Collaborative Novelty Detection for Distributed Data by a Probabilistic Method

Akira Imakura IMAKURA@CS.TSUKUBA.AC.JP
Xiucai Ye YEXIUCAI@CS.TSUKUBA.AC.JP
Tetsuya Sakurai SAKURAI@CS.TSUKUBA.AC.JP
University of Tsukuba, 1-1-1 Tennodai, Ibaraki, Tsukuba 305-8573, Japan

Editors: Vineeth N Balasubramanian and Ivor Tsang

Abstract

Novelty detection, which detects anomalies based on a training dataset consisting of only the normal data, is an important task in several applications. In addition, in the real world, there may be situations where data is owned by multiple parties in a distributed manner but cannot be shared with each other due to privacy and confidentiality requirements. Therefore, how to develop distributed novelty detection while preserving privacy is essential. To address this challenge, we propose a probabilistic collaborative method that allows distributed novelty detection for multiple parties without sharing the original data. The proposed method constructs a collaborative kernel based on a collaborative data analysis framework, by which intermediate representations are generated from each party and shared for collaborative novelty detection. Numerical experiments demonstrate that the proposed method obtains better performance compared with the individual novelty detection in the local party.

Keywords: Novelty detection, Distributed data, Collaborative data analysis framework, Intermediate representation, Collaborative kernel.

1. Introduction

1.1. Motivation

Novelty detection detects anomalies in test data which significantly differ from the training data. Since the training data consists of only the normal data, novelty detection is considered as a challenging and important task in several applications such as medical diagnostic problems (Clifton et al. (2011)), sensor networks (Zhang et al. (2010)), video surveillance (Diehl and Hampshire (2002)), and detection of masses in mammograms (Tarassenko et al. (1995)). In addition, in the real world, there may be situations where data is owned by multiple parties in a distributed manner but cannot be shared with each other due to privacy and confidentiality requirements. Distributed data analysis methods, which aim to analyse such kind of distributed data without sharing local data between parties, have recently attracted significant attention.

A motivating example is a failure and defect detection for distributed manufacturing datasets. When the same type of products are manufactured by multiple companies, the product data are owned in distributed manner. In addition, for each product, the manufacturing data and trial data will be owned by the manufacturing company and the purchasing

company. Because of a limited number of samples and features, it is difficult to detect anomalies only using a data in a single party. Centralizing the data from multiple parties could help to achieve a high-quality novelty detections; however, it is difficult to share manufacturing and trial data between multiple parties due to confidentiality requirements. A similar situation occurs in medical and financial data.

Thus, a development of a novelty detection method for horizontal and vertical distributed data, which are partitioned according to samples and features, is essential.

1.2. Main purpose and contributions

Some methods have been proposed for detecting anomalies from distributed data but without considering the privacy-preserving problem (Chatzigiannakis et al. (2006); Deshmeh and Rahmati (2008)). On the other hand, typical techniques including cryptographic computations (or secure multi-party computation) (Jha et al. (2005); Cho et al. (2018); Gilad-Bachrach et al. (2016)) and differential privacy methods (Abadi et al. (2016); Ji et al. (2014); Dwork (2006)) can be applied to protect the privacy of the original data. However, the high computational cost or the randomization implementation may cause inefficiency for distributed novelty detection. Recently, the federated learning systems been proposed for distributed data analysis and privacy preserving. The federated learning systems can be classified into model share-type federated learning (Li et al. (2019); Konečnỳ et al. (2016a,b); McMahan et al. (2016); Yang et al. (2019)) and non-model share-type collaborative data analysis (Imakura and Sakurai (2020); Imakura et al. (2021c); Ye et al. (2019); Takahashi et al. (2021); Imakura et al. (2021b)).

To realize a novelty detection method for distributed data without sharing the original data, in this paper, we propose a probabilistic novelty detection method based on the non-model share-type collaborative data analysis framework. The proposed method utilizes least square probabilistic analysis using *a collaborative kernel* which is constructed based on the intermediate representations from individual data in local parties. Finally, a novelty score based on the collaborative kernel is computed to detect the anomalous data.

The main contributions of the proposed method are summarized as follows:

- The proposed method utilizes a least square probabilistic analysis using a collaborative kernel which contains the relationship information between data in all parties and thus allows collaborative novelty detection without revealing the private data for horizontal and vertical data distribution.
- The obtained novelty detection model is constructed based on all features and samples of the distributed data, which is impossible in individual analysis in a local party.

Numerical experiments on both artificial and real-world data indicate that the proposed method obtains better performance than individual novelty detection and comparable to that of centralized novelty detection.

1.3. Organization of the paper

The remainder of this paper is organized as follows. In Section 2, we briefly introduce probabilistic novelty detection and federated learning systems as related works. In Section 3, we propose a novel collaborative novelty detection method for distributed data. Numerical

results are reported in Section 4. Finally, in Section 5, we summarize our results and conclude the paper.

2. Related works

2.1. Probabilistic novelty detection

Typical novelty detection methods include probabilistic, distance-based, domain-based, reconstruction-based, and information theoretic methods (Chandola et al. (2009)).

Here, we focus on probabilistic novelty detection methods. Probabilistic novelty detection computes a novelty score $w(\boldsymbol{x}) \in \mathbb{R}$ for the data $\boldsymbol{x}$ and detects whether $\boldsymbol{x}$ is anomalous using a prescribed threshold $\tau \in \mathbb{R}$. If $w(\boldsymbol{x}) > \tau$, $\boldsymbol{x}$ is anomalous otherwise $\boldsymbol{x}$ is normal. Some non-parametric methods have been proposed (Parzen (1962); Sugiyama et al. (2008)), which contains no information on the underlying distribution. However, when training data is multimodal, their performance decrease. To deal with the multimodal data, a novelty detection method based on least square probabilistic analysis (ND-LSPA) has been proposed (Yoda et al. (2020)).

