Research Article

Utility Optimization of Federated Learning with Differential Privacy

Jianzhe Zhao , Keming Mao , Chenxi Huang , and Yuyang Zeng

Software College, Northeastern University, Shenyang, Liaoning Province 110169, China

Correspondence should be addressed to Jianzhe Zhao; zhaojz@swc.neu.edu.cn

Received 13 August 2021; Revised 30 August 2021; Accepted 14 September 2021; Published 8 October 2021

Academic Editor: Shi Cheng

Copyright © 2021 Jianzhe Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Secure and trusted cross-platform knowledge sharing is significant for modern intelligent data analysis. To address the trade-off problems between privacy and utility in complex federated learning, a novel differentially private federated learning framework is proposed. First, the impact of data heterogeneity of participants on global model accuracy is analyzed quantitatively based on 1-Wasserstein distance. Then, we design a multilevel and multiparticipant dynamic allocation method of privacy budget to reduce the injected noise, and the utility can be improved efficiently. Finally, they are integrated, and a novel adaptive differentially private federated learning algorithm (A-DPFL) is designed. Comprehensive experiments on redefined non-I.I.D MNIST and CIFAR-10 datasets are conducted, and the results demonstrate the superiority of model accuracy, convergence, and robustness.

1. Introduction

Nowadays, artificial intelligence faces two main challenges: data silos and privacy concerns [1, 2]. Meanwhile, with the widely used edge and Internet of things devices, the federated learning framework attracts extensive attentions, which can train the global model based on the mode of decentralization and cooperation. However, there are still severe privacy protection problems in federated learning, although models are trained by sharing updates (such as gradient information) rather than raw data [3, 4]. Recent studies show that, by analyzing the parameter differences between training and uploading, privacy can still be leaked to a certain extent, such as the weight of neural network training [5, 6]. Federated learning requires different participants to upload and aggregate parameters repeatedly to train the global model. This process leads to more privacy information disclosure. Therefore, personal privacy can be obtained through the model-inversion attack.

Differential privacy (DP) provides a strict, quantifiable, and context-independent privacy protection method for machine learning. For its information theory guarantee, DP is also widely used to enhance data privacy with its simplicity and low cost [7–10]. However, as the fundamental challenge

of privacy-preserved methods, DP mechanisms inevitably cause model performance degradation and system utility loss. Traditional differential privacy injects bounded noise into model to protect privacy, and privacy budget is a crucial factor in measuring the level of protection and the amount of noise. Existing differentially private methods allocate the same amount of privacy budget for each participant and each iteration in model updating, which leads to ubiquitous trade-off problems between privacy and utility [7, 9, 10]. This problem is mainly because federated learning is a complex training process with many participants and iterative updating, aggregation, and broadcasting. The guarantee of its privacy comes from adding a certain amount of noise at each communication. Therefore, the methods which allocate the privacy budgets equally will lead to severe utility damage.

In global model training, allocating privacy budgets for different participants and iterative process is an essential means of utility optimization under certain privacy protection levels. However, how to allocate the privacy budget is a crucial issue. At the participants' selected level, due to personalized use and collection of local data for each participant, the assumption of independent and identically distributed (I.I.D) of underlying data may not exist. Non-independent and identically distributed (non-I.I.D) makes

of data heterogeneity for federated learning. Data heterogeneity, also called statistical heterogeneity, refers to the heterogeneity of the underlying data distribution for various participants [1, 2, 9]. The existing theoretical research only concentrates on proofing the convergence and privacy protection level. The analysis for the internal relationship between the utility of federated learning and data heterogeneity is insufficient. Therefore, it is a sort of theoretical basis for privacy budget allocation that utility improvement oriented. At the iterative training process level, for each participant, with the convergence of the local model, the parameter update speed slows down, and adding the same amount of noise will have a significant impact on the convergence performance. In addition, at a certain level of global privacy, adjusting privacy budgets in different training stages will improve accuracy. In general practice, a similar strategy is applied to deep learning, which reduces the learning rate as the training goes on, rather than using a constant one. Therefore, in model convergence, adding gradually reduced noise is an effective way to improve the performance of a federated learning system.

Works on the trade-off between privacy and utility in federated model training become the research hot spot, and many state-of-the-art differentially private algorithms have been proposed [11–17]. Geyer et al. [11] proposed a DP-SGD algorithm. It analyzed the convergence performance boundary, which showed that the privacy protected level and data size affected the convergence performance. Zhang et al. [12] improved DP-SGD by tracking the privacy loss. It obtained an accurate estimation of the overall privacy loss. Kerkouche et al. [13] proposed an effective federated learning algorithm for privacy protection under a given privacy level and plenty of customers involved. Wei et al. [14] made theoretical research on differential privacy level and convergence performance. However, it does not study the allocation strategy. Authors of [15–17] proposed personalized privacy budget allocation methods to improve the accuracy of the federated learning according to the different contributions of participants to the global accuracy under a scenario of unbalanced data. Although these methods are promising, there remain several issues of the method design. First, these methods only consider the privacy budget allocation technology for different participants, but they do not consider the privacy budget cost on the process of the global model iteration. Second, inevitable existing data heterogeneity in real federated scenarios brings new challenges to the theoretical analysis. The existing research is limited to the research of technical methods, and there is a lack of method design based on theoretical research.

In order to solve problems of utility optimization in federated learning, a novel differentially private federated learning framework is proposed in this paper. The impact of data heterogeneity of participants on global model accuracy is first analyzed. Then, a multilevel and multiparticipant privacy budget dynamic allocation method is designed. Based on the proposed methods, a novel adaptive differentially private federated learning algorithm (A-DPFL) is also devised to balance privacy and utility. The main contributions of this paper are three folds:

- (1) We quantitatively analyze the impact of the participants' data heterogeneity on the global model accuracy based on 1-Wasserstein distance, and then, we propose an appropriate participant evaluation method for participants' selection in federated learning to improve the system's effectiveness.
- (2) A multilevel and multiparticipant privacy budget dynamic allocation method is designed. At the global iteration process, the noise scale adjustment method is proposed to inject less noise in pace with the convergence of the global model, and the global optimal accuracy is gained at a local optimum position. Moreover, in each training iteration, participants' active selection is achieved to obtain better utility.
- (3) A new federated learning algorithm with differential privacy is proposed. It dynamically allocates the privacy budget for multiple participants and multiple iterations in view of the data heterogeneity of participants in a federated system. It can improve system utility while keeping privacy and adaptively balance privacy and utility.

