

一种改进的基于 EASI 的语音分离算法

孟东霞, 马建芬, 乔永凤

MENG Dong-xia, MA Jian-fen, QIAO Yong-feng

1. 太原理工大学 理学院, 太原 030024

2. 太原理工大学 计算机与软件学院, 太原 030024

1. College of Sciences, Taiyuan University of Technology, Taiyuan 030024, China

2. Institute of Computer and Software, Taiyuan University of Technology, Taiyuan 030024, China

E-mail: mdx2002@163.com

MENG Dong-xia, MA Jian-fen, QIAO Yong-feng. Improved speech separation algorithm based on EASI. Computer Engineering and Applications, 2007, 43(33): 214-216.

Abstract: Independent component analysis is a kind of widely used signal processing method developed in recent years because it does not need to obtain the prior knowledge of signals. In this paper we first have introduced the basic method of the ICA and EASI algorithm. Then we have provided an improved one used in speech separation based on neural network.

Key words: speech separation; independent component analysis; EASI algorithm; neural network

摘要: 独立分量分析是近年来发展起来的一门新的数字信号处理方法, 因为不需要知道信号的先验信息而得到广泛应用。论文简单介绍了 ICA 的原理及 EASI 算法, 并根据神经网络理论提出一种改进的 EASI 语音分离算法。

关键词: 语音分离; 独立分量分析; EASI 算法; 神经网络

文章编号: 1002-8331(2007)33-0214-03 **文献标识码:** A **中图分类号:** TN911.7

1 引言

独立分量分析(Independent Component Analysis, ICA)是近年来发展起来的一种新的信号分离技术, 是由 Herault 和 Jutten 在 1983 年首次提出, 并在 1991 年提出了基于神经网络的 H-J 算法^[1]。1994 年, Comon 首次提出了 ICA 的概念^[2], 其基本思想是从多个源信号的混合中将原始信号分离出来。除了假设源信号是相互独立统计外, 没有任何其他先验知识。ICA 是伴随着盲信号问题而发展起来的, 故又称为盲分离(Blind Source Separation, BSS)。ICA 在通信、图像处理、生理医学信号处理、语音信号处理、特征提取等领域有着广泛的应用。

2 ICA 的定义及数学模型

设有 m 个信号源和传感器, 观测信号和源信号之间有如下关系式:

$$X(t) = AS(t) + n(t) \quad (1)$$

其中观测信号 $X(t) = [x_1(t), x_2(t), \dots, x_m(t)]^T$, 它是 n 个未知源信号 $S(t) = [s_1(t), s_2(t), \dots, s_n(t)]^T$ 的瞬时线性混迭, $s_i(t)$, $i=1, 2, \dots, n$, 为 n 个相互独立的随机信号, 混迭矩阵为 $A = (a_{ij})$, $i=1, 2, \dots, m$, $j=1, 2, \dots, n$ 上式中 $n(t)$ 为加性观测噪声。源分离问题就是要估计系统 W , 使 $X(t)$ 通过 W 时的输出为 $Y(t)$, 即有

$$Y(t) = WX(t) = WAS(t) = PDS(t) \quad (2)$$

式中, P 为置换矩阵, 即矩阵的每一行和每一列有且仅有一个

元素为非零, D 为任意非奇异对角阵, $Y(t)$ 是 $X(t)$ 的一个估计, W 是输入层与输出层之间的权系数。对这样一个模型, 如果没有关于 A 和源信号的任何先验信息, 仅从观测信号 $X(t)$ 恢复出 $S(t)$ 是不可能的, 一般对模型作如下假设:

- (1) $m \geq n$, 即观测变量的个数大于或等于信号独立源的个数; 且矩阵 A 为满秩的, 即 $\det A \neq 0$;
- (2) 信号源的各个分量之间相互统计独立;
- (3) 至多有一个独立源满足高斯分布;
- (4) 没有噪声或者噪声能量很小。

3 EASI 算法

一般来讲, 盲分离算法是基于“正交化对比函数”标准的, 选择不同的对比函数, 可以得到不同的独立分量分析算法。ICA 的算法很多, 这里介绍一种典型的独立分量分析算法: 等变量自适应分离算法(EASI)。

