A novel blind source separation (BSS) algorithm based on the combination of negentropy and signal noise ratio (SNR) is presented to solve the deficiency of the traditional independent component analysis (ICA) algorithm after the introduction of the principle and algorithm of ICA. The main formulas in the novel algorithm are elaborated and the idiographic steps of the algorithm are given. Then the computer simulation is used to test the performance of this algorithm. Both the traditional FastICA algorithm and the novel ICA algorithm are applied to separate mixed signal data. Experiment results show the novel method has a better performance in separating signals than the traditional FastICA algorithm based on negentropy. The novel algorithm could estimate the source signals from the mixed signals more precisely.

Keywords: blind source separation; independent component analysis; negentropy; signal noise ratio

1 Introduction

The independent component analysis (ICA) is a signal processing method developed in the 1990s. This method is capable of separating the signals which are mixed from some independent signals by the difference of the statistical characterization factors with the criterion of the signals’ statistical independence, when the source signals and other transmit parameters are unknown [1]. It is widely used in signal processing [2], image processing [3], data compression [4], characteristic identification [5] and many other domains.

Besides the application research in those domains, many researchers are studying idiographic algorithms. An ICA algorithm based on an adaptive artificial nerve network is proposed in Ref. [6], which avoids the morbidity matrix problem and reduces the complexity of studying a matrix and estimating independent components. In Ref. [7], kurtosis is taken as the measurement of negative gauss and the changing rules of mixed signals’ negative gauss are analyzed when the mixing degree of signals changes. Based on these analyses, a fast searching algorithm of ICA is designed. In Ref. [8], the similarity of stochastic variables is defined from the concept of stochastic variables’ differential coefficient entropy. The method of seeking similarity’s culmination is discussed to realize the maximizing of the negative gauss of linearity combination of data, which can estimate the components in the ICA model. Bach proposed the concept of kernel independent component analysis in Ref. [9], in which signals data vectors are projected into a reproducing kernel Hilbert space and mixed data are separated by canonical correlation analysis combined with mutual information theory. This thesis introduces the basic principal and relative algorithm, and then combines the negative entropy with signal noise ratio to improve the performance of ICA method.

2 Basic principal of ICA

The ICA is a method to find statistically independent
components from multidimensional data and estimate the original signals from the mixed signals \[^{[10]}\] , the common model is

\[ X(t) = AS(t) \quad (1) \]

in which \( X(t) = [x_1(t), x_2(t), \ldots, x_N(t)]^T \) denotes the mixed signals; \( S(t) = [s_1(t), s_2(t), \ldots, s_N(t)]^T \) denotes the source signals; and \( A \) denotes the mixing matrix with the rank of \( M \times N \) and \( N \leq M \).

The purpose of ICA is to find a matrix \( W \) by data matrix \( X \) when source signals \( S \) and mixing matrix \( A \) are unknown. The product of \( W \) and \( X \) should be the estimate of source signals \( S \), that is

\[ Y(t) = WX(t) = WAS(t) \approx S(t), \quad (2) \]

in which \( Y(t) = [y_1(t), y_2(t), \ldots, y_N(t)]^T \) is the estimate of source signals. When \( M = N \), \( W = A^{-1} \) and \( Y(t) = S(t) \) in the perfect situation. But in common situations, \( S(t) \) and matrix \( A \) are unknown, so we make some hypotheses and restrict conditions to acquire matrix \( W \), and get the output \( Y(t) \) approach the source signals \( S(t) \).

### 3 Algorithm of ICA

#### 3.1 Measurement function of independence

A most important step in ICA is to construct or select the function to measure the independence of signals. It could be seen in (2) that every element of \( Y \) is the linear combination of source signals \( S \), which are independent to each other. According to the central limit principal, the distribution of the combination of several independent stochastic variables tends towards the gauss distribution. So we can use the character of negative gauss to measure the independence between the output signals \( Y \), and the problem of ICA turns out to be the maximum of the character of negative gauss of \( Y \).

The main functions used to measure the character of negative gauss are kurtosis and negative entropy. But the value of kurtosis may be influenced by a few observation values that may be wrong or outlying, which means the robustness of this kurtosis is not good. So the negative entropy is used in many situations.