2. Related Works

2.1. Utility Optimization of Federated Learning. Many works have been done to improve the utility of global models in federated learning, such as global accuracy, convergence property, time consumption, and robustness. These studies are mainly based on the following different assumptions: number of participants, data heterogeneity, communication properties, the properties of loss functions, and the bound of gradient noise. In the device-based federated learning system, some personalized federated learning studies are derived for system heterogeneity [18–24]. They involve the optimization objectives of federated systems such as distributed mobile user interaction, communication costs in mass distribution, unbalanced data distribution, and device reliability. The key directions of this kind of research can be classified as asynchronous communication [18], scheduling policy [21], and model compression [24].

Federated learning is generally based on gradient information, such as gradient descent (GD) and stochastic gradient descent (SGD). Among them, the most used is the federated average (FedAvg) [23], which updated the average of local stochastic gradient descent. The convergence of FedAvg performs well when data follows the I.I.D hypothesis. However, research shows that even in a classical federated scenario, the underlying data does not necessarily conform to the I.I.D. It demonstrates that data heterogeneity may lead to model divergence. Some distributed gradient descent algorithms (Dist. GD) and related variants of distributed stochastic gradient descent (Dist. SGD) have been proposed, which achieve local updates similar to the federated average. For data heterogeneity, FedProx was proposed recently [24]. FedProx made a slight modification to FedAvg to ensure convergence in both theory and practice. Many advanced optimization methods are also available for federated system optimization [25–28].

The above studies achieve good results in optimizing the federated algorithm, but the algorithms only concern one or two utility indicators. Nevertheless, the method proposed in this paper boosts the possibility of improving more comprehensive utility indicators through the active selection of participants. In addition, these existing algorithms have insufficient research on privacy issues. Besides the utility loss caused by federated model training, privacy protection also brings accuracy loss and convergence performance decline additionally. At last, since the utility optimization will be different when performing privacy protection, the theoretical research on the utility optimization of privacy-preserved federated learning is still insufficient.

2.2. Privacy-Preserved Federated Learning. A series of theoretical and methodological research on privacy-preserved federated learning has been conducted, and many research results have been obtained. Privacy in federated learning can be divided into global differential privacy [14] (GDP) and local differential privacy (LDP) [7]. Global privacy requires that the model updates generated in each round be privacy-protected for all untrusted third parties except the central server. In contrast, local privacy further requires that the update be privacy-protected for the server. The privacy protected method research in federated learning continues and extends in traditional machine learning, which are mainly based on multiparty secure computing [29, 30] and differential privacy [7, 31, 32]. Among them, multiparty secure computing is a kind of the lossless method, which can maintain the original accuracy and make a strong privacy guarantee. However, the approach results significant additional communication costs. Considering the high communication cost in the federated system, differential privacy has low system overhead. Existing studies include federated learning algorithms that satisfy LDP [7], differentially private stochastic gradient descent algorithm (DP-SGD) [31], and metalearning with DP [33].

However, as the fundamental challenge of privacy-preserved methods, DP inevitably causes utility loss. Focusing on the trade-off between privacy and utility in the model training process, many state-of-the-art differentially private federated algorithms have been proposed [11–17]. There are still several issues of the method design. First, these methods only consider the privacy budget allocation for different participants, but they do not consider the privacy budget cost on the process of the global model iteration. Second, inevitable data heterogeneity in real federated scenarios brings new challenges on both the theoretical analysis and pro. The existing research is limited to technical methods, and there lacks theoretical research.

3. Preliminary

3.1. Differential Privacy. Differential privacy (DP) designs a mechanism to add noise to the target dataset so that the statistical information loss of the released dataset and the original dataset is in a small range, which ensures that the modification of an individual record in the dataset does not have a significant impact on the statistical results.

Definition 1 (differential privacy). For any dataset D and D' differing on at most one record and for any possible sanitized dataset $r \in \text{Range}(M)$, a random mechanism M satisfies ϵ -differential privacy if

$$\text{DP}(M) = \sup_{D, D', S} \log \frac{\Pr[r \in S|D]}{\Pr[r \in S|D']} \leq \epsilon, \quad (1)$$

where ϵ refers to the privacy budget that controls the privacy level of the mechanism M . The lower ϵ represents the higher privacy level.

Definition 2 (global sensitivity). For any function $Q: D \rightarrow \mathbb{R}^d$, for all D and D' differing in at most one record, the global sensitivity of Q is

$$\Delta Q = \max_{D, D'} \|Q(D) - Q(D')\|_1. \quad (2)$$

Definition 3 (mechanism). Mechanism M is associated with the global sensitivity. This measures the maximal change on the result of query Q when removing one record from the dataset D . Let D be a database, where a randomized function $M(D)$ is a (randomized) perturbation mechanism on D , if the output $r = M(D)$ follows a conditional distribution $\Pr(r \in S|D)$.

Definition 4 (Gaussian mechanism). A privacy mechanism M guarantees (ϵ, δ) -differential privacy if, for any two adjacent dataset D and D' , the following inequality holds:

$$\Pr(r \in S|D) \leq e^\epsilon \Pr(r \in S|D') + \delta. \quad (3)$$

Gaussian mechanism is used to guarantee (ϵ, δ) -differential privacy. We present the following DP mechanism by adding artificial Gaussian noise $nN(0, \sigma^2)$, where $\sigma \geq (\Delta Q/\epsilon)\sqrt{\ln(1.25/\delta)}$.

3.2. Federated Learning. The concept of federated learning was first proposed by Google [34]. It is an algorithm framework for building machine learning models, which generally includes the following features: two or more participants jointly train a global model; each participant has some local data that can be used to train the global model; during the global model training, data are retained locally by participants; during local parameter communicating, privacy protection can be used to prevent privacy leakage; the accuracy of the federated model is an optimal approximation of the accuracy of the ideal model constructed from the centralized data. There are two implicit concerns, include the utility of the global model and privacy issues in the communication process.

In application scenarios, federated learning can be divided into horizontal federated learning, vertical federated learning, and federated transfer learning. Among them, vertical and transfer federated learning devote to train different local models for aggregation. These kinds of research focus on multitask learning and metalearning, and this is different from classical federated learning in nature

[35]. In a classical federated learning scenario, samples from different participants are introduced into the same feature space shared by the dataset, called horizontal federated learning, also known as sample-based federated learning. Figure 1 illustrates a classical federated learning system, in which participants can be organizations, nodes and clients, and terminals or devices in the Internet of Things according to different application scenarios.