基于信息几何理论的 EASI 算法采用自然梯度标准, 算法的收敛特性不依赖于 $R_{xx}(t) = R_{xx}$ 的特征值, 具有优于普通梯度标准的收敛性能。在该算法中, 采用对比函数

$$\varphi(B) = \sum_{i=1}^n E[y_i^4] \quad (3)$$

定义 $\varphi(B)$ 是 $n \times m$ 维矩阵 B 的目标函数, 对 B 的各个元素是可微的。对比函数 $\varphi(B)$ 在点 B 的一阶展开为

基金项目: 山西省高校科技开发项目(No.197-12909083)。

作者简介: 孟东霞(1966-), 副教授, 主要研究方向: 计算机应用与语音信号处理; 马建芬(1967-), 女, 副教授, 主要研究方向: 语音信号处理及神经网络与小波理论。

$$\varphi(\mathbf{B}+\varepsilon)=\varphi(\mathbf{B})+\langle \frac{\partial \varphi}{\partial \mathbf{B}}(\mathbf{B})|\varepsilon \rangle+o(\varepsilon) \quad (4)$$

其中 $\nabla \varphi(\mathbf{B})=\frac{\partial \varphi}{\partial \mathbf{B}}(\mathbf{B})\mathbf{B}^T$ 称为相对梯度。与一般的随机梯度不同, 它计算的是矩阵的相对变化量, 所以相对梯度能更好地满足连续迭代的要求。

和所有的梯度准则一样, “相对梯度准则”就是在与相对梯度相反的方向上调整“相对偏差” ε 。换句话说, 也就是将 \mathbf{B} 调整到 $\mathbf{B}+\varepsilon\mathbf{B}$, 其中, $\varepsilon=-\lambda \nabla \varphi(\mathbf{B})$, λ 是一个正的尺度因子。

$$\begin{aligned} \varphi(\mathbf{B}-\lambda \nabla \varphi(\mathbf{B})) &= \varphi(\mathbf{B})+\langle \nabla \varphi(\mathbf{B})|-\lambda \nabla \varphi(\mathbf{B}) \rangle+ \\ o(\lambda \nabla \varphi(\mathbf{B})) &= \varphi(\mathbf{B})-\lambda \|\nabla \varphi(\mathbf{B})\|_{Fro}^2+o(\lambda) \end{aligned} \quad (5)$$

通过消除相对梯度中的期望值算子得到式(6):

$$\mathbf{B}_{k+1}=\mathbf{B}_k-\eta(k) f'(Y_k(t)) Y_k^T(t) \mathbf{B}_k \quad (6)$$

上式所表示的是一个连续迭代的算法。

EASI 以自然梯度下降为标准, 文献[4]中给出了等变量自适应分离算法(EASI)迭代公式:

$$\mathbf{B}(t+1)=\mathbf{B}(t)+\eta(t)\{\mathbf{I}-\phi(y(t))y^T(t)+y(t)\phi^T(y(t))-y(t)y^T(t)\}\mathbf{B}(t) \quad (7)$$

实验中采用两路自行录制的语音信号 $s_1(t), s_2(t)$, (数据采样率为 22 050 Hz, 长度 30 000), 图 1(a)为源信号的波形图。经

矩阵 $\mathbf{A}=\begin{bmatrix} 0.4 & 0.5 \\ 0.6 & 0.3 \end{bmatrix}$ 混合后得到混迭信号为 $x_1(t), x_2(t)$, 波形如图 1(b)所示; \mathbf{A} 为 [0, 1] 之间均匀分布的随机数, 学习步长 $\eta(t)=0.0001$, 选取目标函数 $\varphi(y)=2 \tanh(y)$, 随机选取权系数的初值, 这里取 $\mathbf{B}_0=\mathbf{I}$, 利用规则 (7) 训练后得到分离矩阵 $\mathbf{B}=\begin{bmatrix} 10.5556 & -7.0736 \\ -8.0527 & 13.4887 \end{bmatrix}$, 并由此分离得到 $y_1(t), y_2(t)$ 。计算可得

