Australian Journal of Basic and Applied Sciences, 5(9): 2152-2156, 2011
ISSN 1991-8178

Separation of the EEG Signal using Improved FastICA Based on Kurtosis Contrast Function

[1]Tahir Ahmad, [2]Hjh.Norma Binti Alias, [3]Mahdi Ghanbari

[1]ibnu Sina Institute For Fundamental Sciences,
[1,2,3]Theoretical and Computational Modelling for Complex Systems Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia 81310 Skudai, Johor, MALAYSIA.

Abstract: Independent component analysis (ICA) is a computational method to solve blind source separation (BSS) problem. In this paper, an improved FastICA based on Kurtosis was proposed to separate EEG signal. It was adopted from the gradient descent algorithm to improve the converge rate. Improved FastICA method was applied to separate EEG signal from a patient who had epilepsy seizure. The separation results proved that the method was feasible and effective.

Key words: independent component analysis (ICA), EEG signal, kurtosis.

INTRODUCTION

Blind source separation (BSS) is a process used in retrieving sources without any prior information given except the sensor observations. The output of the set of sensors typically gives observations where every sensor receives different mixture of source signals (A. Hyvarinen et al., 2001). Independent component analysis (ICA) is a fundamental technique that is most widely used for solving the blind source separation problem (A. Hyvarinen et al., 2001, J.V. Stone, 2004).

In this paper, we first briefly introduce ICA model and ICA algorithm with Kurtosis maximization (KM). Then, an improved FastICA based on Kurtosis contrast function is proposed. Finally, after preprocessing EEG signal by principle component analysis (PCA), new algorithm is applied to separate EEG signal. The new algorithm is compared with FastICA algorithm in term of convergence rate.

Ica Model:

In early 1980s, ICA was introduced by J.H´erault ,C.Jutten, and B.Ana but with different name (J.H´erault and B. Ans, 1984; J. H´erault et al., 1985; B. Ans et al., 1985). This technique was reviewed by Juttenand J. H´erault (1991). Let us denote the number of underlying source signal by m, and the number of observed signals by n $(n \geq m)$. We can denote the observed signals by$x_1^t, x_2^t, \dots, x_n^t$, which are amplitude of the recorded signals at time point $t$and original signals by $s_1^t, s_2^t, \dots, s_m^t$. The x_i^t are the weighted sums of s_i^t:

$$\begin{pmatrix} x_1^t \\ x_2^t \\ \vdots \\ x_n^t \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} s_1^t \\ s_2^t \\ \vdots \\ s_m^t \end{pmatrix} = A \begin{pmatrix} s_1^t \\ s_2^t \\ \vdots \\ s_m^t \end{pmatrix} \tag{1}$$

The mixing matrix is denoted by A such as elements of A are denoted by a_{ij} which are the constant coefficients. The blind source separation (BSS) problem is when we want to find the original signals from the mixtures $x_1^t, x_2^t, \dots, x_n^t$. We use the word blind, because we do not know any information about the original sources.

We consider matrix $\boldsymbol{w} = (w_1, \dots, w_n)^T$ where $w_i$has coefficients $w_{ij}, j = 1, \dots, n$. Then, we separate the estimate of s_i as:

$$\boldsymbol{y} = \boldsymbol{w}X \tag{2}$$

Where $\boldsymbol{y} = (y_1, y_2, \dots, y_m)$ and y_i is an estimate of s_i. If $m = n$ and matrix A is known then $\boldsymbol{w}$ can be found as the inverse of the matrix.

Figure 1 illustrates both the mixing and un-mixing processes involved in ICA model. The independent sources are mixed by the unknown matrix A. Vector y approximates s by estimating the un-mixing matrix **w**.

2.1 Source Separation Based On Independent:

One statistical principle used in determining matrix **w** is independence. In other words, for choosing the matrix **w** the components y_i should be statistically independent. Independent component means that we cannot obtain any information about the value of a component from another component (James V. Stone, 1999; A.