Entropy is a basic concept in information domain and a measurement of stochastic variables’ uncertainness. The calculation formula is

\[ H(y) = - \int p(y) \log p(y) \, dy. \quad (3) \]

The more independent several stochastic variables are, the larger the entropy of these variables is. It is pointed out in information domain that the entropy of gauss variables is the largest of all. So the negative entropy is defined to measure the character of negative gauss of variables:

\[ J(y) = H(y_{\text{gauss}}) - H(y), \quad (5) \]

where \( y_{\text{gauss}} \) is a gauss variable; so the value of \( J(y) \) will be plus unless \( y \) is also a gauss variable.

#### 3.2 FastICA algorithm

Aapo et al. advanced some approximate calculations of negative entropy and advanced the FastICA algorithm based on negative entropy in Ref. [10]. The steps of this algorithm are listed below.

1) Center the data, making the mean of the data zero;
2) Whiten the data obtaining \( x \);
3) Let \( W = (w_1, w_2, \ldots, w_N)^T \), and initialize the vectors and pick out one of them as \( w_i \). If this is not the first time to pick out a vector, \( w_i \) should subtract the projection of other picked out vector by

\[ w_i = w_i - \sum_{j=1}^{i-1} (w_j^T w_i)w_j. \quad (5) \]

4) Update \( w_i \) by

\[ w_i \rightarrow E \left( xg \left( w_i^T x \right) \right) - E \left( g^\prime \left( w_i^T x \right) \right) w_i, \quad (6) \]

where function \( g \) could be

\[ g_1(y) = ye^{-\frac{y^2}{2}}, \quad (7) \]

\[ g_2(y) = y^3. \quad (8) \]

5) Set a threshold \( \varepsilon \) and if \( |w_i^T w_i| - 1 > \varepsilon \), come to step 4; otherwise estimate a component by \( y_i^* = w_i^T X \).
and continue to calculate other vectors \( w \) until all the independent components are established.

4 Improved BSS algorithm

Although the FastICA algorithm can constrain the result fast, the algorithm has a low degree of separation precision and the results are stochastic when calculated every time. So the negative entropy is used to multiply the SNR and taken as a new independence measurement function to improve the BSS algorithm. This combination not only demands the separation of mixed signals but also minimizes the noise in signals, which means the precision of separation. The SNR calculation formula is

\[
\text{SNR} = 10 \log \frac{\sum s_i^2}{\sum (s_i - y_i)^2}.
\]

(9)

A BSS algorithm based on maximization the SNR has been advanced in Ref. [12] where the SNR calculation formula is

\[
\text{SNR} = 10 \log \frac{y^T y'}{(y' - y)(y' - y)'},
\]

(10)

where \( y'(t) = \frac{1}{p} \sum_{j=0}^{p-1} y(t-i) \), \( p < 100 \). It can be seen that the estimated signal \( y \) is used to substitute the source signals to simplify the calculation in the numerator part. But the performance of the algorithm will be bad when separating two sub-gauss signals, even sometimes this method will not be effective [13]. So \( y' \) the slippage means of estimated signals should be used to substitute the source signals.

At the same time, \( y \) is a column vector with many signal elements based on the assumption before. So \( yy^T \) and \( (y' - y)(y' - y)' \) are both matrixes of rank \( N \times N \). The calculation result of matrix division is still a matrix, so the result of (10) is not the SNR obviously and the calculation process in Ref. [12] is wrong. The right calculation process is reckoned in this thesis on the basis of careful analysis on the fault in Ref. [12]. Based on the assumption we have \( y^T = w^T X \), which could be predigested by removing the subscript to \( y^T = w^T X \).

\[
\text{SNR} = 10 \log \frac{w^T X' X' w}{w^T (X' - X)(X' - X)^T w} = 10 \log \frac{U}{V} = 10(\log U - \log V),
\]

(11)

where \( U = w^T X' X' w \), and \( V = w^T (X' - X)(X' - X)^T w \).

