Blind Source Separation for Fusion of Medical Imaging Data

Yuri Levin-Schwartz¹, Advisor: Tülay Adalı¹ ¹Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250 USA

Overview

The ability of different sensors to provide complementary views of complex systems has driven the collection of data from multiple sources for use at the same time. The fields where such data is exploited include: weather and financial analyses, video surveillance, and biomedical imaging. This data can be either from sensors of the same type, referred to as multiset data, or from sensors of different types, as in multimodal data. In both cases, full utilization of the common information across datasets requires a joint analysis. However, in many applications, little *a priori* information about the true latent sources is available. This stimulates the use of datadriven blind source separation methods, such as: independent component analysis (ICA) [1], joint ICA [2], [3], and group ICA [4], which minimize the assumptions placed on the data through the use of simple generative models, for the analysis of multiset and multimodal data. The success of these methods depends on the validity of their modeling assumptions, and care must be taken before applying a method to a particular problem. Our current research has been the development of novel data-driven methods for performing exploratory analyses on multiset as well as multimodal data and the determination of the effects of modeling assumptions on the fusion of medical imaging data.

CURRENT RESULTS

Order Selection for Multimodal Data, See reference: [5]

Since most medical imaging data is of high dimension and quite noisy, reliable determination of a signal subspace, or order, is critical in order to avoid over-fitting in the model. However, most order estimation techniques are only valid for a single dataset, for example [6], and almost none are designed for the sample-poor regime inherent to multimodal data fusion. In addition, the potential success of any data fusion method is dependent on commonalities shared across datasets, making it desirable to determine the strength of these commonalities prior performing fusion. In this work, we

- Developed a method to jointly estimate the order of multimodal datasets as well as their commonalities named principal component analysis and canonical correlation analysis (PCA-CCA);
- Demonstrated that PCA-CCA has superior performance to traditional methods on simulated data;

- Applied PCA-CCA to pairwise combinations of functional magnetic resonance imaging (fMRI), structural magnetic resonance imaging (sMRI), and electroencephalogram (EEG) data drawn from 14 patients with schizophrenia and 22 healthy controls;
- Determined that the sMRI and EEG datasets share the least commonality, whereas the fMRI and sMRI datasets share the most;
- Showed that the level of commonality obtained by PCA-CCA is predictive of the degree of significance found for components generated using canonical correlation analysis as shown in the Table I.

TABLE I: Number of significant components at two significance thresholds and PCA-CCA estimated common components, for the three pairwise combinations of modalities.

	fMRI-sMRI	fMRI-EEG	sMRI-EEG
Significant components, $p = 0.05$	2	1	1
Significant components, $p = 0.1$	5	4	2
Common components, PCA-CCA	4	3	2

Transposed Independent Vector Analysis for Data Fusion, See references: [7]–[10]

Though flexible and able to minimize the assumptions placed on the data, the success of data-driven blind source separation methods, such as independent vector analysis (IVA) [11], is intimately tied to the legitimacy of their modeling assumptions. This is particularly true for methods used for the fusion of neural imaging data, since improperly applied techniques may lead to incorrect conclusions and thus hinder the understanding of neural processes. This stimulates the development of novel blind source separation methods that can alleviate the limitations of previous multivariate methods and the exploration of the effects of modeling assumptions on the result. To this end, we

- Developed a novel method, transposed independent vector analysis (tIVA), for the fusion of multimodal neurological data and exploratory data analysis;
- Applied tIVA to fMRI data, sMRI data, and EEG data drawn from 14 patients with schizophrenia and 22 healthy controls;
- Demonstrated, through simulations, the importance of algorithm choice and the stability to model mismatch of tIVA compared to the popular fusion method, joint independent component analysis;

- Compared the results of univariate analyses with multivariate analyses to explore the interaction of different datasets in the context of data fusion and determine their contribution to the final result;
- Showed that the differences in the results when combining the fMRI and EEG datasets compared with the combination of all three datasets is minimal; highlighting that the sMRI contributes little to the final result and supporting the conclusions found using PCA-CCA;
- Compared the performance of tIVA with that of spatial IVA (sIVA) and individual ICAs applied to multitask fMRI data derived from 121 patients with schizophrenia and 150 healthy controls during the performance of three tasks: auditory oddball (AOD), Sternberg item recognition paradigm (SIRP), and sensorimotor task (SM);
- Showed using global difference maps (GDMs) that though sIVA has, in general, higher spatial variability than tIVA, tIVA appears more sensitive to group differences, as shown in Figure 1.