Here, we briefly introduce ND-LSPA for multimodal data. Assume that the datasets $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ is multimodal. Let L be the number of classes and $y_k \in \{1, 2, \ldots, L\}$ be the label for each training data $\boldsymbol{x}_k$. A novelty score is defined as

$$w(\boldsymbol{x}) := p(y = 0|\boldsymbol{x}) = 1 - \sum_{\ell=1}^{L} p(y = \ell|\boldsymbol{x}), \tag{1}$$

where $p(y = 0|\boldsymbol{x})$ is a posterior probability for anomalous and $p(y = \ell|\boldsymbol{x})$ is a posterior probability for each class of normal data. Then, the posterior probability $p(y = \ell|\boldsymbol{x})$ $(\ell = 1, 2, \ldots, L)$ is modeled as

$$p(y = \ell|\boldsymbol{x}) \approx q(y = \ell|\boldsymbol{x}; \boldsymbol{\alpha}_\ell) = \boldsymbol{\alpha}_\ell^{\mathrm{T}} \boldsymbol{\phi}(\boldsymbol{x}),$$

with a weight vector $\boldsymbol{\alpha}_\ell \in \mathbb{R}^n$ and a basis vector $\boldsymbol{\phi}(\boldsymbol{x}) = [K(\boldsymbol{x}, \boldsymbol{x}_1), K(\boldsymbol{x}, \boldsymbol{x}_2), \ldots, K(\boldsymbol{x}, \boldsymbol{x}_n)]^{\mathrm{T}} \in \mathbb{R}^n$, where

$$K(\boldsymbol{x}, \boldsymbol{x}_k) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}_k\|_2^2}{2\sigma^2}\right), \quad k = 1, 2, \ldots, n$$

is a kernel function with a band width σ. The weight vector $\boldsymbol{\alpha}_\ell$ is set to minimize the following squared loss,

$$J(\boldsymbol{\alpha}_\ell) = \frac{1}{2} \int \left(q(y = \ell|\boldsymbol{x}; \boldsymbol{\alpha}_\ell) - p(y = \ell|\boldsymbol{x})\right)^2 p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{2}$$

2.2. Federated learning systems

The federated learning systems can be classified into model share-type federated learning (Li et al. (2019); Konečnỳ et al. (2016a,b); McMahan et al. (2016); Yang et al. (2019)) and non-model share-type collaborative data analysis (Imakura and Sakurai (2020); Imakura et al. (2021c); Ye et al. (2019); Takahashi et al. (2021); Imakura et al. (2021b)).

The model share-type federated learning has been first proposed by Google (Konečnỳ et al. (2016b,a)), which are mainly based on (deep) neural network. All parties share the model

and update it in each iteration without sharing the local data. The federated averaging (FedAvg) algorithm is widely used for model updating, which takes either the model average or the gradient average of the local model weight or gradient updates from all parties (Li et al. (2019)). However, for model share-type federated learning, we may need to care for a privacy of the original dataset due to the shared functional model (Yang (2019)).

Instead of sharing the model, the non-model share-type *collaborative data analysis* was proposed in (Imakura and Sakurai (2020)) to utilizes *intermediate representations* from each party. By allowing different functions to be used in different parties to generate intermediate representations, the collaborative data analysis ensure both the recognition performance and privacy as analyzed in (Imakura et al. (2021a)). Unlike the model-sharing federated learning, the collaborative data analysis does not require iterative computing with cross-organization communications. The performance comparison between the non-model share-type collaborative data analysis and the model share-type federated learning has been reported in (Bogdanova et al. (2021)).

3. Probabilistic collaborative novelty detection for distributed data

3.1. Distributed novelty detection

Let m, n, and s denote the numbers of features, training data samples, and test data samples. Novelty detection aims to detect whether the test data $X^{\mathrm{test}} = [\boldsymbol{x}_1^{\mathrm{test}}, \boldsymbol{x}_2^{\mathrm{test}}, \ldots, \boldsymbol{x}_s^{\mathrm{test}}]^{\mathrm{T}} \in \mathbb{R}^{s \times m}$ is normal or anomalous based on only normal training dataset $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]^{\mathrm{T}} \in \mathbb{R}^{n \times m}$.

In this paper, for distributed novelty detection, we consider the horizontal and vertical distribution, i.e., data samples are partitioned into c parties and features are partitioned into d parties as follows:

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,d} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ X_{c,1} & X_{c,2} & \cdots & X_{c,d} \end{bmatrix}, \quad X^{\mathrm{test}} = \begin{bmatrix} X_{1,1}^{\mathrm{test}} & X_{1,2}^{\mathrm{test}} & \cdots & X_{1,d}^{\mathrm{test}} \\ X_{2,1}^{\mathrm{test}} & X_{2,2}^{\mathrm{test}} & \cdots & X_{2,d}^{\mathrm{test}} \\ \vdots & \vdots & \ddots & \vdots \\ X_{c,1}^{\mathrm{test}} & X_{c,2}^{\mathrm{test}} & \cdots & X_{c,d}^{\mathrm{test}} \end{bmatrix}. \tag{3}$$

Then, the (i,j)-th party has partial dataset $X_{i,j} \in \mathbb{R}^{n_i \times m_j}$ and $X_{i,j}^{\mathrm{test}} \in \mathbb{R}^{s_i \times m_j}$. All parties do not want to share the original data $X_{i,j}, X_{i,j}^{\mathrm{test}}$, but aim to detect the test data X^{test}. Since the training dataset is distributed in multiple parties, the training dataset is considered to be multimodal.

Individual novelty detection using only the dataset in a local party may not have high-quality novelty detection results due to a lack of feature information or insufficient samples. If we can centralize the datasets from multiple parties and analyze them as one dataset, i.e., *centralized novelty detection*, then we expect to achieve a high-quality novelty detection. However, it is difficult to share the individual data for centralization due to privacy and confidentiality concerns.

3.2. Main concept

Here, we derive a probabilistic collaborative novelty detection for distributed dataset (3). The proposed method utilizes least square probabilistic analysis using *a collaborative kernel*

which is constructed based on the intermediate representations from individual data in local parties. The main concepts of the proposed method are the followings.

- The novelty score $w(\boldsymbol{x})$ is computed by (1) based on the idea of ND-LSPA method to deal with multimodal data from multiple parties.
- A collaborative kernel is constructed without sharing the local data by the non-model share-type collaborative data analysis framework.

3.3. Derivation of practical algorithm

For distributed novelty detection, we need to compute the kernel function without sharing the local data $X_{i,j}$. To realize this, we newly introduce a collaborative kernel based on the idea of collaborative data analysis framework as follows. Based on the framework of the collaborative data analysis, the proposed method is operated by two roles: *worker* and *master*. Each worker, who has the private dataset $X_{i,j}$, individually constructs a dimensionally-reduced intermediate representation and shares it to the master. The master transforms the shared intermediate representations to the collaboration representations and analyzes them to compute a novelty score.

