Mixtures of Independent Component Analyzers for EEG Prediction

Gonzalo Safont[1], Addisson Salazar[1], Luis Vergara[1], Alberto Gonzalez[1], Antonio Vidal[2]

[1] Instituto de Telecomunicaciones y Aplicaciones Multimedia
[2] Departamento. de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia, Camino de Vera s/n, 46022, Valencia, Spain
gonsaar@upvnet.upv.es, asalazar@dcom.upv.es, lvergara@dcom.upv.es

Abstract. This paper presents a new application of independent component analysis mixture modeling (ICAMM) for prediction of electroencephalographic (EEG) signals. Demonstrations in prediction of missing EEG data in a working memory task using classic methods and an ICAMM-based algorithm are included. The performance of the methods is measured by using four error indicators: signal-to-interference (SIR) ratio, Kullback-Leibler divergence, correlation at lag zero and mean structural similarity index. The results show that the ICAMM-based algorithm outperforms the classical spherical splines method which is commonly used in EEG signal processing. Hence, the potential of using mixtures of independent component analyzers (ICAs) to improve prediction, as opposed on estimating only one ICA is demonstrated.

Keywords: ICA, EEG, prediction, working-memory task.

1 Introduction

Independent Component Analysis (ICA) is a research area that is progressively finding more applications for blind source separation (BSS) and feature modeling/extraction. The objective of ICA is finding a linear transformation of the observed signals, such that the resulting signals (the sources or generators of observed data) are as statistically independent as possible [1].

The ICA method is extended in the ICA mixture model (ICAMM), a kind of nonlinear ICA, where multiple ICA models are learned and weighted in a probabilistic manner. ICAMM is a model of conditional independence, i.e. the assumption of independence only holds for data within the same class, with dependencies between classes being possible [2, 3]. ICAMM has been proposed recently as a flexible approximation for modeling mixtures of arbitrary probability densities with non-Gaussian distributions for the independent components (i.e. relaxing the restriction of modeling each component as a multivariable Gaussian) [4].

There is extensive literature of ICA applications in electroencephalographic (EEG) signal processing. For instance, study of developmental differences in the saccadic contingent negative variation [5], EEG and event-related potential (ERP) data [6, 7], and removal of artifacts in the EEG signal [8]. However, there are a few references of

the use of ICAMM for EEG signal processing. Recently, ICAMM was extended to the case of having sequential dependence in the feature observation record [9] in order to model the nonlinear dynamical system underlying the EEG signals [10].

This paper proposes an application of ICAMM to prediction of EEG signals. The application is demonstrated in the field of scanning human short-term memory during a working memory task [11, 12]. Although there are several applications of ICA as preprocessor in prediction of temporal series (see for instance [13]), the prediction itself has not been done considering an underlying ICA model of the data density. In this work, we propose a new ICAMM-based algorithm for EEG signal prediction. Data are assumed to come from several mutually-exclusive classes that are each generated by a different ICA model. This strategy allows for linear local projections that can adapt to partial segments of a data set, while keeping generalization (i.e. the capability of nonlinear modeling) given the mixture of the different ICA models.

2   Materials and methods

2.1   Working memory task

We implemented the Sternberg task that is a classic test of multi-item short-term memory [11]. Each participant was shown series (or trials) of one to five symbols (or items), each one taken at random from a fixed set shown in Fig. 1. Each symbol was displayed during 0.2 seconds (item stage), after which the screen was cleared for 1 second (pause stage), and then the following symbol appeared on screen. The length of the series varied at random from trial to trial. After the last symbol, there followed a further 1-second delay (retention stage), a warning signal, and then a test symbol. The subject was required to decide whether or not the test symbol is one of the symbols shown in this series (probe stage). Positive and negative responses were required with equal frequency. Finally, after the subject had decided, there was a pause of 0.5 s until the next trial. Each experiment comprised 30 trials.

