MULTI-TASK FMRI DATA FUSION USING IVA AND PARAFAC2

Isabell Lehmann[1]*, Evrim Acar*[2]*, Tanuj Hasija*[1]*, M.A.B.S. Akhonda*[3]*, Vince D. Calhoun*[4]*, Peter J. Schreier*[1]*, Tülay Adali*[3]

[1] Signal and System Theory Group, Paderborn University, Germany
[2] Simula Metropolitan Center for Digital Engineering, Oslo, Norway
[3] Dept.of CSEE, University of Maryland, Baltimore County
[4] Tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State, Georgia Tech, and Emory University, Atlanta, GA

ABSTRACT

Data fusion—the joint analysis of multiple datasets—through coupled factorizations has the promise to enable enhanced knowledge discovery, and hence is an active area. Various formulations of coupled matrix factorizations have been proposed, each with its own modeling assumptions. In this paper, we study two such methods, namely Independent Vector Analysis (IVA), i.e., extension of Independent Component Analysis (ICA) to multiple datasets, and PARAFAC2, a tensor factorization approach. We demonstrate the modeling assumptions of IVA and PARAFAC2 using simulations, revealing that both methods can accurately capture the latent components, albeit with certain differences in capturing the corresponding subject scores. By making use of a rich multi-task functional Magnetic Resonance Imaging (fMRI) dataset, we show how the two methods can be used for achieving two important goals at once, namely capturing group differences between patients with schizophrenia and healthy controls with interpretable components, as well as understanding the relationship across multiple tasks. This is achieved through the definition of source component vectors across datasets.

Index Terms— data fusion, independent vector analysis, tensor decompositions, PARAFAC2, multi-task fMRI

1. INTRODUCTION

The interest in data fusion, which refers to the joint analysis of multiple related datasets, has grown in recent years in various research areas [1, 2, 3] such as biomedicine [4, 5] or wireless sensor networks [6], among others. Data fusion approaches allow us to leverage the complementary information in multiple datasets [7] by letting them fully interact and inform each other without putting strong constraints on them [2, 8]. Data-driven methods, especially methods based on matrix and tensor decompositions, are particularly attractive for data fusion [2, 7, 8]. IVA [9] is an extension of the well-known ICA to multiple datasets, and is thus a good candidate for data fusion by making use of the dependence across datasets [10]. If multiple datasets are stacked along a third mode to form a higher-order array (also referred to as a higher-order tensor), one can also take advantage of the inherent relationship across the datasets using tensor decompositions [11]. The PARAFAC2 [12] tensor model has proved useful for jointly analyzing datasets [13, 14, 15] as it does not impose strong constraints on them, in contrast to the well-known CANDECOMP/PARAFAC (CP) [16, 17] tensor decomposition method.

In recent years, fusion of medical imaging data has received growing attention and has led to the identification of novel biomarkers for disorders such as schizophrenia [4, 18]. One of the important fusion tasks is the use of data collected from the same subjects while they are performing different tasks, as in the case of multi-task fMRI data [19]. Previous studies have concentrated on analyzing multiple datasets jointly to be able to distinguish between two groups, e.g., patients with schizophrenia and healthy controls, and made use of a single image per task, see, e.g., [20]. It is of interest to analyze multiple feature data per task to also examine the relationship between different brain networks for these datasets that report on multiple aspects of a task.

In this paper, we address the relationship between two important models, namely IVA-G [21], which is IVA using a multivariate Gaussian model, and PARAFAC2, for data fusion, and we demonstrate their modeling assumptions through simulations. We apply these two models for the analysis of multi-task fMRI data, such that the importance and relevance of different modeling assumptions as well as their role for interpretability of the results can be clearly assessed. We see that both methods not only recover components that discriminate between patients and controls, but also enable discovery of the relationship across different tasks, thus, *fully* make use of the complementary information across datasets. Furthermore, to the best of our knowledge, this is the first application of the PARAFAC2 model to analyze multiple datasets with the goal of understanding the relationship across datasets in terms of biomarkers. We analyze 13 datasets from the MCIC collection [22], which are collected from 271 subjects that perform three different tasks with well defined relationship among them. The datasets are acquired using different regressors for the three tasks. As such, these datasets allow us to interpret the outcome of both methods, especially for an analysis that has a focus on finding their relationship.