First, all workers generate the same anchor dataset $X^{\mathrm{anc}} \in \mathbb{R}^{r \times m}$, which is a shareable data consisting of public data or dummy data randomly constructed, and partition it by features as $X^{\mathrm{anc}} = [X^{\mathrm{anc}}_{:,1}, X^{\mathrm{anc}}_{:,2}, \ldots, X^{\mathrm{anc}}_{:,d}]$. The usage of a large number of anchor data is expected to improve recognition performance as shown in (Bogdanova et al. (2021)) for classification problems. On the other hand, the computational costs increase with increasing the number of anchor data. Also, the usage of statics of training dataset for constructing anchor data is expected to improve recognition performance as shown in (Takahashi et al. (2021)) for a medical dataset, although the collaborative data analysis shows a good performance even with random anchor data (Imakura and Sakurai (2020); Imakura et al. (2021c)).

Then, each worker constructs the intermediate representations,

$$\widetilde{X}_{i,j} = f_{i,j}(X_{i,j}) \in \mathbb{R}^{n_i \times \widetilde{m}_{i,j}}, \quad \widetilde{X}^{\mathrm{anc}}_{i,j} = f_{i,j}(X^{\mathrm{anc}}_{:,j}) \in \mathbb{R}^{r \times \widetilde{m}_{i,j}},$$

where $f_{i,j}$ denotes a linear or nonlinear row-wise mapping function and centralize the intermediate representations to the master. A typical setting for $f_{i,j}$ is an unsupervised dimensionality reduction (Pearson (1901); He and Niyogi (2004); Maaten and Hinton (2008)), with $\widetilde{m}_{i,j} < m_{i,j}$. If some workers have label information for classification of workers' local data, they can use supervised methods (Fisher (1936); Sugiyama (2007); Li et al. (2017); Imakura et al. (2019)).

In the master-side, the mapping function g_i for the collaboration representation is constructed satisfying

$$\widehat{X}^{\mathrm{anc}}_i = g_i(\widetilde{X}^{\mathrm{anc}}_i) \in \mathbb{R}^{r \times \widehat{m}} \quad \text{s.t. } \widehat{X}^{\mathrm{anc}}_i \approx \widehat{X}^{\mathrm{anc}}_{i'} \quad (i \neq i'),$$

in some sense, where $\widetilde{X}^{\mathrm{anc}}_i = [\widetilde{X}^{\mathrm{anc}}_{i,1}, \widetilde{X}^{\mathrm{anc}}_{i,2}, \ldots, \widetilde{X}^{\mathrm{anc}}_{i,d}]$. In practice, we use a linear function $\widehat{X}_i = g_i(\widetilde{X}_i) = \widetilde{X}_i G_i$, as the same manner as written in (Imakura and Sakurai (2020); Imakura et al. (2021c)). Using the low-rank approximation based on singular value decomposition of the matrix $[\widetilde{X}^{\mathrm{anc}}_1, \widetilde{X}^{\mathrm{anc}}_2, \ldots, \widetilde{X}^{\mathrm{anc}}_c] \approx U_1 \Sigma_1 V_1^{\mathrm{T}}$, the matrix G_i is computed as

$G_i = (X_i^{\mathrm{anc}})^{\dagger} U_1$, where $\dagger$ denotes pseudo-inverse. We also set $\widetilde{X}_i = [\widetilde{X}_{i,1}, \widetilde{X}_{i,2}, \ldots, \widetilde{X}_{i,d}]$ and $\widehat{X}_i = g_i(\widetilde{X}_i)$.

Then, the obtained collaboration representations $\widehat{X}_i$ $(i = 1, 2, \ldots, c)$ can be analyzed as one dataset,

$$\widehat{X} = [\widehat{\boldsymbol{x}}_1, \widehat{\boldsymbol{x}}_2, \ldots, \widehat{\boldsymbol{x}}_n]^{\mathrm{T}} = \begin{bmatrix} \widehat{X}_1 \\ \widehat{X}_2 \\ \vdots \\ \widehat{X}_c \end{bmatrix} \in \mathbb{R}^{n \times \widehat{m}}.$$

Using the collaboration representation, we define the collaborative kernel function $\widehat{K}(\boldsymbol{x}, \boldsymbol{x}_k)$ as

$$\widehat{K}(\boldsymbol{x}, \boldsymbol{x}_k) = \exp\left(-\frac{\|\widehat{\boldsymbol{x}} - \widehat{\boldsymbol{x}}_k\|_2^2}{2\sigma^2}\right), \quad k = 1, 2, \ldots, n$$

with a band width σ, where $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{x}}_k$ are the corresponding collaboration representation of $\boldsymbol{x}$ and $\boldsymbol{x}_k$, respectively. Note that the collaborative kernel $\widehat{K}(\boldsymbol{x}, \boldsymbol{x}_k)$ is defined by the collaboration representations. Therefore, $\widehat{K}(\boldsymbol{x}, \boldsymbol{x}_k)$ can be computed without sharing the original data via the intermediate representation. On the other hand, $\widehat{K}(\boldsymbol{x}, \boldsymbol{x}_k)$ is an approximation of a standard kernel $K(\boldsymbol{x}, \boldsymbol{x}_k)$ using the original data in some sense. We will evaluate its approximation in numerical experiments.

