Blind Source Separation Based on Multiple Algorithm Fusion Using FSS-kernel and FastICA

Pengju He, Xingxing Chou, Xiaomeng Chen and Huazhi Zeng
Department of Automation, Northwestern Polytechnical University,
Xi'an 710129,China,
e-mail: hepengju@nwpu.edu.cn,xingxing212@163.com

***Abstract* —This paper presents a multiple algorithm fusion using FSS-kernel and FastICA to improve the accuracy of blind source separation. FastICA algorithm has a faster and more robust convergence speed than the traditional ICA algorithm. But its recovery results are not satisfied. Inspired by various successful applications of kernel and spectral clustering methods in machine learning and data mining community, probability density function of the source signal is estimated by the FSS-kernel algorithm, and then to restore the blind separation of mixed signals, FastICA algorithm is used, the negative entropy is the objective function. The simulation results show that the signal aliasing could be separated effecti- vely by this method. It is proved that the method has higher separation accuracy and adaptive capacity, by contrasting with the traditional ICA algorithms and FastICA algorithm.**

***Keywords*—FastICA algorithm, FSS-kernel algorithm, blind source separation, algorithm fusion.**

I. INTRODUCTION

Blind source separation is to determine a transformation from the observed data of mixing signal to restore the original signal or source [1]. Blind source separation is the most commonly used assumptions in the statistical independence of the source signal, when the components are Independent of each other, become Independent Component Analysis (Independent Component Analysis, the ICA). Traditional ICA algorithm using gradient method [2] usually to optimization of objective function, but the gradient method, slow convergence speed, easily trapped in local minima point. The independent component analysis (ICA) is performed on the embedding matrix to decompose the single channel recording into its underlying independent components; At last subjective methods are then used to identify component of interest that are then projected back onto the measurement space in isolation. FastICA batch calculation method [3], improves the processing speed, but involves the selection problem of nonlinear function, so the separation accuracy is not ideal.

To solve above problems, this paper calculates smart method as a optimization algorithm of blind source separation [4], [5], namely a limited support sample kernel function (FSS - kernel) with FastICA fusion of blind source separation algorithm. This method not only inherited the FastICA algorithm is rapid and steady convergence speed advantage, and it has higher separation accuracy[6]-[9], improve the separation performance of BSS algorithm [14], [15].

II. FASTICA AND FSS-KERNEL

Separating mixing signal of single channel need to use more potential information that may be statistics quantity, exact quantity, system quantity and signal quantity, because usually input, system and output need to know any two of quantities to estimate the third quantity, so the key is how to excavate and utilize the a priori knowledge of the problem. For a specific problem some assumptions are required. In order to attempt to get a general method, some categories of single channel source separation methods are surveyed.

A. Basic theory of BSS

Assume observed signal from m sensors is x, where $x=[x_1, x_2, \cdots x_m]^T$. Mixed signal is s, which is composed with n independent unknown signals, $s=[s_1, s_2, \cdots s_n]^T$. So linear instant mixture BSS problem can be described with the equation:

$$x = As + n \tag{1}$$

A is mixture matrix, and $A \in R^{m*n}$, $n=[n_1, n_2, \cdots n_m]^T$, and n is composed with m noisy signal. Generally, noisy signal can be one of the original signals or be ignored. Equation (1) can be simplified as follows:

$$x = A \times s \tag{2}$$

The goal of BSS is find such a separation matrix W that makes $y = Wx$ approach to the original signal s. The model can be described as Fig.1[10]:

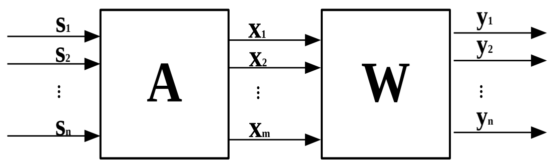

Fig. 1. Model of BSS

B. FastICA algorithm

The computing process of FastICA uses a batch processing method, while the derivation of computing way uses self-adapted processing method. So we can regard the FastICA as the combination of batch processing and self-adaption. Thus this method has a high processing method than batch processing or some self-adapted algorithm. FastICA algorithm doesn't separate all components one time, it extracts component one by one. That is, FastICA extracts one component and wipe out it, then it pick up component from remaining data. We can adopt Gram-Schmidt to achieve the above goal.