2. METHODS

2.1. IVA-G

IVA [9] is an extension of ICA to multiple datasets. It assumes the Blind Source Separation (BSS) model for K datasets,

$$\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\mathbf{S}^{[k]}, \quad k = 1, \ldots, K, \tag{1}$$

where $\mathbf{X}^{[k]} \in \mathbb{R}^{I \times J}$ is the observed data in the k^{th} dataset, $\mathbf{S}^{[k]} \in \mathbb{R}^{I \times J}$ is the source matrix consisting of I source components with each J samples, and $\mathbf{A}^{[k]} \in \mathbb{R}^{I \times I}$ is a mixing matrix. The goal of IVA is to estimate the source matrices $\mathbf{U}^{[k]} = \mathbf{W}^{[k]}\mathbf{X}^{[k]} \in \mathbb{R}^{I \times J}$, $k = 1, \ldots, K$, where $\mathbf{W}^{[k]} \in \mathbb{R}^{I \times I}$ is the demixing matrix for the k^{th} dataset.

We define the i^{th} Source Component Vector (SCV) $\mathbf{S}_i$, by concatenating the i^{th} source component (row) of each $\mathbf{S}^{[k]}$, as

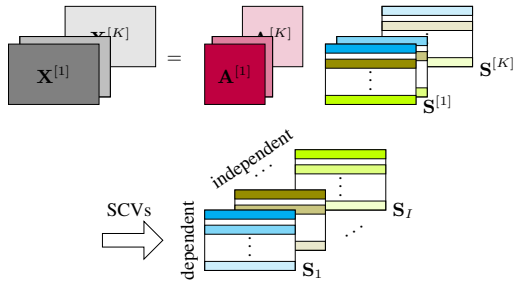


Fig. 1: Illustration of IVA (from [23]). The i^{th} row (source component) of all $\mathbf{S}^{[k]}$ forms the i^{th} SCV $\mathbf{S}_i$. Source components are made maximally dependent within an SCV and maximally independent across SCVs in IVA (and maximally uncorrelated in IVA-G).

$\mathbf{S}_i = \left[\left(\mathbf{s}_{i:}^{[1]}\right)^T \left(\mathbf{s}_{i:}^{[2]}\right)^T \dots \left(\mathbf{s}_{i:}^{[K]}\right)^T\right]^T \in \mathbb{R}^{K\times J}$, as shown in Figure 1. IVA minimizes the mutual information among SCVs, which also leads to increasing the dependence within an SCV through an appropriate selected multivariate density model [24]. IVA-G [24] assumes a multivariate Gaussian distribution for the SCVs, and thus only takes second-order statistical information into account.

2.2. PARAFAC2

PARAFAC2 [12, 25] is a more flexible version of the commonly used CP tensor model. An R-component CP model represents a third-order tensor $\mathcal{X} \in \mathbb{R}^{I\times J\times K}$ as follows [16, 17]:

$$\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \tag{2}$$

with the factor matrices $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_R] \in \mathbb{R}^{I\times R}$, $\mathbf{B} = [\mathbf{b}_1 \dots \mathbf{b}_R] \in \mathbb{R}^{J\times R}$, $\mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_R] \in \mathbb{R}^{K\times R}$, and the outer product denoted by $\circ$. We use $\mathbf{a}_r$, $\mathbf{b}_r$, $\mathbf{c}_r$ as a more compact notation of $\mathbf{a}_{:r}$, $\mathbf{b}_{:r}$, $\mathbf{c}_{:r}$, the r^{th} column of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, respectively. With $\mathbf{X}^{[k]} \in \mathbb{R}^{I\times J}$ being the k^{th} frontal slice of $\mathcal{X}$, (2) can be rewritten as

