Secure distribution of Factor Analysis of Mixed Data (FAMD) and its application to personalized medicine of transplanted patients

Sirine Sayadi, Estelle Geffard, Mario Südholt, Nicolas Vince, Pierre-Antoine Gourraud

Secure Distribution of Factor Analysis of Mixed Data (FAMD) and Its Application to Personalized Medicine of Transplanted Patients

Sirine Sayadi[1,2(✉)], Estelle Geffard[2], Mario Südholt[1], Nicolas Vince[2], and Pierre-Antoine Gourraud[2]

[1] STACK Team, IMT Atlantique, Inria, LS2N, Nantes, France
{Sirine.Sayadi,Mario.Sudholt}@imt-atlantique.fr
[2] University of Nantes, Nantes University Hospital, INSERM, Research Center in Transplantation and Immunology, UMR 1064, ATIP-Avenir, Nantes, France
Estelle.Geffard@etu.univ-nantes.fr,
{Nicolas.Vince,Pierre-Antoine.Gourraud}@univ-nantes.fr

Abstract. Factor analysis of mixed data (FAMD) is an important statistical technique that not only enables the visualization of large data but also helps to select subgroups of relevant information for a given patient. While such analyses are well-known in the medical domain, they have to satisfy new data governance constraints if reference data is distributed, notably in the context of large consortia developing the coming generation of personalised medicine analyses.

In this paper we motivate the use of distributed implementations for FAMD analyses in the context of the development of a personalised medicine application called KITAPP. We present a new distribution method for FAMD and evaluate its implementation in a multi-site setting based on real data. Finally we study how individual reference data is used to substantiate decision making, while enforcing a high level of usage control and data privacy for patients.

1 Introduction

Big data analysis techniques are increasingly popular to extract new information from massive amounts of data to improve decision making, notably in the medical sector. A major challenge for clinicians consists in safely making correct treatment decisions based on ever growing amounts of patient data. This problem calls for new analysis techniques and algorithms, in particular, for precision medicine.

Precision Medicine, that is using genetic or molecular profiling for optimizing care of small patient groups, will probably become the standard of care in the next decades. It represents a deep revolution in health care also because not only patients will be covered but also healthy individuals. For this (r)evolution of medicine to become reality, it is necessary to deliver the evidence of the efficiency and cost-effectiveness of precision medicine. This, in turn, requires decisions concerning patients to be substantiated

L. Barolli et al. (Eds.): AINA 2021, LNNS 225, pp. 507–518, 2021.
https://doi.org/10.1007/978-3-030-75100-5_44

by an analysis of large-scale reference data that is relevant for their personal situation compared to others [2].

Dimension reduction is a major technique for transforming large multi-dimensional data spaces into a lower dimensional subspaces. This is while preserving significant characteristics of the original data. Among dimension reduction methods, the most common method is Principal Component Analysis (PCA) [3], which enables dimension reduction for quantitative data variables. Other methods are Factor Analysis of Mixed Data (FAMD) [5], which performs dimension reduction for mixed (quantitative and qualitative) data variables, and dictionary learning (DL) [4] one of the most powerful methods of extracting features from data.

FAMD analysis provides simplified representations of multi-dimensional data spaces in the form of a point cloud within a vector subspace of principal components. If two points are close to each other in this cloud, a strong global similarity exists between them with respect to the selected principal components. In the biomedical field, this kind of analysis is frequently used to present patients groups in a simplified and visual way for a large range of complex clinical data encompassing quantitative data (for example obtained from biological exams) and qualitative data (for example gender information). The result are actionable representations of each patient's individual characteristics compared to those of others.

In the context of a French public-private partnership KTD-innov and a H2020 EU project EU-train, FAMD has been used for dimension reduction as part of the clinical decision support system KiTApp (the kidney transplant application) [6]. This precision medicine web-application computes predictive scores and represents distributions of patients' variables in a subgroup of reference patients after kidney transplantation. The application is conceived to relay the intuition and experience of clinicians by means of on-demand computations and graphical representations. One of KiTApp's key functionalities consists in the "contextualization" of patients relative of a population of reference (POR). To this end, it first uses FAMD for dimension reduction, then applies a percentile statistical modelling [21] algorithm, and visualizes the relations of patients to the POR.