Fig. **2** shows an outline of the stages in a Sternberg working memory task. There were a total of three subjects, with 2 to 5 experiments per subject. The subjects showed a success rate of 98 %, with an average response time of 1.17 s.
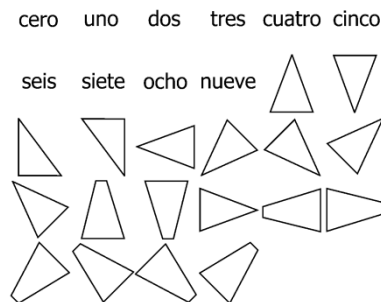

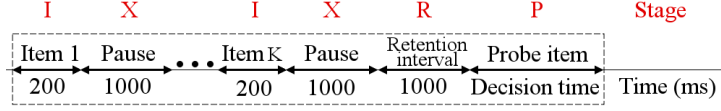**Fig. 1.** Set of 28 symbols used for the experiment.

Fig. 2. Stage diagram in a single trial of the Sternberg task.

2.2 EEG recording

EEG signals were recorded from electrodes using water-based gel coupling and placed according to the 10-10 system. Sixty-four channels were recorded, using Cz as reference. The signals were recorded using a Biosemi device with active electrodes. The sampling rate for data acquisition was 512 Hz. All channels were band-pass filtered between 1 and 70 Hz, with an additional narrow notch filter at 50 Hz.

2.3 Independent Component Analysis Mixture Modeling

Independent component analysis is a technique that aims to maximize statistical independence between hidden variables in observation vectors [1]. The standard noiseless ICA model assumes that the observations, $\mathbf{x} = (x_1, x_2, ..., x_M)^T$, are composed by linear mixtures of random variables that are mutually independent, the sources $s = (s_1, s_2, ..., s_N)^T$. That is, for some matrix $\mathbf{A}$:

$$\mathbf{x} = \mathbf{A}\,\mathbf{s} \tag{1}$$

ICA methods estimate the so-called mixing matrix, $\mathbf{A}$, or its inverse, $\mathbf{W} = \mathbf{A}^{-1}$. This matrix contains the coefficients of the linear transformations from sources to observations. Thus, the mixing matrix can be applied to separate each of the sources $\mathbf{s}_i = \mathbf{w}_i\,\mathbf{x}$, $i = 1, ..., N$, where $\mathbf{w}_i$ is the i-th row of $\mathbf{W}$. For the sake of simplicity, we will assume the square problem (the same number of sources as mixtures).

ICA mixture modeling (ICAMM) is proposed in the framework of pattern recognition, considering that the observed data come from a mixture model and that they can be categorized in K mutually exclusive classes. Each one of these classes is modeled as an ICA, i.e. linear combinations of independent non-Gaussian sources. The general formulation of ICAMM is:

$$\mathbf{x}_k = \mathbf{A}_k\mathbf{s}_k + \mathbf{b}_k, \ \ k = 1, ..., K \tag{2}$$

where each class k is described by an ICA model with mixing matrix $\mathbf{A}_k$ (or its inverse $\mathbf{W}_k = \mathbf{A}_k^{-1}$) and bias vector $\mathbf{b}_k$. Essentially, $\mathbf{b}_k$ determines the location of the cluster and the mixing matrix determines its shape. The goal of an ICAMM algorithm is to determine the parameters for each class.

ICAMM was introduced in [3] considering a source model switching between Laplacian and bimodal densities. Recently, a generalization of the ICAMM framework was proposed which includes non-parametric density estimation, semi-

supervised learning, using any ICA algorithm for parameter updating, and correction of residual dependencies [2].

2.4 ICAMM-based algorithm for prediction

Let us consider data vector $\mathbf{x}$ of size $[N \times 1]$ which can be modeled through an ICAMM model such as the one in equation (2). Assuming that N_2 values of vector $\mathbf{x}$ are unknown, known and unknown values can be grouped into two smaller vectors, $\mathbf{y}$ (known values) and $\mathbf{z}$ (unknown values). Thus, we can write:

$$\mathbf{x} = \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} \tag{3}$$

The parameters of the ICAMM model are estimated previously using training data. Once the model is estimated, the probability density function of data vector $\mathbf{x}$ can be expressed as a function of the probability densities conditioned to each class. The proposed predictor is a maximum a posteriori (MAP) estimator of $\mathbf{z}$ that maximizes the joint probability density of known and unknown data, $p(\mathbf{y}, \mathbf{z})$. This joint probability density can be derived from (2) and (3). If we derivate it:

$$\frac{\delta p(\mathbf{y}, \mathbf{z})}{\delta \mathbf{z}} = \sum_{k=1}^{K} \det(\mathbf{W}_k) \cdot \frac{\delta p(\mathbf{s}_k)}{\delta \mathbf{z}} \cdot p(C_k) \tag{4}$$

where C_k denotes class k, and $s_{k,n}$ is the n-th source of class k. The calculation of the derivative of $p(\mathbf{s}_k)$ is complex for most distributions. By applying the principle of independence of the sources $\mathbf{s}_k$ (one of the assumptions of the ICAMM model), this probability can be obtained as the product of the probability densities of each individual source, $p(s_{k,n})$. In this work, $p(s_{k,n})$ and its derivatives were found using a non-parametric estimator. The target function is then maximized using a steepest ascent algorithm to achieve the prediction, $\hat{\mathbf{z}}$. We have called the algorithm above Predicamm.

2.5 Spherical splines

Spherical splines are a classical method for interpolation of EEG data first introduced in [16]. Let $V(\mathbf{r})$ be the potential at some position $\mathbf{r}$ on the surface of a sphere of radius r, and let $\mathbf{r}_i$ be the location of one of the N measurement electrodes. Spherical splines assume that the potential at any point $\mathbf{r}$ on the surface of the sphere can be approximated by $V(\mathbf{r}) = c_0 + \sum_{i=1}^{N} c_i \, g_m(\mathbf{r} \cdot \mathbf{r}_i)$ and $g_m(x) = \frac{1}{4\pi} \sum_{n=1}^{\infty} \frac{2n+1}{(n(n+1))^m} P_n(x)$, where $P_n(x)$ is the ordinary Legendre polynomial

of order n. In practice, the sum is truncated at some finite number of terms, n_{max}, depending upon the value of the parameter m. The best results are achieved when m equals 3 or 4.

The coefficients c_i are determined by satisfying two conditions. First, the interpolation function must return the data when evaluated at the original data points. Second, the coefficients must add up to zero. These two conditions can be combined into a single linear system of equations.

3   EEG data prediction using ICAMM

A case was considered where the values at one or more of the electrodes were considered unusable and had to be predicted. This would be the case, e.g. if some of the electrodes had been disconnected, or for artifact-removal purposes. To test the performance of Predicamm compared with spline method, a Monte Carlo experiment with 1,000 steps was applied, with the end result being the average of each of the step results. At each step, one or more of the channels (randomly chosen) were removed, and their values were predicted using the proposed algorithm.

To estimate the ICA mixture model in the EEG data, we used the on-line ICAMM algorithm in [14]. Four figures of merit or indices were used to evaluate the quality of the prediction: Signal-to-Interference Ratio (SIR), the Kullback-Leibler divergence (KLD), the correlation at lag zero (CORR) and the Mean Structural Similarity (MSSIM). SIR measures the prediction error, CORR considers the temporal correlation between the prediction and the true data, both indices considering the amplitude of the signals. KLD, on the other hand, measures the distance between the probability densities of predicted and true data. Finally, MSSIM is an index, commonly used in image processing, that measures the structural similarity between predicted and true data [17].

Fig. 3 shows the average prediction results for an increasing number of missing channels for sphere splines, and for Predicamm using either a single-class or a multiple-class model. Predicamm achieves the best result for all the cases considering SIR, CORR and MSSIM. Note that most of the index values are significantly improved with an increasing number of missing channels, more notably for MSSIM and SIR. Both methods achieve a similar result for the KLD. Even though both ICAMM models achieve a good result, the multiple-classes model always achieves a better result than the single-class model. As we will consider in the following section, the multiple-class model achieves this improved result by modeling the local behavior of the EEG signals. Higher-amplitude channels (e.g. frontal channels) are easier to interpolate correctly than lower-amplitude ones. Since the missing channels were chosen at random, with an increasing number of missing channels comes an increase in the probability of easy-to-predict missing channels, thus increasing overall performance. SIR, on the other hand, is not affected by this effect.