$$\mathbf{X}^{[k]} \approx \mathbf{A}\,\mathrm{diag}\left(\mathbf{c}_{k:}\right)\mathbf{B}^T, \quad k = 1, \dots, K, \tag{3}$$

where $\mathbf{c}_{k:}$ is the k^{th} row of $\mathbf{C}$, $\mathrm{diag}\left(\cdot\right)$ denotes a diagonal matrix with the corresponding vector on the main diagonal, and $(\cdot)^T$ denotes the transpose. In (3), we see that the CP model assumes common factor matrices $\mathbf{A}$ and $\mathbf{B}$ for all K slices.

In contrast to this, PARAFAC2 allows for changes of the $\mathbf{B}$ matrix, meaning that the factor matrix $\mathbf{B}^{[k]} \in \mathbb{R}^{J\times R}$ can be different in each frontal slice (dataset) as shown in Figure 2:

$$\begin{aligned} \mathbf{X}^{[k]} &\approx \mathbf{A}\,\mathrm{diag}\left(\mathbf{c}_{k:}\right)\left(\mathbf{B}^{[k]}\right)^T \\ &\text{s.t. } \left(\mathbf{B}^{[k]}\right)^T \mathbf{B}^{[k]} = \mathbf{M}, \quad k = 1, \dots, K, \end{aligned} \tag{4}$$

where $\mathbf{M}$ is an arbitrary matrix, and with the constraint introduced to preserve uniqueness of the components up to permutation and scaling ambiguities [12, 25]. We use a Frobenius-norm based loss function when fitting the model [25].

2.3. Comparison of the methods

Equations (1) and (4) show that IVA-G and PARAFAC2 models are equivalent when $\mathbf{A}^{[k]} = \mathbf{A}\mathrm{diag}(\mathbf{c}_{k:})$ and $\mathbf{S}^{[k]} = (\mathbf{B}^{[k]})^T$. However, each model has different cost functions and hence different assumptions, i.e., PARAFAC2 is more constrained in terms of $\mathbf{A}^{[k]}$,

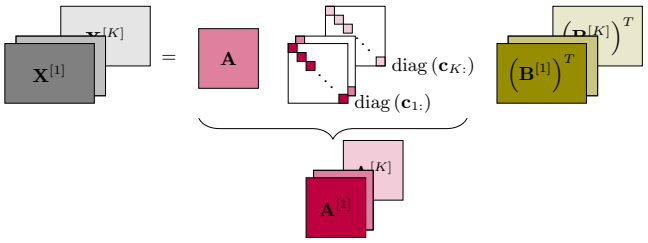


Fig. 2: Illustration of a PARAFAC2 model. The $\mathbf{A}^{[k]}$ are more constrained than in IVA-G; however, the $\mathbf{B}^{[k]}$ are less constrained by only fulfilling $\left(\mathbf{B}^{[k]}\right)^T \mathbf{B}^{[k]} = \mathbf{M}$.

while IVA-G is more constrained in terms of $\mathbf{S}^{[k]}$. If $\mathbf{M} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, the assumptions of the two models become more similar as both assume uncorrelatedness of the sources. We can define the r^{th} SCV in PARAFAC2, by concatenating the r^{th} column of all $\mathbf{B}^{[k]}$, as $\mathbf{B}_r = [\mathbf{b}_r^{[1]}\ \mathbf{b}_r^{[2]} \dots \mathbf{b}_r^{[K]}]^T \in \mathbb{R}^{K\times J}$. As PARAFAC2 has a smaller set of parameters to estimate than IVA-G, it is likely to be more robust against noise.

