APPLYING INDEPENDENT VECTOR ANALYSIS ON EEG-BASED MOTOR IMAGERY CLASSIFICATION

Caroline P. A. Moraes[*], *Bruno Aristimunha*[*,†], *Lucas Heck dos Santos*[*], *Walter Hugo Lopez Pinaya*[†], *Raphael Yokoingawa de Camargo*[*], *Denis G. Fantinato*[∓], *Aline Neves*[*]

[*]Center for Engineering, Modeling and Applied Social Sciences,
Federal University of ABC (UFABC), Santo André, Brazil
[†]Department of Biomedical Engineering, School of Biomedical Engineering & Imaging Sciences, King's College London (KCL), London, United Kingdom
[∓]State University of Campinas (UNICAMP), Campinas, Brazil

ABSTRACT

Joint Blind Source Separation (JBSS) is an essential and versatile research topic that has attracted the attention of researchers in the last decade. Independent Vector Analysis (IVA) is an exciting approach in the context of the JBSS method since it is an extension of Independent Component Analysis (ICA) towards the exploitation of the statistical dependency between different datasets through the use of Mutual Information. In this work, we propose an original approach of IVA as a feature extraction step for Brain-Computer Interfaces, focused on the Motor Imagery (MI) paradigm. For this, we use the BCI Competition IV - Dataset 1. Since the participants of the experiment are performing the same MI tasks, we assume that the channels related to MI present correlated signals across subjects that might be explored by IVA techniques. The results show that the algorithm could classify the MI movements using a consolidated and low-cost classifier, Support Vector Machine, achieving an accuracy of 85%.

Index Terms— Brain-Computer Interfaces, Electroencephalogram, Independent Vector Analysis, Machine Learning, Motor Imagery

1. INTRODUCTION

The Blind Source Separation (BSS) problem, in its linear case, has been extensively studied given its versatility and applications in many areas, from engineering to neuroscience. This problem consists of retrieving sources from signals previously submitted to a mixing process. Since both sources and the mixture process are generally unknown, the source separation task is carried out in an unsupervised fashion, based only on the samples of the mixed signals [1]. However, a multimodal approach is often used to improve feature extraction and classification, for example, in scenarios with multi-subject/multimodal data fusion as in neuro diagnostics [2]. In that sense, Joint Blind Source Separation (JBSS) extends the problem to a multimodal approach, assuming a dependence across datasets and independence of latent sources within a dataset [3]. Independent Vector Analysis (IVA) [4] is one particular approach to the JBSS problem and is a recent extension of the Independent Component Analysis (ICA) to multiple datasets. IVA has shown, in most cases, superior performance in capturing variability in spatial components across datasets, e.g., sound separation [5], denoising electroencephalogram (EEG) [6], and in functional Magnetic Resonance Imaging (fMRI) applications [7].

Brain-Computer Interface (BCI) provides an alternative communication channel for disabled patients and possible automatic diagnoses. However, the efficient classification and selection of the signals remain challenging in different areas [8]. EEG is a non-invasive technique that measures electrical activity in the brain by attaching electrodes to the scalp. The feature extraction over EEG signals for BCI systems has been a crucial stage for several classification problems. Previous studies with EEG signals showed a prominent advantage compared with biomedical signals due to the wide range of practical applications [9]. EEG analysis has attracted the attention of the science community, especially in the diagnosis of several neurological and neuropsychiatric disorders, such as epilepsy [10], Alzheimer's disease [11], and Parkinson's disease [12]. Motor Imagery (MI) is amongst the most popular EEG-based BCI paradigms. It refers to the active cognitive process in which movements are rehearsed mentally without overt body movements. In some cases, MI can be seen as a novel therapeutic tool in neurological rehabilitation that can be used for relearning motor control after damage to the motor system [13, 14].