In order to extract component, we need determine a objective function. And the rule of determination is "the distance between probability density of separated data and Gaussian distribution is farthest” [16]. In theory, negentropy is a variable used measure the degree of non-gaussian for any probability density. Our algorithm will make the negentropy as the objective function. Theoretically, negentropy of Gaussian is the biggest. Negentropy J of random variable y can be defined as:

$$J(y) = H(y_{guass}) - H(y) \tag{3}$$

Where y_{guass} and y are Gaussian random variables with the same covariance, $H(.)$ is differential quotient of y’s probability density $p(y)$. It is defined as:

$$H(y) = -\int p(y)\lg p(y)dy \tag{4}$$

According to the definition, because the probability density of random variables is not easy to estimate, we often use the approximation:

$$J(W) = E\left\{W^T z\right\} - E\left\{G(v)\right\} \tag{5}$$

W is a n-dimensional vector which satisfies $\|W\| = 1$, Z is a orthogonal matrix whose covariance $E\left\{zz^T\right\} = I$,V is a Gaussian random variable whose mean value is 0 and variance is 1.G is not a quadratic function, we often choose $G(u) = \tanh(u)$, $G(u) = ue^{-u^2/2}$ or $G(u) = u^3$. Based on the rule of “separate one by one”[17], the ICA of n original signal becomes the solution of n optimization problems:

$$W_i = \arg\min \sum_{i=1}^{n} J_G(W_i) \qquad \text{i=1,2,3….n} \tag{6}$$

The constrained condition is:

$$E\left\{(W_k^T X)(W_j^T X)\right\} = \delta_{jk} \tag{7}$$

We can solve the above optimization problem with the following fixed-point algorithm:

$$W_p = W_p - \sum_{j=1}^{p-1} W_p^T W_j W_j \quad , \quad W_p \leftarrow \frac{W_p}{\|W_p\|} \tag{8}$$

We need use the iterative calculation to compute the equation (8) until W_p converges. Practices have proved the above iterative calculation cannot converge very well. In order to improve convergence performance, we use fast iteration algorithm. It is easy to understand, the biggest value of $J(W)$ can be achieved from a optimal value of [15] $E\left\{W^T x\right\}$. We can learn from Kuhn-Tucker condition and $\|W\| = 1$, the optimal value can be obtained from the following equation:

$$f(W) = E\left\{zG(W^T z)\right\} + \beta W = 0 \tag{9}$$

where β is a constant, $g(.)$ is first-order derivation of G . In order to get the solution of the above equation, we use Newton iterative equation:

$$W_{n+1} = W_n - \frac{f(W_n)}{f'(W_n)} \tag{10}$$

where

$$f'(W_n) = \frac{\partial f}{\partial W_n} = E\left\{zz^T g(W_n^T z)\right\} + \beta \tag{11}$$

then we can get

$$W_{n+1} = W_n - \frac{E\left\{zG(W_n^T z)\right\} + \beta W_n}{E\left\{zz^T g(W_n^T z) + \beta\right\}} \tag{12}$$

Because z are albino data, there is $E\left\{zz^T\right\} = I$. Then iterative equation can be reduced as:

$$W_{n+1} = W_n - \frac{E\left\{zG(W_n^T z)\right\} + \beta W_n}{E\left\{g(W_n^T z)\right\} + \beta} \tag{13}$$

After simplification, we can get fast iterative algorithm:

$$W_{n+1} = E\left\{g(W_n^T z)\right\}W_n - E\left\{zG(W_n^T z)\right\}$$
$$W_{n+1} \leftarrow \frac{W_{n+1}}{\|W_{n+1}\|} \tag{14}$$

C. FSS-kernel algorithm

According to equation (4) and equation (6), in order to separate original signals from mixed signal, we need know the accurate probability distribution function $G(y)$ .Traditional BSS method use some specified non-linear function as $G(y)$, however, the non-linear function is difficult satisfying all original signals’ statistical property and recovered original signals are not desired. Sometimes, traditional BSS even cannot realize the separation[11].

In order to solve the problem, based on Kernel function method, from observed signals, every window function in this text use M finite supported samples in N observed samples. We use the received probability density function as the rule of minimum mutual information to realize blind separation of mixed signals.

Firstly, we rank samples in non-decreasing sequence and we get $\left\{y_1, y_2, \cdots y_N\right\}$, probability density function of finite supported samples kernel function can be estimated as:

$$\hat{G}(y_i) = \frac{1}{N-M} \sum_{i=1}^{N-M} \phi(y_j - \hat{y}_i) \tag{15}$$

where $\hat{y}_i = (y_{i+M} + y_i)/2$ [18], it is the center of M finite supported samples’ kernel function, M is the width of window.

