An Accuracy-Shaping Mechanism for Competitive Distributed Learning

Chao Huang, Justin Dachille, and Xin Liu

University of California, Davis, CA 95616, USA

Email: {fchhuang, jtdachille, xinliu}@ucdavis.edu

Abstract. In competitive distributed learning, organizations face the challenge of collaboratively training machine learning models without sharing sensitive raw data, while competing for the same customer base using model-based services. Federated learning is an extensively studied distributed learning approach, but it has been shown to discourage collaboration in a competitive environment. The reason is that the shared global model is a public good, which can lead to intense organization competition and hence small incentives for collaboration. To address this issue, this paper uses SplitFed learning (SFL) for model training and proposes an accuracy-shapring mechanism to incentivize inter-organizational collaboration. SFL divides the global model into two components: one trained by the organizations and the other by a main server. After convergence, the mechanism introduces customized noise into the main server's model, enabling the provision of differentiated models to each organization. Both our theoretical analysis and numerical experiments validate the efficacy of SFL and the proposed mechanism, showing significant improvements in both model accuracy and social welfare at equilibrium.

Keywords: distributed machine learning · split federated learning · business competition · mechanism design.[1]

1 Introduction

Distributed learning enables learning clients (e.g., organizations) to collaboratively train machine learning models without sharing raw data [23]. While prior work focused on improving learning performance, the important aspect of **business competition** is yet fully explored and understood. That is, organizations may use the shared model to offer model-based services to compete for the same pool of customers [11] (see Fig. 1). For instance, multiple healthcare providers could train a shared predictive model for patient outcomes without sharing individual patient records, yet each could use the model to offer personalized healthcare services. Similarly, financial institutions might jointly develop a credit scoring model without revealing customer financial histories, and then compete by

[1] This work was supported through USDA-2020-67021-32855 and NSF OIA-2134901. Corresponding author: Chao Huang.

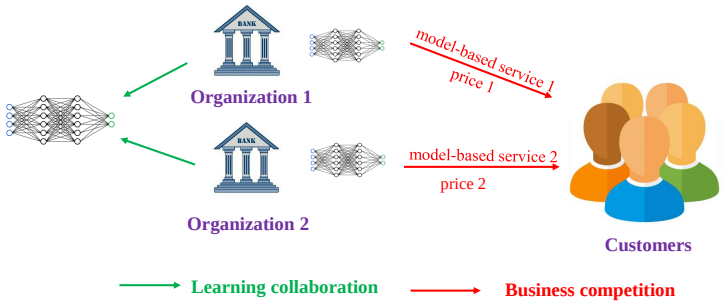

Fig. 1. Business competition in distributed learning.

offering tailored loan services. This dual focus on collaboration without data sharing and competition through model-based services makes distributed learning more challenging and intriguing, which is the focus of this paper.

Federated learning (FL) is a recent distributed learning approach, which has received extensive attention from both academia and industry [15]. Many studies looked at the client drift problem caused by data heterogeneity in FL and proposed effective solutions to this end (e.g., [27, 22, 21]). Some other studies focused on label noise in FL (e.g., [17, 35]) and devised solutions to improve algorithm robustness (e.g., [34, 14]). However, recent research has raised concerns about the incentive structure within the FL framework. Specifically, the global model generated through FL is essentially *a public good*, which is accessible to all participating organizations [29]. In this case, organizations may be reluctant to contribute to (or even participate in) FL, which could also benefit their competitors.

To address the aforementioned challenges posed by FL, this paper proposes to use SplitFed learning (SFL) to incentivize collaborative model training among competing organizations. SFL operates by partitioning a global model into two segments, in which organizations are responsible for training one segment, while the main server performs the training of the remaining one [30]. After multiple iterations of training, an auxiliary fed server is responsible for aggregating the models maintained by the organizations.

It is important to note that during the training phase of SFL, organizations are precluded from accessing the main-server-side model. Consequently, this restriction enables the main server to allocate varying model versions to different organizations. This has a potential to ensure that each organization maintains competitive uniqueness and hence has incentives to collaborate. For instance, one organization may specialize in refining demographic analytics, while another focuses on consumer purchasing patterns. The tailored/differentiated model by SFL not only preserves the competitive edge of each organization but also fosters a more collaborative environment, as organizations are incentivized to collaborate without compromising their unique insights. Conversely, the open-access nature of the global model in FL intensifies competition, as organizations know

that their competitors have equal access to the global model. This paper aims to answer the following questions.

Question 1. How to model business competition and what is its impact on SplitFed learning?

Question 2. How to incentivize collaboration among competing organizations in SplitFed learning?

To answer these questions, we consider a duopoly scenario involving two organizations,[2] a main server, and a continuum of customers. We formulate a four-stage game model to characterize their game-theoretical interactions. In Stage I, the main server designs an accuracy-shaping mechanism that provides tailored main-server-side model to organizations post convergence to maximize social welfare. In Stages II and III, each revenue-maximizing organization decides whether to participate in SFL as well as the unit price for the model-based service. In Stage IV, each individual customer makes informed decisions regarding whether to purchase the service and, if so, from which organization.

To derive insights, we study the equilibrium of the four-stage model via two metrics: *model accuracy* and *social welfare*. Model accuracy serves as an important indicator of whether SFL and the proposed mechanism work. Social welfare, defined as the summation of organization revenue and customer surplus, provides a lens of the overall societal impact of SFL and the accuracy-shaping mechanism. We consider two types of benchmarks: *FL* (e.g., FedAvg [24], FedProx [22], and MOON [21]) and *local learning (LL)*, where local learning means the organizations do not collaborate in either SFL or FL, and instead train a local model using their own data. We will show that under a carefully designed mechanism, SFL outperforms FL and LL w.r.t. the above two metrics at equilibrium.

