

# 基于遗传算法的盲源分离算法

李良敏

(西安交通大学机械工程学院, 710049, 西安)

**摘要:** 针对现有盲源分离算法的性能依赖于对比函数选择的现象, 提出了一种基于遗传算法的盲源分离算法, 该算法直接从信号的样本序列中估计出信号的概率分布, 解决了信号间互信息的求解问题. 通过遗传算法最小化信号的互信息, 实现了对线性混叠信号的分离. 对模拟信号的分离结果表明, 该算法可以成功地分离混叠信号, 同时与快速独立分量分析算法相比, 该算法的性能对源信号的概率密度性质没有依赖, 因而对亚高斯和超高斯信号的混合信号表现出更加优异的分离能力.

**关键词:** 盲源分离; 遗传算法; 互信息; 超高斯; 亚高斯; 快速独立分量分析算法

**中图分类号:** TP18 **文献标识码:** A **文章编号:** 0253-987X(2005)07-0740-04

## New Blind Source Separation Method Based on Genetic Algorithm

Li Liangmin

(School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

**Abstract:** The performance of existing blind source separation methods is highly affected by the non-linear contrast functions that are selected according to the distribution of original signals, and the separation results are not always ideal, especially for the mixture of super-Gaussian signal and sub-Gaussian signal. To solve this problem, a new blind source separation method based on genetic algorithm is proposed, where the probability of separated signals is estimated directly from their samples, so the mutual entropy can be easily evaluated, and genetic algorithm is applied to find the separation matrix to minimize the mutual entropy. The simulated results show that the proposed method is superior to FastICA in separating the mixture of super-Gaussian signal and sub-Gaussian signal.

**Keywords:** *blind source separation; genetic algorithm; mutual entropy; super-Gaussian; sub-Gaussian; FastICA*

近年来盲源分离<sup>[1]</sup> (BSS) 已成为信号处理学界的研究热点之一, 其实质为在没有任何先验知识的前提下, 仅仅依靠源信号的一些基本统计特征, 如统计独立, 便可从混叠信号中恢复提取出源信号来. 独立分量分析<sup>[2,3]</sup> (ICA) 是最常见的实现盲源分离的算法. 围绕解决 ICA 问题已经出现了许多算法<sup>[4-6]</sup>, 大多是围绕源信号的独立性提出各种独立性准则, 结合各种优化方法来寻找分离矩阵, 如 1995 年 Bell 和 Sejnowski 提出的基于信息最大化准则的独立分量分析方法 (InfoMax 算法), 1999 年 Hyvarinen 提

出的基于负熵最大化准则的快速独立分量分析算法 (FastICA) 等. 但是, 无论是基于哪种独立性准则, 现有的 ICA 算法多涉及非线性函数的选取问题, 如 FastICA 中的  $G(u)$  函数模型选择, InfoMax 算法中的  $g(y)$  函数模型等, 这些函数模型的选择主要取决于源信号的概率密度性质, 也即峭度性质. 然而在工程应用中, 源信号的峭度性质在信号被分离前是无从得知的, 这种矛盾极大地影响了 ICA 算法的分离能力, 尤其是对于亚高斯和超高斯的混合信号.

本文针对盲源分离算法的缺点, 提出了一种基

于遗传算法的盲源分离算法,该算法以信号间的互信息为优化目标,采用直方图法<sup>[7]</sup>直接从信号的样本分布中估计出信号的概率分布,解决了互信息的求解问题,通过遗传算法最小化信号的互信息,实现了对线性混叠信号的分离.由于该算法对于源信号的峭度性质没有依赖,因此适用于对各种类型混合信号的分离.

## 1 ICA 的原理介绍

ICA 问题可以用下面的方程描述

$$\left. \begin{aligned} x(t) &= As(t) \\ s(t) &= [s_1(t), s_2(t), \dots, s_n(t)]^T \\ x(t) &= [x_1(t), \dots, x_m(t)]^T \end{aligned} \right\} \quad (1)$$

式中: $s(t)$ 为源信号; $x(t)$ 为观测向量,其中的元素为各个传感器的实测信号; $A$ 为混合矩阵.ICA的实质是在 $s(t)$ 和 $A$ 均未知的情况下,根据 $x(t)$ 找到一个分离矩阵 $W$ ,使 $y(t) = Wx(t)$ 成为 $s(t)$ 的拷贝或估计.