In recent years, researchers developed novel techniques

This study was financed by grants #2020/10014-2 and #2020/09838-0, São Paulo Research Foundation (FAPESP), grant #88887.595656/2020-00 Coordination for the Improvement of Higher Education Personnel (CAPES), and WHLP which is supported by Wellcome Innovations WT213038/Z/18/Z.

for movement classification using the MI paradigm. The BCI competition IV[1] is a traditional dataset that provides open access to high-quality neuroscientific data [15]. In [16], the authors explore the spectral information existing in the bispectrum of MI-related EEG signals, called bispectrum-based channel selection (BCS), to classify movements from the BCI competition IV Dataset 1 (DS1). They suggest that the spectral information existing in the bispectrum of MI-related EEG signals may aid the identification of which channels contain redundant information. Haihong Zhang et al. [17] also worked with the DS1 dataset, but using an optimum spatio-spectral filtering network (OSSFN) that optimizes the spatial filters in conjunction with a band-pass filter by maximizing the mutual information between the feature vectors.

This paper proposes a novel application of IVA as a feature extraction technique for MI classification. In our method, we combine IVA with Autoregressive model and Machine Learning methods to classify MI movements. This method allows the retrieval of independent components and the understanding of the relationship across different datasets. Exploring statistical dependency between datasets through mutual Information helps in MI classification since it allows a generic and homogeneous treatment of the whole data, which facilitates the classification task. Furthermore, to the best of our knowledge, this is the first application of IVA as feature extraction in the motor imagery classification context. Additionally, the features obtained from IVA reduced the need for complex classifiers, achieving good results with a Support Vector Machine.

2. METHODS

2.1. Independent Vector Analysis

Similar to ICA, the components extracted with IVA from a particular dataset are assumed to be maximally independent. In contrast, IVA also maximizes the dependence between correlated components from different datasets. We can describe the general model with K datasets, each containing N samples formed from linear mixtures of M independent sources. The mixing process $\mathbf{x}^{[k]}(n)$ can be modeled by:

$$\mathbf{x}^{[k]}(n) = \mathbf{A}^{[k]}\mathbf{s}^{[k]}(n), \;\; 1 \leq n \leq N, 1 \leq k \leq K, \quad (1)$$

where $\mathbf{s}^{[k]}(n) = [s_1^{[k]}(n), \ldots, s_M^{[k]}(n)]^T \in \mathbb{R}^M$ is the concatenated source vector in each dataset and $\mathbf{A}^{[k]} \in \mathbb{R}^{M\times M}$ is the k-th invertible mixing matrix, being both unknown.

Following [3], we use data pre-whitening. The whitening matrix $\mathbf{V}^{[k]}$ was obtained by computing $\mathbf{V}^{[k]} = \mathbf{E}^{[k]}\mathbf{D}^{[k]^{-0.5}}$ $\mathbf{E}^{[k]^T}$, where $\mathbf{D}^{[k]}$ is a diagonal matrix with the eigenvalues and $\mathbf{E}^{[k]}$ is a matrix with the eigenvectors of the correlation matrix from the mixing vector $\mathbf{x}^{[k]}$ for each dataset. With $\mathbf{V}^{[k]}$, the whitening process results in $\mathbf{z}^{[k]}(n) = \mathbf{V}^{[k]}\mathbf{x}^{[k]}(n)$.

[1]DS1 is available in: https://www.bbci.de/competition/iv/

The demixing process aims to find K matrices $\mathbf{W}^{[k]}$ and the corresponding source vector estimates $\mathbf{y}^{[k]}(n)$ for each dataset. This separation system is given by:

$$\mathbf{y}^{[k]}(n) = \mathbf{W}^{[k]}\mathbf{z}^{[k]}(n), \;\; 1 \leq n \leq N, \; 1 \leq k \leq K. \quad (2)$$

IVA minimizes the mutual information among Sources Component Vectors (SCV). In this paper, we work with IVA-G [3, 7] that assumes a multivariate Gaussian distribution for the SCVs, and thus only takes second-order statistical information into account. In [7], the authors define the m-th SCV as: $\mathbf{s}_m(n) = [s_m^{[1]}(n), s_m^{[2]}(n), \ldots, s_m^{[k]}(n)]^T \in \mathbb{R}^K$. Source components are made maximally dependent within an SCV and maximally independent across SCVs in IVA. This process is exemplified by Fig.1.