3. EXPERIMENTS

Using experiments on both simulated and real data, we demonstrate that IVA-G and PARAFAC2 can capture the relationship across task datasets in the SCV covariance matrices as well as find source components that discriminate between two groups, in our case patients with schizophrenia and healthy controls.

3.1. Simulated data

3.1.1. Data generation

We have simulated two scenarios with $K = 12$ datasets, the first following (4), and the second following (1) and violating $\mathbf{A}^{[k]} = \mathbf{A}\mathrm{diag}(\mathbf{c}_{k:})$. In both scenarios, for the *voxels mode*, the sources $\mathbf{S}^{[k]} = (\mathbf{B}^{[k]})^T \in \mathbb{R}^{4\times 5000}$ are generated using a multivariate Gaussian distribution with the covariance matrix for each SCV shown in Figure 3. For the *task/datasets mode*, the factor matrix $\mathbf{C} \in \mathbb{R}^{12\times 4}$ is also generated equally for both scenarios. The entries of the first three columns are drawn from $\mathcal{N}(1.5, 0.01)$. From the first four datasets in component 2 and the last eight datasets in component 3, we subtract 1 to simulate that the component is not present in some datasets. The last column is drawn from $\mathcal{N}(1.5, 0.25)$ to make sure that the columns of $\mathbf{C}$ are not too similar. For the *subjects mode*, in scenario 1, the first and third column of $\mathbf{A} \in \mathbb{R}^{300\times 4}$ are generated according to a normal distribution $\mathcal{N}(0, 1)$, with 0.5 being added to the last 150 values of the columns to simulate a difference between patients and controls. The second and fourth column are also distributed normally, with mean 0 and standard deviation as the average of the standard deviations of the first and third component. In scenario 2, $\mathbf{A}$ has only a step in the first column, and then all $\mathbf{A}^{[k]}$ are calculated. After that, a step of height 0.5 is added the third column of $\mathbf{A}$, and the third column of $\mathbf{A}^{[k]}$ is recalculated for the first four datasets. This way, $\mathbf{A}^{[k]} \neq \mathbf{A}\mathrm{diag}(\mathbf{c}_{k:})$ for the third component.

3.1.2. Performance evaluation

For PARAFAC2, non-negativity constraints are imposed on the *task/datasets mode* to solve the potential sign indeterminacy per

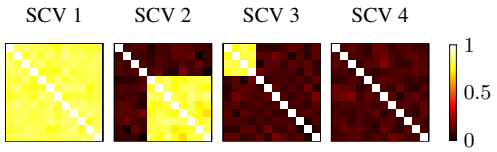
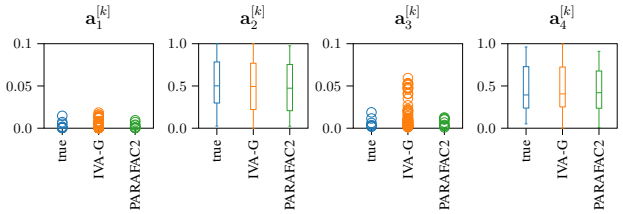
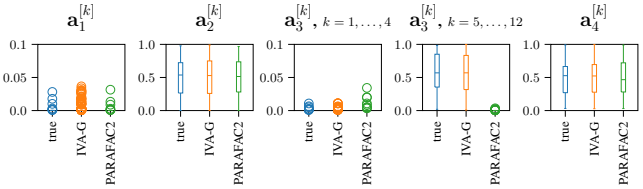


Fig. 3: Covariance matrices (12×12) of simulated sources for each SCV, which are estimated reliably by both models (not shown).



(a) Scenario 1



(b) Scenario 2

Fig. 4: Distribution of the true and estimated p-values in $\mathbf{A}^{[k]}$.

component [12]. We find the correct permutation of the PARAFAC2 components by maximizing the cosine similarity of the concatenated true and estimated $\mathbf{B}^{[k]}$ matrices. To test for statistically significant difference ($p < 0.05$) between patients and controls (first and second half of a column in $\mathbf{A}^{[k]}$), a two-sample t-test (unequal variances) is applied on each column of $\mathbf{A}^{[k]}$, and the corresponding p-values are calculated. The distributions of the true and estimated p-values are compared in boxplots. Furthermore, we calculate the average correlation of the true and estimated sources across datasets for each SCV.

