IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 March 2024, accepted 23 April 2024, date of publication 2 May 2024, date of current version 10 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3396146

RESEARCH ARTICLE

Toward the Application of Differential Privacy to Data Collaboration

HIROMI YAMASHIRO[1], KAZUMASA OMOTE[2], AKIRA IMAKURA[2], AND TETSUYA SAKURAI[2]

[1]Graduate School of Science and Technology, University of Tsukuba, Tsukuba 305-8577, Japan
[2]Institute of Systems and Information Engineering, University of Tsukuba, Tsukuba 305-8577, Japan

Corresponding author: Hiromi Yamashiro (s2220544@u.tsukuba.ac.jp)

This work was supported in part by the New Energy and Industrial Technology Development Organization (NEDO), and in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP22K19768.

ABSTRACT Federated Learning, a model-sharing method, and Data Collaboration, a non-model-sharing method, are recognized as data analysis methods for distributed data. In Federated Learning, clients send only the parameters of a machine learning model to the central server. In Data Collaboration, clients send data that has undergone irreversibly transformed through dimensionality reduction to the central server. Both methods are designed with privacy concerns, but privacy is not guaranteed. Differential Privacy, a theoretical and quantitative privacy criterion, has been applied to Federated Learning to achieve rigorous privacy preservation. In this paper, we introduce a novel method using PCA (Principal Component Analysis) that finds low-rank approximation of a matrix preserving the variance, aiming to apply Differential Privacy to Data Collaboration. Experimental evaluation using the proposed method show that differentially-private Data Collaboration achieves comparable performance to differentially-private Federated Learning.

INDEX TERMS Differential privacy, dimension reduction, distributed machine learning, federated learning, principal component analysis.

I. INTRODUCTION

Privacy-preserving data analysis methods for distributed data are becoming increasingly important. In machine learning and data mining, the size of the dataset is directly related to the quality of the results obtained. However, a single institution cannot always collect a sufficient amount of data. Sharing data among multiple institutions is one way to build large datasets, but the privacy of the data is a problem. Therefore, there is a growing interest in data analysis methods capable of analyzing distributed datasets across multiple institutions or individuals without compromising the confidentiality of personal information.

There are two types of data analysis methods for distributed data: model-sharing and non-model-sharing. A central server transmits a machine learning model to participating clients in model-sharing methods. The clients train the machine learning model with their own data and then send the parameters to the central server. In non-model-sharing methods, the clients send anonymized data to the central server, and then the central server trains the model with the centralized data. Federated Learning [1] is known as a model-sharing method. The central server sends a machine learning model to each client in Federated Learning. The clients train the model with their own data and send back updated parameters to the central server. Data collaboration [2], [3], [4] is known as a non-model-sharing method. In Data Collaboration, the clients anonymize their own data by dimension reduction and then send it to the central server. The central server transforms the data sent by the clients into a representation that can be handled as a single dataset and then trains a machine learning model using the integrated data.

These methods are designed to preserve privacy, but they do not rigorously preserve privacy. Both of these two methods employ central server architecture, which potentially introduce privacy risks if the server is compromised. They use different techniques to allow the clients to avoid the need to fully trust the server. Federated Learning attempts to protect privacy by the clients sending only the parameters of the model instead of data. Data Collaboration attempts

The associate editor coordinating the review of this manuscript and approving it for publication was Sedat Akleylek.

to protect privacy by the clients sending data irreversibly transformed by dimension reduction. However, a model inversion attack for Federated Learning in which an attacker infers about the training data by analyzing the parameters has been reported [5]. Data Collaboration utilizes dimension reduction as a means of privacy protection. While evaluations of primary component analysis as a privacy-preserving method have been reported [6], [7], these researches only experimentally show that an attack accuracy decreases by dimension reduction, lacking theoretical privacy analysis.

Differential privacy (DP) [8] is a standard method of protecting privacy in data analysis. DP makes it difficult to infer original data by adding artificial noise with appropriately tuned parameters and randomizing an algorithm's output. DP can also achieve greater and more rigorous privacy protection for data analysis on distributed data. An application of DP to Federated Learning has already been proposed [9]. However, there is no research on an application of DP to Data Collaboration.

Data Collaboration exploits dimension reduction to preserve privacy. Intuitively, inferring personal information contained in the original data from the dimension reduced data is hard. The privacy of Data Collaboration is based on this intuition. However, dimension reduction does not inherently provide privacy protection, and no theoretical analysis of the privacy or additional privacy-preserving techniques have been studied. DP brings rigorous and quantitative privacy to data collaboration.

In this paper, we propose the first application of DP to Data Collaboration as a privacy-preserving and non-model-sharing data analysis method for distributed data. We apply DP to Data Collaboration by using a dimension reduction algorithm that satisfies DP when a client performs dimension reduction on the data. Especially, we used PCA (Principal Component Analysis), which finds low-rank approximation of the data matrix preserving the variance, for dimension reduction method of Data Collaboration. In general, when DP is applied, the utility of the results obtained is lower than that of non-private cases. In Data Collaboration, we expect that applying DP will reduce the utility of the results, but the degree of the decrease is unclear. Therefore, we experimentally evaluated the impact of DP on the utility of Data Collaboration and compared the degree of the impact with that of using DP for Federated Learning. The results showed that the decrease in the utility of Data Collaboration due to DP was about the same as in the case of Federated Learning.

II. PRELIMINARIES

A. DATA COLLABORATION

Data Collaboration is a non-model-sharing type data analysis method for distributed data. Distributed data are not shared among clients but are aggregated and processed on a central server to analyze distributed data while preserving privacy.

