

INVESTIGATION OF ICA ALGORITHMS FOR FEATURE EXTRACTION OF EEG SIGNALS IN DISCRIMINATION OF ALZHEIMER DISEASE

Jordi Solé-Casals

*Signal Processing Group, University of Vic, Sagrada Família 7, 08500 Vic, Spain
jordi.sole@uvic.cat*

François Vialatte, Zhe Chen, Andrzej Cichocki

*RIKEN Brain Science Institute, LABSP, 2-1 Hirosawa, Saitama, 351-0106 Wako-Shi, Japan
{fvialatte, zhechen, cia}@brain.riken.jp*

Keywords: EEG, Alzheimer disease, ICA, BSS, Feature extraction

Abstract: In this paper we present a quantitative comparisons of different independent component analysis (ICA) algorithms in order to investigate their potential use in preprocessing (such as noise reduction and feature extraction) the electroencephalogram (EEG) data for early detection of Alzheimer disease (AD) or discrimination between AD (or mild cognitive impairment, MCI) and age-match control subjects.

1 INTRODUCTION

Independent component analysis (ICA) is a method for recovering underlying signals from linear mixtures of those signals. ICA draws upon higher-order signal statistics to determine a set of "components" which are maximally independent of each other.

The aims of this paper is to investigate which ICA algorithm is best adapted to deal as a preprocessing stage with EEG signals. In order to do that, we made different experiments with EEG data from Alzheimer and age-match control subjects. The evaluation was calculated in terms of measure of receiver operating characteristic (ROC) score.

The paper is organized as follows: in Section 2 we present experimental data characteristics used in the experiments. Section 3 is devoted to procedure and ICA algorithms used. In Section 4 we explain the measure that we will use for obtaining the experimental results, that are presented in Section 5. Finally, conclusions are presented in Section 6.

2 EXPERIMENTAL DATA

In the course of a clinical study, multichannel EEG recordings (Deltamed EEG machine) were recorded from 33 elderly patients affected by Alzheimer's disease and followed clinically (labeled AD set) and

from 39 age-matched controls (labeled Control set), with electrodes located on 19 sites according to the 10-20 international system. This database was recorded in normal routine. Reference electrodes were placed between Fz and Cz, and between Cz and Pz. The sampling frequency was 256 Hz, with band-pass filter 0.17-100 Hz. Three periods of 5 seconds were selected in a "rest eyes-closed" condition for each patients. In selecting these three independent sessions, an artifact rejection procedure was used to help minimize the artifact effect.

3 ICA AND BSS

3.1 Procedure

At the first stage, we apply principal component analysis (PCA) to perform dimensionality reduction. At the second stage, an ICA algorithm is implemented to perform BSS. The estimated output signal \mathbf{y}_t are assumed to be the source signals of interest up certain scaling and permutation ambiguity.

In addition, if we are only interested in denoising or getting rid of specific component, we can set that specific output signal (say y_i) to zero while keeping other components intact, and apply back projection procedure to recover the original scene. In our experiments, in ranking the output components, we al-

ways select the one that has the least absolute kurtosis value (i.e., the one close to Gaussian by assuming zero kurtosis statistic for Gaussian signal, positive kurtosis statistic for super-Gaussian signal, and negative kurtosis for sub-Gaussian signal).

3.2 Selection of candidate algorithms

For comparison, we have selected seven representative ICA algorithms. The selection criteria of them are based on several factors: (i) computationally efficiency; (ii) robustness; (iii) fewer degree of freedom (such as the choices of learning rate parameter, non-linearity, or number of iterations); (iv) preference to batch method.

Specifically, the following seven ICA/BSS algorithms are among some of most popular BSS methods in the literature: AMUSE, SOBI, JADE, Pearson-ICA, Thin-ICA, CCA-BSS and TFD-BSS.

The detailed description of algorithms are neglected here; for relevant references, see (Cichocki and Amari, 2002). All of algorithms are implemented in MATLAB, some of them are available for download from the original contributors (Cichocki et al.,).

For each algorithm, we have varied the number of independent components (namely, n), from 3 to 10, to extract the resultant uncorrelated or independent components.

4 PERFORMANCE EVALUATION

In signal detection/classification theory, a receiver operating characteristic (ROC) is a graphical plot of the sensitivity vs (1-specificity) for a binary classifier system as its discrimination threshold is varied. The ROC can also be represented equivalently by plotting the fraction of true positives (TP) vs the fraction of true negatives (TN). Nowadays, the usage of ROC has become a common measure to evaluate the discrimination ability of the feature or classifier. Roughly, the discrimination ability or performance is measured by the area value underneath the ROC curve, the greater the value, the better is the performance (with 1 denoting perfect classification, and 0.5 denoting pure random guess).