4. Problem Setting

In this paper, we take the horizontal federated learning system as the research object, so vertical federated learning and federated transfer learning can be regarded as exceptional cases of horizontal federated learning. Aiming at the problem of decreasing system utility caused by the privacy protection method and the new challenges in theoretical analysis and the method design caused by the data heterogeneity of federated learning, we focus on how to inject an appropriate differential privacy noise in the global model training to achieve privacy protection and optimize utility through Collaborator's privacy budget dynamic allocation. In this section, we first formalize the federated learning system (Section 4.1) and then analyze the possible attack in the federated system (Section 4.2), and finally, a global differential privacy model is designed to solve the privacy issues setting above (Section 4.3).

4.1. Formalization of Federated Learning. We consider a federated learning system, as shown in Figure 2, which trains the global model by introducing different samples from local datasets into the common feature space, namely, $X_i = X_j, Y_i = Y_j, I_i \neq I_j$, and $\forall D_i, D_j, i \neq j$. The system consists of a trusted collaborator C and N participants. D_i represents local dataset of participant $i, i \in \{1, 2, \dots, N\}$. Defining $f_i: R^d \rightarrow R$ as the loss function of participant i , federated learning can be formalized as the following optimization problem:

$$\min_{w \in R^d} f(w) := \frac{1}{N} \sum_{i=1}^N f_i(w). \quad (4)$$

In order to optimize problem 4, the federated learning process of training a global model is designed as follows:

- (1) Local model training: the participants conduct a round of parameters' training according to the local data and upload the local parameters of the current round to the collaborator
- (2) Secure aggregating: the collaborator performs secure aggregation for the current round of local parameters uploaded by the participants
- (3) Global parameter broadcasting: the collaborator broadcasts the aggregation parameters of the current round to the participants
- (4) Local model updating: the participants update the model according to the global parameters of the current round and conduct the next round of training

4.2. Threat Model. This paper analyzes and assumes the following two possible attacks based on the above-federated learning scenario. The collaborator (server) is assumed to be honest; however, external adversaries aim to obtain participants' private information, as demonstrated by model-inversion attacks. Moreover, the participants are assumed to be semihonest; adversaries can perform poison attacks on the federated system by uploading malicious parameters to interfere with the aggregation of the global model.

4.3. Global Differential Privacy Model. According to the possible threat model and solving the optimization problem formalized mentioned above, we design a differentially private federated learning framework that satisfies global differential privacy for the federated system and optimizes the utility by dynamic allocating differential privacy budgets, as shown in Figure 2. By injecting Gaussian noise to local parameters, the designed federated system achieves (ϵ, δ) -differential privacy. Furthermore, the honest collaborator conducts the privacy budgets dynamic allocation according to the utility performance of participants (data heterogeneity) and iteration process. Meanwhile, if the participant's evaluation is below a certain threshold, it signifies the participant may be an adversary, and the collaborator refuses to update its parameters against poison attacks.

5. Utility Optimization of DPFL Framework

5.1. Framework Outline. In the paper, a novel differentially private federated learning framework is proposed to provide global differential privacy and optimize the system utility for the federated system. Including this framework, the participants' evaluation method is proposed based on analyzing the impact of the data heterogeneity of participants on the global model accuracy, a multilevel and multiparticipant privacy budget dynamic allocation method is designed, and a novel adaptive differentially private federated learning algorithm (A-DPFL) is also devised to balance the privacy and utility. The research framework proposed in this paper mainly proposes an appropriate solution to the problem of equal allocation of privacy budget by existing methods and designs a privacy allocation scheme for multiple participants and multiple iterations in complex federal systems. It mainly includes participant selection in a federated system with non-I.I.D and dynamic adjustment of privacy parameters in global iteration. The following is an explanation of the scheme design at each level.

Due to the general data heterogeneity in the complex federated system and the significant impact of data heterogeneity on the accuracy and convergence performance of the global model [18, 22, 24], the participants make a significant difference in the contributions of the global training model. Therefore, participants' active selection is an effective means to improve the utility. Allocating privacy budgets for the selected participants will improve the utility under the premise of a privacy guarantee. Our method quantitatively analyzes the

FIGURE 1: Framework of classical federated learning.

FIGURE 2: Formalization of utility optimization federated learning framework.

impact of the participant on the global model accuracy based on 1-Wasserstein. It proposes an appropriate participant evaluation method used as the basis for participants selection to improve the system's utility. We design a general approach to participants' selection. We take improving global accuracy as an example to expand the description in this paper. In fact, we can also choose participants with better convergence performance or other utility indicators to train the global model.

At present, federated learning with differential privacy is realized by adding quantitative random noise to the parameters in each iteration to protect privacy. However, as the

model converges, the parameter update speed slows down, and adding the same amount of noise will significantly impact the final convergence of the model. Moreover, one main concern of model training is the accuracy achieved under a certain private level. Maintaining certain global private levels and reducing the noise injected as the iteration process will achieve higher accuracy. The practice of general deep learning applies a similar strategy to achieve higher accuracy, which reduces the learning rate as the training progresses instead of using a constant learning rate in all periods [36, 37]. Therefore, to approach the local optimum better and achieve higher model accuracy, gradually adding

reduced noise during the model convergence is an effective means to improve the utility of the federated learning system. We design a privacy budget dynamic allocation method that adjusts the noise scale dynamically as the model converges so that less noise is injected to achieve better model performance under a certain level of privacy.

In the following section, we conduct a theoretical analysis of participants' statistical heterogeneity in the federated system (Section 5.2) and then design a multilevel and multiparticipant privacy budget dynamic allocation method (Section 5.3) to allocate privacy budget adaptively in the federated learning training process. Finally, we propose an adaptive differentially private federated learning algorithm (Section 5.4) to optimize the utility of federated learning. The designed framework achieves a good balance between privacy and utility for federated learning by these necessary means of the framework.

5.2. Data Heterogeneity Analysis of Participants. To theoretically analyze the participants' data heterogeneity in federated learning, we firstly declare the federated system's data heterogeneity, then quantitatively analyze the effect of participant data heterogeneity on the global accuracy, and finally propose the participants' evaluation method.