Our key contributions are summarized as follows:

- We propose to use SFL to address the business competition issue in distributed learning. To understand the impact of business competition, we formulate a four-stage model to characterize the strategic interactions among the main server, the competing organizations, and the customers.
- We analyze the model's equilibrium and show that organizations are reluctant to participate in SFL if the main server assigns them the same version of main-server-side model. Providing the same models to organizations intensifies their business competition and resulting in small incentives to collaborate.
- To incentivize collaboration, we design an accuracy-shaping mechanism where the main server assigns different versions of main-server-side models to organizations post convergence. Both theoretical analysis and numerical results show that SFL outperforms both FL and LL in model accuracy and social welfare.

[2] Duopoly exists in practice, e.g., SwissRe and WeBank collaborate via FL to provide reinsurance services [1]. The anaysis of more than two organizations is much more math involved and its discussions are left to the online Appendix 3.1 [2].

1.1 Related Work

Federated Learning (FL). FL is a popular distributed learning approach that has received tremendous attention. See [15, 36, 12] for a few excellent surveys. We only review the most relevant work pertaining to incentives in FL. Sun *et al* [28] proposed customized mechanisms catering to clients' privacy costs. Donahue and Kleinberg in [3, 4] studied the clients' optimal collaboration strategy under data heterogeneity. However, none of these studies looked at business competition.

Until recently, some research attention has been drawn to business competition in FL [11, 33, 5, 6]. Wu *et al* [33] proposed a game-theoretical decision framework for FL competitors to maximize their market share. Gradwohl and Tennenholtz [6] analyzed clients' optimal data sharing strategy in the presence of competition. Huang *et al* in [11] proposed a revenue-sharing framework to promote collaborations between competing organizations. Dorner *et al* in [5] devised mechanisms to incentivize truthful model updates. However, these studies require monetary rewards/transfers among organizations, which may not be feasible in practice. Our study presents an accuracy-shaping mechanism that does not require any monetary transfer. One recent work [9] designed a model differentiation mechanism without monetary transfer, but the method was tailored for FL and cannot be directly used for SFL.

Split Learning (SL) and SplitFed Learning (SFL). SL is another distributed learning paradigm in which the model is split into two segments, where clients train one segment in a round-a-robin fashion and the server trains the other segment [32, 31]. SL reduces the computation overhead of clients but suffers from a large latency due to sequential client training. To this end, Thapa *et al* in [30] proposed SFL that combines FL and SL, enabling parallel training among clients (like FL) and model splitting (like SL). Recent studies on SFL focused on convergence analysis (e.g., [8]), reducing the communication overheads (e.g., [25, 26, 7]) and improving privacy/security [18, 20, 19]. Different from prior work, we study business competition in SFL, and propose game-theoretical mechanisms to promote inter-organization collaboration.

2 Model

Section 2.1 introduces SFL. Sections 2.2-2.4 model the customers, organizations, and the main server. Section 2.5 models their game-theoretical interactions.

2.1 SplitFed Learning Process

We consider a group of two organizations, denoted as $\mathcal{N} = \{1, 2\}$, who seek to cooperatively develop a machine learning model without exchanging raw data. Each organization, represented by n, possesses a private dataset $\mathcal{D}_n$ of size $D_n = |\mathcal{D}_n|$. In SFL, organizations execute model training with the help from two servers: (1) **fed-server**, responsible for averaging the local models from the organizations similar to FL; (2) **main-server**, responsible for conducting a segment of the model training.

To be more concrete, consider a global model characterized by $\boldsymbol{x}$, split into an organization-side model $\boldsymbol{x}_c$ and a main-server-side model $\boldsymbol{x}_s$, with $\boldsymbol{x} = \{\boldsymbol{x}_c, \boldsymbol{x}_s\}$. Assume that the global model comprises L layers. Here, the organizations train the organization-side model, i.e., the initial L_c layers, while the main-server focuses on the main-server-side model, i.e., the residual $L_s = L - L_c$ layers. The L_c-th layer is referred to as the *cut layer*. It's important to note that every organization sustains its individual organization-side model $\boldsymbol{x}_{c,n}$, whereas the main-server maintains a model corresponding to each organization $\boldsymbol{x}_{s,n}$.[3] The loss function of organization n is represented as $f_n(\boldsymbol{x})$. The goal of SFL is to minimize the global loss function $F(\boldsymbol{x})$ below:

$$\min_{\boldsymbol{x}} F(\boldsymbol{x}) = \sum_{n=1}^{N} a_n f_n(\boldsymbol{x}), \tag{1}$$

where $a_n = D_n / \sum_{n' \in \mathcal{N}} D_{n'}$.

A typical SFL process undergoes a total of T rounds to minimize $F(\boldsymbol{x})$. In each round t, the organizations first acquire their initialized models from the fed-server. Each round is bifurcated into two phases:

Phase 1: Training the Model. Organizations, along with the main-server, engage in training the model through τ iterations. In every iteration $i < \tau$:
1. *Organization Forward Propagation:* Every organization n selects a mini-batch of data from $\mathcal{D}_n$, calculates the intermediate representations (e.g., activation values at the cut layer) over its current model $\boldsymbol{x}_{c,n}^{t,i}$, and transmits the intermediate representations (together with labels) to the main-server. Note that the organizations execute forward propagation in parallel.
2. *Main-server Training:* The main-server computes the loss and the corresponding gradients for organization n, and updates the server-side model $\boldsymbol{x}_{s,n}^{t,i}$ corresponding to organization n. The N main-server-side models are updated in parallel as well.
3. *Organization Backward Propagation:* Each organization n calculates the gradient and updates its model accordingly.