Multiple clients with datasets and a central server that trains machine learning models using the aggregated datasets perform Data Collaboration. Each client $i(i = 1, \ldots, n)$ has its own data $\boldsymbol{X}_i$ and auxiliary data $\boldsymbol{X}^{anc}$ common to all clients. Clients transform data by arbitrary mapping function f_i, and then send the mapped data $\widetilde{\boldsymbol{X}}_i = f_i(\boldsymbol{X}_i)$, $\widetilde{\boldsymbol{X}}_i^{anc} = f_i(\boldsymbol{X}^{anc})$ to the central server. The mapping functions are row-by-row transformations defined arbitrarily by each client. In Data Collaboration, mapping functions are typically dimension reduction. Mapped and aggregated data cannot be integrated and used to train a machine learning model because the relation of records in the original data is not preserved in the mapped representation $\widetilde{\boldsymbol{X}}_i$. The central server generates a mapping g_i that maps $\widetilde{\boldsymbol{X}}_i$ into a space that reproduces the relation in the original data, exploiting that the auxiliary data mapped by each client, $\widetilde{\boldsymbol{X}}_i^{anc}$, are originally the same records. That is, the central server constructs g_i such that

$$g_i(\widetilde{\boldsymbol{X}}_i^{anc}) \simeq g_j(\widetilde{\boldsymbol{X}}_j^{anc})$$

for any i, j by solving an optimization problem. The data $\hat{\boldsymbol{X}}_i = g_i(\widetilde{\boldsymbol{X}}_i)$ have the same properties as the original data $\boldsymbol{X}$. The central server integrates $\hat{\boldsymbol{X}}_i$ from each client and trains a machine learning model using the integrated dataset as a single dataset.

Since the original auxiliary data $\boldsymbol{X}^{anc}$ are not shared with the central server, it does not know the input and output for f_i simultaneously. Therefore, f_i is said not to be inferable. In addition, since the mapping functions f_i are set arbitrarily by each client and are not shared by the central server, it is said that the original data $\boldsymbol{X}_i$ cannot be reconstructed from the mapped data $\widetilde{\boldsymbol{X}}_i$. However, the privacy of the mapped data has not been theoretically analyzed, and personal information may be statistically inferred. Therefore, in this paper, we have applied DP to Data Collaboration to guarantee its privacy.

B. DIFFERENTIAL PRIVACY

Differential Privacy [8] is a quantitative measure of the degree of privacy protection for randomized algorithms. In less rigorous terms, an algorithm satisfying DP guarantees theoretically that if only one sample in the dataset differs, an attacker can infer little change from the algorithm's output. DP is the most widely used privacy criterion. For example, the U.S. Census Bureau adopted DP as "the gold standard for privacy protection in computer science and cryptography" [10] when publishing the results of the 2020 census.

The formal definition of (ϵ, δ)-DP is as follows. $\Pr[\cdot]$ is propability. ϵ is a privacy budget, set by a user of a randomized algorithm. Smaller ϵ means greater noise. In (ϵ, δ)-DP, the constraint is allowed to break with probability δ. When $\delta = 0$, it is just denoted ϵ-DP.

Definition: A randomized algorithm A satisfies (ϵ, δ)-DP if the following inequalitiy holds for $\epsilon \in (0, \infty]$, $\delta \in [0, 1)$, any adjacent datasets D, D', and any subspace O of the range of A:

$$\Pr[A(D) \in O] \leq e^{\epsilon} \Pr[A(D') \in O] + \delta.$$

An adjacent dataset is a dataset that differs in only one record. The definition of an adjacent dataset varies depending

on which structures in the dataset are considered unitary records. The definition of adjacency concerns the granularity of privacy protected by DP.

Privacy mechanisms or, simply, mechanisms are methods to randomize an output of algorithms by adding noise at some stage of the algorithms. Several general-purpose mechanisms that apply DP to any algorithm under certain conditions and mechanisms tailored for specific algorithms have been proposed. Because privacy mechanisms perturb data or algorithms, DP usually decreases the utility of the data or algorithms. There is a trade-off between the utility and the privacy in DP. In other words, the stronger the privacy protection, the greater the utility loss.

Privacy mechanisms are classified according to which stage the original algorithm is disturbed. Mechanisms classified as output perturbation add noise to a non-disturbed output of an algorithm. Laplace and Gaussian mechanisms [11] are known as the output perturbation mechanisms. These mechanisms achieve DP by adding i.i.d. noise drawn from Laplace distributions or Gaussian distributions to an output of any function that outputs a real vector. The noise scale depends on the L1-sensitivity of a function for Laplace mechanism and the L2-sensitivity for Gaussian mechanism. For adjacent data D, D', the Ln-sensitivity Δ is defined as follows:

$$\Delta = \sup_{D \simeq D'} \|A(D) - A(D')\|_n.$$

In our experiments, we used Analytic Gaussian mechanism [12], a variation of Gaussian mechanism that yields a smaller variance. Let $Z \sim \mathcal{N}(0, \sigma^2 I)$ ($\mathcal{N}$ is Gaussian distribution). Analytic Gaussian mechanism exploits that $f(x)+Z$ satisfies (ϵ, δ)-DP when the following inequality (1) holds. Φ is the cumulative distribution function of the standard Gaussian distribution.

$$\Phi\left(\frac{\Delta}{2\sigma} - \frac{\epsilon\sigma}{\Delta}\right) - e^{\epsilon}\Phi\left(-\frac{\Delta}{2\sigma} - \frac{\epsilon\sigma}{\Delta}\right) \leq \delta. \quad (1)$$

The smallest σ that satisfies this constraint can be found quickly by numerical computation using binary search.

DP applications to PCA have been extensively studied [13], [14], [15]. To the best of our knowledge, they compute a differential-private projection matrix of PCA, not a low-rank approximation of the data. These methods assume the projector will be public and the projection will be kept private. For example, on recommendation systems, a recommendation made to a user is based on the global covariance information and the user's own information, and there is no need to hide their own information from the user [14]. We also propose a DP application to PCA, but it publishes the projection or dimension-reduced data differential-privately.

III. RELATED WORKS

A. LOCAL DIFFERENTIAL PRIVACY

Local differential privacy (LDP) [16] is an application of DP for when data aggregators are not trusted. The formal definition is as follows (this notation is from [17]).