Then, for computing the novelty score $w(\boldsymbol{x})$ by (1), we model the posterior probability $p(y = \ell|\boldsymbol{x})$ $(\ell = 1, 2, \ldots, L)$ as

$$p(y = \ell|\boldsymbol{x}) \approx q(y = \ell|\boldsymbol{x}; \boldsymbol{\alpha}_\ell) = \boldsymbol{\alpha}_\ell^{\mathrm{T}} \widehat{\boldsymbol{\phi}}(\boldsymbol{x}), \tag{4}$$

with a basis vector using the collaborative kernel,

$$\widehat{\boldsymbol{\phi}}(\boldsymbol{x}) = [\widehat{K}(\boldsymbol{x}, \boldsymbol{x}_1), \widehat{K}(\boldsymbol{x}, \boldsymbol{x}_2), \ldots, \widehat{K}(\boldsymbol{x}, \boldsymbol{x}_n)]^{\mathrm{T}} \in \mathbb{R}^n.$$

For computing $\boldsymbol{\alpha}_\ell$ in (4), the squared loss $J(\boldsymbol{\alpha}_\ell)$ (2) is rewritten as

$$\begin{aligned} J(\boldsymbol{\alpha}_\ell) &= \frac{1}{2}\int q(y = \ell|\boldsymbol{x}; \boldsymbol{\alpha}_\ell)^2 p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &\quad - \int q(y = \ell|\boldsymbol{x}; \boldsymbol{\alpha}_\ell) p(y = \ell|\boldsymbol{x}) p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + \frac{1}{2}\int p(y = \ell|\boldsymbol{x})^2 p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &= \frac{1}{2}\int \boldsymbol{\alpha}_\ell^{\mathrm{T}} \widehat{\boldsymbol{\phi}}(\boldsymbol{x}) \widehat{\boldsymbol{\phi}}(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{\alpha}_\ell p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} - \int \boldsymbol{\alpha}_\ell^{\mathrm{T}} \widehat{\boldsymbol{\phi}}(\boldsymbol{x}) p(\boldsymbol{x}, y = \ell) \mathrm{d}\boldsymbol{x} + \mathrm{Const}, \end{aligned}$$

where we used the model (4) and

$$p(y = \ell|\boldsymbol{x}) = \frac{p(\boldsymbol{x}, y = \ell)}{p(\boldsymbol{x})}.$$

Then, using approximations based on sample average and sample rate of the training datasets,

$$\begin{aligned} \int \boldsymbol{\alpha}_\ell^{\mathrm{T}} \widehat{\boldsymbol{\phi}}(\boldsymbol{x}) \widehat{\boldsymbol{\phi}}(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{\alpha}_\ell p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} &\approx \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\alpha}_\ell^{\mathrm{T}} \widehat{\boldsymbol{\phi}}(\boldsymbol{x}_i) \widehat{\boldsymbol{\phi}}(\boldsymbol{x}_i)^{\mathrm{T}} \boldsymbol{\alpha}_\ell, \\ \int \boldsymbol{\alpha}_\ell^{\mathrm{T}} \widehat{\boldsymbol{\phi}}(\boldsymbol{x}) p(\boldsymbol{x}, y = \ell) \mathrm{d}\boldsymbol{x} &\approx \frac{1}{n}\sum_{y_i = \ell} \boldsymbol{\alpha}_\ell^{\mathrm{T}} \widehat{\boldsymbol{\phi}}(\boldsymbol{x}_i), \end{aligned}$$

Algorithm 1 Probabilistic collaborative novelty detection

Input: $X_{i,j} \in \mathbb{R}^{n_i \times m_j}$ and $\boldsymbol{x}^{\text{test}}$ individually
Output: $w(X_i^{\text{test}})$ for each i

	worker-side (i,j)		*master-side*
	——— Training phase ———		
1:	Generate $X_{i,j}^{\text{anc}}$ and share to all workers		
2:	Set $X_{:,j}^{\text{anc}}$		
3:	Generate $f_{i,j}$		
4:	Compute $\widetilde{X}_{i,j} = f_{i,j}(X_{i,j})$		
5:	Compute $\widetilde{X}_{i,j}^{\text{anc}} = f_{i,j}(X_{:,j}^{\text{anc}})$		
6:	Share $\widetilde{X}_{i,j}$ and $\widetilde{X}_{i,j}^{\text{anc}}$ to master	$\rightarrow$	Get $\widetilde{X}_{i,j}$ and $\widetilde{X}_{i,j}^{\text{anc}}$ for all i and j
7:			Set $\widetilde{X}_i$ and $\widetilde{X}_i^{\text{anc}}$
8:			Construct g_i from $\widetilde{X}_i^{\text{anc}}$ for all i
9:			Compute $\widehat{X}_i = g_i(\widetilde{X}_i)$ for all i
10:			Set $\widehat{X}$
11:			Construct w from $\widehat{X}$
	——— Prediction phase ———		
12:	Set $\widetilde{X}_{i,j}^{\text{test}} = f_{i,j}(X_{i,j}^{\text{test}})$		
13:	Share $\widetilde{X}_{i,j}^{\text{test}}$ to master	$\rightarrow$	Get $\widetilde{X}_{i,j}^{\text{test}}$ and set $\widetilde{X}_i^{\text{test}}$
14:			Compute $w(X_i^{\text{test}}) = w(g_i(\widetilde{X}_i^{\text{test}}))$
15:	Get $w(X_i^{\text{test}})$	$\leftarrow$	Return $w(X_i^{\text{test}})$ to worker

and 2-norm regularization with a parameter λ, the approximated weight vector $\widetilde{\boldsymbol{\alpha}}_\ell$ is obtained by minimizing

$$\frac{1}{2n}\sum_{i=1}^{n} \boldsymbol{\alpha}_\ell^{\mathrm{T}} \widehat{\boldsymbol{\phi}}(\boldsymbol{x}_i)\widehat{\boldsymbol{\phi}}(\boldsymbol{x}_i)^{\mathrm{T}}\boldsymbol{\alpha}_\ell - \frac{1}{n}\sum_{y_i=\ell} \boldsymbol{\alpha}_\ell^{\mathrm{T}}\widehat{\boldsymbol{\phi}}(\boldsymbol{x}_i) + \frac{\lambda}{2}\|\alpha_\ell\|_2.$$

Then, taking the derivative, $\widetilde{\boldsymbol{\alpha}}_\ell$ is computed by solving the linear system of size n,

$$(\widehat{\Phi}^{\mathrm{T}}\widehat{\Phi} + \lambda n I_n)\widetilde{\boldsymbol{\alpha}}_\ell = \widehat{\Phi}^{\mathrm{T}}\boldsymbol{z}_\ell,$$

where $\widehat{\Phi} \in \mathbb{R}^{n\times n}$ is a collaborative kernel matrix $\widehat{\Phi} = [\widehat{\boldsymbol{\phi}}(\boldsymbol{x}_1), \widehat{\boldsymbol{\phi}}(\boldsymbol{x}_2), \ldots, \widehat{\boldsymbol{\phi}}(\boldsymbol{x}_n)]^{\mathrm{T}}$ and $\boldsymbol{z}_\ell = [\delta_{\ell,y_1}, \delta_{\ell,y_2}, \ldots, \delta_{\ell,y_n}]^{\mathrm{T}}$ with the Kronecker delta $\delta_{i,j}$. Note again that since $\widehat{\Phi}$ is defined by the collaborative kernel, $\widehat{\Phi}$ can be computed without sharing the original data via the intermediate representations.

