Blind source separation with dynamic source number using adaptive neural algorithm

Tsung-Ying Sun\textsuperscript{a,}\textsuperscript{*}, Chan-Cheng Liu\textsuperscript{a}, Shang-Jeng Tsai\textsuperscript{a}, Sheng-Ta Hsieh\textsuperscript{b}

\textsuperscript{a}Department of Electrical Engineering, National Dong Hwa University, 97401 Hualien, Taiwan, ROC
\textsuperscript{b}Department of Electrical Engineering, Oriental Institute of Technology, 220 Taipei County, Taiwan, ROC

Keywords:
- Adaptive learning rate
- Adaptive neural algorithm
- Blind source separation
- Dynamic number of sources

Abstract

A difficult blind source separation (BSS) issue dealing with an unknown and dynamic number of sources is tackled in this study. In the past, the majority of BSS algorithms familiarize themselves with situations where the numbers of sources are given, because the settings for the dimensions of the algorithm are dependent on this information. However, such an assumption could not be held in many advanced applications. Thus, this paper proposes the adaptive neural algorithm (ANA) which designs and associates several auto-adjust mechanisms to challenge these advanced BSS problems. The first implementation is the on-line estimator of source numbers improved from the cross-validation technique. The second is the adaptive structure neural network that combines feed-forward architecture and the self-organized criterion. The last is the learning rate adjustment in order to enhance efficiency of learning. The validity and performance of the proposed algorithm are demonstrated by computer simulations, and are compared to algorithms with state of the art. From the simulation results, these have been confirmed that the proposed ANA performed better separation than others in static BSS cases and is feasible for dynamic BSS cases.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

In general, the goal of blind source separation (BSS) is to distinguish source signals from mixed signals received by sensors. The adjective “blind” emphasizes the fact that, the source signals are not observed, and, no information is available about the mixing situation. But, the assumption is often held physically that the source signals are mutually independent. Recently, BSS in signal processing has received considerable attention from researchers, due to its numerous promising applications in the areas of biomedical signal processing, digital communications and speech signal, sonar, image processing, and monitoring (Cichocki & Unbehauen, 1996; Schiessl et al., 2000; Tonazzini, Bedini, & Salerno, 2006; Wei, Woo, & Dlay, 2007; Yilmaz & Rickard, 2004; Zhang & Kassam, 2004).

Since the pioneering work of Herault and Juten (1986), a variety of algorithms have been proposed for different BSS subjects. In general, the existing algorithms can be divided into several major categories according to criterion: independent component analysis (Cichocki, Karhunen, Kasprzak, & Vigario, 1999; Cichocki & Unbehauen, 1996; Herault & Juten, 1986; Liu, Sun, Lin, & Chou, 2006; Schiessl et al., 2000; Tonazzini et al., 2006; Wei et al., 2007; Ye, Zhu, & Zhang, 2004; Yilmaz & Rickard, 2004; Zhang & Kassam, 2004), density model-based algorithms (Amari, Chen, & Cichocki, 1997; Comon, 1994; Lee, Girolami, & Sejnowski, 1999; Vigliano & Uncknin, 2003), algebraic algorithms (Belouchrani, Abed-Meraim, Cardoso, & Moulines, 1997; Even & Moisan, 2005; Li & Wang, 2002; Lou & Zhang, 2003), information-theoretic algorithms (Pajunen, 1998; Pham, 2002; Pham & Vrins, 2005) and time-frequency (or sparseness) based algorithms (Araki, Sawada, Mukai, & Makino, 2007; Lee, Lewicki, Girolami, & Sejnowski, 1999; Liu, Sun, Li, Hsieh, & Tsai, 2007; Li & Zhang, 2006; Yilmaz & Rickard, 2004).

A common assumption, the numbers of sources are known a priori, has been applied by most of BSS algorithms in the past. However, this assumption could not be ensured in some advanced applications. The mechanics identifying the number of sources was developed upon time-frequency based algorithms, such as the one discussed in Liu et al. (2007). However, the method could not perform an instant identification because of its inherent shorts from huge computation. Neural network (NN) based algorithms are often utilized in operations of on-line separation, because of its ability to produce output signals almost instantly with its simple operations of gradient iteration. For taking into account the non-stationary environment, an adaptive algorithm is necessary when dealing with BSS problems with dynamic source numbers.