5.2.1. Data Heterogeneity Declaration. Many factors cause non-I.I.D; among them, most extensive research focuses on label distribution skew [38, 39]. Since participants are associated with specific geographic areas, the label distribution of local data is quite different from each other. For example, kangaroos are only found in Australia or zoos, and each specific face only appears in a few places in the world. Therefore, we declare data heterogeneity in federated learning as the labels skewed distribution of different participants i and j as

$$p_i(y) \neq p_j(y) \neq p(y). \quad (5)$$

If formalizing the federated learning as a supervised learning, we have the expected loss under the local data distribution of the participant i :

$$f_i(w) := E_{(x,y) \sim p_i} [l_i(w; x; y)], \quad (6)$$

where the loss function $l_i(w; x; y)$ is used to measure the error of local model parameters w judging the true label $y \in Y_i$ based on the given input $x \in X_i$. p_i is the probability distribution of the local sample space $X_i \times Y_i$, and data heterogeneity is manifested as a skewed distribution of $p_i(x, y)$. More intuitively, assuming that the value of the classification problem is discrete, the participant's expected loss can be formalized as

$$f_i(w) = \sum_{(x,y) \in D_i} p_i(x, y) l_i(w; x; y). \quad (7)$$

5.2.2. Quantitative Analysis between Data Heterogeneity and Global Accuracy. According to the setting of the research

scope, the participants of the federated system in this paper have the same loss function, which is recorded as

$$l(z; w), z := (x; y) \in Z := X \times Y. \quad (8)$$

Then, the loss function of the participant can be recorded as

$$f_i(w) := E_{(z \sim p_i)} [l(z; w)]. \quad (9)$$

According to the formalized definition of the research problem, that is, equation (7), it can be concluded that the global accuracy is closely related to the similarity between the local distribution and the global distribution. 1-Wasserstein distance is one of the effective methods to measure the similarity of two probability distributions. When $d(z_1, z_2)$ is distance function, $Q(q_1, q_2)$ is the set of measures of the marginal distributions q_1 and q_2 in the space $Z \times Z$. The 1-Wasserstein distance of the two probability distributions q_1 and q_2 in the matrix space Z can be defined in

$$W(q_1, q_2) := \inf_{q \in Q(q_1, q_2)} \int_{Z \times Z} d(z_1, z_2) dq(z_1, z_2). \quad (10)$$

Therefore, assuming that, for any w , $\nabla_w l(z; w)$ has β -Lipschitz continuity and ρ -Lipschitz smoothness, we quantitatively analyze and prove the influence of the 1-Wasserstein distance between the local data distribution and the global data distribution on the global accuracy, which can be summarized and simply proofed as

$$\|\nabla_w l(z_1; w) - \nabla_w l(z_2; w)\| \leq L_Z(z_1, z_2). \quad (11)$$

5.2.3. Participants' Evaluation Method. Based on the above theoretical research on the impact of participants' local distribution on global accuracy, we propose the following methods for evaluating participant data heterogeneity. The participant evaluation method based on 1-Wasserstein distance is evaluating the data heterogeneity of participants by solving equation (12). The value γ_1 is the average degree of data heterogeneity for all participants. Then, comparing the values of $W(p_i, p)$, γ_1 is defined as evaluation score SC_i :

$$\gamma_1^2 = L_Z^2 \frac{1}{N} \sum_{i=1}^N W(p_i, p)^2. \quad (12)$$

5.3. Dynamic Allocation Method of Privacy Budgets. Since the complex federated learning differs from the previous machine learning, we design a multilevel and multiparticipant privacy budget dynamic allocation method to improve the utility. It reduces noise intake from two levels and improves the utility of the federated system. First, in one round of the iterative process, according to the heterogeneity evaluation of the participants, that is, the degree of contribution to the global model, it allocates privacy budget for selected participants. Second, according to the degree of convergence in the global iteration process, it allocates the privacy budget for global iteration by dynamically adjusting the noise scale.

5.3.1. Participants Active Selection. In practice, it is unrealistic for global participants in the federated learning system to participate in each iteration. In theory, at the expense of system time utility and robustness, increasing the number of participants can achieve better global accuracy. However, in one round of iteration, the introduction of differential privacy protection operation makes greater amount of noise with more participants, resulting in a decrease in global accuracy. Therefore, the number of participants K of each iteration has a significant impact on the utility of the federated system. Meanwhile, according to the data heterogeneity of participants, selecting different participants to participate in each round of iteration will have a different impact on the system's utility. Based on the above analysis, we design a participants' active selection method that actively allocates the privacy budget to the selected participants to reduce global noise intake and obtain better utility.

The specific method of allocating privacy budgets for K participants in one round of iteration is based on the participant's probability of being selected. The probability of participant i being selected is denoted as p_i^t , and p_i^t is calculated according to evaluation result SC_i . The probability of global participants satisfies

$$\sum_{i=1}^N p_i^t = 1. \quad (13)$$

According to the ranking of the probability of being selected, the decision of the K participants in round t is made, see equation (14). Then, the participants are allocated privacy budgets for this round to achieve privacy protection:

$$S(t) \triangleq \{S(t) : |S(t)| = K\}. \quad (14)$$

5.3.2. Noise Scale Dynamic Adjustment. Theoretically, when all global participants N participate in training, the convergence performance of the federated system is related to the number of iterations T and the noise level. As T increases, the convergence performance improves, but the accumulated noise level also increases, which leads the convergence performance decrease. Therefore, this paper designs a privacy budget allocation method, i.e., noise scale dynamic adjustment method; under a certain global privacy level, adjusting the noise scale dynamically as each round of convergence changing and reducing the noise scale dynamically by verifying the accuracy of the current model. We can use the calculating average of the local accuracy as the verification accuracy. Another way is to use a public dataset accessible to the server, monitor the accuracy of the model, and reduce the noise scale when the accuracy stops increasing. Every time the increase in verification accuracy is lower than the threshold, the noise scale is reduced by k times, until the total privacy budget is exhausted. Here, k is the noise scale adjusting parameter. The iterative round of performing the inspection mentioned above operations is called dynamic adjustment round. When the accuracy improvement $S_t - S_{t-1}$ between two dynamic adjustment rounds is less than the threshold α , the server will reduce the

noise parameter to $\sigma'_t = k\sigma_e$ ($k \in (0, 1)$). If the accuracy increase is less than the threshold α , the noise scale remains unchanged. Subsequently, the updated noise scale will be used in the subsequent training until the next dynamic adjustment round $t + 1$ is reached. The noise scale σ'_t update equation is as follows:

$$\sigma'_t = \begin{cases} k\sigma_t, & S_t - S_{t-1} \leq \alpha, \\ \sigma_t, & \end{cases} \quad (15)$$

where $k \in (0, 1)$ and $S_0 = 0$.

5.4. Adaptive Differentially Private Federated Learning. In this section, an adaptive differentially private federated learning algorithm (A-DPFL) is proposed combining the above theories and methods. By injecting Gaussian noise to the local weights to provide global differential privacy, it aggregates the updating private local weights for federated global model training. The multilevel and multi-participant privacy budget dynamic allocation method is performed in the traditional federated learning process to reduce the noise injected from the perspectives of participants selecting and noise scale adjusting, thereby improving the utility. Algorithm 1 presents the detail of A-DPFL.