$\mathbf{C}=\mathbf{B}\mathbf{A}=\begin{bmatrix} 0.0219 & 3.1557 \\ 4.8721 & 0.0203 \end{bmatrix}$, 它是一广义置换矩阵, 源信号较好的得以分离。图 2 画出了 $\mathbf{C}=\mathbf{B}\mathbf{A}$ 的迭代曲线。图 1(c)为分离输出的信号。

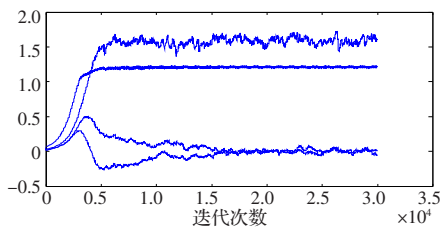


图 2 广义矩阵 $\mathbf{C}=\mathbf{B}\mathbf{A}$ 的迭代曲线

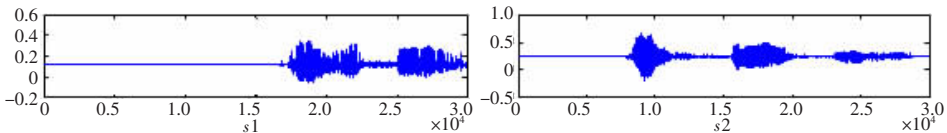


图 1(a) 源信号

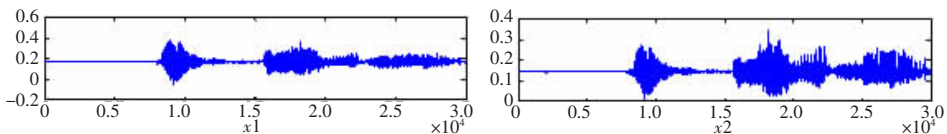


图 1(b) 观测信号

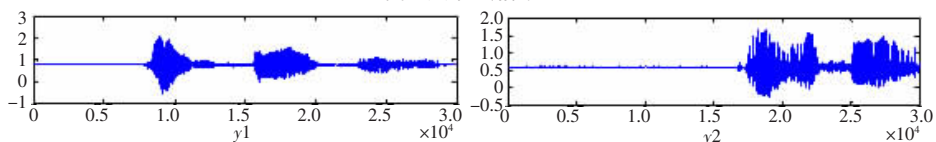


图 1(c) 分离信号

从试验结果可以看出, 分离语音信号从波形上看很好地保持了原始信号的波, 而且实际分离的信号在听觉上也很好的实现了分离, 几乎没有什么失真。用信噪比作为性能标准衡量算法的分离效果。语音的信噪比提高值为分别为:

表 1 分离后信噪比的改善程度 dB

SNR _{in1}	SNR _{out1}	提高	SNR _{in2}	SNR _{out2}	提高
-14.70	-10.28	4.42	-2.95	2.87	5.82

4 改进的基于神经网络的语音盲分离算法

把经过用 EASI 算法分离出来的信号再经神经网络进行优化, 优化函数为 ψ_i 。函数 ψ_i 用来估计输出分量 y_i 的累积概率函数(CPF)(Cumulative Probability Function)。输出为 $Z_i=\psi_i(y_i)$ 。 ψ_i 是在 [0, 1] 之间单调递增的, 且为非线性可逆, 则 Z_i 是 [0, 1] 之间的均匀分布。这使得 Y 和 Z 的互信息相等, 即 $I(Y)=I(Z)$, 且 $H(Z_i)=0$, 因此有

$$I(Y)=I(Z)=\sum_i H(Z_i)-H(Z)=-H(Z)$$

输出熵 $H(Z)$ 的最大化导致估计分量的互信息量最小。

整个网络以输出熵最大为训练目的, 首先由 Infomax 算法, 有:

$$H(\mathbf{Z})=H(\mathbf{X})+E[\log|\det \mathbf{J}|]$$

$\mathbf{J}=\partial \mathbf{Z} / \partial \mathbf{X}$ 是经由网络传递的雅各比行列式, $H(\mathbf{X})$ 不随网络参数而变, 所以只需要最大化 $E[\log|\det \mathbf{J}|]$ 。因此, 目标函数选为:

$$V=\frac{1}{K} \sum_{k=1}^K \log|\mathbf{J}^k| \approx E[\log|\det \mathbf{J}|]$$