Medical studies often involve large national or international collaborations (such as our KTD-innov and EU-Train projects). Simple centralization schemes for the placement of data and computations are frequently not applicable in this context because data and computations may not be shared due to legal reasons, security/privacy concerns and performance issues). To deploy this kind of medical services in larger contexts distributed systems and algorithms for precision medicine have to be provided. One of the main lines of research around the KTD-innov and EU-TRAIN projects is the implementation of a reference database integrated into a distributed computing infrastructure allowing a secure access to data while respecting the European GPDR data protection regulation [12]. However, very few distributed algorithms have been developed for and applied to the domain of precision medicine.

Data sharing and analysis placement are generally difficult due to governance, regulatory, scientific and technical reasons. Analyses are often only possible "on premise." Furthermore, researchers and institutions may be averse to lose control over both data usage. In addition, huge volumes of data are intrinsically difficult to share or transfer,

notably because of cost of the use of computational, storage and network resources.

On the other hand distributed architectures enable more flexible data governance strategies and analysis processes by freeing them from centralization constraints [9]. Decentralized databases enable performing local calculations on patient data, without any individual data circulating outside the clinical centers generating the data. To this end, one may strive for distributed statistical calculations. Fully-distributed analyses have been proposed, see Sect. 2, for contextualizing the state of a patient relative to POR data stored in a distributed database. Such algorithms have to meet requirements of scalability, security and confidentiality [10], as well as availability properties and right to privacy properties [11]. These criteria are difficult to satisfy, however, because the statistical significance and accuracy of analyses often directly depend on the number of cases or individuals included in the database.

A solution to these problems can be based on harnessing distributed analyses that manipulate sensitive data on the premises of their respective owners and harness distributed computations if non sensitive, aggregated, summarized or anonymized data is involved.

In this paper, we present two main contributions:

- We motivate and define requirements for distributed algorithms for dimension reduction in the context of the KITAPP project.
- We present a novel distributed FAMD algorithm for dimension reduction in the presence of sensitive data in precision medicine and apply it to a contextualization problem.

The rest of this article is organized as follows. Section 2 presents related works. Section 3 presents the kidney transplantation application (KITAPP) and its use of FAMD dimension reduction. Section 4 presents our distributed algorithm FAMD and a corresponding implementation. Section 5 provides experimental results, notably concerning privacy requirements, and a performance evaluation. Finally, Sect. 6 summarizes our findings and proposes some future work.

2 Related Work

Parallel versions of PCA dimension reduction algorithms have already been proposed. Liang *et al.* [14] propose a client-server system and send singular vectors and singular values $U\Sigma V$ from the client to the server. Feldman *et al.* [15] have shown how to compute PCAs by sending smaller matrices $U\Sigma$ instead of sending all matrices of a singular value decomposition, thus improving on the communication cost. Wu *et al.* [16] have introduced an algorithm that improves on the storage and data processing requirements and harnesses Cloud computing for PCA dimension reduction. These proposals send matrices of synthesized data of the original data and not real data. This is very interesting for biomedical analyses in order to ensure privacy of patient data. Imtiaz *et al.* [17] have improved Feldman *et al.s'* proposal by adding privacy guarantees using differential privacy.

To the best of our knowledge, no distributed FAMD algorithm has already been proposed. In this paper, we propose a distributed FAMD on the basis of a distributed PCA algorithm. Our algorithm is structured into two parts (similar to Pagès [5]):

1. Transform qualitative data into quantitative one using complete disjunctive tables [13], thus transforming the original FAMD dimension reduction problem into a PCA one.
2. Perform the dimension reduction of the distributed PCA based on Feldman *et al. s'* proposal [15].

3 The Kidney Transplantation Application (KiTapp)

Chronic kidney failure affects approximately 10% of the world population and can progressively lead to end-stage kidney disease requiring replacement therapy (dialysis or transplantation). Kidney transplantation is the best treatment for end-stage kidney disease [19].