The main steps are described as follows. Comparing with the traditional process of DPFL, after initialization, a participant selection step is added, and after several rounds of aggregation, noise scale adjustment step is carried out.

- (1) Initialization: the collaborator initializes global model parameters such as learning rate, batch size, and noise parameter.
- (2) Participant selection: the collaborator evaluates the available participants and, according to the optimal indicator, decides who is selected to join in the current training round of federated learning.
- (3) Parameter broadcast: the collaborator broadcasts the global model and the noise parameters of the current round to the selected participants. The participants synchronously update the local model as the global one.
- (4) Local model update: the participant privately and locally performs a round of model training by adding random Gaussian noise to the local weights based on the local data.
- (5) Local models upload and aggregation: the selected participants upload the private local model to the collaborator. After that, the collaborator aggregates the received weights as a new global model.
- (6) Noise scale adjustment: the collaborator obtains verification accuracy to verify the global model convergence situation and decides whether adjusting the noise parameter or not in the current predetermined dynamic adjustment round. If so, an adjusted noise parameter will be broadcast in the next training round.

Input: Dataset $D = \{D_1, \dots, D_n\}$, initial $\sigma_0 w^{(0)}$, maxRound, $t = 0$, number of client chosen K , $p_threshold$, learning rate η , batch size B , clipping threshold C , σ_t , max_local_round, noise scale adjusting $k \in (0, 1)$, adjusting threshold α , $S_0 = 0$

Output: Global model weight $w^{(t)}$

```

(1)  $D' \leftarrow D$ 
(2)  $w^{(t)} = w^{(0)}$ 
(3)  $\sigma_t = \sigma_0$ 
(4) while  $t < \text{max Round}$  do
(5)    $D' = \text{Client choose}(D', K) // 2$ 
(6)   Broadcast( $w^{(t)}$ )
(7)   for  $i \in D', i = 1, \dots, n$  do
(8)      $\tilde{w}_i^{(t)} = \text{Client\_update}(D_i, \sigma_t) // \text{Algorithm 3}$ 
(9)     Upload( $\tilde{w}_i^{(t)}$ )
(10)  end
(11)   $w^{(t)} = \text{Aggregate}(w_i^{(t)}) // i = 1, \dots, n$ 
(12)  if round = dynamic_adjust_round then
(13)     $\sigma_t = \text{Dynamic\_adjust}(D', \sigma_t) // \text{Algorithm 4}$ 
(14)  end
(15)   $t = t + 1$ 
(16) end
(17) return  $w^{(t)}$ 

```

ALGORITHM 1: A-DPFL.

Input: D , number of participant selected K , $p_threshold$

Output: Clients chosen D'

```

(1) for client in client list do
(2)   score = evaluate client( $D$ )
(3)   score list.add(score)
(4) end
(5)  $p\_list = \text{calculate\_p}(\text{score list})$ 
(6) sort( $p\_list$ )
(7) calculate  $p\_threshold(K)$ 
(8) for  $p$  in  $p\_list$  do
(9)   if  $p \geq p\_threshold$  then
(10)     $D'$  add(client)
(11)  end
(12) end
(13) return  $D'$ 

```

ALGORITHM 2: Client choose.

Input: $D_i (i = 1, \dots, n)$, learning rate η , batch size B , clipping threshold C , σ_t , max_local_round

Output: Private local weight $\tilde{w}_i^{(t)}$

```

(1) Local round  $j = 0$ 
(2) while  $j \leq \text{max\_local\_round}$  do
(3)   Forward pass( $B_j$ )
(4)    $g_t(B_i) \leftarrow \nabla_{w_i^{(t)}} L(w_i^{(t)}, B_i) // \text{Compute gradient}$ 
(5)    $\tilde{g}_t \leftarrow (1/B) (\sum_i g_t(B_i) + N(0, \sigma_t^2 C^2)) // \text{Add noise}$ 
(6)    $\hat{g}_t(B_i) \leftarrow \tilde{g}_t / \max(1, (\|\tilde{g}_t\|_2 / C)) // \text{Clip gradient}$ 
(7)    $\tilde{w}_i^{(t)} \leftarrow w_i^{(t)} - \eta \hat{g}_t // \text{Apply gradient}$ 
(8)  $j++$ 
(9) end
(10) return  $\tilde{w}_i^{(t)}$ 

```

ALGORITHM 3: Client update.

Input: $D, w^{(t)}, \sigma_t$, noise adjusting scale $k \in (0, 1)$,
adjusting threshold $\alpha, S_0 = 0$
Output: Adjusted noise parameters σ_t

- (1) $S_{t-1} = S_0$
- (2) $S_t = \text{Validate}(D, w^{(t)})$
- (3) **if** $S_t - S_{t-1} \leq \alpha$ **then**
- (4) $\sigma_t = k\sigma_t$
- (5) **else**
- (6) $\sigma_t = \sigma_t$
- (7) **end**
- (8) $S_t = S_0$
- (9) **return** σ_t

ALGORITHM 4: Dynamic adjust.

Algorithm 1 includes four significant steps: participant selection, local model update, local models' upload and aggregation, and noise scale adjustment. The following is a detailed description of these steps.

5.4.1. Participant Selection. Algorithm 2 presents the participant active selection process.

In the participant selection step, the participant evaluation method described in Section 5.2.3 is utilized to determine the contribution to the federated model training. Equation (12) shows the detail of evaluation. We calculate the probability p_i^t of participant i being selected according to its evaluation result SC_i and sort them and select the participants whose probability of being selected is higher than a certain probability threshold and add them to current round of iteration.

5.4.2. Local Model Update. In the participant local model update step, each participant selected in the iteration will train a local model for a specified number of times to obtain the updated local model. In A-DPFL, DP-SGD is utilized, and Gaussian noise is added to grantee privacy. Algorithm 3 presents the process in detail.

According to DP-SGD, it adjusts the model weights w to minimize the error function $\mathcal{L}w$ of the model. We also adopt the idea of the minibatch gradient descent method and randomly extract a batch of data with the number of B from all the local data of the participants for model training. The gradient firstly is calculated based on the training local data x_i with its label i :

$$g_t(x_i) \leftarrow \nabla_{w_i^{(t)}} L(w_i^{(t)}, x_i), \quad (16)$$

where $i \in B$ and $g_t(x_i)$ represents the gradient of the loss function calculated by training data x_i to the weight w in the t th iteration.