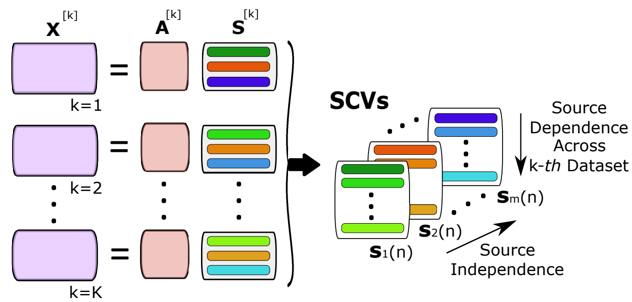

Fig. 1. Independent Vector Analysis concept.

2.2. Machine Learning Classifiers

Following [18], we selected the Support Vector Machine (SVM) and the Logistic Regression (LR) to classify features obtained through the use of IVA. SVM is an efficient supervised algorithm based on statistical learning theory that can be used for classification or regression problems [19]. When the IVA is combined with SVM, the algorithm is called IVA-S. In our experiments, we used a Bayesian optimization to automatically tune the hyper-parameters of the SVM model, with the hyper-parameter search space including the product of the gamma options, the degrees, the number of kernels, and the type of kernel [20].

For the Logistic Regression [21] applied in IVA (IVA-LR), we use a Random Grid Search for accuracy optimization, analyzing the regularization hyper-parameters. When evaluating the effectiveness of the accuracy improvement during training, we profiled a vast range of hyper-parameters values. In this case, we used 1000 iterations, and the search space included the inverse of regularization strength C (with 10 search values between [0.001, 10]) and penalty (L1, L2).

3. EXPERIMENTS

3.1. Dataset 1 from BCI competition IV

Our experiments used the motor imagery dataset 1 from BCI Competition IV [15]. This dataset has seven subjects, three artificial (labeled as c, d, and e), and four real subjects (a, b, f, and g). Each of them imagined the movement of the left hand (mandatory) and either the right hand or foot/feet. In this paper, we select only the real subjects since the aim is to explore the real MI features. Each subject performed 200 trials (100 per class), acquired over four different sessions. The recording setup used 59 EEG electrodes and at first, the dataset was sampled at 1000 Hz and band-pass filtered between 0.05 and 200 Hz, but a downsampled version was also provided by the competition. We use the downsampled (100 Hz) version of the data for computational efficiency.

3.2. Procedure

We compare the proposed IVA feature extraction technique with two methods, BCS [16] and OSSFN [17], which are the state-of-the-art methods for decoding this specific motor imagery dataset. Initially, to evaluate the algorithm performance, the EEG signals were submitted to a holdout evaluation technique where 80% of the data was used for training and 20% for testing, following the competition rules. Training data was then whitened, considering each subject and each class separately, leading to the definition of matrices $\mathbf{V}_c^{[k]}$, with $k = \{1, 2, 3, 4\}$ and $c = \{1, 2\}$.

After the pre-processing stage, the IVA was applied to the training data for each class separately to obtain the $\mathbf{W}_c^{[k]}$ matrices where $k = \{1, 2, 3, 4\}$ and $c = \{1, 2\}$, that corresponds to the extraction of the main features for the c-th class and k-th subject. Then, considering the k-th subject, training data was multiplied by $\mathbf{W}_1^{[k]}$ and $\mathbf{W}_2^{[k]}$, followed by each corresponding whitening matrix $\mathbf{V}_1^{[k]}$ and $\mathbf{V}_2^{[k]}$, resulting in $\mathbf{y}_c^{[k]}$ (Algorithm 1). Using matrices of both classes is necessary at this point considering that test data is unknown.

Next, $\mathbf{y}_1^{[k]}$ and $\mathbf{y}_2^{[k]}$ were stacked, and Autoregressive (AR) modeling with 4 coefficients (the number of coefficients was chosen based on a sweeping survey) was applied to each extracted feature, which aside from preserving the important attributes from the time series, also reduces the data dimension. The resulting AR parameters are used as the classifiers inputs (SVM and Logistic Regression). Finally, the classifiers predict the class for each subject.