3.1 KiTapp Overview

Data from approximately 1500 renal transplantation, including clinical and immunological items, were collected since 2008 as part of a French national project.

KiTapp enables personalized contextualization algorithm to be harnessed to compare data trajectories of a given patient (POI) to a sub-population with similar characteristics (POR) selected by filters or distance measures. The information relative to a graft is selected from similar cases at the time of the graft. With the help of clinicians and knowledge of the existing body of research, we defined a set of variables to select the sub-population of reference.

We propose three population contextualization algorithms: compare a given patient's data to PORs with

1. similar characteristics selected by filters or approaches based on statistical analysis;
2. the nearest neighbor method or
3. the cluster method.

With our filter approach (1), the POR is defined according to selected filters made available to the clinician, such as age, gender and Body Mass Index (BMI). Methods (2) and (3) are based on the results of an FAMD. Following this analysis, we can then select a POR by close neighbor method (2): by selecting the N individuals most similar to a POI or we can select a POR by clustering (3): by selecting the individuals in the same cluster as our POI.

The visualization of contextualized information is done by comparing a POI's biological data (creatinemia) and its evolution over time post-transplantation (clinical visits) to a POR that is represented by their median and percentile values.

3.2 Motivation for Distributed Analyses for KITAPP

We intend to harness the KITAPP application as part of large-scale cooperations with many (national and international) partners. To meet the challenge of harnessing medical data while keeping sensitive data on premise or ensure strong data protection if data is moved, computations are often performed today over distributed databases that are linked to a computation integrator that enables a center to interact with and access some data from remote sites. Each clinical center collects, stores and controls their own patients' data. The founding principle of the architecture is that no data of individuals circulates outside the centers. However, this sharing paradigm is very restrictive and inhibits a large range of potential analyses to be performed - either because sensitive data cannot be appropriately protected or the analysis cannot be performed sufficiently efficient.

The need for local storage and distribution of reference data is motivated by the actionable value and publication value. It provides the possibility of controlling locally who has accessed to data, what are the usages of the local data and how to limit then should it be needed. The use of distributed infrastructure is a central element of multi-stakeholders data governance.

We are therefore working on more general distributed analysis architectures and implementations that ease collaboration as part of multi-centric research projects, where each center can control and account for their own patients' data usage even if located remotely. Contextualization then has to be performed with respect to large-scale distributed medical databases that are maintained at different sites.

4 Distributed FAMD

In the following we first provide an overview over the architecture and properties of our algorithm before defining it in detail.œ

4.1 Overall Architecture and Properties

Factor analysis of mixed data (FAMD) [5] is a method of dimension reduction of variables including mixed quantitative and qualitative data into fewer components for information synthesis reasons. This analysis can defined, for instance using matrix operations, as follows:

$$FAMD = PCA + MCA \tag{1}$$

where PCA is a principal component analysis dimension reduction for quantitative variables and MCA is a multiple correspondence analysis dimension reduction for qualitative variables.

Overall, our algorithm works as follows. As a first step, we transform the qualitative variables into quantitative ones using complete disjunctive coding (CDC) [20] that is performed locally on each site. In a second step we perform dimension reduction by means of a distributed PCA in order to obtain a secure and distributed FAMD algorithm.

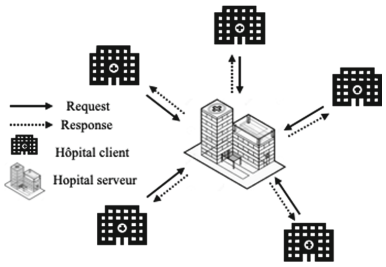

Fig. 1. Collaboration architecture

We harness the distributed cooperation architecture shown in Fig. 1. Data transformation and dimension reduction analysis are performed locally at multiple sites separately. The coordination between the sites is done by an aggregator site, which receives synthetic data and performs the overall dimension reduction. Imtiaz *et al.* [17] have proposed a secure and distributed algorithm for PCA dimension reduction. This algorithm uses differential privacy as a security technique and synthetic data for communication between nodes.