Gaussian noise is added to the gradient calculated by the data in the current batch, and then, the average gradient of the current batch is calculated:

$$\tilde{g}_t \leftarrow \frac{1}{B} \left(\sum_i g_t(x_i) + N(0, \sigma_t^2 C^2) \right), \quad (17)$$

where $\mathcal{N}(0, \sigma_t^2 C^2)$ represents Gaussian noise with mean 0 and variance $\sigma_t^2 C^2$. σ is the key parameter to control the noise scale. The greater the σ , the greater the variance of the normal distribution and the greater the amount of Gaussian noise.

Following adding noise, the gradient is clipped, as shown in equation (18). Gradient clipping ensures that the second norm of the gradient does not exceed the clipping threshold C . If $\|\tilde{g}_t\|_2 \leq C$, the value of $\|\tilde{g}_t\|_2$ will not change. If $\|\tilde{g}_t\|_2 > C$, the value of $\|\tilde{g}_t\|_2$ will be reduced to C . Gradient clipping is a common step in deep learning, which can effectively prevent the phenomenon of gradient explosion that results in poor model convergence performance:

$$\hat{g}_t(x_i) \leftarrow \frac{\tilde{g}_t}{\max(1, (\|\tilde{g}_t\|_2/C))}. \quad (18)$$

Finally, we have the processed gradient $\hat{g}_t(x_i)$ and updated the model according to the learning rate η to obtain the local parameters after training:

$$\tilde{w}_i^{(t)} \leftarrow w_i^{(t)} - \eta \hat{g}_t. \quad (19)$$

5.4.3. Local Models' Upload and Aggregation. The collaborator receives the local weight $\tilde{w}_i^{(t)}$ uploaded by each participant and aggregates to update the global model as equation (20). In the step, FedAvg is considered to aggregate local weights as a new global model:

$$w^{(t)} \leftarrow \sum_{i=1}^N \tilde{w}_i^{(t)}. \quad (20)$$

5.4.4. Noise Scale Adjustment. In this step, the noise scale dynamic adjustment method presented in Section 5.3.2 is utilized to dynamically adjust the noise scale σ . In the dynamic adjustment round, the collaborator obtains verification accuracy and compares with the verification accuracy

of the last dynamic adjustment round. If the difference is less than the set threshold α , it can be considered that the current global model is gradually converging. Therefore, to achieve better model accuracy, it is necessary to reduce the allocation of privacy budget by decreasing noise parameter σk times. If the difference is greater than the threshold, σ will remain unchanged and continue training with current noise scale. The algorithm description is shown in Algorithm 4.

6. Evaluation

In this section, we evaluate the performance of our method and against data heterogeneity and different parameters. We first compare convergence properties and the accuracy properties between our method and classical method and then study the impact of data heterogeneity and the trade-off between accuracy and privacy.

6.1. Experimental Setting. This paper used TensorFlow-Federated, a federated learning library in TensorFlow. We develop the algorithm secondarily based on Python and built a real federated learning environment by deploying the algorithm to multiple edge devices. In the experimental process, two classical datasets commonly used in deep learning, the MNIST dataset [40] and CIFAR-10 dataset [41], were chosen for the experiments:

- (i) MNIST: the dataset contains 70,000 gray-scale images of handwritten digits, divided into 60,000 training images and 10,000 test images. The size of each image is $28 * 28$, and the labels are numbers between 0 and 9.
- (ii) CIFAR-10: this dataset contains ten categories of RGB color images, divided into 50,000 training images and 10,000 test images. The size of each image is $32 * 32$. The images have the following ten categories: airplane, car, bird, cat, deer, dog, frog, horse, boat, and truck.

To simulate the characteristics of participants' data heterogeneity in the real environment, we performed the non-I.I.D setting common in the bottom data of federated learning, i.e., the label skewed distribution setting. The skewness of data labels refers to the inhomogeneous degree in the division of labels among different participants. In the I.I.D setting, the data distribution of different labels is uniformly distributed, while in the non-I.I.D setting of label skewness, some labels are present only in some participants. Because partitions are tied to particular geo-regions, the distribution of labels varies across partitions. For example, a 20% skewness means that 20% of the participants' data are divided by labels and 80% of the data are randomly selected with different labels. In contrast, a 100% skewness means that all the participants' data are of the same label. We first preprocessed all the training data that were read and assigned data of different label skewness to all participants during the experiments. Taking the MNIST dataset as an example, the data distribution of the participants is shown in Figure 3. Participant 1 has 80% label skewness and 80% of

FIGURE 3: Label distribution skew of participants.

the training data labels are 1; participant 2 has 50% label skewness and half of the training data labels are 4; participant N has 0 label skewness and the data labels it owns are more uniformly distributed.

Due to convolutional neural networks' excellent feature extraction ability, convolutional neural networks were used as the structure of the neural network of the participant's local model in the experiments. In the experiments for both datasets, the convolutional neural network with two convolutional layers was used, and its specific structure was set up as follows: a convolutional layer with the convolutional kernel of $5 * 5$, the output channel number of 32, and a pooling layer of $2 * 2$ and next was a fully connected layer with 1024 nodes, ReLU as the activation function, and an output layer with 10 nodes.

Some of the parameters used in the experiments are shown in Table 1. The CIFAR-10 dataset is an RGB three-channel color image with a larger total data volume compared to MNIST. Due to hardware limitations, the number of participants was set to a smaller number of 100 in the experiments on the CIFAR-10 dataset. In training on the CIFAR-10 dataset, there were more model parameters. In order to make the gradient after gradient cropping keeps the same direction as the original gradient as much as possible, the gradient cropping threshold needed to be set to a larger value. In this paper, the gradient cropping threshold was set to 7.0 in the experiments on the CIFAR-10 dataset.

6.2. Utility evaluation. In order to verify the global model accuracy and convergence performance of the algorithm designed in this paper, the comparison experiments were conducted to compare the designed algorithm (A-DPFL) with the classical differentially private federated learning optimization algorithm (CL-FL) [11], and the differentially private federated learning algorithm without optimization (DPFL). Also, a nonprivate federated learning algorithm was included to observe the effect of the differential privacy operation on the accuracy. The comprehensive experiments were conducted on the algorithm using MNIST and CIFAR-10 datasets to verify the effectiveness of the designed algorithm on different image classification datasets. Figures 4(a) and 4(b) present the variation of global model accuracy with increasing iteration rounds on the MNIST dataset and CIFAR-10.

TABLE 1: Parameters setting on different datasets.

