

峭度自适应学习率的盲信源分离

孙守宇, 郑君里, 吴里江, 赵 莹

(1. 清华大学电子工程系, 北京 100084; 2. 北京信息技术研究所, 北京 100088)

摘要: 本文提出了一种自适应学习率盲信源分离的自然梯度算法, 自适应学习率仅依赖于神经网络输出峭度平方和的负指数. 开始阶段由于小的峭度, 学习率大收敛速度快. 之后, 随着峭度变大, 学习率慢慢变小, 产生小的稳态误差. 在线性无记忆混合的情况下, 用欠高斯信源进行的模拟实验表明, 与固定学习率相比, 本文提出的峭度自适应学习率盲信源分离算法具有收敛速度快和稳态误差小的特点.

关键词: 自适应; 学习率; 盲信源分离; 峭度

中图分类号: TN911 **文献标识码:** A **文章编号:** 0372-2112 (2005) 03-0473-04

Blind Source Separation with Kurtosis Adaptive Learning Rate

SUN Shouyu, ZHENG Junli, WU Lijiang, ZHAO Ying

(1. Department of Electronic Engineering, Tsinghua University, Beijing 100084, China;

2. Beijing Institute of Information Technology, Beijing 100088, China)

Abstract: A natural gradient algorithm with an adaptive learning rate for blind source separation is proposed. The adaptive learning rate only depends on the negative exponential of the square sum of the kurtosis of neural networks outputs. The initial convergence speed is very fast because of large learning rate (according to small kurtosis). After an initial period, the learning rate decreases slowly due to large kurtosis, giving rise to small steady state errors. Simulation with under Gaussian as sources and in the instantaneous mixing case shows that the proposed algorithm has faster convergence and smaller steady state error than those without adaptive learning rate.

Key words: adaptive; learning rate; blind source separation; kurtosis

1 引言

盲信源分离(BSS) 也称作独立元分析(ICA), 被看作主元分析(PCA) 的扩展, 其差别在于, PCA 只利用了观察数据的二阶统计量且要求各输出分量正交, 而 ICA 利用了观察数据的高阶统计量且要求各输出分量统计独立. 对于高斯分布的信号, 二阶统计量足以描述其特性, 但是对于常见的通信信号、图象信号、语音信号, 其分布通常是非高斯的, 二阶统计量不足以描述其特性, 必须利用高阶统计量.

近年来, 有关盲信源分离和独立元分析研究的文献很多, 值得注意的方法^[1~9]有: 最大负熵法、高阶累计量法、非线性 PCA、最大似然估计法、贝叶斯 ICA、最大后验参数估计法、信息几何法、最小互信息法(MMI). 这些盲信源分离算法基本上可以归类为 LMS 型算法, 都存在一个学习率(learning rate) 参数的优选问题, 如何提高算法的收敛速度和改进算法的稳态性能一直是盲信源分离研究的热点问题之一.

当学习率(或步长) 取常数时, LMS 型算法存在收敛速度和稳态性能之间的矛盾^[13]: 步长小, 算法的稳态性能(即源信

号的恢复质量) 好, 但算法收敛慢; 反之, 则收敛快, 但稳态性能差. 解决上述矛盾的最简单做法是令学习速率随时间递减, 但这又会产生新的问题, 如果信号没有完全分离, 学习率降至很小时, 未分离的信号会因学习率太小而长时间甚至永远得不到分离. 解决上述矛盾的另一类做法是采用自适应步长参数, 如基于辅助变量的自适应步长算法和梯度自适应步长算法等, 这些算法的自适应步长的选择取决于辅助变量, 而与神经网络输出之间的分离状态无关. 能否将自适应学习率与信号分离状态联系起来呢? 文献[10] 在这方面进行了尝试, 提出了一种具有信号自适应学习率的分阶段学习算法. 算法在学习过程中, 分为三个阶段, 在不同阶段取不同的学习率.