The resulting algorithm has two important properties :

- *Low communication cost:* The communication cost of parallel and distributed FAMD algorithms essentially depends on the size of the matrices transferred between sites. Many dimension reduction algorithms require matrices of size $D \times D$ to be sent, where D is the number of data items to be analyzed (that is, transplantation data in our case). In contrast, Imtiaz *et al.*' [17] algorithm requires matrices to be sent of type $D \times R$ where R is the number of variables and (typically) $R \ll D$.
- *Security/privacy awareness:* Our algorithm satisfies two interesting characteristics: (1) differential privacy is used for data protection and (2) communication between the sites and the aggregator involves only synthesized data P_S and not the original data, which minimizes possibilities of data theft and also supports data protection.

We harness the same principle and properties while providing two new contributions: (1) a transformation of qualitative variables into quantitative variables to obtain a secure and distributed algorithm FAMD dimension reduction and (2) a scalable distributed implementation and evaluated it by analyzing real-world biomedical data on a realistic grid environment.

4.2 Algorithm Definition

Algorithm 1 presents our secure and distributed FAMD algorithm. Lines 1–16 implement the first step, the transformation of a full FAMD problem into a (qualitative) PCA problem. Each site begins by calculating, for each quantitative variable, the corresponding mean μ_k, standard deviation σ_k and Centering and Reduction Function $X_{i,k} = \frac{1}{\sigma_k}(x_{i,k} - \mu_k)$. For each qualitative variable, Complete Disjunctive Coding using

Algorithm 1: Distributed FAMD Algorithm

Input : Data matrix $X_s \in \mathbb{R}^{D \times N_s}$ *for* $s \in [S]$ of N elements and P variables, with C quantitative variables and M qualitative variables; ε, δ: privacy parameters; j : reduced dimension;

Output: V_j: Matrix of eigenvectors on top j

1 **foreach** *site* $s \in S$ **do**
2 **foreach** *element* $i \in N$ **do**
3 **foreach** *element* $k \in C$ **do**
4 Compute the mean of the variable μ_k;
5 Compute the standard deviation of the variable σ_k;
6 Compute the Centering and Reduction Function $X_{i,k} = \frac{1}{\sigma_k}(x_{i,k} - \mu_k)$;
7 **return** $X_{i,k}$;
8 **end**
9 **foreach** *element* $k \in M$ **do**
10 Apply the Complete Disjunctive Coding using (ade4 package on R);
11 Compute the effective of the modality N_k;
12 Compute the proportion $p_k = N_k/N$;
13 Compute the Indicator Weighting Function $X_{i,k} = \frac{x_{i,k}}{\sqrt{p,k}}$;
14 **return** $X_{i,k}$;
15 **end**
16 **end**
17 Compute $A_s = \frac{1}{N_s} X_s X_s^T$;
18 Generate $D \times D$ symmetric Matrix E where $E_{i,j} : i \in [D], j \leq i$ drawn i.i.d. from $N(0, \Delta^2_{\varepsilon,\delta})$ where $\Delta_{\varepsilon,\delta} = \frac{1}{Ns_\varepsilon}\sqrt{2log(\frac{1.25}{\delta})}$, $E_{i,j} = E_{j,i}$;
19 Compute $A_s = A_s + E$;
20 Perform $SVD(A_s) = U\Sigma U^T$;
21 Compute $P_s = U\Sigma^{1/2}$;
22 Send P_s to the aggregator;
23 **end**
24 Compute $A = \frac{1}{s}\sum_{s=1}^{s} P_s P_s^T$;
25 Perform $SVD(A) = V\Lambda V^T$;
26 Send V_j to all sites ;
27 **return** V_j;

the ade4 package from the R language is then applied in order to transform the qualitative variables into a quantitative variable, followed by the computation of the modalities N_k and proportions $p_k = N_k/N$ to compute the indicator weighting function $X_{i,k}$.

