Competitive-Cooperative Multi-Agent Reinforcement Learning for Auction-based Federated Learning

Xiaoli Tang and **Han Yu**
School of Computer Science and Engineering, Nanyang Technological University, Singapore
{xiaoli001, han.yu}@ntu.edu.sg

Abstract

Auction-based Federated Learning (AFL) enables open collaboration among self-interested data consumers and data owners. Existing AFL approaches cannot manage the mutual influence among multiple data consumers competing to enlist data owners. Moreover, they cannot support a single data owner to join multiple data consumers simultaneously. To bridge these gaps, we propose the Multi-Agent Reinforcement Learning for AFL (MARL-AFL) approach to steer data consumers to bid strategically towards an equilibrium with desirable overall system characteristics. We design a temperature-based reward reassignment scheme to make trade-offs between cooperation and competition among AFL data consumers. In this way, it can reach an equilibrium state that ensures individual data consumers can achieve good utility, while preserving system-level social welfare. To circumvent potential collusion behaviors among data consumers, we introduce a bar agent to set a personalized bidding lower bound for each data consumer. Extensive experiments on six commonly adopted benchmark datasets show that MARL-AFL is significantly more advantageous compared to six state-of-the-art approaches, outperforming the best by 12.2%, 1.9% and 3.4% in terms of social welfare, revenue and accuracy, respectively.

1 Introduction

Due to user privacy and data confidentiality requirements, Federated Learning (FL) has recently attracted significant research interest from academia and industry alike [Yang *et al.*, 2019; Liu *et al.*, 2020; Liu *et al.*, 2022; Lyu *et al.*, 2022]. As data owners (a.k.a. FL clients) are self-interested entities who need to take a complex set of considerations (e.g., costs, potential utility gains) into account to determine which FL data consumer to join, FL incentive mechanism design [Khan *et al.*, 2020; Zhan *et al.*, 2021] has been brought to the forefront to motivate them to join FL with rewards.

Auction-based federated learning (AFL) constitutes an important category of FL incentive mechanism design research due to their promising ability to achieve efficiency and fairness. Nevertheless, AFL is still in its infancy. Existing works [Jiao *et al.*, 2020; Zeng *et al.*, 2020; Ying *et al.*, 2020] mostly focus on designing data consumer-data owner matching and winner payment schemes to achieve desired objectives (e.g., social welfare maximization, social cost minimization) for the FL ecosystem. However, it is challenging to design bidding strategies for data consumers to guide their bids for data owners while preserving the health of the FL ecosystem, especially when multiple data consumers are recruiting from the same pool of data owners. Firstly, it is imperative to strike a trade-off between cooperation and competition relations among data consumers. On one hand, in a fully competitive AFL marketplace, the objective of the most competitive bidding data consumer is optimized at the expense of others, leading to unfairness in the FL auction ecosystem. On the other hand, a fully cooperative AFL ecosystem may sacrifice the objectives of some data consumers for system-level social welfare [Yu *et al.*, 2017], thereby hurting the motivation of these data consumers to participate in the future. Secondly, data consumers might collude (e.g., bidding extremely low prices together) to improve their own payoff.

To address these limitations, we model the AFL ecosystem as a multi-agent system and propose the Multi-Agent Reinforcement Learning for AFL (MARL-AFL) approach. It learns bidding strategies for data consumers, with the aim of coordinating their bidding behaviors towards an equilibrium that ensures fairness towards all data consumers in the FL ecosystem, while guarding against collusion among data consumers. Firstly, to enhance fair treatment towards the data consumers in an FL ecosystem [Shi *et al.*, 2023], we design a temperature-based reward reassignment mechanism for MARL-AFL, which reassigns the total auction reward among all participating data consumers in accordance with their contributions (which is captured by a temperature-based softmax function). Secondly, to dissuade collusive behaviors among data consumers (i.e., bidding lower bid price maliciously), we introduce a bar agent for each bidding agent of the data consumer to learn the personalized bidding patterns [Tan *et al.*, 2022]. Different from the bidding agents which aim to lower the cost for their data consumers, the bar agents aim to increase the bidding lower bounds to boost overall ecosystem revenue. These two types of agents are trained in an adversarial manner until an equilibrium is reached.

To the best of our knowledge, MARL-AFL is the first multi-agent reinforcement learning approach to support multiple data consumers to compete to recruit from a same pool of data owners, while allowing each data owner to join multiple data consumers simultaneously. Extensive experiments based on six commonly adopted benchmark datasets show that MARL-AFL is more advantageous than six state-of-the-art approaches, outperforming the best by 12.2%, 1.9% and 3.4% in terms of average social welfare, revenue and model accuracy, respectively.

2 Related Works

The most related domain to our research is AFL incentive mechanism design.

The combinatorial and double auction-based approaches [Krishnaraj *et al.*, 2022; Zavodovski *et al.*, 2019; Hong *et al.*, 2020; Yang, 2020; Bahreini *et al.*, 2018; Gao *et al.*, 2019; Jiao *et al.*, 2018; Jiao *et al.*, 2019] are designed to help the FL auctioneer to achieve desired goals (e.g., maximizing social welfare, minimizing social cost). They can support scenarios involving multiple data consumers and data owners. In [Yang, 2020], a multi-round sequential combination auction model was adopted to allocate data owners with limited resources to data consumers with heterogeneous resource requirements. Under this approach, data consumers first publicize their respective resource requirements and bidding values for different data owners in a sequential manner. The data consumer-data owner matching and corresponding payments are then optimized.