注意到信号在分离过程中自组织神经网络输出的互信息逐步下降, 而互信息与信号的高阶统计量(峭度) 有关. 根据这一思想, 本文提出了峭度自适应学习率的盲信源分离算法.

2 盲信源分离

盲信源分离试图从一组混合观察数据中恢复未知的独立源信号. 这里之所以称作“盲”, 是因为观察信号是恢复原始信

号的唯一可利用信息,对于源信号和这些信号的混合过程一概不知.盲信源分离的目标是确定一稳定的神经网络逆系统(也称重构系统),估计原始输入信号.神经网络逆系统必须具有自组织特性,以便能够在非平稳环境下也具有很好的跟踪能力.下面用数学公式描述该问题.

假设有一组彼此相互独立的未知信源信号矢量 $s(k) = [s_1(k) s_2(k) \dots s_N(k)]^T$, 以未知方式线性混合产生观察矢量 $x(k) = [x_1(k) x_2(k) \dots x_N(k)]$, 在不考虑噪声的情况下, 它们的关系为

$$x(k) = As(k) \quad (1)$$

其中, A 是一个 $N \times N$ 的未知非奇异混合矩阵. 关于信源作如下假设:

- (1) 每个信源信号 $s_j (1 \leq j \leq N)$ 是独立同分布的随机过程;
- (2) 信号 s_j 在任何时刻是统计独立的, 它们的联合概率密度等于各信号概率密度的乘积;
- (3) 在所有信源信号中, 最多只能有一个高斯概率分布的信号.

由于独立信源信号 $s(k)$ 经过线性混合式(1)后, 观察矢量 $x(k)$ 不再独立. 为了达到分离的目的, 观察矢量 $x(k)$ 经过神经网络产生输出信号矢量 $y(k) = [y_1(k) y_2(k) \dots, y_N(k)]^T$

$$y(k) = Wx(k) \quad (2)$$

其中, W 是一个 $N \times N$ 的权矩阵. 神经网络通过学习调整权矩阵 W , 使输出 $y(k)$ 变得尽可能独立, 即它们的联合概率密度等于各信号概率密度的乘积

$$p_y(y) = p_{y_1}(y_1) p_{y_2}(y_2) \dots p_{y_N}(y_N) \quad (3)$$

一个普遍使用的有效方法是定义式(3)两边概率分布的失配程度(即 Kullback Leibler 分歧度)作为代价函数:

$$D_{KL}(w) = D\left(p_y(y) \parallel \prod_{i=1}^N p_i(y_i)\right) = \int p_y(y) \log \frac{p_y(y)}{\prod_{i=1}^N p_i(y_i)} dy \quad (4)$$

对 $D_{KL}(W)$ 关于 W 求导数, 得

$$\frac{\partial D_{KL}(W)}{\partial W} = (W^{-1})^T - \Phi(y) x^T \quad (5)$$

因此, 基于统计梯度的自适应算法为

$$W(n+1) = W(n) + \mu \{ (W(n))^{-1} - \Phi[y(n)] x^T(n) \} \quad (6)$$

其中, μ 为学习率. 但是, 该算法受混合矩阵 A 的影响较大^[8], 这是普通梯度算法进行盲信源分离时的一个缺点. Crchck^[9] 和 Amar^[12] 等人在式(5)后乘矩阵 $W^T W$ 提出了自然梯度算法

$$W(n+1) = W(n) + \mu \{ I - \Phi[y(n)] y^T(n) \} W(n) \quad (7)$$

其中, $\Phi[y(n)] = \begin{bmatrix} -\frac{p'_{y_1}(y_1)}{p_{y_1}(y_1)} - \frac{p'_{y_2}(y_2)}{p_{y_2}(y_2)} \dots - \frac{p'_{y_N}(y_N)}{p_{y_N}(y_N)} \end{bmatrix}^T$ 称

作活动函数(activation function). 自然梯度算法几乎不受混合矩阵 A 的影响, 从而使算法更有效. 当算法收敛后, 输出近似独立, 互信息最小, 且逼近 0.

