

# 基于自然梯度算法的盲信源分离研究

孙守宇<sup>1</sup>, 郑君里<sup>1</sup>, 吴德伟<sup>2</sup>

(1. 清华大学电子工程系, 北京 100084; 2. 空军工程大学 电讯工程学院, 陕西 西安 710077)

**摘要:**盲信源分离试图从给定的一组混合观察数据中恢复未知的独立信源。文中介绍盲信源分离的一种非常重要的算法——自然梯度算法。对通信信号和自然语音信号采用不同的活动函数进行了盲信源分离的计算机模拟实验,结果显示该算法能够分别有效地分离这两类随机混合的信号。

**关键词:**盲信源分离;自然梯度算法;活动函数

**中图分类号:**TN713 **文献标识码:**A **文章编号:**1009-3516(2003)03-0050-05

盲信源分离是近年来信号处理领域研究的热点,在许多领域具有广泛潜在的应用前景。例如,天线阵列处理、图像重构和储存、声控机器、医疗信号处理、语音处理和通信技术。

## 1 基本原理

1985年,Herault和Jutten首先提出盲源分离问题<sup>[1]</sup>,其目的是从一组观察信号中,求得一组不可观察的信源信号,除了信源数之外,对信源和混合过程一无所知。

信源混合和分离的方框图如图1所示。假设 $N$ 个未知零均值信源信号 $s_i(n) (i=1,2,\dots,N)$ 在任何时间相互独立,且其中最多只有一个高斯信源。无噪观察模型是

$$x(n) = As(n) \tag{1}$$

这里 $A$ 是一个 $N \times N$ 的未知非奇异混合矩阵, $s(n) = [s_1(n) s_2(n) \dots s_N(n)]^T$ 是包含信源信号的一个 $N \times 1$ 维矢量, $x(n) = [x_1(n) x_2(n) \dots x_N(n)]^T$ 是一个 $N \times 1$ 维观察矢量,它是唯一可利用的信息。

为了从观测矢量中分离出独立源信号,需要一个 $N$ 输入、 $N$ 输出的神经网络(这里采用前向网络结构),自适应调整去混合方阵 $W$ ,以便使

$$u(n) = Wx(n) = WAs(n) \tag{2}$$

这里 $u(n) = [u_1(n) u_2(n) \dots u_N(n)]^T$ 是信源信号的一个估计。注意到,如果没有别的信息,准确地估计信源信息的次序和幅度是不可能的,因为混合矩阵 $A$ 和信源矢量 $s(n)$ 都是未知的。我们定义一性能矩阵

$$P = WA \tag{3}$$

如果 $P$ 归一化并调整次序,完全分离时它是一个单位阵。

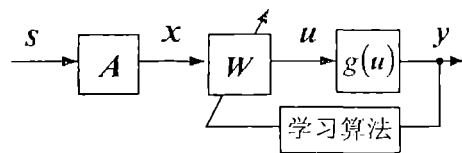


图1 信源混合和分离的方框图

收稿日期:2002-09-06

基金项目:国家自然科学基金资助项目(60072001)

作者简介:孙守宇(1962-),男,江苏沛县人,高级工程师,博士生,主要从事通信、信息处理和人工智能应用研究;  
郑君里(1938-),男,天津人,清华大学电子工程系教授,博士生导师,主要从事通信和人工智能等领域教学与研究。

## 2 信息最大原理

Bell 和 Sejnowski 把信息最大原理应用于盲分离问题<sup>[2]</sup>,证明了一个神经网络输出的联合熵  $H(\mathbf{y}; \mathbf{W})$  最大近似于输出分量  $y_i = g_i(u_i)$  ( $i=1, 2, \dots, N$ ) 之间的互信息最小,当输出的互信息为 0 时,输出变量是统计独立的。非线性变换  $y_i = g_i(u_i)$  是为了把熵限制在有限的范围之内。我们假设  $y_i = g_i(u_i)$  有唯一的逆变换  $u_i = g_i^{-1}(y_i)$ 。矢量  $\mathbf{y}$  的联合熵

$$H(\mathbf{y}; \mathbf{W}) = - \int p(\mathbf{y}; \mathbf{W}) \log p(\mathbf{y}; \mathbf{W}) d\mathbf{y} = - E\{\log p(\mathbf{y}; \mathbf{W})\} \quad (4)$$