The reverse auction-based approaches are designed to help the data consumer recruit data owners, while achieving desirable goals (e.g., utility maximization) [Jiao *et al.*, 2020; Zeng *et al.*, 2020; Ying *et al.*, 2020; Le *et al.*, 2020; Thi Le *et al.*, 2021; Zhang *et al.*, 2021; Roy *et al.*, 2021; Deng *et al.*, 2021; Zhang *et al.*, 2022a; Zhang *et al.*, 2022b]. They can be combined with various techniques such as reputation, blockchain, deep reinforcement learning and graph neural networks. Generally, these methods are designed for a monopoly market (i.e., with only one data consumer and multiple data owners). For instance, in [Zhang *et al.*, 2021], RRAFL was proposed by incorporating reputation and blockchain into reverse auction. In RRAFL, the data consumer first publicizes its FL task. Then, the data owners bid for it. After receiving the bids from all data owners, the data consumer determines the winning data owners based on their reputation values which are derived based on their reliability and data quality [Chen *et al.*, 2020] track records in the blockchain.

Nevertheless, existing works do not support multiple data consumers bidding for data owners in a competitive AFL marketplace, nor do they support an individual data owner to join multiple FL tasks simultaneously if its local computational resources allow. MARL-AFL bridges these gaps.

3 Preliminaries

AFL typically includes three types of participants: 1) data owners with useful but potentially sensitive data; 2) data consumers who require data to build AI models; and 3) an FL auctioneer who matches data owners with data consumers. The auctioneer triggers an auction for interested data consumers to bid for a given data owner with the required data resources. Each data consumer sets the bid value according to its objective and constraints, as well as the utility of the data owner's data resources. After receiving the bids from all interested data consumers, the auctioneer follows an auction mechanism (e.g., the first-price auction, the second-price auction) to determine the winner and the payment. When there is no eligible data owners left, the procedure terminates. Then, each data consumer initiates FL model training with the data owners it has recruited.

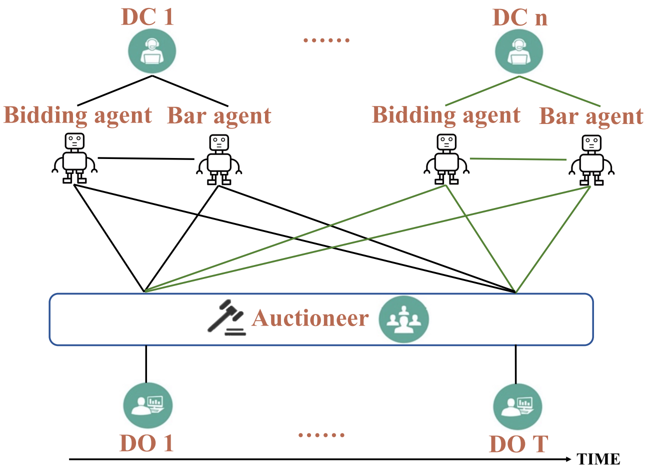

Figure 1: The MARL-AFL system architecture. DC and DO represent data consumer and data owner, respectively.

In this paper, we assume that there are n data consumers competing to recruit data owners at any point in time in an AFL ecosystem. We further assume that a data owner can simultaneously join more than one data consumer for FL model training as long as its local computational resources can handle the load. For a data consumer $i \in [1, n]$, its objective is to maximize its accumulated profit (i.e., the utility of attracting data owners) within its budget B_i in a competitive setting. However, whether a data consumer wins or not hinges on the bidding strategies of all the data consumers involved. In this sense, competitive AFL is inherently a multi-agent system (MAS) with each agent acting on behalf of a data consumer.

We formulate the interaction among these n bidding agents as a Partially Observable Markov Decision Process (POMDP) [Littman, 1994] $\langle \mathcal{S}, P, \{Z_i, \mathcal{O}_i, \mathcal{A}_i, r_i\}_{i=1}^{n}, \gamma \rangle$. In the context of AFL, at each time step t, a bidding agent i determines the bid price $b_i^t = \pi_i(o_i^t)$ for its data consumer according to the policy π_i based on observation $o_i^t = (B_i^t, v_i^t, ts_i^t)$, where B_i^t is the remaining budget, v_i^t is the utility of the data resources for i, and $ts_i^t = T - t$ is the remaining time steps in this episode that comprises T auctions. The reward for the winning bidder is v_i^t (i.e., $r_i^t = v_i^t$), and the payment p^t is the second-highest bid among all the bids received. Then, each agent obtains its new observation $o_i^{t+1} = (B_i^t - p^t \times x_i^t, v_i^{t+1}, ts_i^t - 1)$ with $x_i^t \in \{0, 1\}$ indicating whether it has won the current auction. The objective of i is to maximize its accumulated expected utility:

$$\max_{\pi_i} \mathbb{E}[\sum_{t=1}^{T} \gamma^{t-1} r_i^t]. \tag{1}$$

The system architecture of MARL-AFL is shown in Fig. 1.

4 The Proposed MARL-AFL Approach

We adopt a deep neural network (DNN) to model the action-value function $Q_i(o_i, b_i)$ of agent i, parameterized by θ_i and conditioned on the agent's observations o_i and bids b_i. To improve stability during training, we pair this network with a similar DNN architecture parameterized by $\hat{\theta}_i$, known as the target network, which also approximates $Q_i(o_i, b_i)$. To update θ_i, we minimize the following loss function:

$$\mathcal{L}(\theta_i) = \frac{1}{2} \mathbb{E}_{(o_i, b_i, r_i, o'_i) \sim \mathcal{D}} [(y_i - Q_i(o_i, b_i; \theta_i))^2]. \tag{2}$$

The replay buffer $\mathcal{D}$ is a storage mechanism for transition tuples $(o_i, b_i, r_i, o'i)i = 1^n$, where o'_i is the new observation of agent i following its bid b_i based on the observation o_i, resulting in reward r_i. This buffer allows the agent to learn from its past experiences by randomly sampling batches of transitions during training. In the loss function defined in E-quation (2), y_i represents the temporal difference target and is computed as $y_i = r_i + \gamma \max_{b'_i} Q(o'_i, b'_i; \hat{\theta}_i)$, where γ is the discount factor, $\hat{\theta}_i$ represents the parameters of the target network associated with agent i, and $Q(o'_i, b'_i; \hat{\theta}_i)$ is the predicted action-value function of agent i for its next observation o'_i and all possible bids b'_i. This target network is used to stabilize the learning process by providing a fixed target during training, which is updated periodically to match the current action-value network.