Definition: A randomized algorithm A satisfies ϵ-LDP if the following inequalitiy holds for $\epsilon \in (0, \infty]$, any two data records X, Y, and any output $\tilde{X}$ of the range of A:

$$\Pr[A(X) = \tilde{X}] \leq e^{\epsilon}\Pr[A(Y) = \tilde{X}].$$

In DP, data owners trust data aggregators, but in LDP, they do not trust the aggregators and perturb their own data records on the client side. Some LDP based methods employ dimension reduction [17], [18]. These methods aggregate locally perturbed data records, estimate the original distribution of the data, and then synthesize artificial data drawn from the estimated distribution. Due to the curse of dimensionality, it is difficult to estimate the distribution of high-dimensional data. These methods use dimension reduction to estimate the distribution, not to preserve privacy. Data Collaboration differs from these methods in this respect.

Our proposed method can be seen as a type of LDP but not a typical one. LDP is client-side and per-record DP, but in Data Collaboration, clients are also data aggregators. When the data aggregators cooperatively train a machine learning model, they perturb their aggregated data per record. We assume the data owners trust the aggregators and give them original data.

B. PRIVACY OF DATA COLLABORATION

When we consider privacy violations by a malicious central server, the privacy of Data Collaboration depends on the privacy of dimension-reduced data. Studies on PCA to add noise to dimension-reduced data to increase privacy have been reported. In the study by Chen et al. [7], an attacker attempts to reconstruct the original data by performing the inverse transformation on the dimension-reduced data. That is, for original data X, dimension-reduced data $\widetilde{X}$, transformation matrix of PCA W, and reconstructed data X_r, dimension reduce of PCA is performed as follows

$$\widetilde{X} = W \times X.$$

The attacker attempts to reconstruct the original data by

$$X_r = \widetilde{X} \times W^{\top}.$$

However, they only re-transform the dimension-reduced data from a latent space to the original space and do not restore the information lost through the dimension reduction. Chen et al. experimentally evaluated the privacy increase for such attackers by adding noise to the dimension-reduced data. A success rate of the Re-identification attack [19] was used for the privacy evaluation. This attack links corresponding records $x_i \in X$ and $\widetilde{x_i} \in \widetilde{X}$ for the original data X and anonymized data $\widetilde{X}$. In the experiment by Chen et al., they performed the Re-identification attack on the original data and the restored dimension-reduced data. The results showed that even if the cumulative ratio of the variation of PCA was 99%, i.e., 99% of the information was preserved before and

after the dimension reduction, the attack success rate was still less than 70%. This result indicated that the dimension reduction was robust against the Re-identification attack. While the attack success rate decreased as the cumulative ratio of the variation decreased, the performance of machine learning models trained on the dimension-reduced data also decreased, confirming that there was a trade-off between the utility of data and privacy.

Following the results of Yamashiro et al. evaluated the privacy of PCA with a more realistic attacker [6]. In the attack model of Chen et al., an attacker uses a transformation matrix of PCA that is actually used to dimension-reduce to reconstruct original data. However, in practice, this matrix is not always publicly available and accessible to attackers. Therefore, Yamashiro et al. proposed an attack model in which the attacker estimates W using auxiliary information and experimentally evaluated the privacy of PCA using this model. In this model, the attacker obtains data drawn from the same distribution as the original data and uses these to estimate W. Experimental results with this model confirmed that the success rate of the Re-identification attack significantly decreased when the attacker used the estimated W.

In the two studies above, the dimension-reduced data is mapped to the original space to reconstruct the original data, but information lost through the dimension reduction is not reconstructed. In the study by Yamazoe et al. [20], they proposed a method to recover the original data from the latent space using GAN. In GAN, a generator $G(z)$ generates false data from the input, and a discriminator D classifies whether the given data is false data produced by the generator or true data, which are trained simultaneously. If the training is successful, the generator can estimate the generative distribution of the data with high accuracy. In the proposed method, an attacker has external data X_{ex} as auxiliary information. These external data are not necessarily the same in distribution or domain as the original data and, thus, are not necessarily labeled in the same way as the original data. Therefore, the attacker first trains a classification model with dimension-reduced data $\widetilde{X}_i$ aggregated from each client and then labels the external data by classifying X_{ex} with the model. The classification model is trained by integrating X_{ex} and $\widetilde{X}_i$ into a representation that can be handled as a single data set through the procedure of Data Collaboration. GAN is trained by taking the labeled external data as true data. The generator takes the dimension-reduced data $\widetilde{X}_i$ as input. In the experiment, the original data was MNIST, a handwritten digit data set. As external data, ''digits,'' handwritten digit data sets like MNIST, ''letters,'' handwritten alphabet data sets, and ''balanced'' data sets with a mixture of ''digits'' and ''letters'' were used. The experimental results showed that when ''digits,'' which were in the same domain as the original data, were used as external data, the reconstructed data was successfully restored to a level where the digits could be visually distinguished. When ''letters,'' which were in a different domain from the original data, were used as external data, the restored data was difficult to identify visually. In the case of ''balanced,'' the results were intermediate between ''digits'' and ''letters.'' It was confirmed that there is a privacy risk that the original data is recovered from the dimension-reduced data. The risk varies depending on whether an attacker can estimate the domain of the data and collect data from that domain.

C. FEDERATED LEARNING

Federated Learning is a model-sharing type data analysis method for distributed data. A central server transmits a machine learning model to clients. The clients train the model with local data and send back only the model parameters to the central server.

There are several variations of Federated Learning. A method called federated average trains the model by repeating the following steps.

1) The central server transmits a global training model to clients.
2) The clients train the global model with their own local data.
3) After a specified number of training epochs, the clients send the updated parameters to the central server.
4) The central server takes a weighted average of the clients' parameters, updates the global model with the averaged parameters, and transmits the new global model to the clients.

The central server need not always transmit the global model to all clients. Instead, it may only transmit the model to randomly selected clients.

Privacy threats in Federated Learning and the application of DP as a countermeasure are known. Federated Learning preserves privacy by allowing clients to transmit only the parameters of machine learning models. However, model inversion attack [21] reconstructs the data used for training from a trained machine learning model. In Federated Learning, the training data can be inferred from the global model transmitted to the clients even if the central server is honest. Therefore, applications of DP to Federated Learning have been proposed. In the method by Wei et al. [9], used in our experiments, the clients train the global model with their own data and then clip the updated parameters so that the norm is below a threshold. The clients add noise to the clipped parameters and send the noised parameters to the central server. The central server takes the average of the parameters sent by the clients and adds noise.