K 为训练次数, \mathbf{J}^k 是第 K 次训练后得到的雅各比行列式。

$$\frac{\partial V}{\partial \mathbf{J}}=(\mathbf{J}^{-1})^T$$

T 代表矩阵转置。 \mathbf{J} 的计算如图 3 所示^[2,3]:

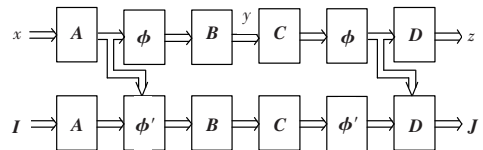
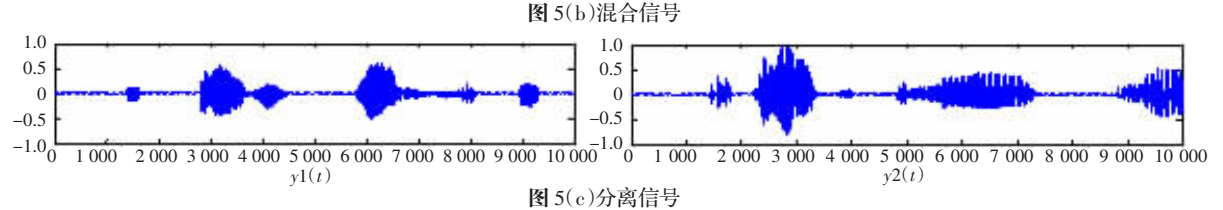
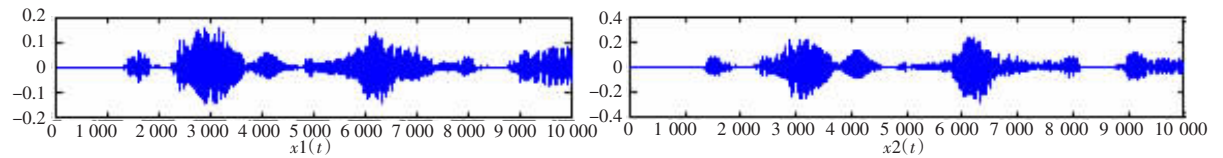
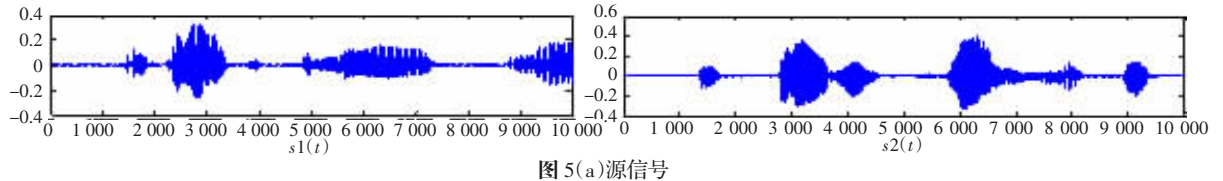
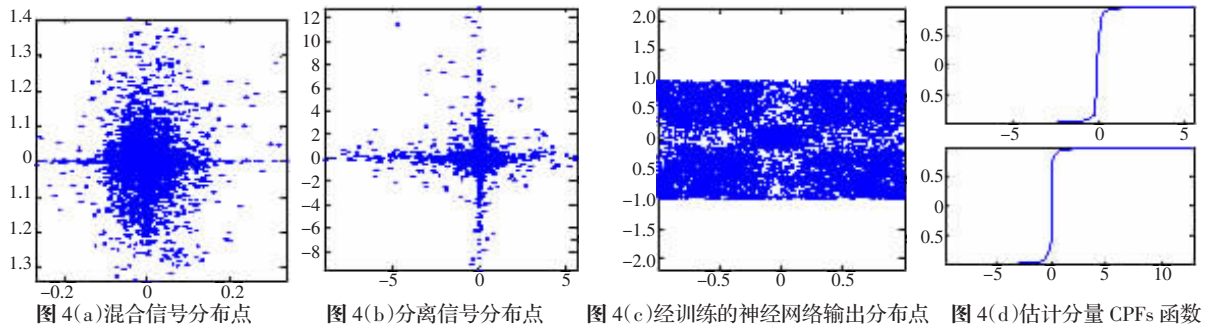