3.1.3. Results of the simulated data analysis

We reduce the dimension of the simulated dataset to 4 via Principal Component Analysis (PCA), and we apply IVA-G on the dimension-reduced dataset with multiple initializations and choose the most consistent run[1] [26]. We run PARAFAC2 ($R = 4$ components) with non-negativity constraints[2] for multiple initializations and choose the run with the smallest reconstruction error. The results for the simulated data analysis, averaged across 50 independent Monte-Carlo runs, show that IVA-G and PARAFAC2 both achieve a high correlation between the true and estimated sources. Thus, PARAFAC2 is able to estimate $\mathbf{B}^{[k]}$ correctly despite the assumption on $\mathbf{A}^{[k]}$ being violated. Figure 4 shows the distributions of the true and estimated p-values for both scenarios. As $\mathbf{A}$ is the same for all datasets in scenario 1, the p-value for all datasets is also the same, while in scenario 2, it is different between the first four and last eight datasets in component 3. IVA-G is more flexible as it estimates a different p-value for every dataset, while PARAFAC2 finds the same p-value for all datasets because it assumes the same mixing matrix. We see in Figure 4(a) that, for scenario 1, in component 3, IVA-G overestimates the p-value for some datasets, while PARAFAC2 estimates the p-value small enough matching the ground truth. However, in scenario 2, IVA-G can capture the different p-values in component 3 (Figure 4(b)), which PARAFAC2 cannot due to its model design.

[1]Python code available at: `https://github.com/SSTGroup/independent_vector_analysis`

[2]Python code available at: `https://github.com/tensorly/tensorly` [15, 27]

3.2. Real data

3.2.1. Dataset

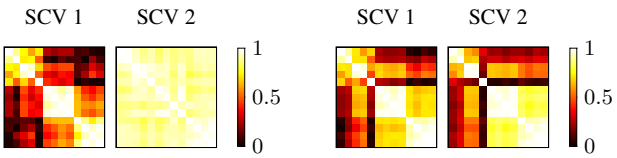
We analyze 13 fMRI datasets from the MCIC collection [22], which are collected from 271 subjects (121 patients with schizophrenia and 150 healthy controls) that perform three different tasks: Auditory Oddball (AOD), Sensory Motor (SM), and Sternberg Item Recognition Paradigm (SIRP). The lower-dimensional multivariate features for each subject and task were extracted using regressors that were created by convolving the hemodynamic response function (HRF) in SPM [28] with desired predictors, as in [20]. The resulting regression coefficient maps, i.e., features, for all subjects are then concatenated vertically to form the task datasets $\mathbf{X}^{[k]} \in \mathbb{R}^{271 \times 48546}$, $k = 1, \ldots 13$, where 48546 is the number of voxels.

The AOD task required subjects to listen to three different types of stimuli, i.e., standard, novel and target, coming in pseudorandom order, and to press a button whenever the target stimulus occurs. Regressors were created to model the target (T), the target with the standard (TS), the novel (N), and the novel with the standard (NS) stimuli, thus resulting in four task datasets for the AOD task. In the SM task, subjects had to listen to 16 different audio tones and to press a button after each tone change. The regressor was created to model the entire increase and decrease block, thus resulting in only one task dataset. The SIRP task consists of the encoding and probe phase. Subjects were asked to memorize a set of 1, 3, and 5 integer digits, randomly selected from 0 to 9, in the encoding phase and to press a button whenever a digit from the memorized set was presented in the probe phase. The regressors were created for both the encoding (E) and probe (P) phase (for 1, 3, 5 digits, and the averaged data), resulting in eight task datasets for the SIRP task.

