Sensor Applications

ICA and IVA for Data Fusion: An Overview and a New Approach Based on Disjoint Subspaces

Tulay Adali[1*], M.A.B.S. Akhonda[1], and Vince D. Calhoun[2*]

[1] Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, MD
[2] University of New Mexico and the Mind Research Network, Albuquerque, NM
[*] Fellow, IEEE

Manuscript received Sep. 14, 2018; revised XXXX, 2018; accepted XXXX, 2018. Date of publication XXXX, 2018; date of current version XXXX, 2018.

Abstract—Data-driven methods have been very attractive for fusion of both multiset and multimodal data, in particular using matrix factorizations based on independent component analysis (ICA) and its extension to multiple datasets, independent vector analysis (IVA). This is primarily due to the fact that independence enables (essentially) unique decompositions under very general conditions and for a large class of signals, and *independent components* lend themselves to easier interpretation. In this paper, we first present a framework that provides a common umbrella to previously introduced fusion methods based on ICA and IVA, and allows us to clearly demonstrate the tradeoffs involved in the design of these approaches. This then motivates the introduction of a new approach for fusion, that of *disjoint subspaces* (DS). We demonstrate the desired performance of DS using ICA through simulations as well as application to real data, for fusion of multi-modal medical imaging data—functional magnetic resonance imaging (fMRI),and electroencephalography (EEG) data collected from a group of healthy controls and patients with schizophrenia performing an auditory oddball task.

Index Terms—Data fusion, independent component analysis, fMRI, EEG, multimodality.

I. INTRODUCTION

Information about a phenomenon or a system of interest can be obtained through various types of instruments, experimental setups, and sources. We refer to a related set of measurements as a dataset, and our focus in this paper is the *fusion of multiple datasets, i.e,*, extraction of multivariate *interpretable* features from multiple related datasets that can be used for classification, prediction, detection, and change analysis, among other tasks.

Since typically very little is known about the interaction of multiple datasets, data-driven methods based on matrix decompositions have been the natural solution. Among those, ICA has proven especially attractive and is our focus in this short overview along with its generalization to multiple datasets, IVA. We first introduce a general umbrella that includes previously introduced methods for data fusion based on ICA and IVA, and allows us to classify those based on two key properties, level of dataset interaction and implementation flexibility. We then use this general umbrella to introduce an effective new approach, *disjoint subspaces* (DS) that addresses limitations of previous methods, and present comparison studies to highlight the desirable properties of the new approach along with the tradeoffs across these methods.

In the next section, we formalize our definition of fusion by first making an important distinction among multiple datasets, those that are multiset and multimodal, and then formally introduce the key concepts of interpretability, uniqueness and diversity. Then in Section III, we introduce our general framework for data fusion and describe how previous methods fall under this umbrella, discuss their properties and introduce the DS approach.

Corresponding author: T. Adali (e-mail: adali@umbc.edu).

This work was supported by the NSF under Grants No. NSF-CCF 1618551 and NSF-NCS 1631838.
Digital Object Identifier 10.1109/LSENS.2018.2884775

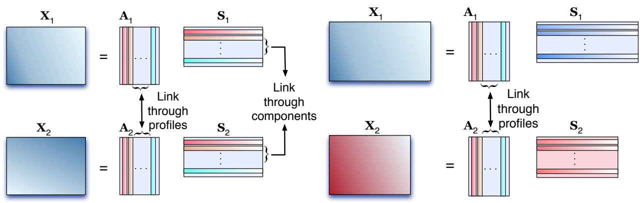
II. PRELIMINARIES

In this section, we present key definitions that provide the necessary background and form the basis of the general fusion framework we later introduce.

Multiset data refers to multiple sets of data that are all collected using the same modality at different conditions, observation times, using multiple experiments, tasks, or subjects. Hence such datasets are all of the same type, resolution, and dimension. Examples include a single modality data such as fMRI data collected from different subjects or at different time points, or multispectral images from different color channels. On the other hand, *multimodal data*, refers to information collected about the same phenomenon using different types of modalities or sensors, where the modalities provide complementary information. Examples include medical imaging data such as EEG and fMRI data that both measure the brain function, EEG by recording electrical activity through electrodes placed on the scalp and fMRI by imaging the brain hemodynamic response.