As shown in Eq. (1), the objective of data consumer i is to maximize its accumulated reward (i.e., the accumulated expected utility of the data resources it recruited). In the following sections, we first describe how to estimate the utility of the data resources at each time step. Then, we describe how to model the cooperation-competition relationships among data consumers. Finally, we analyze how to set personalized lower bounds for bids to mitigate potential collusive behaviors among data consumers.

Algorithm 1 Learning Θ_i in Eq. (3)

```
INPUT: Training Data $H$, training round $Z$, learning rate $\eta_\Theta$
OUTPUT: Utility estimation function $v_i(\cdot)$
for $z = 1$ to $Z$ do
    for sample $(\boldsymbol{q}_m, y_m)$ in $H$ do
        Calculate the gradient of $\Theta_i$ according to Eq. (6)
        Update $\Theta_i$ according to Eq. (5)
    end for
end for
return $v_i(\cdot)$
```

4.1 Utility Estimation

Following [Zhan *et al.*, 2020], we define the utility estimation function of data consumer i with respect to data resources being auctioned at time step t in the form of

$$v_i^t = v_i(\boldsymbol{q}_t) \triangleq \ln(1 + \Theta_i^T \boldsymbol{q}_t), \tag{3}$$

where Θ_i represents the learnable parameter. Θ_i is a high-dimensional feature vector the entries of which comprise features related to the data resources at time step t (e.g., identities of the data owners, data quantity). For clarity, we denote $v_i(\boldsymbol{q}_t)$ as $v_i(\cdot)$ in subsequent derivations.

Over multiple rounds of auctions, data consumers in an AFL ecosystem accumulate historical data H which can be used to derive the utility of the current data resources. Such data can be recorded in the form of $(\boldsymbol{q}_m, y_m)$ ($(\boldsymbol{q}_m, y_m) \in H$), where y_m denotes the real utility of data resources at time step m. Then, we leverage the squared error (SE) loss [Ren *et al.*, 2017] to train the utility estimation function $v_i(\cdot)$,

$$\mathcal{L}(v_i(\cdot)) = \frac{1}{2} \sum_{(\boldsymbol{q}_m, y_m)} (y_m - v_i(\boldsymbol{q}_m))^2. \tag{4}$$

The parameter Θ_i in Eq. (3) can be obtained via gradient descent:

$$\Theta_i \leftarrow \Theta_i - \eta_{\Theta_i} \frac{\partial \mathcal{L}(v_i(\cdot))}{\partial \Theta_i}, \tag{5}$$

where η_Θ is the learning rate. $\frac{\partial \mathcal{L}(s_j(\cdot))}{\partial \Theta_j}$ is derived as:

$$\frac{\partial \mathcal{L}(v_i(\cdot))}{\partial \Theta_i} = \sum_{(\boldsymbol{q}_m, y_m)} [\ln(1 + \Theta_i^T \boldsymbol{q}_m) - y_m] \frac{\boldsymbol{q}_m}{1 + \Theta_i^T \boldsymbol{q}_m}. \tag{6}$$

The detailed process for learning θ_i in the utility estimation function is shown in Algorithm 1.

4.2 Cooperation-Competition Modeling

To establish the cooperation and competition among bidding agents, we design a temperature-based reward reassignment mechanism. This mechanism is based on setting a parameter α_i that weights the contribution of each agent i to the total reward. Specifically, the reassigned reward for agent i, denoted as r_i^{reassign}, is computed as

$$r_i^{\text{reassign}} = \alpha_i \times r^{\text{total}}, \tag{7}$$

where

$$\alpha_i = \frac{\exp\{b_i/\tau\}}{\sum_{j=1}^{n} \exp\{b_j/\tau\}}, \tag{8}$$

and r^{total} is the sum of rewards obtained by all agents and is formulated as:

$$r^{\text{total}} = \sum_{i=1}^{n} r_i. \tag{9}$$

In Eq. (8), τ balances the level of cooperation and competition, with higher bids having a greater impact on the value of r^{total}. Consequently, the reassigned reward for agent i increases with its bid b_i, which is directly proportional to the revenue received by each agent. Incorporating τ in our approach has two key advantages.

1. Firstly, it enables the reward function to capture both social welfare and revenue considerations. Eq. (7) shows how the social welfare component is incorporated into the reward function. Meanwhile, the bids submitted by each agent implicitly express the revenue generated by the auction, since revenue is directly proportional to bids under the generalised second-price auction. This allows the reward function to take into account both social welfare and revenue, resulting in a more comprehensive measure of performance.
2. Secondly, the use of τ in Eq. (8) allows for a balance between competition and cooperation. By adjusting the value of τ, we can trade off between these two relationships in a flexible and convenient way. This allows for a more nuanced and adaptable approach to multi-agent learning, which can be particularly useful in complex and dynamic environments where the relationship between agents may change over time. By using τ to balance competition and cooperation, we can achieve better overall performance and avoid undesirable outcomes such as collusion or excessive competition.

4.3 Bidding Lower Bounds

The cooperation enhances social welfare. However, it might lead to collusion among bidding agents (e.g., to lower bid prices together), which can reduce the revenue of the AFL ecosystem. In the following, we describe how MARL-AFL addresses such behaviors with Bidding Lower Bounds.