Since the primary purpose here is to evaluate the features extracted from different ICA algorithms, we have focused on the comparison between ICA algorithms and the choice of number of independent components. In order to obtain the baseline, we choose two simple yet popular linear classifiers—the linear discriminant analysis (LDA) and linear perceptron.

In calculating the ROC score, we have employed the leave-one-out (LOO) procedure.

The features we use to feed the linear classifier are the power values extracted from different frequency bands (θ, α, β , and γ). The ROC score is first calculated by using raw EEG data without any ICA preprocessing; this ROC score is regarded as a baseline for further comparison. For ICA feature extraction, we conduct the procedures of dimensionality reduction, source separation, component rejection, followed by backward projection. For each algorithm, we calculate their ROC score by varying the number of independent components from 3 to 10. Note that all the discrimination tasks are binary classification: AD against control subjects.

5 EXPERIMENTAL RESULTS

First, we calculated the ROC score for all ICA algorithm with varying number of independent components. All algorithms follow the similar-shape trend: compared to baseline, there is a positive gain in high-frequency bands using ICA; while for low-frequency bands, there is no need for using ICA because of their negative gains. In fact, the result is consistent with what was expected: since the SNR is poor in high-frequency bands, eliminating the independent component with the least absolute value of kurtosis would lead to a gain in SNR; consequently, the ROC score or its gain is greater.

Next, the comparison was conducted on three individual 5-second sessions' EEG recordings. By averaging these three independent data set, we also obtain the performance comparison. It can be seen from these results that for all independent data sets, the performance depends on the choice of the ICA algorithm as well as the choice of components. On the other hand, it is also obvious that by using ICA algorithms for feature extraction, it is possible to boost the ROC score performance (w.r.t. the baseline) around $\frac{0.7467-0.6193}{0.6193} = 20.6\%$ (data set 1), 15.6% (data set 2), and 10.2% (data set 3), assuming the best ICA algorithm (with optimum number of IC) is employed. This improvement is quite significant. The averaged ROC score against the number of independent components is plotted in Figure 1.

From Table 1, several noteworthy observations are in order:

- It seems that the optimum number of IC is 4, obtaining the highest mean ROC score (averaged over all ICA algorithms) 0.6536, followed by 0.6447 (IC=6). Overall, it seems the optimal range for the number of IC is between 4 to 7.

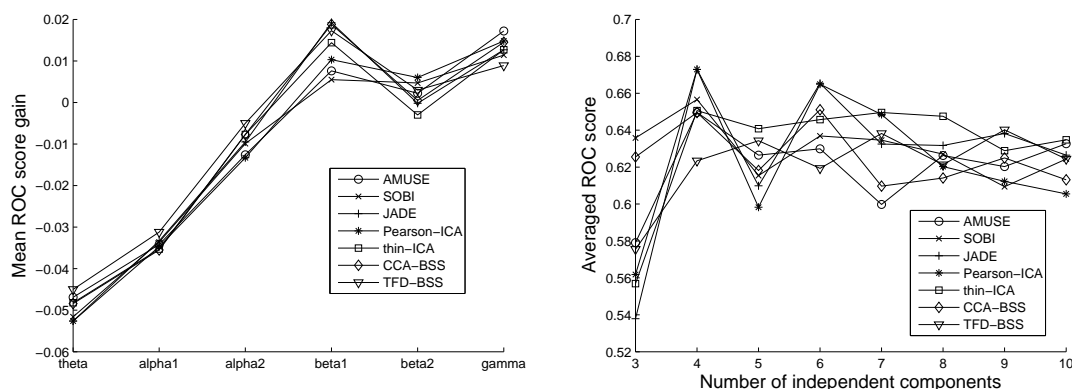


Figure 1: Left: The mean ROC score gain (averaged over 3 data bases and the number of independent components; with 0 as baseline) against frequency bands for 7 algorithms. Right: The averaged ROC score (over 3 data bases) comparison between different ICA algorithms with varying number of independent components.

- By averaging different numbers of IC, it seems the overall best ICA algorithms are Pearson-ICA and JADE (averaged from 4 to 7 components), or Thin-ICA and SOBI (averaged from 3 to 10 components).
- Overall, JADE and SOBI seem to give quite consistent performance for different number of components.

In addition, we can also compare the correct classification rate between different ICA algorithms with different setups. The results using LDA and linear perceptron classifiers are summarized in Table 2. Likewise, compared to the baseline correct classification rate, the performance with appropriate ICA algorithm give some more or less improvement.

6 CONCLUSION

In this work, we have proposed a measure or criteria to compare several popular ICA algorithms in the investigation of feature extraction of eeg signals in discrimination of Alzheimer disease. As a powerful signal processing tool used in the preprocessing step, ICA was found useful in artifact rejection, improving SNR, and noise reduction, all of which are important for the feature selection at the later stage. The ICA algorithms and the optimum choice of independent components are extensively investigated using simple linear classifiers and LOO procedure for calculating the resultant ROC scores and correct classification rate, both compared to their baselines.

