2015 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 17–20, 2015, BOSTON, USA

LARGE SCALE COLLABORATION WITH AUTONOMY: DECENTRALIZED DATA ICA

Bradley T. Baker[⋆] *Rogers F. Silva*[†] *Vince D. Calhoun*[†] *Anand D. Sarwate*[‡] *Sergey M. Plis*[†]

[⋆] New College of Florida [†] Mind Research Network [‡] Rutgers University

ABSTRACT

Data sharing for collaborative research systems may not be able to use contemporary architectures that collect and store data in centralized data centers. Research groups often wish to control their data locally but are willing to share access to it for collaborations. This may stem from research culture as well as privacy concerns. To leverage the potential of these aggregated larger data sets, we would like tools that perform joint analyses without transmitting the data. Ideally, these analyses would have similar performance and ease of use as current team-based research structures. In this paper we design, implement, and evaluate a decentralized data independent component analysis (ICA) that meets these criteria. We validate our method on temporal ICA for functional magnetic resonance imaging (fMRI) data; this method shares only intermediate statistics and may be amenable to further privacy protections via differential privacy.

Index Terms— multi-site collaboration, decentralized data, ICA

1. INTRODUCTION

Research groups studying complex phenomena (such as certain diseases) often focus on specific questions but gather data that could be used to answer questions beyond the scope of the original study. For example, a mental health study may collect a brain scan using magnetic resonance imaging (MRI) from all enrolled subjects but may only examine a particular aspect of the MRI data. The whole scan is saved as part of the data set associated with that study and could therefore be used in other studies. Technological advances have dramatically increased the complexity of data per measurement while lowering the cost. Researchers hope to leverage data across multiple research groups to achieve sufficiently large sample sizes that may uncover important, relevant, and interpretable features that characterize the underlying complex phenomenon. Many research communities have proposed collaborative research systems to help enable such joint analyses.

The standard industry solution to data sharing involves each group uploading data to a shared-use data center such as a cloud-based service. This solution is not possible for many research applications. For example, since neuroimaging data is from human subjects, data sharing may be limited or prohibited due to issues such as *(i)* local administrative rules, *(ii)* local desire to retain control over the data until a specific project has reached completion, *(iii)* ethical concerns of data re-identification. The last point is particularly acute in scenarios involving genetic information, patient groups with rare diseases, and other identity-sensitive applications.

Recently, two data sharing architectures have emerged: centralized data sharing via a repository, and case-specific, agreement-based collaborations. In the latter case, even estimating simple statistical metrics from multiple sites currently requires significant manual labor. For example, in contemporary neuroscience research, we could find no automated approaches that can use modern computational techniques to enable collaborative and interactive feature-learning from multiple sites.

In this paper, we take a step in the development of algorithms for decentralized feature learning, by designing a decentralized data independent component analysis (ICA) [1] algorithm, a widely-used method in neuroimaging applications. Specifically, we design a decentralized temporal ICA algorithm for use with functional magnetic resonance imaging (fMRI) data. In resting-state fMRI studies, we can assume that the spatial maps remain stable across subjects and experiment duration [2], while activation of certain neurological regions varies over time and across subjects. Temporal ICA finds temporally independent components, hopefully corresponding to the subjects' intrinsic common spatial networks. Temporal ICA typically requires more data than can be produced by a single research group because of computational complexity as well as statistical sample size; the ratio of spatial to temporal dimensions often requires the aggregate temporal dimension to be similar to the voxel dimension.

To overcome these challenges, we suggest an approach that allows for the computation of aggregate spatial maps and local independent time courses across decentralized data stored at different servers belonging to independent labs. The approach combines individual computation performed locally with global process to obtain both local and global results. Our approach produces results with a similar performance to the pooled-data case.

This work was sponsored in part by NSF award CCF-1453432 and NIH award 1R01DA040487-01A1.

978-1-4673-7454-5/15/$31.00 ©2015 IEEE

2. METHODS

In this section, we cover some preliminary concepts and describe our new algorithm, *decentralized joint ICA* (djICA). Let $\mathbf{1}$ denote a column vector whose entries are all equal to 1.

