Challenges in Multimodal Data Fusion

Dana Lahat, Tülay Adali, Christian Jutten

CHALLENGES IN MULTIMODAL DATA FUSION

Dana Lahat[1]*, Tülay Adalı*[2]*, and Christian Jutten*[1]

[1]GIPSA-Lab, UMR CNRS 5216
Grenoble Campus, 38400 Saint Martin d'Hères, France

[2]Department of CSEE, University of Maryland, Baltimore County
Baltimore, MD 21250, USA

ABSTRACT

In various disciplines, information about the same phenomenon can be acquired from different types of detectors, at different conditions, different observations times, in multiple experiments or subjects, etc. We use the term "modality" to denote each such type of acquisition framework. Due to the rich characteristics of natural phenomena, as well as of the environments in which they occur, it is rare that a single modality can provide complete knowledge of the phenomenon of interest. The increasing availability of several modalities at once introduces new degrees of freedom, which raise questions beyond those related to exploiting each modality separately. It is the aim of this paper to evoke and promote various challenges in multimodal data fusion at the conceptual level, without focusing on any specific model, method or application.

***Index Terms*—** Data fusion, multimodality

1. INTRODUCTION

Information about a phenomenon or a system of interest can be acquired using different types of instruments, measurement techniques, experimental setups, etc. Due to the rich characteristics of natural processes and environments, it is rare that a single acquisition method can provide complete understanding thereof. The increasing availability of multiple datasets that contain information, obtained using different acquisition methods, about the same system, introduces new degrees of freedom that raise questions beyond those related to exploiting each dataset separately. Despite the evident *potential* benefit, and massive work that has already been done in the field (see, for example, [1–7] and references therein), the knowledge of *how* to actually exploit the additional diversity that multiple datasets offer is currently at its very preliminary stages [1, 2].

Data fusion is a challenging task for several reasons. *First*, the data are generated by very complex systems (biological, environmental and psychological, among others), driven by numerous underlying processes that depend on a large number of variables to which we have no access. *Second*, due to the augmented diversity, the number, type and scope of new research questions that can be posed is potentially very large. *Third*, fitting together heterogeneous datasets such that the respective advantages of each dataset are maximally exploited, and drawbacks suppressed, is not an evident task. We elaborate on the second and third issues in Sec. 2 and Sec. 3, respectively. Most of these questions have been devised only in the very recent years, and, as we show in the sequel, only a fraction of their potential has already been exploited. Hence, we refer to them as "challenges".

A rather wide perspective on challenges in data fusion is presented in [1], which presents linked-mode decomposition models within the framework of chemometrics and psychometrics, and [2], which focuses on "automated decision making" with special attention to multisensor information fusion. In practice, however, challenges in data fusion are most often brought up within a framework dedicated to a specific application, model and dataset (examples will be given below).

In this paper, we bring together a comprehensive (but definitely not exhaustive) list of challenges in data fusion. We consider this list to be of interest to communities beyond signal processing. Following from [1, 2], and further emphasized by our discussion below, it is clear that at the appropriate level of abstraction, the same challenge in data fusion can be relevant to completely different and diverse applications, goals and data types. Consequently, a solution to a challenge that is based on a sufficiently model-free approach may turn out useful in very different domains. Therefore, there is an obvious interest in opening up the discussion of data fusion challenges to include and involve disparate communities, so that each community could inform the other. Our goal is to stimulate and evoke the relevance and importance of a perspective based on challenges to advanced data fusion. More specifically, we would like to promote data-driven approaches, that is, approaches with minimal and weak priors and constraints, such as sparsity, nonnegativity, and independence, among others, that can be applied to more than one specific application or dataset. Hence, we present these challenges in quite a general framework that is not specific to an application, goal or data type. We also give examples and motivations from different domains.

This work is supported by the project CHESS, 2012-ERC-AdG-320684 (D. Lahat and Ch. Jutten) and by the grant NSF-III 1017718 (T. Adalı). GIPSA-Lab is a partner of the LabEx PERSYVAL-Lab (ANR–11-LABX-0025).

In order to contain our discussion, we focus on datasets in which a phenomenon or a system is observed using multiple instruments. In this case, each acquisition framework is denoted as *modality* and the setup is known as *multimodal*. As already noted, "data fusion" is quite a diffuse concept, which takes different interpretations with applications and goals. Therefore, within the context of this paper, and in accordance with the types of problems on which we focus, our emphasis is on the following tighter interpretation [8]:

Data fusion is an approach to the analysis of multimodal data, in which different datasets can interact and inform each other. The latter terms will be given a more concrete meaning in Sec. 3.3.