Phase 2: Aggregating the Model. Post τ iterations of model training with the main-server, every organization n forwards its updated organization-side model (or gradients) to the fed-server. The fed-server aggregates the received models (possibly through averaging) for organizations to download in the following round. The main server also aggregates all the server-side-models to be initialized in the next round.

After T rounds, each organization n obtains $\boldsymbol{x}_{c,n}^T$, and the main server obtains $\{\boldsymbol{x}_{s,n}^T, \forall n\}$. The main server assigns a main-server-side model $\tilde{\boldsymbol{x}}_{s,n}^T$ (which can differ from $\boldsymbol{x}_{s,n}^T$) to each organization n, who obtains a full global model $\tilde{\boldsymbol{x}}_n^T = \left\{x_{c,n}^T, \tilde{\boldsymbol{x}}_{s,n}^T\right\}$ with accuracy $A_n \in [0, 1]$. Organizations utilize their local models to produce services related to the model (e.g., disease diagnoses by hospitals).

[3] This is another major version of SFL in which the main server only maintains one model when interacting with all organizations. We include this version in Sec. 5, and provide its description in Appendix 3.2 [2].

Subsequently, organizations enter the market to competitively offer these services to consumers. Next, we model the decision problems of customers, organizations, and the main server.

2.2 Customer's Decision Problem

We consider that the two organizations can reach the same pool of customers, who are normalized to a population size of one. A customer's valuation for a service related to the model is denoted by θ. We model θ as a random variable with a PDF $h(\theta)$ and a CDF $H(\theta)$ on support $[0, \theta_{\max}]$. Similar to prior studies [11, 13], we assume that the distribution of valuations is known to the organizations, for example due to market research, while individual valuations remain unknown.

Two organizations in competition might provide services of varying quality (as they have different local models) and at different prices. Based on the competing qualities and prices, a customer decides whether to buy a service and, if so, from which organization. A customer's decision is denoted as d_θ, where $d_\theta = \emptyset$ means no buying, and $d_\theta = n$ means buying from organization n. A customer's surplus is the utility derived from the service minus the payment made to the organization. If no service is purchased, the surplus is zero. If a service is bought from organization n, the customer enjoys the utility from the service and pays an amount p_n to organization n. Specifically, a customer with valuation θ has the surplus function as follows:

$$u_\theta(d_\theta; \mathcal{M}, \boldsymbol{q}, \boldsymbol{p}) = \begin{cases} 0, & \text{if } d_\theta = \emptyset, \\ \theta \cdot V(A_n(\mathcal{M}, \boldsymbol{q})) - p_n, & \text{if } d_\theta = n. \end{cases} \tag{2}$$

We assume $V(\cdot)$ to be a strictly increasing function, meaning that the customer obtains a higher utility $\theta \cdot V(A_n(\mathcal{M}, \boldsymbol{q}))$ from enjoying a service of higher quality (which corresponds to a larger $A_n(\mathcal{M}, \boldsymbol{q})$). Here, $\mathcal{M}$ is the accuracy-shaping mechanism and $\boldsymbol{q}$ represents the organizations' participation decision in SFL, which will be introduced in Sec. 2.3-2.4.

We define a type-θ customer's decision problem as:

$$\mathbf{P1}: \quad \max_{d_\theta} \quad u_\theta(d_\theta; \mathcal{M}, \boldsymbol{q}, \boldsymbol{p}). \tag{3}$$

2.3 Organizations' Decision Problem

Each organization n has two decisions to make. First, it needs to decide whether to participate in SFL, and we use $q_n \in \{0, 1\}$ to model the participation decision, where 0 means no participation and 1 means participation. Second, each organization decides a unit price $p_n \geq 0$ for its provided model-based service to each customer. Denote $\boldsymbol{q} = (q_1, q_2)$ and $\boldsymbol{p} = (p_1, p_2)$.

The revenue that an organization n earns from an individual customer is the price p_n it sets. Therefore, the total revenue that organization n accumulates

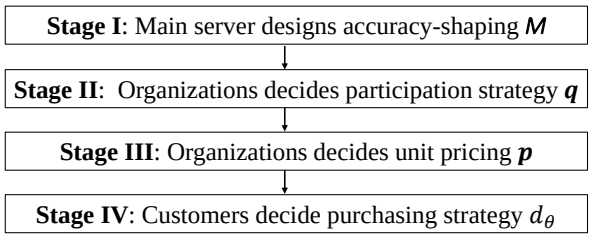

Fig. 2. Four-stage game model.

from the entire customer base is calculated as follows:

$$R_n(\mathcal{M}, \boldsymbol{q}, \boldsymbol{p}) = \int_0^{\theta_{\max}} p_n \cdot \mathbb{1}_{d_\theta = n}(\mathcal{M}, \boldsymbol{q}, \boldsymbol{p}) \cdot h(\theta) d\theta, \tag{4}$$

where $\mathbb{1}$ is an indicator function, meaning $\mathbb{1}_{d_\theta=n} = 1$ if and only if $d_\theta = n$. Participating in SFL also involves computation costs, such as model training, and communication costs, such as the transmission of organization-side model with the fed-server and activations/gradients with the main-server. Similar to [10, 11], we normalize the costs to zero, as the organizations usually have strong computation capabilities and sufficient communication resources. Refer to Appendix 3.3 for a discussion on the computation/communication of SFL.