ICA Model. ICA is a popular method for blind source separation: it attempts to decompose mixed signals into underlying sources. Empirically, ICA applied to brain imaging data produces robust features which are physiologically interpretable and markedly reproducible across studies. While justification for successful ICA of fMRI results had been previously attributed to sparsity [3], it has been shown [4] that statistical independence between the underlying sources is in fact the driving mechanism of ICA algorithms. In linear ICA, we model a data matrix $\mathbf{X} \in \mathbb{R}^{d \times N}$ as a product $\mathbf{X} \approx \mathbf{AS}$, where $\mathbf{S} \in \mathbb{R}^{r \times N}$ is composed of N observations from r *statistically independent* components, each representing an underlying signal source.

We interpret ICA in terms of a generative model in which the independent sources $\mathbf{S}$ are submitted to a linear mixing process described by a mixing matrix $\mathbf{A} \in \mathbb{R}^{d \times r}$, forming the observed data $\mathbf{X}$. Most algorithms attempt to recover the "unmixing matrix" $\mathbf{W} = \mathbf{A}^{-1}$ assuming the matrix $\mathbf{A}$ is invertible. They do this by trying to maximize independence between rows of the product $\mathbf{WX}$. Maximal information transfer (Infomax) is a popular heuristic for estimating $\mathbf{W}$ that results in maximizing an entropy functional related to $\mathbf{WX}$. More precisely, with some abuse of notation, let

$$g(z) = \frac{1}{1 + e^{-z}} \tag{1}$$

be the sigmoid function with $g(\mathbf{Z})$ being the result of elementwise application of $g(\cdot)$ on the entries of a matrix or vector $\mathbf{Z}$. The entropy of a random vector Z with joint density p is

$$h(Z) = -\int p(Z) \log p(Z) dZ. \tag{2}$$

The objective of Infomax ICA then becomes

$$\widehat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmax}}\, h(g(\mathbf{WX})). \tag{3}$$

As we will show, Infomax ICA is well-suited to decentralization. However, note that many other approaches for (centalized) ICA exist, e.g., FastICA [5], which has faster yet less robust convergence, and the RADICAL algorithm [6], which performs well on artifact removal for certain modalities of biomedical data [7]. It would be interesting to see how amenable these approaches are to decentralized implementation; we leave this for future work.

Decentralized Joint ICA. Our goal is to design an ICA algorithm that can be applied to decentralized data. Currently, a number of extensions of ICA exist for the purpose of joining together various data sets [8] and performing simultaneous decomposition of data from a number of subjects and modalities [9]. Group spatial ICA (GICA) stands out as the leading approach for multi-subject analysis of task- and resting-state fMRI data [10], building on the assumption that the spatial map components ($\mathbf{S}$) are common (or at least similar) across subjects. Another approach, called joint ICA (jICA) [11], is popular in the field on multimodal data fusion and assumes instead that the mixing process ($\mathbf{A}$) over a group of subjects is common between a pair of data modalities. A largely unexplored area of fMRI research is group temporal ICA, which, like spatial ICA, also assumes common spatial maps but pursues statistical independence of timecourses instead. Consequently, like jICA, the common spatial maps from temporal ICA describe a common mixing process ($\mathbf{A}$) among subjects. While very interesting, temporal ICA of fMRI is typically not investigated because of the small number of time points in each data set, which leads to unreliable estimates. Our decentralized jICA approach overcomes that limitation by leveraging information from data sets distributed over multiple sites.

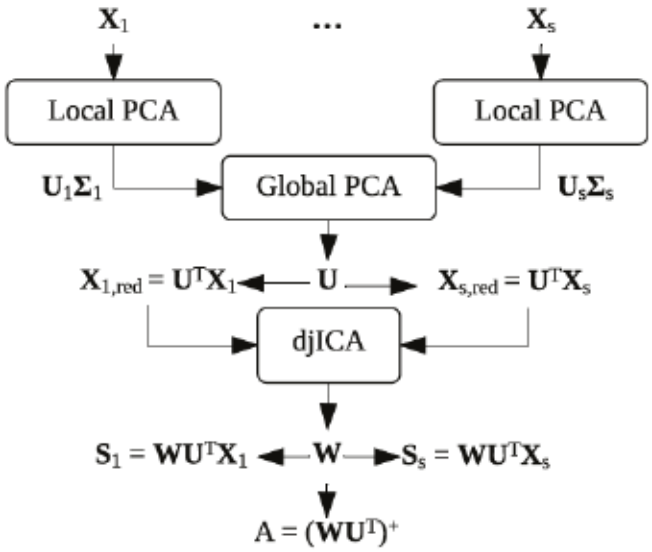

Fig. 1. djICA algorithm overview. The superscript '+' indicates the pseudo inverse.