Taking into consideration that the Bidding Lower Bound needs to be flexible to achieve satisfactory performance while accommodating the budgets and utilities of different bidding agents, we introduce a bar agent with policy $\bar{\pi}_i$ to set a personalized Bidding Lower Bound $\bar{b}_i$ for each bidding agent i with policy π_i. The observation of $\bar{\pi}_i$ is identical to that of policy π_i. At each time step t, the bar agents generate the Bidding Lower Bounds $\{\bar{b}_i\}_{i=1}^{n}$, while the bidding agents generate the bids $\{b_i\}_{i=1}^{n}$. Then, the bids $\{b_i\}_{i=1}^{n}$ are submitted to the auctioneer as the bid response while the lower bounds $\{\bar{b}_i\}_{i=1}^{n}$ are kept locally to restrict the minimum bidding limit. After the payment p^t is determined by the auctioneer and rewards $\{r_i\}_{i=1}^{n}$ are obtained, $\{r_i^{\text{reassign}}\}_{i=1}^{n}$ is reassigned according to Eq. (7). Next, we employ an indicator function $z_i = z(\bar{b}_i, b_i) \in \{0,1\}$ to determine if the bid generated by the bidding agent is higher than the Bidding Lower Bound set by the corresponding bar agent:

$$z(\bar{b}_i, b_i) = \begin{cases} 1, & \text{if } b_i \geqslant \bar{b}_i, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Then, the rewards for π_i and $\bar{\pi}_i$ are

$$r_i = z(\bar{b}_i, b_i) \times r_i^{\text{reassign}}, \tag{11}$$

and

$$\bar{r}_i = z(\bar{b}_i, b_i) \times p^t. \tag{12}$$

The bidding agents and their corresponding bar agents are trained adversarially at the same time: $\bar{\pi}_i$ is trained to optimize the Bidding Lower Bound $\bar{b}_i$ so as to raise the ecosystem's revenue, while π_i attempts to reduce the bid price to set aside more budget when colluding with others. $z(\bar{b}_i, b_i)$ ties these two disparate objectives together by enforcing the Bidding Lower Bound $\bar{b}_i$ generated by the bar agent to be the highest lower bound of the bid b_i output by the corresponding bidding agent.

The detailed training process of the bidding agents and bar agents of MARL-AFL is provided in Algorithm 2.

Algorithm 2 MARL-AFL

```
Initialize $Q_i(o_i, b_i; \theta_i)$ and $\bar{Q}_i(o_i, \bar{b}_i; \bar{\theta}_i)$ and their target networks with parameters $\hat{\theta}_i \leftarrow \theta_i$ and $\hat{\bar{\theta}}_i \leftarrow \bar{\theta}_i$, for $\forall i \in [1, n]$; the replay memory $\mathcal{D}$, the max-episode, and training batch size $m$, the update frequency of target networks $C$
for episode $= 1$ to max-episode do
    for $t = 1$ to $T$ do
        for each agent $i$ do
            Compute $b_i$ according to $\epsilon$-greedy policy w.r.t $Q_i$
            Compute $\bar{b}_i$ according to $\epsilon$-greedy policy w.r.t $\bar{Q}_i$
            Submit $b_i$ to the auctioneer
        end for
        Obtain rewards $\{r+1, \cdots, r_n\}$ and the payment $p$
        Calculate $r_i^{\text{reassign}}$ based on Eq. (7)
        Obtain $z_i$ based on Eq. (10)
        Obtain $r_i$ and $\bar{r}_i$ based on Eq. (11) and Eq. (12), respectively
        Store transition tuples of all agents in $\mathcal{D}$
        for each agent $i$ do
            Sample a random minibatch of $m$ samples from $\mathcal{D}$
            $y_i = r_i + \gamma \max_{b_i'} Q(o_i', b_i'; \hat{\theta}_i)$
            $\bar{y}_i = \bar{r}_i + \gamma \max_{\bar{b}_i'} \bar{Q}(o_i', \bar{b}_i'; \hat{\bar{\theta}}_i)$
            Update $\theta_i$ by minimizing $\sum_m [(y_i - Q_i(o_i, b_i; \theta_i))^2]$
            Update $\bar{\theta}_i$ by minimizing $\sum_m [(\bar{y}_i - \bar{Q}_i(o_i, \bar{b}_i; \bar{\theta}_i))^2]$
            $\hat{\theta}_i \leftarrow \theta_i, \hat{\bar{\theta}}_i \leftarrow \bar{\theta}_i$ every $C$ episodes
        end for
    end for
end for
```

5 Experimental Evaluation

5.1 Experiment Setup

To evaluate the performance of MARL-AFL, we conduct experiments based on six commonly used datasets in FL studies, including MNIST (http://yann.lecun.com/exdb/mnist/), CIFAR-10 (https://www.cs.toronto.edu/kriz/cifar.html), Fashion-MNIST (i.e., FMNIST) [Xiao *et al.*, 2017], EMNIST-digits (i.e., EMNIST-D), EMNIST-letters (i.e., EMNIST-L) [Cohen *et al.*, 2017] and Kuzushiji-MNIST (i.e., KMNIST) [Clanuwat *et al.*, 2018]. In each experiment, each data owner has a training set size that is determined at random be between 1,000 and 10,000 samples. Both the test set and the validation set for each data consumer include 2,000 samples. Following [Zhang *et al.*, 2021], tasks on MNIST, FMNIST, EMNIST-D and KMNIST are processed by a base model with an input layer containing 784 nodes, a hidden layer containing 50 nodes and an output layer containing 10 nodes. For tasks on EMNIST-L, the base model is similar to the aforementioned network but with the output layer having 26 nodes. The tasks on CIFAR-10 are processed by the streamlined VGG11 network [Simonyan and Zisserman, 2015], where the number of convolutional filters and the size of the hidden fully-connected layers are $\{32, 64, 128, 128, 128, 128, 128, 128\}$ and 128, respectively.

