



OPEN

A federated learning differential privacy algorithm for non-Gaussian heterogeneous data

Xinyu Yang & Weisan Wu[✉]

Multi-center heterogeneous data are a hot topic in federated learning. The data of clients and centers do not follow a normal distribution, posing significant challenges to learning. Based on the assumption that the client data have a multivariate skewed normal distribution, we improve the DP-Fed-mv-PPCA model. We use a Bayesian framework to construct prior distributions of local parameters and use expectation maximization and pseudo-Newton algorithms to obtain robust parameter estimates. Then, the clipping algorithm and differential privacy algorithm are used to solve the problem in which the model parameters do not have a display solution and achieve privacy guarantee. Furthermore, we verified the effectiveness of our model using synthetic and actual data from the Internet of vehicles.

In the present increasingly developed technology, various intelligent terminals collect personal data, and because of improvements in storage devices and communication technology, people have begun to focus on the collection and release of high-dimensional complex data; however, the data often contain private information. Considering this situation, Google proposed the federated learning (FL) model¹ such that the clients do not send the original data to the central server, but only need to train the data locally and thereafter send the trained parameters to the central server. Although the raw data are not sent, an attacker can still infer the original data based on the sent training parameters, thereby causing privacy leakage. To ensure client-level privacy, the fashion method uses differential privacy (DP) with rigorous mathematical proofs. Several successful algorithms have combined DP and FL.

Heterogeneous and heavy-tailed data are often encountered in high-dimensional data, and the effects of traditional privacy protection algorithms on such data are often unsatisfactory. To protect client privacy, researchers have attempted to combine DP with different FL models such as those of Agarwal et al.², Asoodeh et al.³, and Geyer et al.⁴. In 2022, Balelli adopted a DP-FL algorithm (**DP-Fed-mv-PPCA**)⁵ for multi-view data. The algorithm assumes that data and latent variables follow a Gaussian distribution, handles statistical heterogeneity and missing view data in local datasets using the Bayesian method, and proves that it guarantees privacy and convergence. FL can be used in various fields, such as using FL in combination with the Internet of Things, edge computing blockchains, and 5G networks, as well as for studying more complex instances with heterogeneous servers.⁶⁻⁸. Simultaneously, we also noted that in the field of vehicle networks, the use of FL methods can not only effectively integrate data generated by various users and vehicles but also protect users' personal privacy. There have been some detailed studies on this subject⁹⁻¹¹, but they are all based on the strong assumption that the data follow a normal distribution, whereas in practice, the data are often asymmetric, heavy-tailed, or long-tailed.

In this study, we use a multivariate skewed normal distribution instead of a Gaussian distribution to improve the robustness of the statistical inference model for client data with heavy-tail and asymmetric characteristics, and to ensure the privacy of client data. A skewed normal distribution was proposed by Azzalini¹² in 1985 on the basis that the data did not meet the requirements of the normal distribution. This distribution has a flexible skewness parameter that describes data asymmetry. After the bias normal distribution was proposed, remarkable success was achieved in data fitting. Based on this, Azzalini proposed a multivariate normal distribution¹³ in 1996 and skewed distribution¹⁴ in 2005. Subsequently, new skewed distributions have been proposed continuously, for instance, skew-laplace distribution¹⁵, skew-cauchy distribution¹⁶, and skew-logistic distribution¹⁷. The main purpose of these distributions is to describe the degree of skew and thicker tail of the data by relying on flexible hyperparameters.

The remainder of this paper is organized as follows: In “Related works” section, we present the work on heterogeneous FL and the contributions of our work. In “Preliminaries” section, we introduce some necessary concepts,

School of Mathematics and Statistics, Baicheng Normal University, Baicheng 137000, China. [✉]email: wuws009@outlook.com

theories, and notation symbols, and provide a detailed analysis of client data with skewed normal distribution, combined with the design based on these conclusions. In “Methods” section, we skew the normal distribution difference of the privacy federal study expectation maximization (EM) algorithm. Furthermore, we verify the effect of our model using synthetic and real data from the Internet of Vehicles in “Performance evaluation” section. Finally, “Conclusions” section summarizes the advantages of the model and discusses future research.

Related works

Several studies have been conducted on heterogeneous data federation learning. The FedProx algorithm proposed by Li et al. in 2018¹⁸ is an improved FedAvg algorithm for partial local work that avoids data heterogeneity by introducing an approximation term. Li considered multi-center scenarios in 2020¹⁹. In the past two years, Zhu considered the knowledge distillation method in 2021²⁰ and Guo studied heterogeneous FL in 5G networks using the dynamic scheduling method in 2022²¹. Similarly, Wang also developed the device sampling method of heterogeneous FL and the node sampling theory of graph convolutional neural network in 2021²². Shen studied a fast algorithm for heterogeneous federation learning in 2022²³ when clients were mixed, which could effectively reduce the variance of the model.

For more theoretical and applied research, please read Fed2KD: Heterogeneous FL for Pandemic Risk Assessment via Two-way Knowledge Distillation by Sun et al.²⁴, HFedMS: Heterogeneous FL with Memorable Data Semantics in Industrial Metaverse by Zeng et al.²⁵, FedRolex: Model-heterogeneous FL with Rolling Sub-model Extraction by Alam et al.²⁶, Lazy Aggregation for Heterogeneous FL by Xu et al.²⁷ and FedFOR: Stateless Heterogeneous FL with First-order Regularization by Tian et al.²⁸ in detail.

All these studies are generalized studies on heterogeneous FL, either to improve computing speed or to focus on the heterogeneity of the client or central server. However, there have been no in-depth studies on irregular data with non-normal scores. Therefore, our study contributes to the literature in several ways.

- We use the EM algorithm to study the parameter estimation problem of client and server data with biased normal structure FL.
- By calculating the sensitivity of the parameters, our algorithm makes the parameters have a small estimation error and ensures the differential privacy during the EM algorithm iteration.
- We applied our algorithm to both synthetic data and real data from the Internet of vehicles. The results demonstrate that our algorithm is more robust and can ensure the accuracy of estimation while achieving DP.

Preliminaries

Model setting. We assume that there are I mutually independent service centers and that each service center $i, i \in \{1, 2, \dots, I\}$ contains the local original dataset $D_i = \{x_{i,n}\}_{i=1}^{n_i}$. Furthermore, we assume that the local dataset is incomplete and missing at least one view in the dataset. For every $g \in 1, \dots, G$, the dimensions of the g thview are d_g , and $d := \sum_{g=1}^G d_g$. We denote by $x_{i,n}^{(g)}$ the raw data of subject n in center i corresponding to the g th-view, and $x_{i,n} = (x_{i,n}^{(1)}, \dots, x_{i,n}^{(g)})$.

Federated multi-views PPCA and multivariate skew normal distribution. In this section, we first review the federated multi-view PPCA model given by Balelli⁵ and assume that the client data have a multivariate skewed normal distribution. Then, we present related concepts and lemmas. Based on Arellano Valle's study²⁹, we extend the theories of joint density and conditional density of multivariate skew normal random variables.

We consider a situation in which the underlying data distribution has a hierarchical structure and assume that the global parameter is $\tilde{\theta}$ and the corresponding density function is $f(\tilde{\theta})$. For each center, the local parameter θ_i is generated by the distribution $f(\theta_i|\tilde{\theta})$. Furthermore, the local data x_i are generated from the local distribution $f(x_i|\theta_i)$. The federated algorithm, known as **Fed-mv-PPCA**, can be used to solve the inverse problem from the local data to the central server in a hierarchical structure using a Bayesian method, and the parameters of this model can be estimated by maximizing the likelihood.