这里  $p(\mathbf{y}; \mathbf{W})$  是由  $\mathbf{W}$  和  $\{g_i\}$  确定矢量  $\mathbf{y}$  的联合概率密度函数(pdf),  $E\{\cdot\}$  表示数学期望。式(4)可以重写为

$$H(\mathbf{y}; \mathbf{W}) = \sum_{i=1}^N H(y_i; \mathbf{W}) - I(\mathbf{W}) \quad (5)$$

这里

$$H(y_i; \mathbf{W}) = - E\{\log p(y_i; \mathbf{W})\} \quad (6)$$

是边缘熵,  $I(\mathbf{W})$  是他们的互信息。

输出密度  $p(y_i; \mathbf{W})$  和信源估计密度  $p(u_i; \mathbf{W})$  之间的影射关系描述为<sup>[3]</sup>

$$p(y_i; \mathbf{W}) = \frac{p(u_i; \mathbf{W})}{|\partial y_i / \partial u_i|} = \frac{p(u_i; \mathbf{W})}{|g'(u_i)|} \quad (7)$$

将式(7)代入式(6),式(5)变为

$$H(\mathbf{y}; \mathbf{W}) = - I(\mathbf{W}) + \sum_{i=1}^N E\left\{\log \frac{p(u_i; \mathbf{W})}{|g'(u_i)|}\right\} \quad (8)$$

由式(8)可见,如果  $p(u_i; \mathbf{W}) = |g'(u)|$ , 即估计信源的密度函数是非线性函数  $g_i$  微分的绝对值,那么最大输出的联合熵等效于输出分量之间的互信息最小。因此,式(8)的第二项对最大熵(maximum entropy, ME)原理的自适应算法的收敛性产生影响。假设总能找到非线性函数使关系:  $p(u_i; \mathbf{W}) \approx |g'(u)|$  成立。

## 3 基于 ME 的统计梯度下降规则

最大信息学习算法可以通过使输出熵最大推出

$$H(\mathbf{y}; \mathbf{W}) = - E\{\log p(\mathbf{y}; \mathbf{W})\} = - E\left\{\log \frac{p(\mathbf{x})}{|J(\mathbf{x}, \mathbf{y})|}\right\} \quad (9)$$

这里  $J(\mathbf{x}, \mathbf{y}) = \det(\mathbf{W}) \prod_{i=1}^N |g'(u_i)|$  是从矢量  $\mathbf{x}$  变换为  $\mathbf{y}$  的雅可比函数。这样输出熵可以写成

$$H(\mathbf{y}; \mathbf{W}) = H(\mathbf{x}) + \log |\det(\mathbf{W})| + \sum_{i=1}^N E\{\log |g'(u_i)|\} \quad (10)$$

容易证明

$$\frac{\partial \log |\det(\mathbf{W})|}{\partial \mathbf{W}} = (\mathbf{W}^{-1})^T$$

和

$$\frac{\partial \sum_{i=1}^N E\{\log |g'(u_i)|\}}{\partial \mathbf{W}} = - E\{\Phi(\mathbf{u}) \mathbf{x}^T\}$$

这里

$$\Phi(\mathbf{u}) = \begin{bmatrix} -\frac{\dot{g}_1'(u_1)}{g_1'(u_1)} & -\frac{\dot{g}_2'(u_2)}{g_2'(u_2)} & \dots & -\frac{\dot{g}_N'(u_N)}{g_N'(u_N)} \end{bmatrix}^T \quad (11)$$

称作算法的活动函数(activation function)<sup>[4]</sup>。输出熵  $H(\mathbf{y}; \mathbf{W})$  关于  $\mathbf{W}$  的统计梯度是

$$\frac{\partial H(\mathbf{y}; \mathbf{W})}{\partial \mathbf{W}} = (\mathbf{W}^{-1})^T - \Phi(\mathbf{u}) \mathbf{x}^T$$

文献[2] 相对应的自适应算法是

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \{ (\mathbf{W}(n)^{-1})^T - \Phi[\mathbf{u}(n)] \mathbf{x}^T(n) \} \quad (12)$$

