

基于相关熵的盲源分离算法

成昊, 唐斌

(电子科技大学电子工程学院 成都 611731)

【摘要】提出了基于相关熵的盲源分离算法。与传统独立成分分析(ICA)方法利用四阶统计量或时间结构的盲源分离不同,该算法从信息理论学习中的相关熵概念出发,利用相关熵中蕴涵的各偶数阶统计信息,通过参数化中心相关熵与独立性测度的关系,建立代价函数,并通过优化算法对其进行寻优,从而得到解混矩阵并分离出源信号。仿真结果表明,在分离超高斯混合源和次高斯混合源时,分离性能优于传统的ICA方法。

关键词 盲源分离; 相关熵; 独立性测度; 信息理论学习; 参数化中心相关熵

中图分类号 TN912

文献标志码 A

doi:10.3969/j.issn.1001-0548.2013.02.007

Blind Source Separation Algorithm Based on Correntropy

CHENG Hao and TANG Bin

(School of Electronic Engineering, University of Electronic Science and Technology of China Chengdu 611731)

Abstract A blind source separation algorithm based on correntropy is presented. Unlike the traditional independent component analysis (ICA) method which utilizes the forth-order statistics or temporal structure to achieve the blind source separation. This algorithm is motivated from the notion of correntropy in the information theoretic learning, utilizing the even statistics implied in correntropy. The cost function is established according to the relationship between the parametric centered correntropy and the independence measure, and then minimized by using the optimization algorithm to acquire the demixing matrix and separate the signal. Simulations show that the performance is better than the traditional ICA method when separating the mixture of the super-Gaussian source and sub-Gaussian source.

Key words blind source separation; correntropy; independence measure; information theoretic learning; parametric centered correntropy

盲源分离起源于鸡尾酒会问题^[1],即在嘈杂环境下如何通过不同传感器接收到的混合信号,分离出各个说话者的语音信号。在这过程中,并无先验信息或仅有较少的先验信息可用。随后,盲源分离被广泛扩展到了生物医学、雷达、通信和地震信号处理等领域。

为解决该问题,自文献[2]提出了H-J算法后,独立成分分析(independent component analysis, ICA)的理论体系逐渐成熟并被广泛应用到各个领域。ICA主要利用了源信号间的独立性假设,通过不同的角度对假设进行数学近似,从而达到盲源分离的目的。通常情况下,ICA算法多要求信号源中至多只有一个服从高斯分布。从信息极大化的角度出发,文献[3]提出了infomax算法;为避免自适应过程中的矩阵求逆,文献[4]提出了自然梯度算法;文献[5]利用四阶累计量矩阵,提出了特征矩阵联合近似对角

化(joint approximative diagonalization of eigenmatrix, JADE)算法;为提高计算效率,文献[6]提出了固定点ICA算法(fixed point ICA, FPICA),又称为快速ICA(FastICA)算法等。这些算法的出发点在于利用一定的统计信息对源信号的独立性进行近似体现,在此基础上实现盲源分离。通常认为,采用四阶的统计信息即可较好地实现盲源分离,现有算法也大多基于四阶的统计信息。因此,从统计角度来说,盲源分离算法还有一定发展空间。

在统计信号处理方面,文献[7-12]提出了信息理论学习(information theoretic learning, ITL)的理论体系。ITL从信息论的角度,通过将传统的香农熵(Shannon's entropy)到雷尼熵(Renyi's entropy)的扩展,探索了有别于传统的矩和累计量的统计信息,并获得了大量应用。文献[13]指出将相关熵(correntropy)和时延信息相结合的相关熵函数

(correntropy function), 可以应用于对超高斯分布源和次高斯分布源组成的混合信号的分离, 即信号源在峭度上需具有不同的符号, 但算法中多个不同时延相关熵函数的估计使得算法复杂度较高。更进一步地, 文献[14]提出并证明了基于相关熵的独立性测度, 使得这一概念有了更完善的理论依据。

本文提出了基于相关熵的盲源分离算法, 该算法不需要多个不同时延的相关熵函数的估计, 而是通过利用相关熵近似体现各个源信号之间的独立性, 建立合适的代价函数, 实现盲源分离。