D. COMPARING DATA ANALYSIS METHODS FOR DISTRIBUTED DATA

Federated Learning, a model-sharing distributed data analysis method, and Data Collaboration, a non-model-sharing distributed data analysis method, have different characteristics in practical use. Figs. 1 and 2 show the overview of Data

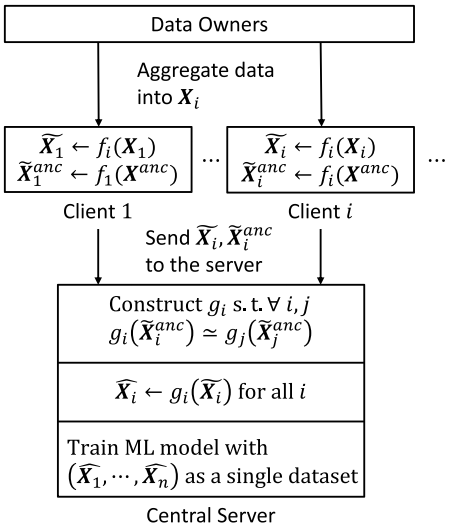

FIGURE 1. Overview of data collaboration.

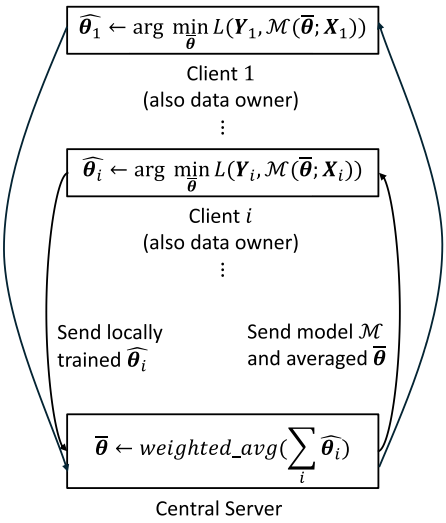

FIGURE 2. Overview of federated learning.

Collaboration and Federated Learning for comparison of these two methods.

Regarding communication cost, Federated Learning sends and receives machine learning models between a server and clients at each learning iteration. Hence, its communication cost depends on the size of the machine learning model, the number of clients, and the number of iterations. In Data Collaboration, communication is limited to aggregating anonymized data to a central server. Its communication cost depends on the data size. Although their communication cost depends on different factors, typically, Data Collaboration requires less time and network resources.

Regarding computational resources, Federated Learning requires that all clients with data have enough computational resources to train a machine learning model. In Data Collaboration, clients with data must have enough computational resources to perform dimension reduction, and resources capable of training machine learning models are required only at the central server.

Data Collaboration tends to require fewer resources than Federated learning. Also, Data Collaboration performs better than Federated Learning when the number of clients is small [22]. Therefore, Data Collaboration is lightweight and suitable for relatively small cases. Testing the effectiveness of distributed learning across multiple institutions with Data Collaboration easy to introduce is a possible scenario.

It is also noted that both Federated Learning and Data Collaboration have low explainability for machine learning models obtained. The explainability of the models is important to interpret the relation between the inputs and outputs of the models. It allows us to prevent the models from learning bias or prejudice in real-world. SHAP [23] is a widely used model-agnostic explainability method. SHAP calculates the importance of each attribute on a single record by averaging differences in a machine learning model's outputs caused by adding an attribute to all possible combinations of the other attributes. By calculating this record-wise importance for all records, SHAP obtains the importance of each attribute on the model. Because SHAP is relative to a dataset used to compute the attributes' importance, it may not calculate proper importance when one dataset holder performs it on distributed datasets. A method that extends Data Collaboration to obtain highly explainable machine learning models using distributed data has been proposed [24]. This method utilizes the auxiliary data X^{anc} of Data Collaboration that is common to all clients.

It is difficult to describe the accuracy and security of trained machine learning models based on the characteristics of the two data analysis methods because they depend highly on use cases. For example, in typical use cases, Federated Learning involves thousands or tens of thousands of unspecified clients. In contrast, in Data Collaboration, learning occurs with relatively few institutions. At this point, which actors are trusted and which are not depends on the use case and acceptable security risks.

IV. METHOD

A. APPLYING DIFFERENTIAL PRIVACY TO DATA COLLABORATION

We propose a new method for applying DP to Data Collaboration. There are two possible cases where privacy can be violated in Data Collaboration.

1) Malicious central serversinfer original data from data anonymized by dimension reduction.
2) Clients infer data held by another client from a trained machine learning model.

To prevent (1), the clients must share the data mapped by a differentially-private method with the central server. Applying DP at this stage will also prevent (2). Therefore, as an application of DP to Data Collaboration, we consider the sharing of mapped data by clients in a manner that satisfies

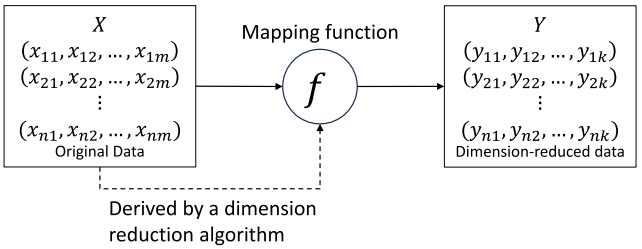

FIGURE 3. Deriving a dimension-reduction function.

DP. If only (2) is a problem, after aggregating data through Data Collaboration, a differentially-private training method such as DP-SGD [25] can be used.

The specific method of applying DP to Data Collaboration depends on a dimension reduction algorithm a client uses. Regardless of the dimensionality reduction algorithm, the following two points should be considered.

1) It is dimension-reduced data itself that is shared. To the best of our knowledge, most differentially-private dimension-reduction methods compute a function that maps original data to a latent space. For example, differentially-private PCA has been widely studied. However, they privately compute a transformation matrix that projects the original data to the principal components and does not guarantee DP for the dimension-reduced data itself.
2) Privacy mechanisms cannot simply be applied to a function that maps original data to a latent space. As shown in Fig. 3, the function that performs the dimension reduction is derived from the original data and receives the original data as input. Therefore, the function's sensitivity depends on the original data, but privacy mechanisms require a global sensitivity that is independent of the data.

