IDIAP RESEARCH REPORT

Generative Independent Component Analysis for EEG Classification

Silvia Chiappa and David Barber

IDIAP–RR 04-77

January 2005

submitted for publication

IDIAP Research Institute **www.idiap.ch**
Rue du Simplon 4 P.O. Box 592 1920 Martigny − Switzerland
Tel: +41 27 721 77 11 Fax: +41 27 721 77 12 Email: info@idiap.ch

IDIAP Research Report 04-77

Generative Independent Component Analysis for EEG Classification

Silvia Chiappa and David Barber

January 2005

submitted for publication

e present an application of Independent Component Analysis (ICA) to the discrimination of mental tasks for EEG-based Brain Computer Interface systems. ICA is most commonly used with EEG for artifact identification with little work on the use of ICA for direct discrimination of different types of EEG signals. By viewing ICA as a generative model, we can use Bayes' rule to form a classifier. This enables us also to investigate whether simple spatial information is sufficiently informative to produce state-of-the-art results when compared to more traditional methods based on using temporal features as inputs to off-the-shelf classifiers. Experiments conducted on two subjects suggest that knowing 'where' activity is happening alone gives encouraging results.

1 Introduction

EEG-based Brain Computer Interface (BCI) systems allow a person to control devices by using the electrical activity of the brain, recorded by electrodes placed over the scalp. In the particular case of systems based on spontaneous brain activity, the user concentrates on different mental tasks (e.g. imagination of hand movement) which are associated with different device commands. Tasks are usually selected so that different brain areas become active while performing each one. In addition to 'where' activity is, 'what' the activity is (or indeed, absence of activity) may also be characteristic for a certain task. Activity is usually indicated by the absence of rhythmic components, which are mostly prominent in the α band (8-13 Hz). Standard approaches extract the frequency content of the signal, which is then processed by a static classifier (see [5] for a general introduction on BCI research). In this paper we try to answer the question whether the discrimination of mental tasks can be based essentially on spatial information alone.

Signals v_t^j recorded at time t at scalp electrodes $j = 1, \ldots, V$ are commonly considered as a linear and instantaneous superposition of sources h_t^i, $i = 1, \ldots, H$, in the cortex. For technical reasons, we assume that this process is noiseless, so that $v_t^j = \sum_{i=1}^{H} w_{ji} h_t^i$. Furthermore, we make the assumption that different independent brain processes underly the observed signal v_t. For these reasons ICA seems an appropriate model for EEG signal and has been extensively applied to related tasks, such as the identification of artifacts and the analysis of the underlying brain sources.

The central aim of this paper is to use directly a simple ICA generative model of EEG signals as a classifier. This is in sharp contrast to more traditional approaches, which commonly view ICA-type methods only as a preprocessing step. Some work in this direction has already been presented in [4], where the authors introduced a combination of Hidden Markov Models and Independent Component Analysis for biomedical signals. In that model, the distribution of the observations depends on a hidden variable, which is associated with an ICA model. The authors present a graphical demonstration of the method's applicability to the detection of change between two mental conditions: baseline activity and imaginary movement. Using that paper as a basis, we further investigate the use of ICA for classification. However, we use a simplified model with no temporal dependence, since we are here interested critically in whether or not the spatial information is a reliable indicator of the task, without the need to explicitly search for the presence of task-dependent temporal features.
Our approach will be to fit, for each person, an ICA generative model to each separate task, and then use Bayes' rule to form directly a classifier. This will be compared with two more standard techniques: the Multilayer Perceptron (MLP) and Support Vector Machine (SVM) [1] trained with power spectral density features.

2 ICA Model

Generative Independent Component Analysis is a probabilistic model in which a vector of observations v_t is considered to be generated by statistically independent (hidden) random variables h_t via an instantaneous linear transformation, $v_t = Wh_t + \eta_t$. For reasons of computational tractability, we restrict ourself to the limit of zero noise $\eta_t = 0$. Hence $p(v_t|h_t) = \delta(v_t - Wh_t)$, where δ is the delta function. It is also convenient to consider square W, so that $V = H$.

