An Empirical Measurement of Client Contribution for Federated Learning using Data Size Diversification

Sung Kuk Shyn, Donghee Kim, and Kwangsu Kim

IEEE Access

2022.10.28 가천대학교

Sung Kuk Shyn (신성국)

Master's Student

성균관대학교 인공지능융합연구실

Research Interest

Federated Learning, Explainable AI, AI Applications in Social Science

Education

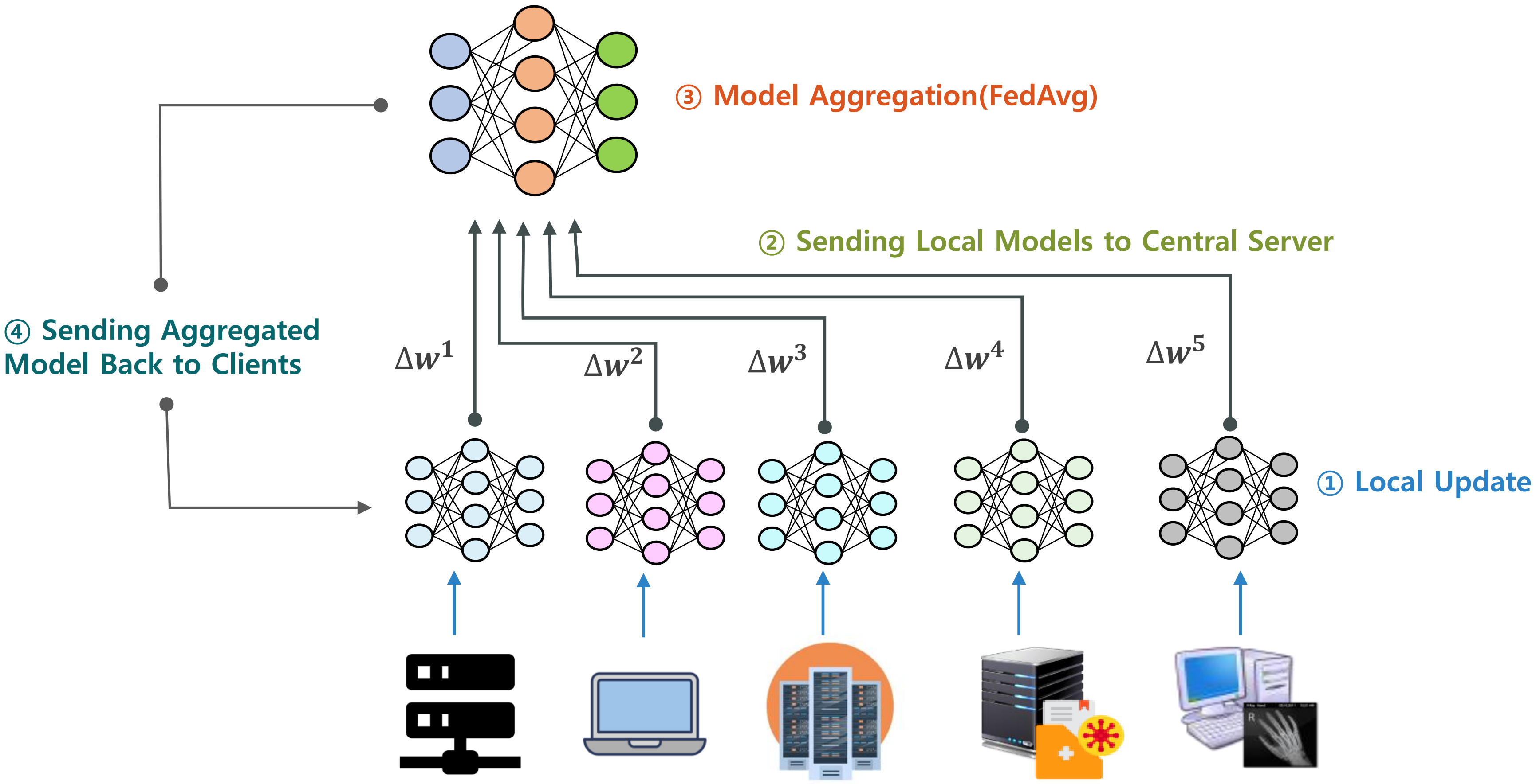
B.A. in Economics, SKKU
B.S. in Computer Science, SKKU

Publication

An Empirical Measurement of Client Contribution for Federated Learning with Data Size Diversification, 2022 IEEE Access

Work in Progress

Similarity-based Regulation of Federated Learning for Mitigating Client Drift and Global Model Overfitting
(with Dongwon Kim, Donghee Kim, Kwangsu Kim)



Federated Learning

Aggregating updated models of clients in distributed environments without accessing private data.

Main Goals of Federated Learning

HIGH PERFORMANCE

REACH THE PERFORMANCE CLOSE TO THE CENTRALIZED MODEL.

EFFICIENT LEARNING

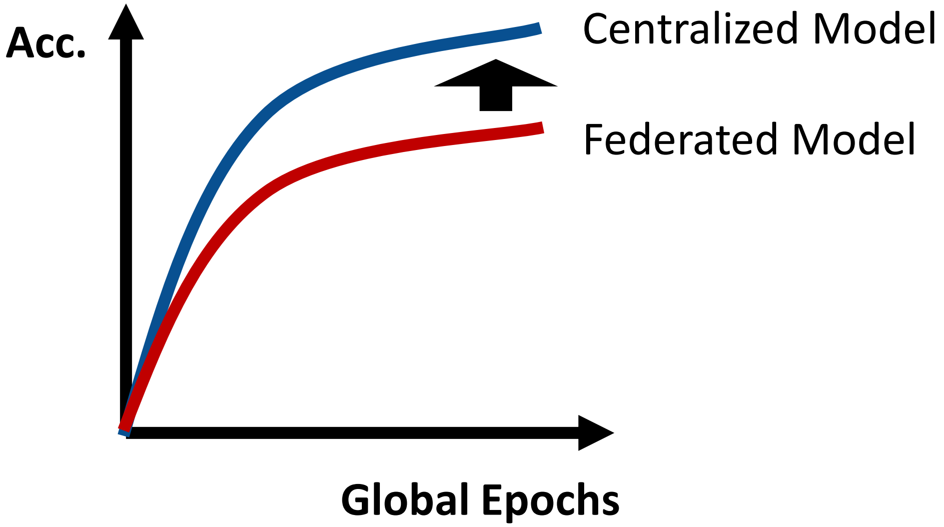
CREATE COMMUNICATION COSTS AND COMPUTATION COSTS AS EFFICIENTLY AS POSSIBLE.

PRIVACY PROTECTION

MUST PRESERVE LOCAL PRIVACY AND CONFIDENTIALITY.

HIGH PERFORMANCE

REACH THE PERFORMANCE CLOSE TO THE CENTRALIZED MODEL.



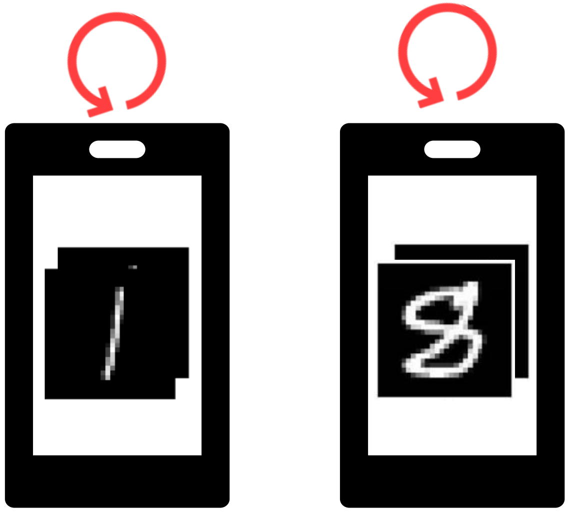
The main bottleneck for achieving high performance is **data heterogeneity**

Main Bottleneck: Data Heterogeneity

Data Distribution

- Not Independently and Identically Distributed (non-IID)
- Provoke **Client Drift Problem**

Clients containing a very different dataset from each other



Client 1 **Client 2**

SCAFFOLD, S. P. Karimireddy, et. al, ICML'20

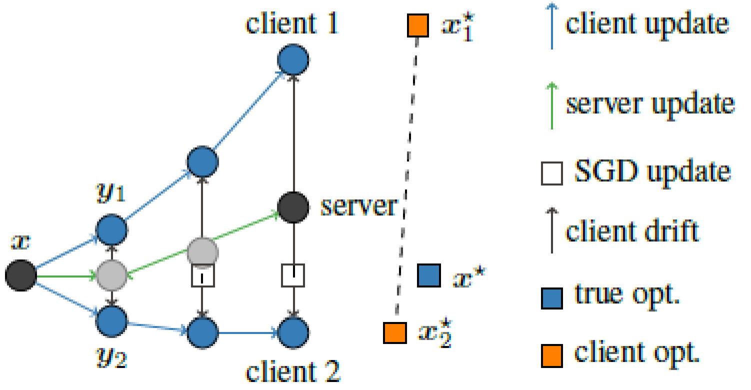


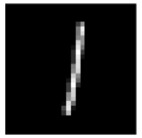
Figure 1. Client-drift in FEDAVG is illustrated for 2 clients with 3 local steps ($N = 2$, $K = 3$). The local updates y_i (in blue) move towards the individual client optima $x_i^\star$ (orange square). The server updates (in red) move towards $\frac{1}{N}\sum_i x_i^\star$ instead of to the true optimum $x^\star$ (black square).

Main Bottleneck: Data Heterogeneity

Data Noise

- Situations where **poisoning attacks(backdoor attacks)** have occurred in some clients

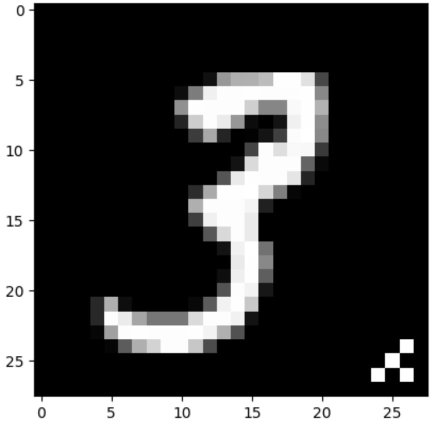
Image **Label**



→ 5

→ 0

Noise to labels

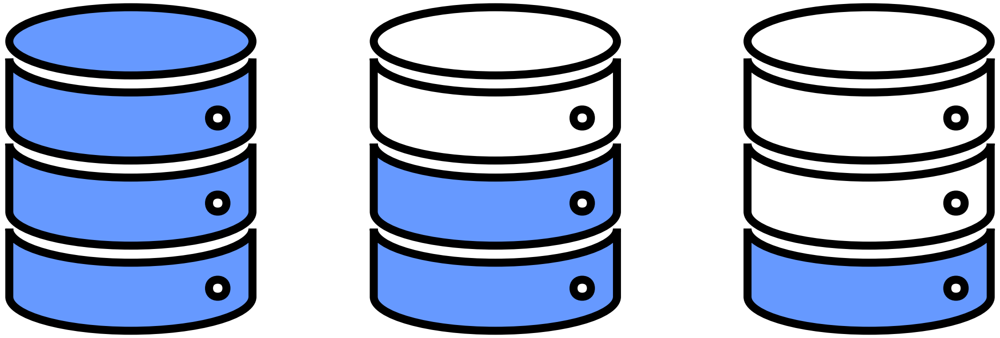


Noise to images

Main Bottleneck: Data Heterogeneity

Data Size Imbalance

- Not Independently and Identically Distributed(non-IID) **in Size**



Federated Learning should not treat clients **equally with different data sizes**.