B. METHOD

An overview of the proposed method for PCA, which differential-privately publishes dimension-reduced values, is shown in Fig. 4. The dataset is represented as a data matrix $X \in \mathbb{R}^{n \times m}$ where row vectors represent the records.

1) Normalizing each component of the row vectors of the data matrix to range L. That is, normalizing to $[L_1, L_2](L_2 - L_1 = L)$.
2) Applying the regular PCA procedure to the normalized X to obtain the dimension-reduced data $Y \in \mathbb{R}^{n \times k}$.
3) By Gaussian mechanism, adding noise to the dimension-reduced data. Let the L2-sensitivity of the function be L.

In (3), letting the mapping matrix obtained by PCA be W, the privacy mechanism is applied to the function $f : \mathbb{R}^m \to \mathbb{R}^k$ such that $f(\boldsymbol{x}) = W\boldsymbol{x}$.

It is sufficient for the privacy analysis of our proposed method to show that the sensitivity of $f(\boldsymbol{x}) = W\boldsymbol{x}$ is less than or equal to L since we apply DP to PCA using the existing

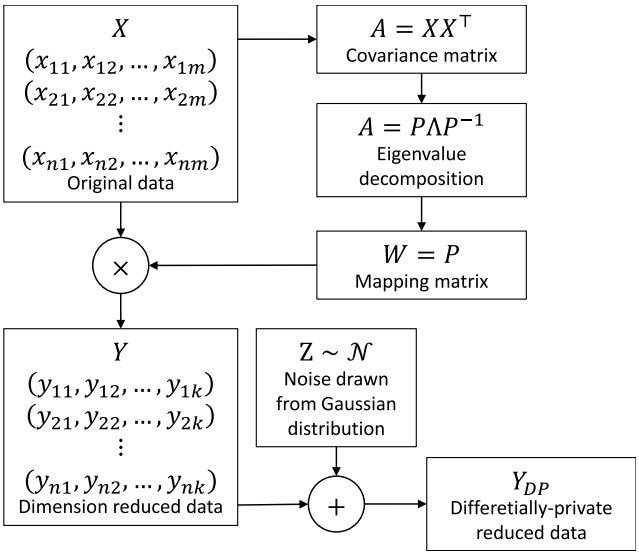

FIGURE 4. Applying DP to data collaboration using PCA.

privacy mechanism. A spectral norm for matrixes is defined as follows:

$$\|W\|_2 = \sup_{x \neq 0}\left(\frac{\|Wx\|_2}{\|x\|_2}\right).$$

Since $\boldsymbol{x}$ is normalized to the range L, for adjacent, that is, different only in one component $\boldsymbol{x}, \hat{\boldsymbol{x}}$, the max of $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|$ is L. Therefore, since the following inequality holds for the L2-sensitivity $\Delta = \sup_{\boldsymbol{x} \sim \hat{\boldsymbol{x}}} \|f(\boldsymbol{x}) - f(\hat{\boldsymbol{x}})\|_2$, we have $\Delta \leq L\|W\|_2$.

$$\begin{aligned}\|f(\boldsymbol{x}) - f(\hat{\boldsymbol{x}})\|_2 &= \|W\boldsymbol{x} - W\hat{\boldsymbol{x}}\|_2 \\ &= \|W(\boldsymbol{x} - \hat{\boldsymbol{x}})\|_2 \\ &\leq \|W\|_2\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 \\ &\leq L\|W\|_2.\end{aligned}$$

Here, W is an orthogonal matrix, and it is known that the spectral norm of an orthogonal matrix is 1, so we have $\Delta \leq L\|W\|_2 = L$. W is an orthogonal matrix because W is a matrix that consists of eigenvectors of a covariance matrix of X, and eigenvectors of a real symmetric matrix like a covariance matrix are orthogonal.

V. EXPERIMENTS

A. SETTINGS

We compared the performance of classification models trained by Data Collaboration and Federated Learning. We used MNIST [26], a handwritten digit dataset and Fashon-MNIST [27], a clothing image dataset for evaluation. Each sample of MNIST and Fashion-MNIST is a grayscale 28×28 image. We used these two datasets to evaluate the proposed method on tasks with different difficulty; Fashion-MNIST is compatible with MNIST but harder to classificate than MNIST. For both Data Collaboration and Federated Learning, the number of participating clients was 10, and the

number of overall training data was 1,000 and 10,000. In other words, the number of training data per client is either 100 or 1,000. For testing, we used 10,000 data sets that differed from the training data.

A neural network consisting of fully connected and activation layers was used for classification. This network takes flattened vectors of records as input, while original records have a two-dimensional structure. Specifically, the network takes a 784-dimensional vector as an input and outputs a 10-dimensional vector through hidden layers of 512 and 128 nodes. We used ReLU for activation layers.

For Data Collaboration, we varied the following conditions.

- Number of post-reduce dimensions: Varying from 10 to 100 by 10.
- Differential Privacy: We used the regular non-private and our proposed differentially-private methods.

In the experiments, clients on Data Collaboration used PCA as a dimension reduction algorithm with the common parameters among the clients. For Federated Learning, we used a regular non-private method (fed-avg) [1] and differentially-private method [9].

In any settings, we used $\epsilon = 50, \delta = 0.01$ when DP was involved. $\epsilon = 50$ is larger than the typical magnitude of privacy budgets. We took this epsilon from the experiments in literature of the differentially-private Federated Learning by Wei et al. In the literature, the authors evaluated the method they proposed using multiple epsilons, and $\epsilon = 50$ was the minimum one. In addition, when we used smaller epsilon, say $\epsilon = 2$, the differentially-private Federated Learning could not be trained properly. We have assumed that the larger privacy budget than the typical one is acceptable for the experimental evaluations and comparison.

