W}), \text{ where } \mathcal{F}(\mathcal{W}) = E[\mathcal{L}(\mathcal{W}, \mathcal{D})] + \lambda * \mathcal{G}(\mathcal{W})$$



Situation 1. 연합학습을 시작하는 연구자

- 연합학습을 실행하는 방법?
- 접근성 높은 연합학습 환경?

Situation 2. 연합학습 데이터 및 모델 연구자

- 데이터셋 유무?
- 커스텀 데이터셋 및 모델 적용?

Situation 3. 연합학습 알고리즘 연구자

- 프레임워크 별 성능차이?
- Regularization term 추가 방법?

Situation 4. 연합학습 실세계 적용 연구자

- 연합학습 클라이언트 관리 및 모니터링?
- Edge 디바이스 연합학습?

Situation 1. 연합학습을 시작하는 연구자

연합학습을 시작하는 연구자는 연합학습을 시도해보는 단계의 상황으로 연합학습 목적함수 최적화에 큰 의미를 가지지 않는다. 연합학습이라는 개념을 실제로 구현하고 실현되는 것을 목표로 한다.

Q1-1.연합학습을 실행하는 간단한 방법은 무엇인가?

Q1-2.접근성 높은 연합학습 환경을 지원하는가?

1. Use PaddleFL in docker

We highly recommend to run PaddleFL in Docker

```
#Pull and run the docker
docker pull paddlepaddle/paddlefl:1.1.2
docker run --name <docker_name> --net=host -it -v $PWD:/paddle <image id> /bin/bash
```

2. Install PaddleFL via installation package

We provide compiled PaddlePaddle and PaddleFL installation packages, you can download and install them directly.

First, install PaddlePaddle

```
#Install PaddlePaddle
wget https://paddlefl.bj.bcebos.com/paddlepaddle-1.8.5-cp*-cp*-linux_x86_64.whl
pip3 install paddlepaddle-1.8.5-cp*-cp*-linux_x86_64.whl
```

	Easy start	Tutorial Doc	Model Library
Flower	○	○	○
FATE	✓	○	○
FedML	○	○	✓
PaddleFL	○	○	X
TFF	✓	○	✓
EasyFL	○	✓	○
FedBench	✓	X	✓

(표 1) 연합학습 초기 연구자를 위한 연합학습 프레임워크 비교 분석.
(Table 1) Comparative analysis of federated learning frameworks.
*✓는 제한적 지원을 의미

Situation 2. 연합학습 데이터 및 모델 연구자

이 상황의 연구자들은 연합학습을 적용하여 로컬 업데이트와 글로벌 업데이트의 기댓값을 향상에 집중한다.

Q2-1.데이터가 존재하지 않을 경우 데이터셋을 제공 하는가?

Q2-2.커스텀 데이터 및 모델을 쉽게 적용할 수 있는가?

Params

client_number

train_data_num

test_data_num

train_data_global

test_data_global

train_data_local_num_dict

train_data_local_dict

test_data_local_dict

class_num

```
logger.info("load_data. dataset_name = %s" % dataset_name)
(
    client_num,
    train_data_num,
    test_data_num,
    train_data_global,
    test_data_global,
    train_data_local_num_dict,
    train_data_local_dict,
    test_data_local_dict,
    class_num,
) = fedml.data.load(
    args.batch_size,
    train_path=args.data_cache_dir + "/MNIST/train",
    test_path=args.data_cache_dir + "/MNIST/test",
)
```

Framework	Sample Dataset	Data Partition	Custom format
Flower	2	O	informal
FedML	17	O	formal
FedScale	20	X	formal
FedBench	5	X	informal
Fate	13	X	formal
Leaf	6	O	informal
PaddleFL	10	X	formal
EasyFL	3	O	formal

(표 2) 연합학습 프레임워크 데이터 샘플 및 파티션, 커스텀 형식 비교 표
(Table 2) FL Framework Data Sample and Partition, Custom Format Comparison Table