It was found, in general, ICA algorithms are particularly useful for feature extraction in high frequency bands, especially on high alpha and beta

ranges; where in low frequency bands, little gain has been obtained compared to the baselines. This is more or less anticipated, because EEG signals are usually contaminated by noise at high frequency bands, but are more resistant to noise at low frequency bands. Moreover, the optimum number of selected components seem to depend on the selected algorithms, but overall observations seem to indicate the number should be in the range from 4 to 7. Interestingly, this number is consistent with our early independent investigations (Vialatte et al., 2005)). In terms of overall average performance, it seem that the JADE, SOBI, thinICA, and CCABSS algorithms give more consistent and better results.

ACKNOWLEDGEMENTS

This work has been partly funded by the Direcció General de Recerca de la Generalitat de Catalunya and by the Ministerio de Educación y Ciencia under the grant TEC2007-61535/TCM

REFERENCES

Cichocki, A. and Amari, S. (2002). *Adaptive Blind Signal and Image Processing*. Wiley, New York.

Cichocki, A., Amari, S., Siwek, K., and et al., T. T. Icalab toolboxes. <http://www.bsp.brain.riken.jp/ICALAB>.

Vialatte, F., Cichocki, A., Dreyfus, G., Musha, T., Rutkowski, T., and Gervais, R. (2005). Blind source separation and sparse bump modelling of time frequency representation of eeg signals: New tools for early detection of alzheimer's disease. In *Proc. IEEE Workshop on Machine Learning for Signal Processing*, pp. 27–32.

Table 1: The ROC score comparison between ICA algorithms by averaging the results from three 5-second sessions. The baseline value (without ICA) of ROC score is 0.63. The bold fonts indicate the top two winners or the maximal two values in each column.

no. IC	AMUSE	SOBI	JADE	Pearson-ICA	Thin-ICA	CCA-BSS	TFBSS
3	0.5791	0.6358	0.5379	0.5618	0.5569	0.6255	0.5755
4	0.6496	0.6566	0.6729	0.6729	0.6506	0.6496	0.6234
5	0.6265	0.6154	0.6099	0.5983	0.6408	0.6182	0.6343
6	0.6299	0.6369	0.6654	0.6649	0.6457	0.6511	0.6193
7	0.5998	0.6346	0.6325	0.6485	0.6496	0.6097	0.6382
8	0.6263	0.6270	0.6317	0.6203	0.6475	0.6141	0.6213
9	0.6203	0.6094	0.6382	0.6123	0.6289	0.6250	0.6402
10	0.6327	0.6244	0.6265	0.6055	0.6348	0.6131	0.6244
	average from {2,3,4,5} rows						
	0.6292	0.6210	0.6498	0.6544	0.6385	0.6094	0.6364
	average from all rows						
	0.6205	0.6300	0.6269	0.6231	0.6318	0.6258	0.6221

Table 2: Classification results using leave-one-out procedure. The bold font indicate the maximum value in each column.

no. IC	AMUSE	SOBI	JADE	Pearson-ICA	Thin-ICA	CCA-BSS	TFBSS
	LDA			baseline value (without ICA) 75%			
3	65.2778	65.2778	65.2778	63.8889	62.5000	61.1111	62.5000
4	69.4444	73.6111	70.8333	69.4444	68.0556	69.4444	61.1111
5	68.0556	66.6667	77.7778	69.4444	75.0000	73.6111	69.4444
6	72.2222	72.2222	76.3889	75.0000	70.8333	77.7778	69.4444
7	68.0556	73.6111	70.8333	76.3889	72.2222	72.2222	72.2222
8	73.6111	76.3889	72.2222	73.6111	72.2222	70.8333	70.8333
9	76.3889	73.6111	68.0556	73.6111	76.3889	69.4444	70.8333
10	70.8333	75.0000	73.6111	72.2222	77.7778	72.2222	76.3889
	linear perceptron			baseline value (without ICA) 62.5%			
3	59.7222	51.3889	54.1667	54.1667	45.8333	68.0556	54.1667
4	62.5000	70.8333	65.2778	70.8333	68.0556	56.9444	54.1667
5	59.7222	62.5000	56.9444	56.9444	68.0556	62.5000	62.5000
6	65.2778	62.5000	70.8333	65.2778	68.0556	65.2778	62.5000
7	59.7222	56.9444	65.2778	65.2778	62.5000	65.2778	65.2778
8	65.2778	62.5000	59.7222	59.7222	65.2778	65.2778	62.5000
9	62.5000	62.5000	70.8333	56.9444	68.0556	62.5000	62.5000
10	62.5000	62.5000	68.0556	59.7222	62.5000	65.2778	59.7222