3 峭度自适应学习率的盲信源分离算法

实际应用中, 几乎所有盲信源分离或 ICA 的文献都用“串

音”(cross talk) 误差来评价分离算法的性能

$$E_{\alpha} = \sum_{i=1}^N \left(\sum_{j=1}^N \frac{|p_{ij}|}{\max_k |p_{ik}|} - 1 \right) + \sum_{j=1}^N \left(\sum_{i=1}^N \frac{|p_{ij}|}{\max_k |p_{kj}|} - 1 \right) \quad (8)$$

这里 $P = (p_{ij}) = WA$ 表示整个混合分离合成系统的传递矩阵, 该性能曲线随迭代次数呈指数趋势下降. 用串音误差 E_{α} 控制自适应学习率 μ , 能够反映信号的分离程度. 信号分离的初始阶段, 串音误差 E_{α} 较大, 相应地学习速率较大, 信号分离的后期阶段, 串音误差 E_{α} 变小, 相应地学习速率较小. 这样兼顾了算法的收敛速度和信号分离的质量. 但是, 实际应用中混合矩阵 A 是未知的, 无法计算每次迭代的串音误差 E_{α} . 能否有其它参数既能反映信号分离状态又便于参数估计的物理量呢? 分离信号之间的互信息是一个很好的选择. 分离开始阶段分离信号间的互信息较大, 这时适用较大的学习率, 随着学习进展, 互信息逐渐减小, 学习率跟着变小, 直到完全分离时, 互信息为 0. 对于去除二阶相关的数据, 输出互信息等于其负熵减去边缘负熵的代数和^[2]:

$$I(y) = J(y) - \sum_{i=1}^N J_i(y_i) \quad (9)$$

其中, $I(y)$ 是 y 的互信息, $J(y)$ 是 y 的负熵, $J_i(y_i) (1 \leq i \leq N)$ 是 y 的边缘负熵. 显然, 边缘负熵和最大, 互信息最小. 边缘负熵可以计算, Comon^[2] 证明多变量负熵由下式近似

$$J_i(y_i) \cong \frac{1}{12} k_3^2 + \frac{1}{48} k_4^2 + \frac{7}{48} k_3^4 - \frac{1}{8} k_3^2 k_4 \quad (10)$$

其中, k_3 表示信号的三阶累积量, k_4 表示四阶累积量. 假设信号的概率分布是对称的, 那么 $k_3 = 0$, 于是

$$J_i(y_i) \cong \frac{1}{48} k_4^2 \quad (11)$$

将式(11)代入式(9), 得

$$I(y) = J(y) - \sum_{i=1}^N \frac{1}{48} k_4^2(i) \quad (12)$$

其中, $k_4^2(i)$ 表示神经网络输出矢量 y 的第 i 个分量的四阶累积量的平方. 四阶边缘累积量 $k_4(i)$ 的归一化计算公式

$$k_4(i) = \frac{E\{y_i^4\}}{(E\{y_i^2\})^2} - 3 \quad (13)$$

称作信号的峭度(kurtosis), 用来度量信号偏离高斯信号的程度. 高斯信号的峭度等于 0, 欠高斯信号的峭度小于 0, 超高斯信号的峭度大于 0, 但是峭度的平方 $k_4^2(i)$ 非负. 因为在信号分离过程中计算互信息是困难的, 所以可以通过计算分离信号边缘负熵总和的方法代替互信息来控制学习率自适应变化. 如果记边缘负熵的总和

$$J_{all}(k) = \sum_{i=1}^N \frac{1}{48} k_4^2(i) \quad (14)$$

那么, 学习率

$$\eta_k = \eta_0 e^{-\alpha J_{all}(k)} \quad (15)$$

其中, η_0, α 是经验常数.