3.2.2. Results with the real data

We assume that the fMRI datasets can be modeled with a lower-dimensional set of latent sources, as in the simulations. Therefore, again a dimension reduction is performed via PCA before applying IVA-G. PARAFAC2 may have an upper limit on the number of components to uniquely estimate, which is $K \geq R(R+1)(R+2)(R+3)/24$ [25]. With IVA, typical orders are significantly higher, which allows for a more detailed decomposition that is easier to interpret, as order selection is highly linked to the interpretability of the source components. Given the space constraints of this paper and to have a fair comparison, we choose the PCA rank as the number of components in PARAFAC2. We have studied different orders for PARAFAC2 and have decided for $R = 2$ to obtain the most interpretable source components. The best run of IVA-G and PARAFAC2 is found as described in Section 3.1.3. The source components estimated by IVA-G and PARAFAC2 are normalized to unit variance.

Figure 5 shows the absolute values of the SCV covariance matrices estimated by both IVA-G and PARAFAC2, which summarize the relationship across the task datasets. Figure 5(a) shows that the first SCV in IVA-G shows high correlations only across subsets of



(a) Estimated by IVA-G (b) Estimated by PARAFAC2

Fig. 5: Covariance matrices of the estimated SCVs with real data.

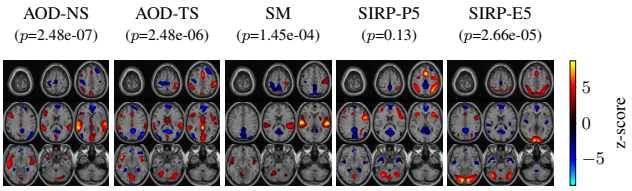


Fig. 6: Subset of fMRI maps from SCV 1, estimated by IVA-G.

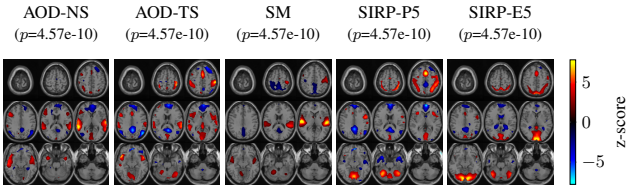


Fig. 7: Subset of fMRI maps from SCV 1, estimated by PARAFAC2.

task datasets, while the second SCV shows high correlations across all task datasets. On the other hand, as shown in Figure 5(b), both SCVs in PARAFAC2 show high correlations only across subsets of task datasets. Thus, although PARAFAC2 does not model the dependence within an SCV, in the final decomposition, the estimated SCVs also yield the information related to the relationship of the task datasets. For both methods, the block-structured covariance matrices indicate that correlations across task datasets form two distinct groups, one across the AOD and SM datasets and another across the SIRP datasets. This is because the AOD and SM tasks are more similar to each other compared with the SIRP task [20].

A subset of the source components, i.e., fMRI maps, corresponding to the block-structured covariance matrix of SCV 1 in IVA-G and the corresponding SCV 1 in PARAFAC2, are visualized (thresholded at $|z| = 2$) in Figure 6 and Figure 7, respectively. To overcome the sign ambiguity of IVA-G, we first made sure that the t-value of each significant source component is positive or is made positive by multiplying the source component and the corresponding column of $\mathbf{A}^{[k]}$ with -1, and then we adjusted the signs of the remaining source components in the same way such that the activated regions have the same color in each SCV. This way, red / yellow voxels indicate a higher activation in controls than in patients, and blue means the opposite. The p-value is the same for all task datasets in PARAFAC2 and is very small, thus all task datasets provide a high discrimination between patients and controls. However, IVA-G can identify the task datasets that provide significantly different activations between patients and controls ($p < 0.05$), which improves the interpretability, especially when IVA-G is implemented with more components. Overall, group-discriminant source components estimated for the AOD and SM datasets are showing higher activations in controls in the auditory and motor regions (red focal areas in slices 5-7 and 2-3), while the source components in the