The proposed method utilizes fully connected neural networks with three hidden layers each containing 64 nodes to generate bid prices for data owners on behalf of their respective data consumers. The action-value functions Q_i and $\bar{Q}_i$ are trained using a replay buffer $\mathcal{D}$ with a size of 5,000. During training, the agents explore the environment using an ϵ-greedy policy with an annealing rate from 1.0 to 0.05. To update Q_i, 32 episodes uniformly sampled from $\mathcal{D}$ are used for each training step, and $\bar{Q}_i$ is updated twice after each episode to speed up convergence. The target networks of Q_i and $\bar{Q}_i$ are updated once every 20 training episodes. We use RMSprop with a learning rate of 0.0005 to train all neural networks, and set the discount factor γ to 0.99 and the temperature hyperparameter τ to 4. As discussed in Section 3, the bid price b_i^t of each agent i is determined based on observations of the utility of each data owner. We adopt the utility evaluation method in [Zhang *et al.*, 2021] to compute the utility of each data owner.

5.2 Comparison Approaches

We compare MARL-AFL with the following six state-of-the-art bidding strategies experimentally:

1. **Constant Bid (Const)** [Zhang *et al.*, 2014]: Each data consumer offers the same bid for all data owners. The bid value by each data consumer can differ.
2. **Randomly Generated Bid (Rand)** [Zhang *et al.*, 2021; Zhang *et al.*, 2022b]: This strategy is commonly used in AFL. Under this approach, each data consumer randomly generates a bid from a fixed range of values for each bid request.
3. **Below Max Utility Bid (Bmub)**: This approach is adapted from bidding below max eCPC [Lee *et al.*, 2012] from the field of advertising. The utility of each bid request from a data owner is set as the upper bound of the bid values from data consumers. For each bid request, the bid price is randomly generated within the range of 0 and this upper bound.
4. **Linear-Form Bid (Lin)** [Perlich *et al.*, 2012]: The bid values generated by data consumers are directly proportional to the estimated utility of the bid requests, which can be formulated as $b^{Lin}(v_i^t) = \lambda_{Lin} v_i^t$.
5. **Bidding Machine (BM)** [Ren *et al.*, 2017]: This approach is commonly used in the field of online advertising, especially real-time bidding advertising. It maximizes the profit of one specific advertiser by jointly optimising the click-through rate prediction, ad cost estimation and the bidding strategy.
6. **Reinforcement Learning-based Bid (RLB)** [Cai *et al.*, 2017]: This method treats the bid decision process in the field of online advertising as a reinforcement learning problem by adopting a Markov Decision Process framework to learn the optimal bidding policy for one specific advertiser to maximize the number of clicks on its ads.

As there is no public dataset related to AFL yet, in our experiments, we track the behaviours of the agents over time during the simulations to gradually accumulate such data. Specifically, we create four settings, each of which contains 10,000 data owners and 120 data consumers. Data consumers under each setting adopt one of the bidding strategies listed in the Compared Approaches section. Under the first setting, the percentage of data consumers adopting each of the six baseline bidding methods is one sixth. Under the second and third settings, considering that MARL-AFL is based on multi-agent reinforcement, we set the percentage of data consumers adopting the reinforcement learning-based baseline bidding approach (i.e., RLB) higher than those adopting the other five baselines, with 40 data consumers adopting RLB and 16 data consumers adopting each of the other five baselines under the second setting; while 60 data consumers adopting RLB and 12 data consumers adopting each of the other five baselines under the third setting. Considering that BM and RLB are equipped with AI techniques similar to MARL-AFL, under the fourth setting, data consumers adopting either of these two bidding strategies account for one third of the total population, while data consumers adopting each of the remaining four baselines accounting for one twelfth of the total population. We adopt the generalised second-price sealed-bid forward auction mechanism.

In our experiments, data consumers assess the utility of each data owner from the perspective of both the quantity and the quality of local data resources, the latter of which is set to be related to their IDs. Specifically, we first assign a sequential ID number to each data owner. Then, the data resources of the data owners with IDs ranked in the first half of the population are added with Gaussian noise, while those of the rest of the data owners remain unchanged. Therefore, data resources provided by the first half of the data owners are of low quality, while those provided by the second half are of high quality. Based on the estimated utility of the auctioned data resources, each data consumer can join the auction process by submitting a bid price calculated according to its bidding strategy. After that, each data consumer utilizes the data resources it obtained through the current auction to train its FL model. Then, based on the actual utility of each winning record obtained after training the FL model, each data consumer trains its own utility estimation function following Algorithm 1.

To evaluate the effectiveness of the proposed MARL-AFL, we create seven AFL ecosystems, each of which includes five data consumers (to support FL task reassignment) and adopts the generalised-second price auction mechanism to determine the winners and payments. Data consumers from a given ecosystem adopt one of the aforementioned bidding approaches to compete for the same pool of data owners.

Method	MNIST		CIFAR-10		FMNIST		EMNIST-D		EMNIST-L		KMNIST	
	SW	Rev	SW	Rev	SW	Rev	SW	Rev	SW	Rev	SW	Rev
Const	283.3	375.7	568.6	648.1	284.6	374.2	102.3	573.9	126.8	211.7	386.3	341.5
Rand	201.4	314.1	178.3	537.5	486.1	517.5	400.4	279.5	94.8	158.3	149.6	216.3
Bmub	774.6	528.5	601.7	433.3	835.9	795.6	723.8	712.4	595.8	497.7	713.6	616.2
Lin	928.2	669.8	734.2	652.6	886.1	702.6	804.7	668.2	745.4	595.4	726.6	684.4
BM	1,001.0	893.7	1,092.5	**885.2**	1,086.3	797.4	953.8	**901.6**	996.3	851.7	1,168.7	804.6
RLB	1,056.1	865.1	1,035.6	724.7	946.2	807.5	974.6	897.4	1,035.1	858.9	1,257.2	842.7
MARL-AFL	**1,094.7**	**903.3**	**1,204.2**	857.8	**1,300.4**	**855.8**	**1,131.1**	887.9	**1,245.2**	**901.6**	**1,295.2**	**873.8**

Table 1: Auctioning performance comparison on different datasets. SW denotes the social welfare (the higher, the better), while Rev denotes the revenue of the AFL ecosystem (the higher, the better).