该算法的缺点是,收敛性受混合矩阵的严重影响。自然梯度的概念首次由文献[5] 提出,后来又经文献[6] 修正。自然梯度在熵梯度的基础上后乘方阵  $\mathbf{W}^T - \mathbf{W}$  得到

$$\frac{\partial H(\mathbf{y}; \mathbf{W})}{\partial \mathbf{W}} = \mathbf{W}^T \mathbf{W} = [\mathbf{I} - \Phi(\mathbf{u}) \mathbf{u}^T] \mathbf{W}$$

这里  $\mathbf{I}$  表示单位阵。对应的自适应算法是

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \{ \mathbf{I} - \Phi[\mathbf{u}(n)] \mathbf{u}^T(n) \} \mathbf{W}(n) \quad (13)$$

自然梯度自适应算法有2个重要特性:①算法不受混合矩阵的影响;②算法阻止  $\mathbf{W}(n)$  变为奇异阵。结果加快了算法的收敛速度。

## 4 模拟实验

在盲信号处理的文献中,峭度是一个非常重要的概念,它由信号分布的高阶统计量来确定,用符号  $k_4$  表示,其归一化形式

$$k_4 = \frac{E\{s^4(n)\}}{E\{s^2(n)\}^2} - 3 \quad (14)$$

高斯分布信号的峭度为0。在下面的模拟实验中通信信号的峭度小于0(峭度小于0的信号称作欠高斯信号)。自然语音信号的峭度大于0(峭度大于0的信号称作超高斯信号)。

### 4.1 通信信号的盲分离

图2(a)是6个未知通信信号,信号的排列顺序由下向上,由一随机生成的混合矩阵  $\mathbf{A}$  混合,混合信号如图2(b)所示。该随机生成的混合矩阵是

$$\mathbf{A} = \begin{bmatrix} 0.6904 & -0.9608 & -0.5255 & -0.4734 & -0.9881 & 0.1176 \\ 0.5280 & 0.9643 & -0.1525 & 0.6108 & -0.9347 & 0.9028 \\ -0.7666 & 0.9978 & -0.0940 & -0.2288 & -0.3748 & 0.0589 \\ 0.0080 & -0.6895 & -0.4737 & 0.3672 & 0.9310 & 0.9512 \\ 0.7145 & -0.6906 & -0.4511 & -0.2777 & -0.1662 & 0.3777 \\ -0.0899 & -0.8841 & -0.7720 & 0.4761 & 0.0510 & 0.4570 \end{bmatrix}$$

混合信号通过 11 kHz 的采样率采样,取 2 000 个样点,由式(13)的算法模拟。这些信号分布的峭度均小于0(如表1所示),式(11)的活动函数选择:  $\phi(u) = u^3$ 。其它参数:学习率  $\mu = 0.01$ ,初始矩阵  $\mathbf{W}(0) = 0.1\mathbf{I}$ 。算法的分离结果如图2(c)所示。