This text adopt “leave-one-out” to determine the best M is $\sqrt{N}$. The choose of Kernel function $\phi(y)$ should satisfy finite supported samples boundary condition and the basic condition of probability density function.

When $y \le y_i$ or $y \ge y_{i+M}$, $\phi(y) = 0$;The derivation of $\phi(y)$ in y_i and y_{i+M} is 0;The integration of $\phi(y)$ in the domain $y_i \le y \le y_{i+M}$ equal to 1[19];Kernel function is continuous differentiable functions satisfied the above conditions which has the least order is :

$$\phi(y)=\begin{cases} C_i (y-\mu_i)^2 (y+\mu_i)^2 & y\in[-\mu_i,\mu_i] \\ 0 & \end{cases} \quad (16)$$

Where $\mu_i=(y_{i+M}-y_i)/2$. According to condition (3), we can get $C_i=15\mu_i^{-5}/16$.With $\hat{y}_i$, μ_i, C_i, equation (15) and equation (16), we can obtain the derivation for the probability density $p(y)$ of random variable y:

$$H(y)\approx -\lg\sum_{i=1}^{N-M}\sum_{j=i+1}^{i+M-1}\frac{30(y_j-y_{i+M})^2(y_j-y_i)^2}{N(N-m)(y_{i+M}-y_i)^5} \quad (17)$$

D. The principle and steps of fusion algorithm

The blind source separation method fused with FSS-Kernel and FastICA algorithm, starts from the observation samples, then estimate the probability density function of sources to get in line with the source signal statistical characteristics of nonlinear function, then applies with negative entropy to be the objective function, then combine with the advantages of batch computing FastICA algorithm, thus the realization of fast and steady mixed signal blind source separation.

The process of blind source separation method fused with FSS-Kernel and FastICA algorithm, as follows:

Step1: make the observation data centralized, and make its mean value to zero;

Step2: Whiten the data, remove the correlation of data, $X\rightarrow Z$;

Step3: Choose to estimate the number of components, set the number of iterations, $P\leftarrow 1$;

Step4: Choose an initial weight vector (random) W_p;

Step5: Use FSS - kernel algorithm to estimate the probability density function and calculate the nonlinear function;

Step6: Let $W_p=E\{g'(W_p^TZ)\}W-E\{Zg(W_p^TZ)\}$;

Step7:Let $W_p=W_p-\sum_{j=1}^{p-1}(W_p^TW_j)W_j$;

Step8: Let $W_p=W_p/\|W_p\|$, if W_p convergence, return to step 6;

Step9: If $p\le m$, let $p=p+1$, return to step 4;

III. Simulation And Performance

In order to verify the feasibility and effectiveness of algorithm ,the experiment applies with two Sub_Gaussian signals, a Gaussian white noise signal and two Super Gaussian signal as the unknown source signal, as follows:

$$s_1=5sign(\cos(310\pi t/f_s))$$

$$s_2=5\sin(180\pi t/f_s)$$

$$s_3=25(t/f_s)^2$$

$$s_4=5\sin(18\pi t/f_s)\sin(600\pi t/f_s)$$

$$s_5=1-2rand(1,4000)$$

In the formula, F_s=10000 is Sampling frequency, and the number of signal sample points is 4000. Source signal time domain waveform is shown in Fig.2.

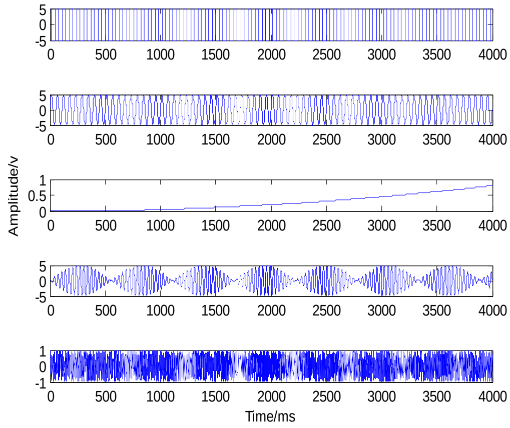

Fig. 2. The original signal

The randomly generated matrix A mixes with the original signal to be a mixed signal, whose waveform is shown in Fig. 3.