1 盲源分离模型

令 t 为离散的采样时刻, 盲源分离模型为:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{1}$$

式中, 在 t 固定为某一采样时刻 t_0 时, $\mathbf{x} \in \mathbb{R}^m$ 为 m 路接收到的混合信号; $\mathbf{s} \in \mathbb{R}^n$ 为 n 路源信号; $\mathbf{A} \in \mathbb{R}^{m \times n}$ 为混合矩阵, 且通常要求 $m = n$ 。盲源分离的目的就是通过某种算法得到混合矩阵 \mathbf{A} 或者分离矩阵 \mathbf{W} 的估计, 从而通过:

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \tag{2}$$

得到源信号的估计 $\mathbf{y}(t)$ 。

2 信息理论学习的信息度量

2.1 雷尼熵和信息势

信息理论学习采用了雷尼熵作为基础和出发点。雷尼熵定义为:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log[V_\alpha(X)] \tag{3}$$

式中, X 为随机变量; $V_\alpha(X)$ 为对应的信息势 (information potential)。当 X 为离散随机变量时, 信息势定义为:

$$V_\alpha(X) = \sum_k p_k^\alpha = E_X[p_k^{\alpha-1}] \tag{4}$$

式中, $P\{X = x_k\} = p_k$ 为 X 的分布律; $k = 1, 2, \dots$ 。

当 X 为连续随机变量时, 信息势定义为:

$$V_\alpha(X) = \int p^\alpha(x) dx = E_X[p^{\alpha-1}(x)] \tag{5}$$

式中, $p(x)$ 为连续随机变量 X 的概率密度函数。本文主要研究 X 为连续随机变量的情况。当参数 $\alpha = 2$ 时, 雷尼熵和信息势变为最为常用的二次熵和二次信息势。

采用宽度 σ 的高斯核, 利用Parzen的概率密度函数估计方法^[15], 可得到二次信息势的估计为:

$$\hat{V}_{2,\sigma}(X) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G_\sigma(x_i - x_j) \tag{6}$$

式中, x_i 为随机变量 X 的 N 个样本点; $i = 1, 2, \dots, N$; $G_\sigma(\cdot)$ 为宽度为 σ 的高斯核函数。将其扩展到两个变量的情况, 得到:

$$\hat{V}_{2,\sigma}(X, Y) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G_\sigma(x_i - y_j) \tag{7}$$

式中, x_i 和 y_j 分别为随机变量 X 和 Y 的 N 个样本点; $i = 1, 2, \dots, N$; $j = 1, 2, \dots, N$ 。

2.2 相关熵

在信息理论学习基础上, 作为随机变量之间的广义相似程度的度量, 定义相关熵为:

$$v(X, Y) = E_{XY}[\kappa(X, Y)] = \iint \kappa(x, y) p_{XY}(x, y) dx dy \tag{8}$$

式中, κ 为满足 $\kappa \in L_\infty$ 的任意正定核。该相关熵指两个随机变量之间的互相关熵, 为简明起见, 在下文的分析中, 均直接以相关熵进行说明。

传统的相关定义即为选取 $\kappa(x, y) = xy$ 的特殊情况, 即只考虑二阶统计信息。在相关熵中, 通常选取 κ 为高斯核 $G_\sigma(\cdot)$, 得到:

$$v_\sigma(X, Y) = \iint G_\sigma(x - y) p_{XY}(x, y) dx dy \tag{9}$$

在实际的联合概率密度函数未知的情况下, 利用有限数据 $\{(x_i, y_i)\}_{i=1}^N$, 可得到相关熵的估计为:

$$\hat{v}_\sigma(X, Y) = \frac{1}{N} \sum_{i=1}^N G_\sigma(x_i - y_i) \tag{10}$$

将式(9)作泰勒级数展开, 可得:

$$v_\sigma(X, Y) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} E_{XY}[(x - y)^{2n}] \tag{11}$$

可以看到, 相关熵可以分解为 $X - Y$ 的偶数阶矩的加权, 因此相关熵中蕴涵了比传统的相关更多的信息量。此外, 相对于传统ICA方法多利用四阶统计信息, 相关熵更广泛地考虑到了各偶数阶统计信息, 提供了更有效实现盲源分离的可能。

若高斯核宽度 σ 缩减为 0, 即 $\lim_{\sigma \rightarrow 0} G_\sigma(x) = \delta(x)$ ($\delta(x)$ 为一Dirac Delta函数)时, 相关熵为:

$$\lim_{\sigma \rightarrow 0} v_\sigma(X, Y) = \lim_{\sigma \rightarrow 0} \iint G_\sigma(x - y) p_{XY}(x, y) dx dy = \iint \delta(x - y) p_{XY}(x, y) dx dy = \int p_{XY}(x, x) dx \tag{12}$$