In order to demonstrate the difference in performance, a 32-missing-channel case was performed. Fig. 4 shows the original scalp map, the predicted scalp maps obtained by the spline and Predicamm methods and the error achieved by the predictions. Note that the error achieved by spline method is much higher than the

error achieved by Predicamm. The high amplitudes recorded at prefrontal and parietal-occipital areas of the scalp map are incorrectly recovered by the spline method. The prediction assessment indices measure these differences (see Table 1).
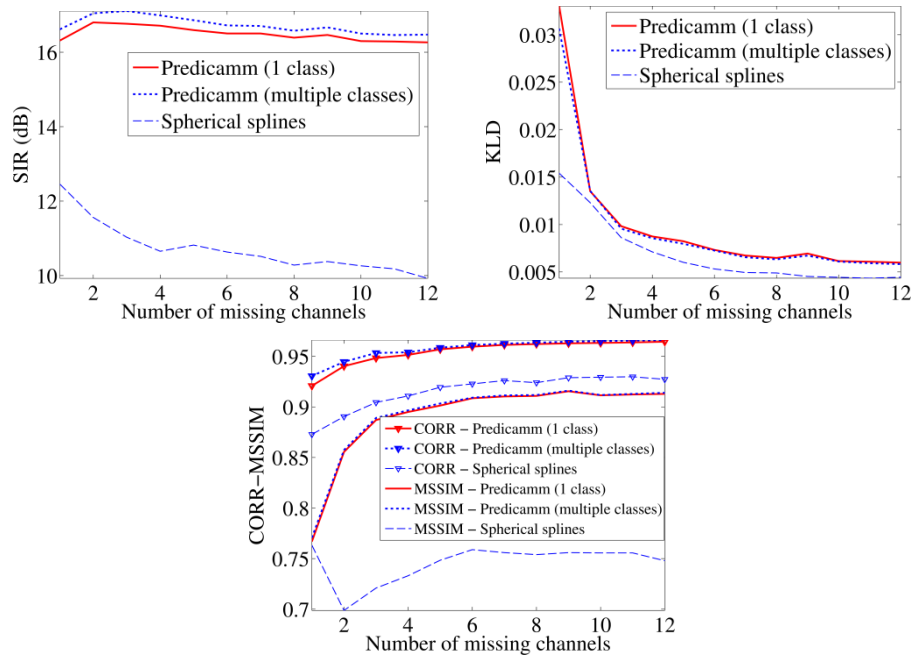

Fig. 3. Evolution of the indices for an increasing number of missing channels.
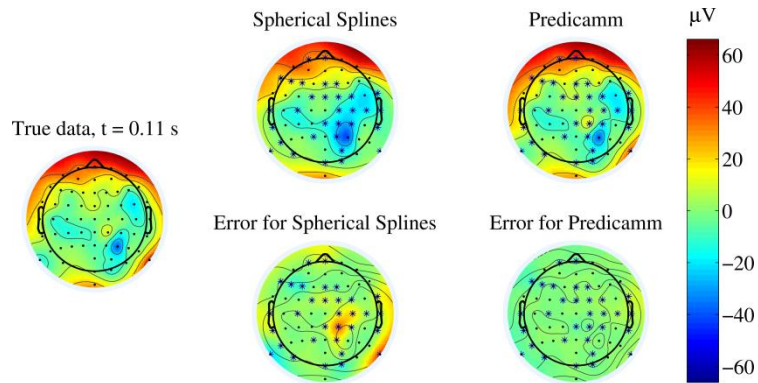

Fig. 4. Prediction results for a 32-missing-channel case for the proposed and spline methods. The missing channels positioning are highlighted by asterisk.

Table 1. Indices for the 32-missing-channel case.

	SIR (dB)	KLD	CORR	MSSIM
Splines	4.15	0.1614	0.5727	0.3771
Predicamm	15.41	0.0006	0.9337	0.9070

Conclusion

A novel ICAMM-based algorithm (Predicamm) was proposed for the prediction of EEG signals. The results of Predicamm clearly outperformed the classic spline method commonly used in EEG signal processing. ICA mixture models allow for greater flexibility, modelling local nonlinearities while keeping the general structure of the data, and thus improving the result of the prediction.