Parameters	MNIST	CIFAR-10
σ	1.0	1.0
k	0.7	0.7
Batch size	20	20
Clipping threshold	4.0	7.0
Learning rate	0.05	0.05
Number of participants	1000	100
Training round	50	50

FIGURE 4: (a) MNIST. (b) CIFAR-10. Accuracy of different methods on different datasets.

The experimental results on the two datasets show that the nonprivate federated learning algorithm has the highest global accuracy on different datasets, and the global accuracy increases more steadily during the convergence process. It indicates that the privacy-protected operation will have some impact on the federated system performance. The global model accuracies of the remaining private algorithms are lower than that of the nonprivate algorithms, and the global model accuracy fluctuates to some extent during the convergence process because of adding random noise. Federated learning for privacy protection optimization algorithms, such as the classical CL-FL algorithm, will improve the system utility to some extent due to the personalized allocation of the privacy budget considered. In contrast, the traditional DPFL algorithm differs significantly from the optimization algorithm in terms of global accuracy and convergence performance. However, the convergence performance (convergence speed) and the global model accuracy of our method (A-DPFL) outperform CL-FL on

both datasets due to the dynamic allocation method of privacy budget designed. It autonomously selects participants with higher model accuracy contribution and reduces the noise intake as training proceeds, thus enhancing the utility of federated learning from two levels.

6.3. Impact of Data Heterogeneity. When performing federated learning, it is often desirable that each participant has I.I.D local data so that each participant's local model can be adequately trained to obtain better global model accuracy after aggregation. However, there is usually non-I.I.D of local data due to the differences in device usage. In this set of experiments, we simulated the case of label skewed distribution of the participants' data and observed the difference of the global accuracy under different data heterogeneity settings. Table 2 shows the comparison of the global accuracy of A-DPFL on two datasets.

As shown explicitly, the global model accuracy still retains high usability in data label skewed distribution, although there is some degradation compared to the ideal case.

TABLE 2: Global accuracy of A-DPFL under different label skew distributions.

Dataset	Accuracy on non-I.I.D	Accuracy on I.I.D
MNIST	0.942	0.989
CIFAR-10	0.641	0.704

FIGURE 5: Noise scale σ varies with the training rounds.

The experiment illustrates that data heterogeneity impacts the training accuracy and verifies that our method has good robustness in a data heterogeneity practical scenario.

6.4. Trade-Off between Accuracy and Privacy. The necessary means of our method that significantly impacts the global model utility enhancement and privacy utility balance is the dynamic adjustment of the noise scale. In this section, we design experiments to show the changes in the accuracy enhancement with the convergence process and the sensitivity of the global accuracy to the noise to verify the effectiveness of our method.

6.4.1. Noise Scale Dynamic Adjustment Method Evaluation. Firstly, to verify the noise scale dynamic adjustment method proposed in this paper, the change of noise scale σ during the training process of the designed algorithm using the MNIST dataset is observed in this paper. The trend of noise scale σ with the increase of training rounds is shown in Figure 5.

It can be observed that the time to keep the noise scale σ constant in each round decreases as the training proceeds. In the beginning, σ is kept at 1.0 for 20 training rounds; then, it is kept at 0.7 for 4 rounds, 0.49 for 4 rounds, and 0.34 for 7 rounds, and the following holding time is gradually shortened. The reason for this phenomenon is that, as the global model converges, the speed of accuracy rise of the global model on the validation dataset gradually decreases, and the amount of validation accuracy rise does not exceed the given threshold α for more time, resulting in a frequent adjustment of the noise scale σ . At the same time, as the noise scale σ decreases faster, the noise intake decreases during training.

FIGURE 6: Global accuracy varies with k .

The experiment indicates that the global model can obtain better global model accuracy in the process of convergence to the local optimum, proving the effectiveness of the method designed in this paper.

6.4.2. Impact of Noise Scale Adjusting Parameter. In our method, the noise scale adjusting parameter k controls the noise variation and effects on global accuracy. In noise scale dynamic adjustment method design, k determines the speed of noise size reduction with training rounds, so k weighs the level of privacy protection against the global model accuracy. In the last experiments, we measured the accuracy of the

global model corresponding to k , keeping other parameters' constant. This paper conducted experiments with k ranging from 0.3 to 0.9 and observed the trend of global model accuracy with k on the MNIST dataset. The specific results are shown in Figure 6.

As we can observe from the figure, the global model accuracy decreases as k increases. An increase in k represents a slow reduction in the noise scale, which means that more noise is added during the training process, causing a decrease in the global model accuracy. However, the decrease of k will make the noise decrease too fast and reduce the privacy level. Through the experiments, we can find that the global model accuracy decreases slowly when the value of k is in the range of 0.3–0.7, but the decreasing trend speeds up significantly when it is in the range of 0.8–0.9. According to the above analysis, $k = 0.7$ is a more suitable value to balance privacy protection and global model accuracy. In other experiments of this paper, the value of $k = 0.7$ is also adopted.

7. Conclusion and Future Works

This study makes research on utility optimization of federated learning with differential privacy. A novel framework is designed for solving the inherent problem of balancing between privacy and utility in privacy-preserved machine learning. In order to deal with utility damage through equal allocation to each participant and iteration in federated learning, a multilevel and multiparticipant privacy budgets' dynamic allocation method based on theoretical analysis of data heterogeneity is proposed. Implementing the framework of this study, a novel adaptive differentially private federated learning algorithm (A-DPFL) is also devised, and comprehensive experiments are conducted to demonstrate the improvement of the global accuracy and convergence speed and the robustness in non-I.I.D scenario on the MNIST and CIFAR-10 dataset.

Our future works will focus on these aspects. (1) More comparison algorithms should be added to verify the method's effectiveness. (2) In the complex federated learning, due to devices' performance deference, system heterogeneity should be considered to improve the system robustness.

Data Availability

The mnist and cifar-10 datasets used to support the findings of this study are available at tensorflow/g3doc/tutorials/mnist/ and <http://www.cs.toronto.edu/~kriz/cifar-10-python.tar.gz>.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No. 62102074, the Natural Science Foundation of Liaoning Province of China under

Grant No. 2020-MS-091, and the Fundamental Research Funds for the Central Universities of China under Grant No. N2017015.