Accordingly, we suggest the following *operative* definition for the special type of diversity that is associated with multimodality and data fusion:

Diversity (due to data fusion) is the property that allows to enhance the uses, benefits and insights (discussed in Sec. 2) in a way that cannot be achieved with a single modality.

2. WHY DO WE NEED MULTIMODALITY?

For living creatures, multimodality is a very natural concept. Living creatures use external and internal sensors, sometimes denoted as "senses", in order to detect and discriminate among signals coming from the environment, communicate, cross-validate, disambiguate, and add robustness to numerous life-and-death choices and responses that must be taken rapidly and in a dynamic and constantly changing internal and external environment. As a result, certain multimodal applications are based on imitating natural multimodality: the most prevalent is audio-video [9, 10]. Below, we list some of the prominent uses, benefits and insights that can be obtained from properly exploiting multimodal data, especially as opposed to single-set and uni-modal data.

2.1. Exploratory

Despite the well-accepted paradigm that certain natural processes and phenomena can express themselves under completely different physical guises (this is the *raison d'être* of multimodal data fusion), often very little is known about the underlying relationships between the modalities. Such is the case, for example, in neurological activity observed via electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) [11], and atrial fibrillation measured via surface and intra-cardiac electrodes [12]. In other cases, one would like to propose new modalities and combinations thereof [13]. Therefore, the most obvious and essential endeavour to be undertaken in any multimodal data analysis task is exploratory: to learn about relationship between modalities, their complementarity, common and modality-specific information content, among others. Applications are diverse: understanding brain functionality [6, 8], health monitoring [13], developing non-invasive medical diagnosis techniques [12], exploring the relationship between tasks for human-machine interaction (HMI) [10], and so forth. As an active initiative, we point out the yearly data fusion contest of the IEEE geoscience and remote sensing society (GRSS), which aims at investigating the potential use of various remote-sensing modalities: participants are encouraged to consider various open problems on multisensor data fusion, and to use the provided data sets to demonstrate novel and effective approaches to solve these problems [14].

2.2. Uniqueness, Identifiability and Disambiguation

Multimodality provides redundancy that can be exploited to resolve otherwise ill-posed problems. We illustrate this powerful property and its potential impact through examples. It is well known that statistically independent sources with real-valued Gaussian independent and identically distributed (i.i.d.) samples cannot be blindly separated from a single arbitrary linear instantaneous invertible mixture [15]. If *several* such mixtures are considered simultaneously, however, and certain statistical dependencies are allowed *across* mixtures (without changing the assumptions *within* each separate blind source separation (BSS) problem), then it has been proved that a unique and identifiable solution, within each mixture, up to the unavoidable scaling ambiguity, exists [16]. This model, when not restricted to Gaussian i.i.d. samples, is known as independent vector analysis (IVA) [17]. It has been shown that the IVA framework provides a fixed permutation to all the corresponding component estimates and thus significantly alleviates the permutation ambiguity problem inherent to classical BSS [16, 17]. In this example, multiple datasets are allowed to interact (in the sense of Sec. 3.3). This interaction provides yet another type of diversity [16], which results in identifiability of an otherwise unidentifiable model, and permutation disambiguation that so far did not have any closed-form solution. Another example is the EEG inverse problem, which is underdetermined: infinitely many spatial current patterns can give rise to identical measurements. An identifiable and unique solution can be obtained using spatial constraints from fMRI [6].

3. CHALLENGES

Thanks to recent advances, the *availability* of multimodal data is now a fact of life. The *acquisition* of multimodal data, nevertheless, is only a first step. In this section, we discuss some of the issues that should be addressed in the actual processing of multimodal data. These challenges may be partitioned into the following groups: *data*, *level of data fusion*, *model*, and *theoretical validation*.

3.1. Data

The first group of challenges are those *imposed* by the data. They can be partitioned into challenges at the *acquisition and observation level* and those due to various *uncertainties in the data*. Complicating factors of the first type *must* be accounted for, and working with features (Sec. 3.2), instead of raw data, is a possible remedy.

Non-commensurability: Different instruments are sensitive to different physical phenomena and consequently, report on different aspects of the problem. As a result, the data is represented by heterogeneous physical units. This situation, known as *non-commensurability*, is probably the principal and first obstacle to tackle. Examples include EEG, which measures the propagation of electrical fields generated by direct neurological activity, vs. fMRI, which measures induced changes in magnetization between oxygen-rich and oxygen-poor blood [6], and spectral content (hyperspectral imaging) vs. information about spatial (3D) geometry (LiDAR) [14].