Approaches to high performance

FL Aggregation Optimization

- **Develop the federated averaging(FedAVG) algorithm**
- FedProx(T. Li, et al., MLSys'20), SCAFFOLD(S. P. Karimireddy, et. al, ICML'20)

Personalized Optimization

- **Change the learning algorithms during the local training stage**
- MOON(Q. Li, et. al, CVPR'21), FedBABU(J. Oh, et. al, ICLR'22), Classifier Calibration(M. Luo, et. Al, NeurIPS'21)

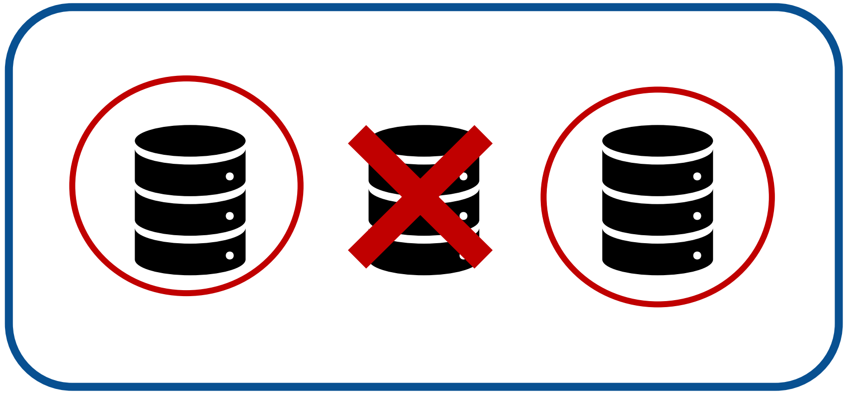
Client-level Optimization

- **Focus on clients by evaluating whether they are good clients or bad clients.**
- Client Contribution Evaluation/Client Selection/Adversarial Attack Detection/Incentive Mechanism

Client-level Optimization : Client Contribution Evaluation

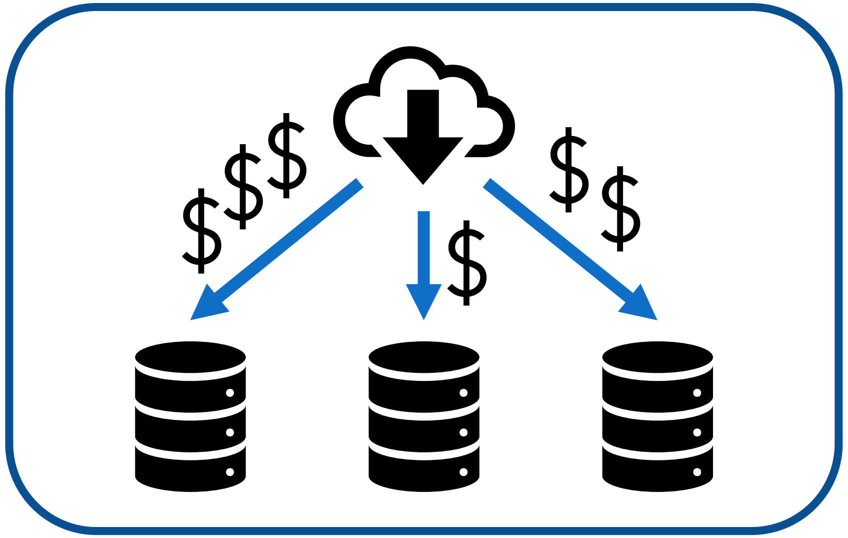
Client Contribution

Is a client good or bad?



Client Selection

How much does a client contribute?



Incentive Allocation

Client-level Optimization : Client Contribution Evaluation

How do we evaluate the client contribution in the federated learning setting?

Measure the Impact of Data Heterogeneity

- **Measure Data Distribution Difference**
- **Measure Data Noise Proportion**
- **Measure Data Size**

Client-level Optimization : Client Contribution Evaluation

Limitation 1 : Local Information Inaccessibility

FedAVG

local data size

$$\theta_r^G = \sum_{i=1}^{n} \frac{d_{r,s}^{(i)}}{\sum_j d_{r,s}^{(j)}} \times \theta_r^{(i)}$$

global model weights/gradients

local model weights/gradients

$\theta_r^{(i)}$

local model weights/gradients

Information We Have

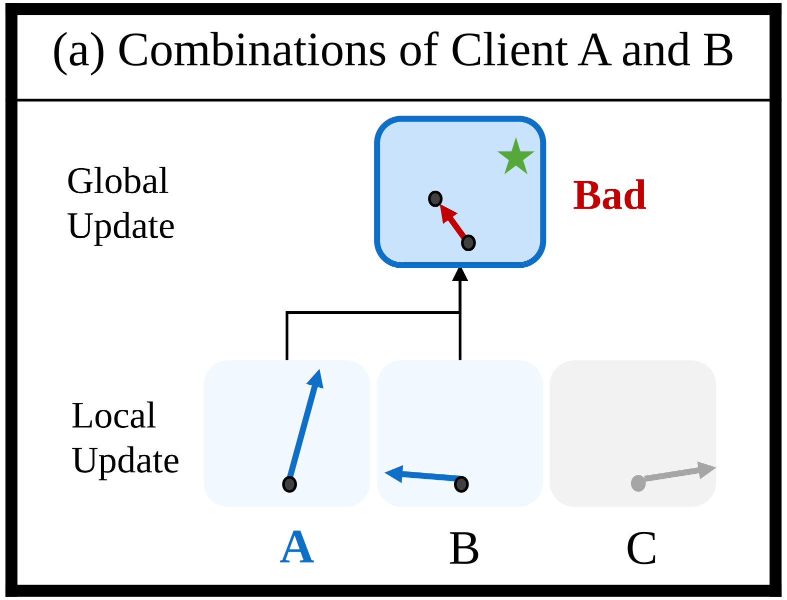
$d_{r,s}^{(i)}$

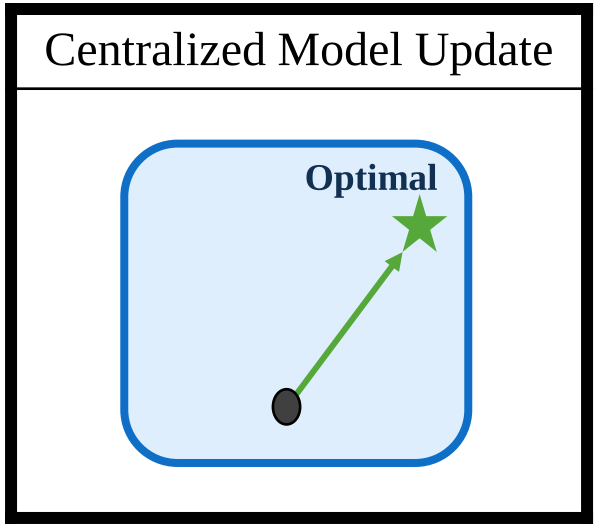
local data size

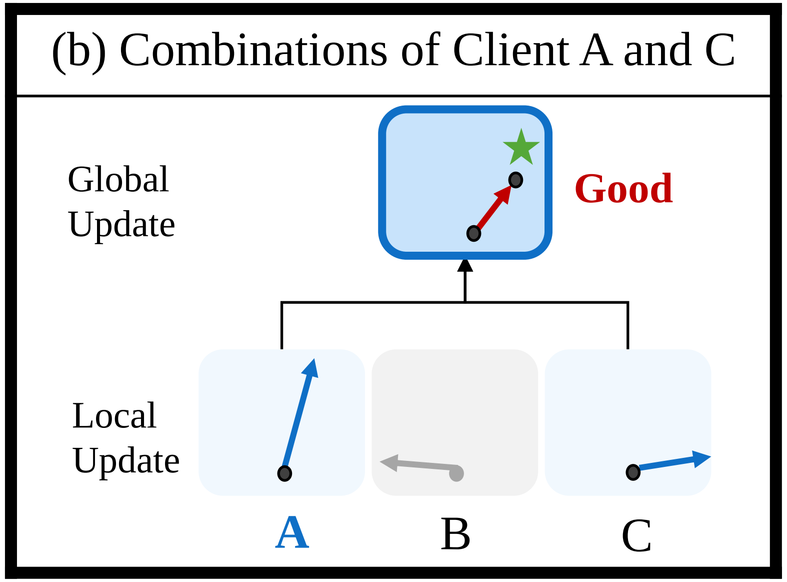
The other local information cannot be observed unless we access the local dataset.
ex) Data Distribution, Data Noise Proportion

Client-level Optimization : Client Contribution Evaluation

Limitation 2 : Unclear Combinatorial Impact of Data Heterogeneity







Is Client A Good or Bad?

Existing Methods for Client Contribution Evaluation

Client Contribution Evaluation in Federated Learning

Game-Theoretic Methods

- **Leave-one-out**
  - G. Wang, et. al, IEEE *BigData'*19
- **Shapley Value**
  - T. Wang, et. al, *Federated Learning, Springer, 2020*

FL-specific Methods

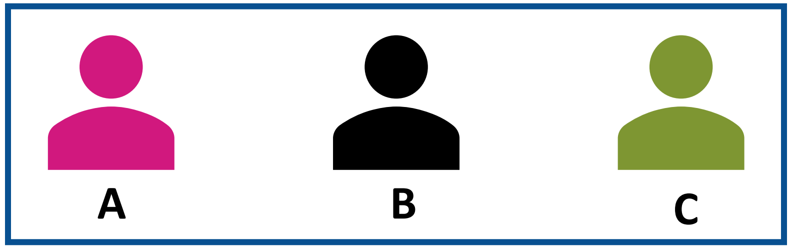
- **Data Quantity**
  - Y. Zhan, et. al, IEEE *Internet of Things Journal*, 2020
- **Weight Difference**
  - J. Zhang, et. al, *WWW*'21
- **Deep Reinforcement Learning**
  - J. Zhao, *ICASSP*'21

Game Theoretic Methods

Pretend that the clients are playing a game.

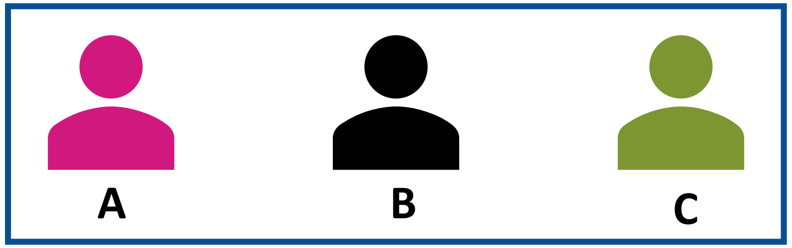
Is Federated Learning
Competition Game or Cooperation Game?