Each organization's decision problem is defined as:

$$\mathbf{P2}: \quad \max_{q_n, p_n} \quad R_n(\mathcal{M}, \boldsymbol{q}, \boldsymbol{p}). \tag{5}$$

2.4 Main-Server's Decision Problem

As discussed, the main server assigns each organization n a main-server-side model $\tilde{\boldsymbol{x}}_{s,n}^T$ which may differ from $\boldsymbol{x}_{s,n}^T$. To be concrete, we use an accuracy shaping mechanism $\mathcal{M} : \left\{\boldsymbol{x}_{s,n}^T, \forall n\right\} \to \left\{\tilde{\boldsymbol{x}}_{s,n}^T, \forall n\right\}$, that maps from the main-server-side models to the assigned versions to organizations. In practice, the main server can assign customized models to organization via noise perturbation [16].

The main server aims to maximize the social welfare, i.e., the summation of organizations' revenues and customer surplus. The problem is formulated below.

$$\mathbf{P3}: \max_{\mathcal{M}} \sum_{n \in \mathcal{N}} R_n(\mathcal{M}, \boldsymbol{q}, \boldsymbol{p}) + \int_\theta u_\theta(d_\theta; \mathcal{M}, \boldsymbol{q}, \boldsymbol{p}) h(\theta) d\theta. \tag{6}$$

2.5 Four-Stage Game Formulation

We model the interactions among the main server, organizations, and customers as a four-stage game (Fig. 2). The main server designs the accuracy-shaping

mechanism in Stage I to solve **P3**. The organizations choose participation in Stage II and pricing in Stage III to solve **P2**. Each customer decides purchasing in Stage IV to solve **P1**. We analyze the game via backward induction.

3 Customers' and Organizations' Decisions

3.1 Customers' Optimal Purchasing in Stage IV

Given $\mathcal{M}$, $\boldsymbol{q}$, and $\boldsymbol{p}$, each customer aims to solve **P1** by choosing his purchasing strategy. We summarize the optimal decisions in Lemma 1.

Lemma 1. *Let $-n \triangleq \mathcal{N} \setminus \{n\}$. A type-$\theta$ customer's optimal decision is*

$$d_\theta^*(\mathcal{M}, \boldsymbol{q}, \boldsymbol{p}) = \begin{cases} n, \text{ if } \theta V(A_n(\mathcal{M}, \boldsymbol{q})) - p_n \geq \max\{\theta V(A_{-n}(\mathcal{M}, \boldsymbol{q})) - p_{-n}, 0\}, \\ \emptyset, \text{ else.} \end{cases} \tag{7}$$

Refer to the Appendix 1 [2] for all technical proofs.

Lemma 1 shows that in a duopoly market, a customer of type-θ will purchase service from organization n if the payoff from organization n is both non-negative and greater than that from the other one.

3.2 Organization Optimal Pricing in Stage III

In Stage III, based on the accuracy-shaping mechanism $\mathcal{M}$ in Stage I, and the participation strategy $\boldsymbol{q}$ in Stage II, the two organizations decide their unit pricing $\boldsymbol{p}$ to optimize their own revenue as defined in (4), while anticipating the customers' optimal purchasing in Stage IV.

We assume without loss of generality that in Stage III, organization 1 has a better model than organization 2, i.e., $A_1(\mathcal{M}, \boldsymbol{q}) > A_2(\mathcal{M}, \boldsymbol{q})$. Based on Lemma 1, we can derive the organizations' expected revenue functions below:

$$R_1(\mathcal{M}, \boldsymbol{q}, \boldsymbol{p}) = p_1 \left[1 - H\left(\frac{p_1 - p_2}{V(A_1(\mathcal{M}, \boldsymbol{q})) - V(A_2(\mathcal{M}, \boldsymbol{q}))} \right) \right], \tag{8}$$

$$R_2(\mathcal{M}, \boldsymbol{q}, \boldsymbol{p}) = p_2 H\left(\frac{p_1 - p_2}{V(A_1(\mathcal{M}, \boldsymbol{q})) - V(A_2(\mathcal{M}, \boldsymbol{q}))} \right) - p_2 H\left(\frac{p_2}{V(A_2(\mathcal{M}, \boldsymbol{q}))} \right). \tag{9}$$

where $H(\cdot)$ is the CDF of customers' type distribution.

Based on (8)-(9), we model the pricing competition as a game.

Game 1. *(Price Competition in Stage III) The price competition game in Stage III is defined as a tuple $< \mathcal{N}, \mathcal{P} = \prod p_n, \mathcal{R} = \prod R_n >$, where each organization n in $\mathcal{N}$ decides the unit pricing p_n to maximize R_n in (4).*

We aim to find Game 1's Nash equilibrium (NE).

Definition 1. *A strategy profile* $\boldsymbol{p}^*(\mathcal{M}, \boldsymbol{q}) \triangleq (p_n^*(\mathcal{M}, \boldsymbol{q}), p_j^*(\mathcal{M}, \boldsymbol{q}))$ *is an NE of Game 1 if for all* $n \in \{1, 2\}$, $p_n'(\mathcal{M}, \boldsymbol{q}) \neq p_n^*(\mathcal{M}, \boldsymbol{q})$,

$$R_n(p_n^*(\mathcal{M}, \boldsymbol{q}), p_j^*(\mathcal{M}, \boldsymbol{q})) \geq R_n(p_n'(\mathcal{M}, \boldsymbol{q}), p_j^*(\mathcal{M}, \boldsymbol{q})), \tag{10}$$

where $j \neq n$ *and* $j \in \{1, 2\}$.

NE is a stable outcome as no organization can be better off via unilaterally changing its pricing strategy. Note that an arbitrary choice of $H(\cdot)$ can render the problem analytically intractable. To facilitate analysis, we make minor assumptions on $H(\cdot)$ (and $h(\cdot)$).