Different resolutions: Datasets may share the same coordinates but at very disparate resolutions. One example is fusing EEG, which has excellent temporal but low spatial resolution, with fMRI, which has a fine spatial resolution but a very large integration time [6]. This is also the case in "pan-sharpening" [18] [4, Chapter 9]: merging a high-spatial low-spectral (monochromatic) resolution panchromatic and lower-spatial higher-spectral (four bands) resolution multispectral images, in order to generate a new synthetic image having both the higher spectral and spatial resolution of the two.

Number of dimensions (ways): Different acquisition methods may yield datasets with different structures. For example, matrices vs. higher-order tensors. This topic is further elaborated, with emphasis on applications in metabolomics and psychometrics, in [1, 19].

We now turn to discussing uncertainties in the data. Any real-world set of observations is prone to various uncertainties. The fact that the presence of heterogeneous datasets creates *new types of uncertainties* implies that these new uncertainties cannot (or should not) be treated by uni-modal, single-dataset tools. We argue that in such cases, it is the complementary quality of multimodal data that should be exploited to resolve these challenges.

Noise: Thermal noise, calibration imprecision (alternatively: finite precision), or any other quality degradation in the measurements is unavoidable and present in all datasets. For simplicity, we denote all these nuisance phenomena as "noise". Naturally, each measurement procedure produces not only heterogeneous types of desired data, but also different amounts and types of errors [1]. The question of how to *jointly* weigh or balance the *different* errors is discussed in a number of scenarios, see, e.g., [2, 20].

Missing data: Another uncertainty in the observations may be due to "missing data". This term may stand for various cases. *First*, certain samples can be unreliable, discarded or simply missing due to faulty detectors. *Second*, sometimes a modality can report only on part of the system (w.r.t. the other modalities), as with EEG and magnetoencephalography (MEG) [6] or nuclear magnetic resonance (NMR) and liquid chromatographymass spectrometry (LC-MS) [19]. *Third*, data may be regarded as structurally missing if samples at different modalities are not taken at comparable sampling points [1] and we would like to construct a more complete picture from the entire sample set. An approach to the missing data problem within coupled matrix and tensor factorization (CMTF) is discussed in [19].

Conflicting, contradicting or inconsistent data: Obviously, this problem can occur only in the presence of multiple datasets. If data is fused at the decision level (as is the case, for example, in fusion of different classification maps in remote sensing), then a decision or voting [1] rule may be applied. When only two datasets are confronted, more elaborate approaches may be requried. Other approaches, related to multisensor data fusion, are discussed in [2]. An obvious challenge is to devise a suitable compromise. A more fundamental challenge, however, is *identifying* the inconsistencies.

3.2. Level of Data Fusion

At first thought, it may seem that fusing multiple datasets at the raw-data level should always yield the best inference. In practice, however, due to the complex and largely unknown nature of the underlying phenomena, various complicating factors, and the specific research question, it may be preferable to fuse the datasets at a higher abstraction level and after certain simplification and reduction steps. Some of the existing approaches are presented below. In general, in the presence of multiple datasets, there exist different possible levels of jointly processing their information:

Data integration implies parallel processing pipelines for each modality, followed by a decision-making step. This procedure may be preferred when modality-specific information is greater than that shared by the modalities, as in the search for true single-trial EEG-fMRI coupling [11], or when modalities are completely non-commensurable, as with hyperspectral imaging and LiDAR [14]. The latter example is often related to classification tasks. Contraindications to data integration are discussed in [8].

Processing modalities sequentially, where one (or more) modality(ies) are used to constrain another. This approach can be found in certain audio-visual applications [9, 10], as well as in the fMRI-constrained solution for the otherwise-underdetermined, ill-posed EEG inverse problem [6].

True fusion, which lets modalities fully interact and inform each other as defined in Sec. 1.

Within "true fusion" there are varying degrees:

Fusion using high-level fatures: Significantly reduce the dimensionality by associating each modality with a small number of variables. In this case, inference is typically of clas-

sification type. This practice can be found in certain mutlisensor [2], HMI [10] and remote-sensing [14] applications.

Fusion using multivariate features: This approach may accommodate for heterogeneities across modalities, as well as significantly reduce the number of samples involved, while leaving the data sufficiently multivariate within each modality (which now is in feature form) such that data in each modality can fully interact [8, 21]. In neuroimaging, common features are task-related spatial maps from fMRI, gray matter images from structural magnetic resonance imaging (sMRI), and event-related potentials (ERP) from EEG, extracted for each subject [8]. In audio-visual applications, features often correspond to speech spectral coefficients and visual cues such as lip countours or speaker's presence in the scene [9].