Suppose that we have s total sites; each site i has a data matrix $\mathbf{X}_i \in \mathbb{R}^{d \times N_i}$ consisting of a total time course of length N_i time points over d voxels. Let $N = \sum_{i=1}^{s} N_i$ be the total length. We model the data at each site as coming from a common (global) mixing matrix $\mathbf{A} \in \mathbb{R}^{d \times r}$ applied to local data sources $\mathbf{S}_i \in \mathbb{R}^{r \times N_i}$. Thus, the total model can be written as

$$\mathbf{X} = [\mathbf{AS}_1\ \mathbf{AS}_2\ \cdots\ \mathbf{AS}_s] \in \mathbb{R}^{d \times N}. \tag{4}$$

We design a new algorithm, decentralized joint ICA (djICA), that uses locally computed gradients to estimate a common, global unmixing matrix $\mathbf{W} \in \mathbb{R}^{r \times d}$ corresponding to the Moore-Penrose pseudo-inverse of $\mathbf{A}$ in (4), denoted $\mathbf{A}^+$.

Fig. 1 summarizes the overall algorithm in the context of

temporal ICA for fMRI data. Each site i has data matrices $\mathbf{X}_{i,m} \in \mathbb{R}^{d \times n_i}$ corresponding to subjects $m = 1, 2, \ldots, M_i$ with d voxels and n_i time samples. Sites concatenate their local data matrices temporally to form a $d \times n_i M_i$ data matrix $\mathbf{X}_i$, so $N_i = n_i M_i$. In a two-step distributed principal component analysis (dPCA) framework, each site performs local PCA (Algorithm 2) by means of singular value decomposition (SVD), with matrices $\mathbf{U}_i \in \mathbb{R}^{d \times k}$ and $\mathbf{\Sigma}_i \in \mathbb{R}^{k \times k}$ corresponding to the top singular vectors and values, respectively. The sites then compute a global PCA (Algorithm 3) to form a common projection matrix $\mathbf{U} \in \mathbb{R}^{d \times r}$. Alternatively, in a one-step dPCA framework, we can compute the global $\mathbf{U}$ directly but at the expense of communicating a large $d \times d$ matrix between sites. Finally, all sites project their data onto the subspace corresponding to $\mathbf{U}$ to obtain reduced datasets $\mathbf{X}_{i,\text{red}} \in \mathbb{R}^{r \times N_i}$. The projected data is the input to the iterative djICA algorithm that estimates the unmixing matrix $\mathbf{W} \in \mathbb{R}^{r \times r}$ as described in Algorithm 1. The full mixing matrix for the global data is modeled as $\mathbf{A} \approx (\mathbf{W}\mathbf{U}^\top)^+ \in \mathbb{R}^{d \times r}$.

After initializing $\mathbf{W}$ (for example, as the identity matrix), the djICA algorithm iteratively updates $\mathbf{W}$ using a distributed natural gradient descent procedure [12]. At each iteration j the sites update locally: in lines 4 and 5, the sites adjust the local source estimates $\mathbf{Z}_i$ by the bias estimates $\mathbf{b}(j-1)\mathbf{1}^\top \in \mathbb{R}^{r \times N_i}$, followed by the sigmoid transformation $g(\cdot)$; they then calculate local gradients with respect to $\mathbf{W}_i$ and $\mathbf{b}_i$ in lines 6 and 7. Here $\mathbf{y}_{m,i}(j)$ is the m-th column of $\mathbf{Y}_i(j)$.

The sites then send their local gradient estimates $\mathbf{G}_i(j)$ and $\mathbf{h}_i(j)$ to an aggregator site, which aggregates them according to lines 11-13. After updating $\mathbf{W}(j)$ and $\mathbf{b}(j)$, the aggregator checks if any values in $\mathbf{W}(j)$ increased above an upper bound of 10^9 in absolute value. If so, the aggregator resets the global unmixing matrix, sets $j = 0$, and anneals the learning rate by $\rho = 0.9\rho$. Otherwise, before continuing, if the angle between $\Delta_{\mathbf{W}}(j)$ and $\Delta_{\mathbf{W}}(j-1)$ is above 60^o, the aggregator anneals the learning rate by $\rho = 0.9\rho$, preventing $\mathbf{W}$ from scaling down too quickly without learning the data. The aggregator sends the updated $\mathbf{W}(j)$ and $\mathbf{b}(j)$ back to the sites. Finally, the algorithm stops when $\|\Delta_{\mathbf{W}}(j)\|_2^2 < t$.