盲源分离主要建立在下列前提假设的基础上:

① 源信号各分量相互统计独立,且其中最多只能有一个分量服从高斯分布;② 传感器数目不小于源信号的数目.

在概率论中,独立性是通过变量间的联合概率密度与边缘概率密度之间的关系来衡量的,但计算变量的概率密度分布并非易事.事实上,并不直接用这种方法作为衡量变量独立的标准,而是采用了其他的一些准则,如互信息、负熵等等,其中以互信息最为常用.利用熵的概念,定义变量 $y_1(t), y_2(t), \dots, y_m(t)$ 的互信息为

$$\left. \begin{aligned} I(y_1, y_2, \dots, y_m) &= \\ &\sum_{i=1}^m H(y_i) - H(y_1, y_2, \dots, y_m) \\ H(y_i) &= - \int p(y_i) \lg p(y_i) dy_i \end{aligned} \right\} \quad (2)$$

式中: $H(y_i)$ 为变量的熵; $p(y_i)$ 为变量 $y_i$ 的概率密度函数. $I$ 总是非负的,当且仅当变量统计独立时它才为0,因此可以作为变量间相互独立性的一种度量.

## 2 基于遗传算法的盲源分离

简单地讲,ICA是一个优化问题,其目的是通过优化 $W$ ,以获得 $y(t)$ ,使其各分量间的独立性最强.当只有信号的有限样本时,很难对变量的概率密度以表达式的形式予以描述,传统的基于梯度的寻优

方法无法直接对独立性判据进行优化,而需引入非线性函数对独立性判据进行近似,使优化目标的梯度信息可以求解,但这同时引出另一个问题,即所选用的非线性函数对于信号的峭度性质存在很大的依赖性,解的质量无法得到保证.

遗传算法<sup>[8]</sup>作为一种全局并行搜索算法,其寻优不依赖问题的梯度信息,对于盲源分离问题,无需给出变量概率密度函数的表达式形式,只要能够解决信号间互信息的计算问题,然后通过遗传算法寻找使互信息最小的 $W$ ,即可实现对混合信号的分离.

### 2.1 生成初始群体

首先需要确定个体的编码方式,由于优化变量数目较少,故采用二进制编码方式.在确定优化变量的取值范围时,由于改变 $W$ 中各行元素的大小或行向量的符号均不会改变分解信号间的独立性,因此可以将分离矩阵中各元素的取值范围限定在 $[-1, 1]$ 之间.然后,随机生成 $M$ 组定长的二进制串,即可形成群体规模为 $M$ 的初始群体.

### 2.2 个体的适应度评价

个体的优劣可用式(2)来评价, $I$ 越小,表明个体越优秀,但这样又引发了如何计算变量间互信息的问题,因为互信息与各变量的熵和变量间的联合熵有关,所以要计算互信息就必须求出各变量的概率密度和变量间的联合概率密度.常见的概率密度估计方法包括直方图法、Rosenblatt法、Parzen的核估计法及最近邻估计法等<sup>[7]</sup>,其中以直方图法最为简单易行,且当样本数较大时所得到的图像能够显示密度的基本特征,因此选择用直方图法进行密度估计.

以二维向量为例,具体计算过程如下.

(1) 样本点数偏少时,可先采用Bootstrap统计模拟方法<sup>[9]</sup>对样本进行扩充.

(2) 以 $\Delta t$ 为长度,将第 $i$ 个分量所在的样本空间等分为 $N_i$ 个区间,则整个样本空间被分为 $N_1 \times N_2$ 个子区域.

(3) 统计样本落入各个子区域的点数 $D_{ij}, i=1, \dots, N_1; j=1, \dots, N_2$ .