First, we assume that the local data of subject n corresponding to the g thview, $x_{i,n}^{(g)}$, are generated as follows:

$$x_{i,n}^{(g)} = \mu_i^{(g)} + \varepsilon_i^{(g)} + A_i^{(g)} z_{i,n}, \quad (1)$$

where $\varepsilon_i^{(g)} \sim N(0, \sigma_i^{(g)2} I_{d_g})$ denotes a Gaussian noise, $z_{i,n} \sim SN_{d_g}(0, I, \Lambda_i^{(g)})$ denotes a q -dimension latent variable, and we assumed $q < \min_g d_g$ for computation. $A_i^{(g)}$ and $\mu_i^{(g)}$ are the coefficients and offsets with respect to view g , respectively. We find that $A_i^{(g)} z_{i,n}$ follows the multivariate skew normal distribution by Arellano Valle (2005) in Corollary 2.3 as follows²⁹:

$$A_i^{(g)} z_{i,n} \sim SN_{d_g}(0, A_i^{(g)} A_i^{(g)T} \Lambda_i^{(g)}), \quad (2)$$

and

$$\varepsilon_i^{(g)} + \mu_i^{(g)} \sim N_{d_g}(\mu_i^{(g)}, \sigma_i^{(g)2} I_{d_g}), \quad (3)$$

$$x_{i,n}^{(g)} \sim SN_{d_g}(\mu_i^{(g)}, \Sigma_i^{(g)}, \tilde{\Lambda}_i^{(g)}), \quad (4)$$

where $\tilde{\Lambda}_i^{(g)} = (1 + \eta^T \Omega_X^{-1} \eta)^{-\frac{1}{2}} \omega_X \Omega_X^{-1} \Omega \eta$, $\eta = \omega^{-1} A_i^{(g)} \Lambda_i^{(g)}$, $\omega_X = (\Omega_X \odot I)^{\frac{1}{2}}$, $\Sigma_i^{(g)} = A_i^{(g)} A_i^{(g)T} + \sigma_i^{(g)2} I_{d_g}$, and \odot is the Hadamard product. Then, concatenate all views $x_{i,n}^{(g)}$ into a compact formulation from Equation (1) as follows:

$$x_{i,n} = \mu_i + \Phi_i + A_i z_{i,n}, \quad (5)$$

where A_i, μ_i are obtained by concatenating all $A_i^{(g)}, \mu_i^{(g)}$, and Φ_i denotes a block diagonal matrix with the g th block from $\varepsilon_i^{(g)}$. From the described analysis process, we can denote the local parameter as follows:

$$\theta_i := \{\mu_i^{(g)}, A_i^{(g)}, \sigma_i^{(g)2}, \Lambda_i^{(g)}\}. \quad (6)$$

We consider a Bayesian approach to (5) in which θ_i is regarded as random with a prior distribution that reflects the degree of belief in the different values of these quantities. The prior distributions selected are weakly informative and subject to vague prior knowledge, and this avoids non-integrable posterior distributions. The prior distributions for the model are as follows:

$$\begin{aligned} (\mu_i^{(g)} | \tilde{\mu}^{(g)}, \sigma_{\tilde{\mu}}) &\sim N_{d_g}(\tilde{\mu}^{(g)}, \sigma_{\tilde{\mu}}^2), \\ (A_i^{(g)} | \tilde{A}^{(g)}, \sigma_{\tilde{A}^{(g)}}) &\sim N_{d_g}(\tilde{A}^{(g)}, \sigma_{\tilde{A}^{(g)}}^2), \\ (\sigma_i^{(g)2} | \tilde{\beta}^{(g)}, \tilde{\gamma}^{(g)}) &\sim \Gamma(\tilde{\beta}^{(g)}, \tilde{\gamma}^{(g)}), \\ (\delta(\lambda_i^{(g)}) | \lambda^{(g)}) &\sim U(-1, 1). \end{aligned} \quad (7)$$

Thus, we denote the global prior distribution by parameter $\tilde{\theta}_i = \{\tilde{\mu}^{(g)}, \sigma_{\tilde{\mu}}, \tilde{A}^{(g)}, \sigma_{\tilde{A}^{(g)}}, \tilde{\beta}^{(g)}, \tilde{\gamma}^{(g)}, \tilde{\Lambda}^{(g)}\}$.

Second, we need to concrete analyze the joint and conditional density functions of the observed and latent variables. Therefore, we introduce the Arellano Valle 2005 standard skewed normal distribution theory and extend it to general location and scale cases³⁰.

Lemma 3.1 *If $Z_1 \sim SN_m(\Lambda_1)$, $Z_2 \sim SN_m(\Lambda_2)$, then the conditional density of Z_1 given $Z_2 = z_2$ is as follows:*

$$f_{Z_1|Z_2=z_2}(z_1) = \phi_{m_1}(z_1) \Phi_{m_1+m_2}(\Lambda_1^T z_1 + \Lambda_2^T z_2 | I_{m_1+m_2} - \Lambda_1^T \Lambda_1 - \Lambda_2^T \Lambda_2) / \Phi_{m_2}(\Lambda_2^T z_2 | I_{m_2} - \Lambda_2^T \Lambda_2) \quad (8)$$

Theorem 3.2 *If $W_1 = \mu_1 + \Sigma_1^{\frac{1}{2}} Z_1$, $Z_1 \sim SN_{m_1}(\Lambda_1)$, $W_2 = \mu_2 + \Sigma_2^{\frac{1}{2}} Z_2$, $Z_2 \sim SN_{m_2}(\Lambda_2)$, then the conditional density of W_1 given $W_2 = w_2$ is as follows:*

$$\begin{aligned} f_{W_1|W_2=w_2}(w_1) = & |\Sigma_1 - \Sigma_{12} \Sigma_2^{-1} \Sigma_{12}|^{-\frac{1}{2}} |\Sigma_2|^{-\frac{1}{2}} \phi_{m_1}(\Sigma_1^{-\frac{1}{2}}(w_1 - \mu_1)) \\ & \Phi_{m_1+m_2}(\Lambda_1^T \Sigma_1^{-\frac{1}{2}}(w_1 - \mu_1) + \Lambda_2^T \Sigma_2^{-\frac{1}{2}}(w_2 - \mu_2) | I_{m_1+m_2} - \Lambda_1^T \Lambda_1 - \Lambda_2^T \Lambda_2) / \\ & \Phi_{m_2}(\Lambda_2^T \Sigma_2^{-\frac{1}{2}}(w_2 - \mu_2) | I_{m_2} - \Lambda_2^T \Lambda_2). \end{aligned} \quad (9)$$