Fig. 1: GDMs for the AOD, SIRP, and SM tasks using the methods ICA, sIVA, and tIVA. The GDMs for the same method are in the columns, while the GDMs for the same dataset across methods are in the rows. These spatial maps correspond to *z*-maps thresholded at z = 2.7, where red and orange represent an increase in activation for controls versus patients and blue represent an increase in activation in patients over controls. Note that the *p*-value associated with each GDM, which assesses the significance of the decomposition, is shown above the corresponding spatial map.

FUTURE WORK AND EXPECTED RESULTS

Quantifying the Value of Datasets and of Fusion

The extraction of information from multiple sets of data is a problem inherent to many fields, thus spurring the development of a variety of methods to achieve this goal. Such techniques can fall into one of two main categories: data fusion, where the datasets are analyzed jointly, or data integration, where the datasets are analyzed separately and the results are combined. However, the selection of the optimal method is dependent on the relationships between the datasets, information that is very hard to obtain *a priori*. In order to determine the effect of different modeling assumption on the final result, multitask fMRI data drawn from 121 patients with schizophrenia and 151 healthy controls during the performance of three tasks, two auditory and one visual, is analyzed. In this work, we expect to show that

- The contribution of each task to the final result can be quantified by comparing the results when it is included in the analysis to the results where the dataset is not included in the analysis;
- There is an improvement when datasets are combined with data fusion compared to when datasets are combined through integration;
- Each task contributes differently to the final result and the gain by combining two auditory tasks should be smaller than combining a visual task with an auditory task.

References

- M. J. Mckeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, "Analysis of fMRI Data by Blind Separation Into Independent Spatial Components," *Human Brain Mapping*, vol. 6, pp. 160–188, 1998.
- [2] M. Svensén, F. Kruggel, and H. Benali, "ICA of fMRI group study data," *NeuroImage*, vol. 16, no. 3, pp. 551–563, 2002.
- [3] V. D. Calhoun, T. Adalı, 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," *Human Brain Mapping*, vol. 27, no. 1, pp. 47–62, 2006.
- [4] V. D. Calhoun, T. Adalı, G. D. Pearlson, and J. J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Human Brain Mapping*, vol. 14, no. 3, pp. 140– 151, 2001.
- [5] Y. Levin-Schwartz, Y. Song, P. J. Schreier, V. D. Calhoun, and T. Adalı, "Sample-poor estimation of order and common signal subspace with application to fusion of medical imaging data," *NeuroImage*, vol. 134, pp. 486–493, 2016.
- [6] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech and Signal Process.*, vol. 33, no. 2, pp. 387–392, April 1985.
- [7] Y. Levin-Schwartz, V. D. Calhoun, and T. Adalı, "Data-driven fusion of EEG, functional and structural MRI: A comparison of two models," in *48th Annual Conference on Information Sciences and Systems (CISS)*, March 2014, pp. 1–6.
- [8] T. Adalı, Y. Levin-Schwartz, and V. D. Calhoun, "Multi-modal data fusion using source separation: Two effective models based on ICA and IVA and their properties," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1478–1493, 2015.
- [9] —, "Multi-modal data fusion using source separation: Application to medical imaging," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1494– 1506, 2015.
- [10] Y. Levin-Schwartz, V. D. Calhoun, and T. Adalı, "Two models for fusion of medical imaging data: Comparison and connections," in 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), (Accepted), 2017.
- [11] J.-H. Lee, T.-W. Lee, F. A. Jolesz, and S.-S. Yoo, "Independent vector analysis (IVA): Multivariate approach for fMRI group study," *NeuroImage*, vol. 40, no. 1, pp. 86–109, 2008.