因此, 得到本文提出的峭度自适应学习率的盲信源分离算法

$$W(k+1) = W(k) + \eta_k \{ I - \Phi[y(k)] y^T(k) \} W(k) \quad (16)$$

其中, $\eta_k = \eta_0 e^{-\alpha J_{ad}(k)}$, $J_{ad}(k) = \sum_{i=1}^N \frac{1}{48} k_4^i(n)$. 在线(online)盲分离时,峭度可以由下式自适应更新

$$k_4^i(n+1) = k_4^i(n) - \mu_0(k_4^i(n) - y_i^4(n) + 3y_i^2(n)) \quad (17)$$

其中, $k_4^i(n)$ 表示第 i 个分量在第 n 步迭代时的四阶累积量, μ_0 是峭度更新自适应步长.

4 计算机仿真

为了验证上节提出的峭度自适应学习率的盲信源分离法的有效性,对以下 8 个欠高斯信号进行计算机仿真:

$s_1(t)$ 随机二进码 (-1 或 +1, 码元内有 10 个样点);

$s_2(t) = \sin(2\pi 1000t)$, (高频信号);

$s_3(t) =$ 频移键控 (2FSK) 信号 ($f_1 = 400\text{Hz}, f_2 = 800\text{Hz}$);

$s_4(t) = \text{sign}(\sin(2\pi 280t))$, (方波信号);

$s_5(t) = \cos(2\pi 48t)$, (低频信号);

$s_6(t) = \sin(2\pi 900t) \cos(2\pi 40t)$, (调幅信号);

$s_7(t) =$ 区间 [-1, +1] 内均匀分布的噪声;

$s_8(t) =$ 二元移相键控 (BPSK) 信号 ($f = 600\text{Hz}$).

仿真中,混合矩阵 A 的各元素在区间 [-1, 1] 随机生成,信号采样率为 12kHz,活动函数选择简单的 $\Phi(y) = y^3$. 作为比较,同时运行取固定常数的学习率和峭度自适应的学习率,仿真结果如图 1 所示.对于固定学习率,当取较大学习率 0.01 时,算法收敛快但稳态误差较大;当取较小学习率 0.001 时,算法收敛慢但稳态误差较小.而峭度自适应学习率,具有收敛速度快和稳态误差最小的优点.峭度自适应仿真时的其它参数: $\eta_0 = 0.01, \alpha = 8, \mu_0 = 0.001, W(0) = 0.1I$. 另外,还可以看出固定学习率,算法达到稳态时学习曲线波动较大,而自适应学习率的曲线较平稳.

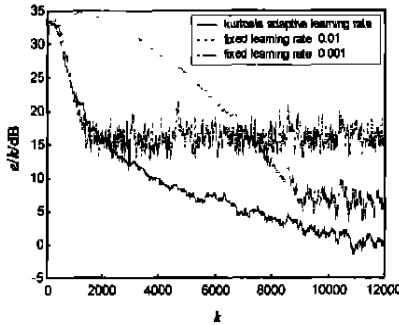


图 1 三种算法性能比较

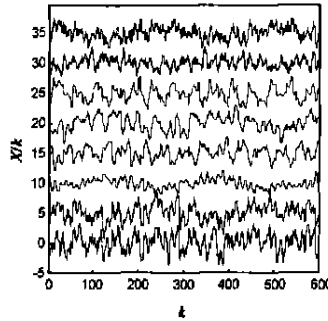


图 2 混合信号

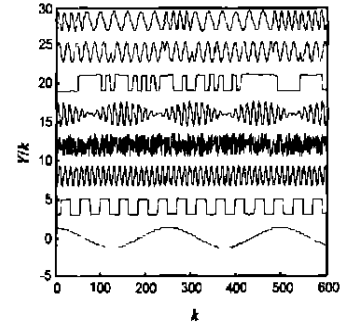


图 3 分离结果

图 2 是上述 8 个信源信号 ($s_1(t) \sim s_8(t)$) 由随机生成的混合矩阵 A 线性混合后的信号,利用本文提出的峭度自适应学习率算法,神经网络将混合信号的分离结果如图 3 所示.