Team Project



Reward : $300

$$v(\mathrm{A}) = 4, \quad v(\mathrm{B}) = v(\mathrm{C}) = 5, \quad v(\mathrm{A},\mathrm{B}) = v(\mathrm{A},\mathrm{C}) = 12, \quad v(\mathrm{B},\mathrm{C}) = 9, \quad v(\mathrm{A},\mathrm{B},\mathrm{C}) = 15$$



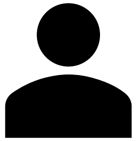
$$v(\mathrm{A}) = 4, \quad v(\mathrm{B}) = v(\mathrm{C}) = 5, \quad v(\mathrm{A,B}) = v(\mathrm{A,C}) = 12, \quad v(\mathrm{B,C}) = 9, \quad v(\mathrm{A,B,C}) = 15$$

Leave-one-out

Exclude the main person I want to know.

A

$v(\mathrm{A,B,C}) - v(\mathrm{B,C}) = \mathbf{6}$



B

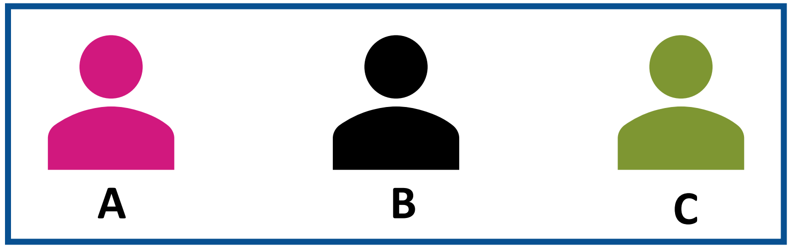
$v(\mathrm{A,B,C}) - v(\mathrm{A,C}) = \mathbf{3}$

C

$v(\mathrm{A,B,C}) - v(\mathrm{A,B}) = \mathbf{3}$

The team will share the reward by **2:1:1**.

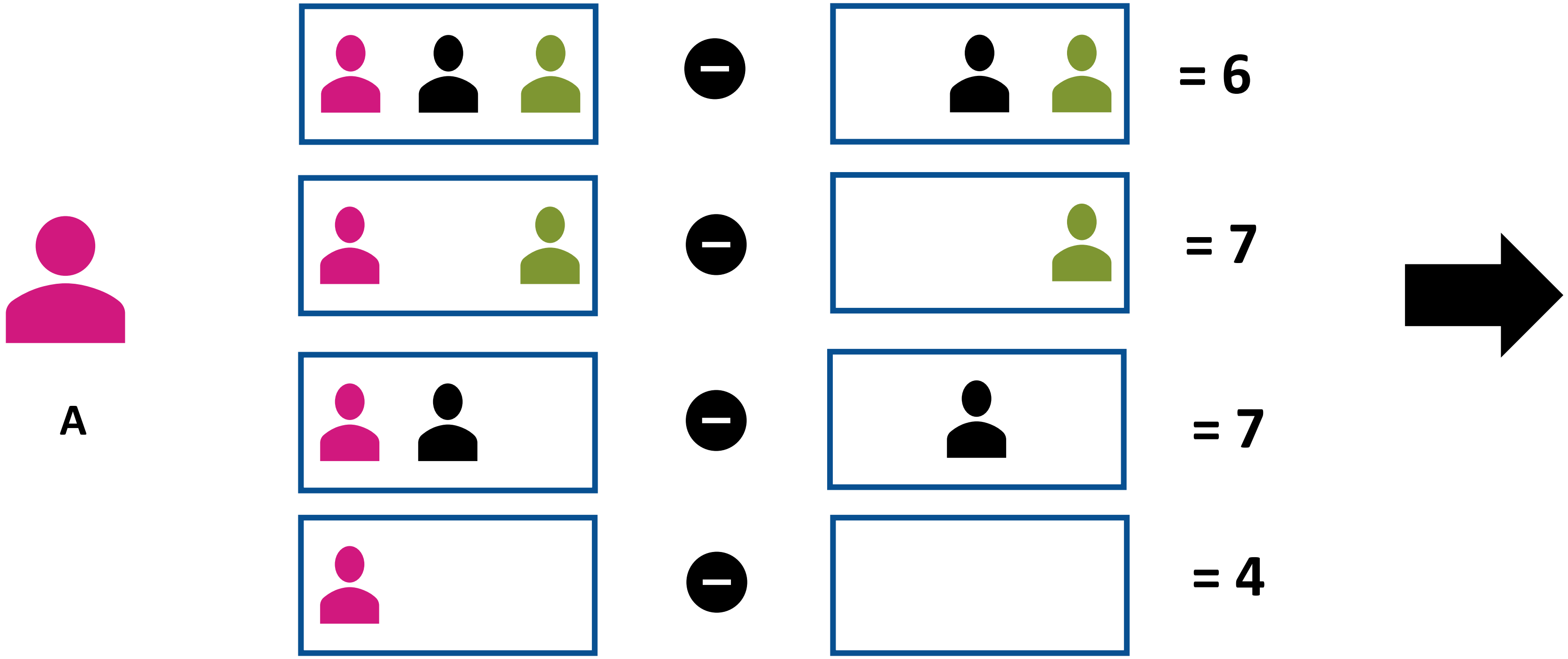
Who thinks that it is unfair?



$v(\mathrm{A}) = 4, \quad v(\mathrm{B}) = v(\mathrm{C}) = 5, \quad v(\mathrm{A,B}) = v(\mathrm{A,C}) = 12, \quad v(\mathrm{B,C}) = 9, \quad v(\mathrm{A,B,C}) = 15$

Shapley Value

Consider every synergy of sub-sets of the main person



Weighted Average

5.67

A **B** **C**

$v(\mathrm{A}) = 4, \quad v(\mathrm{B}) = v(\mathrm{C}) = 5, \quad v(\mathrm{A,B}) = v(\mathrm{A,C}) = 12, \quad v(\mathrm{B,C}) = 9, \quad v(\mathrm{A,B,C}) = 15$

Shapley Value

Consider every synergy of sub-sets of the main person

$$SV_i = \frac{1}{N!} \sum_{S \subseteq N \setminus \{i\}} |S|!\,(|N| - |S| - 1)!\,(v(S \cup \{i\}) - v(S))$$

L.S. Shapley, A Value of n-person Games, Classics in Game Theory, 1997

A

Contribution : **5.67**

B

Contribution : **4.67**

C

Contribution : **4.67**

The team will share the reward by **5.67:4.67:4.67**.

Client Contribution Evaluation - Game Theoretic Methods

Federated LOO

G. Wang, et. al, Measure Contribution of Participants in Federated Learning IEEE *BigData*'19

Measure the **marginal performance of a single client's participation** (with all other clients participating) in every round.

Algorithm 1 Approximating influence estimation for each party in horizontal FML

Input
number of parties K, model f
instance subsets $D_1,\ldots,D_K$
Output
Influence measure Influence^{-D_k} for $k=1,\ldots,K$
for all k=1,...,K **do**
delete D_k from training dataset
retrain model f'
compute $\text{Influence}^{-D_k} = \frac{1}{n}\sum_j |\hat{y}_j - \hat{y}_j^{-D}|$
end for
return Influence^{-D_k} for $k=1,\ldots,K$

Use local weights/gradients

Limitations : No considerations of combinatorial effects

Advanced Research : *Fed-Influence in Accuracy(FIA) & Fed-Influence in Loss(FIL)*

Y. Xue, et. al, Toward Understanding the Influence of Individual clients in Federated Learning, AAAI'21

Client Contribution Evaluation - Game Theoretic Methods

Federated SV

T. Wang, et. al, A Principled Approach to Data Valuation for Federated Learning, *Federated Learning*, 2020

Measure the **weighted mean of the marginal performance** difference of all possible subsets with a single client's participation.

$$s_t^{\nu}(i) = \frac{1}{|I_t|} \sum_{S \subseteq I_t \setminus \{i\}} \frac{1}{\binom{|I_t|-1}{|S|}} \left[\nu(I_{1:t-1} + (S \cup \{i\})) - \nu(I_{1:t-1} + S)\right] \; \mathit{if}\ i \in I_t \quad (3)$$

Use local weights/gradients

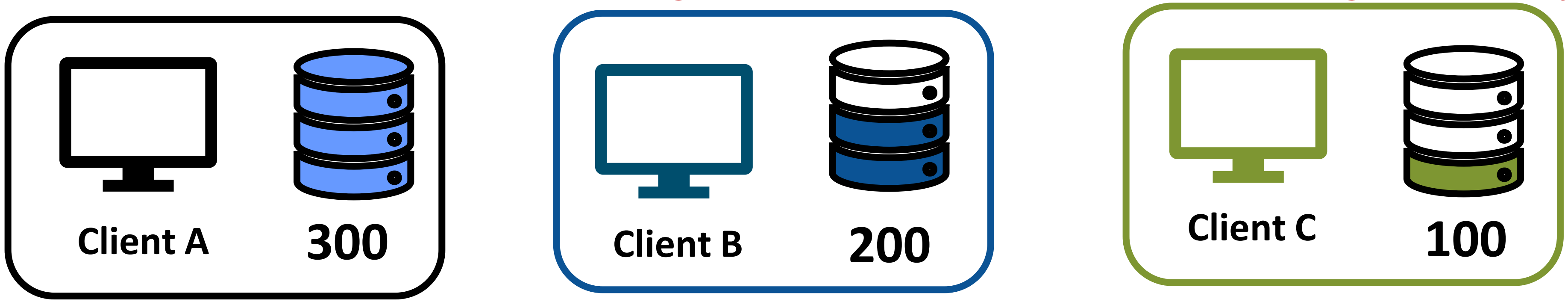
Limitations : Massive Computation Costs($O(2^n)$) when the number of clients is huge.
Unstable Measurement of combinatorial effects **(our focus)**

Advanced Research : *Monte-Carlo SV, gradient-based SV, One- & Multi-Round SV, GTG-Shapley, etc.*

Client Contribution Evaluation - FL-specific Methods

Data Quantity(Data Size)

Y. Zhan, et. al, A Learning-based Incentive Mechanism for Federated Learning, IEEE *Internet of Things Journal*, 2020



Use local data size

Contribution **3/6** **2/6** **1/6**

A simple calculation only for easy incentive allocation.

Assumption: All clients are in IID condition.