Corresponding Author: Tahir Ahmad, Theoretical and Computational Modelling for Complex Systems Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia 81310 Skudai, Johor, ALAYSIA.
E-mail: tahir@ibnusina.utm.my (Tahir Ahmad)

Hyv¨arinen et al, 2001). We know the goal of blind source separation problem is to find a linear representation so that components can be statistically independent.

We assume $(x_1^t, x_2^t, \dots, x_n^t)$ as a set of observations of random variables, where t denotes the time. We also assume that they are generated as a linear mixture of independent components as in equation (1). The purpose of independent component analysis is to estimate both the matrix mixture A ands_i^t, where we do not have any prior information just by sensor observation x_i^t. On the other hand, independent component analysis now consists of finding a linear transformation given by a matrix W as in equation (2) and the random variabley_i that they are independent. To simplify this model, we suppose that the number of independent components y_i is equal to the number of observed variables.

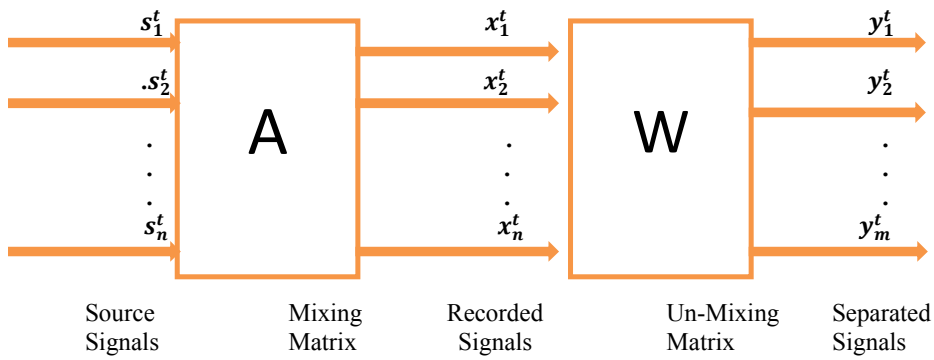

Fig. 1: ICA model block diagram. $s^t$are the sources, x^t are the recordings, y^t are the estimated sources, A is mixing matrix, W is un-mixing matrix.

2.2 Measuring Nonguassianity by Kurtosis:

When we want to use nonguassianity in ICA estimation, we should have a classic measure of non-Gaussian random variable. One of the classic measures for estimation of nongaussianity of random variable for ICA is kurtosis. We denoted the kurtosis of y by $Kurt(y)$ and it is defined by

$$Kurt(y) = E[y^4] - 3(E[y^2])^2 \tag{3}$$

Where y is a random variable with zero mean. If y is normalized, variance of y is equal to one, $E[y^2] = 1$. Then, the above formula is simplified to:

$$Kurt(y) = E[y^4] - 3 \tag{4}$$

The concept of nongaussianity for independent component analysis was first used by N. Delfosse and P. Loubaton (1995). A natural algorithm based on gradient descent was used by N. Delfosse and P. Loubatonto minimize or maximize $Kurt(w^T X)$. The convergence of new separation algorithm was proved analytically. They proved the maximal property of kurtosis as higher order statistics property.

2.3 Principal Component Analysis:

Principal component analysis (PCA) is a statistic technique to data analysis, feature extraction, and data compression(T.W.Aanderson, 1958; P.Devijver, 1982; K.I.Diamantaras and S.Y.Kung. 1996, E.Oja, 1983). The aim of PCA is to find a smaller set of variables with less redundancy from a set of multivariable measurement. The concept of correlation is used in PCA. It is much weaker than the concept of independence used in ICA because the concept of correlation in PCA is second-order statistics only. However, PCA is a useful preprocessing for ICA.