The IVA and classifiers weights obtained in the training stage were kept fixed and applied to the test data. In that sense, the parameters optimization for IVA, AR modeling, and classifiers were performed only with the training data. This procedure, exemplified by Fig. 2, was used to classify the MI movements between left (L) or right/foot (R/F). The whole procedure is summarized in Algorithm 1.

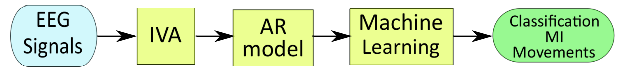

Fig. 2. General Algorithm block diagram. The EEG data from each subject is transformed using IVA, AR modeling and classified using ML methods.

Algorithm 1: IVA-Machine Learning

Initialization of IVA-S, IVA-LR:
$\mathbf{W}_c^{[k]}$= **IVA**($\mathbf{z}^{[k]}$); $c = 1, 2; k = 1, \ldots, 4$
for *each subject* k **do**
  for *each class* c **do**
    $$\mathbf{y}_c^{[k]} = \begin{bmatrix} \mathbf{W}_1^{[k]}\mathbf{V}_1^{[k]}\mathbf{z}_c^{[k]} \\ \mathbf{W}_2^{[k]}\mathbf{V}_2^{[k]}\mathbf{z}_c^{[k]} \end{bmatrix}$$
  end
  $\mathbf{y}^{[k]} \leftarrow [\mathbf{y}_1^{[k]} \;\; \mathbf{y}_2^{[k]}]$
  $\hat{\mathbf{y}}^{[k]}$=**AR model**($\mathbf{y}^{[k]}$)
  SVM or LR classifier - input: $\hat{\mathbf{y}}^{[k]}$;
end

3.3. Analysis of feature extraction

In the previous sections, we presented the Sources Component Vectors and their relation with IVA. Based on the SCV concept and following [7], in Fig. 3 we present some examples of SCVs covariance matrices obtained for the MI movements classification. In Fig. 3(a), we considered class 1 (left-L), showing $\mathbf{s}_1(n)$ and $\mathbf{s}_3(n)$ covariance matrices ($K = 4$), while Fig. 3(b) considers class 2 (right/foot-R/F), $\mathbf{s}_2(n)$ and $\mathbf{s}_{15}(n)$ covariance matrices. In Fig. 3(a) the high covariance values between subjects "a", "b", and "f" are noticeable in $\mathbf{s}_1(n)$, while they are more important for subjects "a", "b", and "g" in $\mathbf{s}_3(n)$. On the other hand, the covariance is more prominent between subjects "a" and "b" for $\mathbf{s}_2(n)$, and among "a", "b", and "g" subjects for $\mathbf{s}_{15}(n)$ (Fig. 3(b)). In that sense, it is possible to verify that IVA is capable of exploiting the correlation between different datasets, using such information to aid in the extraction of features.

3.4. Analysis of the Machine Learning algorithms parameters

The IVA-S was trained from scratch with a maximum of 1000 iterations. The Bayesian technique searched the parameter C from 0.01 to 1, the degree from 1 to 12, the kernel among linear, polynomial, radial basis, or sigmoid, and the gamma as scaling or automatic, for 50 distinct combinations.

The partial dependence obtained in this search, for subject f, is seen in Fig. 4 as an example and Tab. 1 shows the selected parameters for each subject. This makes it clear that the C and kernel had higher relevance in the model evaluation, while the

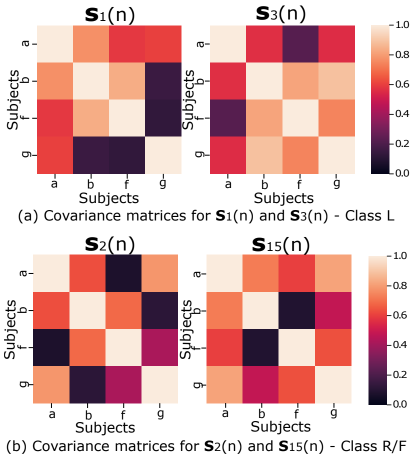

Fig. 3. Examples of SCVs covariance matrices obtained through IVA for the DS1 dataset.

other parameters had no relevant impact.