In order to recover the statistically independent source estimates $\mathbf{S}_i$, each site computes

$$\mathbf{S}_i \approx \mathbf{W}\mathbf{X}_{i,\text{red}}. \tag{5}$$

For the pooled-data case, Amari et. al [13] demonstrate that $\mathbf{W}$ will converge asymptotically to $\mathbf{A}^{-1}$ in Infomax ICA. In the decentralized-data case, djICA acts as in the pooled-data case in terms of convergence. The assumption of a common mixing matrix assures that the global gradient sum will be identical to the pooled-data gradient on average, likewise moving the global weight matrix towards convergence too.

PCA preprocessing. Here, we describe dPCA algorithms for dimension reduction and whitening. This serves as a preprocessing step to standardize the data prior to djICA, also without communicating full data sets outside of local sites.

Algorithm 1 decentralized joint ICA (djICA)

Require: data $\{\mathbf{X}_{i,\text{red}} \in \mathbb{R}^{r \times N_i} : i = 1, 2, \ldots, s\}$, where r is the same across sites, tolerance level $t = 10^{-6}$, $j = 0$, maximum iterations J, $\|\Delta_{\mathbf{W}}(0)\|_2^2 = t$, initial learning rate $\rho = 0.015/ln(r)$

1: Initialize $\mathbf{W} \in \mathbb{R}^{r \times r}$ ▷ for example, $\mathbf{W} = \mathbf{I}$
2: **while** $j < J$, $\|\Delta_{\mathbf{W}}(j)\|_2^2 \geq t$ **do**
3: **for all** sites $i = 1, 2, \ldots, s$ **do**
4: $\mathbf{Z}_i(j) = \mathbf{W}(j-1)\mathbf{X}_i + \mathbf{b}(j-1)\mathbf{1}^\top$
5: $\mathbf{Y}_i(j) = g(\mathbf{Z}_i(j))$
6: $\mathbf{G}_i(j) = \rho\left(\mathbf{I} + (1 - 2\mathbf{Y}_i(j))\mathbf{Z}_i(j)^\top\right)\mathbf{W}(j-1)$
7: $\mathbf{h}_i(j) = \rho \sum_{m=1}^{N_i}(1 - 2\mathbf{y}_{m,i}(j))$
8: Send $\mathbf{G}_i(j)$ and $\mathbf{h}_i(j)$ to the aggregator site.
9: **end for**
10: At the aggregator site, update global variables
11: $\Delta_{\mathbf{W}}(j) = \sum_{i=1}^{s} \mathbf{G}_i(j)$
12: $\mathbf{W}(j) = \mathbf{W}_i(j-1) + \Delta_{\mathbf{W}}(j)$
13: $\mathbf{b}(j) = \mathbf{b}(j-1) + \sum_{i=1}^{s} \mathbf{h}_i(j)$
14: Check upper bound and learning rate adjustment.
15: Send global $\mathbf{W}(j)$ and $\mathbf{b}(j)$ back to each site
16: **end while**

First, Balcan et al. [14] use subspace embeddings that decrease the runtime for dPCA while controlling for the accuracy of $\mathbf{U}$. Bai et al. [15], however, bypass local data reduction in their approach (i.e., one-step dPCA), which was appealing and motivated its choice for some of our experiments. Lastly, an alternative two-step dPCA approach was considered based on the STP and MIGP [16] approaches recently developed for large PCA of multi-subject fMRI data. Its advantage is the small whitening matrix $\mathbf{P} \in \mathbb{R}^{d \times k}$ that is transmitted across sites, compared to the large $d \times d$ $\mathbf{R}$-matrix in Bai's algorithm [15]. The downside is that there are no bounds on the accuracy of $\mathbf{U}$, and results can vary slightly with the order in which sites and subjects are processed. Nonetheless, our results suggest that the two-step dPCA approach, summarized in Algorithms 2 and 3, yields a fairly good estimate of $\mathbf{U}$.