Assumption 1. *On support* $[0, \theta_{\max}]$, *(i)* $h(\theta) > 0$ *and is continuous; (ii)* $h(\theta)/[1 - H(\theta)]$ *is increasing in* θ.

Assumption 1 holds for commonly adopted distributions, e.g., normal, gamma, and uniform distributions. We can establish the equilibrium existence below.

Lemma 2. *Under Assumption 1, Game 1's NE exists.*

To obtain clean insights, we use a uniform distribution to compute the closed-form equilibrium:

$$\begin{cases} p_1^*(\mathcal{M}, \boldsymbol{q}) & = \frac{2[V(A_1(\mathcal{M},\boldsymbol{q})) - V(A_2(\mathcal{M},\boldsymbol{q}))]V(A_1(\mathcal{M},\boldsymbol{q}))\theta_{\max}}{4V(A_1(\mathcal{M},\boldsymbol{q})) - V(A_2(\mathcal{M},\boldsymbol{q}))}, \\ p_2^*(\mathcal{M}, \boldsymbol{q}) & = \frac{[V(A_1(\mathcal{M},\boldsymbol{q})) - V(A_2(\mathcal{M},\boldsymbol{q}))]V(A_2(\mathcal{M},\boldsymbol{q}))\theta_{\max}}{4V(A_1(\mathcal{M},\boldsymbol{q})) - V(A_2(\mathcal{M},\boldsymbol{q}))}. \end{cases} \tag{11}$$

Based on (11), we can see that $p_1^*(\mathcal{M}, \boldsymbol{q}) > p_2^*(\mathcal{M}, \boldsymbol{q})$. It implies that organization 1 with a better model will set a larger price to attract customers with larger valuations. Importantly, the optimal prices of both organizations depend on the main-server's mechanism $\mathcal{M}$.

3.3 Organization Optimal Participation in Stage II

In Stage II, given the mechanism $\mathcal{M}$ from Stage I, the organizations decide participation strategy $\boldsymbol{q}$ for SFL, anticipating the responses from Stages III and IV. We model the two organizations' interactions as a participation game.

Game 2. *(Participation Game in Stage II) The participation game is a tuple* $< \mathcal{N}, \mathcal{Q} = \prod q_n, \mathcal{R} = \prod R_n >$, *where each organization* n *decides* q_n *in SFL to maximize* $R_n(\mathcal{M}, \boldsymbol{q}, \boldsymbol{p}^*(\mathcal{M}, \boldsymbol{q}))$, *where* $\boldsymbol{p}^*(\mathcal{M}, \boldsymbol{q})$ *is the NE of Game 1.*

We are interested in solving Game 2's NE defined below.

Definition 2. *A profile* $\boldsymbol{q}^* \triangleq (q_n^*, q_j^*)$ *is an NE of Game 2 if* $\forall n$, $\forall q_n' \neq q_n^*$,

$$R_n(\mathcal{M}, \boldsymbol{q}^*, \boldsymbol{p}^*(\mathcal{M}, \boldsymbol{q}^*)) \geq R_n(\mathcal{M}, (q_n', q_j^*), \boldsymbol{p}^*(\mathcal{M}, (q_n', q_j^*))), \tag{12}$$

where $j \neq n$ *and* $j \in \{1, 2\}$.

Next, we characterize the NE in Proposition1.

Proposition 1. $(1,1)$ *is Game 2's NE if and only* $R_n(\mathcal{M},(1,1),\boldsymbol{p}^*(\mathcal{M},(1,1))) \geq R_n(\mathcal{M},(0,0),\boldsymbol{p}^*(\mathcal{M},(0,0))),\forall n$. *Otherwise,* $(0,0)$ *is the NE.*

Proposition 1 means that the two organizations will participate in SFL if and only if both obtain a higher revenue than not participation. If any one of the organization does not participate, then the SFL process ceases to exist. Instead, each organization will use its own data to train a local model. Next, we characterize a somewhat counter-intuitive result on the equilibrium participation.

Theorem 1. *Assume that* $\tilde{\boldsymbol{x}}_{s,n}^T = \boldsymbol{x}_{s,n}^T$, $\forall n$, *and let Assumption 1 hold. There exist scenarios where* $(0,0)$ *is the NE of Game 2, even if choosing* $(1,1)$ *leads to a higher model accuracy for both organizations.*

Theorem 1 implies that if the main server assigns the main-server-side model to each organization without accuracy shaping, then even if participating in SFL improves both organizations' model accuracy, they may choose not to participate.[4] Next, we design an accuracy-shaping mechanism to tackle this issue.

4 Main Server's Accuracy-Shaping Mechanism in Stage I

4.1 Accuracy-Shaping Mechanism

To maximize social welfare, we propose an accuracy-shaping mechanism where the main server can allocate different versions of main-server-side model to organizations. To design an effective mechanism, we need to incorporate the organization contribution. More specifically, let C_n denote the contribution index of organization n, and $\boldsymbol{C} = (C_1, C_2)$. In practice, the main server can use various metrics (e.g., the number of training data, the label distribution, and the gradient divergence) of organizations to estimate their contributions to SFL. Intuitively, an organization with a larger contribution index should be assigned a better model. More specifically, we propose the following mechanism.

Mechanism 1. *Denote the mapping below as the accuracy-shaping mechanism:*

$$\mathcal{M} : \{\boldsymbol{q}, \boldsymbol{C}\} \mapsto \{\epsilon_n, \forall n\}, \tag{13}$$

where ϵ_n *is the model accuracy degradation of organization* n, *and it has the following form:*

$$\epsilon_n(\boldsymbol{q},\boldsymbol{C}) = \begin{cases} 0, & \text{if } q_1 q_2 = 0, \\ \epsilon_0 \left(1 - \frac{C_n}{\max\{C_1,C_2\}}\right), & \text{if } q_1 q_2 = 1, \end{cases} \tag{14}$$

where $\epsilon_0 \geq 0$ *is a tunable parameter.*

[4] We provide an example in Appendix 1.4 to illustrate the intuition.