B. RESULTS

Figs. 5 to 8 show the accuracy of the classification models. The y-axis of the graphs is the accuracy of the classification model (%), and the x-axis is the number of dimensions after dimension reduction in Data Collaboration. Figs. 5 and 7 shows the results for Data Collaboration and Federated Learning, where each client has 100 records each, for a total of 1,000 data items. Figs. 6 and 8 shows the results where each client has 1,000 records each, for a total of 10,000 records. In the legend of the graphs, `DC` and `DC-DP` denote the cases trained with regular Data Collaboration and one with DP, respectively. Similarly, `FL` and `FL-DP` denote the case trained by regular Federated Learning and one with DP, respectively. `individual` denotes the case where the classification model is trained directly on 100 or 1,000 records without using distributed data. This case corresponds to the case where the client trained the model individually and served as a baseline for evaluating the results using distributed data analysis. `centralized` denotes the case where the model was trained by aggregating all the institutions' data in one place. Since no dimensionality reduction is involved in `FL`, `FL-DP`, `individual`, `centralized`, they are straight lines parallel to the x-axis.

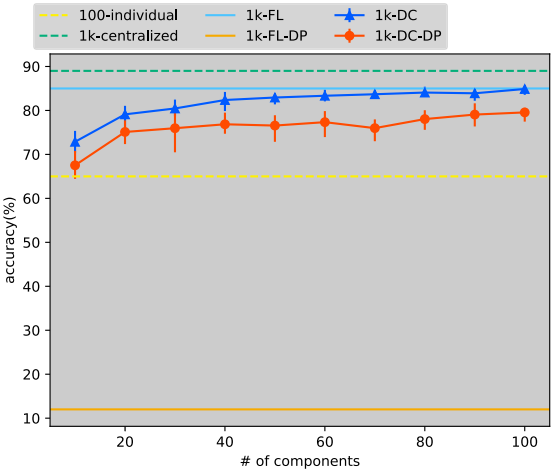

FIGURE 5. The classification accuracies when # of data is 100 per client on MNIST.

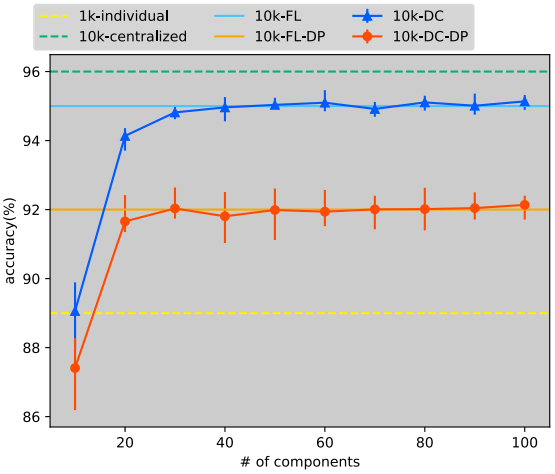

FIGURE 6. The classification accuracies when # of data is 1,000 per client on MNIST.

The experimental results are shown as tables in Tables 1 to 6. Tables 1, 2, 4 and 5 show the accuracy of the classification models trained on Data Collaboration. In Tables 1, 2, 4 and 5, Non-DP and DP denote the case without and with DP, respectively. Tables 3 and 6 show the accuracy of the model trained by Federated Learning and a single institution. In Tables 3 and 6, FL(Non-DP) and FL(DP) denote the case when DP is not applied in Federated Learning and when DP is applied, respectively. Individual denotes the results for a single institution.

Compared to the baseline case `individual`, both Data Collaboration and federated learning, regardless of DP, outperform the accuracy in almost all conditions. Thus, these distributed data analysis methods are more useful than training a machine learning model on a single organization, even with the privacy preservation of DP. The loss of the accuracy due to DP is only a few percentage points.

TABLE 1. The classification accuracies of data collaboration when # of data is 100 per client on MNIST (%).

	# of components									
	10	20	30	40	50	60	70	80	90	100
Non-DP	72.87	79.11	80.45	82.37	82.94	83.34	83.68	84.07	83.9	84.88
DP	67.52	75.11	75.96	76.85	76.56	77.33	75.98	78.04	79.06	79.55

TABLE 2. The classification accuracies of data collaboration when # of data is 1,000 per client on MNIST (%).

	# of components									
	10	20	30	40	50	60	70	80	90	100
Non-DP	89.06	94.14	94.82	94.96	95.04	95.1	94.92	95.11	95.01	95.14
DP	87.41	91.66	92.03	91.81	91.99	91.94	92.01	92.01	92.04	92.14

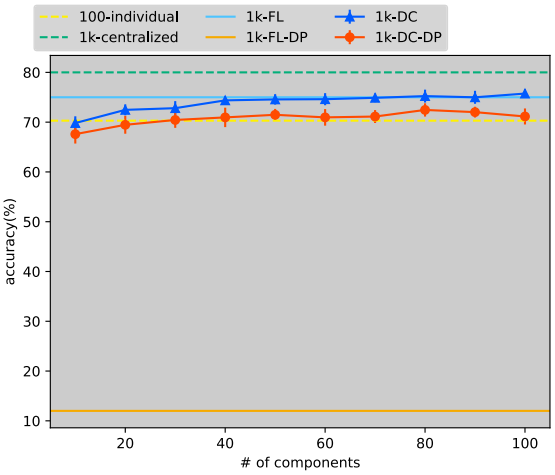

FIGURE 7. The classification accuracies when # of data is 100 per client on Fashion-MNIST.

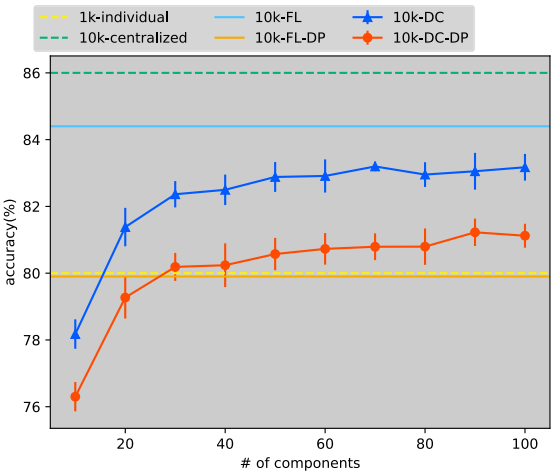

FIGURE 8. The classification accuracies when # of data is 1,000 per client on Fashion-MNIST.