Algorithm 2 Local PCA algotithm (LocalPCA)

Require: data $\mathbf{X} \in \mathbb{R}^{d \times N}$ and intended rank k

1: Compute the SVD $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}$.
2: Let $\mathbf{\Sigma}^{(k)} \in \mathbb{R}^{k \times k}$ contain the largest k singular values and $\mathbf{U}^{(k)} \in \mathbb{R}^{d \times k}$ the corresponding singular vectors.
3: Save $\mathbf{U}^{(k)}$ and $\mathbf{\Sigma}^{(k)}$ locally and return $\mathbf{P} = \mathbf{U}^{(k)}\mathbf{\Sigma}^{(k)}$.

Algorithm 3 uses a peer-to-peer scheme to iteratively refine $\mathbf{P}(j)$, with the last site broadcasting the final $\mathbf{U}$ to all sites. $\mathbf{U}$ is the matrix containing the top r' columns of $\mathbf{P}(s)$ with largest L_2-norm, but normalized to unit L_2-norm instead. Following the recommendation in [16], we set $r = 20$ and $k = 5 \cdot r$ for our simulations.

Algorithm 3 Global PCA algorithm (GlobalPCA)

Require: s sites with data $\{\mathbf{X}_i \in \mathbb{R}^{d\times N_i} : i = 1, 2, \ldots, s\}$, intended final rank r, local rank $k \geq r$.

Choose a random order π for the sites.
$\mathbf{P}(1) = \mathsf{LocalPCA}(\mathbf{X}_{\pi(1)}, \min\{k, \mathrm{rank}(\mathbf{X}_{\pi(1)})\})$
for all $j = 2$ to s **do**
  $i = \pi(j)$
  Send $\mathbf{P}(j-1)$ from site $\pi(j-1)$ to site $\pi(j)$
  $k' = \min\{k, \mathrm{rank}(\mathbf{X}_i)\}$
  $\mathbf{P}' = \mathsf{LocalPCA}(\mathbf{X}_i, k')$
  $k' = \max\{k', \mathrm{rank}(\mathbf{P}(j-1))\}$
  $\mathbf{P}(j) = \mathsf{LocalPCA}([\mathbf{P}'\ \mathbf{P}(j-1)], k')$
end for
$r' = \min\{r, \mathrm{rank}(\mathbf{P}(s))\}$
$\mathbf{U} = \textsc{NormalizeTopColumns}(\mathbf{P}(s), r')$ ▷ At last site
Send $\mathbf{U}$ to sites $\pi(1), \ldots, \pi(s-1)$.
for all $i = 1$ to s **do**
  $\mathbf{X}_{i,\mathrm{red}} = \mathbf{U}^\top \mathbf{X}_i$ ▷ The locally reduced data
end for

3. RESULTS

We examined five different scenarios involving synthetic data to understand if and when djICA would have comparable performance to a pooled analysis.

Synthetic sources. The $\mathbf{S}$ signals were simulated using a generalized autoregressive (AR) conditional heteroscedastic (GARCH) model [17, 18], which has been shown to be useful in models of causal source separation [19] and time-series analyses of data from neuroscience experiments [19, 20], especially resting-state fMRI time courses [21, 22]. We simulated fMRI time courses using a GARCH model by generating an AR process (no moving average terms) randomly such that the AR series converges. We chose a random order between 1 and 10 and random coefficients $\{\alpha[\ell]\}$ such that $\alpha[0] \in [0.55, 0.8]$ and $\alpha[\ell] \in [-0.35, 0.35]$ for $\ell > 0$. For the error terms $\delta_t = \sigma_t \epsilon_t$, we used an ARMA model driven by ϵ_t from a generalized Normal distribution with shape parameter 100 (so it was approximately uniform on $[-1, 1]$) and $\sigma_t^2 = 0.1 + 0.1y[t-1]^2 + 0.75\sigma[t-1]^2$. For each of 2048 simulated subjects, we generated 20 time courses with 250 time points, each after a "burn-in" period of 20000 samples, checking that all pair-wise correlations between the 20 time courses was below 0.35. We generated a total of 2048 mixed datasets for each experiment.