Limitations: No Considerations of **other data heterogeneity conditions.**

Client Contribution Evaluation - FL-specific Methods

J. Zhang, et. al, Incentive Mechanism for Horizontal Federated Learning based on Reputation and Reverse Auction, *WWW'21*

Weight Difference

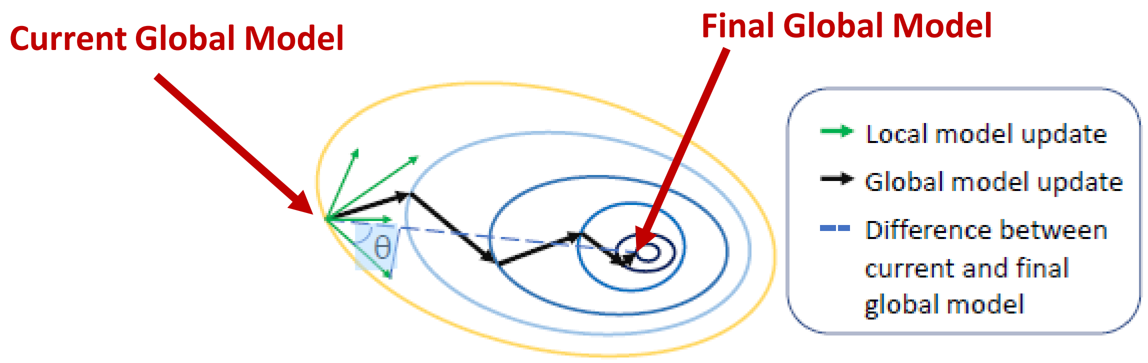


Figure 3: Contribution Measurement

Use local weights

Inner Angle ⬆ **Contribution** ⬇

Measured by cosine similarity between the local update and the total global update.

Limitations: Does not consider **the case that clients are partly excluded**. Lose combinatorial effects.

Client Contribution Evaluation - FL-specific Methods

J. Zhao, et. al, Efficient Client Contribution Evaluation for Horizontal Federated Learning, *ICASSP*'21

Deep Reinforcement Learning

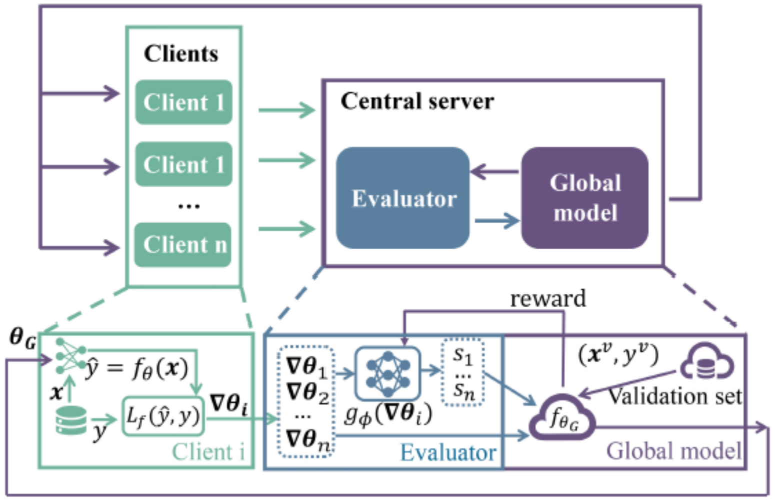


Fig. 1. The block diagram of F-RCCE method.

Use local gradients

RL ingredients

State: A **local gradients set** after local update($\nabla\theta_i$)
Action: Client Selection(s_i) (equals to client contribution)
Reward: Performance of the global model

Use REINFORCE model to find out the best strategy for client selection

Limitations: Needs **numerous simulations** to make the DRL model precise.

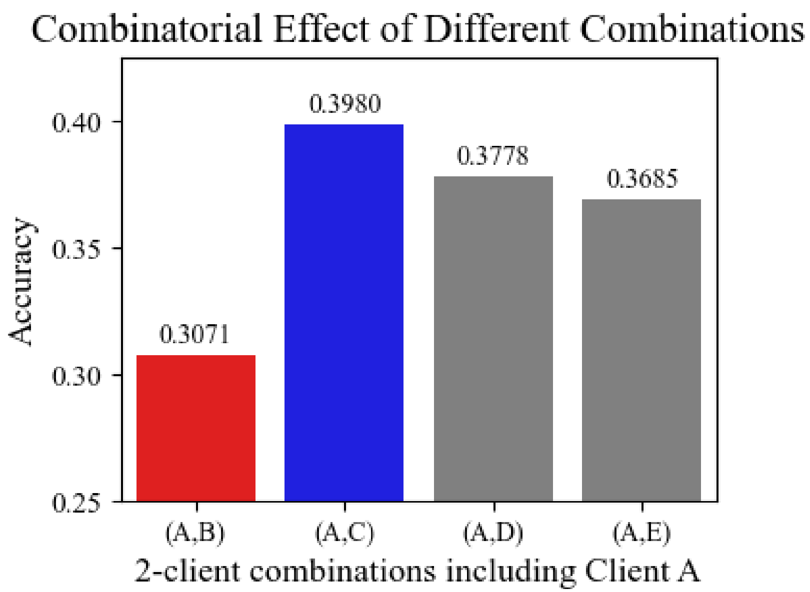
Client Contribution Evaluation - Summary

Information Used	Keyword	Literature	Description	Comparison
Local Weights/ Local Gradients	Leave-one-out	[14], [25]	Measure the marginal performance difference of a specific client's participation.	**Preciseness - , Computation Cost -**
	Shapley Value	[8], [21], [26], [27]	Measure the weighted mean of the marginal performance difference of all possible subsets with a specific client's participation.	**Preciseness ↑ , Computation Cost ↑**
	Weight Difference	[28]	Use client contribution based on the directional difference of local weights/gradients for incentive allocation.	**Preciseness - , Computation Cost ↓**
	DRL Models	[29], [30]	Empirically predict each client contribution using REINFORCE or DQN models with local weights/gradients.	**Preciseness ? , Computation Cost ?**
Local Data Size	Data Quantity	[12]	Simply define a local data size as a total value of each local dataset for incentive allocation.	**Preciseness ↓ , Computation Cost ↓**
	FedCCEA	Ours	Empirically predict an averaged impact of each local dataset using deep learning models with diverse cases of data size.	-

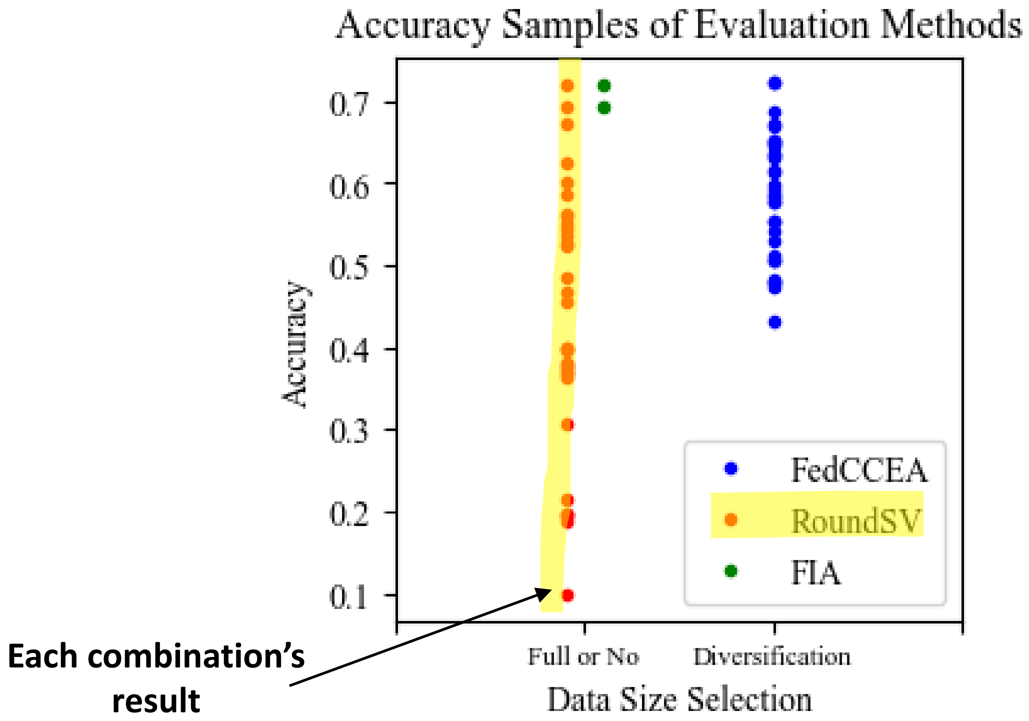
Question: Is SV-based Client Contribution precise?

Close-up to Shapley Value

Noisy client A's combinatorial effect



Noisy client A's accuracy variation with all combinations



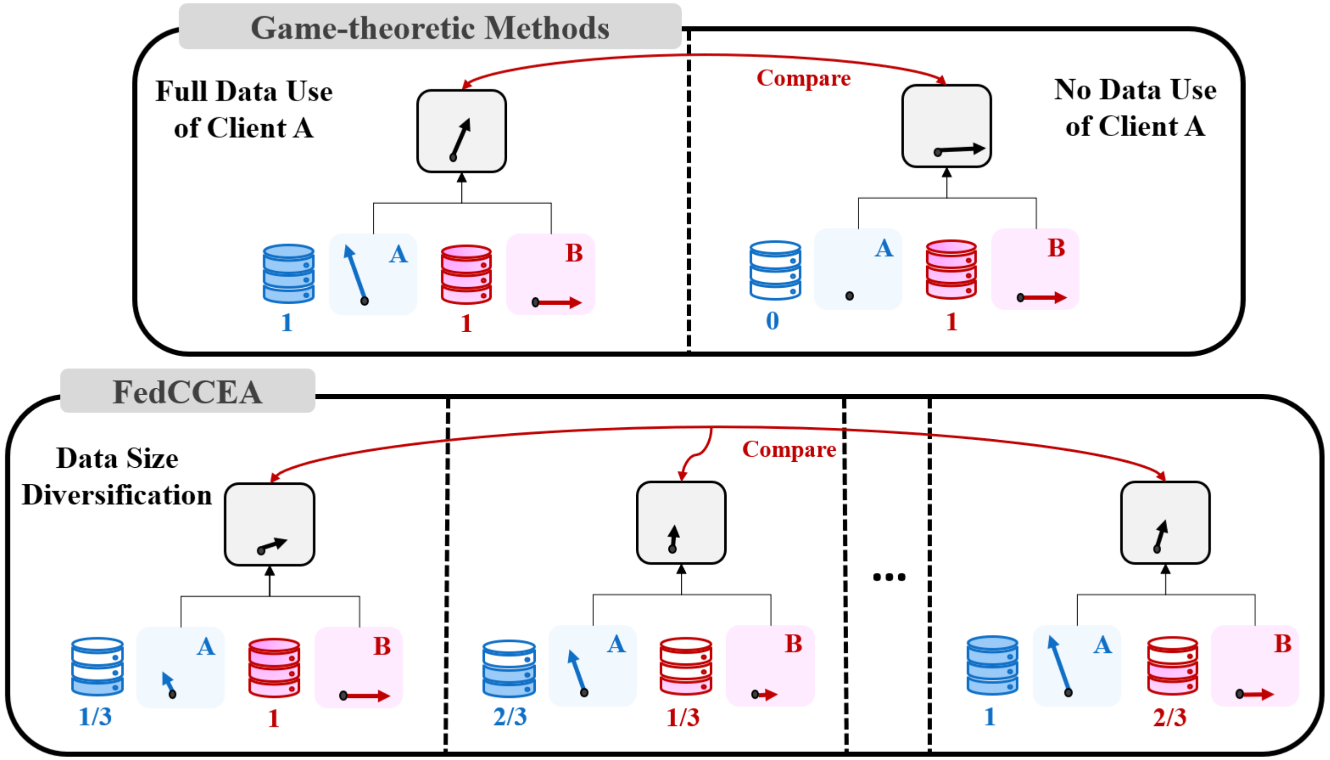
Shapley-based evaluation causes **unstable measurement** when data heterogeneity is high.