Comparing the results of Data Collaboration with those of Federated Learning, the accuracy is equal to or slightly worse than the corresponding Federated Learning with respect to DP, provided that the number of dimensions after dimension reduction is sufficient. As a differentially-private data analysis method for distributed data, Data Collaboration can achieve the same level of utility as Federated Learning. When the number of data per client is 100 (Figs. 5 and 7), the accuracy of Federated Learning with DP is significantly lower, but the accuracy did not change when the number of local and global epochs was adjusted.

TABLE 3. The classification accuracies of federated learning and a single institution on MNIST (%).

# of data per client	100	1,000
FL (Non-DP)	85.43	95.1
FL (DP)	11.97	89.3
individual	65.81	92.5
centralized	89.0	96.1

VI. DISCUSSION

This work has two major limitations. First, we don't adequately consider the effect of dimension reduction on Data Collaboration. On the privacy aspect, it is obvious intuition that a smaller number of dimensions provides greater privacy protection. On the performance aspect, we can see that the number of dimensions significantly affects the performance of Data Collaboration in the experimental results. The size of latent space is an important parameter of Data Collaboration that greatly affects privacy and performance. However, the proposed method doesn't take into consideration the effect of dimension reduction.

Second, this work lacks the implications for real-world applications of Data Collaboration. This work does not give a criterion for choosing the parameters in Data Collaboration, i.e., how small a privacy budget and a number of dimensions should be or how much privacy risk Data Collaboration has.

Thus, two types of studies about data collaboration are needed. The first is a more sophisticated method to apply DP to Data Collaboration. Our approach is straightforward and can be more efficient regarding noise magnitude. Also, as described above, a method that takes the effect of

TABLE 4. The classification accuracies of data collaboration when # of data is 100 per client on Fashion-MNIST (%).

	# of components									
	10	20	30	40	50	60	70	80	90	100
Non-DP	69.82	72.47	72.81	74.38	74.57	74.61	74.89	75.24	74.99	75.74
DP	78.18	81.38	82.36	82.5	82.88	82.91	83.2	82.95	83.05	83.17

TABLE 5. The classification accuracies of data collaboration when # of data is 1,000 per client on Fashion-MNIST (%).

	# of components									
	10	20	30	40	50	60	70	80	90	100
Non-DP	78.18	81.38	82.36	82.5	82.88	82.91	83.2	82.95	83.05	83.17
DP	76.3	79.27	80.19	80.24	80.57	80.73	80.79	80.8	81.22	81.12

TABLE 6. The classification accuracies of federated learning and a single institution on Fashion-MNIST (%).

# of data per client	100	1,000
FL (Non-DP)	74.7	84.4
FL (DP)	12.1	79.9
individual	70.3	80
centralized	80	85.9

dimension reduction on Data Collaboration into account and handles privacy budgets of DP and the size of latent space in a unified manner may reduce the magnitude of noise. The second is practical privacy analysis for Data Collaboration. The need for privacy protection and the appropriate degree of protection are based on actual use cases, attacker models, and privacy risks derived from these two. DP provides Data Collaboration with theoretical and quantitative privacy protection against attackers DP assumes, but it is unclear what kind of and how much risk Data Collaboration faces.

VII. CONCLUSION

In this paper, we have proposed a method of differentially-private Data Collaboration, which is one of the data analysis methods for distributed data, and experimentally evaluated its performance. As a result, we found that the performance of the proposed method is comparable to that of differentially-private Federated Learning, which is also a distributed data analysis method. Thus, we concluded that Data Collaboration is as useful as Federated Learning as a differentially-private data analysis method for distributed data. However, the proposed method is limited to the case where PCA is used as dimension reduction, and the use of other dimension reduction methods is a future work. It is also a future work to consider attack and privacy models in Data Collaboration and to provide the appropriate number of dimensions and privacy budgets from the privacy perspective based on these models.

REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, vol. 54, A. Singh and J. Zhu, Eds., 2017, pp. 1273–1282. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html

[2] A. Imakura and T. Sakurai, "Data collaboration analysis framework using centralization of individual intermediate representations for distributed data sets," *ASCE-ASME J. Risk Uncertainty Eng. Syst., A, Civil Eng.*, vol. 6, no. 2, Jun. 2020, Art. no. 04020018. [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/AJRUA6.0001058

[3] A. Imakura, X. Ye, and T. Sakurai, "Collaborative data analysis: Non-model sharing-type machine learning for distributed data," in *Knowledge Management and Acquisition for Intelligent Systems*, H. Uehara, T. Yamaguchi, and Q. Bai, Eds. Cham, Switzerland: Springer, 2021, pp. 14–29.

[4] A. Imakura, R. Tsunoda, R. Kagawa, K. Yamagata, and T. Sakurai, "DC-COX: Data collaboration cox proportional hazards model for privacy-preserving survival analysis on multiple parties," *J. Biomed. Informat.*, vol. 137, Jan. 2023, Art. no. 104264. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046422002696

[5] L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov, "Exploiting unintended feature leakage in collaborative learning," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2019, pp. 691–706.

[6] H. Yamashiro, K. Omote, A. Imakura, and T. Sakurai, "A study of the privacy perspective on principal component analysis via a realistic attack model," in *Proc. 18th Int. Conf. Comput. Intell. Secur. (CIS)*, Dec. 2022, pp. 376–380.

[7] Z. Chen and K. Omote, "A privacy preserving scheme with dimensionality reduction for distributed machine learning," in *Proc. 16th Asia Joint Conf. Inf. Secur. (AsiaJCIS)*, Aug. 2021, pp. 45–50.

[8] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography*, S. Halevi and T. Rabin, Eds. Berlin, Germany: Springer, 2006, pp. 265–284.

[9] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, and H. V. Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3454–3469, 2020.