We define *data fusion* as the analysis of multimodal or multiset data where all datasets are allowed to *fully interact and inform each other*. Hence, we distinguish fusion from an integration type approach where there are separate processing chains for each dataset, and at the end, they are brought together so that one informs the other, which is also referred to as *late fusion*. In our case, we emphasize the fact that the datasets are formed such that the variability in the data is maximally preserved.

We form the datasets by stacking N, each V_k dimensional measurements or multivariate features vertically to yield $\mathbf{X}_k \in \mathbb{R}^{N \times V_k}$, $k, \ldots K$. For the multiset scenario, we have $V_k = V$ for all k, while for the multimodal one, this is not typically the case. However, in both cases we assume that there is a common dimension across which the datasets can be linked. Then the main assumption is that the datasets are generated as mixture of N components—referred to as



(a) Multiset data (b) Multimodal data

Fig. 1. Associations for fusion of two types of data

sources in blind source separation—given by the rows of matrix $\mathbf{S}_k$ and linearly mixed through $\mathbf{A}_k$

$$\mathbf{X}_k = \mathbf{A}_k \mathbf{S}_k \text{ for } \quad k = 1, \ldots, K. \tag{1}$$

We refer to the columns of matrix $\mathbf{A}$ as *profiles* as they quantify the contributions to the overall mixture, $\mathbf{X}_k$. The goal is, given $\mathbf{X}_k$, to identify these two factors—matrices—and to interpret them, *i.e.*, to attach a physical meaning to them. A simple example is in spatial fMRI analysis where the components correspond to functional networks and the corresponding column of the mixing matrix to their temporal modulation, which also provides a good model match and hence produced results that have been especially useful [1].

An important consequence of the requirement on interpretability is that we require the decomposition to be *essentially unique* by which we mean that the two matrices in the decomposition are unique up to a common permutation and scaling/counter-scaling of the columns. It is easy to see that, in general, matrix decompositions are highly nonunique, and with only additional assumptions such as sparsity nonnegativity, and orthogonality, one can guarantee (essential) uniqueness. We define any property or constraint that enables a unique decomposition as *diversity*. A very useful type of diversity is independence as it enables unique solutions for a very general class of signals including multiple Gaussians [2]. Given the linear generative model in (1), ICA ($K = 1$) recovers the sources by estimating the demixing matrix $\mathbf{W}$ such that we have $\mathbf{U} = \mathbf{W}\mathbf{X}$ where $\mathbf{U}$ includes the source estimates. A direct measure of independence is mutual information rate, which enables one to account for multiple statistical properties of the sources simultaneously, such as non-Gaussianity, non-whiteness (sample dependence), non-stationarity, and non-circularity [2].

III. A GENERAL FRAMEWORK FOR DATA FUSION USING ICA AND IVA, AND A NEW METHOD

Given multiple datasets with the linear generative model in (1), we perform data fusion using matrix decomposition by establishing associations across the datasets in two ways, either through the components and/or the profiles as shown in Figure 1. We refer to these associations as *links* across the datasets, which can either be *hard* where we assume that a given set of components and/or profiles are common—*i.e.*, exactly shared—across the datasets, or *soft* which implies that the association is statistical, *i.e.*, that the components/profiles are statistically dependent. As shown in Fig. 1(a), for multiset data, we have the option of establishing links both through the components and the profiles. However, in the case of multimodal data, since the components are of different nature, we can neither assume that they can be exactly shared nor are dependent across the datasets. Thus, the only possibility is linking them through profiles as in Fig. 1(b), covariations across, *e.g.*, subjects or samples.

