Reference를 갖는 ICA를 이용한 자동적 P300 검출

최희열[O], 최승진
포항공과대학교 컴퓨터공학과
{hychoi[O], seungjin}@postech.ac.kr

Automatic P300 Detection using ICA with Reference

Heeyoul Choi[O] and Seungjin Choi
Dept. of Computer Science and Engineering, POSTECH, Korea

Abstract

The analysis of EEG data is an important task in the domain of Brain Computer Interface (BCI). In general, this task is extremely difficult because EEG data is very noisy and contains many artifacts and consists of mixtures of several brain waves. The P300 component of the evoked potential is a relatively evident signal which has a large positive wave that occurs around 300 msec after a task-relevant stimulus. Thus automatic detection of P300 is useful in BCI. To this end, in this paper we employ a method of reference-based independent component analysis (ICA) which overcomes the ordering ambiguity in the conventional ICA. We show here that ICA incorporating with prior knowledge is useful in the task of automatic P300 detection.

1. Introduction

It is very interesting to control a system only by thinking. To make such a BCI system, EEG analysis is an important task. But EEG data is very noisy and contains many artifacts and consists of mixtures of several brain waves, so this task is very difficult. Contrary to other components of ERP, P300 is relatively evident component to analyze. Moreover in a BCI system under Virtual Reality environment, the automatic detection of P300 plays an important role.

Kalman Filter or ICA was successfully used in the detection of P300 [1]. However, these methods require a manual processing which selects the best component resembling P300. This manual processing might be cumbersome in the automatic detection of P300.

In this paper, we apply a method of *ICA with reference* [3] where the prior knowledge for the signal of interest is used, in order to build an automatic P300 detection system.

Section 2 describes what P300 is and Section 3 reviews the conventional ICA and its limitation. In Section 4, we describe the constrained ICA and ICA with reference. Finally, experimental results confirm that our proposed method is useful in the task of automatic P300 detection.

2. P300

Event-Related Potentials (ERP) comprises a part of Electroencephalogram (EEG), and P300 is one of ERP components (See Fig 1).

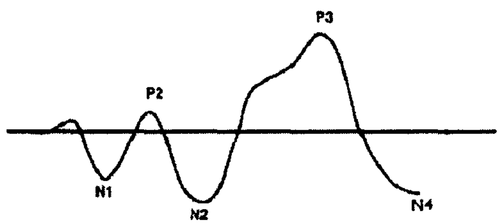

Fig 1. Components of ERP

The P300 component is a large positive wave that peaks around 300msec after presentation of a stimulus when the subject is actively attending to

that stimulus. It is modality-nonspecific and the less probable event is, the larger amplitude. It seems to be generated in the hippocampal region and was discovered in 1965 by Sutton, Braren, Zubin, John.
It is associated with memory and learning and reflects the updating of working memory that is called context updating. And the latency reflects the duration of stimulus evaluation
It is related to the aging process, dementia (Alzheimer's disease), degenerative disorders (Parkinson's disease).

3. Conventional ICA

To detect P300 component more exactly, we need to separate sources independently. That is, BSS methods can be applied.
BSS or ICA can be illustrated as a probability density matching problem. Let us denote the observed density and model density by $p^o(x)$ and $p(x)$, respectively. As an optimization function to find A which best match $p^o(x)$ and $p(x)$, the Kullback-Leibler divergence is considered [4]. This gives the risk R that has the form

$$R = KL\left[p^0(x) \| p(x)\right] = \int p^0(x) \log \frac{p^o(x)}{p(x)} dx \qquad (1)$$

And, the loss function L is

$$L = \log\left|\det A\right| - \sum_{i=1}^{n} \log p_i(s_i) \qquad (2)$$

where log $p^0(x)$ was neglected since it does not depend on A. Popular ICA algorithms were derived from the minimization of the loss function (2) using the natural gradient [5].
In ICA, the number of hidden sources is equal to the number of observed channels. This is the reason why we have to manually select one channel that reflects P300 after ICA. This conventional ICA is not enough to develop a system that performs real-time EEG signal analysis in order to generate control commands for environmental control, communication, or even simple driving commands.

4. ICA with Reference

Conventional ICA only defines the directions of independent components. That is, ICA has such an inherent indeterminacy on dilation and permutation. We can give some constraints such as order and scale to ICA. And then Lagrange multiplier methods are used.

4.1 Constrained ICA [2]

*. Ordering (inequality constraint)
Object: Minimize Mutual information
Subject to: $g(W) <= 0$ where $g_i(W) = I(u_i+1) - I(u_i)$

$I(u_i)$ is the index of some statistical measures of output component u_i, e.g. variance, normalized kurtosis

$$W = W + W^T + \Psi(u)\, x^T$$

Variance : ordered by magnitude of information
Kurtosis : Super Gaussian, Gaussian, Sub Gaussian
The signal with the most variance shows the majority of information. However, Kurtosis has been widely used.