Using the data as is, or with minimal reduction: In fact, working with features implies a two-step approach: at the first step, features are calculated using a certain criterion; at the second step, features are fused using a different, second criterion. An approach that merges the two, and thus can better exploit the whole raw data, is proposed in [22]. An application in which raw data must be used is pan-sharpening. Here, it is natural to work on raw data because modalities are commensurable.

Related to the open issue of choosing the most appropriate level of data fusion is that of **order selection**. As in non-multimodal analysis, a dimension reduction step may be required in order to avoid over-fitting the data. In a data fusion framework, this step must take into consideration the possibly different representations of the latent variables across datasets. As an example, a solution that maximally retains the joint information while also ensuring that the decomposed sources are independent from each other, in the context of a joint ICA-based approach, is proposed in [23].

3.3. Model

Data fusion, as described in Sec. 1, is all about enabling modalities to fully interact and inform each other. Hence, a key point is choosing an analytical model that faithfully represents the link between modalities and yields a meaningful combination thereof, without imposing phantom connections or suppressing existing ones. Very little is known about the underlying relationships between different modalities. Hence, it is important to be data driven as much as possible. In practice, this means making the least assumptions and using the simplest models, both within and across modalities. "Simple" means, for example, linear relationships between underlying latent variables and/or use of model-independent priors such as sparsity, nonnegativity, statistical independence, low-rank and smoothness. As far as the link among modalities is concerned, existing models can be very roughly partitioned into two classes. In one class of models, modalities deterministically share a certain element. A common mode in CMTF [19, 20], the mixing matrix in *joint ICA* or the signal subspace in *group ICA*; the last two, as well as other relevant examples, are discussed in [21, 24]. A second class holds a separate set of variables for each dataset. In this case, the link is due to a statistical correspondence among the latent variables. These can be generally partitioned into models where the link is via maximizing covariations of corresponding factors, as in multi-way partial least squares (PLS) or multimodal canonical correlation analysis (CCA) (these models are explained in [24] and references therein), and models where the latent sources are statistically dependent across modalities, as multiset CCA (M-CCA) [21] and IVA [16, 17]. Models that allow more flexibility per modality generally yield better inferences, especially regarding modality-specific vs. common latent sources; others may be better at identifying covariations [1, 19, 21, 24]. These are, however, very general observations, which should be tested along with other model choices. The variety of existing solutions, together with the fact that new solutions are constantly being devised for the data [14, 21, 24] with no single solution yet proving to be optimal for any real-world data fusion problem, suggests that the choice of model is a challenge far from being exhausted.

3.4. Theoretical Validation

Despite accumulating empirical evidence of the benefits of data fusion discussed in Sec. 2, there is still very little theoretical validation and quantitative measure of its gain [19]. In addition, as argued in Sec. 3.3, choosing an appropriate model is a widely open question, and approximate and highly simplified models are often preferred. Therefore, a validation step is indispensable. In particular, we are interested in (1) An absolute measure of success: how good is our method or algorithm? (2) Comparing alternative models in order to decide what the most appropriate level of data fusion, model order, analytical model within and across modalities and choice of prior are. (3) Lower bounds on the best achievable error: how far are we from the best possible result (for a given dataset and task)? (4) Theoretical results on the reliability and practical usefulness of the method: Can we prove that the model is identifiable? Is the solution unique? Is the output physically meaningful? Are the results sufficiently interpretable? As an example for a comprehensive performance analysis that answers many of the above questions, we refer to [16] (and references therein) which deals with IVA, and uses also information-theoretic tools.

Although the above questions are not specific to multimodal data fusion, they take special interpretations in the presence of multiple datasets. For example, (1) What is the mathematical formulation of "success", "optimality" [1] and "error" when heterogeneous modalities and uncertainties are involved? What is the most appropriate target function and criterion of success? (2) How can the figure of merit inform us how to exploit the advantages of each modality without suffering from its deficiencies w.r.t. the other modalities? (3) How to evaluate performance of exploratory tasks? Due to the

heterogeneous characteristics of the data, and particularly in exploratory tasks, the interpretability of the output should be given special care. A class on their own are questions regarding the choice of modalities: Should all available modalities be used, and/or given equal importance [8]? How much (information, diversity, redundancy) does each modality bring in to the total equations? How to quantify this "extra contribution"? Information theory seems like a natural direction, as discussed in [6]. Attention should be paid, for example, when modalities are too close to each other: in this case, they may not really convey new information; in addition, they may be exposed to similar noise, and thus bias results [2].