(4) 当样本点数足够多,而 $\Delta t$ 足够小时,可以用样本落在区间 $(i, j)$ 的概率来近似估计变量的联合概率

$$P(y_{1i}, y_{2j}) = \frac{D_{ij}}{l} \quad (3)$$

而各变量的边缘概率可以用下式进行估计

$$P(y_{1i}) = \frac{\sum_j D_{ij}}{l} \quad (4)$$

$$P(y_{2j}) = \frac{\sum_i D_{ij}}{l} \quad (5)$$

其中  $l$  为样本的长度。

(5) 当变量的联合概率和边缘概率估计求出后,可以通过下式计算相应的熵

$$H(y) = - \sum_i P(y_i) \lg P(y_i) \quad (6)$$

(6) 将求得的熵代入式(2),即可求出变量的互信息。

由于求解变量的联合概率需要对多维空间进行划分,会出现计算量大的问题,随着变量维数的增加,这个问题更加明显。为了解决这个问题,下面给出一种简化的互信息计算方法。

当  $m=n$  时,对于  $y(t)=Wx(t)$ ,有下式成立

$$H(y) = H(x) + \lg |\det W| \quad (7)$$

那么

$$\begin{aligned} I(y_1, y_2, \dots, y_m) &= \\ \sum_{i=1}^m H(y_i) - H(y_1, y_2, \dots, y_m) &= \\ \sum_{i=1}^m H(y_i) - H(x) - \lg |\det W| & \quad (8) \end{aligned}$$

由于  $H(x)$  为定值,则优化目标可以简化为

$$\min \left( \sum H(y_i) - \lg |\det W| \right) \quad (9)$$

式(9)中无须求解变量的联合概率,而边缘概率的求解只需对一维空间进行划分,因此计算量得到明显减少。

在对种群进行适应度评价后,执行选择、杂交和变异等遗传操作,生成新的种群,然后判断是否满足终止条件,如满足,输出最优解,否则,重复执行上述操作,直至满足终止条件。

### 3 仿真试验

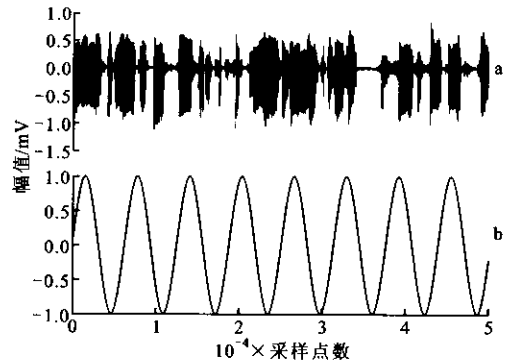
为了验证本文提出算法的可行性,将本文算法应用于一组模拟信号的分离,并将分离效果与目前最流行的 ICA 算法——FastICA<sup>[6]</sup>的分离效果进行了比较。为了简化描述,将以式(2)作为优化目标的 ICA 算法简称为 I-GaICA,将以式(9)作为优化目标的 ICA 算法简称为 MI-GaICA。

仿真试验采用一组声音信号与正弦信号作为源信号,波形见图 1。由于 2 组信号的峭度值分别为 1.1750, -0.3770,说明这是一个超高斯与亚高斯混

合信号的分离问题,经过一个随机产生的混合矩阵

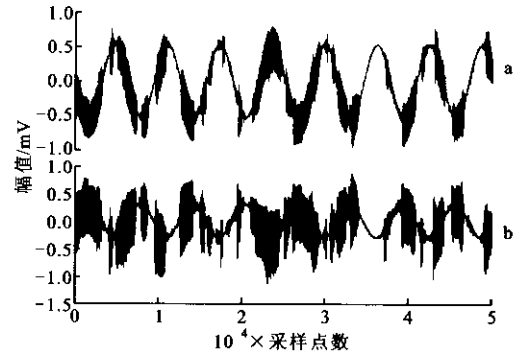
$$A = \begin{bmatrix} 0.44 & -0.53 \\ 0.80 & 0.30 \end{bmatrix}$$