For IVA-LR, penalty L1 was used and we analyzed the Receiver Operating Characteristic (ROC) curve on label 1. In this case, we have the Area Under the Curve with 80%, indicating good coverage and no imbalance between the ratio of true positives and false positives.

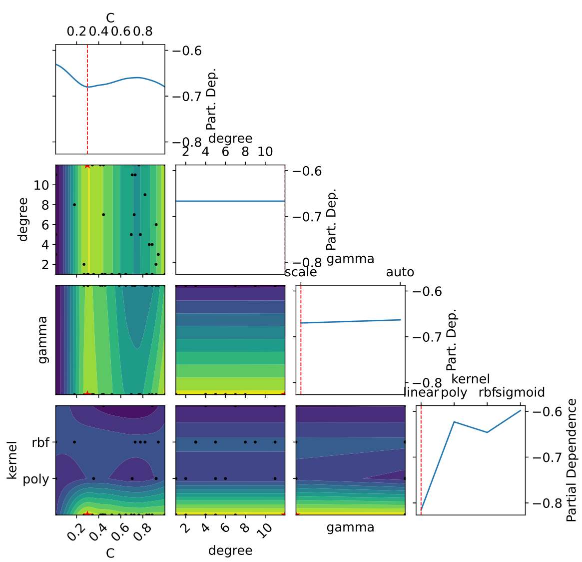

Fig. 4. Partial dependence plot of the SVM's Bayesian hyper-parameters search.

3.5. Results

Tab. 2 shows the accuracy based on 10 simulations by using IVA with Logistic Regression and SVM, compared with the literature results (OSSFN [17] and BCS [16]). The IVA-S was able to find the best results for subjects "a" and "g" with an accuracy of 92.5%, in both cases. For subject "f" the best result was achieved through the IVA-LR with an accuracy of 95%. Subject "b" was the most difficult to classify, for all methods, including the literature ones, achieving a maximum accuracy of 86.8% with the OSSFN method. Several possible variables could contribute to this lower performance of subject "b" such as interferences as artifacts or noise; EEG signals from subject "b" not presenting correlation or not as strong as the others.

On average, the IVA-S overpasses the second best state-of-art result and presents a small difference of only 5.5% compared to the OSSFN algorithm, achieving an average accuracy of 85%. Additionally, our method achieved good results with the application of a simpler and consolidated machine learning method, i.e., SVM, where the number of adjusted weights is considerably smaller when compared with the benchmark ones.

Table 1. The chosen SVM's Bayesian hyper-parameters.

Parameters / Subjects	C	Kernel	Gamma	Degree
a	0.514	linear	scale	12
b	0.896	linear	scale	5
f	0.297	linear	scale	12
g	0.935	linear	scale	6

Table 2. Performance on the BCI Dataset 1 measured based on individual, average (Aver.) and standard deviation (SD) accuracy [%].

Methods / Subjects	OSSFN	BCS	IVA-LR	IVA-S
a	89.8	78.5	80.0	**92.5**
b	**86.8**	77.5	75.0	65.0
f	93.0	92.0	**95.0**	90.0
g	92.4	87.0	80.0	**92.5**
Aver.±SD	**90.5**±5.8	83.8±7.0	82.50± 7.5	**85.0**±13.4

4. CONCLUSION

We proposed a method to use IVA to perform feature extraction for Motor Imagery EEG signals. The model performed competitively compared with state-of-art methods in the same dataset, while using classic low-cost machine learning methods. Our method holds promise in scenarios where structured data captured by the brain signals are somewhat correlated, such as in cases where different subjects execute the same task, as in BCI experiments. To the best of our knowledge, this is the first time IVA has been used as a feature extraction method in an EEG dataset for the MI problem. We believe that IVA has the potential to be even further improved, bringing more advances to the Multivariate Pattern Analysis (MVPA) area through a solid source separation step.