Situation 3. 연합학습 데이터 및 모델 연구자

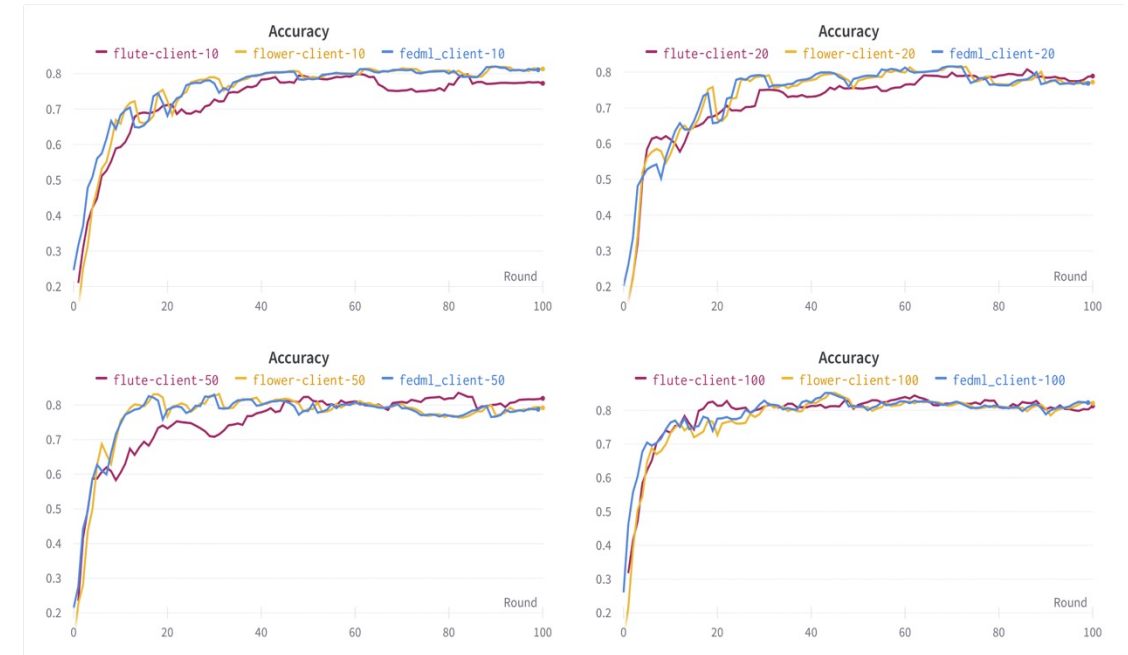
클라이언트의 업데이트 방식을 개선하거나 서버의 모델 통합 과정에서 가중치를 추가하여 성능을 향상시키는 것이 목표이다.

Q3-1. 프레임 워크 별 성능 차이가 존재하는가?

Q3-2. 집계 최적화, 개인화, 클라이언트 선택 등 규제 조건을 추가하기 용이한가?

Frame work	Example of Aggregation Optimization	Client Selection Module
Flower	O	O
FedML	O	X
FedScale	O	O
FedBench	O	O
Fate	X	X
Leaf	X	O
FLUTE	X	O
PaddleFL	X	X
EasyFL	X	O

(표 4) 연합학습 프레임워크의 집계 최적화 예시 및 클라이언트 선택 모듈화 비교 표



(그림 3) 시뮬레이션 환경에서 연합학습 프레임워크 성능 비교

Situation 4. 연합학습 실세계 적용 연구자

자신이 앞의 과정에서 겪은 경험과 기반 지식을 통해서 실제 연합학습 목적함수의 최적화를 달성하고 이 모델을 실 세계에 적용하고 테스트하는 것이 목표이다.

Q4-1.연합학습의 클라이언트를 관리, 모니터링 할 수 있는 기능이 존재하는가?

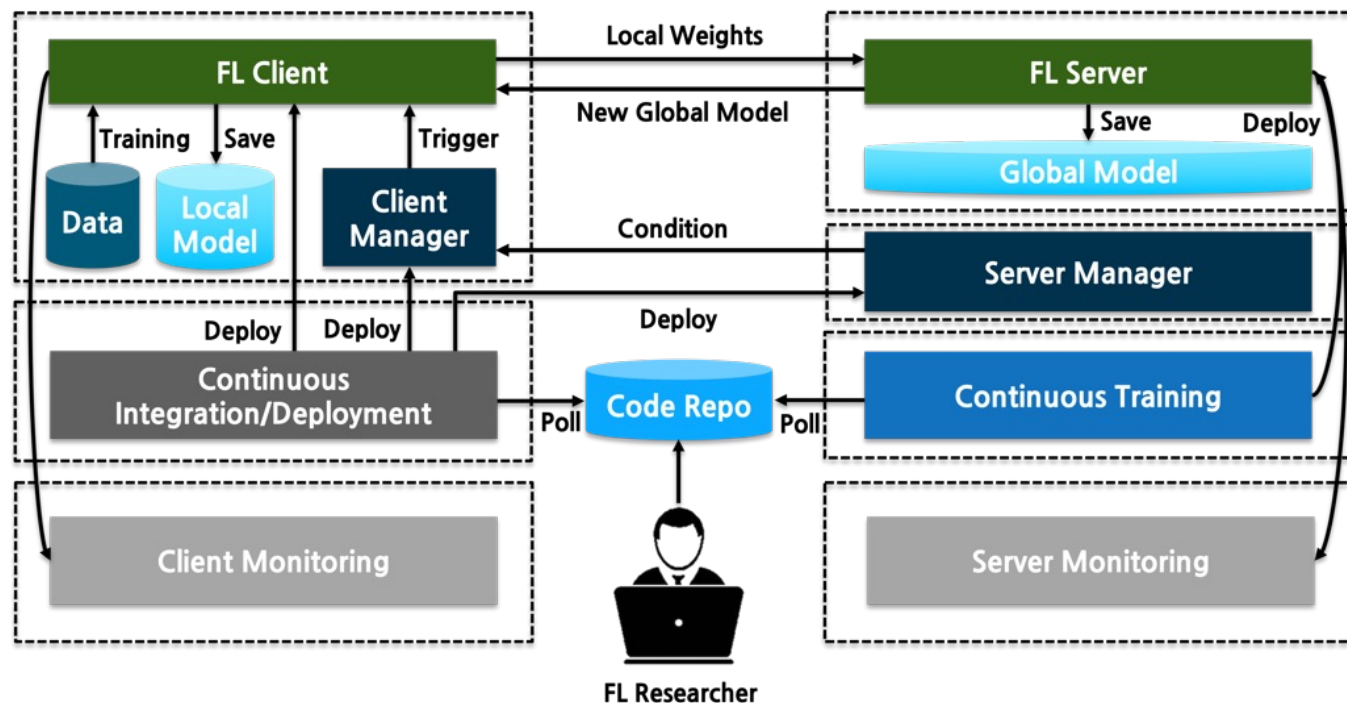
Q4-2.Edge 디바이스에서 연합학습이 가능한가?