Proposed Method

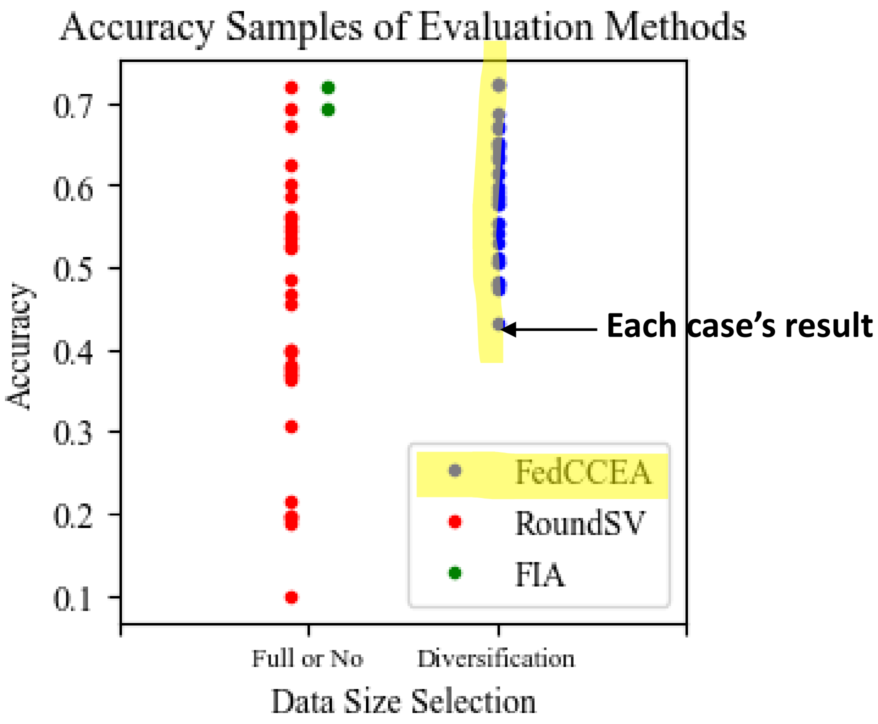
Federated Client Contribution Evaluation through Accuracy approximation(FedCCEA)

FedCCEA Descriptions

Data Size Diversification

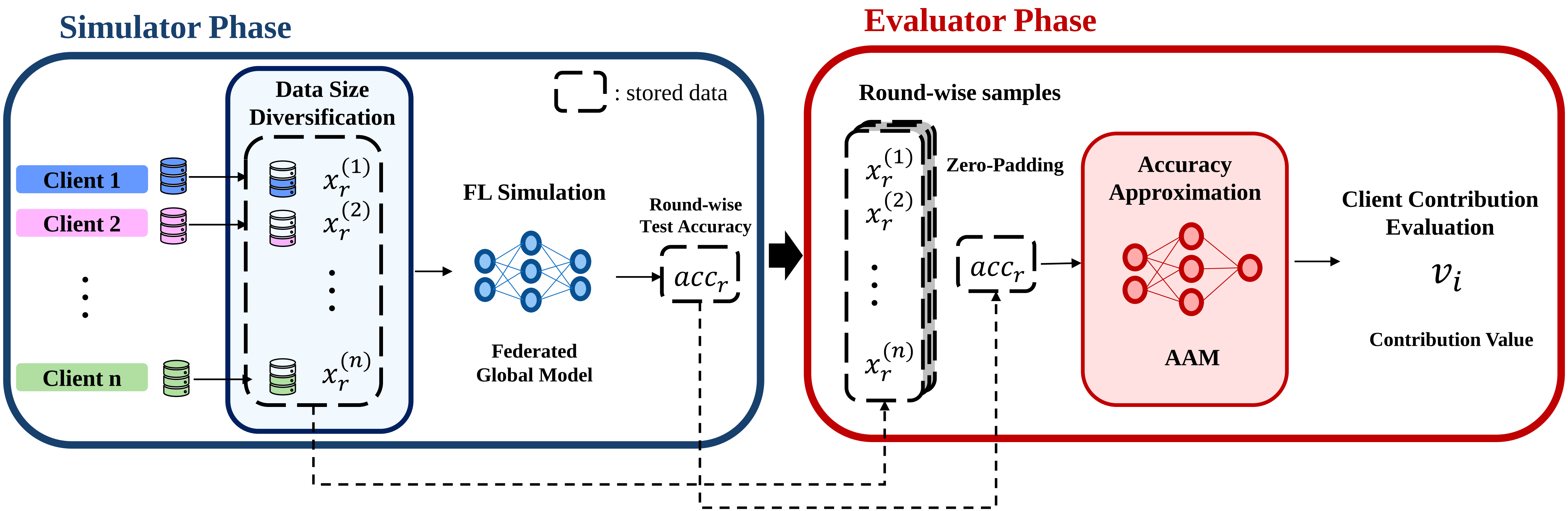


Result of Data Size Diversification

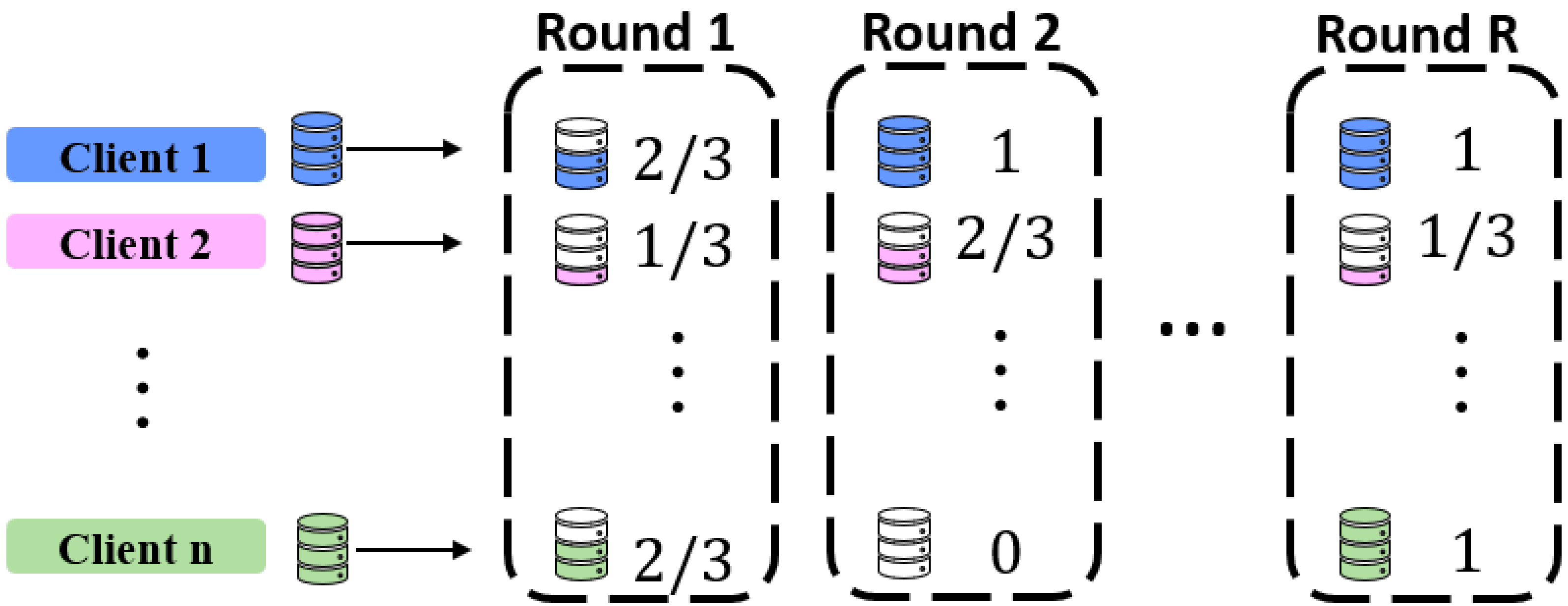


- Allows **partial participation** by **data size diversification**.
- With data size samples, we build an **accuracy approximation deep learning model**.
- Empirically predict **the impact of data heterogeneity** within the approximation model.

Overall Framework of FedCCEA



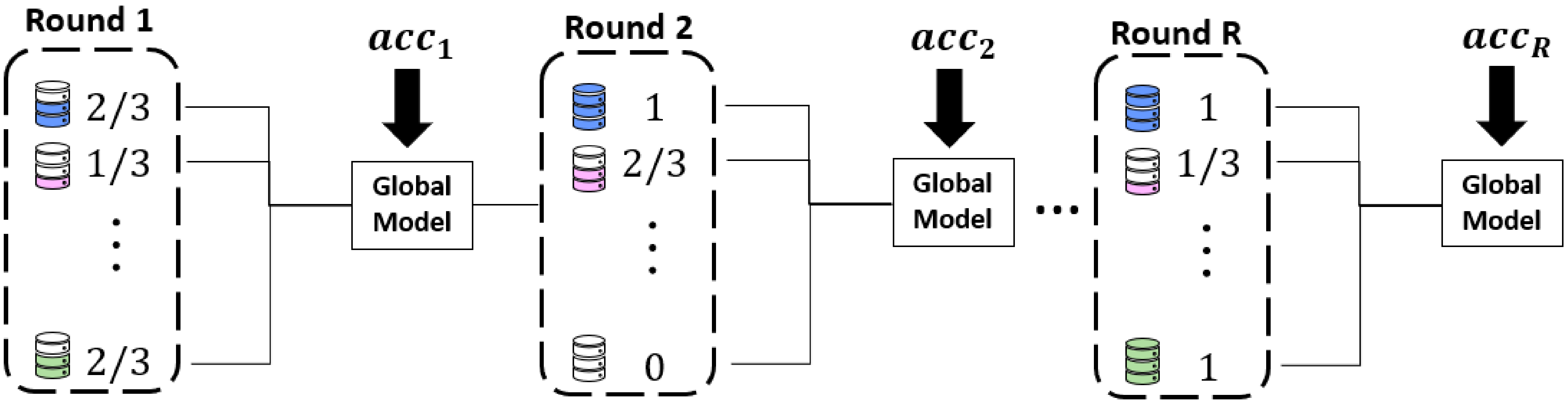
FedCCEA: 1. Simulator



Data Size Diversification

Each client **randomly selects** the proportion of data size in every round.