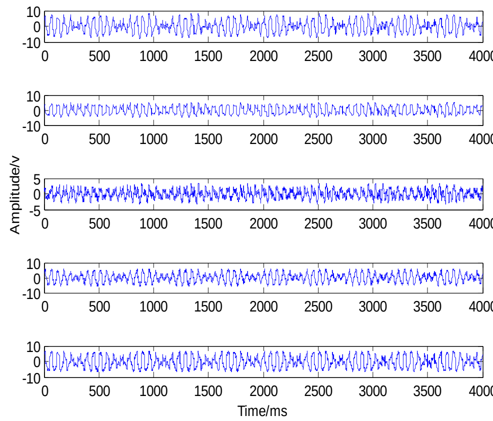

Fig. 3. The mixed-signal

$$A=\begin{bmatrix} 0.8147 & 0.0975 & 0.1576 & 0.1419 & 0.6557 \\ 0.9058 & 0.2785 & 0.9706 & 0.4218 & 0.0357 \\ 0.1270 & 0.5469 & 0.9572 & 0.9157 & 0.8491 \\ 0.9134 & 0.9575 & 0.4854 & 0.7922 & 0.9340 \\ 0.6324 & 0.9649 & 0.8003 & 0.9595 & 0.6787 \end{bmatrix}$$

With the blind source separation algorithm in the paper, the mixed-signal can be separated to recover the source signal, shown in Fig. 4.

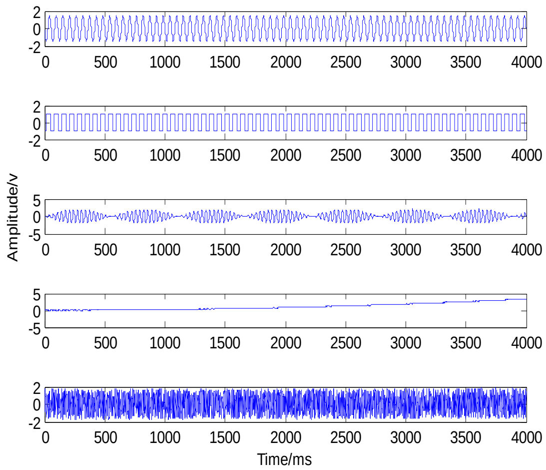

Fig. 4. The separated signal

We can see visually from figure 4, the proposed method can achieve a good separation of ultra-blind, Gaussian, sub-Gaussian mixed signal separation and has a better effect comparing other method[20]. In order to study quantitatively the separation performance of the proposed method, this algorithm is compared with the best traditional ICA algorithms and FastICA algorithm. And different sample points are employed into simulation. The signal to interference ratio(SIR) is also introduced to evaluate the separation performance. The formula as follows:

$$U_{SIR} = 10\lg \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} s_i^2(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{s}_i(j) - s_i(j))^2} \qquad (18)$$

In the formula, S represents the source signal, $\hat{S}$ represents the separated source estimated signals, the unit is dB. It can be seen from the definition that the greater U_{SIR} is, the more obvious separation effect is indicated. Normally, When U_{SIR} is less than 10 ~ 12 dB, it is described mixed-signal failed to achieve effective separation. We take 30 times simulations, then mean these results, we can see the comparison of three kinds of algorithm lastly.

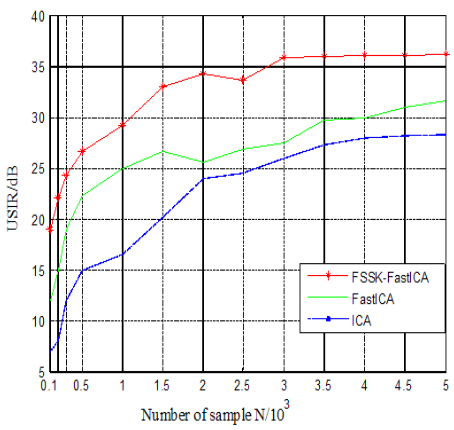

Fig. 5. The comparison of different algorithms performance

From Fig. 5, it is clearly shown that when the sample point is not many, the separation performance is not satisfactory. While the sample point is over 5000, these algorithm can realize the separation of the mixed-signal, and with the increase of sample points, the performance grows gradually. Among these methods, the traditional ICA algorithm performs worst, and the effect of the FastICA algorithm displays better, when the sample points increase. a blind source separation method fused with FSS-Kerneland FastICA algorithm receive the best separation performance.

Ⅳ. CONCLUSION

This paper proposes a blind source separation method fused with the limited support sample kernel function and FastICA algorithm. In the method, Maximum negative entropy is applied to be the search direction , And the fast iterative optimization algorithm also is employed.