The novelty score is finally computed as

$$w(\boldsymbol{x}) = 1 - \sum_{\ell=1}^{L} q(y = \ell | \boldsymbol{x}; \widetilde{\boldsymbol{\alpha}}_\ell)$$

with some normalizations for $q(y = \ell|\boldsymbol{x}; \widetilde{\boldsymbol{\alpha}}_\ell)$ and $w(\boldsymbol{x})$ such that the maximum value of q for all training datasets is 1 and the minimum value of w for all test datasets is 0. Here,

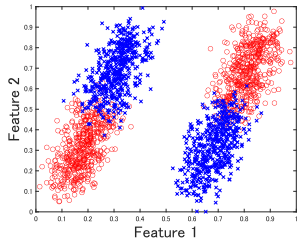


Figure 1: Features 1 and 2 of the training datasets. Markers ○ and × show samples in group 1 and 2, respectively.

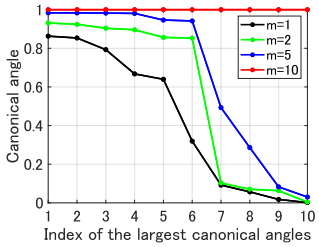


Figure 2: 10 largest canonical angles.

for computing the novelty score of the test data, the test data $\boldsymbol{x}^{\mathrm{test}}$ is also transformed to the collaborative representation $\widehat{\boldsymbol{x}}^{\mathrm{test}}$ via centralizing its intermediate representation.

The algorithm of the proposed collaborative novelty detection is summarized in Algorithm 1. In Algorithm 1, the local data $X_{i,j}, X_{i,j}^{\mathrm{test}}$ and the function $f_{i,j}$ are not shared to others. Instead, the dimensionality-reduced intermediate representations $\widetilde{X}_{i,j}, \widetilde{X}_{i,j}^{\mathrm{test}}$ are shared only to the master. Note that this algorithm requires a class label information of each training data. In the case when there is no class information, we employ some clustering method to $\widehat{X}$ in Step 11.

4. Numerical experiments

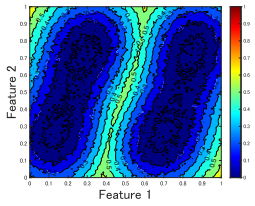
This section evaluates the performance of the proposed collaborative novelty detection (Algorithm 1) and compares it with those of centralized and individual novelty detections. Note that centralized novelty detection is considered as an ideal case since the private datasets $X_{i,j}$ cannot be shared in our target situation. The proposed collaborative novelty detection aims to achieve a better performance than individual novelty detection.

For the centralized and individual novelty detection, we used ND-LSPA (Yoda et al. (2020)). For the proposed method, we used PCA for constructing intermediate representations. The anchor data X^{anc} is constructed as a random matrix on the range of each feature and the number of the anchor data is set as $r = 1000$. For all methods, we set $\lambda = 0.01$ as a regularization parameter. The bandwidth σ is set based on the local scaling (Zelnik-Manor and Perona (2005)).

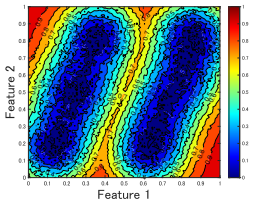
All the numerical experiments were performed on Windows 10 Pro, Intel(R) Core(TM) i7-10710U CPU @ 1.10GHz, 16GB RAM using MATLAB2019b. All random values are generated by Mersenne Twister in MATLAB.

4.1. Experiment I: artificial data

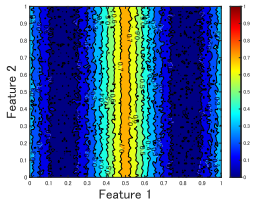
Herein, we used a 20-dimensional artificial data for performance evaluation of novelty detection methods. Figure 1 depicts the first two dimensions of all the training dataset, where the number of samples is $n = 2000$. The other 18 dimensions were random values in $[0, 0.1]$.



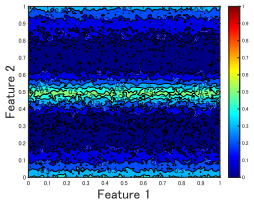
(a) Centralized novelty detection.



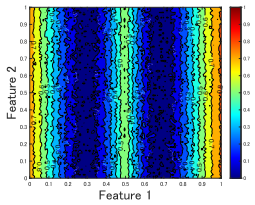
(b) Collaborative novelty detection.



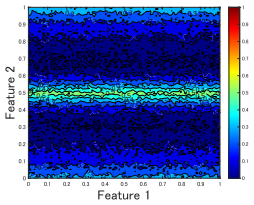
(c) Individual novelty detection for $(1,1)$-party.



(d) Individual novelty detection for $(1,2)$-party.



(e) Individual novelty detection for $(2,1)$-party.



(f) Individual novelty detection for $(2,2)$-party.

Figure 3: Novelty scores obtained by centralized, individual and collaborative novelty detections.

We considered the case in which the dataset in Figure 1 is distributed into four parties: $c = d = 2$. The $(1,1)$ and $(2,1)$-parties have features $1, 3, 5, \ldots, 19$ of $\circ$ and $\times$ samples in Figure 1, respectively. The $(1,2)$ and $(2,2)$-parties have features $2, 4, 6, \ldots, 20$ of $\circ$ and $\times$ samples in Figure 1, respectively.

Firstly, we evaluate the approximation of the collaborative kernel matrix $\widehat{\Phi}$ regarding canonical angles of eigenvectors corresponding to the 10 largest eigenvalues, i.e.,

$$\sigma_i(\widehat{U}^{\mathrm{T}}U), \quad i = 1, 2, \ldots, 10,$$

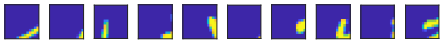
where σ_i is the i-th largest singular value of the matrix and $\widehat{U}$ and $U \in \mathbb{R}^{n\times 10}$ are the orthogonal matrices whose columns are eigenvectors corresponding to the 10 largest eigenvalues of the collaborative and standard kernel matrices $\widehat{\Phi}$ and $\Phi = [\boldsymbol{\phi}(\boldsymbol{x}_1), \boldsymbol{\phi}(\boldsymbol{x}_2), \ldots, \boldsymbol{\phi}(\boldsymbol{x}_n)]$, respectively. Note that if all $\sigma_i = 1$ then $\widehat{U}$ and U have the same range space.