The second step, the dimension reduction proper, is implemented on lines 17–23. Each site calculates the (second moment) matrix $A_s = \frac{1}{NS} X_s X_s^T$. The application of the scheme of differential privacy (following Dwork *et al.*' proposal [18]) is performed by generating the noise matrix E of size $D \times D$ and the estimated differential privacy matrix $A_s = A_s + E$ on line 19. Each site then performs the Singular Value Decomposition $SVD(A_s)$) of matrix A_s to compute the matrix $(P_s = U\Sigma^{1/2})$ and broadcast it to the aggregator.

At the aggregator site, the server computes, see lines 24–26, the matrix $A = \frac{1}{s}\sum_{s=1}^{s} P_s P_s^T$ of all sites. It performs next the global Singular Value Decomposition $SVD(A) = V\Lambda V^T$.

5 Experimentation

In this section we report on experiments involving analyses over real medical data that we have carried out on a real heterogeneous large-scale grid infrastructure. We report on our setup, and evaluate our implementation w.r.t. three criteria.

5.1 Setup

Our experiments have been carried out on renal transplantation data available in the European database Divat [22]. In order to compare with results from the KITAPP project, we have applied our distributed algorithm to its analyses on 11,163 transplantation data. We started by divided the data file before transfer and analysis on the different sites.

We have implemented our distributed algorithm and executed it in a grid-based environment featuring different distributed architectures, ranging from placing all clients on different (geo-distributed) machines to placing them as one cluster on only one machine. This distributed environment constitutes a realistic architecture of a medical collaboration involving the research and clinical centers, the partners of the KITAPP project. We have implemented our distributed algorithm using the Python and R programming language using 860 lines of code. The whole distributed system can be deployed and executed on an arbitrary number of sites of the Grid'5000 infrastructure using a small script of only eight commands.

The Grid'5000 platform is a platform, built from eight clusters in two European countries for research in the field of large-scale distributed systems and high performance computing. For our experiment, we have reserved a machine as a server (aggregator) executing a Python program to manage the analysis, client interactions and generation of the final result. To create a number of client sites we have reserved machines distributed over five different sites in France.

5.2 Results and Performance Evaluation

The KITAPP-motivated FAMD dimension reduction analysis we employed as a test case has been executed on the basis of 11,163 transplantation operations characterized using 27 qualitative and quantitative variables distributed over five sites. We have set the dimension reduction parameter j on the server to two. Figure 2 shows the resulting two-dimensional subspace after application of our distributed FAMD analysis.

In order to distribute the POR selection for POI contextualization between sites, we have applied the K-means unsupervised clustering technique to the result of the distributed FAMD dimension reduction analysis. FAMD and clustering enables grouping of patient data according to their similarity and proximity relative to principal components. Figure 3 presents the result of k-means clustering, three independent data clusters

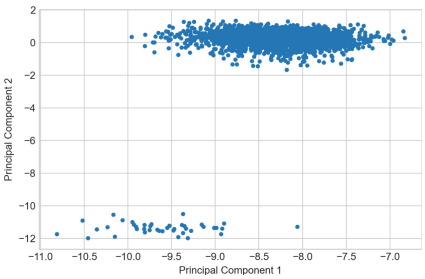

Fig. 2. Distributed FAMD for 5 sites.

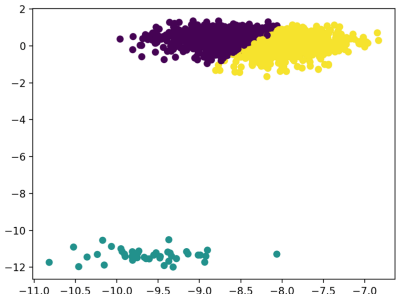

Fig. 3. Clustering of distributed FAMD.

that correspond exactly to the result of the (centralized) sequential algorithm that is used as part of the KITAPP project.

Each cluster is characterized by specific variables combinations. The green cluster corresponds to living donors. The yellow cluster corresponds to deceased donors and the purple one to deceased donors with expanded criteria, such as aged $we > 50$ years or subject to hypertension or creatinine levels $\geq 133\,\mu$mol/L. Note that we always obtain the same clusters independent from the number of sites that participate in the distributed FAMD analysis if we operate it with the same data, which shows a strong scalability potential of our proposed algorithm.