So the SNR is a function with the variables of \( w \) and \( X \), and the vector gradient of this function is

\[
\frac{\partial F(W, x)}{\partial W} = \frac{\partial U}{\partial W} - \frac{1}{V} \frac{\partial V}{\partial W},
\]

(12)

where

\[
\frac{\partial U}{\partial w} = \frac{\partial w^T X' X' w}{\partial w}, \quad \frac{\partial V}{\partial w} = \frac{\partial w^T (X' - X)(X' - X)^T w}{\partial w}.
\]

The definition of vector gradient is

\[
\frac{\partial g}{\partial w} = \left[ \frac{\partial g}{\partial w_1}, \frac{\partial g}{\partial w_2}, \ldots, \frac{\partial g}{\partial w_M} \right]^T.
\]

(14)

According to the definition of vector gradient, the following formula could be obtained:

\[
\frac{\partial U}{\partial w} = \frac{\partial w^T X' X' w}{\partial w} = \frac{\partial \left( \sum_{i=1}^{M} \sum_{j=1}^{M} w_{ij} x'_i x'_j \right)}{\partial w} = 2 \left[ \sum_{j=1}^{M} w_{j1} x'_j, \sum_{j=1}^{M} w_{j2} x'_j, \ldots, \sum_{j=1}^{M} w_{jM} x'_j \right]^T.
\]

(15)
It could be seen that the formula is a product of some matrixes and it could be decomposed:

\[ \frac{\partial U}{\partial w} = 2 \begin{bmatrix} x_1^{T} & x_2^{T} & \cdots & x_M^{T} \\ x_1^{T} & x_2^{T} & \cdots & x_M^{T} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{T} & x_2^{T} & \cdots & x_M^{T} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{bmatrix} = 2 \begin{bmatrix} x_1^{T} \\ \vdots \\ x_M^{T} \end{bmatrix} \begin{bmatrix} x_1^{T} & x_2^{T} & \cdots & x_M^{T} \end{bmatrix} w. \] (16)

The following formula could be obtained in the same way.

\[ \frac{\partial V}{\partial w} = \frac{\partial w^T (X' - X)(X' - X)^{T}}{\partial w} = 2(X' - X)(X' - X)^{T} w. \] (17)

Bring them back to formula (12), and we can obtain

\[ \frac{\partial F(w, X)}{\partial w} = 20 \left( \frac{1}{U} \frac{\partial U}{\partial w} - \frac{1}{V} \frac{\partial V}{\partial w} \right) = 20 \left( \frac{X'X'^{T}w}{w^{T}X'X'^{T}w} - \frac{(X' - X)(X' - X)^{T}w}{w^{T}(X' - X)(X' - X)^{T}w} \right). \] (18)

The vector gradient of SNR will be used in the following advanced BSS algorithm. Aapo et al. give an approximate calculation of negative entropy in Ref. [10]:

\[ J(y) \approx \left[ E[G(y)] - E[G(\nu)] \right]^{2}. \] (19)

So a new contrast function is built:

\[ \Psi(w) = \left[ E[G(y)] - E[G(\nu)] \right]^{2} E[SNR], \] (20)

where \( E[\cdot] \) is expectation operation, \( G(\cdot) \) is any non-quadratic function, and symbol \( \nu \) is a gauss variable. By maximizing \( \Psi(w) \), the needed signals can be separated. Calculate the vector gradient of \( \Psi(w) \):

\[ \frac{\partial \Psi(w)}{\partial w} = 2 \frac{\partial E[G(y)] - E[G(\nu)]}{\partial w} E[10\lg U - 10\lg V^{'},. \] (21)

The following formula can be obtained by the calculation characteristic of vector gradient and the formulas deduced above.

\[ \frac{\partial \Psi(w)}{\partial w} = 20\gamma E\left( Xg(w^{T}X) \right) E[\lg U - \lg V^{'}, \right. \] (22)

where \( \gamma = E[G(y)] - E[G(\nu)] \). We can simplify this formula in some aspects. The variable \( \gamma \) does not change the steady point of algorithm while its sign does, so we used its sign function \( \text{sign}(\gamma) \) to take its place. The constant 20 can be ignored because it does not affect the characteristic of algorithm. So we get an advanced gradient algorithm estimating independent components:

1) Center the data, making the mean of the data zero;
2) Whiten the data and obtain \( X \);
3) Let \( W = \{w_1, w_2, \ldots, w_N\}^{T} \), initialize the vectors and pick out one of them as \( w_1 \). If this is not the first time to pick out a vector, \( w_i \) should subtract the projection of other picked out vector by

\[ w_i = w_i - \sum_{j=1}^{i-1} (w_j w_j^{T}) w_j. \] (23)

4) Update the \( w_i \) by

\[ \Delta w_i \propto \text{sign}(\gamma_i) \left[ E \left( Xg(w_i^{T}X) \right) E[\lg U - \lg V^{'}, \right. \] (24)

\[ w_i \leftarrow w_i + \Delta w_i. \] (25)

5) Update \( \gamma_i \) by the new \( w_i \):

\[ \gamma_i \leftarrow E[G(w_i^{T}X)] - E[G(\nu)]. \] (26)

6) Set a threshold \( \varepsilon \) and if \( |w_i^{T}w_j| - 1 > \varepsilon \), come to step 4, otherwise estimate a independent component by
\[ y_i^T = w_i^T X \]
and continue to calculate other vectors \( w \) until all the independent components are established.

5 Simulations and results

5.1 Separation of mixed signals with Gauss signal

The FastICA algorithm and ICA algorithm advanced in this thesis are used to separate 4 simulated signals that are different in distribution. Fig. 1 is the graph of the 4 signals, which are sine wave signal, pulse signal, saw tooth signal and gauss signal. Fig. 2 is the graph of the signals mixed by a random matrix \( A \). Fig. 3 is the graph of the estimated signals by the FastICA algorithm based on negative entropy. Fig. 4 is the graph of the estimated signals by the algorithm built in this work.

5.2 Results and discussion

It could be seen from the figures that two methods both separate the probable source signals, but the order and sign change too, which is the problem ICA could not solve right now. By contrast we could see the signals separated by algorithm in this work are smoother than the signals separated by FastICA, which means there is much noise in the FastICA result.

We use mean square error (MSE), signal-to-noise ratio (SNR) and performance index (PI) to compare the two algorithms. The definition of MSE, SNR and PI are listed below:

\[
MSE = \frac{1}{N} \sum_{i=1}^{N} (s_i - y_i)^2 ,
\]
\[
SNR = 10\log \frac{\sum_{i=1}^{N} s_i^2}{\sum_{i=1}^{N} (s_i - y_i)^2};
\]

\[
PI = \frac{1}{m^2} \left[ \sum_{i=1}^{m} \left( \sum_{j=1}^{m} \frac{|p_{ij}|}{\max p_{ik}} - 1 \right) + \sum_{j=1}^{m} \left( \sum_{i=1}^{m} \frac{|p_{ij}|}{\max p_{jk}} - 1 \right) \right],
\]

where \( p_{ij} \) is the element in position \((i, j)\) of the matrix \( P = WA \). It is obvious that the separating performance of the algorithm is better when \( PI \) is smaller [14].

Taking the sine wave signal, pulse signal and saw tooth signal as needed signals, the MSE, SNR and PI of the FastICA algorithm's result are calculated, which turn out to be MSE=0.0290, SNR=35.567 9 dB, and PI=3.448; and for the algorithm in this work, MSE=0.001 6, SNR=64.258 0 dB, and PI=3.203. It is shown  that the algorithm in this work is better than the FastICA and can separate the mixed signals more exactly.

6 Conclusions

The negative entropy and SNR are combined to improve the ICA algorithm on the basis of introducing the basic principle and algorithm of ICA. The new ICA algorithm has better performance when separating mixed signals in computer simulation and the source signals are estimated precisely. But the speed of new algorithm is slower than the FastICA because it's more complex than FastICA Research is needed on simplifying the variables to improve the calculation speed.

References


冯燕,何明一,宋江红.基于独立成分分析的高光谱图像数据降维及压缩[J].电子与信息学报[ISSN 1009-5896],2009,27(12):2871-2875.


张旭秀,邱天爽.基于随机变量相似度的ICA 方法[J].电子与信息学报[ISSN 1009-5896],2009,31(9):2104-2108.


Edited by LUO Min