Figure 2 demonstrates 10 largest canonical angles for $\widetilde{m}_{ij} = 1, 2, 5$ and 10. These results indicate that even small $\widetilde{m}_{ij}$, the collaborative kernel matrix $\widehat{\Phi}$ well approximates the standard kernel matrix Φ regarding the subspace spanned by some principle eigenvectors.

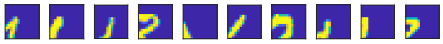
(a) Training datasets.

(b) Test datasets. ($*$ denotes anomalous data.)

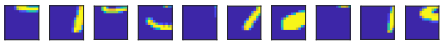
Figure 4: Part of training and test datasets.



(a) $(1,1)$ sub-picture



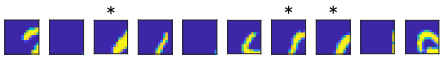
(b) $(1,2)$ sub-picture



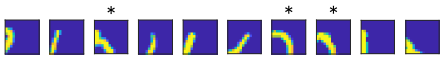
(c) $(2,1)$ sub-picture

(d) $(2,2)$ sub-picture

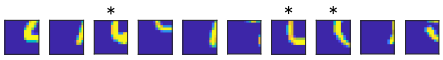
Figure 5: Feature distributions for training dataset.



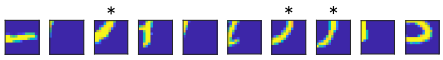
(a) sub-picture set $(1,1)$



(b) sub-picture set $(1,2)$



(c) sub-picture set $(2,1)$



(d) sub-picture set $(2,2)$

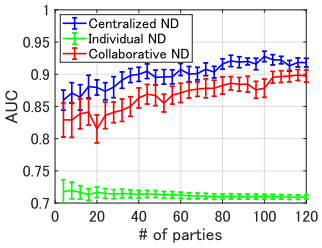
Figure 6: Feature distributions for test dataset.

Next, we computed novelty scores of the test datasets in which each test sample has values in $[0,1]$ for features 1 and 2 and in $[0,0.1]$ for other features. Here, we set $\widetilde{m}_{ij} = 5$ for all parties.

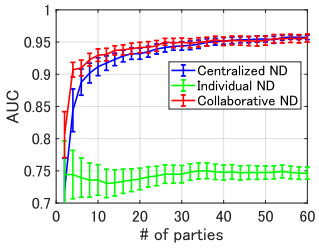
In Figure 3, we show the novelty scores obtained by centralized, individual and collaborative novelty detections. Since the individual novelty detections used only a part of features and of a part of samples, they computed completely different novelty scores (Figure 3(c)–(f)) from that of the centralized novelty detection (Figure 3(a)). We observed that, in this case, we cannot detect anomalous data correctly, even if we share novelty scores of the individual novelty detections. Instead, the proposed collaborative novelty detection obtained similar novelty score (Figure 3(b)) as the centralized novelty detection, even the original datasets remain distributed. Note that the novelty score of the collaborative novelty detection shows larger values than that of the centralized novelty detection for the regions with no training data. This may be because the dimensionality reduction in the collaborative novelty detection reduces the effect of noise.

4.2. Experiment II: performance v.s. number of parties

We used a handwritten digit dataset (MNIST) (LeCun (1998)) and a credit rating dataset "CreditRating_Historical" from the MATLAB Statistics and Machine Learning Toolbox for evaluating the recognition performance v.s. the number of parties.



(a) MNIST



(b) CreditRating_Historical

Figure 7: Average AUC with 95% confidence interval v.s. number of parties.

For MNIST, we set the training dataset with four digits "1" to "4" and set the test dataset additionally with "0" as anomalous data as shown in Figure 4. We split each picture with 28×28 pixels into 2×2 sub-pictures with 14×14 pixels as shown in Figures 5 and 6. Thus, the number of partitions for features is $d = 4$ and each party has a sub-picture set ($m_j = 196$) of $n_i = 25$ samples.

CreditRating_Historical has five financial ratios: Working capital / Total Assets (WC_TA), Retained Earnings / Total Assets (RE_TA), Earnings Before Interests and Taxes / Total Assets (EBIT_TA), Market Value of Equity / Book Value of Total Debt (MVE_BVTD), Sales / Total Assets (S_TA), and industry sector labels from 1 to 12. We set the training dataset with credit ratings from "AAA" to "B" and set the test dataset additionally with "CCC" as anomalous data. We split each sample into two parties: {WC_TA, RE_TA, and EBIT_TA} and {MVE_BVTD, S_TA, and Industry sector label}. Each party has $n_i = 20$ samples.

Then, we evaluated the area under curve (AUC) for each test dataset by increasing the number of partitions for samples c from 1 to 30. For the proposed method, we set $\widetilde{m}_{ij} = 25$ for MNIST and $\widetilde{m}_{ij} = 2$ for CreditRating_Historical.

We evaluated 50 trials and show the average AUC with 95% confidence interval in Figure 7(a) for MNIST and in Figure 7(b) for CreditRating_Historical. It is observed that the performance of the proposed method increases with an increase in the number of parties and achieves a higher recognition performance than the individual novelty detection and comparable to the centralized novelty detection.

4.3. Experiment III: real-world data

We evaluated the performance of the novelty detection methods for mislabeled detection of the binary and multi-class classification problems obtained from (Samaria and Harter (1994)) and feature selection datasets [1]. Let $Y \in \mathbb{R}^{n \times \ell}$ and Y^{test} be label of training data X and test data X^{test}. Mislabeled detection aims to find mislabeled data in test datasets $X^{\text{test}}, Y^{\text{test}}$ from the training datasets X, Y only with the correct label.