Proof Because of $W_1 = \mu_1 + \Sigma_1^{\frac{1}{2}} Z_1$, $Z_1 \sim SN_{m_1}(\Lambda_1)$, $W_2 = \mu_2 + \Sigma_2^{\frac{1}{2}} Z_2$, $Z_2 \sim SN_{m_2}(\Lambda_2)$, we know about $W_1 \sim SN_{m_1}(\mu_1, \Sigma_1, \Lambda_1)$, $W_2 \sim SN_{m_2}(\mu_2, \Sigma_2, \Lambda_2)$. Thus, the random variable W_1 has the following density function:

$$\begin{aligned} f_{W_1^*}(w_1) = & 2^m |\Sigma_1|^{-\frac{1}{2}} \phi_{k_1}(\Sigma_1^{-\frac{1}{2}}(w_1 - \mu_1)) \Phi_{m_1}(\Lambda_1^T \Sigma_1^{-\frac{1}{2}}(w_1 - \mu_1) | I_{m_1} - \Lambda_1^T \Lambda_1) f_{W_2^*}(w_2) \\ = & 2^m |\Sigma_2|^{-\frac{1}{2}} \phi_{k_2}(\Sigma_2^{-\frac{1}{2}}(w_2 - \mu_2)) \Phi_{m_2}(\Lambda_2^T \Sigma_2^{-\frac{1}{2}}(w_2 - \mu_2) | I_{m_2} - \Lambda_2^T \Lambda_2). \end{aligned} \quad (10)$$

W_1, W_2 have a joint density function, as follows:

$$\begin{aligned} f_{W_1, W_2}(w_1, w_2) = & 2^{m_1+m_2} |\Sigma_1 - \Sigma_{12} \Sigma_2^{-1} \Sigma_{12}|^{-\frac{1}{2}} |\Sigma_2|^{-\frac{1}{2}} \phi_{m_1}(\Sigma_1^{-\frac{1}{2}}(w_1 - \mu_1)) \phi_{k_2}(\Sigma_2^{-\frac{1}{2}}(w_2 - \mu_2)) \\ & \Phi_{m_1+m_2}(\Lambda_1^T \Sigma_1^{-\frac{1}{2}}(w_1 - \mu_1) + \Lambda_2^T \Sigma_2^{-\frac{1}{2}}(w_2 - \mu_2) | I_{m_1+m_2} - \Lambda_1^T \Lambda_1 - \Lambda_2^T \Lambda_2). \end{aligned} \quad (11)$$

Furthermore, the conditional density function of W_1 given $W_2 = w_2$ is as follows:

$$f_{W_1|W_2=w_2}(w_1) = \frac{\phi_{m_1}(\Sigma_1^{-\frac{1}{2}}(w_1 - \mu_1)) \Phi_{m_1+m_2}(\Lambda_1^T \Sigma_1^{-\frac{1}{2}}(w_1 - \mu_1) + \Lambda_2^T \Sigma_2^{-\frac{1}{2}}(w_2 - \mu_2) | I_{m_1+m_2} - \Lambda_1^T \Lambda_1 - \Lambda_2^T \Lambda_2)}{|\Sigma_1 - \Sigma_{12} \Sigma_2^{-1} \Sigma_{12}|^{\frac{1}{2}} |\Sigma_2|^{\frac{1}{2}} \Phi_{m_2}(\Lambda_2^T \Sigma_2^{-\frac{1}{2}}(w_2 - \mu_2) | I_{m_2} - \Lambda_2^T \Lambda_2)}. \quad (12)$$

Specifically, we can express the $(x_{i,n}, z_{i,n})$ joint density as follows:

$$\begin{aligned}
f(x_{i,n}, z_{i,n}) = & 2^{d+q} |A_i A_i^T + \sigma_i^2 I_{d_g}|^{-\frac{1}{2}} \phi_{d_g}(|A_i A_i^T + \sigma_i^2 I_{d_g}|^{-\frac{1}{2}} (x_{i,n} - \mu_i)) \phi_q(z_{i,n}) \\
& \Phi_{d_g+q}(\tilde{\Lambda}_i^T |A_i A_i^T + \sigma_i^2 I_{d_g}|^{-\frac{1}{2}} (x_{i,n} - \mu_i) + \tilde{\Lambda}_i^T z_{i,n} |I_{d_g+q} - \tilde{\Lambda}_i^T \tilde{\Lambda}_i - \Lambda_i^T \Lambda_i|) \\
= & 2^{d+q} |A_i A_i^T + \sigma_i^2 I_{d_g}|^{-\frac{1}{2}} 2^d \exp -\frac{1}{2} (x_{i,n} - \mu_i) |A_i A_i^T + \sigma_i^2 I_{d_g}|^{-\frac{1}{2}} (x_{i,n} - \mu_i)^T \\
& \Phi_{d_g+q}(\tilde{\Lambda}_i^T |A_i A_i^T + \sigma_i^2 I_{d_g}|^{-\frac{1}{2}} (x_{i,n} - \mu_i) + \tilde{\Lambda}_i^T z_{i,n} |I_{d_g+q} - \tilde{\Lambda}_i^T \tilde{\Lambda}_i - \Lambda_i^T \Lambda_i|).
\end{aligned} \tag{13}$$

Therefore, the conditional density function of $z_{i,n}$ given $x_{i,n}$ is as follows:

$$f(z_{i,n} | x_{i,n}) = 2^q \phi_q(z_{i,n}) \frac{\Phi_{d_g+q}(\tilde{\Lambda}_i^T |A_i A_i^T + \sigma_i^2 I_{d_g}|^{-\frac{1}{2}} (x_{i,n} - \mu_i) + \tilde{\Lambda}_i^T z_{i,n} |I_{d_g+q} - \tilde{\Lambda}_i^T \tilde{\Lambda}_i - \Lambda_i^T \Lambda_i|)}{\Phi_{d_g}(\tilde{\Lambda}_i^T |A_i A_i^T + \sigma_i^2 I_{d_g}|^{-\frac{1}{2}} (x_{i,n} - \mu_i) |I_{d_g} - \tilde{\Lambda}_i^T \tilde{\Lambda}_i|)}.
\tag{14}$$

By combining Equations (13) and (14), we obtain the posterior logarithmic likelihood function of the complete data as follows:

$$\begin{aligned}
\log f(x_{i,n}, z_{i,n} | \theta_i) = & c + \sum_{n=1}^{n_i} \sum_{g=1}^G \left[\frac{d_g}{2} \log |A_i A_i^T + \sigma_i^2 I_{d_g}| + \frac{1}{2} \text{tr}((A_i A_i^T + \sigma_i^2 I_{d_g}))^{-1} \|x_{i,n} - \mu_i\|^2 \right. \\
& + \log [\Phi_{d_g+q}(\tilde{\Lambda}_i^T |A_i A_i^T + \sigma_i^2 I_{d_g}|^{-\frac{1}{2}} (x_{i,n} - \mu_i) \\
& \left. + \tilde{\Lambda}_i^T z_{i,n} |I_{d_g+q} - \tilde{\Lambda}_i^T \tilde{\Lambda}_i - \Lambda_i^T \Lambda_i|) + \frac{1}{2} \text{tr}(\langle z_{i,n}, z_{i,n}^T \rangle) \right].
\end{aligned} \tag{15}$$

□

Methods

In this section, we present the detailed process and algorithm of parameter estimation and the DP algorithm with a privacy guarantee.

Federated multi-views PPCA for non-Gaussian data. Based on the described analysis of the observed and latent variables, we use the EM and pseudo-Newton algorithms to estimate the parameters locally, release all the local parameters to the master, aggregate the parameters at the master, and estimate the global parameters using maximum likelihood estimation (Algorithm 1).