4. CONCLUSION

We enter a "Big Data" era, in which the abundance of diverse sources of information makes it practically impossible to ignore the presence of multiple datasets that are possibly related. It is very likely that an ensemble of related datasets is "more than the sum of its parts", in the sense that it contains precious information that is lost if these relations are ignored. The information of interest that is hidden in these datasets is usually not easily accessible, however. We argue that the road to this added value must go through first understanding and identifying the particularities of multimodal data, as opposed to other types of aggregated datasets. A second message is that the encountered challenges are ubiquitous, whence the incentive that both challenges and solutions be discussed at a level that brings together all involved communities.

REFERENCES

[1] I. Van Mechelen and A. K. Smilde, "A generic linked-mode decomposition model for data fusion," *Chemometrics and Intelligent Laboratory Systems*, vol. 104, no. 1, pp. 83–94, Nov. 2010.

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

[3] L. Xie et al., "Multimodal joint information processing in human machine interaction: recent advances (guest editorial)," *Multimedia Tools and Applications*, pp. 1–5, Nov. 2013.

[4] T. Stathaki, *Image fusion: algorithms and applications*, Elsevier, 2008.

[5] H. B. Mitchell, *Data fusion: concepts and ideas*, Springer, 2nd edition, 2012.

[6] F. Bießmann et al., "Analysis of multimodal neuroimaging data," *IEEE Rev. Biomed. Eng.*, vol. 4, pp. 26–58, 2011.

[7] T. Adalı et al., Eds., *Special Section on Multimodal Biomedical Imaging: Algorithms and Applications*, vol. 15. IEEE Trans. Multimedia, Aug. 2013.

[8] V. D. Calhoun and T. Adalı, "Feature-based fusion of medical imaging data," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 5, pp. 711–720, Sep. 2009.

[9] B. Rivet et al., "Audio-visual speech source separation," *IEEE Signal Process. Mag.*, May 2014, Special Issue: Source Separation and its Applications.

[10] S. T. Shivappa et al., "Audiovisual information fusion in human-computer interfaces and intelligent environments: A survey," *Proc. IEEE*, vol. 98, no. 10, pp. 1692–1715, Oct. 2010.

[11] M. De Vos et al., "The quest for single trial correlations in multimodal EEG–fMRI data," in *Proc. EMBC'13*, Osaka, Japan, Jul. 2013, pp. 6027–6030.

[12] M. Garibaldi and V. Zarzoso, "Exploiting intracardiac and surface recording modalities for atrial signal extraction in atrial fibrillation," in *Proc. EMBC'13*, Osaka, Japan, Jul. 2013, pp. 6015–6018.

[13] A. Van de Vel et al., "Non-EEG seizure-detection systems and potential SUDEP prevention: State of the art," *Seizure: European Journal of Epilepsy*, vol. 22, no. Issue 5, pp. 345–355, Jun. 2013.

[14] Ch. Debes et al., "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.*, vol. 7, 2014, to appear.

[15] J.-F. Cardoso, "The three easy routes to independent component analysis; contrasts and geometry," in *Proc. ICA 2001*, San Diego, CA, Dec. 2001, pp. 1–6.

[16] T. Adalı et al., "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Process. Mag.*, May 2014, Special Issue: Source Separation and its Applications.

[17] T. Kim et al., "Independent vector analysis: An extension of ICA to multivariate components," in *Independent Component Analysis and Blind Signal Separation*, Heidelberg, 2006, vol. 3889 of *LNCS*, pp. 165–172, Springer.

[18] L. Alparone et al., "Data fusion contest: Fusion of panchromatic and multispectral images," in *Proc. IGARSS*, Jul. 2006, pp. 3814–3815.

[19] E. Acar et al., "Understanding data fusion within the framework of coupled matrix and tensor factorizations," *Chemometrics and Intelligent Laboratory Systems*, vol. 129, pp. 53–63, 2013.

[20] U. Şimşekli et al., "Optimal weight learning for coupled tensor factorization with mixed divergences," in *Proc. EUSIPCO*, Marrakech, Morocco, Sep. 2013.

[21] N. M. Correa et al., "Canonical correlation analysis for data fusion and group inferences," *IEEE Signal Process. Mag.*, vol. 27, no. 4, pp. 39–50, Jul. 2010.

[22] N. M. Correa et al., "Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI," *NeuroImage*, vol. 50, no. 4, pp. 1438–1445, May 2010.

[23] J. Sui et al., "Three-way (N-way) fusion of brain imaging data based on mCCA+jICA and its application to discriminating schizophrenia," *NeuroImage*, vol. 66, pp. 119–132, Feb. 2013.

[24] J. Sui et al., "A review of multivariate methods for multimodal fusion of brain imaging data," *Journal of Neuroscience Methods*, vol. 204, no. 1, pp. 68–81, 2012.