Mechanism 1 has two intuitions. First, if organizations do not participate in SFL (i.e., $q_1 q_2 = 0$), then trivially the accuracy-shaping mechanism will not be implemented. Second, the organization with a larger contribution index C_n will receive noise-free model from the main server. The other organization with a smaller contribution index will receive a degenerated model. As will be shown, compared to the case where organizations always receive the same noise-free main-server-side model, accuracy shaping can better incentivize training participation and induce higher social welfare at equilibrium.

4.2 Mechanism Properties

Mechanism 1 satisfies a few properties. We start with the definition.

Definition 3. *A mechanism is (i) feasible if $\epsilon_n \geq 0, \forall n$; (ii) individually rational if $R_n(\mathcal{M}, \boldsymbol{q}^*, \boldsymbol{p}^*) \geq 0, \forall n$; (iii) incentive compatible if each organization n achieves a no smaller revenue from participating in SFL than without participation (i.e., local learning).*

Feasibility means that the main server can only degrade the model performance by adding noise to the converged (server-side) model. It cannot improve the model performance due to lack of raw data. Individual rationality means that at equilibrium, each organization can achieve a non-negative payoff. Incentive compatibility means that each organization achieves a no-smaller payoff than when training a local model without SFL.

It is straightforward to show that Mechanism 1 is feasible and individual rational. We are interested in understanding the incentive compatibility property, which is summarized in Theorem 2.

Theorem 2. *Assume that 1) participating in SFL leads to a higher accuracy than local learning for both organizations, and 2) customers' type θ follows a uniform distribution. Then, there exists a positive ϵ_0 such that Mechanism 1 is incentive compatible, and it achieves a higher social welfare than local training.*

Note that even if the analysis of incentive compatibility depends on the second assumption of uniform distribution, our numerical results using other distributions (e.g., normal distribution) are consistent with Theorem 2.

5 Numerical Results

5.1 Simulation Setup

We train ResNet-18 on the CIFAR-10. We consider that the two organizations A and B have $5k$ and $8k$ data, which are sampled using the Dirichlet distribution with a controlling parameter $\beta \in \{0.1, 0.5, 1, \infty\}$. We consider two versions of SFL, i.e., SFL-V1 and SFL-V2 [30], and split the model after the fourth residual block. For the benchmarks, we use vanilla FL (FedAvg), FedProx, MOON, and local learning (LL) where each organization trains a local model using its own

data without FL or SFL. We use 50 communication rounds for SFL and FL. Each experiment is repeated by 3 runs. More details of hyper-parameters and compute resources are in Appendix 2.1.

5.2 Training Results

Table 1 reports the training results including the values of mean and standard deviation. From this table, we make a few observations below. First, all of FL and SFL algorithms consistently outperform local training across all β values. This indicates the effectiveness of collaborative training in leveraging data from multiple sources, compared to training on a single local dataset. Second, as β decreases, there is a general trend of decreasing performance for all algorithms. This is a commonly observed phenomenon in distributed learning. Importantly, SFL consistently outperforms FL and LL. We will show in the following that SFL also incentivizes learning collaboration and achieves a higher social welfare than FL algorithms.

5.3 Equilibrium Results

Now, we use Table 1 (the mean values) to calculate the equilibrium of the four-stage game. We consider that the customers have a linear valuation function, i.e., $V(A) = A$, and their type θ is uniformly distributed on $[0, 10^4]$. For the accuracy-shaping mechanism, we use the number of data samples as the contribution index, i.e., $C_1 = 5k$ and $C_2 = 8k$. Note that the mechanism only applies to SFL-V1 and SFL-V2 in which the main server holds the main-server-side model that is not accessible to the organizations during training. The accuracy-shaping mechanism does not work for FL, because organizations have full access to the entire global model during training. The mechanism does not apply to LL, either.

More specifically, we consider four types of accuracy-shaping mechanisms for SFL-V1 and SFL-V2 using $\epsilon_0 \in \{0, 0.2, 0.3, 0.4\}$, where $\epsilon_0 = 0$ means no accuracy-shaping (similar as in FL). The equilibriums in Stages IV and III can be calculated by Lemma 1 and Eq. (11), respectively. We calculate the equilibrium participation using Proposition 1. We report the equilibrium participation and the social welfare defined in (6) in Table 2. We make a few observations.

1. *Non-participation and low NE outcomes of FL.* From Table 2, we see that FL (e.g., FedAvg and FedProx) leads to non-participation and low social welfare. This observation can be attributed to the intense price competition induced by the shared global model. Despite potential gains in accuracy, the shared model intensifies competition and crucially reduces organization revenues. Hence, organizations prefer local learning over participation in FL. Similar observation has also been made in prior literature [11].

2. *Accuracy-shaping of SFL encourages participation and improves social welfare.* Where there is no accuracy-shaping, SFL faces similar issue to FL, i.e., the increased competition leads to non-participation and hence low social welfare. When there is accuracy-shaping, i.e., $\epsilon_0 \in \{0.2, 0.3, 0.4\}$, SFL significantly boosts participation and social welfare. This is because the tailored accuracy

Table 1. Training results (in %) under different β. For SFL and FL, we report the global model accuracy. For LL, we report the accuracy for both organizations A and B. Given β, we highlight the best score in bold and the second best underlined.