A. ICA and IVA for Data Fusion

We can study most ICA/IVA based fusion methods under the general umbrella in Fig. 1. For multiset data, the most widely used solutions are Group ICA [1] and, more recently, IVA [2],[3]. Given K datasets, Group ICA vertically concatenates the datasets and performs dimension reduction using PCA. Then, a single ICA is performed on this group subspace, hence providing a hard link of common group subspace across the datasets. Individual dataset estimates are then back-reconstructed. IVA, on the other hand, identifies components that are maximally independent within each dataset but are dependent across the datasets. This is achieved through the definition of a source component vector (SCV) of dimensionality K obtained by concatenating corresponding sources from each dataset. The soft link across the datasets is then established by using a suitable multivariate probability density function (pdf) for the SCV. IVA is then achieved, by minimizing mutual information rate like ICA, but among the SCVs rather than individual sources. The IVA formulation reduces to ICA for a single dataset.

For multimodal data, joint ICA (jICA) [4] concatenates multiple datasets horizontally, and performs a single ICA, assuming that the mixing matrix, *i.e.*, all profiles are shared across all K datasets, thus establishing hard links across the datasets. Parallel ICA(pICA) [5], on the other hand, relaxes the common mixing matrix assumption by jointly performing two separate ICAs while maximizing correlation of fixed number of selected profiles at each iteration through an augmented cost that takes both independence and profile correlation into account. The resulting components are not as independent as in a single dataset ICA decomposition and level of correlation and the number of profiles are parameters that need to be determined.

A more direct approach to transition from hard links to soft links is by simply transposing the model in (1) and using the *transposed* IVA model (tIVA) [6]. While it appears to be the most natural way to link multimodal data, a main disadvantage is that the number of observations now become observation samples and hence to be able to reliably compute higher-order statistics, one needs a significant number of samples. A new data fusion method, consecutive independence and correlation transform (C-ICT) [7], introduced for two datasets like pICA, alleviates the main limitations of jICA and pICA. It is a two step hybrid model based on ICA and CCA to factor and fuse multimodal data. In the initial step, C-ICT performs individual ICAs on each dataset separately thus yielding maximally independent component estimates $\mathbf{U}_k$ for each modality and corresponding $\hat{\mathbf{A}}_k$ for $k = 1, 2$. In the second step, the two sets of results are linked by using CCA by maximizing correlation between the columns of the two estimated $\hat{\mathbf{A}}_k$ matrices. The use of CCA after a separate ICA decomposition, allows extraction of a different number of components for each modality and use of different ICA algorithms for each dataset.

There are constraints imposed by each method and the algorithm employed for the decomposition, discussed in detail in [6] for jICA, IVA, and tIVA. Here we introduce two key tradeoffs for fusion, flexibility in terms of choice of different algorithms and order as well as the level of interaction among the datasets. The first one, flexibility in implementation allows one to pick the most appropriate algorithm for each dataset, and to select the best order for the decomposition. The selection of order refers to identification of the signal subspace size, which is assumed to be less than the number of observations and plays a very important role in final performance [2]. Level of interaction among the datasets is another important factor determining how strongly they influence and inform each other.

Hard links usually enable stronger interaction across the datasets, and approaches in order of decreasing level of interaction are jICA, Group ICA, IVA, tIVA, pICA, and C-ICT. As associations in the component dimension is expected to be stronger, both because of the availability of higher number of samples and richness of the distribution in that dimension, IVA provides stronger links than tIVA. On the other hand, in terms of implementation flexibility C-ICT is the most flexible one, allowing choice of different algorithms and orders, followed by pICA which allows for flexibility in the choice of order, and then IVA, Group ICA, tIVA (all three offering the same level of flexibility), and finally jICA. These tradeoffs motivate the introduction of a new method, which we discuss next. Obviously all methods introduced for multimodal data can be used for multiset fusion as well, and this is also the case for DS, but here, we emphasize its use in multimodal fusion as that is the focus of the special issue.

B. Disjoint Subspaces (DS) for Fusion

The main idea for DS is to identify and split the common and distinct subspaces from the modalities and perform separate analyses. We determine the subspace *common* to both datasets where hard links are justified and use jICA in this common subspace, and separate ICAs in the distinct subspaces. Hence, one can easily choose the most appropriate algorithm and order for each case. The approach builds on PCA-CCA [8], which makes use of consecutive steps of PCA and CCA to identify the order of common and distinct subspaces with small sample support, important in multimodal fusion as the associations in this case are in the sample dimension, N. Given the orders, common and distinct subspaces are then split, and analyzed independently using decomposition methods like ICA, where for the common subspace jICA that assumes shared mixing matrix becomes a very suitable candidate.