即相关熵变为在联合平面上的直线 $x = y$ 上的积分。

针对该特点, 特别是在核宽度 σ 较小时带来的局限性, 对相关熵的概念进行扩展, 得到在笛卡儿坐标系中的参数化相关熵:

$$v_{a,b,\sigma}(X, Y) = E_{XY}[G_\sigma(ax + b - y)] = \iint G_\sigma(ax + b - y) p(x, y) dx dy \tag{13}$$

式中, a 和 b 为任意实数, 且要求 $a \neq 0$ 。该定义的提出允许对联合概率密度函数的积分可以沿平面上任意斜率的直线进行, 且该直线不需要一定经过平面原点, 其估计为:

$$\hat{v}_{a,b,\sigma}(X,Y) = \frac{1}{N} \sum_{i=1}^N G_{\sigma}(ax_i + b - y_i) \quad (14)$$

此外, 由于核的性质造成的非线性变换的存在, 即使输入数据已经被中心化的情况下, 得到的相关熵也并非总是零均值。因此, 定义中心相关熵:

$$u_{\sigma}(X,Y) = E_{XY}[G_{\sigma}(x,y)] - E_X E_Y[G_{\sigma}(x,y)] = \iint G_{\sigma}(x,y)(p_{XY}(x,y) - p_X(x)p_Y(y))dxdy \quad (15)$$

其估计为:

$$\hat{u}_{\sigma}(X,Y) = \frac{1}{N} \sum_{i=1}^N G_{\sigma}(x_i - y_i) - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G_{\sigma}(x_i - y_j) = \hat{v}_{\sigma}(X,Y) - \hat{V}_{2,\sigma}(X,Y) \quad (16)$$

从而将信息熵与相关熵的概念联系起来。

将式(13)的参数化相关熵与式(15)的中心相关熵结合起来, 得到参数化中心相关熵:

$$u_{a,b,\sigma}(X,Y) = E_{XY}[G_{\sigma}(ax+b-y)] - E_X E_Y[G_{\sigma}(ax+b-y)] = \iint G_{\sigma}(ax+b-y)(p_{XY}(x,y) - p_X(x)p_Y(y))dxdy \quad (17)$$

其估计为:

$$\hat{u}_{a,b,\sigma}(X,Y) = \frac{1}{N} \sum_{i=1}^N G_{\sigma}(ax_i + b - y_i) - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G_{\sigma}(ax_i + b - y_j) = \hat{v}_{a,b,\sigma}(X,Y) - \hat{V}_{2,\sigma}(aX + b, Y) \quad (18)$$

2.3 核宽度选取

在信息理论学习中, 核宽度 σ 是其中的一个自由参数。在核宽度的选取上, 一个建议的准则为 Silverman 准则(Silverman's rule):

$$\sigma_{\text{opt}} = \sigma_x [4T^{-1}(2d+1)^{-1}]^{\frac{1}{d+4}} \quad (19)$$

式中, T 为数据长度; d 为数据维度; σ_x 为数据标准差。实际中, 根据不同情况, 选取不同的核宽度, 可能会得到更好的性能。

3 基于相关熵的盲源分离

3.1 相关熵的独立性测度

在概率论中, 传统的对随机变量间的独立性是通过概率密度函数定义的, 即:

$$p_{XY}(x,y) = p_X(x)p_Y(y) \quad (20)$$

在利用信息理论学习对随机变量独立程度的测度上, 文献[14]提出了利用参数化中心相关熵的独立性测度:

$$\Gamma(X,Y) = \sup_{a,b} |u_{a,b}(X,Y)| \quad (21)$$

其中 $a, b \in \mathbb{R}$ 。该文证明了当且仅当随机变量 X 和 Y 相互独立时 $\Gamma(X,Y) = 0$ 。然而, 利用该方法进行独立测度计算时, 需要对参数 a 和 b 进行二维搜索, 具有较高的计算复杂度。因此, 可采用一个近似的独立性测度:

$$\gamma(X,Y) = \max(|u(X,Y)|, |u(-X,Y)|) \quad (22)$$

即只计算 $\{a,b\} = \{1,0\}$ 和 $\{a,b\} = \{-1,0\}$ 时的参数化中心相关熵, 并取二者绝对值的最大值。通过对 $\gamma(X,Y)$ 的计算, 减小了计算复杂度, 得到随机变量间独立性测度的近似估计。