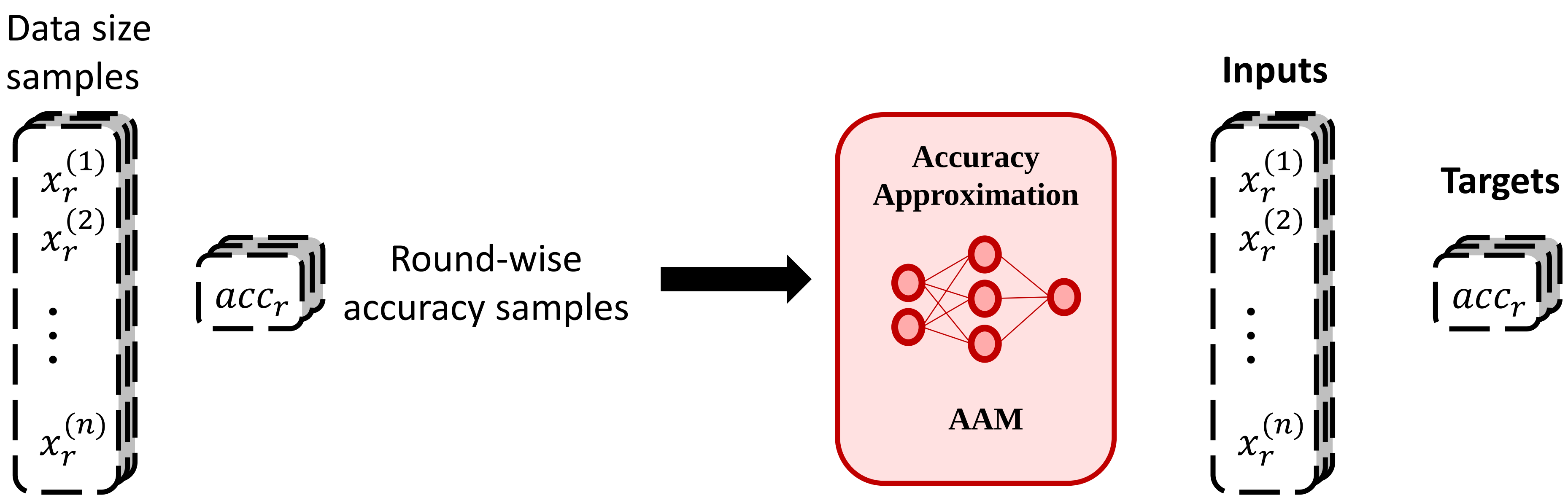
FedCCEA: 1. Simulator



FL Iterations & Testing

Partially participate in FL with selected proportions and calculate the round-wise accuracies. **The data size and the round-wise accuracy samples** are used in the next step.

FedCCEA: 2. Evaluator

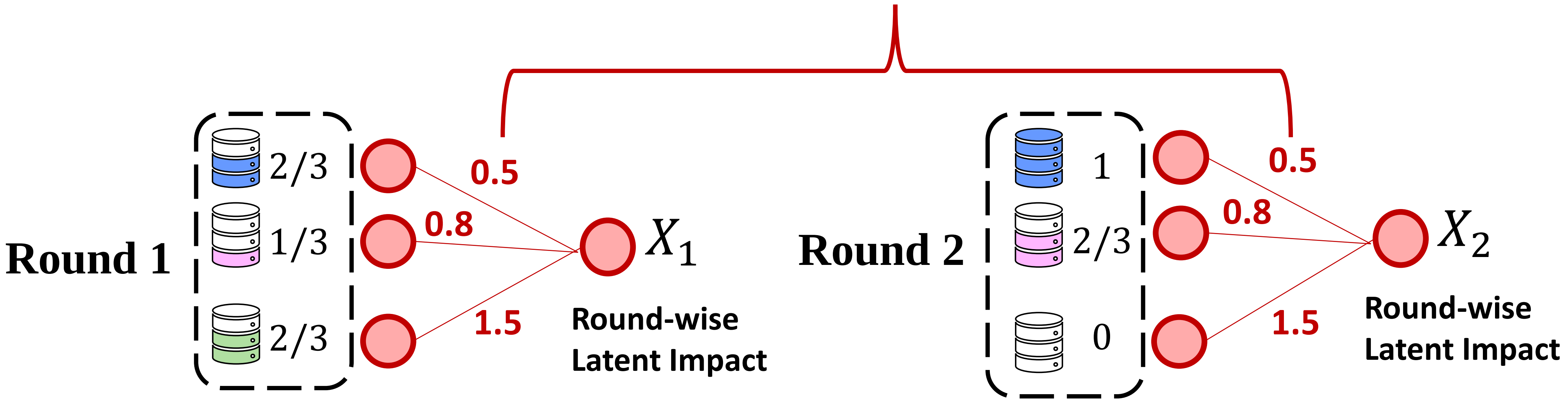


Accuracy Approximation

With given data size samples from the simulator, we **approximate the round-wise accuracy** by a **deep learning model**.

FedCCEA: 2. Evaluator - Accuracy Approximation Model

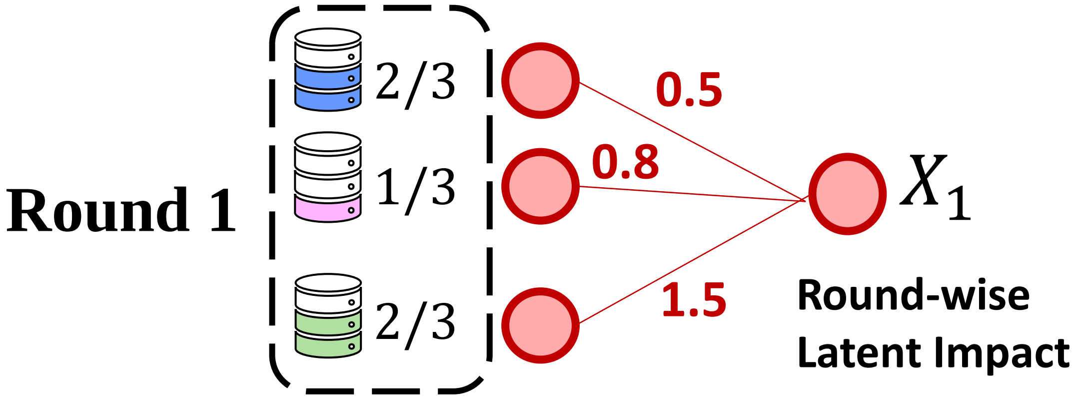
Begin with Linear Regressions

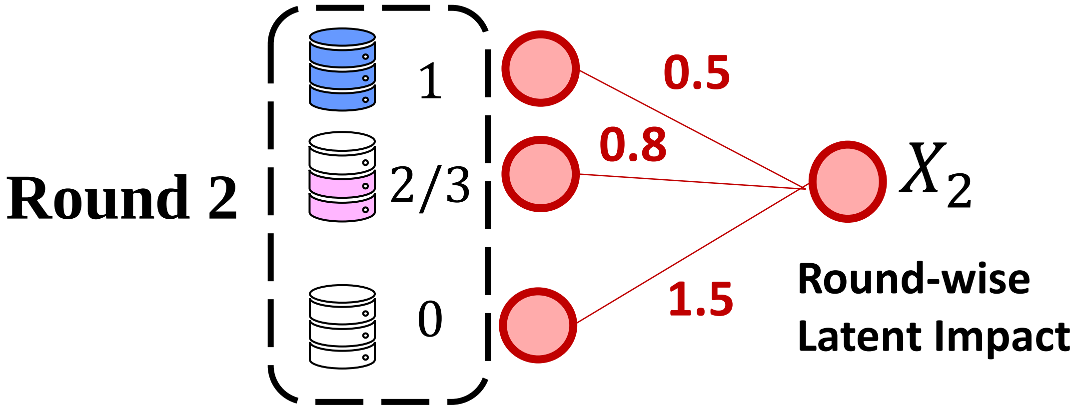


Set the round-wise linear regressions with **shared weights.**

FedCCEA: 2. Evaluator - Accuracy Approximation Model

Begin with Linear Regressions





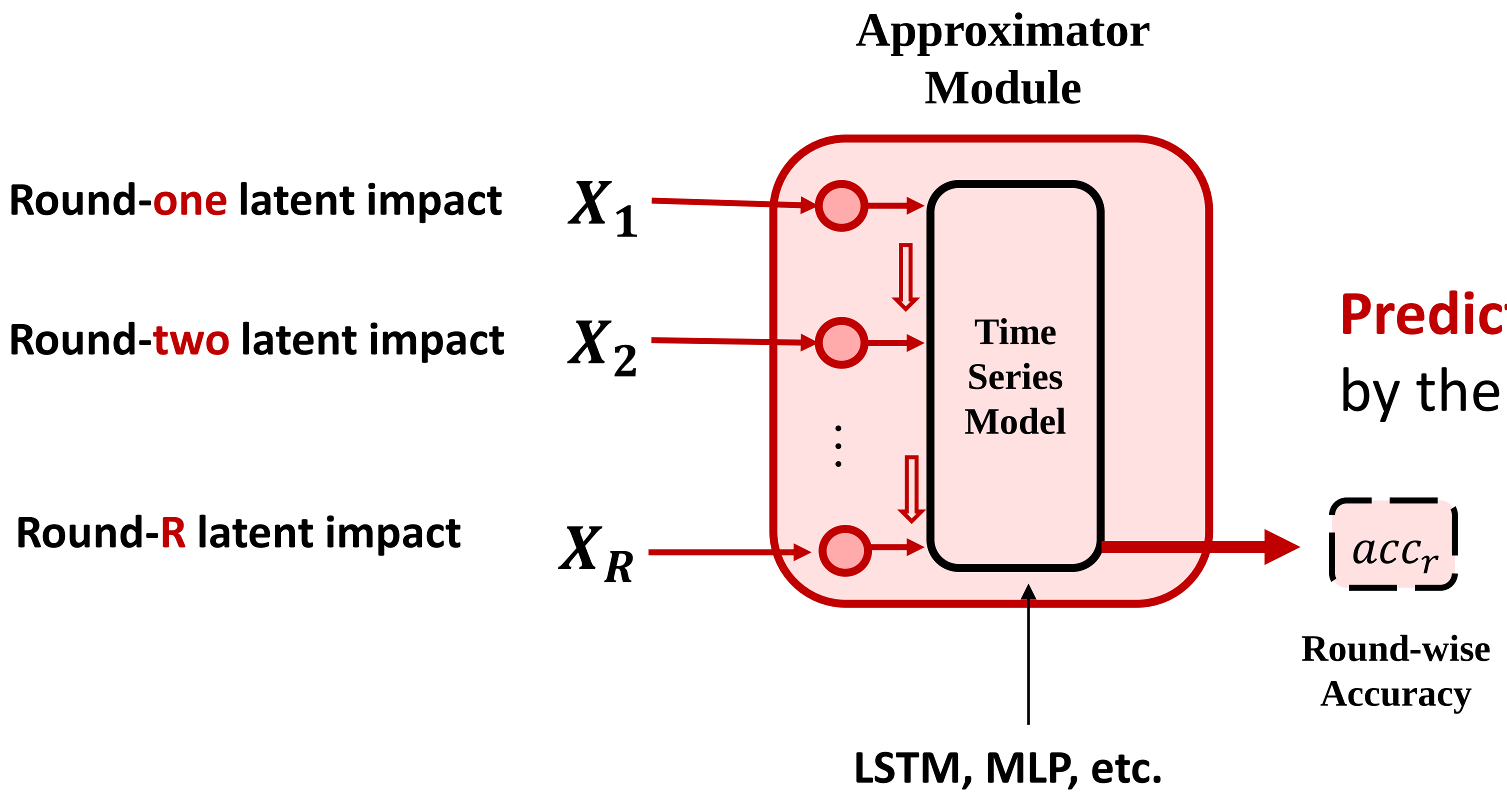
$$\boldsymbol{X_1} = \frac{2}{3} \times 0.5 + \frac{1}{3} \times 0.8 + \frac{2}{3} \times 1.5 = 1.6$$