混合后得到的混叠信号波形如图 2 所示。



a: 声音信号; b: 正弦信号

图 1 源信号波形



a: 混合信号 1; b: 混合信号 2

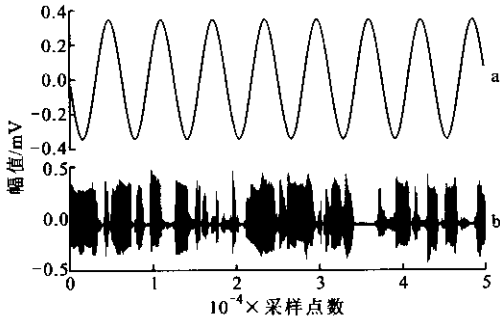
图 2 混合信号波形

本例中遗传算法的各控制参数分别为:群体规模  $M=30$ , 杂交率  $P_c=0.9$ , 变异率  $P_m=0.05$ , 最大进化代数  $N_g=40$ 。根据源信号的峭度性质,采用  $G(u) = \frac{1}{a} \lg \cosh(au)$  作为 FastICA 的  $G(u)$  函数。

3 种 ICA 算法对混合信号的分离效果见图 3~图 5。从图中可以看出, I-GaICA 算法及 MI-GaICA 算法均成功地将源信号分离了出来,尤其是 I-GaICA 算法对正弦信号的分离最成功。对于 FastICA 算法,分离出的声音信号中仍可看出有正弦信号的存在,正弦信号中也混叠着部分声音信号,表明源信号没有被完全分离出来。

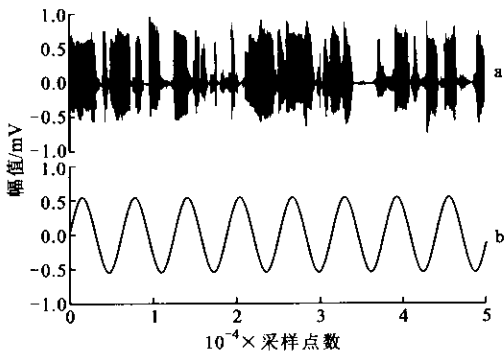
从图中还可以看出, ICA 算法分离的信号与源信号相比,在幅值与相位上存在着差异。从  $y(t) = Wx(t)$  中可以看出,改变  $W$  中各行元素的大小(行向量中各元素按同样比例放大或缩小)、调整行向量的

次序、改变行向量的符号,均不改变分离信号间的独立性,因此在盲源分离过程中,分离信号存在着幅值、相位和次序的不确定性,这种现象存在于大多数 ICA 算法中,目前还没有很好的解决方法。



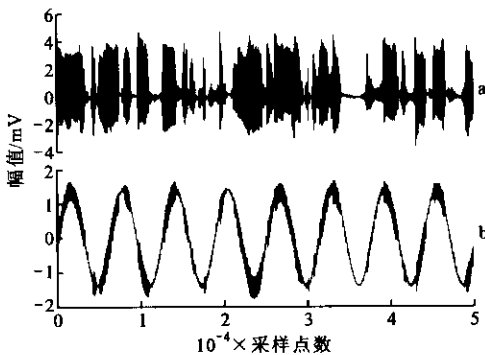
a:分离信号 1;b:分离信号 2

图 3 I-GaICA 算法的分离结果



a:分离信号 1;b:分离信号 2

图 4 MI-GaICA 算法的分离结果



a:分离信号 1;b:分离信号 2

图 5 FastICA 算法的分离结果

对分离结果除了可以用直观的观察法来评估分离效果外,还可以采用一些定量的指标来衡量<sup>[10]</sup>。一种方法是用乘积矩阵  $WA$  的行元素优势度量函数  $f(WA)$  来衡量,其定义式为

$$f(WA) =$$

$$\begin{cases} \max_{j,i} (|q_{ij}/q_{ik}|), & j, i = 1, 2, \dots, n; j \neq k, \text{如果} \\ & WA \text{ 是行元素优势矩阵} \\ 1, & \text{如果 } WA \text{ 不是行元素优势矩阵} \end{cases}$$

$$q_{ik} = \max_j |q_{ij}|, \quad j = 1, 2, \dots, n \quad (10)$$

式中: $q_{ij}$ 为乘积矩阵  $WA$  中的元素; $q_{ik}$ 为第  $i$  行的优势元素. 如果  $WA$  所有行的优势元素不同列,则称  $WA$  为行元素优势矩阵.  $f(WA)$  越小,说明乘积矩阵中每行的最大值和其他值之间的差异越大,盲源分离效果越好。