For mislabeled detection problems, we compute novelty scores of $[X^{\text{test}}, \alpha Y^{\text{test}}]$ by training $[X, \alpha Y]$ to find the mislabeled data in $[X^{\text{test}}, \alpha Y^{\text{test}}]$, where $\alpha \in \mathbb{R}$ is a scaling parameter. We set $\alpha = 1.1 \times \min_{i \neq j} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2$ where $\boldsymbol{\mu}_i$ is the center of data belong to each class.

1. available at http://featureselection.asu.edu/datasets.php.

Table 1: Average AUC with 95% confidence interval for novelty scores by the centralized, individual and proposed collaborative novelty detections.

Dataset	Centralized	Individual	Collaborative
ALLAML	0.96 ± 0.017	0.92 ± 0.011	0.96 ± 0.017
Carcinom	0.93 ± 0.015	0.72 ± 0.009	0.79 ± 0.024
CLL-SUB-111	0.67 ± 0.058	0.49 ± 0.023	0.51 ± 0.059
COLON	0.87 ± 0.046	0.83 ± 0.020	0.84 ± 0.032
GLA-BRA-180	0.82 ± 0.028	0.69 ± 0.012	0.71 ± 0.021
GLI-85	0.93 ± 0.042	0.83 ± 0.020	0.92 ± 0.030
jaffe	0.99 ± 0.004	0.86 ± 0.008	0.92 ± 0.009
leukemia	0.97 ± 0.017	0.89 ± 0.016	0.97 ± 0.017
lung	0.99 ± 0.004	0.95 ± 0.004	0.96 ± 0.008
lymphoma	0.92 ± 0.023	0.70 ± 0.016	0.84 ± 0.025
pixraw10P	0.99 ± 0.006	0.74 ± 0.018	0.93 ± 0.021
Prostate_GE	0.76 ± 0.020	0.82 ± 0.014	0.83 ± 0.040
SMK-CAN-187	0.62 ± 0.020	0.62 ± 0.016	0.66 ± 0.031
TOX-171	0.89 ± 0.027	0.77 ± 0.014	0.87 ± 0.027
warpAR10P	0.84 ± 0.041	0.65 ± 0.019	0.71 ± 0.051
warpPIE10P	0.96 ± 0.008	0.72 ± 0.016	0.83 ± 0.035

Let the X and Y be partitioned as

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & X_{1,3} \\ X_{2,1} & X_{2,2} & X_{2,3} \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}.$$

Then, we consider the case that the dataset $[X, \alpha Y]$ is distributed into six parties as

$$\begin{bmatrix} X_{1,1}, \frac{\alpha}{3} Y_1 \end{bmatrix}, \quad \begin{bmatrix} X_{1,2}, \frac{\alpha}{3} Y_1 \end{bmatrix}, \quad \begin{bmatrix} X_{1,3}, \frac{\alpha}{3} Y_1 \end{bmatrix},$$
$$\begin{bmatrix} X_{2,1}, \frac{\alpha}{3} Y_2 \end{bmatrix}, \quad \begin{bmatrix} X_{2,2}, \frac{\alpha}{3} Y_2 \end{bmatrix}, \quad \begin{bmatrix} X_{2,3}, \frac{\alpha}{3} Y_2 \end{bmatrix}.$$

The performance of each method is evaluated by utilizing five-fold cross-validation. In each training set of the cross-validation, we evaluated 10 trials with random data distribution. We set 10% as the rate of mislabeling in test data. For the proposed method, we set $\widetilde{m}_{ij} = 25$. For the centralized and individual novelty detection, we reduce the dimensionality of the data using PCA that improves the recognition performance in our preliminary experiment.

The numerical results of centralized, individual, and the collaborative novelty detections for 16 test problems are presented in Table 1. We can observe from Table 1 that the proposed method has a recognition performance higher than that of individual novelty detection and competitive to that of centralized novelty detection on most datasets.

4.4. Remarks on numerical results

The experiment I demonstrates that the introduced collaborative kernel matrix well approximates the standard kernel matrix and thus it provides a high recognition performance by the proposed novelty detection. The results of experiments II and III indicate that the proposed collaborative novelty detection achieves a high recognition performance for real-world problems.

Therefore, we conclude that the efficiency of the main aspects of the proposed method is confirmed by these numerical experiments.

5. Conclusions

To address the challenge of distributed novelty detection while preserving privacy in real-world applications, we propose a probabilistic distributed novelty detection method for horizontal and vertical distributed datasets. The proposed method is a probabilistic and non-parametric novelty detection method using a collaborative kernel based on a non-model share-type collaborative data analysis framework. The proposed method generates dimensionally-reduced intermediate representations from individual data in local parties, which are then shared instead of the individual data and models. Then, the proposed method computes a novelty score based on a collaborative kernel via the intermediate representations for detecting anomalous data using whole features.

Numerical experiments on both artificial and real-world data show that the proposed method realizes a novelty detection with better recognition performance than individual novelty detection and comparable to that of centralized novelty detection. It is inferred that the proposed method would become a breakthrough technology for distributed data analysis in the real-world including distributed medical data analysis and distributed manufacturing data analysis.

In our future studies, we will apply the proposed method to practical distributed data in, e.g., medical or manufacturing fields and evaluate its recognition performance. We will investigate to combine the ideas of data collaboration method and other novelty detection method.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive comments. This work was supported in part by the New Energy and Industrial Technology Development Organization (NEDO) and the Japan Society for the Promotion of Science (JSPS), Grants-in-Aid for Scientific Research (Nos. 19KK0255, 21H03451).

References

Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318. ACM, 2016.

Anna Bogdanova, Akie Nakai, Yukihiko Okada, Akira Imakura, and Tetsuya Sakurai. Federated learning system without model sharing through integration of dimensional reduced data representations. In *Proceedings of International Workshop on Federated Learning for User Privacy and Data Confidentiality in Conjunction with IJCAI 2020 (FL-IJCAI'20)*, 2021.

Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58, 2009.

Vasilis Chatzigiannakis, Symeon Papavassiliou, Mary Grammatikou, and B Maglaris. Hierarchical anomaly detection in distributed large-scale sensor networks. In *11th IEEE Symposium on Computers and Communications (ISCC'06)*, pages 761–767. IEEE, 2006.

Hyunghoon Cho, David J Wu, and Bonnie Berger. Secure genome-wide association analysis using multiparty computation. *Nature biotechnology*, 36(6):547, 2018.