Mixing model and algorithms. We evaluated 5 different scenarios for processing our synthetic data, as summarized in the table below:

scenario	preprocessing	mixing matrix A
1. ICA (pooled)	none	i.i.d. Gaussian
2. djICA	none	i.i.d. Gaussian
3. ICA (pooled)	PCA	simTB map
4. djICA	One−Step dPCA	simTB map
5. djICA	Two−Step dPCA	simTB map

For the one-step dPCA scenario, we use Bai et. al's dPCA algorithm without updating [15], and for the two-step dPCA scenario, we use the LocalPCA algorithm (Algorithm 2), followed by GlobalPCA (Algorithm 3). For the first two scenarios, we generated i.i.d. Gaussian mixing matrices $\mathbf{A} \in \mathbb{R}^{r\times r}$. For the higher-dimensional problems (scenarios 3-5), we used the fMRI Simulation Toolbox's simTB spatial maps [23] to generate different $\mathbf{A} \in \mathbb{R}^{d\times r}$ mixing matrices.

As a performance metric we use the Moreau-Amari inter-symbol interference (ISI) index [12], which is a function of the square matrix $\mathbf{Q} = \hat{\mathbf{W}}\mathbf{A}$, where $\hat{\mathbf{W}} = \mathbf{W}\mathbf{U}^\top$.

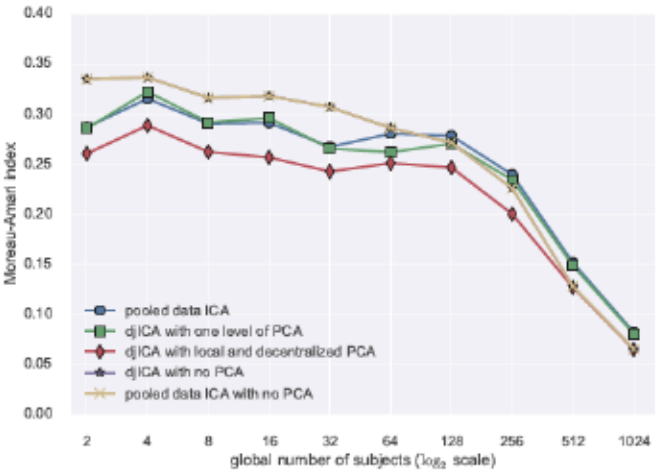

Fig. 2. Increasing the number of subjects at two sites

How do the algorithms compare as we increase the data at a fixed number of sites? We fixed $s = 2$ sites and evaluated our five algorithms. For the distributed settings we split the data evenly per site. Fig. 2 shows ISI versus the total data set size. As the data increases all algorithms improve, and more importantly, the distributed versions perform nearly as well as the pooled-data counterparts. Results are averaged over 10 randomly generated mixing matrices.

How do the algorithms compare as we increase the number of sites with a fixed amount of data sets per site? We fixed the total 2048 subjects but investigated the effect of increasing s where $M_i = 32$ subjects at each site. Results are averaged over 10 randomly generated mixing matrices. Fig. 3 demonstrates the convergence of ISI curve with the increase of the combined data. Again, we see that the performance of djICA is very close to the centralized pooled performance, even for a small number of subjects per site.

How does splitting the data sets across more sites affect performance? We examined splitting the 2048 data subjects

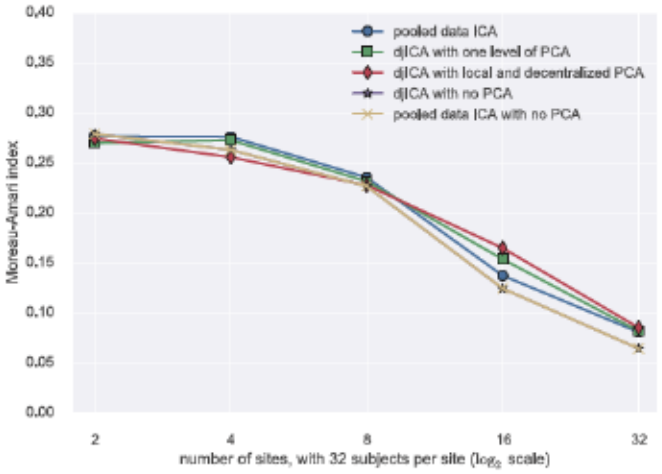

Fig. 3. Each site with 32 subjects

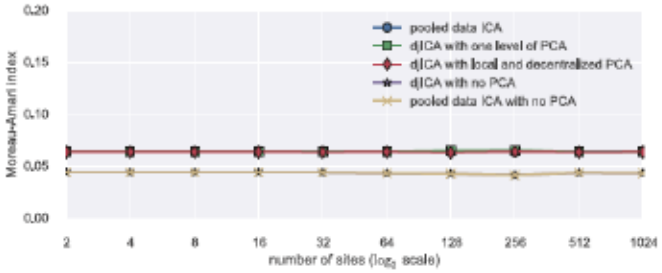

Fig. 4. Gradually splitting 2048 subjects

across more and more sites (increasing s), so for small s each site had more data. Fig. 4 shows that the performance of djICA is very close to that of the pooled-data ICA, even with more and more sites holding fewer and fewer data points. This implies that we can support decentralized data with little loss in performance.