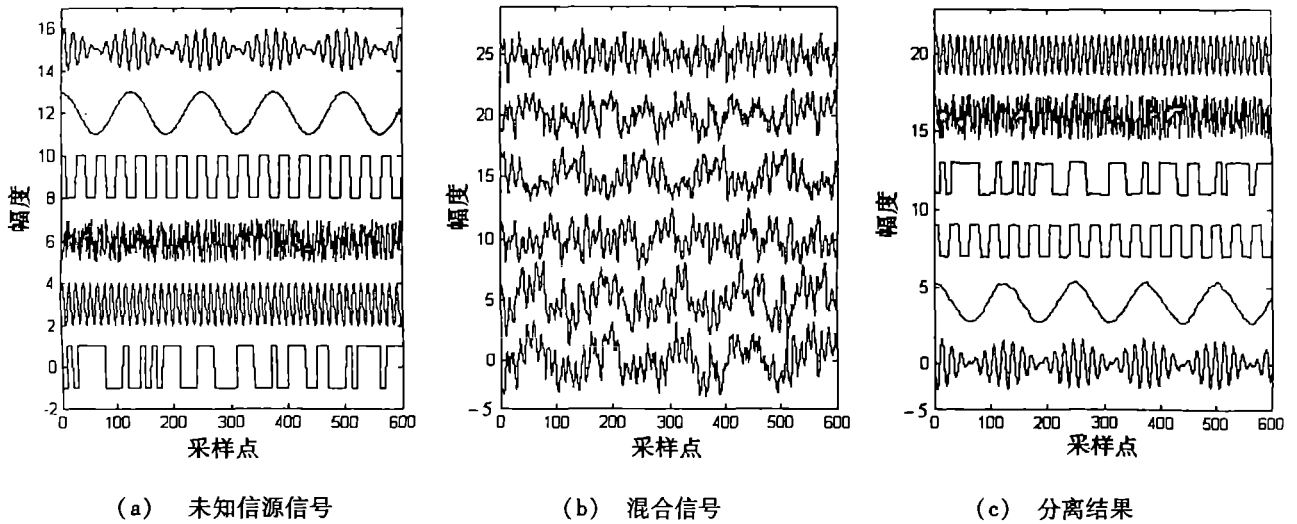


图2 6个未知通信信号混合盲分离的模拟结果

实验也可以用性能矩阵  $P$  来衡量

$$P = WA = \begin{bmatrix} 0.0610 & 0.0037 & 0.0209 & -0.0265 & -0.0016 & -1.6210 \\ 0.0548 & 0.0063 & -0.0134 & -0.0059 & 1.2767 & 0.0294 \\ 0.0227 & 0.0294 & 0.0032 & 0.9988 & 0.0014 & 0.0122 \\ 1.0014 & 0.0012 & -0.0242 & -0.0231 & 0.0372 & -0.0315 \\ 0.0248 & -0.0156 & -1.4725 & -0.0025 & 0.0048 & -0.0126 \\ -0.0020 & 1.2719 & 0.0187 & -0.0287 & 0.0109 & -0.0056 \end{bmatrix}$$

从性能矩阵可见,每行每列仅有1个元素远大于其它元素,这表明算法较好地恢复了源信号,只是分离出的源信号的幅度、位置、相位有变化。但是,这并不影响分离的信息,负号表示反相180°。分离信号与信源信号的对应关系是: $y_1(n) \leftrightarrow s_6(n)$ ,  $y_2(n) \leftrightarrow s_5(n)$ ,  $y_3(n) \leftrightarrow s_4(n)$ ,  $y_4(n) \leftrightarrow s_1(n)$ ,  $y_5(n) \leftrightarrow s_3(n)$ ,  $y_6(n) \leftrightarrow s_2(n)$ 。

从表1可见,混合信号的峭度趋向于0,根据大数定理原理,混合信号更接近高斯分布。分离信号的峭度非常接近信源信号的峭度,同样说明信源恢复较好

表1 信源信号、混合信号和分离信号的峭度对比

信源信号	$s_1(n)$	$s_2(n)$	$s_3(n)$	$s_4(n)$	$s_5(n)$	$s_6(n)$
峭度	-1.9742	-1.5000	-1.2267	-2.000	-1.5000	-1.0453
混合信号	$x_1(n)$	$x_2(n)$	$x_3(n)$	$x_4(n)$	$x_5(n)$	$x_6(n)$
峭度	-0.3955	-0.5013	-0.6089	-0.3024	-0.2664	-0.5077
分离信号	$y_1(n)$	$y_2(n)$	$y_3(n)$	$y_4(n)$	$y_5(n)$	$y_6(n)$
峭度	-0.9951	-1.4654	-1.9752	-1.9408	-1.1973	-1.4953

#### 4.2 自然语音信号的分离

该实验的源信号是5个说话者(有男声有女声,有中文也有英文)的语音,时间4s,采样率8kHz,32000个数据点,训练重复2次(4×2=8s)。所有语音信号预先处理为0均值单位方差。混合矩阵  $A$  在[-1, +1]之间均匀随机生成。盲分离性能选用交叉谈话误差来表示<sup>[7]</sup>。

$$E = \sum_{i=1}^N \left( \sum_{j=1}^N \frac{|P_{ij}|}{\max_k |P_{ik}|} - 1 \right) + \sum_{j=1}^N \left( \sum_{i=1}^N \frac{|P_{ij}|}{\max_k |P_{kj}|} - 1 \right) \quad (15)$$

这里  $P = (p_{ij}) = WA$ 。显然,误差越小性能越好。语音信号的峭度大于0,活动函数选择  $\phi(u) = \tanh(u)$ 。图3显示了模拟结果的性能曲线,性能改善30dB。