5. REFERENCES

[1] Pierre Comon and Christian Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*, Academic press, 2010.

[2] Yuhui Du, Zening Fu, and Vince D Calhoun, "Classification and prediction of brain disorders using functional connectivity: promising but challenging," *Frontiers in neuroscience*, vol. 12, pp. 525, 2018.

[3] Matthew Anderson, Tuelay Adali, and Xi-Lin 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, 2011.

[4] Taesu Kim, Torbjørn Eltoft, and Te-Won Lee, "Independent vector analysis: An extension of ICA to multivariate components," in *International conference on independent component analysis and signal separation*. Springer, 2006, pp. 165–172.

[5] Robin Scheibler and Nobutaka Ono, "Independent vector analysis with more microphones than sources," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2019, pp. 185–189.

[6] Xueyuan Xu, Xun Chen, and Yu Zhang, "Removal of muscle artefacts from few-channel EEG recordings based on multivariate empirical mode decomposition and independent vector analysis," *Electronics Letters*, vol. 54, no. 14, pp. 866–868, 2018.

[7] Isabell Lehmann, Evrim Acar, Tanuj Hasija, MABS Akhonda, Vince D Calhoun, Peter J Schreier, and Tülay Adali, "Multi-task fMRI data fusion using IVA and PARAFAC2," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 1466–1470.

[8] Li Hu and Zhiguo Zhang, *EEG signal processing and feature extraction*, Springer, 2019.

[9] Ashima Khosla, Padmavati Khandnor, and Trilok Chand, "A comparative analysis of signal processing and classification methods for different applications based on EEG signals," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 2, pp. 649–690, 2020.

[10] Zixu Chen, Guoliang Lu, Zhaohong Xie, and Wei Shang, "A unified framework and method for EEG-based early epileptic seizure detection and epilepsy diagnosis," *IEEE Access*, vol. 8, pp. 20080–20092, 2020.

[11] Haitao Yu, Xinyu Lei, Zhenxi Song, Chen Liu, and Jiang Wang, "Supervised network-based fuzzy learning of EEG signals for Alzheimer's disease identification," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 1, pp. 60–71, 2019.

[12] Syed Aamir Ali Shah, Lei Zhang, and Abdul Bais, "Dynamical system based compact deep hybrid network for classification of parkinson disease related EEG signals," *Neural Networks*, vol. 130, pp. 75–84, 2020.

[13] TH Mulder, JBH Hochstenbach, MJG Van Heuvelen, and AR Den Otter, "Motor imagery: the relation between age and imagery capacity," *Human movement science*, vol. 26, no. 2, pp. 203–211, 2007.

[14] Martin Lotze and Ulrike Halsband, "Motor imagery," *Journal of Physiology-paris*, vol. 99, no. 4-6, pp. 386–395, 2006.

[15] Benjamin Blankertz, Guido Dornhege, Matthias Krauledat, Klaus-Robert Müller, and Gabriel Curio, "The non-invasive Berlin brain–computer interface: fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.

[16] Jing Jin, Chang Liu, Ian Daly, Yangyang Miao, Shurui Li, Xingyu Wang, and Andrzej Cichocki, "Bispectrum-based channel selection for motor imagery based brain-computer interfacing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 10, pp. 2153–2163, 2020.

[17] Haihong Zhang, Zheng Yang Chin, Kai Keng Ang, Cuntai Guan, and Chuanchu Wang, "Optimum spatio-spectral filtering network for brain–computer interface," *IEEE Transactions on Neural Networks*, vol. 22, no. 1, pp. 52–63, 2010.

[18] Xindong Wu, Vipin Kumar, J Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J McLachlan, Angus Ng, Bing Liu, Philip S Yu, et al., "Top 10 algorithms in data mining," *Knowledge and information systems*, vol. 14, no. 1, pp. 1–37, 2008.

[19] Vladimir Vapnik, *Statistical learning theory*, Wiley New York, 1998.

[20] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al., "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[21] Thomas Edgar and David Manz, *Research methods for cyber security*, Syngress, 2017.