3.2 盲源分离算法

考虑由超高斯信号和次高斯信号混合得到的两路接收混合信号, 按照常规的分离流程, 首先对其进行球化, 以去除其中的二阶相关并使得各路信号功率相同, 得到预处理后的信号 $\tilde{\mathbf{x}}$ 。此后的解混矩阵为一正交矩阵, 可表示为旋转角度的形式:

$$\tilde{\mathbf{W}} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \quad (23)$$

选取 $\gamma(X,Y)$ 作为代价函数, 通过梯度下降法对其进行最小化, 得到对应的旋转角 θ , 从而获得恢复信号 \mathbf{y} 。算法使得对解混矩阵中四个元素的寻优转化为对标量 θ 的寻优, 降低了计算维数。

算法的具体流程如下:

1) 对接收到的混合信号 \mathbf{x} 进行球化, 得到预处理后的信号 $\tilde{\mathbf{x}}$;

2) 选取初值 θ_0 和小的角度增量 $\Delta\theta$, 计算:

$$\mathbf{y}(\theta_k) = \begin{bmatrix} \cos \theta_k & \sin \theta_k \\ -\sin \theta_k & \cos \theta_k \end{bmatrix} \tilde{\mathbf{x}}$$

$$\mathbf{y}(\theta_k + \Delta\theta) = \begin{bmatrix} \cos(\theta_k + \Delta\theta) & \sin(\theta_k + \Delta\theta) \\ -\sin(\theta_k + \Delta\theta) & \cos(\theta_k + \Delta\theta) \end{bmatrix} \tilde{\mathbf{x}}$$

3) 对 \mathbf{y} 中的两路 y_1 和 y_2 , 计算代价函数:

$$\hat{\gamma}(y_1(\theta_k), y_2(\theta_k)) =$$

$$\max(|\hat{u}(y_1(\theta_k), y_2(\theta_k))|, |\hat{u}(-y_1(\theta_k), y_2(\theta_k))|)$$

和

$$\hat{\gamma}(y_1(\theta_k + \Delta\theta), y_2(\theta_k + \Delta\theta)) =$$

$$\max\left(|\hat{u}(y_1(\theta_k + \Delta\theta), y_2(\theta_k + \Delta\theta))|, |\hat{u}(-y_1(\theta_k + \Delta\theta), y_2(\theta_k + \Delta\theta))|\right)$$

4) 选取步长 η , 计算:

$$\theta_{k+1} = \theta_k - \eta \frac{\hat{\gamma}(y_1(\theta_k + \Delta\theta), y_2(\theta_k + \Delta\theta)) - \hat{\gamma}(y_1(\theta_k), y_2(\theta_k))}{\Delta\theta}$$

5) 选取门限 ε , 当 $|\hat{\gamma}(y_1(\theta_{k+1}), y_2(\theta_{k+1})) - \hat{\gamma}(y_1(\theta_k), y_2(\theta_k))| < \varepsilon$ 时, 得到输出 $\mathbf{y}(\theta_k)$, 否则, $k \leftarrow k + 1$, 转步骤2)。

4 计算机仿真

4.1 仿真参数设置与性能评价指标

在本文的计算机仿真中, 分别采用拉普拉斯分布源和均匀分布源生成超高斯源信号和次高斯源信号。拉普拉斯分布的概率密度函数为:

$$f_L(x) = \frac{\lambda_L}{2} e^{-\lambda_L|x|} \quad (24)$$

选取 $\lambda_L = \sqrt{2}$ 以使得其零均值且方差为1。均匀分布的概率密度函数为:

$$f_U(x) = \begin{cases} \frac{1}{b_U - a_U} & a_U < x < b_U \\ 0 & \text{其他} \end{cases} \quad (25)$$

选取 $\{a_U, b_U\} = \{-\sqrt{3}, \sqrt{3}\}$ 以使得其零均值且方差为1。

在下述除实验1以外的其他仿真中, 采用一随机产生的混合矩阵:

$$\mathbf{A}_1 = \begin{bmatrix} 0.5377 & 0.7588 \\ 0.9622 & 0.7354 \end{bmatrix}$$