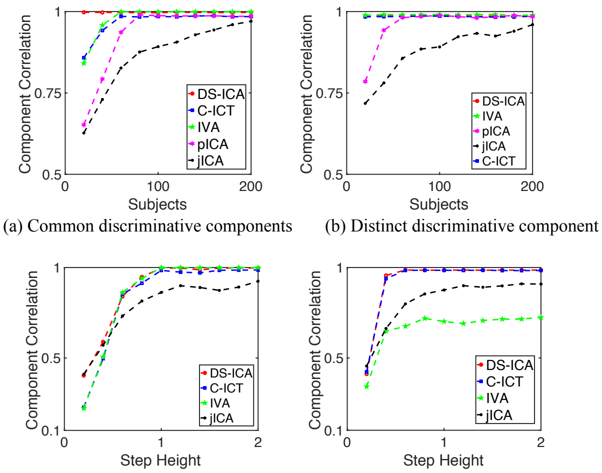
Given $\mathbf{X}_k$, $k = 1, 2$, we perform CCA along the common dimension N, given the common order C determined using PCA-CCA, first C canonical coefficients are used to project the datasets into a space where association between those datasets is maximized. Similarly, we can use the $N-C$ canonical coefficients to project and backreconstruct two datasets between which the associations are minimum. We call these subspaces $\mathbf{X}_{k,D} \in \mathbb{R}^{N-C \times V_k}$ distinct and the common subspaces $X_{k,C} \in \mathbb{R}^{C \times V_k}$, for $k = 1, 2$.

In a second step, the common parts can be analyzed using jICA as in $[\mathbf{X}_{1C}\mathbf{X}_{2C}] = \mathbf{A}[\mathbf{S}_{1C}\mathbf{S}_{2C}]$ and distinct parts can be analyzed using separate ICAs. Using jICA only for the common part provides a good match to the strong common mixing assumption of jICA and yields sources that are strongly associated across the datasets. Separate analyses on the distinct parts on the other hand, yields sources that are distinct to each modality.

IV. RESULTS

A. Simulation Results

In the comparison, we include jICA, pICA, IVA, C-ICT, all four methods discussed in Section III and DS using ICA, which we call DS-ICA. Apart from pICA, which is developed for Infomax, entropy bound minimization (EBM) [9] is used for the other methods as it provides a much more flexible matching of the component pdf.



(a) Common discriminative components (b) Distinct discriminative component

(c) Common discriminative components (d) Distinct discriminative component

Fig. 2. Performance comparison for component estimation

However to be fair to pICA, the components are generated from a Laplacian distribution, for which Infomax provides a good match, and for IVA, we use the version that assumes a Laplacian pdf and takes all-order statistics into account [2]. We generate $N = 10$ sources each with 1000 independent and identically distributed (i.i.d.) samples for two datasets where $C = 3$ sources are correlated across the datasets with correlation values 0.9, 0.7 and 0.5. Latent sources are linearly mixed using mixing matrices $\mathbf{A}_k \in \mathbb{R}^{M \times 10}$, where M is the number of subjects, and the datasets are reduced to 10, the true order, for all methods. Three columns of the mixing matrices are used to simulate the group difference through a step type response, *i.e.*, a difference in the mean value with additive standard Gaussian noise, one pair as common and a single one distinct in the second dataset. We refer to the components corresponding to these profiles as *discriminative components* since they can be identified through a t-test.