Algorithm 1 Fed-mv-PPCA

Input: Rounds R , Iterations T , latent dimensional q ,

Output: $\tilde{\theta}$,

```

for  $r = 0, 1, \dots, R$  do,
  for  $i = 1, \dots, I$  do in parallel,
    Each center  $i$  initializes  $\theta_i$  using  $f(\theta_i | \tilde{\theta})$ ;
     $T$  iterations of MAP estimation of  $\theta_i$  using  $\tilde{\theta}$  as prior;
  end for
  Each center  $i$  returns  $\theta_i$  to the master;
  The master collects  $\theta_i, i = 1, \dots, I$  and estimates  $\tilde{\theta}$  through ML;
  The master sends  $\tilde{\theta}$  to all centers,
end for.

```

The parameter estimation in Algorithm 1 is as follows: For simplicity, we take the parameter μ as an example and assume that in center i , the g th-view is missing, instead of simply removing it. We assume $\forall i, g$.

$$\mu_i^{(g)} \sim N_{d_g}(\tilde{\mu}^{(g)}, \sigma_{\mu}^2 I_{d_g}), \tag{16}$$

Step 1. In each center: Estimate $\mu_i^{(g)}[s+1] | (\tilde{\mu}^{(g)}, \sigma_{\mu}^2)[s]$ from Equation (1). The marginal distribution of $x_{i,n}^{(g)}$ is as follows:

$$x_{i,n}^{(g)} \sim SN_{d_g}(\tilde{\mu}_i^{(g)}, \Sigma_i^{(g)}, \tilde{\Lambda}_i^{(g)}), \tag{17}$$

Recall:

$$f(x_{i,n}^{(g)}) = 2(2\pi)^{-\frac{d_g}{2}} |\Sigma_i^{(g)}|^{-\frac{1}{2}} \exp -\frac{1}{2}(x_{i,n}^{(g)} - \mu_i^{(g)})^T (\Sigma_i^{(g)})^{-1} (x_{i,n}^{(g)} - \mu_i^{(g)}) \Phi((\tilde{\Lambda}_i^{(g)})^T (\Sigma_i^{(g)})^{-1} (x_{i,n}^{(g)} - \mu_i^{(g)})) \quad (18)$$

the log-likelihood function is:

$$\begin{aligned} \mathcal{L}_i^{(g)} = & -\frac{1}{2} \{N_i d_g \log(2\pi) \log |\Sigma_i^{(g)}| + \sum_{n=1}^{N_i} (x_{i,n}^{(g)} - \mu_i^{(g)})^T (\Sigma_i^{(g)})^{-1} (x_{i,n}^{(g)} - \mu_i^{(g)}) + \\ & \zeta_0((\tilde{\Lambda}_i^{(g)})^T (\Sigma_i^{(g)})^{-1} (x_{i,n}^{(g)} - \mu_i^{(g)}))\}, \end{aligned} \quad (19)$$

where $\zeta_0(x) = \log 2x$. Subsequently, we fixed the center order i and considered the following optimization problem for $\forall g \in 1, \dots, G$,

$$\max_{\mu_i^{(g)}} \mathcal{L}_i^{(g)} + \log f(\mu_i^{(g)}), \quad (20)$$

where $\log f(\mu_i^{(g)}) = -\frac{1}{2\sigma_{\tilde{\mu}^{(g)}}^2} (\mu_i^{(g)} - \tilde{\mu}^{(g)})^T (\mu_i^{(g)} - \tilde{\mu}^{(g)}) + c1$, $c1$ is a constant independent of $\mu_i^{(g)}$. Using the maximum-likelihood method for this problem, we obtain the following:

$$\mu_i^{(g)}[s+1] = [N_i I_{d_g} + \sigma_{\tilde{\mu}^{(g)}}^{-2}[s] \Sigma_i^{(g)}]^{-1} \left[\sum_{i=1}^{N_i} x_{i,n}^{(g)} + \sigma_{\tilde{\mu}^{(g)}}^{-2}[s] \Sigma_i^{(g)} \tilde{\mu}^{(g)}[s] \right]. \quad (21)$$

Step 2. In the master: Estimate $(\tilde{\mu}^{(g)}[s+1], \sigma_{\tilde{\mu}^{(g)}}^2[s+1])$ given $\mu^{(g)}[s+1]$ for each i .

$$\begin{aligned} \mathcal{L} = & \sum_{i=1}^I \log f(\mu_i^{(g)}) \\ = & \sum_{i=1}^I \{c1 - \frac{1}{\sigma_{\tilde{\mu}^{(g)}}^2} \|\mu_i^{(g)} - \tilde{\mu}^{(g)}\|^2\}. \end{aligned} \quad (22)$$

Let $\frac{\partial \mathcal{L}}{\partial \tilde{\mu}^{(g)}} = 0$, and $\frac{\partial \mathcal{L}}{\partial \sigma_{\tilde{\mu}^{(g)}}^2} = 0$, we can obtain:

$$\tilde{\mu}^{(g)}[s+1] = I^{-1} \sum_{i=1}^{N_i} \mu_i^{(g)}[s+1], \quad (23)$$

and

$$\sigma_{\tilde{\mu}^{(g)}}^{-2}[s+1] = (Id_g)^{-1} \sum_{i=1}^{N_i} \|\mu_i^{(g)}[s+1] - \tilde{\mu}^{(g)}[s+1]\|^2, \quad (24)$$

respectively.

We performed similar treatments for the other parameters. Notably, although the skewness parameter depends on both the location and scale parameters, the prior distribution of the transformed parameter δ is uniform under the Bayesian framework.

DP-Fed-mv-PPCA algorithm for non-Gaussian data. In this section, we propose an improved DP-Fed-mv-PPCA algorithm based on clients with skewed normal data. Our main goal is to protect the privacy of the training parameters of the local data, which requires the use of different noise mechanisms for different components to perturb the data properly.

First, let us provide the basic concept of DP and classical combinatorial theory. Then, we will gradually use these theories in the algorithm.

Definition 1 Given a data universe \mathcal{X} , we say that two datasets D, D' are neighbors if they differ by only one entry, which is denoted as $D \tilde{D}' \subseteq \mathcal{X}$. A randomized algorithm \mathcal{M} is (ε, δ) -differentially private (DP) if for all neighboring datasets D, D' and for all events S in the output space of \mathcal{M} , we have $\mathbb{P}(\mathcal{M}(D) \in S) \leq e^\varepsilon \mathbb{P}(\mathcal{M}(D') \in S) + \delta$.

Definition 2 Given a function $q : \mathcal{X}^n \rightarrow \mathbb{R}^d$, the Gaussian Mechanism is defined as: $\mathcal{M}(D; q; \cdot) = q(D) + Noise$; where $Noise$ is drawn from a Gaussian distribution $N(0, \sigma^2 I_d)$ with $\sigma \geq \frac{\sqrt{2 \ln 1.25/\delta} \Delta_2(q)}{\varepsilon}$, $\Delta_2(q)$ is the l_2 -sensitivity of the function q , i.e., $\Delta_2(q) = \sup_{D, D'} \|q(D) - q(D')\|_2$. The Gaussian mechanism preserves (ε, δ) -DP.

Definition 3 Given a function $q : \mathcal{X}^n \rightarrow \mathbb{R}^d$, the Laplace mechanism is defined as: $\mathcal{M}(D; q; \cdot) = q(D) + Noise$; where $Noise$ is drawn from a Laplace distribution $Lap(0, \frac{\Delta_1(q)}{\varepsilon})$, $\Delta_1(q)$ is the l_1 -sensitivity of the function q , i.e., $\Delta_1(q) = \sup_{D, D'} \|q(D) - q(D')\|_1$. The Laplace mechanism preserves $(\varepsilon, 0)$ -DP.