*. Nomalization(equality constraint)
Object: Minimize Mutual information
Subject to : $h(W) = 0$ where $h_i(w_i) = w_i^T w_i - 1$

$$W = W + W^T + \Phi(u)\, x^T + \Omega(W)$$

where $\Omega_i(w_i) = 2\,\lambda_i w_i^T$

4.2 ICA with Reference [3]

We can extract a subset of independent sources from multidimensional observations when some a priori information that can be incorporated to the learning algorithm as reference is available
As stated above, Constrained ICA is able to extract independent components in order, but is not able to extract subset of components.
Wei Lu introduced ICA with Reference that extract independent components as many as reference components.
The conventional ICA maximize the negentropy, we give a constraint to conventional ICA. Then, Lagrange multiplier methods are used to solve constrained nonlinear optimization problems.

Object : Maximize Negentropy
Subject to : $g(w) <= 0$, $h(w) = 0$
where $h(w) = E\{y^2\} - 1 = 0$, $g(w) = e(y,r) - d <= 0$,
$e(y,r)$ closeness between output y and reference

r, d is threshold.
Lagrange Function L is

$$L = f + (c/2) \| h \|^2 + \lambda h + 1/(2c) \Sigma\{max(0, G_i)^2 - u_i^2\}$$

f threshold d is too large then we have many local maximum, if d is too small, then the equation will not converge. So it may as well increase from small value.
For $e(y,r)$, $E[(y_i - r_i)^2]$ or $-E[y_i r_i]$ are widely used. The detail algorithm of Constrained ICA or ICA with Reference is in [2] and [3].

5. Experiments

Eight electrodes sites (FZ, CZ, CPZ, PZ, P3, P4, as well as 2 vertical EOG channels) were arranged on the heads of three subjects 5 with a linked mastoid reference. And 45 red light and 90 yellow light trials from each subject were classified. A combination of eye and head movement artifact is subtracted using the linear regression technique [1].
For conventional ICA, JadeR is used and a best channel was selected manually (Channel 5). As classification technique we use simple correlation with each average signals over red light and yellow light (See Fig 2) in order to compare ICA and ICA with reference.

$$Correlation = (sample * ave^T)/(|sample| * |ave|)$$

In the case of the total performance, JadeR is slightly better than ICA with reference (See Table 1). But we can confirm from Table 1 that our proposed algorithm extracted more P300 component from raw data.

6. Discussion

ICA with reference can extract some independent components which are interesting. With this advantage, we built an automatic P300 detection system. In other words, P300 component can be automatically extracted from the EEG data.

7. Acknowledgment

This work was supported in part by Korea Ministry of Science and Technology under Brain Science & Engineering Research Program and BK 21 in POSTECH.

8. References

[1] Jessica D. Bayliss, Dana H. Ballard, "Single Trial P3 Recognition in a Virtual Environment", NRL T.R. 98.1, University of Rochester National Resource Laboratory for the Study of Brain and Behavior, September, 1998
[2] Wei Lu, Jagath C. Rajapakse "Constrained Independent Component Analysis", in Advances in Neural Information Processing Systems 13 (NIPS2000), pp. 570 - 576, December 2000, USA.
[3] Wei Lu, Jagath C. Rajapakse "ICA with Reference", ICA 2001, San Diego, CA, USA, December 2001.
[4] H. Attias and C. E. Schreiner, "Blind source separation and deconvolution: The dynamic component analysis algorithms," Neural Computation, vol. 10, pp. 1373-1424, 1998
[5] S. Amari and A. Cichocki, "Adaptive blind singal processing - neural network approaches," Proc. Of the IEEE, Special Issue on Blind Identification and Estimation, vol. 86, no. 10, pp. 2026-2048, Oct. 1998.

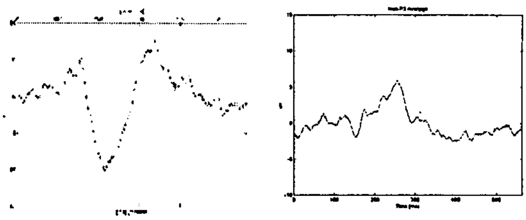

Fig 2. Average signals over Red (left) and Yellow (right)

Threshold d		0.05	0.1	0.15	0.2	0.25
JadeR	Red	74.2	69.7	66.7	62.1	57.6
	Yellow	73.7	75.9	78.8	81.8	83.5
	Total	73.8	74.5	76.1	77.5	77.8
ICA with Ref.	Red	75.8	71.2	68.2	66.7	62.1
	Yellow	60.6	63.1	64.8	68.2	69.9
	total	63.9	64.9	65.6	67.9	68.2

Table 1. Correctness (%) for JadeR and ICA with Reference according to the threshold d

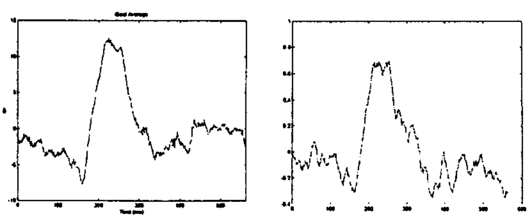

Fig 3. In the case of red light, the selected component from JadeR (left) and ICA with