We assume that $X = (x_1, x_2, \dots, x_n)$ as a random variable with n elements. Firstly, the vector X is centered by subtracting its means:

$$X = X - E[X] \tag{5}$$

Consider a linear combination of the elements of the vector X as follows:

$$y_1 = \sum_{k=1}^{n} w_{k1} x_k = \boldsymbol{w}_1^T X \tag{6}$$

Where $w_{11}, w_{21}, \dots, w_{n1}$ are scalar coefficients elements of vector $\boldsymbol{w}_1$. Then y_1 is the first principal component of X, if the variance of y_1 is maximally large. Since growth of the variance of y_1 is based only on

orientation of the weight vector, $\boldsymbol{w}_1$,therefore we can consider the constraint of the norm $\boldsymbol{w}_1$ is equal to 1. Thus, our goal is to find a weight vector, $\boldsymbol{w}_1$ so that it maximizes the PCA criterion:

$$J_1(\boldsymbol{w}_1) = E[{y_1}^2] = E[(\boldsymbol{w}_1^T X)^2] = \boldsymbol{w}_1^T E[XX^T]\boldsymbol{w}_1 = \boldsymbol{w}_1^T C_x \boldsymbol{w}_1 \quad (7)$$

So that$\|\boldsymbol{w}_1\| = 1$,

where the C_x is the covariance matrix of the X where the covariance matrix for zero-mean vector X is equal to the correlation matrix e.g.$C_x = E[XX^T]$ and the norm of $\boldsymbol{w}_1$is the Euclidean norm that is defined as:

$$\|w_1\| = (w_1^T w_1)^{1/2} = (\textstyle\sum_{k=1}^{n} w_k^2)^{1/2} \quad (8)$$

The solution for the PCA problem based on basic linear algebra (K. I. Diamantaras and S. Y. Kung, 1996; E. Oja, 1983) is given in term of the unit-length eigenvectors $e_1, e_2, \dots, e_n$ of the covariance matrix of X. The sorting of the eigenvectors is such that the corresponding eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$satisfy$\lambda_1 > \lambda_2 > \cdots > \lambda_n$. The solution maximizing the PCA criterion is given by $\boldsymbol{w}_1 = e_1$. Thus, the first principal component of the vector variable X is $y_1 = e_1^T X$.

The m principal component can be generated by maximizing the PCA criterion that m can be any number between 1and n. The mth $(1 \le m \le n)$ principal component is denoted by $y_m = \boldsymbol{w}_m^T X$ where $\boldsymbol{w}_m$ is a unit norm weight vector so where the variance of y_m is maximized under the constraint that y_m is uncorrelated with all the previously found principal components e.g. $E[y_m y_k] = 0, k < m$.To look for the second principle component, we know that $\boldsymbol{w}_1 = e_1$ and our goal is for maximal variance $E[y_2^2] = E[(\boldsymbol{w}_2^T X)^2)]$ in the subspace orthogonal to the first eigenvector of C_x . So the $\boldsymbol{w}_2$ is given by $\boldsymbol{w}_2 = e_2$. Also, respectively, it implies that $\boldsymbol{w}_k = e_k$. Therefore the kth principal component is $y_k = e_k^T X$.

2.4 A Fixed Point Algorithm Based On Kurtosis (Fast ICA):

Assume that random vector $X = (x_1, x_2, \dots, x_n)$ is a collection of linear mixtures of independent source signals according to equation $X = AS$. Assume also that random vector Z is whitened of X. Then, A. Hyv¨arinen and E. Oja(1997) proposed the following fixed -point ICA algorithm:

Step1. Let p=1. Take a random initial vector w (0) of norm 1.

Step2. Let $w(p) = E[Z(w(p-1)^T Z)^3] - 3w(p-1)$. (9)

Step 3.$w(p) = \frac{w(p)}{\|w(p)\|}$ (10)

Step 4.If $|w(p)^T w(p-1)|$ is not close enough to 1, let $p = p + 1$ and go back to step 2. Otherwise, output the vector $w(p)$.