Lemma 4.1 Given any function $q : \mathcal{X}^n \rightarrow \mathbb{R}^d$ and $\varepsilon > 0, \delta \in (0, 0.5)$, the Gaussian mechanism is defined as follows:

$$\mathcal{M}(D; q; \cdot) = q(D) + N(\mathbf{0}, (\frac{(c + \sqrt{c^2 + \varepsilon})\Delta_2(q)}{\sqrt{2\varepsilon}})^2 I_d), \quad (25)$$

where $c = \sqrt{\ln(2/(\sqrt{16\varepsilon + 1} - 1))}$ preserves (ε, δ) -DP.

Lemma 4.2 Given any function $q : \mathcal{X}^n \rightarrow \mathbb{R}^{d \times q}$ and $\varepsilon > 0, \delta \in (0, 0.5)$, the matrix-normal mechanism is defined as follows:

$$\mathcal{M}(D; q; \cdot) = q(D) + N_{d,q}(\mathbf{0}_{d,q}, (\frac{(c + \sqrt{c^2 + \varepsilon})\Delta_2(q)}{\sqrt{2\varepsilon}})^2 I_d), \quad (26)$$

where $c = \sqrt{\ln(2/(\sqrt{16\varepsilon + 1} - 1))}$ preserves (ε, δ) -DP.

Lemma 4.3 For $i = 1, \dots, k$ let \mathcal{M}_i be an $(\varepsilon_i, \delta_i)$ -DP algorithm, and $\mathcal{M}(q) = (\mathcal{M}_1(q), \dots, \mathcal{M}_k(q))$. Then, \mathcal{M} is $(\sum_{i=1}^k \varepsilon_i, \sum_{i=1}^k \delta_i)$ -DP.

Although the algorithm does not show an expression for client-level optimization, we can ensure that the local parameters achieve DP with appropriate clipping methods, such as Zhang et al.³¹ (see Algorithm 2 for the detailed procedure).

Algorithm 2 DP-Fed-mv-PPCA algorithm for Non-Gaussian data (DFmP-SN)

Input: Rounds R , Iterations T , latent dimensional q , parameter ε, δ ,

Output: $\tilde{\theta}$,

for $r = 0, 1, \dots, R$ **do**,

for $i = 1, \dots, I$ **do** in parallel,

 Initialize θ_i using $f(\theta_i | \tilde{\theta}[r])$;

 Update local parameters: T iterations of MAP estimation (EM + prior) to optimize $\theta_i[r+1]$ using $\tilde{\theta}[r]$ as prior;

 Compute difference: $\Delta\theta_i[r+1] = \theta_i[r+1] - \tilde{\theta}[r]$;

 Clip: $\bar{\Delta}\theta_i[r+1] = \Delta\theta_i[r+1] / \max\{1, \|\Delta\theta_i[r+1]\|_p / g(\sigma_{\tilde{\theta}}[r])\}$;

 Add noise: $\mathcal{M}_{\theta_i}[r+1] = \bar{\Delta}\theta_i[r+1] + \text{Noise}(2g(\sigma_{\tilde{\theta}}[r]), \varepsilon, \delta)$;

 Return $\tilde{\theta}_i[r+1] = \mathcal{M}_{\theta_i}[r+1] + \tilde{\theta}[r]$ to the master;

end for.

 The master collects all $\tilde{\theta}_i[r+1]$ and estimates $\tilde{\theta}[r+1]$ through maximized likelihood method;

 The master sends $\tilde{\theta}[r+1]$ to all centers;

end for.

With respect to Algorithm 1, we need to perform appropriate clipping when the local dataset uses the EM and textbf{pseudo-Newton} algorithms to estimate the parameters, and then add noise from different mechanisms (Laplacian or Gaussian³²) to the parameters to ensure that the data released to the master achieve DP³³. The specific execution process is as follows:

1. The client provides the prior distribution in round r and updates the local parameters according to Algorithm 1 based on the prior distribution. The client calculates the difference between the updated parameters in round $r+1$ and the prior parameters in round r .

$$\Delta\theta_i[r+1] = \theta_i[r+1] - \tilde{\theta}[r]. \quad (27)$$

2. The updated difference is clipped according to the standard deviation of the prior values.

$$\bar{\Delta}\theta_i[r+1] = \Delta\theta_i[r+1] / \max\{1, \|\Delta\theta_i[r+1]\|_p / g(\sigma_{\tilde{\theta}}[r])\}, \quad (28)$$

where $g(\sigma_{\tilde{\theta}}[r]) := c2 \cdot \tilde{\theta}[r]$ and $c2$ is a constant fixed by the client. The clipping upper bound of ℓ_p norm of $\bar{\Delta}\theta_i[r+1]$ is $g(\sigma_{\tilde{\theta}}[r])$, and the ℓ_p sensitivity of $\bar{\Delta}\theta_i[r+1]$ at most $2g(\sigma_{\tilde{\theta}}[r])$.

3. We propose a noise mechanism for the clipped difference using

$$\mathcal{M}_{\theta_i}[r+1] := \bar{\Delta}\theta_i[r+1] + \text{Noise}(2g(\sigma_{\tilde{\theta}}[r]), \varepsilon, \delta). \quad (29)$$

We propose to add Gaussian noise to $\bar{\Delta}\mu_i^{(g)}$ and $\bar{\Delta}A_i^{(g)}$, while adding Laplace noise to $(\bar{\Delta}\sigma_i^{(g)})^2$ and $\bar{\Delta}\delta_i^{(g)}$.

4. The client adds a priori and sends $\bar{\Delta}\theta_i[r+1] = \mathcal{M}_{\theta_i}[r+1] + \tilde{\theta}[r]$ to the master.

Figures 1 and 2 shows the graphical model of Fed-mv-PPCA and iterative flow diagram of component parameters of the skew normal mixture model.

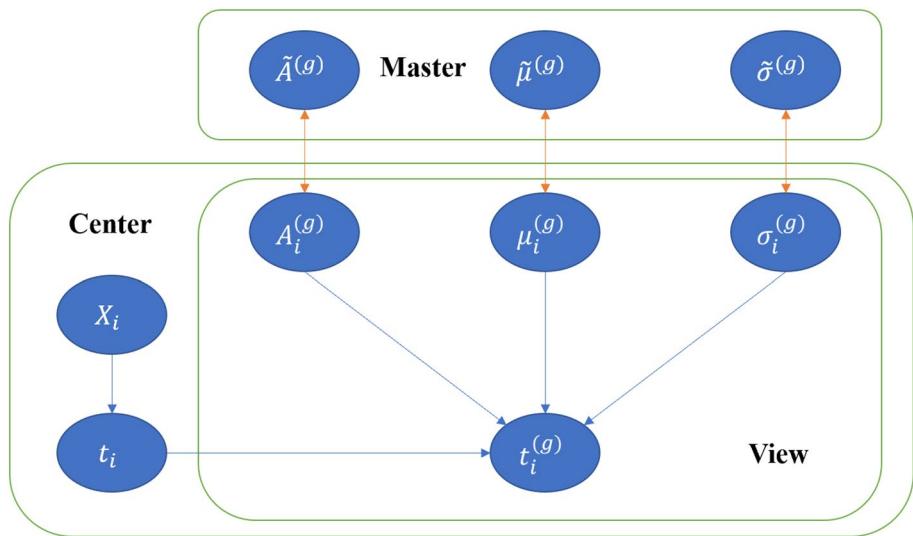


Figure 1. Graphical model of Fed-mv-PPCA.