表 1 显示了源信号,混合信号和分离信号的峭度,可见混合信号的峭度趋向于 0,符合大数定理原理,说明混合信号更接近高斯分布.

分离矩阵 W 与混合矩阵 A 的乘积为

WA =

-0.00228	0.0032	0.0084	0.0009	1.2780	0.0069	-0.0289	0.0057
-0.00244	-0.0020	-0.0022	-1.0005	0.0016	0.0109	-0.0103	-0.0159
-0.0030	-1.2747	0.0006	-0.0084	-0.0018	-0.0114	0.0183	-0.0001
0.0053	-0.0074	0.0288	0.0090	-0.0238	0.0033	1.4886	-0.0107
0.0062	-0.0152	0.0196	-0.0074	0.0045	1.6245	-0.0191	-0.0008
1.0047	0.0055	0.0042	0.0245	-0.0592	0.0037	-0.0019	0.0184
-0.0154	-0.0015	-0.0007	-0.0127	0.0019	-0.0013	0.0174	-1.2759
0.0056	-0.0010	1.2752	0.0040	0.0047	-0.0191	0.0207	-0.0003

可见,矩阵中每行每列仅有一个元素(黑体数字)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

表 1 信源信号、混合信号、分离信号的峭度

信源信号	$s_1(t)$	$s_2(t)$	$s_3(t)$	$s_4(t)$	$s_5(t)$	$s_6(t)$	$s_7(t)$	$s_8(t)$
峭度	-2	-1.5000	-1.5000	-2	-1.5000	-0.7500	-1.1923	-1.5000
观察信号	$x_1(t)$	$x_2(t)$	$x_3(t)$	$x_4(t)$	$x_5(t)$	$x_6(t)$	$x_7(t)$	$x_8(t)$
峭度	-0.4996	-0.2056	-0.5211	-0.5337	-0.3824	-0.4902	-0.5700	-0.7066
分离信号	$y_1(t)$	$y_2(t)$	$y_3(t)$	$y_4(t)$	$y_5(t)$	$y_6(t)$	$y_7(t)$	$y_8(t)$
峭度	-1.4982	-1.9988	-1.4969	-1.1925	-0.7489	-1.9922	-1.4920	-1.4933

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: $y_1(n) \leftrightarrow s_5(n), y_2(n) \leftrightarrow s_4(n), y_3(n) \leftrightarrow s_3(n), y_4(n) \leftrightarrow s_7(n), y_5(n) \leftrightarrow s_6(n), y_6(n) \leftrightarrow s_1(n), y_7(n) \leftrightarrow s_8(n), y_8(n) \leftrightarrow s_2(n)$.

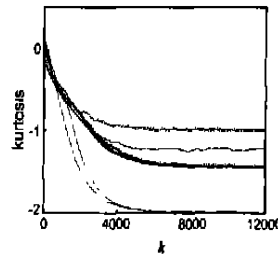


图 4 神经网络输出峭度的自适应收敛曲线

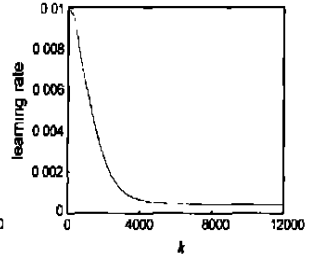


图 5 自适应学习率变化曲线

图 4 显示神经网络各输出峭度的自适应收敛曲线,图 5 显示了算法在仿真实验的迭代过程中,学习率随信号分离状态的自适应变化曲线.开始阶段学习率较大,随后逐渐变小而趋于稳定.

5 结论

盲信源分离是一个无监督自适应学习问题,自组织神经网络通过无监督学习获取数据特征.在盲信源分离

算法研究中,学习率起着重要作用.像 LMS 算法一样,增大学习率能够加快收敛速度,同时也增大了稳态误差.本文提出了一种自适应学习率盲信源分离的自然梯度算法,自适应学习率仅依赖于神经网络各输出峭度平方和的负指数.开始阶段由于小的峭度,学习率大收敛速度快.之后,随着峭度变大学习率慢慢变小,结果稳态误差较小,这样很好地克服了收敛速度和信号恢复质量之间的矛盾.仿真实验证实峭度自适应学习的有效性.