Method	MNIST		CIFAR-10		FMNIST		EMNIST-D		EMNIST-L		KMNIST	
	IID	NIID	IID	NIID	IID	NIID	IID	NIID	IID	NIID	IID	NIID
Const	72.5	65.6	31.8	15.7	70.1	55.4	73.8	61.5	63.6	53.4	56.2	53.6
Rand	77.3	67.1	37.9	15.3	69.3	51.2	72.4	65.9	66.2	55.5	53.8	49.5
Bmub	78.6	70.5	40.6	20.1	71.9	63.0	74.1	72.2	70.7	64.8	57.2	52.8
Lin	78.7	72.8	43.1	19.5	72.7	66.7	78.9	71.6	71.8	64.2	62.9	58.3
BM	79.7	74.4	43.5	22.3	71.7	65.2	81.2	72.8	73.1	65.5	64.8	57.9
RLB	81.1	73.8	44.7	20.8	73.2	67.2	81.4	72.6	72.7	66.9	63.2	58.8
MARL-AFL	**83.6**	**75.3**	**46.2**	**24.9**	**74.9**	**67.8**	**82.8**	**74.5**	**75.2**	**68.3**	**66.8**	**61.2**

Table 2: Test accuracy (%) comparison of the FL models under IID and non-IID (denoted as NIID) settings.

5.3 Results and Discussion

Auctioning Performance

To assess the auctioning performance, the following two evaluation metrics are adopted: 1) Social Welfare (SW), which is the sum of the rewards for all data consumers; and 2) revenue of the AFL ecosystem, which is formulated as the total payment during an experiment. The results are listed in Table 1. We have the following observations:

1. MARL-AFL consistently achieves the best performance in terms of SW. Compared with the best performing baseline, MARL-AFL improves the SW of the AFL ecosystem by 12.2% on average. On datasets CIFAR-10 and EMNIST-D, MARL-AFL slightly underperforms BM in terms of revenue. However, compared to SW, revenue is a less important metric as it can be significantly affected by the auctioning mechanism adopted. In the other four datasets, the proposed MARL-AFL approach outperforms all the baselines in terms of the revenue, improving the revenue by 3.9% on average over the best performing baseline.
2. Among these methods, Const and Rand perform the worst due to random bidding as well as ignoring the mutual influence among multiple data consumers. Different from Const and Rand, the other five methods take the utility of data owners being auctioned into consideration when bidding, making the bidding process more evidence-based and effective. Under the generalised second-price auction mechanism, where the bidder with the highest bid value wins the auction and pays the second highest bid value, the payments from the winning data consumer under these five methods are lower than the winners' bid prices, which are generated based on the utility of the auctioned data owner. Therefore, as shown in Table 1, the SW values achieved by these five methods are higher than their revenues.
3. Lin and Bmub outperform Const and Rand in general, with Lin achieving better performance than Bmub. This is because Bmub includes a degree of randomness in its bidding process. Compared with Lin and Bmub, BM, RLB and MARL-AFL perform significantly better. This can be attributed to the inclusion of machine learning and reinforcement learning frameworks.
4. It can be difficult to compare the performance of BM and RLB in our experiments due to the differences in their main design ideas. Specifically, BM jointly optimizes click-through rate prediction (i.e., utility estimation), cost estimation and the bidding strategy, while RLB adopts model-based reinforcement learning to model the dynamics of the bidding process. They both perform better compared to the other four baselines. However, they underperform MARL-AFL in terms of SW.
5. Both RLB and MARL-AFL are designed based on reinforcement learning. Yet, MARL-AFL outperforms RLB. This can be attributed to the design of the objective functions of these two methods, as well as the solutions to achieve these objectives. As mentioned before, RLB was originally proposed to maximize the number of clicks on ads for a given advertiser, (i.e., the utility gained by a given data consumer in our case); whereas MARL-AFL is designed to maximize the data consumers' utility, while preserving the SW of the FL ecosystem. To achieve these goals, MARL-AFL consists of a specially designed cooperation-competition modeling module for establishing the cooperation and competition among bidding agents as well as the personalized bidding lower bounds in order to mitigate potential col-

lusion (which can significantly decrease the revenue and SW of an AFL ecosystem).

FL Model Performance

Table 2 shows the test accuracy achieved by the FL models trained under different bidding approaches under both the IID and non-IID settings.

It can be observed that MARL-AFL consistently achieves the best performance. Specifically, compared to the best performing baseline, the average test accuracy achieved by MARL-AFL is 2.8% and 4.0% higher under IID and non-IID settings, respectively. Compared to Const, Rand, Bmub and Lin, data consumers under BM, RLB and MARL-AFL bid more strategically in terms of attracting high quality data owners. Different from the baselines, MARL-AFL learns bidding strategies for data consumers with the goal of coordinating their bidding behaviors towards an equilibrium that ensures fairness towards all data consumers while maximizing SW of the AFL ecosystem. In addition, it also guards against collusion among data consumers to mitigate its negative impact on the AFL ecosystem. The accuracy of the FL models produced by all approaches on the CIFAR-10 dataset is generally lower than that on other datasets. This can be attributed to the base model adopted for FL training. As mentioned in Section 5.1, the accuracy reported in Table 2 is with regard to the VGG11 network. Nevertheless, even with such a base model, the proposed MARL-AFL approach still significantly outperforms other baselines.