References

- [1] Y. Cheng, Y. Liu, T. Chen, and Q. Yang, "Federated learning for privacy-preserving AI," *Communications of the ACM*, vol. 63, no. 12, pp. 33–36, 2020.
- [2] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 12–19, 2019.
- [3] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," *Future Generation Computer Systems*, vol. 115, pp. 619–640, 2021.
- [4] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.
- [5] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proceedings of the Conference on Computer and Communications Security*, pp. 1322–1333, Denver, CO, USA, 2015.
- [6] L. Melis, C. Song, and E. De Cristofaro, "Exploiting unintended feature leakage in collaborative learning," in *Proceedings of the IEEE Symposium on Security & Privacy*, pp. 691–706, San Francisco, CA, USA, 2019.
- [7] N. Wang, "Collecting and analyzing multidimensional data with local differential privacy," in *Proceedings of the IEEE 35th International Conference on Data Engineering (ICDE)*, pp. 638–649, Macao, China, 2019.
- [8] Q. Yang, *Federated Learning*, Morgan & Claypool, San Rafael, CA, USA, 2019.
- [9] H. B. McMahan, D. Ramage, K. Talwar, and L. Zhang, "Learning differentially private recurrent language models," in *Proceedings of the International Conference on Learning Representations*, Vancouver, Canada, 2018.
- [10] O. Thakkar, G. Andrew, and H. B. McMahan, "Differentially private learning with adaptive clipping," 2019, <http://arxiv.org/abs/1905.03871>.
- [11] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: a client level perspective," 2017, <http://arxiv.org/abs/1712.07557>.
- [12] J. Zhang, Y. Zhao, J. Wang, and B. Chen, "FedMEC: improving efficiency of differentially private federated learning via mobile edge computing," *Mobile Networks and Applications*, vol. 25, no. 6, pp. 2421–2433, 2020.
- [13] R. Kerkouche, G. Ács, C. Castelluccia, and P. Genevès, "Constrained differentially private federated learning for low-bandwidth devices," 2021, <http://arxiv.org/abs/2103.00342>.
- [14] K. Wei, J. Li, M. Ding et al., "Federated learning with differential privacy: algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020.
- [15] R. Hu, Y. Guo, H. Li, Q. Pei, and Y. Gong, "Personalized federated learning with differential privacy," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9530–9539, 2020.
- [16] X. Liu, H. Li, G. Xu, R. Lu, and M. He, "Adaptive privacy-preserving federated learning," *Peer-to-Peer Networking and Applications*, vol. 13, no. 6, pp. 2356–2366, 2020.

- [17] M. Hao, H. Li, X. Luo, G. Xu, H. Yang, and S. Liu, "Efficient and privacy-enhanced federated learning for industrial artificial intelligence," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6532–6542, 2020.
- [18] F. Sattler, S. Wiedemann, K. Mller, and W. Samek, "Robust and communication-efficient federated learning from non-I.I.D. data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, pp. 1–14, 2019.
- [19] S. Wang, T. Tuor, T. Salonidis et al., "Adaptive federated learning in resource constrained edge computing systems," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1205–1221, 2019.
- [20] H. Yu, R. Jin, and S. Yang, "On the linear speedup analysis of communication efficient momentum SGD for distributed non-convex optimization," in *Proceedings of the Machine Learning Research*, pp. 7184–7193, Long Beach, CA, USA, 2019.
- [21] C. Xie and S. Koyejo, "Asynchronous federated optimization," 2019, <http://arxiv.org/abs/1903.03934>.
- [22] X. Li, K. Huang, and W. Yang, "On the convergence of FedAvg on Non-iid data," in *Proceedings of the International Conference on Learning Representations*, Addis Ababa, Ethiopia, 2020.
- [23] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the Artificial Intelligence and Statistics*, Lauderdale, FL, USA, 2017.
- [24] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proceedings of the ICML 2019 AMTL Workshop*, pp. 1–28, Long Beach, CA, USA, 2019.
- [25] L. Ma, S. Cheng, and Y. Shi, "Enhancing learning efficiency of brain storm optimization via orthogonal learning design," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 99, pp. 1–20, 2020.
- [26] L. Ma, M. Huang, S. Yang, R. Wang, and X. Wang, "An adaptive localized decision variable analysis approach to large-scale multi-objective and many-objective optimization," *IEEE Transactions on Cybernetics*, vol. 99, pp. 1–13, 2021.
- [27] L. Ma, Y. Guo, M. Huang, S. Yang, X. Wang, and H. Zhang, "Learning to optimize: reference vector reinforcement learning adaption to constrained many-objective optimization of industrial copper burdening system," *IEEE Transactions on Cybernetics*, vol. 1, pp. 1–14, 2021.
- [28] L. Ma, K. Hu, Y. Zhu, and H. Chen, "Cooperative artificial bee colony algorithm for multi-objective RFID network planning," *Journal of Network and Computer Applications*, vol. 42, pp. 143–162, 2014.
- [29] K. Bonawitz, V. Ivanov, B. Kreuter et al., "Practical secure aggregation for privacy-preserving machine learning," in *Proceedings of the Conference on Computer and Communications Security*, Dallas, TX, USA, 2017.
- [30] B. Ghazi, R. Pagh, and A. Velingker, "Scalable and differentially private distributed aggregation in the shuffled model," 2019, <http://arxiv.org/abs/1906.08320>.
- [31] N. Wu, F. Farokhi, D. Smith, and M. Ali Kaafar, "The value of collaboration in convex machine learning with differential privacy," in *Proceedings of the IEEE Symposium on Security and Privacy*, pp. 304–317, San Francisco, CA, USA, 2020.
- [32] J. Li, M. Khodak, S. Caldas, and A. Talwalkar, "Differentially private meta-learning," in *Proceedings of the International Conference on Learning Representations*, Addis Ababa, Ethiopia, 2020.
- [33] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2013.
- [34] H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pp. 1273–1282, Lauderdale, FL, USA, 2017.
- [35] Y. Liu, Y. Kang, C. Xing, T. Chen, and Q. Yang, "A secure federated transfer learning framework," *IEEE Intelligent Systems*, vol. 35, no. 4, pp. 70–82, 2020.
- [36] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 1800, Honolulu, HI, USA, 2017.
- [37] L. Yu, L. Liu, C. Pu, M. E. Gursoy, and S. Truex, "Differentially private model publishing for deep learning," in *Proceedings of the IEEE Symposium on Security and Privacy*, pp. 332–349, San Francisco, CA, USA, 2019.
- [38] H. Wang, Z. Kaplan, D. Niu, and B. Li, "Optimizing federated learning on non-IID data with reinforcement learning," in *Proceedings of the IEEE INFOCOM 2020-IEEE Conference on Computer Communications*, pp. 1698–1707, Toronto, Canada, 2020.
- [39] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with Non-IID data," 2018, <http://arxiv.org/abs/1806.00582>.
- [40] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [41] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Technical Report, University of Toronto, Toronto, Canada, 2009.