Algorithm	$\beta = \infty$ (IID)	$\beta = 1$	$\beta = 0.5$	$\beta = 0.1$
SFL-V1	<u>86.57</u> ± 0.26	82.69 ± 2.44	81.25 ± 0.80	75.06 ± 1.18
SFL-V2	**86.72** ± 0.42	**85.36** ± 0.56	**83.01** ± 1.70	**75.69** ± 2.84
FedAvg	85.91 ± 0.22	<u>83.92</u> ± 0.66	<u>81.68</u> ± 1.63	<u>75.61</u> ± 2.76
FedProx	84.27 ± 0.19	81.65 ± 0.63	79.97 ± 1.78	72.84 ± 2.83
MOON	83.34 ± 0.47	80.10 ± 1.09	77.40 ± 2.22	69.85 ± 3.23
LL (A)	58.43 ± 4.04	49.24 ± 3.92	43.06 ± 5.42	34.66 ± 5.06
LL (B)	70.08 ± 2.25	50.94 ± 8.67	46.61 ± 13.65	42.94 ± 5.17

Table 2. Equilibrium results based on Table 1. Each cell contains two values: optimal participation (left) and social welfare (right). Here, ✓ indicates participation, × indicates no participation, and − means not applicable. Given β, we highlight the best score in bold and the second best underlined.

Algorithm	$\beta = \infty$ (IID)	$\beta = 1$	$\beta = 0.5$	$\beta = 0.1$
SFL-V1 ($\epsilon_0 = 0.0$)	× | 3416.91	× | 2536.84	× | 2307.46	× | 2082.90
SFL-V1 ($\epsilon_0 = 0.2$)	× | 3416.91	✓ | 4083.08	✓ | 4010.93	× | 2082.90
SFL-V1 ($\epsilon_0 = 0.3$)	✓ | <u>4246.24</u>	✓ | 4051.47	✓ | 3979.16	✓ | <u>3668.24</u>
SFL-V1 ($\epsilon_0 = 0.4$)	✓ | 4211.70	✓ | 4016.42	✓ | 3943.92	✓ | 3632.06
SFL-V2 ($\epsilon_0 = 0.0$)	× | 3416.91	× | 2536.84	× | 2307.46	× | 2082.90
SFL-V2 ($\epsilon_0 = 0.2$)	× | 3416.91	✓ | **4216.83**	✓ | **4099.11**	× | 2082.90
SFL-V2 ($\epsilon_0 = 0.3$)	✓ | **4253.78**	✓ | <u>4185.50</u>	✓ | <u>4067.53</u>	✓ | **3699.90**
SFL-V2 ($\epsilon_0 = 0.4$)	✓ | 4219.26	✓ | 4150.80	✓ | 4032.53	✓ | 3663.82
FedAvg	× | 3416.91	× | 2536.84	× | 2307.46	× | 2082.90
FedProx	× | 3416.91	× | 2536.84	× | 2307.46	× | 2082.90
MOON	× | 3416.91	× | 2536.84	× | 2307.46	× | 2082.90
LL	− | 3416.91	− | 2536.84	− | 2307.46	− | 2082.90

reduces competition, increases organizational revenues, and makes participation economically appealing. This also contributes to a higher social welfare, as the customers are receiving model-based services with higher qualities (compared to local learning). The social welfare improvement can be up to 66.38%, e.g., SFL-V2 with $\epsilon_0 = 0.2$ at $\beta = 1$.

5.4 More experiments and discussions

We have included more experiments on 1) two different datasets CIFAR-100 and HAM10000, and 2) a different customer scenario with quadratic valuation function and Guassian distributed types. The results are reported in Appendices 2.2-2.3 and consistent with our observations.

So far our results are based on the canonical duopoly competition model. When there are more than two organizations, we can similarly analyze the equilibrium in Stages IV and III. However, the optimal participation in Stage II can

be challenging, and one may apply the coalitional game theory (e.g., as in [3]) to this end. We provide a more detailed discussion in Appendix 3.1.

6 Conclusion

This work studied the complex dynamics of competitive distributed learning, where organizations aim to collaboratively develop ML models while competing for the same base of customers. We showed that sharing the global model can be inefficient in competitive scenarios, as it inadvertently increases competition rather than collaboration. To address this issue, we used SFL and proposed a tailored accuracy-sharing mechanism. Upon convergence, the mechanism induces tailored noise into the main server's model, which facilitates the differentiation of models for each organization. Both theoretical and numerical results show that our proposed mechanism incentivizes collaboration and significantly improves the social welfare, compared to existing FL benchmarks.

For future work, it is interesting to develop more robust SFL algorithms under data heterogeneity. It is also interesting to incorporate privacy enhancing techniques (e.g., differential privacy) into the accuracy-shaping mechanism design.