In the aspect of NN based algorithms, in order to have an exploitable dimension of neural network, the numbers of sources n is assumed a priori. Typically it should be equal to the number of...
sensors and outputs \( m \), termed the determined state. In this state, the natural gradient descent is derived and the property of stability has been theoretically demonstrated by Cichocki & Amari (2003). Under-determined neural algorithms, those that have fewer sensors than sources \( (m < n) \), have been proposed in Li and Wang (2002), Cao and Liu (1996). But when there is insufficient prior information related the mixing matrix or source, the source signals cannot be successfully recovered. The over-determined neural algorithm has more sensors than sources, i.e. \( m > n \). Learning divergence arises from the high correlations of similar or redundant signals from the extra sensors. To overcome the problem, several neural network architectures, together with the associated learning algorithms, were proposed in Cichocki et al. (1999). However, when a pre-whitening layer is applied during estimating the source number to reduce the dimension of the data vectors, separation results may be poor for ill-conditioned mixing matrices or weak source signals. Similar disadvantages exist if possible data compression takes place in the separation layer instead of the pre-whitening layer. A modified natural gradient algorithm using the adaptive scheme of self-normalized matrix has been suggested recently by Ye et al. (2004). Unfortunately, its separating performance is often incommensurate since the adjusting speed of its self-normalized matrix cannot always be applied to every case.

Supervised learning algorithms suffer from a large degree of system complexity because any arbitrary system has to be learned. A large amount of information is required as input to fine tune the system in order to avoid under or overestimation. This kind of prior knowledge is not always available; this restricts the successful application of fixed structure systems. In order to find the optimal system, a NN with self-organized structure was proposed (Tenorio & Manoel, 1990).

The concept behind the proposed method is a system capable of instantly identifying the number of sources, and adjusts the size of the NN accordingly to ensure optimal performance during separation. In 2006 we proposed a self-organized neural algorithm to solve the problem (Liu et al., 2006). According to its properties, the gradient descent value will increase to infinity while \( m > n \), the rule of self-organization will allow the dimension of network to grow gradually at a fixed period, until infinity occurs, and then trim the dimensions one by one until the gradient value steadies again. However, there is no best value when defining a growth period, for example, the algorithm will take a long time to become over-determined when a large value is given; otherwise, it will spend too much time trimming the redundant dimensions when a small value is given. Though the ability of this algorithm to handle a static BSS problem is verified, it could not solve a problem with dynamic source numbers.

In this paper, the focus is on advanced BSS problems which involve unknown and dynamic number of sources. Considering the researchers’ previous experiences in BSS study, an adaptive neural algorithm (ANA) is further proposed. This algorithm is based on the feed-forward NN proposed by Cichocki & Unbehauen (1996), and utilizes an improved cross-validation technique to instantly estimate the number of sources. The size of the NN will be updated accordingly in respect to the estimation. In order to allow greater efficiency when learning in the dynamic NN structure, an adaptive learning rate for the NN by reference to gradient information was proposed. The validity and efficiency of the proposed ANA is demonstrated by several BSS simulations. Ye’s algorithm (Ye et al., 2004) is implemented to perform comparisons on the separation performance. Note that the method proposed in Cichocki et al. (1999) is not able to tackle dynamic BSS problem because its number source estimation – which requires calculation of a large number of mixed signals – cannot be performed instantaneously. This method is not implemented in this paper because it is not appropriate for the function in these simulations.

The remainder of this paper is organized as follows: Section 2 presents the BSS problem formulation, the estimator of source numbers, and the cross-validation technique. Section 3 presents the on-line estimator, the self-organized criterion, and the adjustment of learning rate. The separation performances of the ANA and an existing algorithm are contrasted in several computer simulations, shown in Section 4. Section 5 contains a brief conclusion.

