Joint Blind Source Separation for Multi-modality Data Fusion through Subject Co-variations

Matthew Anderson*, **Nicolle Correa***, **Vince D. Calhoun**†, **and Tülay Adalı***[1]
*University of Maryland, Baltimore County, Dept. of CSEE, Baltimore, MD 21250
†The Mind Research Network, Albuquerque, NM 87131

Analysis of multiple sets of data usually of different type or nature as in multi-modality data, is inherent to many problems in computer science and engineering. Biomedical image analysis is an important one among those and is a particularly challenging one because of the rich nature of the data made available by different imaging modalities. Many biomedical studies collect multiple data-sets, such as functional magnetic resonance imaging (fMRI), electroencephalography (EEG), structural MRI (sMRI), and genetic data in addition to behavioral data and other subject-based assessment parameters. Efficient use of all this information for inference, while minimizing assumptions made about the underlying nature of the data and relationships, is an arduous task, but is one that promises significant gains in challenging and important problems, such as in the understanding of the human brain function. The need to minimize strong modeling assumptions is especially evident when studying brain function in natural states such as rest or sleep, or when performing studying naturalistic tasks such as driving.

Data-driven methods such as blind source separation (BSS), independent component analysis (ICA), and canonical correlation analysis (CCA) that minimize the modeling assumptions on the data and the underlying processes are particularly attractive in this context as they can achieve useful decompositions of the multi-modal or multi-set data without strong modeling assumptions, and can also incorporate reliable prior information whenever available. We have recently introduced the use of multi-dataset canonical correlation analysis (MCCA) [4] for fusion of multi-modality imaging data through subject covariations and showed that the approach is very flexible and easily extendable to application to a wide variety of problems. In addition, we showed that the statistical significance of the difference between patient populations and controls significantly increased when we increased the number of modalities (data types) included in the analysis.

As an example, consider the fusion of three brain imaging modalities: fMRI, structural MRI (sMRI), and EEG. Typically the data acquired through these imaging techniques are analyzed separately. Each of these three modalities records brain structure and function at different scales, and fusing information from such complementary modalities promises to provide additional insight into connectivity across brain networks and changes due to disease. Given data from these three modalities collected on the same subjects, we decompose the data into sets of components and their corresponding modulation profiles across the subjects as shown in Figure 1 (*Left*). The data fusion scheme uses multi-set CCA [4] to determine the transformed co-ordinate system that maximizes the inter-subject covariations across the three modalities, and based on these covariations, we determine the associations among the components across modalities.

In [3], this scheme is applied to fusion of two different types of data and extended to three-way fusion in [2] using data collected from patients diagnosed with schizophrenia and healthy controls. Our results identify changes in the motor and temporal areas associated with the N2/P3 complex in the event related potential (the EEG feature) as shown in Figure 1 (*Right*), all of which have been well known to be affected in schizophrenia [6]. In addition, we noted that the statistical significance of the difference between healthy controls and patients increased significantly when we used three modalities instead of two in the analysis confirming our expectation that increased number of modalities do help identify more discriminative features increasing the overall sensitivity of the analysis.

Though effective, MCCA only takes second-order correlation in the multiple datasets into account. In this contribution, we show that joint BSS, *i.e.*, namely the identification of the dependent sources across the datasets, can be effectively solved by using mutual information as the cost function. In this way, one can take higher-order correlation into account as in independent vector analysis (IVA) [5]—however without assuming uncorrelated sources within a given source component vector, which is a limiting assumption for many applications. We show that when the multivariate Gaussian distribution is used as the prior for

[1]Presenter

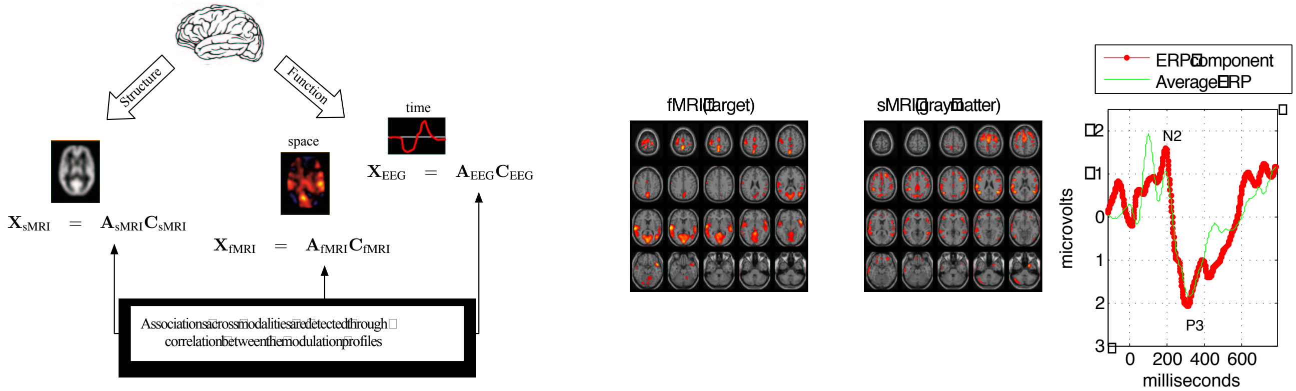

Figure 1: *Left:* MCCA for fusion of data from multiple imaging modalities. *Right:* Set of associated components estimated by MCCA which showed significantly different loading for patients versus controls.

the source component vector as in [1], the resulting cost function is equivalent to one of the cost functions used in MCCA [4], which proposes a number of ad-hoc cost functions for the problem. In addition, we show that effective Newton and quasi-Newton learning procedures can be derived for the problem using a decoupling approach for the weight demixing matrix that alleviates the need to constrain the demixing matrix to be orthogonal. In addition, we present local stability conditions and the identifiability conditions for the problem and give examples to demonstrate the new algorithms provide reliable and superior performance compared to MCCA and other alternatives to JBSS. In addition, we show how the approach can be generalized to handle datasets with different dimensions and hence can be used for multi-set data fusion.

References

[1] M. Anderson, T. Adalı, and X.-L. Li. Nonorthogonal independent vector analysis using multivariate Gaussian model. In *Proc. 2010 Conf. Latent Variable Analysis and Independent Component Analysis (LVA/ICA)*, St. Malo, France, Sept. 2010.

[2] N. Correa, T. Adalı, and V. D. Calhoun. Canonical correlation analysis for data fusion and group inferences: Examining applications of medical imaging data. *IEEE Signal Processing Magazine*, 27(4):39–50, July 2010.

[3] N. M. Correa, Y.-O Li, T. Adalı, and V. D. Calhoun. Canonical correlation analysis for feature-based fusion of biomedical imaging modalities and its application to detection of associative networks in schizophrenia. *IEEE J. Selected Topics in Signal Processing*, 2(6):998–1007, 2009.

[4] J. R. Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58:433–451, 1971.

[5] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee. Blind source separation exploiting higher-order frequency dependencies. *IEEE Trans. Audio Speech Lang. Process.*, 15(1):70–79, Jan. 2007.

[6] A. A. Stevens, P. S. Goldman-Rakic, J. C. Gore, R. K. Fulbright, and B. E. Wexler. Cortical dysfunction in schizophrenia during auditory word and tone working memory demonstrated by functional magnetic resonance imaging. *Arch. Gen. Psychiatry*, 55:1097–1103, 1998.

Topic: Learning algorithms/Other applications
Preference: Oral/Poster