Unlike HMMICA [4] and Contextual ICA [3], we assume temporal independence between the hidden

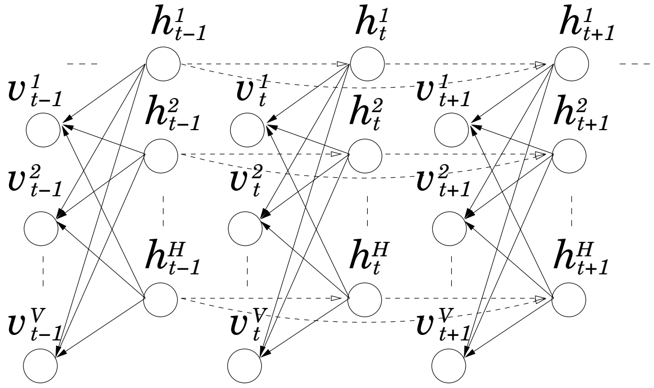

Figure 1: Graphical representation of an ICA model.

variables h_t, and have a separate model for class of task c:

$$p(v_{1:T}, h_{1:T}|c) = \prod_{t=1}^{T} p(v_t|h_t, c) \prod_{i=1}^{H} p(h^i_t|c) = \prod_{t=1}^{T} \delta(v_t - W_c h_t) \prod_{i=1}^{H} p(h^i_t|c)\,. \tag{1}$$

Here $p(h^i_t|c)$ is the prior distribution of the activity of source i, and is assumed to be stationary. Following a Maximum Likelihood (ML) approach, we can integrate (1) over the hidden variables h_t:

$$p(v_{1:T}|c) = \prod_{t=1}^{T} \int_{h_t} \delta(v_t - W_c h_t) \prod_{i=1}^{H} p(h^i_t|c) = |\det W_c|^{-T} \prod_{t=1}^{T}\prod_{i=1}^{H} p(h^i_t|c)\,,$$

where $h_t = W_c^{-1} v_t$. As is well known, it is not necessary to accurately model the source distribution $p(h^i_t|c)$ in order to correctly estimate W_c [2]. Indeed, statistical consistency of estimating W_c can be guaranteed using only two types of fixed prior distributions: one for modeling sub-Gaussian and another for modeling super-Gaussian h^i_t. However the aim of this work is to perform classification, for which an appropriate model for the source distribution is fundamental. A general distribution family which encompasses many types of symmetric and unimodal distributions is the generalized exponential[1]:

$$p(h^i_t|c) = \frac{f(\alpha^{ic})}{\sigma^{ic}} \exp\Big(- g(\alpha^{ic}) \Big|\frac{h^i_t}{\sigma^{ic}}\Big|^{\alpha^{ic}} \Big)\,,$$

where

$$f(\alpha^{ic}) = \frac{\alpha^{ic}\Gamma(3/\alpha^{ic})^{1/2}}{2\Gamma(1/\alpha^{ic})^{3/2}}\,, \quad g(\alpha^{ic}) = \Big(\frac{\Gamma(3/\alpha^{ic})}{\Gamma(1/\alpha^{ic})}\Big)^{\alpha^{ic}/2}$$

and $\Gamma(\cdot)$ is the Gamma function. Although unimodality appears quite a restrictive assumption, our experience on the tasks we consider is that it is not inconsistent with the nature of the underlying sources, as revealed by an analysis of the posterior distribution $p(h^i_t|v_t, c)$. The parameter σ is the standard deviation[2], while α determines the sharpness of the distribution[3]. The log-likelihood as a function of the parameters is:

$$\log p(v_t|h_t, c) = -T \log|\det W_c| + \sum_{t=1}^{T}\sum_{i=1}^{H} \log p(h^i_t|c)\,.$$

[1]We zero mean the data, hence we can assume that the distribution is zero mean.

[2]Due to the indeterminacy of variance of the h^i (h^i can be multiplied by a scaling term a as long as the corresponding column of W_c is multiplied by $1/a$), σ could be set to one in the general model described above. However this cannot be done in the constrained version $W_c = W$ considered in the experiments (see Sec. 3).

[3]$\alpha < 2$, $\alpha = 2$, $\alpha > 2$ describe super-Gaussian, Gaussian or sub-Gaussian pdf respectively.

Assuming independent, identically distributed data, the above log-likelihood function is summed over all training patterns $v_{1:T}$ belonging to each class. We use the scaled conjugate gradient method [1] for updating the parameters after computing the derivatives[4]. After training, a novel test sequence $v^*_{1:T}$ is classified using Bayes' rule $p(c|v^*_{1:T}) \propto p(v^*_{1:T}|c)$, assuming $p(c)$ is uniform.

3 Experiments

EEG potentials were recorded with the Biosemi ActiveTwo system (http://www.biosemi.com), using 32 electrodes located at standard positions of the 10-20 International System, at a sample rate of 512 Hz. The raw potentials were re-referenced to the Common Average Reference in which the mean over the all channels is removed from each channel. Subsequently, the band 6-16 Hz was selected with a Butterworth filter. This preprocessing filter is somewhat inelegant, but is a simple way to remove strong drift terms in the signals, which are artifacts of instrumentation (the so-called DC level), and do not correspond to brain activity. Experimentally, we also found that removing frequencies outside the band 6-16 Hz robustified the performance. Only 19 of the 32 electrodes, namely those covering the temporal-motor cortex were considered for the analysis (see Fig. 2). The data was acquired in an unshielded room from two healthy subjects without any previous experience with BCI systems. During an initial day the subjects learned how to perform the mental tasks. In the following two days, 10 recordings, each lasting around 4 minutes, were acquired for the analysis. During each recording session, every 20 seconds an operator instructed the subject to perform one of three different mental tasks. The tasks were: (1) imagination of self-paced left, (2) right hand movement and (3) mental generation of words starting with a given letter.