Client 1 | Client 2 | Client 3

$$\boldsymbol{X_2} = 1 \times 0.5 + \frac{2}{3} \times 0.8 + 0 \times 1.5 = 1.0333$$

Client 1 | Client 2 | Client 3

FedCCEA: 2. Evaluator - Accuracy Approximation Model



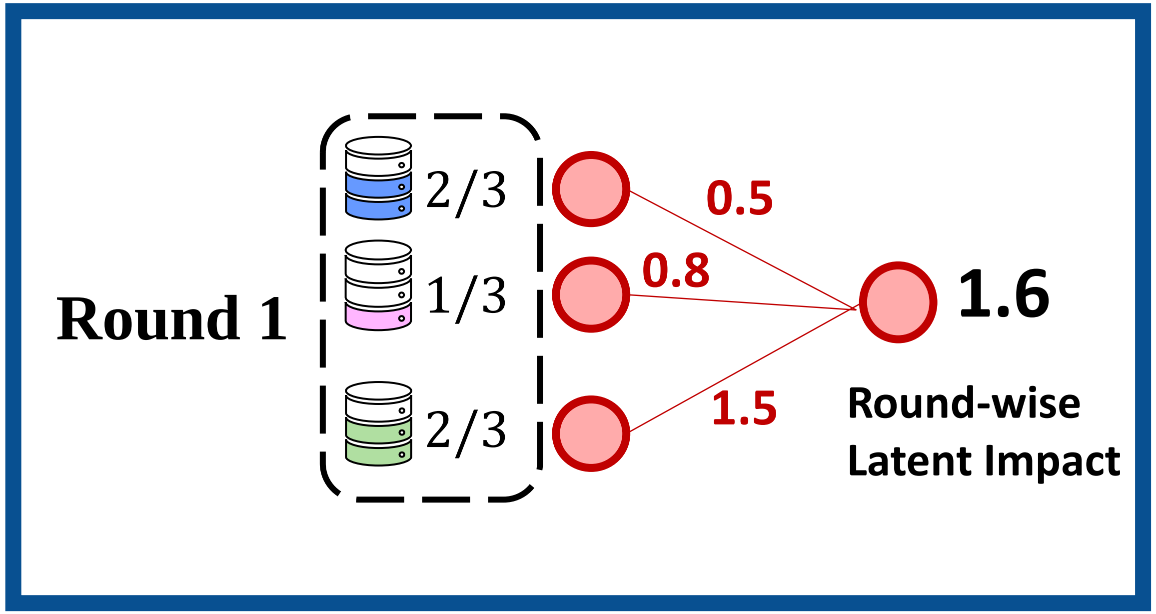
Predict the round-wise accuracy
by the series of latent impacts.

FedCCEA: 2. Evaluator - Client Contribution Measurement

Then, how to extract client contributions from Accuracy Approximation Model?

$$\textit{Client Contribution Value} = \frac{\textit{local data size} * \textcolor{red}{\textit{local heterogeneity}}}{\textit{roundwise latent impact}}$$

from shared weights



$$\frac{\frac{2}{3} \times \textcolor{red}{\mathbf{0.5}}}{1.6} = \frac{1}{4.8}$$

Client 1

$$\frac{\frac{1}{3} \times \textcolor{red}{\mathbf{0.8}}}{1.6} = \frac{0.8}{4.8}$$

Client 2

$$\frac{\frac{2}{3} \times \textcolor{red}{\mathbf{1.5}}}{1.6} = \frac{3}{4.8}$$

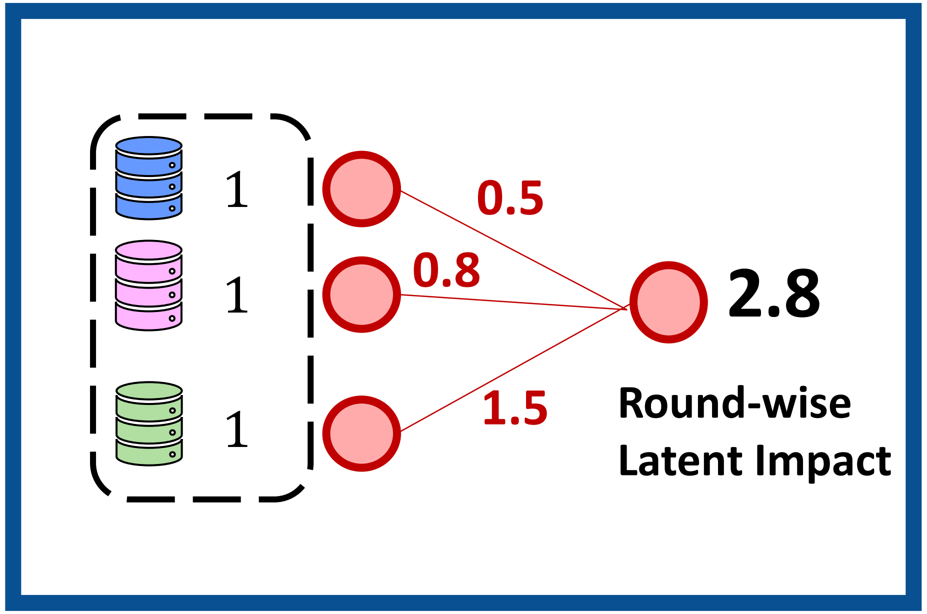
Client 3

FedCCEA: 2. Evaluator - Client Contribution Measurement

If clients use full data…

$$\textit{Client Contribution Value} = \frac{\textit{local data size} * \textbf{\textit{local heterogeneity}}}{\textit{roundwise latent impact}}$$

from shared weights



$$\frac{1 \times 0.5}{2.8} = \frac{1}{2.8}$$

Client 1

$$\frac{1 \times 0.8}{2.8} = \frac{0.8}{2.8}$$

Client 2

$$\frac{1 \times 1.5}{2.8} = \frac{1.5}{2.8}$$

Client 3

FedCCEA: Experiments

We want to answer…

1) How does accuracy variation occur in the Shapley Value evaluation and how does FedCCEA address this problem? **– Already Answered**

2) Is FedCCEA evaluation accurate even in **strong non-IID and noisy environments**?
   - **Experiment 1**

3) Is FedCCEA evaluation accurate even with **partial participation**?
   - **Experiment 2**

Experimental Setups

- **Dataset**: MNIST, EMNIST, CIFAR-10
- **FL Simulations/Rounds/Clients**: 100 simulations / 50 rounds / 20 clients
- **Baselines**
  - **Aggregation Algorithm**: FedAVG
  - **Federated Model**: MLP for MNIST, EMNIST & CNN for CIFAR-10
  - **Accuracy Approximation Model**: MLP for MNIST, CIFAR-10 & LSTM for EMNIST

- **Client Contribution Evaluation Methods** : FedCCEA / RoundSV (T. Wang, et. al, 2020)
  FIA (Y. Xue, et. al, 2021) / RRAFL (J. Zhang, et. al, 2021)
- **Data Heterogeneity**
  - **Data Distribution**: IID / Weak non-IID / Strong non-IID
  - **Data Noise Proportion**: None / 40% label noise to 4 clients

Experiment 1) Client Removal Test

We want to find out **whether each client's contribution is precise** in diverse data settings**.**
However, there are **no exact ground-truth values** to compare directly.

The alternative experiment is…**Client Removal Test**

- **Rank** the clients by contribution
- **Remove** the clients incrementally in order
- **Measure the performance** of each case and **compare** the flows.

① Ranking

Client 4
Client 3
Client 5
Client 2
Client 1

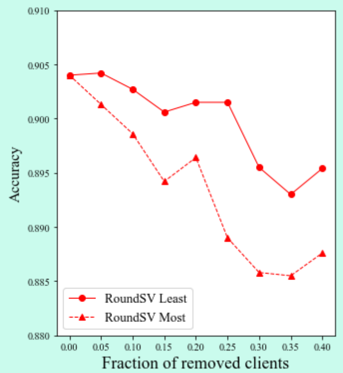
② Incremental Removal

~~Client 4~~
Client 3
Client 5
Client 2
Client 1

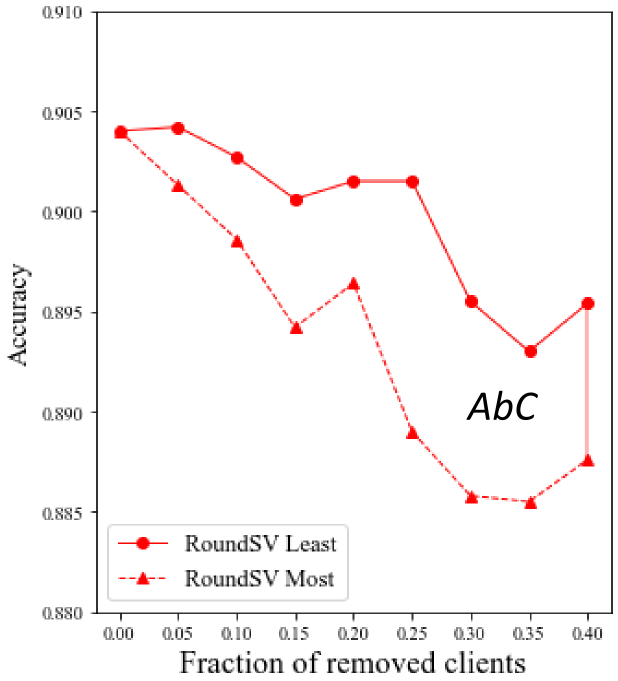
~~Client 4~~
~~Client 3~~
Client 5
Client 2
Client 1

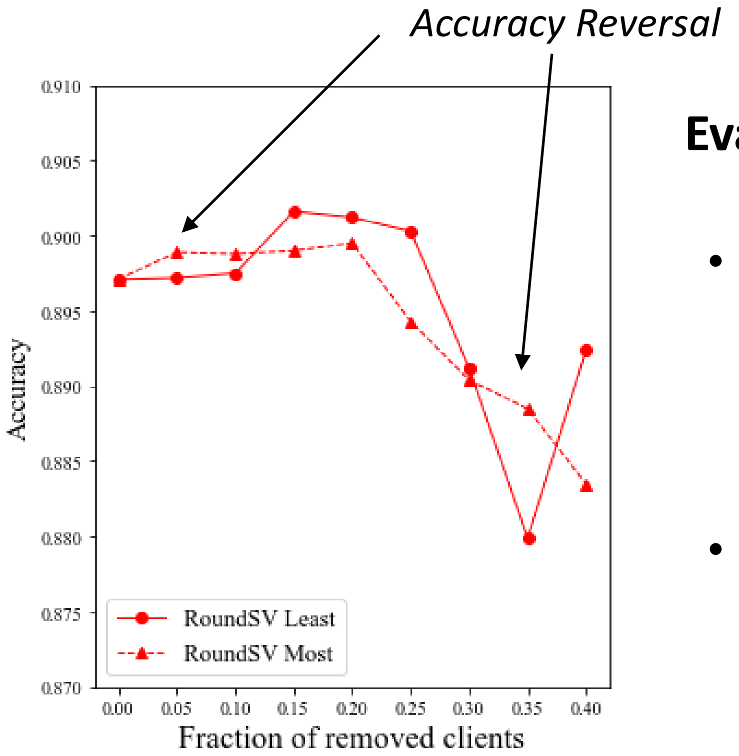
~~Client 4~~
~~Client 3~~
~~Client 5~~
Client 2
Client 1

③ Measure & Compare



Experiment 1) Client Removal Test





Ideally, the model with removed low contributors(straight line) should consistently retain high accuracy and the model with removed high contributors(dashed line) should decrease substantially.