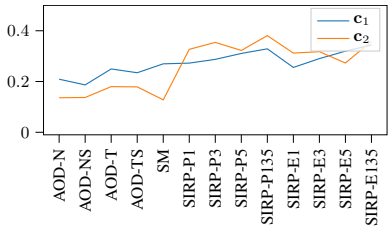


Fig. 8: Dataset covariations estimated by PARAFAC2. The first component is present in all datasets, and the second component is mainly present in the SIRP datasets.

SIRP-P datasets are showing higher activations mostly in the motor and in parts of the visual areas (red focal areas in slices 2-3 and 7-8), and in the SIRP-E datasets in the visual areas (red focal areas in slices 6-8). These are the areas associated with the tasks and known to be affected by schizophrenia from previous studies, and therefore are promising candidates for brain-based biomarkers [29, 30]. The Default Mode Network (DMN) (blue focal areas in slices 4-6), which is present in all but the SM datasets, is activated higher in patients than in controls. This makes sense as the DMN mostly represents the resting state: the less concentrated one is on a given task, the more dominant are the resting state source components, and typically, controls show a higher task-related suppression of the DMN than patients [29, 31]. The DMN is more clearly visible in the SIRP-P dataset for PARAFAC2 than for IVA-G.

PARAFAC2 provides us with an additional summary of the task datasets through the factor matrix $\mathbf{C}$, shown in Figure 8. We see that component 1 is present in all task datasets, thus showing an average component, while component 2 is mainly present in the SIRP datasets.

4. CONCLUSION

We have applied and compared IVA-G and PARAFAC2 in simulation scenarios and for fusing 13 multi-task fMRI datasets. Our simulations reveal that while both methods can accurately capture the underlying source components, PARAFAC2 captures group differences more reliably, providing a compact representation when subject scores differ only up to a scaling across different datasets, and IVA-G performs better when different subject scores are expected in different datasets. In the fMRI data, both methods are able to identify source components that discriminate between patients with schizophrenia and healthy controls, and to capture the relationship across task datasets in the SCV covariance matrices. A more detailed comparison of IVA using its different versions based on higher-order statistics and PARAFAC2 using different constraints could further demonstrate their versatility. How the two models can be objectively compared with an order optimally selected for each, is another point which needs to be further looked into.

5. ACKNOWLEDGEMENTS

This work was supported in part by NSF grants CCF 1618551, NCS 1631838 and NIH grants R01MH123610 and R01MH118695, and in part by the Research Council of Norway through project 300489, and in part by the German Research Foundation (DFG) under grant SCHR 1384/3-2. The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF).

6. REFERENCES

[1] B. Khaleghi, A. Khamis, et al., "Multisensor data fusion: A review of the state-of-the-art," *Information fusion*, vol. 14, no. 1, pp. 28–44, 2013.

[2] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.

[3] E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos, "Tensors for data mining and data fusion: Models, applications, and scalable algorithms," *ACM Transactions on Intelligent Systems and Technology*, vol. 8, no. 2, 2016.

[4] E. Acar, C. Schenker, et al., "Unraveling diagnostic biomarkers of schizophrenia through structure-revealing fusion of multimodal neuroimaging data," *Frontiers in Neuroscience*, vol. 13, 2019.

[5] B. Hunyadi, P. Dupont, et al., "Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 7, no. 1, pp. 1–15, 2017.

[6] H. Luo, H. Tao, et al., "Data fusion with desired reliability in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 3, pp. 501–513, 2010.

[7] E. Acar, R. Bro, and A. K. Smilde, "Data Fusion in Metabolomics Using Coupled Matrix and Tensor Factorizations," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1602–1620, 2015.