Therefore it is not only inherits the advantages of FastICA algorithm which is fast and has robust convergence speed, but also can be estimated from the measured signal source signal probability density function without selecting a nonlinear function and applicable to all types of mixed-signal separation. Simulation shows that this method combines the advantages of FSS-Kernel and FastICA algorithm， which has higher separation accuracy and fast convergence. Considering these advantages, this method has better application in speech signal processing, data mining, image recognition, medical signal processing and radar and communication signals processing and other aspects of the project has better applicability.

REFERENCES

[1]Hopgood J R,Rayner P J W ,“Single channel separation using linear time varying filters: separability of non-stationary stochastic signals,” IEEE International Conference, Vo 3,pp 1449 – 1452, 1999

[2]Pau Bofill , Michael Zibulevsky, “Underdetermined blind source separation using sparse representations,” Signal Processing, vol. 81, no. 11, pp. 2353–2362, 2001.

[3]Hopgood J R,Rayner P J W, “Single channel nonstationary stochastic signal separation using linear time-varying filters,” IEEE Transactions, Vo. 51,on 7,pp 1739 – 1752, 2003

[4]Jang Gil-Jin,Lee Te-Won,Oh Yung-Hwan. Single-channel signal separation using time-domain basis functions[J]. IEEE Signal Processing Letters, Vo.10,on 6, pp168-171, 2003

[5] J. Zhang, W. L. Woo, and S. S. Dlay, “Blind source separation of post-nonlinear convolutive mixture,” IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 8, pp. 2311–2330, Nov. 2007.

[6]J.-T. Chien and H.-L. Hsieh, “Nonstationary source separation using sequential and variational Bayesian learning,” IEEE Trans. Neural Netw.Learn, Syst., vol. 24, no. 5, pp. 681–694, May 2013.

[7]B. Gao, W. L. Woo, and S. S. Dlay, “Variational regularized 2-D nonnegative matrix factorization,” IEEE Trans. Neural Netw. Learn. Syst., vol. 23, no. 5, pp. 703–716, May 2012.

[8]J.-I. Hirayama,S.-I. Maeda,and S. Ishii, “Markov and semi-

Markov switching of source appearances for nonstationary independent component analysis," IEEE Trans. Neural Netw., vol. 18, no. 5, pp. 1326–1342, Sep. 2007.
[9] James C J,Lowe D. Single channel analysis of electromagnetic brain signals through ICA in a dynamical systems framework[C]//EMBS,2001. Istanbul,Turkey:[s.n.], 2001: 1974-1977.
[10]H.S.Kim, R.Eykholt, J.D.Salas. "Nonlinear dynamics, delay times and embedding windows," in Physica D 127,pp.48-60, 1999
[11]Grassberger P, Procaccia I. Characterization of strange attractors [J]. Physical Review Letters, 1983a, 50: 346~ 350.
[12] Jutten C, Herault J. "Blind separation of sources, part I: an adaptive algorithm based on neuromimetic"[J].Signal Processing, 1991, 24(1) : 1- 10.
[13]T.-W. Lee and M. S. Lewicki, "The generalized Gaussian mixture model using ICA," in Proc. Int. Workshop on Independent Component Analysis, Helsinki, Finland, June 2000, pp. 239–244.
[14]H.S.Kim, R.Eykholt, J.D.Salas. "Nonlinear dynamics, delay times and embedding windows," in Physica D, 1999, 127,pp.48-60.
[15]W.A.Brock, W.D.Dechert, J.A.Scheinkman and B.Lebaron. "A test for independence based on the correlation dimension,".in Econ. Rev. 1996,15(3),pp.197-235.
[16]Harold Froehling, J.P. Crutchfield, Doyne Farmer, N.H. Packard, Rob Shaw "On determining the dimension of chaotic flows" Physica D: Nonlinear Phenomena, Volume 3, Issue 3, August 1981, Pages 605–617
[17]W.A.Brock, W.D.Dechert, J.A.Scheinkman and B.Lebaron. A test for independence based on the correlation dimension[J]. Econ. Rev. 1996,15(3):197-235.
[18]P.Grassberger, I.Procaccia. Measuring the strangeness of strange attractors [J]. Physica D, 1983, 9:189-208
[19]M.B.Kennel, R.Brown, H.D.I.Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction[J]. Phys. Rev. A 1992,45:3403.
[20]D.Kugiurmtzis. State space reconstruction parameters in the analysis of chaotic times series – the role of the time window length[J]. Physica D, 1996,95:13-28.