Estimated spatial maps. For three data distribution settings of the last experiment (gradually splitting data over more and more sites), Fig. 5 shows the spatial maps estimated in scenarios 3-5. The maps contain $r = 20$ independent spatial components, which are color coded. The map under the heading 'GT' is the ground-truth simulated map generated with simTB, the map under the heading 'Pooled' is the map estimated by performing centralized pooled ICA, maps under the heading 'djICAp' underwent Bai's (high-bandwidth) one-step dPCA prior to djICA, and maps under the heading djICAp^2 underwent the (lower-bandwidth) two-step dPCA process prior to djICA. The components estimated in each of the maps in Fig. 5 illustrate not only that djICA can estimate components as well as pooled ICA, but that the algorithm can perform well regardless of how the subject datasets are distributed across sites.

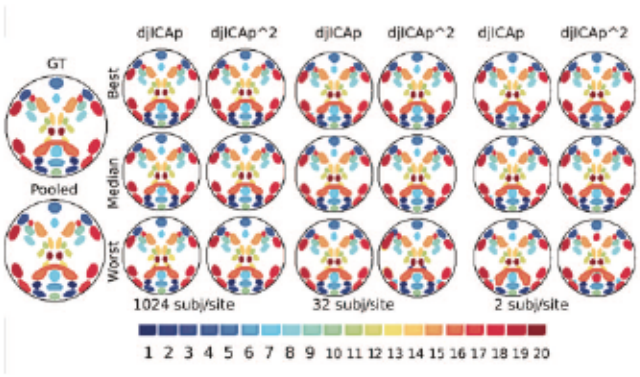

Fig. 5. Spatial pap estimations for 2048 subjects over different site distributions

4. DISCUSSION AND CONCLUSIONS

In contrast to systems optimized for processing large amounts of data by making computation more efficient (Apache Spark, H2O and others), we focus on a different setting common in research collaborations: data is expensive to collect and spread across many sites. To that end we proposed a distributed data ICA algorithm that (in synthetic experiments) finds underlying sources in decentralized data nearly as accurately as its centralized counterpart. This shows that algorithms like djICA may enable collaborative processing of decentralized data by combining local computation and communication. To further validate this proof-of-concept we plan to evaluate this method on real fMRI experiments. Additional extensions include reducing the bandwidth of the method and designing privacy-preserving variants that guarantee differential privacy [24]. There, reducing the iteration complexity will help guarantee more privacy and hence incentivize larger research collaborations.

5. ACKNOWLEDGMENT

The authors thank Dr. Robert Rubin, president and CEO of the Lovelace Respiratory Research Institute, Dr. Patrick McDonald, professor of mathematics at the New College of Florida, and Dr. Tyrone Ryba, assistant professor of bioinformatics at the New College of Florida, for helping organize the partnership between the Mind Research Network and the New College of Florida. They also thank Dr. Zheng-Jian Bai for sharing the MATLAB code for distributed PCA [15].

6. REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?," *Sig. Proc.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[2] V.D. Calhoun, T. Adali, G.D. Pearlson, and J.J. Pekar,

"A method for making group inferences from functional mri data using independent component analysis," *Hum Brain Mapp*, vol. 14, no. 3, pp. 140–151, 2001.

[3] I. Daubechies, E. Roussos, S. Takerkart, M. Benharrosh, C. Golden, K. D'Ardenne, W. Richter, J.D. Cohen, and J. Haxby, "Independent component analysis for brain fMRI does not select for independence," *PNAS*, vol. 106, no. 26, pp. 10415–10422, 2009.

[4] V.D. Calhoun, V.K. Potluru, R. Phlypo, R.F. Silva, B.A. Pearlmutter, A. Caprihan, S.M. Plis, and T. Adali, "Independent component analysis for brain fMRI does indeed select for maximal independence," *PloS One*, vol. 8, no. 8, pp. e73309, 2013.