The final vector $w(p)$ from the algorithm is one of the independent components as the linear combination$w^T Z$.Note that the statement in step 4 means that the old and new values of w point are in the same direction. This algorithm can run for estimating every independent component analysis. But recall that columns of matrix $\boldsymbol{w}$ are orthogonal. One of the methods of orthogonal weight vectors is by projecting current solution $w(p)$ on the space orthogonal to the columns of the matrix $\boldsymbol{w}$ previously found. Define the matrix $\overline{\boldsymbol{w}}$as a matrix whose columns are previously found in columns of matrix$\boldsymbol{w}$.therefor for estimate independent component analysis one by one with the above algorithm, we can add the projection operation in the beginning of step 3:

Let $w(p) = w(p) + \overline{\boldsymbol{w}}\overline{\boldsymbol{w}}^T w(p)$. Divide $w(p)$ by its norms.

Another method of arthogonalizing the weight vectors is a symmetric orthogonalization. This means that after extracting all of the weight vectors with fixed –point algorithm, matrix $\boldsymbol{w}(p) = (w_1(p), w_2(p), \dots, w_m(p))$ can be orthogonal with the following formula:

Let $\boldsymbol{w}(p) = \boldsymbol{w}(p)(\boldsymbol{w}(p)^T \boldsymbol{w}(P))^{-1/2}$ (11)

By using singular value decomposition (SVD) method we have$\boldsymbol{w}(p)^T \boldsymbol{w}(P) = EDE^T$. Therefore, $(\boldsymbol{w}(p)^T \boldsymbol{w}(p))^{-1/2} = ED^{-1/2}E^T$.

2.5 Improved Fastica:

We know that Newton's interval method is used to obtain FastICA algorithm. Newton interval method has some advantages such as fast convergence and in simple forms while its disadvantage is that it has high demand to initial value. Simulation results show that if the initial matrix is random and inappropriate, iteration number is very high and convergence speed are unfavourable.

Our aim here is to complete Newton iteration method by introducing a method that obtains appropriate initial value at first. The gradient descent method can be used to satisfy the request. This algorithm is presented as follow:

Step1. Let p=1. Take a random initial vector w (0) of norm 1.

Step 2. Let $J(p) = E[Z(w(p-1)^T Z)^3]$. (12)

Step 3. $h_{opt} = \text{argmax}_h \left| kurt\big(w(p-1) - hJ(p)\big) \right|$. (13)

Step 4. $w(p) = w(p-1) - h_{opt}\, J(p)$. (14)

Step 5. $w(p) = \frac{w(p)}{\|w(p)\|}$

6. If $\|w(p) - w(p-1)\|$ is not close enough to 0, then $p = p + 1$ go back to 2. Otherwise $w(p)$ is appropriate.

By applying one orthogonal method, we can use this algorithm to find another weight vectors.

3. Simulation Results:

A set of EEG signal recorded from epileptic patient as shown in Figure 1 was sampled during seizure. This signal was sampled at 256 discrete data in every second by using Nicolet One EEG software. After whitening and preprocessing of EEG signal by PCA, 6 principal components were obtain. We used the improved FastICA to estimate separation matrix w. Figure 2 shows the separation results.