另一种方法是采用分离后信号的信噪比  $R_{SN}(y_i)$  作为分离效果的评判指标

$$R_{SN}(y_i) = 10 \lg \frac{E(|s_i|^2)}{E(|y_i - s_i|^2)} \quad (11)$$

信噪比越大,盲源分离效果越好。

表 1 给出了几种盲源分离算法对本例的分离性能指标. 可以看出, FastICA 算法对于超高斯和亚高斯的混合信号虽然有一定的分离能力,但分离结果并不理想. 本文的 2 种算法由于不存在选取非线性函数的问题,可以用直方图法对任何类型的信号进行概率密度估计,因此能适用于对任何类型混合信号的分离。

表 1 几种算法对超高斯与亚高斯混合信号的分离性能比较

分离算法	$f(WA)$	$R_{SN}/\text{dB}$	
		$y_1$	$y_2$
I-GaICA	0.001 9	129.766 5	161.479 2
MI-GaICA	0.009 2	228.110 1	109.415 8
FastICA	0.203 3	81.852 3	48.807 0

## 4 结 论

本文介绍了独立分量分析的基本理论,针对典型实现算法的分离效果依赖于非线性函数选择的缺点,提出了一种基于遗传算法的盲源分离算法,通过遗传算法优化分离矩阵,使得分离信号各分量间的互信息最小. 本文算法的优点在于无须选择非线性函数,对源信号的峭度性质没有依赖,因此适用于对各种类型混合信号的分离,这一点从仿真试验中得到了充分证明,但缺点在于计算速度慢,不适用于信号的在线分离,因此还需要进一步研究以提高算法的计算效率。

量存储和释放! 同时还能够保持分布式发电系统的稳定输出:

参考文献"

#\_ \$ ^8+812! VVDR- V! -<J B: , 9CS@UYQCMU9  
 SLSJT @NCDYJDNRL PEDUNESB# , \$: ENBYUNPS@  
 HW \$\$\$ C555 V D YQD V@ONDPJ @D 59PN@9 d  
 VET@HINERDUNDR#V\$: e@O@R! ?I @B@M! V DA  
 IYI " C555 6N@D V D@N! \$\$\$: \$&= \$&\_:

# \$ 4ISE: 4JPN@F@D@9@, V TTP@D@S#? \$: -8D@D"  
 (Z@BY) @WENSS! =" "%> %/ # %&

# \$ 陈伯时: 电力拖动自动控制系统#? \$: 北京" 机械工业  
 出版社! \$\$\$: \$! \$&&:

#> \$ BJ@D@M@J ? : @D@P@S@Q@F@D@S@ %B@V@&@D@J@N@M@  
 OY@D@P@D@T@P@D@ # \$: C555 2N@D@S@P@E@S@C@  
 JN59PN@B@D@S! =" @G! "%&" > \$ %> \$ :

#& \$ V@M@J B, ! . <J <Y 7: , D@C@B@N@<J I D Y@Q@Z@P@D@S@  
 T@H@Y@B@N@S@C@P@M@Y@N@Q@P@H@D@J T@B@M@Q@S# \$: C555

2N@D@S@P@E@S@C@JN59PN@B@D@S! \$\$\$! =#%&"& -  
 &&#:

#\_ \$ 郭 宏! 王光能: 异步电动机直接转矩控制系统的  
 ? , 2- , ^ 仿真# \$: 北京航空航天大学学报! \$\$\$!  
 \$G%&" #GG-:

## \$ 曹承志! 曲红梅! 路战红! 等: ? , 2- , ^ 软件包中  
 @P) -O+K 环境下直接转矩控制系统的仿真# \$: 电  
 机与控制学报! \$\$\$! &%\$&" == =>:

#G \$ 李华德! 白 晶! 李志民! 等: 交流调速控制系统#? \$:  
 北京" 电子工业出版社! \$\$\$ : =\_ \$\$\$:

# \$ 李 夙: 异步电动机直接转矩控制#? \$: 北京" 机械工  
 业出版社! \$\$\$: >=# :

#> \$ . I@X' ! @V@N@D@ ? ^ ^: V@T@N@S@D@ @Y@O@N@D@H  
 S@C@P@D@R@H@N@D@S@O@Y@P@H@B@N@<J@P@D@S@J@P@W@<J@  
 O@Y@P@D@T@B@S# \$: 59PN@P@ E@S@C@N@ @S@J@T@S@ \*J@  
 S@I@N@W! \$\$\$! \_%\$&" \_! #&:

%编辑 管咏梅&

%上接第#> 页&

参考文献"

#\_ \$ ' @H@D@V! `JN@<P@ : ^@D@S@I@N@H@B@D@S@B@N@S! @N@H"  
 I D I Y @@W I @B@M@T@M@S@Y@B@D@<N@B@T@J@K@I@N@W@H@F@A  
 H@N# \$: @D@9@E@N@P@S@D@R! =" =! \$ %&" => %&

# \$ V@T@T@D@E: @Y@D@Y@D@H@F@T@@D@D@H I D @Q@S@S" I D@C  
 P@D@J@# \$: @D@9@E@N@P@S@D@R! =" ">! \_%&" \$G#! =>:

# \$ `LU@D@D, : 6@N@U@L@B@D@Y@D@Y@D@H@F@T@@D@D@H I D @Q@A  
 S@S# \$: +J@N@I@9@V@T@H@Q@R@6@N@U@L@S! =" ""! \$ %&" "> -  
 = \$G

#> \$ ^@9, ' ! 6@D@B@C@S@Q@2' : , D@O@B@N@I@K@B@T@I@Z@O@L@K@B@  
 I@@B@P@W@H@B@M@D@Y@S@I@N@H@B@D@I@D@Y@M@D@Y@P@B@D@S@H@B@  
 # \$: +J@N@P@B@T@H@Q@R! =" "#! =#%&" \$&>\_:

#& \$ `< 7 `!, T@N@D@6@O, Y@@W @B@D@9@I@N@O@D@R I @B@A  
 N@B@M@T@S@O@N@M@D@Y@S@I@N@H@B@D@ T@I@Z@<T@J@D@N@ @ I D Y  
 T@O@D@<T@T@H@I@9@O@B@N@I@K@B@D# \$: +J@N@I@9@V@T@H@I@B@D

= " #! " %&" = ># = >G\$:

#\_ \$ `LU@D@D, : I@S@H@D@Y@N@B@S@O@Z@Y@@B@D@H@P@B@M@T@S@O@N  
 O@Y@D@Y@D@H@F@T@@D@D@H I D @Q@S@S# \$: C555 2N@D@S@B@D  
 +J@N@I@9@+J@H@C@N@S! =" ""! =%&" \$ -! >:

## \$ 陈希儒! 柴根象: 非参数统计教程#? \$: 上海" 华东师范  
 大学出版社! =" ""! : \$# \$&&:

#G \$ 陈国良! 王煦法! 庄镇泉! 等: 遗传算法及其应用#? \$:  
 北京" 人民邮电出版社! =" ""\_ : \$G-%&

# \$ @B@N@D, ? ! ^@B@S@W@S@W^: 2M@M@B@S@H@N@I@D@Y@G@S@I@@B@A  
 P@H@B@D@S@D@9@@P@S@D@R# \$: C555 @D@9@E@N@P@S@D@R  
 ?I@R@X@D! =" "G =&%&" &\_#\_:

#> \$ 徐明彪! 朱维彰: 关于信号盲分离分离效果评判指标的  
 分析# \$: 杭州电子工业学院学报! \$\$\$! \$%&" \_! \_-:

%编辑 管咏梅&