参考文献:

- [1] Girolami M, Fyfe C. Stochastic ICA contrast maximisation using oja's nonlinear PCA algorithm[J]. International Journal of Neural Systems, 1997, 8(5& 6): 661- 678.
- [2] Comon P. Independent component analysis, a new concept[J]. Signal Processing, 1994, 36(3): 287- 314.
- [3] Oja E. The nonlinear PCA learning rule and signal separation mathematical analysis[J]. Neurocomputing, 1997, 17: 25- 45.
- [4] Cardoso J F. Infomax and maximum likelihood for blind source separation[J]. IEEE Signal Processing Letters, 1997, 4(4): 109- 111.
- [5] Roberts S J. Independent component analysis: Source assessment and separation, a Bayesian approach[J]. IEE Proceedings Vision, Image and Signal Processing, 1998, 145 (3): 149- 154.
- [6] Parra L, Mueller K, Spence C, et al. Unmixing hyperspectral data[A]. Advances in Neural Information Processing Systems[C]. 2000, NIPS* 12: 942- 948.
- [7] Puntonet C, Prieto B. Neural net approach for blind separation of sources based on geometric properties[J]. Neurocomputing, 1998, 18(1 - 3): 141- 164.
- [8] Bell A, Sejnowski T. An information maximisation approach to blind separation and blind deconvolution[J]. Neural Computation, 1995, 7 (6): 1129- 1159.

- [9] Amari S, Chen T P, Cichocki A. Stability analysis of learning algorithms for blind source separation[J]. Neural Networks, 1997, 10(8): 1345- 1351.
- [10] 张贤达, 朱孝龙, 保铮. 基于分阶段学习的盲信号分离[J]. 中国科学(E 辑), 2002, 32(5): 493- 703.
Zhang Xianda, Zhu Xiaolong, Bao Zheng. Grading Learning for blind Source separation[J]. Science in China, Series E, 2002, 32(5): 693- 703.
- [11] Yang H H, Amari S-I. Adaptive online learning algorithms for blind separation: maximum entropy and minimum mutual information[J]. Neural Computation, 1997, 9(7): 1457- 1482.
- [12] A Cichocki, R Unbehauen, L Moszczynski, Rummert E. A new online adaptive learning algorithm for blind separation of source signals[A]. Proc. ISANN[C]. Taiwan: ISANN, 1994. 406- 411.
- [13] Motoaki N, Kawanabe M, Ziehe A, Muller K R, Amari S. On line learning in changing environments with applications in supervised and unsupervised learning[J]. Neural Networks, 2002, 15(4): 743- 760.

作者简介:

孙守宇 男, 1962 年 10 月生于江苏省沛县, 1984 年毕业于空军电讯工程学院, 留校任教, 分别于 1993 年和 2003 年在清华大学电子工程系获得硕士和博士学位, 现任总参某研究所高级工程师, 主要进行数字通信、人工神经网络和盲信号处理等领域的应用研究. E-mail: sunshouyu@tsinghua.org.cn.

郑君里 男, 1937 年生于天津市, 1961 年毕业于清华大学无线电工程系, 现任清华大学电子工程系教授、博士生导师, 主要研究方向: 通信信号处理、盲信号处理、人工神经网络、混沌理论及其应用. E-mail: zhenjunli@atm.mdc.tsinghua.edu.cn.

吴里江 男, 1950 年 10 月生于北京市, 1979 年毕业于清华大学无线电工程系, 现任北京信息技术研究所所长、高级工程师, 主要研究领域包括数字通信、信号处理和纳米技术等. E-mail: wulijiang@tsinghua.org.cn.