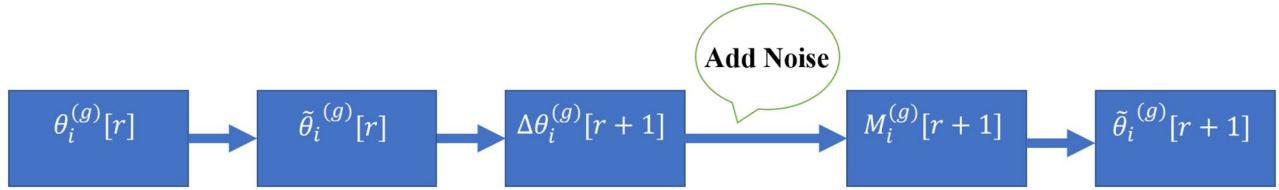


Figure 2. Parameter iteration process of algorithm DFmP-SN.

We provide the privacy budget for Algorithm 1 using the following theorem:

Theorem 4.4 *For simplicity, let us select the same ε, δ for all the mechanisms considered (a generalization to a parameter-specific choice of ε_i, δ_i is straightforward). The total privacy budget for the output of Algorithm 2 is $(4G\varepsilon, 3G\delta)$, where K denotes the total number of views.*

Notably, the data in each center were disjoint. At all centers, we consider the mechanism $\mathcal{M} := \{\mathcal{M}_{\mu_i^{(g)}}, \mathcal{M}_{A_i^{(g)}}, \mathcal{M}_{\sigma_i^{(g)}}, \mathcal{M}_{\delta(\lambda)_i^{(g)}}\}$, where for all g , $\mathcal{M}_{\mu_i^{(g)}}$ and $\mathcal{M}_{A_i^{(g)}}$ are (ε, δ) -DP, whereas for all g , $\mathcal{M}_{\sigma_i^{(g)}}$ and $\mathcal{M}_{\delta(\lambda)_i^{(g)}}$ are ε -DP. We can obtain the results from the composition theorem and post processing of DP.

Performance evaluation

Introduction to dataset. Synthesis data (SD): A total of 10,000 observations were generated using the generation model, which comprised $k = 3$ views with dimensions of $d_1 = 2$, $d_2 = 10$, and $d_3 = 5$. Each view was generated using a common 5-dimensional potential space. We selected the parameters $A^{(g)}, \mu^{(g)}, \sigma^{(g)}$ randomly.

Didi Chuxing GAIA data (DiDi): The data security of connected car users is a concern. In this section, as in [9], we use car-networking data from Didi Chuxing GAIA Initiative (<https://gaia.didichuxing.com>). We selected 10,000 vehicle data points from the driver trajectory data of the Chengdu Second Ring Road from 00 : 00 to 24 : 00 on October 15, 2019. The longitude and latitude coordinates in the dataset were also converted through WGS84 coordinate system and thereafter converted into east-north coordinates in meters using the UTM grid point system. We also used the binary Gaussian random error in [9].

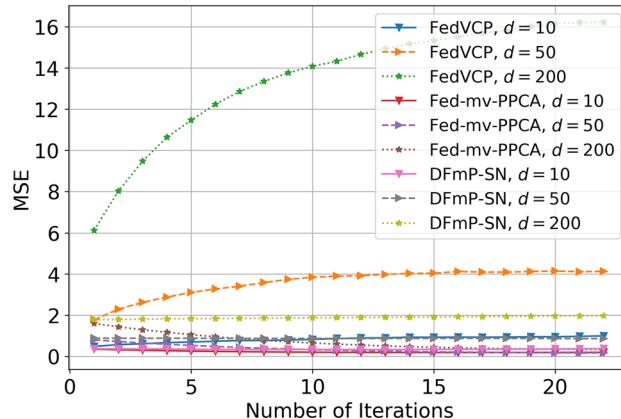
Results. To verify that the non-Gaussian data fitting is more effective in the actual data, in the DiDi dataset, we first performed a mixed biased normal distribution fitting on the data and compared the fitting effect when the number of components was 1–8, which significantly improved the accuracy of the normal fitting compared with [9]. From the results, the best fit was achieved when the component was $g = 5$. For the comparison index of the experiment, we used the common evaluation indices MSE and mean absolute error (MAE). Further results are summarized in Table 1 along with AIC and BIC results, which are caused by the complex traffic condition data structures.

To verify that our algorithm can provide both privacy assurance and data availability, we compared our method with Fed-mv-PPCA in the SD and DiDi datasets and with the FedVCP method in the dataset.

	1	2	3	4	5	6	7	8
k (Gaussian)								
AIC	405.12	415.11	427.55	433.84	439.07	434.91	430.10	431.78
BIC	432.22	445.32	451.29	470.31	481.19	467.28	472.23	473.18
MSE	0.21	0.19	0.15	0.14	0.13	0.12	0.12	0.12
MAE	0.19	0.14	0.13	0.11	0.07	0.07	0.07	0.06
k (Non-Gaussian)								
AIC	472.11	477.36	479.35	480.85	490.01	487.75	476.68	477.79
BIC	500.15	521.12	522.23	531.11	543.33	540.59	540.01	522.55
MSE	0.09	0.07	0.07	0.04	0.02	0.02	0.02	0.02
MAE	0.07	0.06	0.06	0.03	0.01	0.01	0.01	0.01

Table 1. Results of model selection for gaussian and non-gaussian on SD dataset.

	1	2	3	4	5	6	7	8
k (Gaussian)								
AIC	1789.25	1792.11	1840.22	1823.33	1811.05	1766.55	1744.56	1705.63
BIC	1822.05	1827.66	1859.87	1845.55	1833.78	1821.25	1801.01	1796.34
MSE	0.17	0.11	0.09	0.09	0.09	0.09	0.09	0.07
MAE	0.19	0.12	0.10	0.10	0.10	0.10	0.09	0.09
k (Non-Gaussian)								
AIC	1925.23	1935.53	1979.99	1976.52	1927.27	1900.35	1892.25	1888.87
BIC	1975.57	1986.68	2016.64	2008.98	2001.13	1999.76	1989.68	1980.35
MSE	0.11	0.10	0.04	0.04	0.03	0.03	0.02	0.02
MAE	0.05	0.04	0.02	0.02	0.012	0.01	0.01	0.01

Table 2. Results of model selection for gaussian and non-gaussian on DiDi dataset.**Figure 3.** Effect comparison of three algorithms in different dimensions of MSE (FedVCP, Fed-mv-PPCA, DFmP-SN) on DiDi dataset, $\varepsilon = 0.1$.

For the SD dataset and DiDi dataset, the calculation results (see Tables 1 and 2) demonstrate that for different privacy parameters, our algorithm can achieve privacy guarantee and is clearly better than FedVCP. For the DiDi dataset, our algorithm achieved a stronger privacy guarantee. It can be clearly seen from Figs. 3, 4, 5, 6, 7, 8 and 9 that our estimation error is smaller in MSE.