In the following we evaluate three properties of our implementation: (i) the quality of the reduction technique in the presence of noise introduced by the differential privacy technique using a notion of captured energy, (2) execution time and (3) communication cost. For evaluation purposes we consider three architectures: our distributed FAMD reduction technique (denoted"DPdis" below), a more centralized version where W where all the second moment matrix A_s of each clients are aggregated at the server (denoted "fulldis"), and a fully-centralized FAMD version (denoted "pooled").

Captured Energy/Utility. Following Imtiaz *et al.* the captured energy q is used to evaluate the quality of V_j principal directions based on the difference in information utility between the case where all data is centralized q_{pooled}, all second moment matrix A_s of each site are distributed $q_{fulldis}$ and secure distributed proposed FAMD algorithm q_{DPdis} by data size and number of sites.

The captured energy is defined as the matrix multiplication $q = tr(V_j(A)^T A V_j(A)$ measuring the amount of optimal eigenvalues captured in the subspace FAMD. For any other sub-optimal sub-spaces, the value would be less than the optimal value.

– *Energy per site.* We have varied the number of S sites that participate in this analysis by keeping the total number of samples $N = 11163$ (i.e. we decreased the size N_s of each site). Figure 4 shows a deterioration in the performance of $q_{fulldis}$ and q_{DPdis} for an increasing number of sites. This decrease in performance is explained by the decrease in the number of elements per site. In addition, the presence of high variance noise degrades the number of eigen-directions stronger than the noise which

are detected by the PCA instead of capturing all the j directions which present the data.

– *Energy by data size.* Figure 5 shows an increase in performance of captured energy q as a function of the elements number per site. $q_{fulldis}$ and q_{DPdis} have almost the same performance in captured energy for the two variation of sites number and data size, q_{pooled} always keeps better performance.

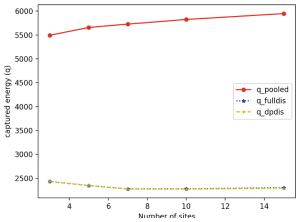

Fig. 4. Captured energy (q) by sites number.

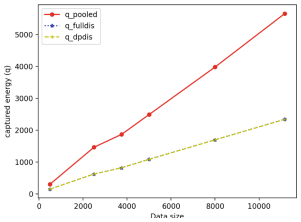

Fig. 5. Captured energy (q) by data size.

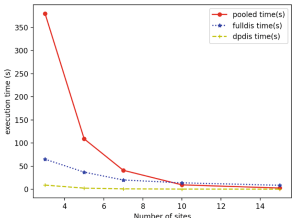

Fig. 6. Execution time by sites number.

Execution Time. We have varied the number of S sites by keeping the global number of samples $N = 11163$ (i.e. we decreased the size N_s of each site) and we have measured the execution time. Figure 6 shows that execution time decreases with increasing sites number. Our proposed approach $dpdis$ always keeps the least execution time. This is due to the Lower communication cost explained in 4.1 section.

Communication Cost/Data Sharing: The lower cost of communication of our proposed algorithm introduced in the previous section allows for a minimum sharing of data (matrices Ps are shared and not A_s). In the case of our experiments with distribution on 5 sites with global samples equal to $11,163$ and 27 features, the quantity of data shared by all clients is equal to $11,163 \times 27 = 301.401$ values instead of sending $sqr(11,163) = 124.612.569$ values.

6 Conclusion and Future Work

FAMD dimension reduction is an important tool for transforming complex data into lower-dimensional sub-spaces while preserving important characteristics of the original data. This technique is generally useful to reduce complexity and support decision making. In this paper, we have motivated the use of dimension reduction for geo-distributed biomedical collaborations that require distributed models and implementations of biomedical algorithms with distributed implementation. Its evaluation on a real geo-distributed grid infrastructure using real data has valided its efficiency, scalability, privacy protection properties.