[5] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[6] E.G. Learned-Miller and J.W. Fisher III, "ICA using spacings estimates of entropy," *J Mach Learn Res*, vol. 4, pp. 1271–1295, 2003.

[7] V. Krishnaveni, S. Jayaraman, P.M. Manoj Kumar, K. Shivakumar, and K. Ramadoss, "Comparison of independent component analysis algorithms for removal of ocular artifacts from electroencephalogram," *Meas Sci Rev*, vol. 5, Section 2, no. 6, pp. 67–78, 2005.

[8] J. Sui, T. Adali, G.D. Pearlson, and V.D. Calhoun, "An ICA-based method for the identification of optimal fMRI features and components using combined group-discriminative techniques," *NeuroImage*, vol. 46, no. 1, pp. 73–86, 2009.

[9] J. Liu and V.D. Calhoun, "Parallel independent component analysis for multimodal analysis: Application to fmri and eeg data," in *Proc. IEEE ISBI: From Nano to Macro*, 2007, pp. 1028–1031.

[10] E.A. Allen, E.B. Erhardt, E. Damaraju, W. Gruner, J.M. Segall, R.F. Silva, M. Havlicek, S. Rachakonda, J. Fries, R. Kalyanam, A.M. Michael, A. Caprihan, J.A. Turner, T. Eichele, S. Adelsheim, A.D. Bryan, J. Bustillo, V.P. Clark, S.W. Feldstein Ewing, F. Filbey, C.C. Ford, K. Hutchison, R.E. Jung, K.A. Kiehl, P. Kodituwakku, Y.M. Komesu, A.R. Mayer, G.D. Pearlson, J.P. Phillips, J.R. Sadek, M. Stevens, U. Teuscher, R.J. Thoma, and V.D. Calhoun, "A baseline for the multivariate comparison of resting state networks," *Front in Syst Neurosci*, vol. 5, no. 2, 2011.

[11] V.D. Calhoun, T. Adali, N.R. Giuliani, J.J. Pekar, K.A. Kiehl, and G.D. Pearlson, "Method for multimodal analysis of independent source differences in schizophrenia: Combining gray matter structural and auditory oddball functional data," *Hum Brain Mapp*, vol. 27, no. 1, pp. 47–62, 2006.

[12] S.-I. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separation," *Advances in NIPS*, pp. 757–763, 1996.

[13] S.-I. Amari, T.-P. Chen, and A. Cichocki, "Stability analysis of learning algorithms for blind source separation," *Neural Networks*, vol. 10, pp. 1345–1351, 1997.

[14] M.-F. Balcan, V. Kanchanapally, Y. Liang, and D.P. Woodruff, "Improved distributed principal component analysis," Tech. Rep. 1408.5823 [cs.LG], ArXiV, 2014.

[15] Z.-J. Bai, R.H. Chan, and F.T. Luk, "Principal component analysis for distributed data sets with updating," in *Proc. APPT*, 2005, pp. 471–483.

[16] V.D. Calhoun, RF. Silva, T. Adali, and S. Rachakonda, "Comparison of PCA approaches for very large group ICA," *NeuroImage, In Press*, 2015.

[17] R.F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.

[18] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J Econometrics*, vol. 31, pp. 307–327, 1986.

[19] K. Zhang and A. Hyvärinen, "Source separation and higher-order causal analysis of MEG and EEG," in *Proc. UAI*, Catalina Island, California, 2010, AUAI Press.

[20] T. Ozaki, *Time Series Modeling of Neuroscience Data*, CRC Press, 2012.

[21] Q. Luo, G. Tian, F. Grabenhorst, J. Feng, and E.T. Rolls, "Attention-dependent modulation of cortical taste circuits revealed by Granger causality with signal-dependent noise," *PLoS Comp. Bio.*, vol. 9, no. 10, pp. e1003265, 10 2013.

[22] M.A. Lindquist, Y. Xu, M.B. Nebel, and B.S. Caffo, "Evaluating dynamic bivariate correlations in resting-state fMRI: A comparison study and a new approach," *NeuroImage*, vol. 101, pp. 531 – 546, 2014.

[23] E.B. Erhardt, E.A. Allen, Y. Wei, T. Eichele, and V.D. Calhoun, "SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability," *NeuroImage*, vol. 59, no. 4, pp. 4160–4167, 2012.

[24] C. Dwork, "Differential privacy," *Automata, languages and programming*, pp. 1–12, 2006.