Conclusions

In this study, we introduce a skewed normal distribution as the non-Gaussian data likelihood and prior. Our application proves that **Fed-mv-PPCA** is robust to increasing levels of heterogeneity in training centers, and provides high-quality data reconstruction that outperforms competing methods in all scenarios. In addition, when DP was introduced, we investigated the performance of the proposed method according to different privacy

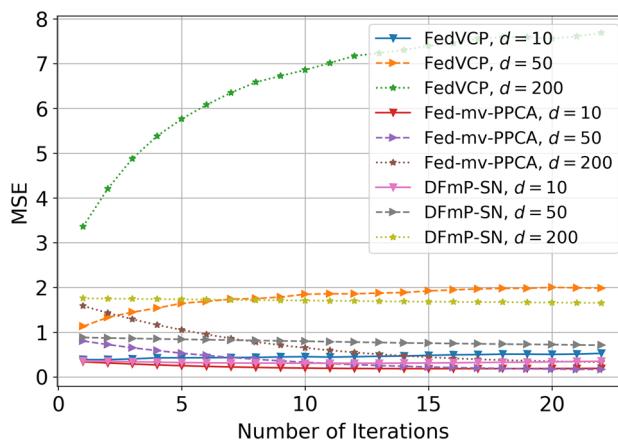


Figure 4. Effect comparison of three algorithms in different dimensions of MSE (FedVCP, Fed-mv-PPCA, DFmP-Sn) on DiDi dataset, $\varepsilon = 0.2$.

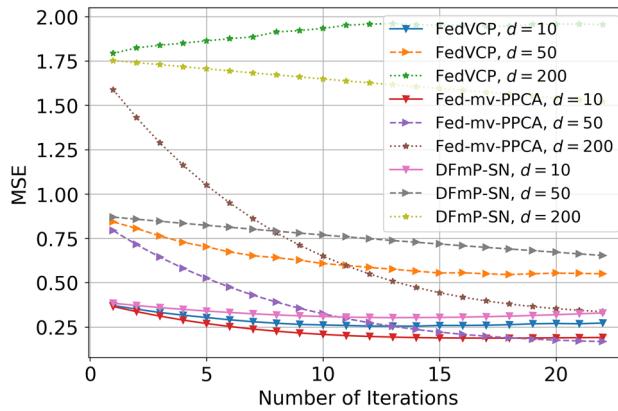


Figure 5. Effect comparison of three algorithms in different dimensions of MSE (FedVCP, Fed-mv-PPCA, DFmP-Sn) on DiDi dataset, $\varepsilon = 0.8$.

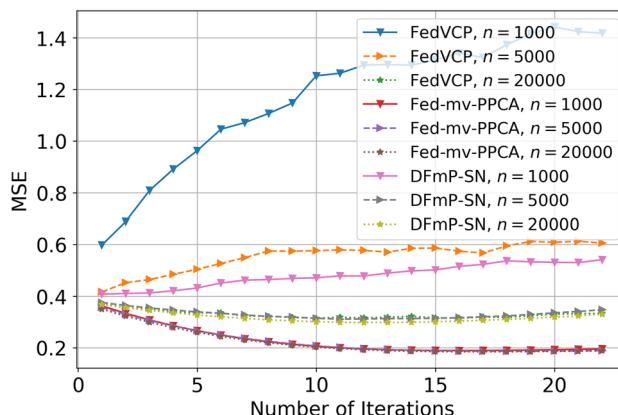


Figure 6. Effect comparison of three algorithms in different samples of MSE (FedVCP, Fed-mv-PPCA, DFmP-Sn) on DiDi dataset, $\varepsilon = 0.1$.

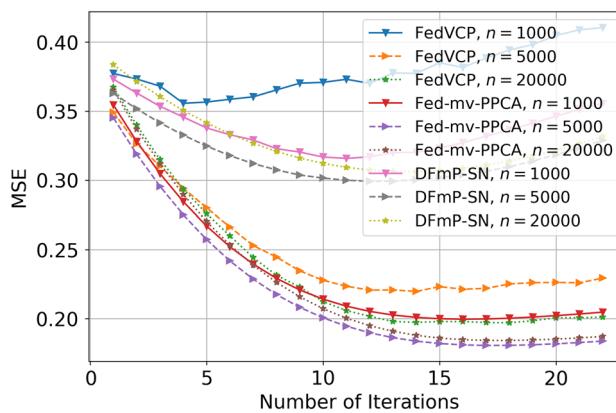


Figure 7. Effect comparison of three algorithms in different samples of MSE (FedVCP, Fed-mv-PPCA, DFmP-SN) on DiDi dataset, $\varepsilon = 0.2$.

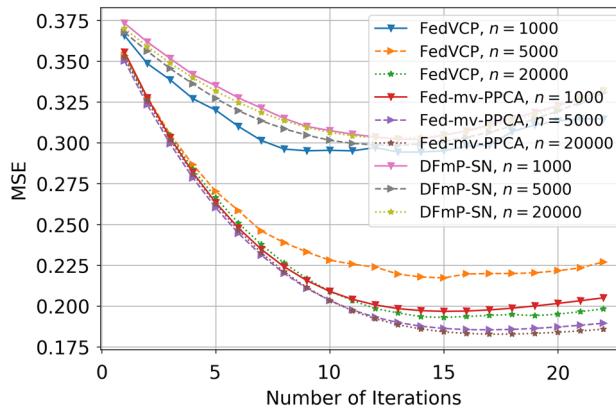


Figure 8. Effect comparison of three algorithms in different samples of MSE (FedVCP, Fed-mv-PPCA, DFmP-SN) on DiDi dataset, $\varepsilon = 0.8$.

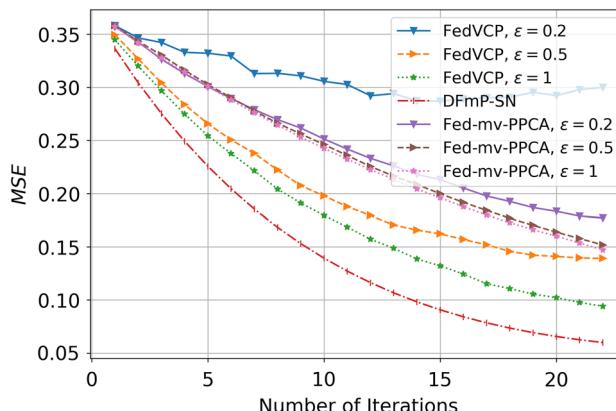


Figure 9. Effect comparison of three algorithms in different budget of MSE (FedVCP, Fed-mv-PPCA, DFmP-SN) on DiDi dataset, $d = 10$.

budget scenarios. Notably, there are four DP hyperparameters that play a key role and may influence the properties of **DP-Fed-mv-PPCA**: the privacy budget parameters (ε, δ), and the clipping and skewness parameters $\delta(\lambda)$. These parameters are closely related and both help determine the magnitude of the noise used to perturb the updated difference $\Delta\theta$. Indeed, increasing δ or decreasing the multiplicative constant in the clipping mechanism

implies the addition of less noise, thereby improving the overall utility of the global model. However, the smaller the values of ϵ and δ , the higher the privacy guarantee.