6 Conclusions and Future Work

This paper focuses on designing bidding strategies for data consumers in auction-based FL (AFL). We approach this problem from the perspective of MASs with a novel multi-objective cooperation-competition bid optimization formulation. The proposed MARL-AFL approach helps data consumers automatically generate bids for data owners with the aim of maximizing the accumulated profit as well as producing high-performance FL models. MARL-AFL is equipped with a temperature-based reward reassignment scheme to flexibly adapt the trade-off between cooperation and competition among data consumers. In addition, to dissuade data consumers from colluding to bid very low prices, we introduce a bar agent for each bidding agent on behalf of a data consumer to set the personalized bidding lower bound. To the best of our knowledge, MARL-AFL is the first AFL approach to support multiple data consumers bidding for data owners in a competitive AFL marketplace, while enabling an individual data owner to join multiple FL tasks simultaneously if its local computational resources allow.

In the future, we will explore the effectiveness of MARL-AFL under various budget settings. In addition, the dynamic tuning of the temperature parameter is needed.

Acknowledgments

This research/project is supported, in part, by the National Research Foundation Singapore and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-RP-2020-019); Nanyang Technological University, Nanyang Assistant Professorship (NAP); the National Key R&D Program of China (No. 2021YFF0900800); the Joint SDU-NTU Centre for AI Research (C-FAIR); and the RIE 2020 Advanced Manufacturing and Engineering (AME) Programmatic Fund (No. A20G8b0102), Singapore.

References

[Bahreini *et al.*, 2018] Tayebeh Bahreini, Hossein Badri, and Daniel Grosu. An envy-free auction mechanism for resource allocation in edge computing systems. In *SEC*, pages 313–322, 2018.

[Cai *et al.*, 2017] Han Cai, Kan Ren, Weinan Zhang, Kleanthis Malialis, Jun Wang, Yong Yu, and Defeng Guo. Real-time bidding by reinforcement learning in display advertising. In *WSDM*, pages 661–670, 2017.

[Chen *et al.*, 2020] Yiqiang Chen, Xiaodong Yang, Xin Qin, Han Yu, Piu Chan, and Zhiqi Shen. Dealing with label quality disparity in federated learning. In *Federated Learning: Privacy and Incentive*, pages 108–121. 2020.

[Clanuwat *et al.*, 2018] Tarin Clanuwat, Mikel Bober-Irizar, Asanobu Kitamoto, Alex Lamb, Kazuaki Yamamoto, and David Ha. Deep learning for classical japanese literature. *arXiv preprint*, page 1812.01718, 2018.

[Cohen *et al.*, 2017] Gregory Cohen, Saeed Afshar, Jonathan Tapson, and Andre Van Schaik. EMNIST: Extending MNIST to handwritten letters. In *IJCNN*, pages 2921–2926, 2017.

[Deng *et al.*, 2021] Yongheng Deng, Feng Lyu, Ju Ren, Yi-Chao Chen, Peng Yang, Yuezhi Zhou, and Yaoxue Zhang. Fair: Quality-aware federated learning with precise user incentive and model aggregation. In *INFOCOM*, 2021.

[Gao *et al.*, 2019] Guoju Gao, Mingjun Xiao, Jie Wu, He Huang, Shengqi Wang, and Guoliang Chen. Auction-based vm allocation for deadline-sensitive tasks in distributed edge cloud. *IEEE Transactions on Services Computing*, 2019.

[Hong *et al.*, 2020] Hsiang-Jen Hong, Wenjun Fan, C Edward Chow, Xiaobo Zhou, and Sang-Yoon Chang. Optimizing social welfare for task offloading in mobile edge computing. In *Networking*, pages 524–528, 2020.

[Jiao *et al.*, 2018] Yutao Jiao, Ping Wang, Dusit Niyato, and Zehui Xiong. Social welfare maximization auction in edge computing resource allocation for mobile blockchain. In *ICC*, pages 1–6. IEEE, 2018.

[Jiao *et al.*, 2019] Yutao Jiao, Ping Wang, Dusit Niyato, and Kongrath Suankaewmanee. Auction mechanisms in cloud/fog computing resource allocation for public blockchain networks. *IEEE Transactions on Parallel and Distributed Systems*, 30(9):1975–1989, 2019.

[Jiao *et al.*, 2020] Yutao Jiao, Ping Wang, Dusit Niyato, Bin Lin, and Dong In Kim. Toward an automated auction framework for wireless federated learning services market. *IEEE Transactions on Mobile Computing*, 20(10):3034–3048, 2020.

[Khan *et al.*, 2020] Latif U. Khan, Shashi Raj Pandey, Nguyen H. Tran, Walid Saad, Zhu Han, Minh N. H. Nguyen, and Choong Seon Hong. Federated learning for edge networks: Resource optimization and incentive mechanism. *IEEE Communications Magazine*, 58(10):88–93, 2020.

[Krishnaraj *et al.*, 2022] Nagappan Krishnaraj, Kiranmai Bellam, B Sivakumar, and Arockiam Daniel. The future of cloud computing: Blockchain-based decentralized cloud/fog solutions–challenges, opportunities, and standards. *Blockchain Security in Cloud Computing*, pages 207–226, 2022.

[Le *et al.*, 2020] Tra Huong Thi Le, Nguyen H. Tran, Yan Kyaw Tun, Zhu Han, and Choong Seon Hong. Auction based incentive design for efficient federated learning in cellular wireless networks. In *WCNC*, pages 1–6, 2020.

[Lee *et al.*, 2012] Kuang-chih Lee, Burkay Orten, Ali Dasdan, and Wentong Li. Estimating conversion rate in display advertising from past erformance data. In *KDD*, pages 768–776, 2012.

[Littman, 1994] Michael L Littman. Markov games as a framework for multi-agent reinforcement learning. In *ICML*, pages 157–163, 1994.