We evaluate performance by changing first the number of subjects—20 to 200 with a step height fixed to 1.5—and then the step-height introduced for distinguishing the two groups in the range [0.2, 2] resulting in correlation values $[0.2, 0.8]$ while keeping the subject count fixed at 50. Results are averaged over 100 runs and shown in Fig. 2. For the first set of simulations, shown in Fig. 2(a) and Fig. 2(b), performance of all methods improves as the number of subjects increase, with C-ICT trailing behind as it allows limited interaction among the datasets. Another approach that does not let datasets fully interact, pICA also exhibits limited performance especially with small number of subjects. DS-ICA makes most efficient use of the data as it performs a joint ICA for the common order of three following a PCA-CCA step to determine the order and provides the best performance. In the second group shown in Fig. 2(c) and Fig. 2(d), as the profile correlation increases with increasing step height, the link becomes stronger and the performance of all algorithms in general improve. Since in this scenario where the association through profiles keeps increasing is a good match for multimodal data, C-ICT designed for the multimodal case performs quite robustly except for very low correlation values as in this case the CCA step might not correctly identify the index of the correlated component, and IVA fails for distinct component estimation primarily because the example challenges one of the assumptions of the IVA model, that the mixing matrices are all full rank as it estimates two separate mixing matrices. DS-ICA provides again the best overall performance.

B. Fusion of fMRI and EEG

We use EEG and fMRI data collected from 38 subjects—16 patients with schizophrenia and 22 healthy controls—performing an auditory odd ball task. The multivariate features used for the study are the task related spatial maps for fMRI and the event related potentials (ERP) calculated from each subject's EEG data from the CZ (midline central position) channel by averaging small window around the target tone across repeated instances of the task. The resulting matrices are $\mathbf{X}_{\text{EEG}} \in \mathbb{R}^{38\times45100}$ and $\mathbf{X}_{\text{fMRI}} \in \mathbb{R}^{38\times60186}$.

Performances are evaluated using jICA, pICA, C-ICT, and DS-ICA. The IVA-based solution for multimodal data, tIVA, is excluded as the sample size of 38 is extremely low for evaluation of higher-order statistics. The ICA algorithm used for jICA, C-ICT and DS-ICA is EBM [9] and for pICA, Infomax [10]. As both EBM and Infomax are of iterative type, cross intersymbol interference [11] is used to select the most consistent estimate, and the entropy rate based order selection scheme [12] is used to determine the order of the signal subspace, and an order $N = 20$ is selected for jICA and 12 for fMRI and 20 for EEG data for the other methods. A two-sample t-test is used to identify the profiles and corresponding components with group difference ($p < 0.05$) and are shown in Fig. 3. The fMRI components are thresholded at $Z = 2$ where color red, orange and yellow indicates higher activation in healthy controls than patients and color blue means decrease in controls than patients.

All four methods identify discriminating components, putative *biomarkers of disease*, and have general agreements in terms of areas indicated in fMRI and ERP characteristics. jICA and pICA estimate one component from each modality, while C-ICT captures two and DSF-jICA captures two fMRI and one EEG component. ERP components estimated by all methods reports peaks around N2-P3 complex, correlating with other studies that indicated changes in motor execution region associated with N2 as well as N2-P3 in patients with schizophrenia [6]. The fMRI component estimated by jICA shows higher activation in sensory motor and auditory region while the fMRI component in pICA shows activation in part of sensory motor area, which is similar to the first component estimated by C-ICT. First fMRI component estimated by DS-ICA shows activation similar to the second C-ICT component, in the auditory and sensory motor region and second one has activation in part of visual and sensory motor region. Overall, components estimated by DS-ICA have p values lower than other methods, better able to differentiate between two groups. In addition, the common and distinct subspace interpretation allows additional information that can be useful depending on the nature of the study.

V. DISCUSSION

We present a general umbrella for fusion methods that are based on ICA and IVA, and introduce DS-ICA, which provides a desirable balance in terms of different assumptions made by previous methods, in particular providing a good match to those of joint ICA. Even though here it is introduced for two datasets and using ICA as the main decomposition method, DS formulation can be used with other matrix decomposition methods, and extended to multiple datasets by generalizing PCA-CCA for multiple datasets. An attractive approach for this is using IVA with multivariate Gaussian model, which is equivalent to multiset canonical correlation analysis [2]. A main difficulty when selecting a given model in applications like fusion of medical imaging data is due to the lack of ground truth. A recent approach proposes the use of classification rate between patients and controls as an objective measure of performance to compare different models [13], however its utility might be limited with small sample size as in the example we presented. Thus, another important direction is development of methods for assessing performance of different fusion methods to provide guidance to the practitioner.