2. Problem formulation

2.1. The description of general BSS

In a situation where the sources are unobservable, \( s(t) = [s_1(t), \ldots, s_n(t)]^\top \) is the zero-mean vector and is mutually (spatially) statistically independent (or as independent as possible), where \( n \) denotes the number of sources and \( t = 1, \ldots, N \) is the instant time of sampling. The available sensor vector \( x(t) = [x_1(t), \ldots, x_m(t)]^\top \), where \( m \) is the number of sensors, is given by

\[
x(t) = As(t)
\]

(1)

where \( A \in \mathbb{R}^{m \times n} \) is a non-singular and unobservable matrix has non-zero determinant. The assumption \( m = n \) is held for convenience when explaining. The goal of BSS is to recover the waveforms of each source from individual output \( y(t) \) such that

\[
y(t) = W(t)x(t)
\]

(3)

Generally, weights of NN are updated by the gradient decent, and the objective function is a nonlinear correlation criterion which is derived from independent component analysis (ICA) Cao & Liu, 1996. Since source signals are zero-mean and mutually independent, the generalized covariance matrix is represented as

\[
R_y = E[yy^\top] - E[y]E[y^\top]
\]

(4)

where \( f(y) \) and \( g(y) \) are different odd nonlinear activation functions. The diagonal elements of \( R_y \) are non-zero and all other terms are zero. The terms \( E[f(y)]E[g^\top(y)] \) will be zero if the probability density functions (pdf) of each source are even. The choice of activation functions, \( f(y) \) and \( g(y) \), is based on the statistically distribution characteristic of source signals, refer to literature (Cao & Liu, 1996) for further detail.

The first updating equation is proposed by Hérald & Juten (1986), as the following:

\[
w_q(t) = w_q(t - 1) + \mu \cdot \Delta w_q(t)
\]

(5)

and

\[
\Delta w_q(t) = f(y_q(t))g(y(t))
\]

(6)

where \( w_q(t) \) are elements of the separating matrix \( W \), and \( \mu \) denotes a learning rate, a lower value is usually given. According to (6), Amari utilized feed-forward NN and proposed the robust updating equation for ill-condition cases, as the following:

\[
\Delta w_q(t) = [\lambda_q - f(y_q(t))g(y(t))]w_q(t)
\]

(7)

where \( \lambda = [\lambda_q] \) denotes the self-normalized matrix (typically \( \lambda = 1 \)).
2.2. Estimation of the number of sources

The cross-validation techniques have been primarily applied in multivariate statistics (Krzanowski, 1987). The basic idea of a cross-validation method is that it divides the data into several subgroups. One group is used to determine some features of the data, and the other groups are used to verify the features. Using this technique, a criterion for estimating the number of sources based on the error of estimated noise variance was proposed by Cao, Murate, Amari, Cichocki, & Takeda (2003).

Let \( C \) denote the covariance matrix given by \( C = xx^T \). The estimation can be obtained as

\[
\Psi = \text{diag}(C - \tilde{A}\tilde{A}^T)
\]

where \( \tilde{A} = U_n\Lambda_n^{1/2} \), and \( \Lambda_n \) is a diagonal matrix whose elements are the \( n \) largest eigenvalues of \( C \). The columns of \( U_n \) are the corresponding eigenvectors. The data matrix \( x \in \mathbb{R}^{m \times N} \) is then divided into several disjoint groups, such as \( x_i \in \mathbb{R}^{m \times N} \), where the group number \( i = 1, ..., K \). There will be a larger error between the noise variance diag(\( \Psi_i \)) and its estimated diag(\( \Psi_{\text{hat}} \)), where \( i \neq m \), making it obvious when the estimation of the source number \( n \) has not been matched to its true value. Based on this property, the criterion for each conjectured source number \( n \) was defined as

\[
\text{Error}(n) = \frac{1}{K} \sum_{i = 1}^{K} \text{tr}[^{\text{diag}}(\Psi_i) - \text{diag}(\Psi_{\text{hat}})^2]
\]

where \( \text{tr} \) is the trace of a matrix. It is necessary to compute all the possible estimates of the source number, the boundary of \( 1 \leq n \leq (1/2)(2m + 1 - \sqrt{8m + 1}) \) is suggested in Cao et al. (2003).