[8] T. Adali, M. A. B. S. Akhonda, and V. D. Calhoun, "ICA and IVA for Data Fusion: An Overview and a New Approach Based on Disjoint Subspaces," *IEEE Sensors Letters*, vol. 3, no. 1, pp. 1–4, 2019.

[9] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ICA to multivariate components," *International conference on independent component analysis and signal separation*, pp. 165–172, 2006.

[10] T. Adali, Y. Levin-Schwartz, and V. D. Calhoun, "Multimodal Data Fusion Using Source Separation: Two Effective Models Based on ICA and IVA and Their Properties," *Proceedings of the IEEE*, vol. 103, no. 9, 2015.

[11] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[12] R. A. Harshman, "PARAFAC2: Mathematical and technical notes," *UCLA working papers in phonetics*, vol. 22, no. 10, pp. 30–44, 1972.

[13] P. A. Chew, B. W. Bader, et al., "Cross-language information retrieval using PARAFAC2," in *KDD '07: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2007, pp. 143–152.

[14] K. H. Madsen, N. W. Churchill, and M. Mørup, "Quantifying functional connectivity in multi-subject fmri data using component models," *Human brain mapping*, vol. 38, no. 2, pp. 882–899, 2017.

[15] M. Roald, S. Bhinge, et al., "Tracing network evolution using the PARAFAC2 model," in *ICASSP'20: IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020, pp. 1100–1104.

[16] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

[17] J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[18] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," *Information Fusion*, vol. 19, no. 1, pp. 4–19, 2014.

[19] V. D. Calhoun, T. Adali, et al., "A method for multitask fMRI data fusion applied to schizophrenia," *Human Brain Mapping*, vol. 27, no. 7, pp. 598–610, 2006.

[20] Y. Levin-Schwartz, V. D. Calhoun, and T. Adali, "Quantifying the Interaction and Contribution of Multiple Datasets in Fusion: Application to the Detection of Schizophrenia," *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1385–1395, 2017.

[21] M. Anderson, T. Adali, and X.-L. Li, "Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1672 –1683, 2012.

[22] R. L. Gollub, J. M. Shoemaker, et al., "The MCIC collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia," *Neuroinformatics*, vol. 11, no. 3, pp. 367–388, 2013.

[23] G. Zhou, Q. Zhao, et al., "Linked Component Analysis from Matrices to High-Order Tensors: Applications to Biomedical Data," *Proceedings of the IEEE*, vol. 104, no. 2, pp. 310–331, 2016.

[24] T. Adali, M. Anderson, and G. S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 18–33, 2014.

[25] H. A. Kiers, J. M. Ten Berge, and R. Bro, "PARAFAC2 - Part I. A direct fitting algorithm for the PARAFAC2 model," *Journal of Chemometrics*, vol. 13, no. 3-4, pp. 275–294, 1999.

[26] Q. Long, C. Jia, et al., "Consistent run selection for independent component analysis: Application to fmri analysis," in *ICASSP'18: IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018, pp. 2581–2585.

[27] J. Kossaifi, Y. Panagakis, et al., "Tensorly: Tensor learning in python," *Journal of Machine Learning Research*, vol. 20, no. 26, pp. 1–6, 2019.

[28] Members and collaborators of the Wellcome Centre for Human Neuroimaging, *Statistical Parametric Mapping Toolbox: SPM12*, 2020.

[29] M.-L. Hu, X.-F. Zong, et al., "A review of the functional and anatomical default mode network in schizophrenia," *Neuroscience bulletin*, vol. 33, no. 1, pp. 73–84, 2017.

[30] W. Du, V. D. Calhoun, et al., "High classification accuracy for schizophrenia with rest and task fmri data," *Frontiers in human neuroscience*, vol. 6, pp. 145, 2012.

[31] S. Whitfield-Gabrieli, H. W. Thermenos, et al., "Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia," *Proceedings of the National Academy of Sciences*, vol. 106, no. 4, pp. 1279–1284, 2009.