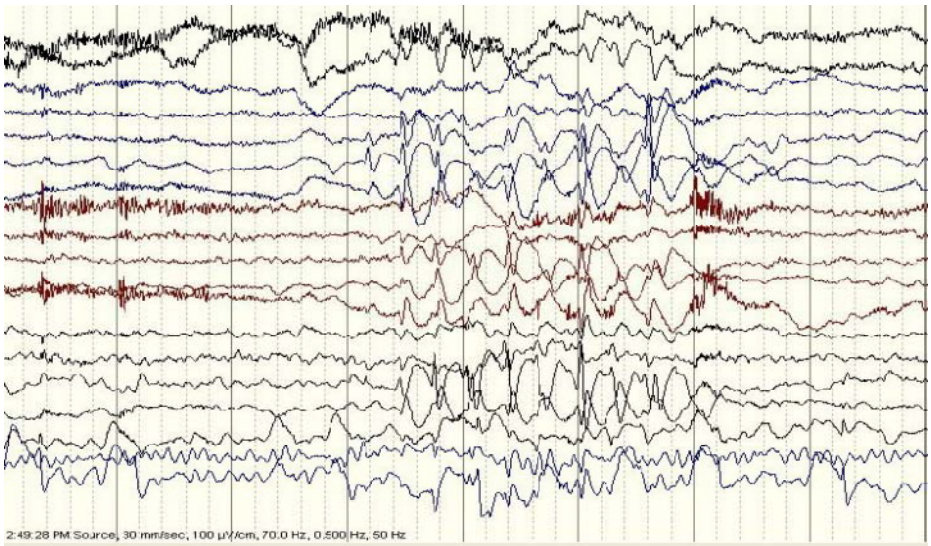

Fig. 1: EEG signal

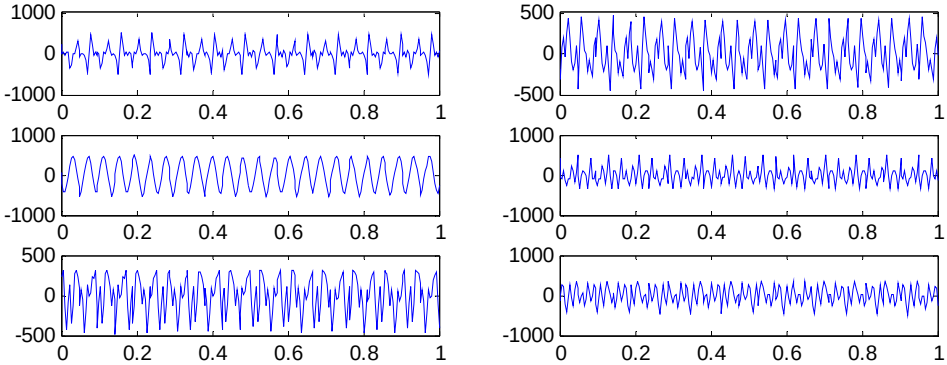

Fig. 2: Separation signals

Table 1 shows the comparison of convergence rates between Newton's iteration in fixed-point FastICA algorithm and iteration in improved algorithm that uses gradient descent method. The minimum, maximum and

average numbers of iterations were taken after running the algorithms ten times. From the table we can see that the number of iteration is decreased by improved algorithm and the coverage rate is improved.

Table 1: Comparison of convergence rates between Newton's iteration in fixed-point algorithm and iteration in improved algorithm

	Minimum number of iterations	Maximum number of iterations	Average number of iterations
Fixed- point algorithm	11	23	16
Improved algorithm	7	12	10

Conclusion:

In this paper, an improved FastICA was proposed and applied to separate EEG signal from an epileptic patient during seizure. Six source signals were obtained after using separation algorithm. It was found that the rate of convergence was improved by the new algorithm.

ACKNOWLEDGEMENT

The researchers would like to thank their family members for their continuous support and Ministry of Science, Technology and Innovation for granting the National Sciences Fellowship scholarship during his study. Authors also appreciate the financial support received from grant vot. 78397

REFERENCES

A. Hyv¨arinen and E. Oja, 1997. A fast fixed-point algorithm for independent component analysis. Neural Computation, 9(7): 1483-1492.

A. Hyv¨arinen, J. Karhunen and E. Oja, 2001. Independent Component Analysis. New York: John Wiley & Sons..

Devijver, P. and J. Kittler, 1982. Pattern Recognition: A Statistical Approach. Prentice Hall.

Delfosse, N. and P. Loubaton, 1995. Adaptive blind separation of independent sources: a deflation approach. Signal Processing, 45: 59-83.

James V. Stone, 1999. Independent component analysis a tutorial introduction, The MIT PressCambridge, Massachusetts. London, England.

Stone, J.V. 2004. Independent Component Analysis: A Tutorial Introduction (Bradford Books). The MIT Press, September.

T.W., 1958. Anderson. An Introduction to Multivariate Statistical Analysis.Wiley.