对本算法和FPICA算法分别进行了球化预处理, 使得算法受混合矩阵影响较小。

在实验中, 采用Amari的输出信干比作为算法性能的衡量。输出信干比定义为:

$$\text{SIR} = \frac{1}{2} \sum_{i=1}^2 10 \lg \left(\frac{\|s_i(t)\|_2^2}{\|y_i(t) - s_i(t)\|_2^2} \right) \quad (26)$$

这里不考虑盲分离中可能会出现分离信号的反号与排序问题, 即选择所有可能组合中最大的平均信干比作为输出。

4.2 仿真结果分析

实验1中, 对于数据长度为1 000的两路源信号, 在不同的旋转角下分别计算代价函数, 关系曲线如图1所示。由于在盲分离中不考虑反号与排序问题, 旋转角范围选择为 $[-\pi/4, \pi/4]$ 。由图1看到, 代价函数在旋转角接近零时最小, 近似体现了源信号之间的独立性; 此外, 代价函数与旋转角的关系曲线仅存在一个全局最小值, 因此采用梯度下降法不会陷入局部最优。

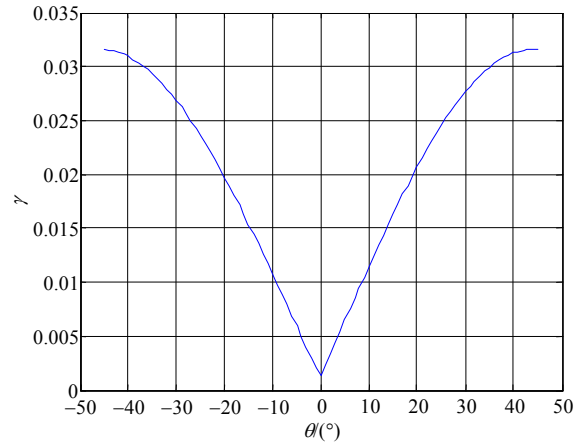


图1 代价函数与旋转角度关系曲线

实验2对本算法的分离效果进行了仿真, 数据长度为1 000的源信号、混合信号以及利用本算法的分离信号如图2所示。可以直观看出, 不考虑分离信号可能会出现反号和排序问题, 本算法有效地进行了分离。此外, 在该次实验中, 通过计算得到输出信干比为23.5 dB。

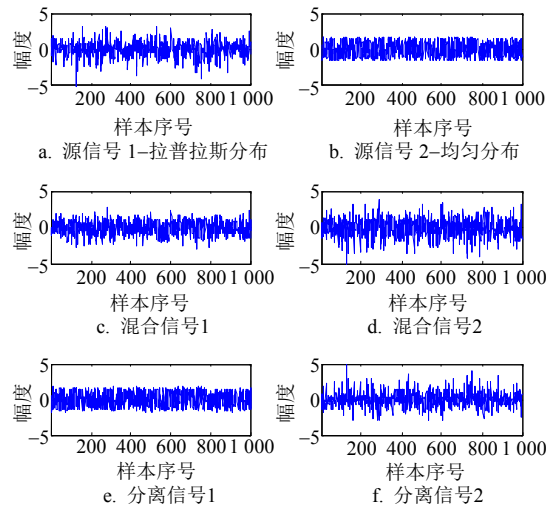


图2 源信号、混合信号与分离信号

实验3研究了核宽度的选取对于算法性能的影响, 关系曲线如图3所示。其结果为100次仿真取平均值, 数据长度为1 000。可以看到, 在有无噪声情况下, 本算法的性能受核宽度影响均不明显。

实验4研究了算法性能与数据长度的关系, 并与传统的FPICA算法进行对比, 其结果为100次仿真取平均值, 如图4所示。可以看到, 无噪声情况下, 本文算法与FPICA算法类似, 数据长度越长, 算法性能越好, 在数据长度接近1 000时, 相对于传统的ICA算法, 本文算法的性能优势开始体现。当输入中带有噪声时, 两种算法性能出现明显降低, 表明本文

算法和传统ICA算法类似,对噪声较为敏感。在信噪比为30 dB时,较无噪声情况,输出信干比下降已较为严重,但本算法在数据长度接近1 000时,性能仍然优于FPICA算法。当信噪比降至20 dB时,两种算法的输出信干比进一步下降,且几乎与数据长度无关。此时的输出信干比约为5 dB,分离效果受到较大影响。在此情况下,本算法相对于FPICA算法优势已经不再明显。