[10] U.S. Census Bureau. (2020). *Census Bureau Adopts Cutting Edge Privacy Protections for 2020 Census*. [Online]. Available: https://www.census.gov/newsroom/blogs/random-samplings/2019/02/censusbureauadopts.html

[11] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, Aug. 2013, doi: 10.1561/0400000042.

[12] B. Balle and Y.-X. Wang, "Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising," in *Proc. Int. Conf. Mach. Learn.*, J. Dy and A. Krause, Eds., 2018, pp. 394–403. [Online]. Available: https://proceedings.mlr.press/v80/balle18a.html

[13] W. Jiang, C. Xie, and Z. Zhang, "Wishart mechanism for differentially private principal components analysis," in *Proc. 13th AAAI Conf. Artif. Intell.* Palo Alto, CA, USA: AAAI Press, 2016, p. 1730–1736.

[14] C. Dwork, K. Talwar, A. Thakurta, and L. Zhang, "Analyze gauss: Optimal bounds for privacy-preserving principal component analysis," in *Proc. 46th Annu. ACM Symp. Theory Comput.* New York, NY, USA: Association for Computing Machinery, May 2014, p. 11, doi: 10.1145/2591796.2591883.

[15] F. Shang, Z. Zhang, T. Xu, Y. Liu, and H. Liu, "Principal component analysis in the stochastic differential privacy model," in *Proc. 37th Conf. Uncertainty Artif. Intell.*, vol. 161, C. de Campos and M. H. Maathuis, Eds. Jul. 2021, pp. 1110–1119. [Online]. Available: https://proceedings.mlr.press/v161/shang21a.html

[16] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Proc. IEEE 54th Annu. Symp. Found. Comput. Sci.*, 2013, pp. 429–438.
[17] X. Ren, C.-M. Yu, W. Yu, S. Yang, X. Yang, J. A. McCann, and P. S. Yu, "LoPub: High-dimensional crowdsourced data publication with local differential privacy," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 9, pp. 2151–2166, Jul. 2018.
[18] T. Wang, X. Yang, X. Ren, W. Yu, and S. Yang, "Locally private high-dimensional crowdsourced data release based on copula functions," *IEEE Trans. Services Comput.*, vol. 15, no. 2, pp. 778–792, Mar. 2022.
[19] T. Mimoto, S. Kiyomoto, S. Hidano, A. Basu, and A. Miyaji, "The possibility of matrix decomposition as anonymization and evaluation for time-sequence data," in *Proc. 16th Annu. Conf. Privacy, Secur. Trust (PST)*, Aug. 2018, pp. 1–7.
[20] T. Yamazoe, H. Yamashiro, K. Omote, A. Imakura, and T. Sakurai, "Image data recoverability against data collaboration and its countermeasure," in *Proc. Sci. Cyber Secur. SciSec Workshops*, C. Su and K. Sakurai, Eds. Singapore: Springer, 2022, pp. 3–15.
[21] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.* New York, NY, USA: Association for Computing Machinery, Oct. 2015, p. 1322, doi: 10.1145/2810103.2813677.
[22] A. Bogdanova, A. Nakai, Y. Okada, A. Imakura, and T. Sakurai, "Federated learning system without model sharing through integration of dimensional reduced data representations," 2020, *arXiv:2011.06803*.
[23] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2017, pp. 4768–4777.
[24] A. Bogdanova, A. Imakura, T. Sakurai, T. Fujii, T. Sakamoto, and H. Abe, "Achieving transparency in distributed machine learning with explainable data collaboration," 2022, *arXiv:2212.03373*.
[25] M. Abadi, A. Chu, I. Goodfellow, H. B. Mcmahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.* New York, NY, USA: Association for Computing Machinery, Oct. 2016, pp. 308–318, doi: 10.1145/2976749.2978318.
[26] L. Deng, "The MNIST database of handwritten digit images for machine learning research [best of the web]," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 141–142, Nov. 2012.
[27] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," 2017, *arXiv:1708.07747*.

HIROMI YAMASHIRO received the B.S. degree in information engineering from the College of Information Science, University of Tsukuba, Japan, in 2022, where he is currently pursuing the master's degree in risk and resilience engineering. His research interests include computer security, privacy in data analysis or machine learning, and data analysis for distributed data.

KAZUMASA OMOTE received the Ph.D. degree in information science from Japan Advanced Institute of Science and Technology (JAIST), in 2002. He was with Fujitsu Laboratories Ltd., from 2002 to 2008, and was engaged in research and development for network security. He was a Research Assistant Professor with JAIST, from 2008 to 2011; an Associate Professor with JAIST, from 2011 to 2016; and an Associate Professor with the University of Tsukuba, from 2016 to 2022. He has been a Professor with the University of Tsukuba, since 2022. His research interests include applied cryptography, network security, and blockchain security. He received the WISTP 2019 Best Paper Award. He was the General Co-Chair of the ACNS 2023 International Conference.

AKIRA IMAKURA received the Ph.D. degree from Nagoya University, Japan, in 2011. He was appointed as Japan Society for the Promotion of Science Research Fellowship for Doctor Course Student (DC2), from 2010 to 2011, and a Research Fellow with the Center for Computational Sciences, University of Tsukuba, Japan, from 2011 to 2013. He is currently an Associate Professor with the Faculty of Engineering, Information and Systems, University of Tsukuba. His current research interest includes developments and analysis of highly parallel algorithms for large matrix computations. Recently, he also investigated matrix factorization-based machine learning algorithms including privacy-preserving machine learning algorithms. He is a member of Japan Society for Industrial and Applied Mathematics (JSIAM), the Information Processing Society of Japan (IPSJ), Japan Association for Medical Informatics (JAMI), and the Society for Industrial and Applied Mathematics (SIAM).

TETSUYA SAKURAI received the Ph.D. degree from Nagoya University, in 1992. He is currently a Professor and the Director of the Center for Artificial Intelligence Research (C-AIR), University of Tsukuba. He is a member of the Society for Industrial and Applied Mathematics (SIAM), Japan Society for Industrial and Applied Mathematics (JSIAM), the Mathematical Society of Japan (MSJ), and the Information Processing Society of Japan (JPSJ). Furthermore, he is a fellow of JSIAM. He was awarded the 2018's Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology of Japan.