[Liu *et al.*, 2020] Yang Liu, Anbu Huang, Yun Luo, He Huang, Youzhi Liu, Yuanyuan Chen, Lican Feng, Tianjian Chen, Han Yu, and Qiang Yang. FedVision: An online visual object detection platform powered by federated learning. In *IAAI*, pages 13172–13179, 2020.

[Liu *et al.*, 2022] Zelei Liu, Yuanyuan Chen, Yansong Zhao, Han Yu, Yang Liu, Renyi Bao, Jinpeng Jiang, Zaiqing Nie, Qian Xu, and Qiang Yang. Contribution-aware federated learning for smart healthcare. In *IAAI*, pages 12396–12404, 2022.

[Lyu *et al.*, 2022] Lingjuan Lyu, Han Yu, Xingjun Ma, Chen Chen, Lichao Sun, Jun Zhao, Qiang Yang, and Philip S. Yu. Privacy and robustness in federated learning: Attacks and defense. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[Perlich *et al.*, 2012] Claudia Perlich, Brian Dalessandro, Rod Hook, Ori Stitelman, Troy Raeder, and Foster Provost. Bid optimizing and inventory scoring in targeted online advertising. In *KDD*, pages 804–812, 2012.

[Ren *et al.*, 2017] Kan Ren, Weinan Zhang, Ke Chang, Yifei Rong, Yong Yu, and Jun Wang. Bidding machine: Learning to bid for directly optimizing profits in display advertising. *IEEE Transactions on Knowledge and Data Engineering*, 30(4):645–659, 2017.

[Roy *et al.*, 2021] Palash Roy, Sujan Sarker, Md Abdur Razzaque, Md Mamun-or Rashid, Mohmmad Mehedi Hassan, and Giancarlo Fortino. Distributed task allocation in mobile device cloud exploiting federated learning and subjective logic. *Journal of Systems Architecture*, 113(2):doi:10.1016/j.sysarc.2020.101972, 2021.

[Shi *et al.*, 2023] Yuxin Shi, Han Yu, and Cyril Leung. Towards fairness-aware federated learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[Simonyan and Zisserman, 2015] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015.

[Tan *et al.*, 2022] Alysa Ziying Tan, Han Yu, Lizhen Cui, and Qiang Yang. Towards personalized federated learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[Thi Le *et al.*, 2021] Tra Huong Thi Le, Nguyen H. Tran, Yan Kyaw Tun, Minh N. H. Nguyen, Shashi Raj Pandey, Zhu Han, and Choong Seon Hong. An incentive mechanism for federated learning in wireless cellular networks: An auction approach. *IEEE Transactions on Wireless Communications*, 20(8):4874–4887, 2021.

[Xiao *et al.*, 2017] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. *arXiv preprint*, page 1708.07747, 2017.

[Yang *et al.*, 2019] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2):12:1–12:19, 2019.

[Yang, 2020] Shi Yang. A task offloading solution for internet of vehicles using combination auction matching model based on mobile edge computing. *IEEE Access*, 8:53261–53273, 2020.

[Ying *et al.*, 2020] Chenhao Ying, Haiming Jin, Xudong Wang, and Yuan Luo. Double insurance: Incentivized federated learning with differential privacy in mobile crowdsensing. In *SRDS*, pages 81–90, 2020.

[Yu *et al.*, 2017] Han Yu, Chunyan Miao, Yiqiang Chen, Simon Fauvel, Xiaoming Li, and Victor R. Lesser. Algorithmic management for improving collective productivity in crowdsourcing. *Scientific Reports*, 7(12541), 2017.

[Zavodovski *et al.*, 2019] Aleksandr Zavodovski, Suzan Bayhan, Nitinder Mohan, Pengyuan Zhou, Walter Wong, and Jussi Kangasharju. Decloud: Truthful decentralized double auction for edge clouds. In *ICDCS*, pages 2157–2167. IEEE, 2019.

[Zeng *et al.*, 2020] Rongfei Zeng, Shixun Zhang, Jiaqi Wang, and Xiaowen Chu. Fmore: An incentive scheme of multi-dimensional auction for federated learning in MEC. In *ICDCS*, pages 278–288, 2020.

[Zhan *et al.*, 2020] Yufeng Zhan, Peng Li, Zhihao Qu, Deze Zeng, and Song Guo. A learning-based incentive mechanism for federated learning. *IEEE Internet of Things Journal*, 7(7):6360–6368, 2020.

[Zhan *et al.*, 2021] Yufeng Zhan, Jie Zhang, Zicong Hong, Leijie Wu, Peng Li, and Song Guo. A survey of incentive mechanism design for federated learning. *IEEE Transactions on Emerging Topics in Computing*, 10(2):1035–1044, 2021.

[Zhang *et al.*, 2014] Weinan Zhang, Shuai Yuan, and Jun Wang. Optimal real-time bidding for display advertising. In *KDD*, pages 1077–1086, 2014.

[Zhang *et al.*, 2021] Jingwen Zhang, Yuezhou Wu, and Rong Pan. Incentive mechanism for horizontal federated learning based on reputation and reverse auction. In *WWW*, page 947–956, 2021.

[Zhang *et al.*, 2022a] Jingwen Zhang, Yuezhou Wu, and Rong Pan. Auction-based ex-post-payment incentive mechanism design for horizontal federated learning with reputation and contribution measurement. *arXiv preprint*, page 2201.02410, 2022.

[Zhang *et al.*, 2022b] Jingwen Zhang, Yuezhou Wu, and Rong Pan. Online auction-based incentive mechanism design for horizontal federated learning with budget constraint. *arXiv preprint*, page 2201.09047, 2022.

[Zou *et al.*, 2020] Liang Zou, Xinhui Yu, Ming Li, Meng Lei, and Han Yu. Nondestructive identification of coal and gangue via near-infrared spectroscopy based on improved broad learning. *IEEE Transactions on Instrumentation and Measurement*, 69(10):8043–8052, 2020.