References

1. https://www.swissre.com/en/china/.
2. https://www.dropbox.com/scl/fi/jjt7evqsp48uah7lsap17/ICANN$_A$$ppendix.pdf?rlkey$ = 53107$mqmizdnc8r9u2l6xeia8dl$ = 0
3. Donahue, K., Kleinberg, J.: Model-sharing games: Analyzing federated learning under voluntary participation. In: Proc. of AAAI. vol. 35, pp. 5303–5311 (2021)
4. Donahue, K., Kleinberg, J.: Optimality and stability in federated learning: A game-theoretic approach. Advances in Neural Information Processing Systems **34**, 1287–1298 (2021)
5. Dorner, F.E., Konstantinov, N., Pashaliev, G., Vechev, M.: Incentivizing honesty among competitors in collaborative learning and optimization. in Proc. of NeurIPS **36** (2024)
6. Gradwohl, R., Tennenholtz, M.: Coopetition against an amazon. Journal of Artificial Intelligence Research **76**, 1077–1116 (2023)
7. Han, D.J., Bhatti, H.I., Lee, J., Moon, J.: Accelerating federated learning with split learning on locally generated losses. In: ICML workshop on federated learning for user privacy and data confidentiality. ICML Board (2021)
8. Han, P., Huang, C., Tian, G., Tang, M., Liu, X.: Convergence analysis of split federated learning on heterogeneous data. arXiv preprint arXiv:2402.15166 (2024)
9. Huang, C., Dachille, J., Liu, X.: When federated learning meets oligopoly competition: Stability and model differentiation. IEEE Internet of Things Journal (2024)
10. Huang, C., Huang, J., Liu, X.: Cross-silo federated learning: Challenges and opportunities. arXiv preprint arXiv:2206.12949 (2022)
11. Huang, C., Ke, S., Liu, X.: Duopoly business competition in cross-silo federated learning. IEEE Transactions on Network Science and Engineering **11**(1), 340–351 (2024)
12. Huang, C., Tang, M., Ma, Q., Huang, J., Liu, X.: Promoting collaborations in cross-silo federated learning: Challenges and opportunities. IEEE Communications Magazine **62**(4), 82–88 (2024)

13. Jain, A.: Sharing demand information with retailer under upstream competition. Management Science **68**(7), 4983–5001 (2022)
14. Ji, X., Zhu, Z., Xi, W., Gadyatskaya, O., Song, Z., Cai, Y., Liu, Y.: Fedfixer: Mitigating heterogeneous label noise in federated learning. In: Proc. of AAAI. vol. 38, pp. 12830–12838 (2024)
15. Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al.: Advances and open problems in federated learning. Foundations and Trends® in Machine Learning **14**(1–2), 1–210 (2021)
16. Karimireddy, S.P., Guo, W., Jordan, M.I.: Mechanisms that incentivize data sharing in federated learning. arXiv preprint arXiv:2207.04557 (2022)
17. Ke, S., Huang, C., Liu, X.: On the impact of label noise in federated learning. In: Proc. of IEEE WiOpt. pp. 183–190 (2023)
18. Khan, M.A., Shejwalkar, V., Houmansadr, A., Anwar, F.M.: Security analysis of splitfed learning. In: Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems. pp. 987–993 (2022)
19. Li, J., Rakin, A.S., Chen, X., He, Z., Fan, D., Chakrabarti, C.: Ressfl: A resistance transfer framework for defending model inversion attack in split federated learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10194–10202 (2022)
20. Li, J., Rakin, A.S., Chen, X., Yang, L., He, Z., Fan, D., Chakrabarti, C.: Model extraction attacks on split federated learning. arXiv preprint arXiv:2303.08581 (2023)
21. Li, Q., He, B., Song, D.: Model-contrastive federated learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10713–10722 (2021)
22. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems **2**, 429–450 (2020)
23. Liu, J., Huang, J., Zhou, Y., Li, X., Ji, S., Xiong, H., Dou, D.: From distributed machine learning to federated learning: A survey. Knowledge and Information Systems **64**(4), 885–917 (2022)
24. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. pp. 1273–1282. PMLR (2017)
25. Mu, Y., Shen, C.: Communication and storage efficient federated split learning. arXiv preprint arXiv:2302.05599 (2023)
26. Shen, J., Cheng, N., Wang, X., Lyu, F., Xu, W., Liu, Z., Aldubaikhy, K., Shen, X.: Ringsfl: An adaptive split federated learning towards taming client heterogeneity. IEEE Transactions on Mobile Computing (2023)
27. Son, H.M., Kim, M.H., Chung, T.M., Huang, C., Liu, X.: Feduv: Uniformity and variance for heterogeneous federated learning. In Proc. of CVPR (2024)
28. Sun, P., Che, H., Wang, Z., Wang, Y., Wang, T., Wu, L., Shao, H.: Pain-fl: Personalized privacy-preserving incentive for federated learning. IEEE Journal on Selected Areas in Communications **39**(12), 3805–3820 (2021)
29. Tang, M., Wong, V.W.: An incentive mechanism for cross-silo federated learning: A public goods perspective. In: Proc. of IEEE INFOCOM (2021)
30. Thapa, C., Arachchige, P.C.M., Camtepe, S., Sun, L.: Splitfed: When federated learning meets split learning. In: Proceedings of the AAAI. vol. 36, pp. 8485–8493 (2022)
31. Vepakomma, P., Gupta, O., Dubey, A., Raskar, R.: Reducing leakage in distributed deep learning for sensitive health data. arXiv preprint arXiv:1812.00564 **2** (2019)

32. Vepakomma, P., Gupta, O., Swedish, T., Raskar, R.: Split learning for health: Distributed deep learning without sharing raw patient data. arXiv preprint arXiv:1812.00564 (2018)
33. Wu, X., Yu, H.: Mars-fl: Enabling competitors to collaborate in federated learning. IEEE Transactions on Big Data (2022)
34. Xu, J., Chen, Z., Quek, T.Q., Chong, K.F.E.: Fedcorr: Multi-stage federated learning for label noise correction. In: Proc. of IEEE/CVF CVPR. pp. 10184–10193 (2022)
35. Yan, Y., Tang, X., Huang, C., Tang, M.: Price of stability in quality-aware federated learning. In: Proc. of IEEE GLOBECOM. pp. 734–739 (2023)
36. Ye, M., Fang, X., Du, B., Yuen, P.C., Tao, D.: Heterogeneous federated learning: State-of-the-art and research challenges. ACM Computing Surveys **56**(3), 1–44 (2023)