The time series obtained from each recording session was split into segments of signal lasting half/one second. ICA was compared with two standard approaches, in which for each segment the power spectral density was extracted and then processed using a (softmax) MLP and a SVM [1][5]. The first three sessions of each day were used for training the models while the other two sessions where used alternatively for validation and testing[6]. Since, we assume that the scalp signal is generated by linear mixing of sources in the cortex, provided the data are acquired under the same conditions, it would seem reasonable to further assume that the mixing should be the same for all classes, $W_c = W$, and this constrained version is also considered. A comparison of the performance of our spatial ICA method against the more traditional methods using temporal features is shown in Table 1. ICA consistently performs at least as well as the temporal feature approach using MLP and SVMs.

For each subject, we used one day's data to select the two hidden components h^i whose distribution

[4]To keep the notation simple, we assume that all training patterns belonging to each class c are concatenated into a single sequence $v_{1:T_c}$. Thus the total log-likelihood L is given by the sum of the log-likelihoods of $v_{1:T_c}$ over all classes. Dropping the component index i and the class index c:

$$\frac{\partial L}{\partial \sigma} = -\frac{T}{\sigma} + \frac{g(\alpha)\alpha}{\sigma^{\alpha+1}}\sum_{t=1}^{T}|h_t|^\alpha \,, \text{that is} \quad \sigma = \Big(\frac{g(\alpha)\alpha}{T}\sum_{t=1}^{T}|h_t|^\alpha\Big)^{1/\alpha}.$$

Using the maximum-likelihood solution for σ we obtain, using $A = W^{-1}$:

$$\frac{\partial L}{\partial \alpha} = \frac{T}{\alpha} + \frac{T}{\alpha^2}\frac{\Gamma(1/\alpha)'}{\Gamma(1/\alpha)} + \frac{1}{\alpha^2}\log\Big(\frac{\alpha\sum_{t=1}^{T}|h_t|^\alpha}{T}\Big) - \frac{T}{\alpha\sum_{t=1}^{T}|h_t|^\alpha}\sum_{t=1}^{T}|h_t|^\alpha\log(h_t)\,,$$

$$\frac{\partial L}{\partial A} = -\sum_{t=1}^{T} b_t v_t^T + T(A^T)^{-1} \quad \text{where} \quad b_t^i = \frac{g(\alpha^i)}{(\sigma^i)^{\alpha^i}}\alpha^i \mathrm{sign}(h_t^i)|h_t^i|^{\alpha^i-1}\,.$$

[5]The best performance was obtained using the following Welch's periodogram method: each pattern was divided into a quarter of second length windows with an overlap of 1/8 of second. Then the average of the power spectral density over all windows was computed.

[6]A one hidden layer MLP was trained using cross-entropy, with a validation set used to choose the number of iterations, number of tanh hidden units (ranging from 1 to 100) and the learning rate. In the SVM, each class was trained against the others, and the standard deviation for the Gaussian SVM found using a validation set (ranging from 1 to 20000).

varied most across the three classes, using the ICA model with a matrix W common to all classes. The projection of each component on the 19 channels (absolute value of the i-th column of W) can indicate which part of the scalp received more activity from the component. The distributions and scalp projections are shown in Fig. 2. Visually, the projections of components a_1 and b_1 are most similar. For these two components, the word task (green) has the strongest activation (width of the distribution), followed by the left task (blue) and the right task (red). Gratifyingly, this suggests that for these two subjects, a similar spatial pattern of activity occurs when they are asked to perform the tasks. To a lesser extent, visually components a_2 and b_2 are similar in their scalp projection, and again the order of class activation in the two components is the same (word task followed by right and left tasks).

	Subject A				Subject B			
	Day 2		Day 3		Day 2		Day 3	
	1/2 s	1 s	1/2 s	1 s	1/2 s	1 s	1/2 s	1 s
ICA W	42.5%	39.0%	39.5%	36.0%	33.7%	29.0%	36.2%	31.3%
ICA W_c	39.3%	37.8%	38.7%	35.5%	32.3%	24.9%	35.4%	30.6%
MLP	44.9%	37.1%	40.4%	38.1%	40.3%	30.5%	44.6%	34.2%
SVM	39.6%	35.1%	42.0%	38.1%	43.0%	32.4%	39.4%	36.6%

Table 1: Test errors in classifying three mental tasks for Subjects A and B using ICA with a matrix W common to all classes (ICA W), ICA with a separate matrix for each class (ICA W_c), MLP and SVM. The first/second column of each day indicates the error rate using half/one second of data (around 840/420 test examples).