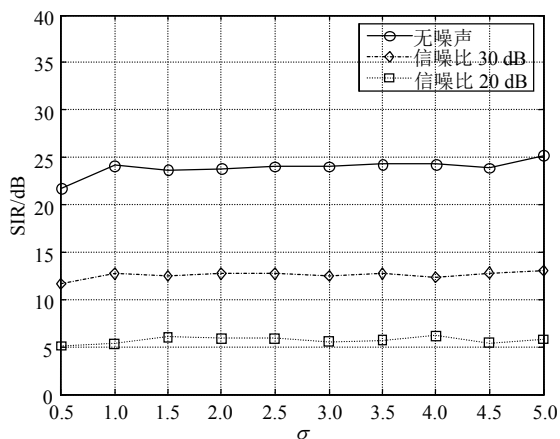


图3 输出信干比与核宽度关系曲线

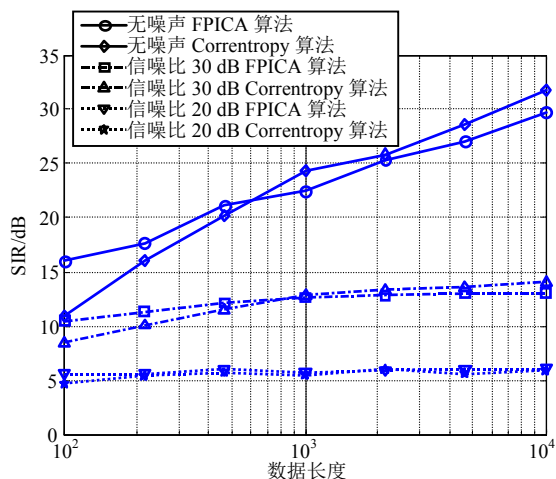


图4 超高斯分布源与次高斯分布源混合时
输出信干比与数据长度关系曲线

实验5在实验4的基础上,将源信号中作为次高斯源的均匀分布源信号替换为一个雷达中常见的线性调频(LFM)信号,使其具有了时间结构。LFM信号的归一化起始频率为0.105,归一化带宽为0.082,仿真结果如图5所示。与实验4类似,在无噪声和信噪比30 dB时,数据长度越长,算法性能越好。在数据点数接近1 000时,相对于传统的ICA算法,本算法具有性能优势。同样,当信噪比降至20 dB时,两种算法分离性能均受到较大影响,且几乎不受数据长度影响,本算法相对于FPICA算法优势已不再明显。

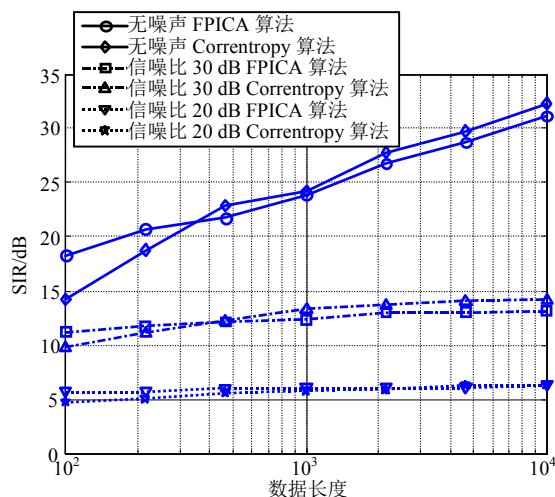


图5 超高斯分布源与LFM信号混合时
输出信干比与数据长度关系曲线

5 结论

本文从信息理论学习原理出发,基于相关熵与独立性测度的关系,在传统ICA方法多使用四阶统计信息的基础上,利用相关熵对其进行了扩展,实现对超高斯混合信号和次高斯混合信号的分离。仿真表明,本文算法优于传统的ICA算法。同时也表明算法受到核宽度选取的影响不大。但本文算法讨论的是对于超高斯分布源和次高斯分布源的分离,下一步的研究内容为如何在源信号峭度符号相同时进行分离,以便扩展到更多数目源信号的盲源分离中。