There are other available methods for approximating the number of sources, such as principal component analysis (PCA) or singular value decomposition (SVD). According to our experiences, the error ratio for the estimations and the cost of computation are slightly less evident with cross-validation than PCA or SVD. Consequently, it is the reason why this study bases its system on cross-validation to develop an instant estimator.

3. The adaptive neural BSS algorithm

The relationship between the unobservable outside situation and our algorithm is illustrated as Fig. 1. The source number estimator informs the adaptive structure NN what size of structure is best. The learning rate adjustment evaluates a proper value according to gradient, and returns it to the updating function of the learning criterion. In this study, ambient noises from the environment are considered, but the subtle noises from transmission are ignored.

3.1. Improvement of cross-validation

According to the cross-validation mentioned in the last section, the unknown number of sources can be calculated in batch. To satisfy the requirements of on-line learning, this paper used the recursive method to calculate the mean and the covariance matrix. In the beginning, a sampling window length \( wd \) is defined. While the number of sampling mixtures is less than \( wd \), the covariance matrix is recursively calculated as following (Lou & Zhang, 2003):

\[
M(t) = \frac{t-1}{t}M(t-1) + \frac{1}{t}x(t)
\]

\[
\Delta x(t) = M(t) - M(t-1)
\]

\[
C(t) = \frac{t-1}{t}[C(t-1) + \Delta x(t)\Delta x(t)^T] + \frac{1}{t}(\Delta x(t) - M(t)|\Delta x(t) - M(t)|)^T
\]

where \( M \) is the mean of \( x \). While \( t > wd \), past \( wd \) mixtures, \( x = \{x(t-(wd+1)), ..., x(t)\} \), used to generate the mean and the covariance matrix, therefore, early mixtures will not affect present estimation when the number of sources has changed.

At each iteration, the instant noise variance \( \Psi(i) \) is obtained from (8), where \( n = 1, ..., m \). Then, at the \( n \)th instant the estimated number of source can be obtained by

\[
n(i) = \min\{\arg\min_n\ n(i)\}
\]

where

\[
J(i) = \sum |\Psi(i) - \Psi(m)|^2
\]

where \( \Psi(m) \) is the last of instant noise variances.

3.2. Self-organized criterion

In the field of neural network research, how to define a proper structure size is an important issue. NN’s with a small structure size cannot perform complex mappings. A large sized NN has not only a high running cost but also unstable performance. The size of the structure is usually defined by trial and error in general applications; however, a self-organized structure is proposed for the dynamic problems such as those discussed in Tenorio and Manoel (1990), Morris and Garvin (1996), Park, Huh, Kim, Seo, and Park (2005). The method proposed in this study utilizes a fully connected feed-forward NN which includes an input layer and an output layer. The two layers contain the same number of nodes, \( n \). Then, a pair of nodes from both layers will be generated or removed at the time \( n \) changes.

The self-organized procedure at each iteration is to evaluate the values of \( n(i) \) at the beginning. When \( n(i) > n(i-1) \), which implies the number of source signals is increasing, then the size of the separating matrix will be enlarged by

\[
w_j(t) = \begin{cases} \frac{w_j(t-1)}{r}, & \text{as } i \leq n(t) \text{ and } j \leq n(t), \\ r, & \text{otherwise.} \end{cases}
\]
where \( r \) is a random value between \([-1,1]\), newly produced elements are added to the end of the matrix \( W(t-1) \). While \( \hat{n}(t) < n(t-1) \), which implies the number of source signals is decreasing, then the size of the separating matrix will be trimmed by
\[
W(t) \leftarrow \{w_{ij}(t-1) | i \leq \hat{n}(t) \text{ and } j \leq \hat{n}(t) \}
\]
(16)
The redundant elements of the separating matrix are removed from the end of the matrix. If \( \hat{n}(t) = n(t-1) \), then the size of separating matrix can no longer be tuned, i.e. \( W(t) = W(t-1) \).