本实验的性能矩阵为

$$P = WA = \begin{bmatrix} \underline{1.5055} & -0.0018 & -0.0197 & -0.0130 & 0.0679 \\ 0.0102 & -0.0099 & -0.0113 & \underline{0.7933} & -0.0357 \\ -0.5010 & \underline{-0.8034} & 0.0014 & 0.0030 & 0.0146 \\ -0.0666 & 0.0017 & 0.0058 & -0.0165 & \underline{1.1958} \\ 0.0130 & 0.0058 & \underline{-0.7453} & 0.0186 & 0.0319 \end{bmatrix}$$

可见每行每列仅有 1 个元素远大于其它元素,表明算法分离效果较理想。如果减小学习速率,误差会更小,但是收敛时间变长。

## 5 结论

盲信源分离是近年来信号处理领域研究的热点,在许多研究领域具有广泛潜在的应用前景。本文针对通信信号和自然语音信号的盲源分离问题进行了深入的研究。计算机模拟结果显示,自然梯度算法采用不同的活动函数能够有效地分离随机混合的通信信号或语音信号。

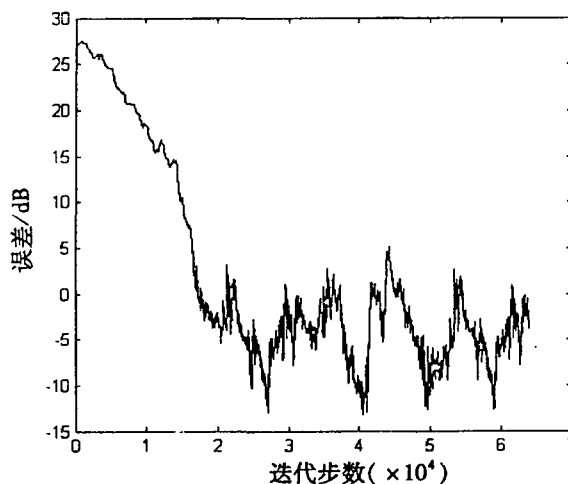


图3 交叉谈话误差曲线

### 参考文献:

- [1] Herault J, Jutten C. Space or time adaptive signal processing by neural network models[A]. in Neural Networks for Computing; Denker J S. AIP Conference Processings 151[C]. New York: American Institute of Physics, 1986.
- [2] Bell A J, Sejnowski T J. An information - maximization approach for blind separation and blind deconvolution[J]. Neural Comput., 1995, 7: 1126 - 1159.
- [3] Papoulis A. Probability, random variables, and stochastic processes[M]. New York: McGraw - Hill, 1991.
- [4] Simon Haykin Neural Networks; a comprehensive foundation[M]. Tsinghua University Press, 1999.
- [5] Cichocki A, Unbehauen R, Moczczynski L, et al. A new online adaptive learning algorithm for blind separation of source signals [A], in Proc. ISANN[C]. Taiwan, 1994. 406 - 411.
- [6] Amari S, Cichocki A, Yang H H. A new learning algorithm for blind signal separation[A]. in Advances in neural information processing systems[C]. Cambridge, 1996. 757 - 763.
- [7] Jacob Benesty. An introduction to blind source separation of speech signals[A]. Acoustic signal processing for telecommunication [C]. Boston: Kluwer Academic, 2000. 321 - 329.

(编辑: 门向生)

## Research on Blind Source Separation Based on Natural Gradient Algorithm

SUN Shou - yu<sup>1</sup>, ZHENG Jun - li<sup>1</sup>, WU De - wei<sup>2</sup>

(1. Department of Electronics Engineering, Tsinghua University, Beijing 100084, China; 2. The Telecommunication Engineering Institute, Air Force Engineering University, Xi'an, Shaanxi 710077, China)

**Abstract:** Blind source separation attempts to recover unknown independent sources from a given set of observed mixtures. The natural gradient algorithm is introduced in this paper, and it is a very important approach for blind source separation. We have examined the algorithm with communication signals and natural speech signals by different activation functions respectively. Simulation results demonstrate the algorithm can effectively separate the two kinds of random mixing signals.

**Key words:** blind source separation (BSS); natural gradient algorithm; activation function