There is much more work to be conducted in the future. First, reducing the computational and sample complexities of the DP calculation between the client and server is the key problem because heterogeneous data with multiple sources often face the problem of small data. Second, different noise mechanisms are also the main direction of subsequent studies, and we note that matrix variables and tensor data are also challenges in modern FL. Finally, we studied distributions with thicker tails and shear methods suitable for high-dimensional data to improve the robustness of the model. Meanwhile, for outlier datasets, the Huber contamination model can be considered for the FL of DP.

Data availability

The datasets generated and/or analyzed in the present study are not publicly available, but are available from the corresponding author upon reasonable request.

Received: 16 December 2022; Accepted: 6 April 2023

Published online: 10 April 2023

References

- McMahan, H. B., Moore, E., Ramage, D., Hampson, S. & y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In *AISTATS* (2017).
- Agarwal, N., Kairouz, P. & Liu, Z. The skellam mechanism for differentially private federated learning. In *NeurIPS* (2021).
- Asoodeh, S., Chen, W.-N., du Pin Calmon, F. & Özgür, A. Differentially private federated learning: An information-theoretic perspective. *2021 IEEE International Symposium on Information Theory (ISIT)* 344–349 (2021).
- Geyer, R. C., Klein, T. & Nabi, M. Differentially private federated learning: A client level perspective. [ArXiv:1712.07557](https://arxiv.org/abs/1712.07557) (2017).
- Balelli, I., SantiagoS.Silva, R. & Lorenzi, M. A differentially private probabilistic framework for modeling the variability across federated datasets of heterogeneous multi-view observations. *n. pag* [arXiv:2204.07352](https://arxiv.org/abs/2204.07352) (2022).
- Xu, C., Qu, Y., Xiang, Y. & Gao, L. Asynchronous federated learning on heterogeneous devices: A survey. [arXiv:2109.04269](https://arxiv.org/abs/2109.04269) (2021).
- Cui, L. *et al.* Security and privacy-enhanced federated learning for anomaly detection in iot infrastructures. *IEEE Trans. Ind. Inform.* **18**, 3492–3500 (2021).
- Wan, Y., Qu, Y., Gao, L. & Xiang, Y. Privacy-preserving blockchain-enabled federated learning for b5g-driven edge computing. *Comput. Netw.* **204**, 108671 (2021).
- Kong, X., Gao, H., Shen, G., Duan, G. & Das, S. K. Fedvcp: A federated-learning-based cooperative positioning scheme for social internet of vehicles. *IEEE Trans. Comput. Soc. Syst.* **9**, 197–206. <https://doi.org/10.1109/TCSS.2021.3062053> (2022).
- Chai, H., Leng, S., Chen, Y. & Zhang, K. A hierarchical blockchain-enabled federated learning algorithm for knowledge sharing in internet of vehicles. *IEEE Trans. Intell. Transp. Syst.* **22**, 3975–3986. <https://doi.org/10.1109/TITS.2020.3002712> (2021).
- Zhou, X., Liang, W., She, J., Yan, Z. & Wang, K.I.-K. Two-layer federated learning with heterogeneous model aggregation for 6g supported internet of vehicles. *IEEE Trans. Veh. Technol.* **70**, 5308–5317 (2021).
- Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **12**, 171–178 (1985).
- Azzalini, A. & DallaValle, A. The multivariate skew-normal distribution. *Biometrika* **83**, 715–726 (1996).
- Azzalini, A. The skew-normal distribution and related multivariate families *. *Scand. J. Stat.* **32**, 66 (2005).
- Aryal, G. R. & Nadarajah, S. On the skew Laplace distribution. *J. Inf. Optim. Sci.* **26**, 205–217 (2005).
- Arnold, B. C. & Beaver, R. J. The skew-Cauchy distribution. *Stat. Probab. Lett.* **49**, 285–290 (2000).
- Nadarajah, S. The skew logistic distribution. *Adv. Stat. Anal.* **93**, 187–203 (2009).
- Sahu, A. K. *et al.* Federated optimization in heterogeneous networks. [arXiv:Learning](https://arxiv.org/abs/1806.07575) (2018).
- Li, T., Sahu, A. K., Talwalkar, A. S. & Smith, V. Federated learning: Challenges, methods, and future directions. *IEEE Signal Process. Mag.* **37**, 50–60 (2019).
- Zhu, Z., Hong, J. & Zhou, J. Data-free knowledge distillation for heterogeneous federated learning. *Proc. Mach. Learn. Res.* **139**, 12878–12889 (2021).
- Guo, K., Chen, Z., Yang, H. H. & Quek, T. Q. S. Dynamic scheduling for heterogeneous federated learning in private 5g edge networks. *IEEE J. Sel. Top. Signal Process.* **16**, 26–40 (2022).
- Wang, S. *et al.* Device sampling for heterogeneous federated learning: Theory, algorithms, and implementation. In *IEEE INFOCOM 2021—IEEE Conference on Computer Communications* 1–10 (2021).
- Shen, G. *et al.* Fast heterogeneous federated learning with hybrid client selection. [arXiv:2208.05135](https://arxiv.org/abs/2208.05135) (2022).
- Sun, C., Jiang, T., Zonouz, S. A. & Pompili, D. Fed2kd: Heterogeneous federated learning for pandemic risk assessment via two-way knowledge distillation. In *2022 17th Wireless On-Demand Network Systems and Services Conference (WONS)* 1–8 (2022).
- Zeng, S. *et al.* Hfedm3s: Heterogeneous federated learning with memorable data semantics in industrial metaverse. [arXiv:2211.03300](https://arxiv.org/abs/2211.03300) (2022).
- Alam, S., Liu, L., Yan, M. & Zhang, M. Fedrolex: Model-heterogeneous federated learning with rolling sub-model extraction. [arXiv:2212.01548](https://arxiv.org/abs/2212.01548) (2022).
- Xu, G., Kong, D., Chen, X. & Liu, X. Lazy aggregation for heterogeneous federated learning. *Appl. Sci.* **6**, 66 (2022).
- Tian, J., Smith, J. & Kira, Z. Fedfor: Stateless heterogeneous federated learning with first-order regularization. [arXiv:2209.10537](https://arxiv.org/abs/2209.10537) (2022).
- Arellano-Valle, R. B. & Genton, M. G. On fundamental skew distributions. *J. Multivar. Anal.* **96**, 93–116 (2005).
- Arellano-Valle, R. B. & Genton, M. G. On fundamental skew distributions. *J. Multivar. Anal.* **96**, 93–116 (2005).
- Zhang, X., Chen, X., Hong, M.-F., Wu, Z. S. & Yi, J. Understanding clipping for federated learning: Convergence and client-level differential privacy. [arXiv:2106.13673](https://arxiv.org/abs/2106.13673) (2022).
- Zhao, J. *et al.* Reviewing and improving the Gaussian mechanism for differential privacy. [arXiv:1911.12060](https://arxiv.org/abs/1911.12060) (2019).
- Wu, W. Differentially private density estimation with skew-normal mixtures model. *Sci. Rep.* <https://doi.org/10.1038/s41598-021-90276-6> (2020).

Acknowledgements

This research was funded by the Natural Science Foundation of Jilin Province (Grant Number YDZJ202201ZYTS666), the Education Science Planning Foundation of Jilin (Grant Number GH21317), the Education Department of Jilin Province industrialization cultivation project (Grant Number JJKH20230020CY), Scientific planning project of Education Department of Jilin Province (Grant Number JJKH20230021KJ) and the Privacy Computing Joint Laboratory.

Author contributions

X.Y. and W.W. contributed equally to this work. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to W.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023