3.3. Learning rate adjustment

Since NN’s structure size is tuned according to the varying number of sources, the weights will be retrained repeatedly. The speed and stability of NN learning are affected by the size of learning rate. Therefore, an adjustment to evaluate an adaptive learning ratio is generated from 50 independent runs of some algorithm for each source signal, we take on the advice of Cao and Liu (1996), and pick activation functions for (20):
\[
\begin{align*}
s_1(t) &= \sin(500t + 5 \cos(60t)) \\
s_2(t) &= \sin(800t) \\
s_3(t) &= \sin(450t) \sin(40t) \\
s_4(t) &= \sin(90t) \\
s_5(t) &= \text{sign}(\cos(2\pi155t)) \\
s_6(t) &= \text{Noise with uniformly distributed in } [-1,1]
\end{align*}
\]
(21)

The cross-talking error proposed by Amari et al. (1997) is utilized to evaluate the separation performance
\[
PI = \frac{1}{m} \sum_{j=1}^{m} \left( \frac{\sum_{q=1}^{n} |c_{pq}|}{\max_{k} |c_{mk}|} - 1 \right) + \frac{1}{n} \sum_{q=1}^{n} \left( \frac{\sum_{p=1}^{m} |c_{pq}|}{\max_{k} |c_{pk}|} - 1 \right)
\]
(22)

The involved parameters of Ye’s algorithm are given by referring to Ye et al. (2004). In ANA, there are eight sensors (i.e. \( m = 8 \)), the initial size of structure is \( 1 \) (i.e. \( n(0) = 1 \)), and the length of the cross-validation slide window is \( wd = 30 \). Since all of the sources in (21) are sub-Gaussian signals, we take on the advice of Cao and Liu (1996), and pick activation functions for (20):
\[
\begin{align*}
f(y_i(t)) &= \text{sign}(y_i(t)) \\
g(y_i(t)) &= \tanh(10y_i(t))
\end{align*}
\]
(23)

All simulation results will be presented as average values which are generated from 50 independent runs of some algorithm for each simulation.

4.1. BSS with static number of sources

This simulation includes two cases whose number of sources is constant. The first is a \( n = 3 \) case which has the first three source signals \( s_i(t) \) of (21), where \( i = 1,\ldots,3 \). Similarly, the second is a \( n = 5 \) case, and the index of source signals is \( i = 1,\ldots,5 \).

Presented in Figs. 3 and 6 are the results of both cases after 10,000 iterations, and the curves of estimation of source signals by the ANA with adaptive learning rates. The waveforms of sepa-
rated signals are presented in Figs. 4 and 7 which display the last 500 separated samples. Figs. 5 and 8 present four curves of the average PI which are obtained by Ye’s algorithm and the ANA with different learning rates. The average and variance of final PI values are listed in Table 1 for comparing.

4.2. BSS with dynamic number of sources

This is a dynamic BSS simulation which uses time-variant numbers of sources in the following equations:

\[(1) \ n = 4, 3, 6 \quad \text{Case: the numbers of sources decrease, then will increase:} \]
\[
\begin{align*}
&n = 4, \ \text{as} \ \mathit{IN} \leq 6000 \\
&3, \ \text{as} \ 6000 < \mathit{IN} \leq 12,000 \\
&6, \ \text{as} \ 12,000 < \mathit{IN} \leq 18,000
\end{align*}
\]

\[(2) \ n = 3, 6, 2 \quad \text{Case: the numbers of sources increase, then will decrease:} \]
\[
\begin{align*}
&n = 3, \ \text{as} \ \mathit{IN} \leq 6000 \\
&6, \ \text{as} \ 6000 < \mathit{IN} \leq 12,000 \\
&2, \ \text{as} \ 12,000 < \mathit{IN} \leq 18,000
\end{align*}
\]

where \( \mathit{IN} \) denotes the iteration number, and \( n \) source signals is the first \( n \) of (21). Both cases have three states of \( n \) which switch every 6000 iterations.