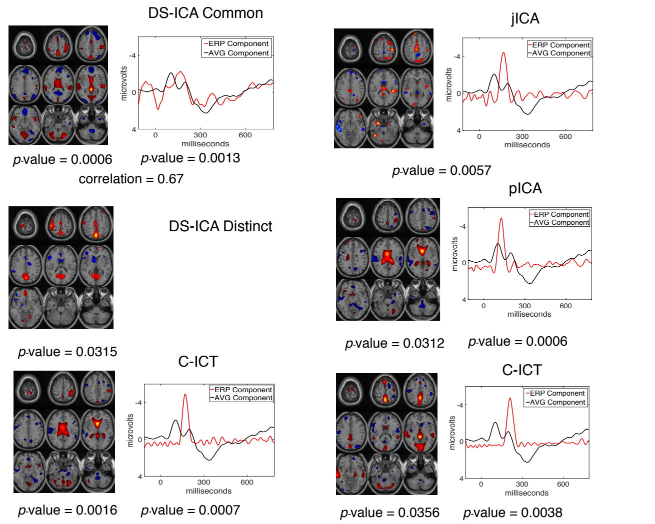


Fig. 3. Discriminating fMRI and ERP components identified using DS-ICA, jICA, pICA, and C-ICT

REFERENCES

[1] V. D. Calhoun and T. Adalı, "Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery," *IEEE Reviews in Biomedical Engineering*, vol. 5, pp. 60–73, Aug. 2012.

[2] T. Adalı, M. Anderson, and G.-S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Proc. Mag.*, vol. 31, no. 3, pp. 18–33, May 2014.

[3] T. Kim, I. Lee, and T.-W. Lee, "Independent vector analysis: Definition and algorithms," in *Proc. 40th Asilomar Conf. Signals, Systems, Comput.*, 2006, pp. 1393–1396.

[4] V. D. Calhoun, T. Adalı, G. D. Pearlson, and K. A. Kiehl, "Neuronal chronometry of target detection: Fusion of hemodynamic and event-related potential data," *NeuroImage*, vol. 30, no. 2, pp. 544–553, April 2006.

[5] J. Liu, G. Pearlson, A. Windemuth, G. Ruano, N. I. Perrone-Bizzozero, and V. D. Calhoun, "Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA," *Human Brain Mapping*, vol. 30, pp. 241–255, 2009.

[6] 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," *Proc. IEEE*, vol. 103, no. 9, pp. 1478–1493, Sep. 2015.

[7] M. A. B. S. Akhonda, Y. Levin-Schwartz, S. Bhinge, V. D. Calhoun, and T. Adalı, "Consecutive independence and correlation transform for multimodal fusion: application to EEG and fMRI data," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, Canada, April 2018.

[8] Y. Song, P. Schreier, D. Ramirez, and T. Hasija, "Canonical correlation analysis of high-dimensional data with very small sample support," *Signal Processing*, vol. 128, pp. 449–458, Nov. 2016.

[9] X.-L. Li and T. Adalı, "Independent component analysis by entropy bound minimization," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5151–5164, Oct. 2010.

[10] A. Bell and T. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995.

[11] Q. Long, C. Jia, Z. Boukouvalas, B. Gabrielson, D. Emge, and T. Adalı, "Consistent run selection for independent component analysis: Application to fMRI analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Calgary, Canada, April 2018.

[12] G.-S. Fu, M. Anderson, and T. Adalı, “Likelihood estimators for dependent samples and their application to order selection,” *IEEE Trans. Signal Processing*, vol. 62, no. 16, pp. 4237–4244, Aug. 2014.

[13] Y. Levin-Schwartz, V. D. Calhoun, and T. Adalı, “Quantifying the interaction and contribution of multiple datasets in fusion: Application to the detection of schizophrenia,” *IEEE Trans. Med. Imaging*, 2017.