Taking into account the success of ICA application to EEG signal processing, the flexibility of ICAMM as a non-linear extension of ICA suggest it for future EEG applications. We demonstrate that the ICA mixture model can detect changes in the EEG signal, particularly in a working memory task. From this work, there are some open issues for future research such as obtaining greater insights into the sources and mixing matrices for EEG signal analysis. One of the problems with these methods is the high computational burden involved, especially if an on-line monitoring is required. Currently, we are researching in the parallelization of the algorithms in order to implement them in multiprocessor structures.

References

1. P. Common, and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, USA, 2010.
2. A. Salazar, L. Vergara, A. Serrano, and J. Igual, "A general procedure for learning mixtures of independent component analyzers," *Pattern Recognition*, vol. 43, no. 1, pp. 69-85, 2010.
3. T.W. Lee, M.S. Lewicki, and T.J. Sejnowski, "ICA mixture models for unsupervised classification of non-gaussian classes and automatic context switching in blind signal separation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1078-1089, 2000.
4. A. Salazar, and L. Vergara, "ICA mixtures applied to ultrasonic nondestructive classification of archaeological ceramics", *Eurasip Journal on Advances in Signal Processing*, vol. 2010, article ID 125201, 11 pages, (2010), doi:10.1155/2010/125201.
5. C. Klein and B. Feige, "An independent component analysis (ICA) approach to the study of developmental differences in the saccadic contingent negative variation," *Biological Psychology*, vol. 70, pp. 105-114, 2005.
6. S. Makeig, M. Westerfield, T.P. Jung, J. Covington, J. Townsend, T.J.Sejnowski and E. Courchesne, "Functionally Independent Components of the Late Positive Event-Related

Potential during Visual Spatial Attention," *Journal of Neuroscience*, vol. 19, no. 7, pp. 2665-2680, 1999.

7. M. Wibral, G. Turi, D.E.J. Linden, J. Kaiser, and C. Bledowski, "Decomposition of working memory-related scalp ERPs: Crossvalidation of fMRI-constrained source analysis and ICA," *Internt J. of Psychol*, vol. 67, pp. 200-211, 2008.

8. N.P. Castellanos and V.A. Makarov, "Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis," *Journal of Neuroscience Methods*, vol. 158, pp. 300-312, 2006.

9. A. Salazar, L. Vergara, and R. Miralles, "On including sequential dependence in ICA mixture models," *Signal Processing*, vol. 90, pp. 2314-2318, 2010.

10. P. Dayan and L.F. Abbot, *Theoretical neuroscience: computational and mathematical modeling of neural systems*, The MIT Press, 2001.

11. S. Sternberg, "High-speed scanning in human memory," *Science*, vol. 153, no. 3736, pp. 652-654, 1966.

12. S. Raghavachari, J.E. Lisman, M. Tully, J.R. Madsen, E.B. Bromfield and M.J.Kahana, "Theta oscillations in human cortex during a working-memory task: evidence for local generators", *J. of Neurophys.*, vol. 95, pp. 1630-1638, 2006.

13. J.M. Gorriz, C.G. Puntonet, G. Salmeron, and E.W. Lang, "Time series prediction using ICA algorithms," in *Proc. of 2nd IEEE Internat. W. on Intellig Data Acquisition and Advanc. Comp. Systems: Tech. and App.*, 2003, pp. 226-230.

14. C.-T. Lin, W.-C. Cheng, and S.-F. Liang, "An On-line ICA-Mixture-Model-Based Self-Constructing Fuzzy Neural Network," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 1, pp. 207-221, January 2005.

15. T.W. Lee, M. Girolami, and T.J. Sejnowski, "Independent component analysis using an extended InfoMax algorithm for mixed sub-gaussian and super-gaussian sources", *Neural Computation*, vol. 11, no. 2, pp. 417-441, 1999.

16. F. Perrin, J. Pernier, D. Bertrand, and J.F. Echallier, "Spherical splines for scalp potential and current density matching," *Electroencep and Clin Neurophys* vol. 72, pp. 184-187, 1989.

17. Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13 no. 4, pp. 600-612, 2004.