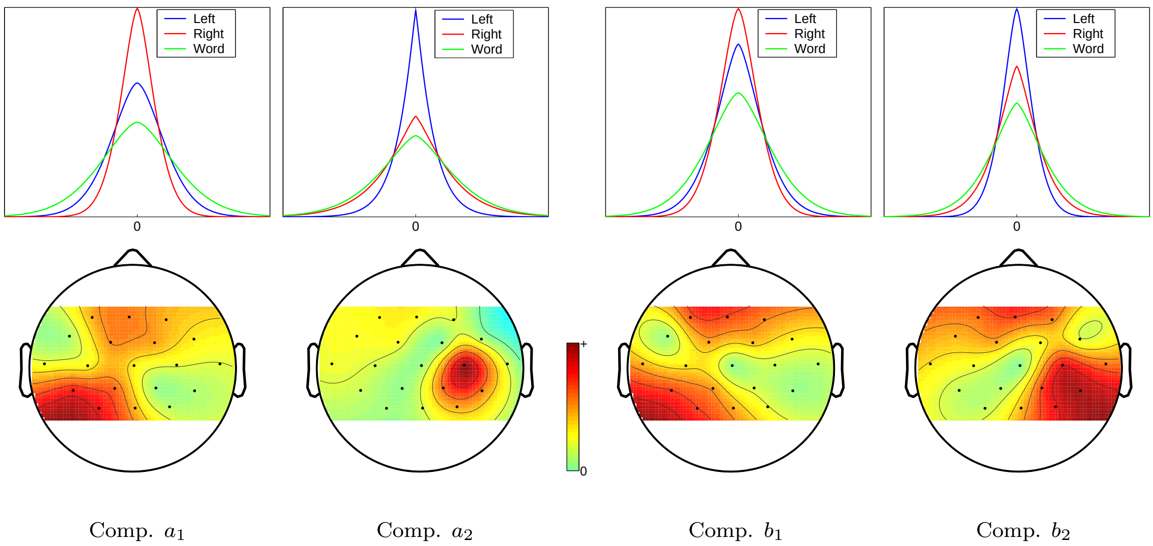

Figure 2: Estimated pdfs of the three classes and absolute projection on the scalp of two hidden components for Subject A, Day 3 (Comp. a1, Comp. a2) and Subject B, Day 2 (Comp. b1, Comp. b2). The topographic plots have been obtained by interpolating the values at the electrode (black dots) using the open source eeglab toolbox (http://www.sccn.ucsd.edu/eeglab). Due to the indeterminacy of variance of the hidden components, axes scale between different figures cannot be compared and has been removed. This also applies to the absolute scalp projection.

4 Conclusions

In this work we have presented a preliminary analysis on the use of a purely spatial Independent Component Analysis model for the discrimination of mental tasks for EEG-based BCI systems. We have compared ICA with two other standard approaches, where temporal information from a window of data (power spectral density) is extracted and then processed using a static classifier. Our results suggest that spatial information alone is indeed powerful enough to produce state-of-the-art performance.

More sophisticated ICA approaches which take into account temporal information have been proposed in the literature. In [3] for example, the hidden components are modeled by an autoregressive process. It would be interesting to investigate whether this information can bring any advantage in terms of discrimination. Additionally, more complex source distributions may bring performance benefits. A key research issue is how to avoid using an initial filtering preprocessing step and make a consistent generative model of the raw data signal.

5 Acknoledment

This work was supported by the Swiss NSF through the NCCR IM2 and by the PASCAL Network of Excellence, IST-20002-506778, funded in part by the Swiss OFES. The authors would like to thank S. Bengio for discussions.

References

[1] C. M. Bishop. *Neural Networks for Pattern Recognition.* Oxford Univ. Press, 1995.

[2] J.-F. Cardoso. On the stability of source separation algorithms. In *Workshop on Neural Networks for Signal Processing*, pages 13–22, 1998.

[3] B. A. Pearlmutter and L. C. Parra. Maximum likelihood blind source separation: A context-sensitive generalization of ICA. In *Advances in Neural Information Processing Systems*, pages 613–619, 1997.

[4] W. D. Penny, S. J. Roberts, and R. M. Everson. Hidden Markov independent components for biosignal analysis. In *International Conference on Advances In Medical Signal and Information Processing*, pages 244–250, 2000.

[5] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113:767–791, 2002.