After the two simulation cases finish, the number of sources obtained by the ANA with \( \mu_{sd} \times \mu_2 \) are presented in Figs. 9 and 12, the waveforms of the last 200 separated signals of each state are presented in Figs. 10 and 13. Several windows have no results because those outputs do not exist. The four PI curves obtained from different methods are compared in Figs. 11 and 14. The average and variance of the final PI values are further presented in Table 1. It should be noted that the values of Ye’s algorithm in the table are all blank because it is always diverging when the number of sources is increasing.

5. Discussion

From the above simulation results, it can be confirmed at first that the on-line source number estimator is functional and has a short identifying time (under 20 iterations which is smaller than the value of \( \mathit{wd} \)). The value of \( \mathit{wd} \) is selected by experience. If \( \mathit{wd} \) is not big enough, the estimation of cross-validation cannot correctly evaluate the source numbers.

About the convergence of the PI average, Ye’s algorithm always converges at a high value because the self-normalized matrix can...
not find the perfect value during its evolution. For ANA, the PI curve decreases monotonously when a small learning rate \( l_1 \) is used. Inversely, the curve fluctuates during decrease and cannot steady when a big learning rate \( l_2 \) is used. However, when the adaptive learning rate \( \mu_{ad} \times l_2 \) is used, the PI speedily and stably converges at a lower value than others.

From Table 1, it can be observed that both algorithms have similar performance in variance, but the ANA has a far superior value on average.

Table 1

<table>
<thead>
<tr>
<th>Case</th>
<th>PI</th>
<th>ANA with ( \mu_{ad} \times l_2 )</th>
<th>Ye’s algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 3 )</td>
<td>Average</td>
<td>0.0476</td>
<td>0.6322</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.0014</td>
<td>0.0116</td>
</tr>
<tr>
<td>( n = 5 )</td>
<td>Average</td>
<td>0.0863</td>
<td>0.8639</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.0116</td>
<td>0.0119</td>
</tr>
<tr>
<td>( n = 4, 3, 6 )</td>
<td>Average</td>
<td>0.1794</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.0202</td>
<td>None</td>
</tr>
<tr>
<td>( n = 3, 6, 2 )</td>
<td>Average</td>
<td>0.0175</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>9.0621e-006</td>
<td>None</td>
</tr>
</tbody>
</table>

Fig. 8. The average performance indexes of Ye’s algorithm and the ANA with different learning rates (50 independent runs of the \( n = 5 \) case).

Fig. 9. The curve of estimated number of sources by the ANA in the \( n = 4, 3, 6 \) case.

Fig. 10. The waveforms of the separated signals by the ANA with \( \mu_{ad} \times l_2 \) in the \( n = 4, 3, 6 \) case (the windows of columns from left to right correspond to all outputs of neural network in the iteration intervals [5801–6000], [11,801–12,000] and [17,801–18,000]).

Fig. 11. The average performance indexes of Ye’s algorithm and the ANA with different learning rates (over 50 independent runs of the \( n = 4, 3, 6 \) case).

Fig. 12. The curve of estimated number of sources by the ANA in the \( n = 3, 6, 2 \) case.

References


8861

6. Conclusion

An adaptive neural algorithm to solve BSS problems with unknown and dynamic source numbers is proposed in this paper. Estimation of the source numbers are based on a system improved the actual value. Next, the self-organized criterion guides the estimation of the source numbers are based on a system improved the actual value. Next, the self-organized criterion guides the system to the determined state. The learning rate adjustment further enhances the speed and stability of NN convergence.

According to the simulation results, the ANA is able to conquer the supplied BSS problems with a separation performance superior to that of Ye’s algorithm. What’s more, the proposed adaptive learning rate benefits from fast and stable training of NN when using fixed learning rates.

Fig. 13. The waveforms of the separated signals by the ANA with $\mu > \mu$ in the $n = 3, 6, 2$ case (the windows of columns from left to right correspond to all outputs of neural network in the iteration intervals [5801–6000], [11,801–12,000] and [17,801–18,000]).

Fig. 14. The average performance indexes of Ye’s algorithm and the ANA with different learning rates (over 50 independent runs of the $n = 3, 6, 2$ case).

References