As future work, we will focus on extensions of federated learning as a more general method for the definition of secure and distributed biomedical analyses.

References

1. Hood, L.: Systems biology and p4 medicine: past, present, and future. Rambam Maimonides Med. J. (2013)
2. Gourraud, P., Henry, R., et al.: Precision medicine in chronic disease management: the multiple sclerosis bioscreen. Ann. Neurol. **76**(5), 633–642 (2014)
3. Jolliffe, I.T.: Principal Component Analysis. Springer Series in Stat. Springer (2002)
4. Shakeri, Z., Sarwate, A.D., Bajwa, W.U.: Sample complexity bounds for dictionary learning from vector and tensor valued data. In: Rodrigues, M., Eldar, Y. (eds.) Information Theoretic Methods in Data Science, Chapter 5. Cambridge University Press (2019)
5. Pagès, J.: Multiple Factor Analysis by Example using R. Chapter 3 (2014)
6. Herve, C., Vince, N., et al.: P218 The kidney transplantation application (KITAPP): a visualization and contextualization tool in a kidney graft patients' cohort. Hum. Immunol. **78**, 216 (2017)
7. KTD-Innov. www.ktdinnov.fr. Accessed 07 Jan 2021
8. EU-TRAIN. `eu-train-project.eu` . Accessed 07 Jan 2021
9. Brous, P., Janssen, M., et al.: Coordinating decision-making in data management activities: a systematic review of data governance principles. In: International Conference on Electronic Government. Springer (2016)
10. Boujdad, F., Gaignard, A., et al.: On Distributed Collaboration for Biomedical Analyses WS CCGrid-Life (2019)
11. Scheel, H., Dathe, H., Franke, T., Scharfe, T., Rottmann, T.: A privacy preserving approach to feasibility analyses on distributed data sources in biomedical research. Stud. Health Technol. Inform. **267**, 254–261 (2019)
12. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). http://data.europa.eu/eli/reg/2016/679/oj. Accessed 07 Jan 2021
13. Greenacre, M., Blasius, J.: Multiple Correspondence Analysis and Related Methods. Chapman and Hall/CRC Press, London (2006)
14. Liang, Y., Balcan, M., Kanchanapally, Y.: Distributed PCA and k-Means Clustering (2013)
15. Feldman D., Schmidt, M., Sohler, C.: Turning big data into tiny data: constant-size coresets for K-means, PCA and projective clustering. In: Proceedings of the Twenty-fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, pp. 1434–1453 (2013)
16. Wu, Z., Member, Li, Y., Plaza, A., Li, J., Xiao, F., Wei, Z.: Parallel and distributed dimensionality reduction of hyperspectral data on cloud computing architectures. IEEE J. Select. Top. Appl. Earth Obser. Remote Sens. **9**, 1–9 (2016)
17. Imtiaz, H., Sarwate, A.: Differentially private distributed principal component analysis. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2206–2210 (2018)
18. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Halevi, S., Rabin, T. (eds) Theory of Cryptography. TCC 2006. Lecture Notes in Computer Science, vol. 3876. Springer, Berlin, Heidelberg (2006)
19. Hill, N., Fatoba, S., et al.: Global prevalence of chronic kidney disease - a systematic review and meta-analysis. PLoS One (2016)
20. Mellinger, M.: Correspondence analysis in the study of lithogeochemical data: general strategy and the usefulness of various data-coding schemes. J. Geochem. Explor. **21**(1–3), 455–469 (1984). ISSN 0375-6742
21. Sayadi, S., Geffard, E., Südholt, M., Vince, N., Gourraud, P.: Distributed contextualization of biomedical data: a case study in precision medicine. In: AICCSA 2020 - 17th IEEE/ACS International Conference on Computer Systems and Applications, pp. 1–6, November 2020

22. Divatfrance. www.divat.fr. Accessed 07 Jan 2021
23. Balouek, D., Carpen Amarie, A., Charrier, G., et al.: Adding virtualization capabilities to the grid'5000 testbed. In: Cloud Computing and Services Science. Springer (2013)