6 致谢

作者在与Jose Principe、Sohan Seth和Ruijiang Li等人的交流中受益匪浅,在此表示感谢。

参考文献

- [1] CHOI S, CICHOCKI A. Adaptive blind separation of speech signals: cocktail party problem[C]//Proc International Conference on Speech Processing. Seoul, Korea: [s.n.], 1997: 617-622.
- [2] JUTTEN C, HERAULT J. Blind separation of sources, part I: an adaptive algorithm based on neuromimetic[J]. Signal Processing, 1991, 24(1): 1-10.
- [3] BELL A J, SEJNOWSKI T J. An information-maximization approach to blind separation and blind deconvolution[J]. Neural Computation, 1995(7): 1129-1159.
- [4] AMARI S. Natural gradient works efficiently in learning[J]. Neural Computation, 1998, 10(2): 251-276.
- [5] CARDOSO J F. High-order contrasts for independent component analysis[J]. Neural Computation, 1999, 11(1): 157-192.

(下转第224页)

- AES-256[C]//Advances in Cryptology-CRYPTO 2009. Berlin: Springer-Verlag, 2009: 231-249.
- [9] DUNKELMAN O, KELLER N, SHAMIR A. Improved single-key attacks on 8-round AES-192 and AES-256[C]//Advances in Cryptology-ASIACRYPT 2010. Berlin: Springer-Verlag, 2011: 158-176.
- [10] MALA H, DAKHILALIAN M, RIJMEN V, et al. Improved impossible differential cryptanalysis of 7-round AES-128[C]//Progress in Cryptology-INDOCRYPT 2010. Berlin: Springer-Verlag, 2011: 282-291.
- [11] BIRYUKOV A, NIKOLIĆ I. Automatic search for related-key differential characteristics in byte-oriented block ciphers: Application to AES, camellia, khazad and others[C]//Advances in Cryptology - EUROCRYPT 2010. Berlin: Springer-Verlag, 2010: 322-344.
- [12] WEN Ji, LEI Hu. Square attack on reduced-round zodiac cipher[C]//Information Security Practice and Experience-ISPEC. Berlin: Springer-Verlag, 2008: 377-391.
- [13] LI Yan-jun, WU Wen-ling, ZHANG Lei. Integral attacks on reduced-round ARIA block cipher[C]//Information Security Practice and Experience-ISPEC. Berlin: Springer-Verlag, 2010: 19-29.
- [14] SUN Bin, LI Ri-lin, QU Long-jiang, et al. SQUARE attack on block ciphers with low algebraic degree[J]. Science China Information Sciences, 2010, 53(10): 1988-1995.

编辑 漆蓉

(上接第218页)

- [6] HYVÄRINEN A. Fast and robust fixed-point algorithms for independent component analysis[J]. IEEE Trans Neural Networks, 1999, 10 (3): 626-634.
- [7] HILD II K, ERDOGMUS D, TORKKOLA K, et al. Feature extraction using information-theoretic learning[J]. IEEE Trans Pat Anal Mach Intell, 2006, 28(9): 1385-1392.
- [8] SANTANA E, PRINCIPE J C, SANTANA E E, et al. Extraction of signals with specific temporal structure using kernel methods[J]. IEEE Transactions on Signal Processing, 2010, 58(10): 5142-5150.
- [9] PRINCIPE J. Information theoretic learning: renyi's entropy and Kernel perspectives[M]. New York: Springer, 2010.
- [10] GUNDUZ A, HEGDE A, PRINCIPE J C. Correntropy as a novel measure for nonlinearity tests[J]. Signal Processing, 2009, 89(1): 14-23.
- [11] JEONG K H, LIU W, PRINCIPE J. The correntropy MACE filter[J]. Pattern Recognition, 2009, 42(5): 871-885.
- [12] VERA P A, ESTEVEZ P A, PRINCIPE J C. Linear projection method based on information theoretic learning[C]//ICANN'10 Proceedings of the 20th international conference on Artificial neural networks: Part III. Thessaloniki, Greece: [s.n.], 2010: 178-187.
- [13] LI R, LIU W, PRINCIPE J C. A Unifying criterion for instantaneous blind source separation based on correntropy[J]. Signal Processing, 2007, 87(8): 1872-1881.
- [14] RAO M, SETH S, XU J, et al. A test of independence based on a generalized correlation function[J]. Signal Processing, 2011, 91(1): 15-27.
- [15] PARZEN E. On the estimation of a probability density function and the mode[J]. Ann Math Statist, 1962(33): 1065-1067.

编辑 税红