Evaluation Metrics

- **Accuracy Reversal**(AR) : The existence of the region where high-contributor removal(Most) exceeds low-contributor removal(Least).

- **Area between the Curves**(AbC):

$$\text{AbC} = \sum_{\text{frac}} \text{acc}_{\text{low,frac}} - \text{acc}_{\text{high,frac}}$$

Experiment 1) Client Removal Test

I: IID / W: weak non-IID / S: strong non-IID
1: no noise / 2: 40% noise to 4 clients

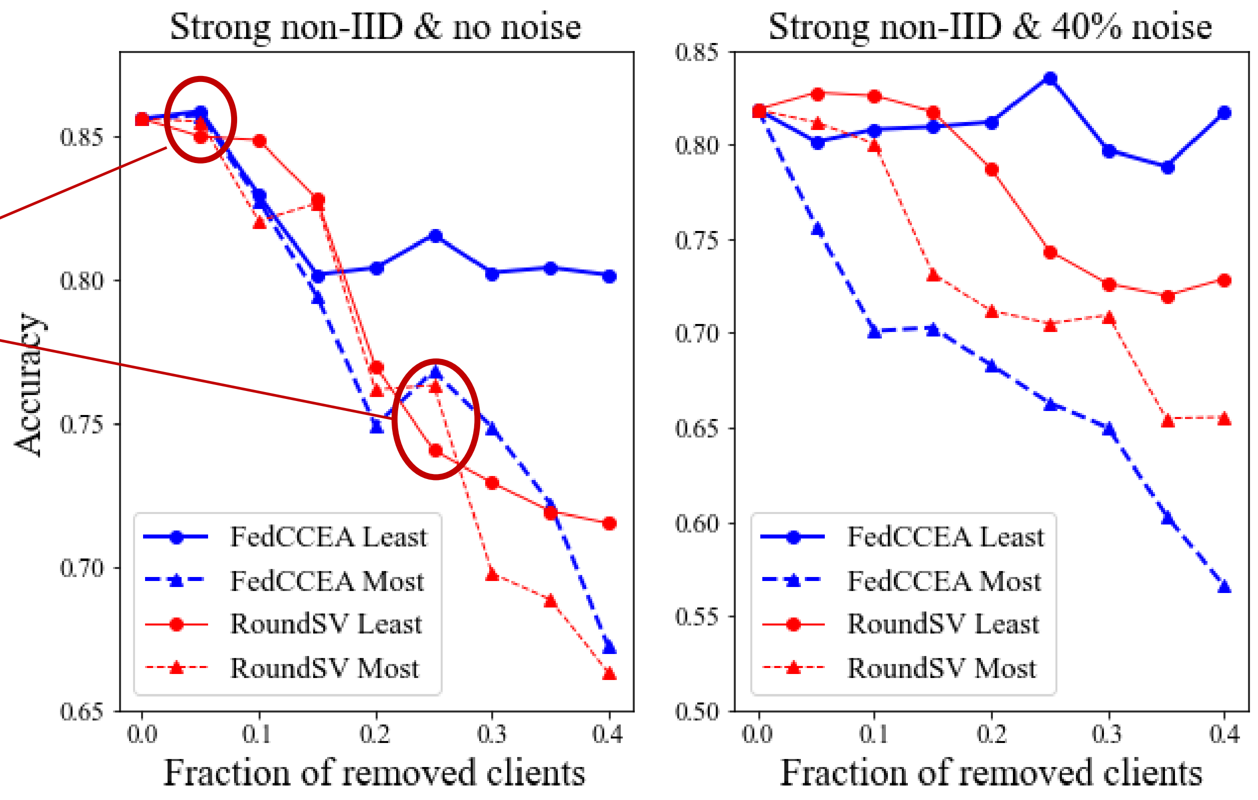
Methods		MNIST					EMNIST					CIFAR-10			Best
		I.1	W.1	W.2	S.1	S.2	I.1	W.1	W.2	S.1	S.2	I.1	W.1	W.2	
FedCCEA	AbC	0.0211	0.1333	0.3314	0.3789	1.1443	0.0593	0.3190	0.3848	0.3732	0.5189	0.0896	0.1651	0.1241	9
	AR	×	×	×	×	×	×	×	×	×	×	×	×	×	
RoundSV	AbC	0.0193	0.056	0.0084	0.1243	0.3964	0.0408	0.2645	0.0106	0.0421	0.2080	-0.1306	0.0556	-0.0742	0
	AR	×	×	√	√	×	×	×	×	√	√	√	√	√	
FIA	AbC	0.0277	0.1165	0.3240	0.1346	1.0995	0.0522	0.3616	0.3093	0.3313	0.7308	0.1132	0.0310	0.0739	3
	AR	×	×	×	√	×	×	×	×	×	×	×	√	√	
RRAFL	AbC	0.0232	-0.0166	0.2499	-0.3588	-0.4707	0.0567	0.2499	0.3796	0.1181	0.4680	0.1162	-0.0246	0.1238	1
	AR	×	√	×	√	√	×	×	×	×	×	×	×	×	

FedCCEA shows **robust measurements** in most of the heterogeneous data environments.

- **RoundSV**
  - Experiences accuracy reversals(AR) in some settings.
  - Negative *AbC*s in some settings.
- **FedCCEA**
  - **No accuracy reversal** in all settings.
  - Receives **9 highest *AbC* out of 13 settings.**

Experiment 1) Client Removal Test

Accuracy Reversal

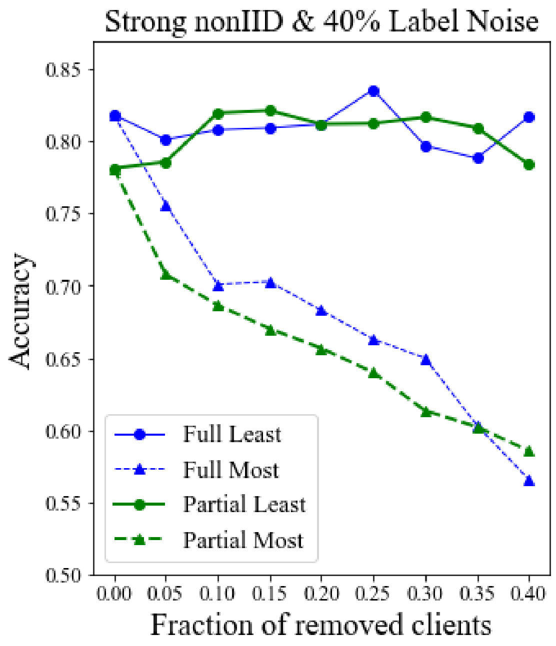


Experiment 2) Client Removal Test for Partial Participation

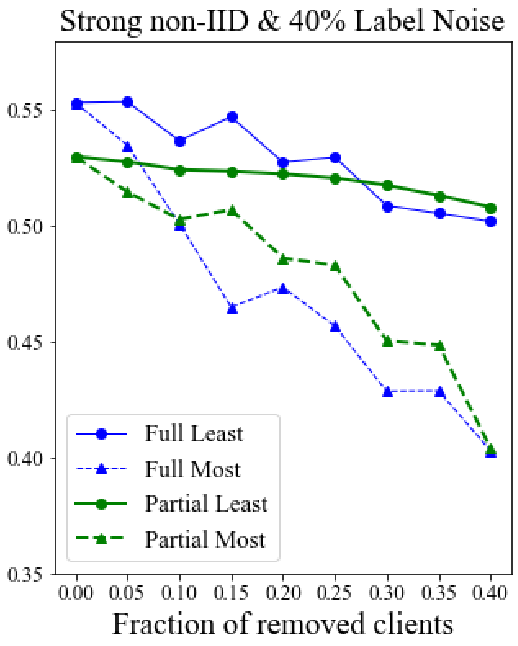
By FedCCEA, we can measure client contributions even if clients partially participate in FL.

Do the results remain consistent? Yes! We can use FedCCEA in practice where partial participation occurs.

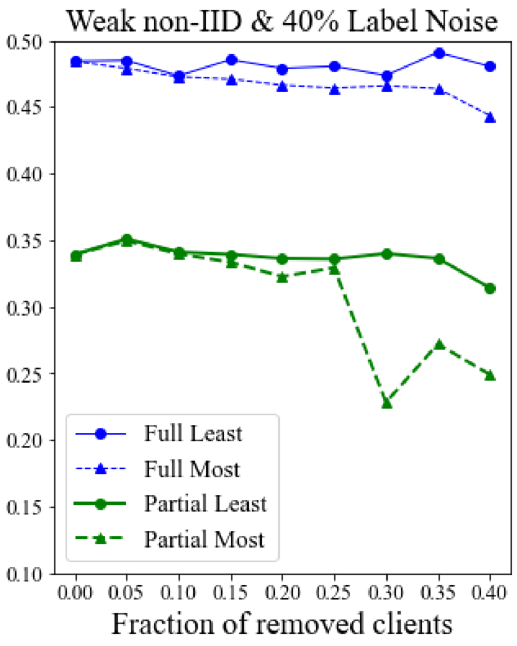
MNIST



EMNIST



CIFAR-10



Summary

- This is the first ***empirical*** method that allows the ***partial participation of clients*** and exploits ***data size sets*** for a client contribution evaluation in FL.
- We use ***data size diversification*** to make diverse data size samples and reduce the accuracy variation so that ***the client contribution can be stabilized***.
- We conduct extensive experiments with heterogeneous data environments: ***diverse data distributions and data noise corruptions***. FedCCEA shows ***more robust results of client removal tests*** than other baseline methods.

Future Work

- **Apply to Incentive Mechanism**
  - In practice, most clients want to maximize profits rather than maximize the federated model performance. (ex. Blockchain)
  - Motivation to participate actively in FL comes from money.
  - Well-structured incentive mechanism using client contribution is necessary.

- **Contribution-based Aggregation**
  - If client contribution is precise, weighted averaging based on client contribution can be applied where FedAVG is placed.

Thank you.