Lei Clifton, David A Clifton, Peter J Watkinson, and Lionel Tarassenko. Identification of patient deterioration in vital-sign data using one-class support vector machines. In *2011 federated conference on computer science and information systems (FedCSIS)*, pages 125–131. IEEE, 2011.

G. Deshmeh and M. Rahmati. Distributed anomaly detection, using cooperative learners and association rule analysis. *Intelligent Data Analysis*, 12:339–357, 2008.

Christopher P Diehl and John B Hampshire. Real-time object classification and novelty detection for collaborative video surveillance. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, volume 3, pages 2620–2625. IEEE, 2002.

C. Dwork. Differential privacy. In *Bugliesi M., Preneel B., Sassone V., Wegener I. (eds) Automata, Languages and Programming. ICALP 2006. Lecture Notes in Computer Science*, volume 4052, 2006.

Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of human genetics*, 7(2):179–188, 1936.

Ran Gilad-Bachrach, Nathan Dowlin, Kim Laine, Kristin Lauter, Michael Naehrig, and John Wernsing. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In *International Conference on Machine Learning*, pages 201–210, 2016.

Xiaofei He and Partha Niyogi. Locality preserving projections. In *Advances in neural information processing systems*, pages 153–160, 2004.

Akira Imakura and Tetsuya Sakurai. Data collaboration analysis framework using centralization of individual intermediate representations for distributed data sets. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 6: 04020018, 2020.

Akira Imakura, Momo Matsuda, Xiucai Ye, and Tetsuya Sakurai. Complex moment-based supervised eigenmap for dimensionality reduction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3910–3918, 2019.

Akira Imakura, Anna Bogdanova, Takaya Yamazoe, Kazumasa Omote, and Tetsuya Sakurai. Accuracy and privacy evaluations of collaborative data analysis. In *Proceedings of The Second AAAI Workshop on Privacy-Preserving Artificial Intelligence (PPAI-21)*, 2021a.

Akira Imakura, Hiroaki Inaba, Yukihiko Okada, and Tetsuya Sakurai. Interpretable collaborative data analysis on distributed data. *Expert Systems with Applications*, 177:114891, 2021b.

Akira Imakura, Xiucai Ye, and Tetsuya Sakurai. Collaborative data analysis: Non-model sharing-type machine learning for distributed data. In *Uehara H., Yamaguchi T., Bai Q. (eds) Knowledge Management and Acquisition for Intelligent Systems. PKAW 2021. Lecture Notes in Computer Science*, volume 12280, pages 14–29, 2021c.

Somesh Jha, Luis Kruger, and Patrick McDaniel. Privacy preserving clustering. In *European Symposium on Research in Computer Security*, pages 397–417. Springer, 2005.

Zhanglong Ji, Zachary C Lipton, and Charles Elkan. Differential privacy and machine learning: A survey and review. *arXiv preprint*, page arXiv:1412.7584, 2014.

Jakub Konečnỳ, H. Brendan McMahan, Daniel Ramage, and Peter Richtarik. Federated optimization: Distributed machine learning for on-device intelligence. *arXiv preprint*, page arXiv:1610.02527, 2016a.

Jakub Konečnỳ, H. Brendan McMahan, Felix X. Yu, Peter Richtarik, Ananda Theertha Suresh, and Dave Bacon. Federated learning: Strategies for improving communication efficiency. In *NIPS Workshop on Private Multi-Party Machine Learning*, 2016b. URL https://arxiv.org/abs/1610.05492.

Yann LeCun. The MNIST database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998.

Q. Li, Z. Wen, Z. Wu, S. Hu, N. Wang, and B. He. A survey on federated learning systems: Vision, hype and reality for data privacy and protection. *arXiv preprint*, page arXiv:1907.09693, 2019.

Xuelong Li, Mulin Chen, Feiping Nie, and Qi Wang. Locality adaptive discriminant analysis. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 2201–2207. AAAI Press, 2017.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of machine learning research*, 9:2579–2605, 2008.

H Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, et al. Communication-efficient learning of deep networks from decentralized data. *arXiv preprint*, page arXiv:1602.05629, 2016.

Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

Karl Pearson. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11): 559–572, 1901.

F. Samaria and A. Harter. Parameterisation of a stochastic model for human face identification. In *Proceeding of IEEE Workshop on Applications of Computer Vision*, 1994.

Masashi Sugiyama. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. *Journal of machine learning research*, 8(May):1027–1061, 2007.

Masashi Sugiyama, Taiji Suzuki, Shinichi Nakajima, Hisashi Kashima, Paul von Bünau, and Motoaki Kawanabe. Direct importance estimation for covariate shift adaptation. *Annals of the Institute of Statistical Mathematics*, 60(4):699–746, 2008.

Yuta Takahashi, Han ten Chang, Akie Nakai, Rina Kagawa, Hiroyasu Ando, Akira Imakura, Yukihiko Okada, Hideo Tsurushima, Kenji Suzuki, and Tetsuya Sakurai. Decentralized learning with virtual patients for medical diagnosis of diabetes. *SN Computer Science*, 2, 2021.

Lionel Tarassenko, Paul Hayton, Nicholas Cerneaz, and Michael Brady. Novelty detection for the identification of masses in mammograms. 1995.

Qiang Yang. GDPR, data shortage and AI, 2019. URL https://aaai.org/Conferences/AAAI-19/invited-speakers/. Invited Talk of The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19).

Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, 10 (2):Article 12, 2019.

Xiucai Ye, Hongmin Li, Akira Imakura, and Tetsuya Sakurai. Distributed collaborative feature selection based on intermediate representation. In *The 28th International Joint Conference on Artificial Intelligence (IJCAI-19)*, pages 4142–4149, 2019.

Hiroyuki Yoda, Akira Imakura, Momo Matsuda, Xiucai Ye, and Tetsuya Sakurai. Novelty detection in multimodal datasets based on least square probabilistic analysis. *International Journal of Machine Learning and Computing*, 10(4), 2020.

Lihi Zelnik-Manor and Pietro Perona. Self-tuning spectral clustering. In *Advances in neural information processing systems*, pages 1601–1608, 2005.

Yang Zhang, Nirvana Meratnia, and Paul Havinga. Outlier detection techniques for wireless sensor networks: A survey. *IEEE communications surveys & tutorials*, 12(2):159–170, 2010.