图 3 雅各比行列式计算框图

在图 3 中, 上部分是 Infomax 算法的另一种画法。模块 A, 左 ϕ, B 完成分离信号的功能, 输出为 y 。 ϕ 是 s 形目标函数。模块 C, 右 ϕ, D 完成网络优化的功能, 输出为 Z 。



下部分是雅各比行列式的计算。输入 I 为单位矩阵, ϕ' 是一一对角阵, ϕ' 是 ϕ 的导数, 输出为雅各比行列式 J :

$$J = D\phi_R' CB\phi_L'A$$

5 试验结果

在试验中, MLP 有 10 个隐层, s 型函数为 $g(x) = a \tan(x)$, ϕ 输出区间为 $[-1, 1]$, 图 4 是经神经网络训练好的两个超高斯混合信号、分离信号和神经网络输出的分布点以及估计分量的 CPFs 函数。可以看出, 训练好的神经网络对函数的 CPFs 的分布范围均在 $[-1, 1]$ 之间。

经过训练的网络来分离混合的语音信号, 如图 5 所示。

用 RBF 和 MLP 两种神经网络来分离超高斯信号, 结果如表 2 所示。

表 2 不同神经网络下分离信号的改进效果

信号	MLP 网络优化			RBF 网络优化		
	输入 SNR	输出 SNR	提高	输入 SNR	输出 SNR	提高
1	9.6	13.5	3.9	9.6	14.8	5.2
2	6.5	11.2	4.7	6.5	12.3	6.8

从两种神经网络对信号的分离效果比较看, 分离效果都非常理想, 平均信噪比提高分别为 4.3 dB 和 6 dB, RBF 网络要比 MLP 网络更好一些。

6 结论

改进的 EASI 算法可以用来解决线性和非线性混合信号分离问题, 其优点是:

- (1) 用同一网络完成盲信号分离和分量统计分布的估计;
- (2) 使用同一目标函数。

这种带有前馈神经网络优化结构, 并用输出熵作为目标函数的 ICA 算法能够很好地提取信号的独立分量, 完成混合信号的分离。试验结果还表明, 对相对平稳混合信号, 分离效果较理想, 但对非线性强混合信号分离还存在一定困难。

(收稿日期: 2007 年 5 月)

参考文献:

- [1] Jutten C, Herault J. Blind separation of source, part I: an adaptive algorithm based on neuromimetic architecture [J]. Signal Processing, 1991, 24(1): 1-10.
- [2] Comon P. Independent component analysis, a new concept? [J]. Signal Processing, 1994, 36: 287-314.
- [3] Cardoso J F, Laheld B. Equivariant adaptive source separation [J]. IEEE Trans Signal Processing, 1996, 44: 3017-3029.
- [4] Comon P. Independent component analysis, a new concept [J]. Signal Processing, 1994, 36(3): 287-314.
- [5] Bell A, Sejnowski T. An information maximization approach to blind separation and blind deconvolution [J]. Neural Computation, 1995, 7(6): 1129-1159.
- [6] 刘璐, 聂开宝, 何振亚. 非线性混叠信号的可分离性及分离方法研究 [J]. 电子与信息学报, 2003, 25(1): 54-61.
- [7] Almeida L B. MISEP - an ICA method for linear and nonlinear mixtures, based on mutual information [C] // Proc 2002 Int Joint Conf on Neural Networks, Honolulu, Hawaii, 2002.
- [8] Almeida L B. MISEP - linear and nonlinear ICA based on mutual information [J]. Journal of Machine Learning Research, 2003, 4: 1297-1318.
- [9] Almeida L B. Linear and nonlinear ICA based on mutual information the MISEP method [J]. Signal Processing, 2004, 84(2): 231-245.