Framework	Data Preprocessing	Client ID	Monitoring	Status Check	Mobile Example
Flower	○	○	X	X	○
FedML	○	○	○	○	○
PaddleFL	○	○	X	X	X
FATE	○	○	○	○	X
EasyFL	○	X	X	X	X
TFF	○	○	○	X	X

(표 5) 연합학습 프레임워크의 클라이언트 관리, 모니터링, 상태 확인, 모바일 환경 지원 비교 표

연합학습 프레임워크 한계 및 실세계 적용 방안

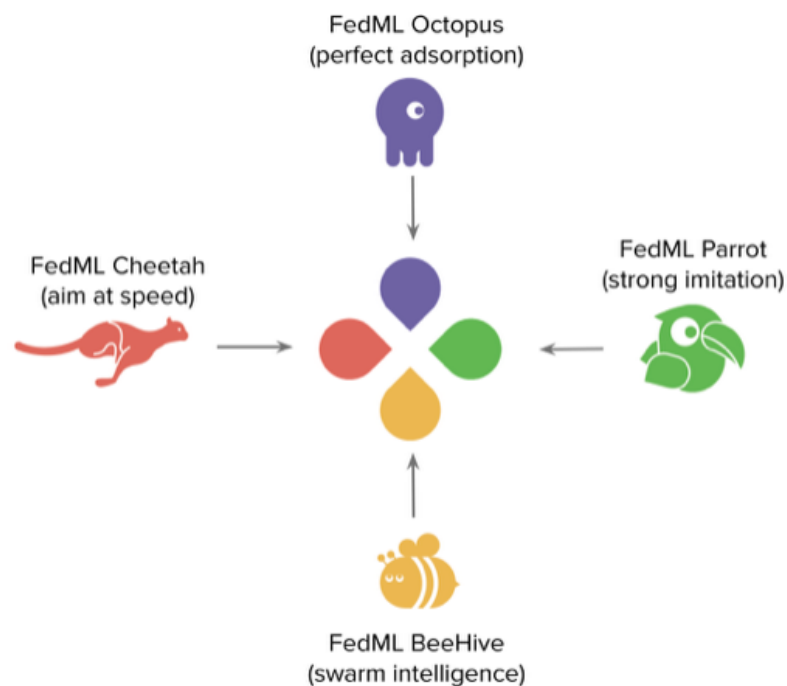
- 다양한 프레임워크가 존재하지만 여러 분야와 다양한 상황에서 통합적으로 사용되는 프레임워크는 없다.
- 사용자 친화적인 UI/UX를 구축하고 다수의 클라이언트에 적합한 운영 관리가 필요하다.
- 실세계 연합학습 적용을 위한 시나리오 정의와 실제 산업 및 연구에서 발생하는 실시간 데이터를 처리할 수 있는 파이프라인이 구축되어야 한다.



(그림 4) 실세계에서 연합학습 운영을 위한 연합학습 수명주기관리 플랫폼 구조

FedML

FedML Feature Overview



- **FedML Cheetah** - Accelerate Model Training with User-friendly Distributed Training
- **FedML Parrot** - Simulating federated learning in the real world (1) simulate FL using a single process (2) MPI-based FL Simulator (3) NCCL-based FL Simulator (fastest)
- **FedML Octopus** - Cross-silo Federated Learning for cross-organization/account training, including Python-based edge SDK.
- **FedML Beehive** - Cross-device Federated Learning for Smartphones and IoTs, including edge SDK for Android/iOS and embedded Linux.
- **FedML MLOps**: FedML's machine learning operation pipeline for AI running anywhere at any scale.
- **Model Serving**: we focus on providing a better user experience for edge AI.

FedML Install

```
git clone https://github.com/FedML-AI/FedML.git
```

```
conda create -n fedml python=3.9
```

```
conda activate fedml
```

```
conda install mpi4py openmpi
```

```
pip install fedml
```