

基于递阶遗传算法的未知源信号个数盲信号分离

陈卫东¹, 舒柏晔²

(1. 天津大学 管理学院, 天津 300072; 2. 华泰证券股份有限公司, 南京 210004)

(cud8291@163.com; shubaixian@gmail.com)

摘要:为了解决在没有源信号个数信息条件下的盲信号分离问题,设计了一种基于递阶遗传算法的盲信号分离算法。用控制层、参数层分别表示分离矩阵有效行数和有效行数的具体数值,将非高斯性作为随机信号相互独立的度量,通过寻求观测信号的一个线性组合,使其峭度最大化或最小化来分离独立分量。计算机仿真验证了所提算法的有效性,对未知源信号个数盲信号处理有一定的指导意义。

关键词:未知源信号个数;递阶遗传算法;峭度;盲信号分离

中图分类号: TP18 **文献标志码:** A

Blind source separation with unknown number of sources based on hierarchical genetic algorithm

CHEN Wei-dong¹, SHU Bo-xian²

(1. School of Management, Tianjin University, Tianjin 300072, China;

2. Huatai Securities Company Limited, Nanjing Jiangsu 210004, China)

Abstract: In order to solve blind source separation without any information of the number of source signals, a new algorithm based on hierarchical genetic was proposed. Control layer and parameter layer represent the available number of rows and the concrete value of the available rows of separation matrix respectively. Non-Gaussian was used to measure the independence of the stochastic signals. The kurtosis was maximized/minimized by searching a linear combination of observation signals, and then the independent variables were separated. Computer simulation verifies the validity of the proposed algorithm.

Key words: unknown number of sources; hierarchical genetic algorithm; kurtosis; blind source separation

0 引言

近年来,盲信号分离已经在语音数字信号处理、多用户通信、阵列处理、生物医学、图像增强和雷达等众多领域中得到广泛的应用。盲信号分离问题的研究始于 B. Widr. W 在 1975 年提出的自适应噪声抵消器。其思想是用得到的有用信号,将噪声与有用信号分离,大大降低了噪声抵消的难度。文献[1]提出一种类神经盲信号分离方法,开创了盲信号分离的先河。该方法基于反馈神经网络,通过选取奇次非线性函数 Hebb 训练,达到分离的目的。文献[2]分析了瞬时混合信号盲分离问题,并明确了独立分量分析的概念。利用测量源信号统计独立性的 Kullbak-Leibler 准则作为对比函数(Contrast Function),通过对概率密度函数的高阶近似,得出用于测量信号各分量统计独立的对比函数。文献[3]基于信息理论,通过最大化输出非线性节点熵,得出一种最大信息传输的准则函数,导出一种自适应盲分离方法,用非线性函数选取逼近源信号的概率分布。该算法只能用于源信号峭度(kurtosis)为正时的信号盲分离(超高斯信号)。文献[4]基于信息理论中概率密度的 Gram-Charlier 展开,利用最小互信息量(Minimum Mutual Information MMI)准则函数,得出一类前馈网络的训练算法,分离具有负峰度的源信号。文献[5]基于源信号非高斯性度量(或峭度),给出一类定点训练算法(fixed point),该算法可以提取单个正或负峭度的源信号。文献[6]给出一种递归最小二乘(RLS)白化算法。文献[7]研究了 RBF 神经网络

的递阶遗传训练新方法。文献[8]研究了源信号数目未知或动态变化情况下的盲信号分离问题。文献[9]提出了基于分阶段学习的盲信号分离方法。虽然人们已经研究出多种信号盲分离的方法,但是每种方法都有一定的局限性,只适用于特定的情况。本文选择 $\sin(8\pi t)$ 、 $\text{sawtooth}(16\pi t)$ 、 $\text{square}(100\pi t)$ 三种规则盲源信号图像,用大系统递阶优化的思想,设计递阶遗传算法,以盲源信息峭度的绝对值和作为适值函数,计算机仿真验证了该方法的有效性。

1 递阶染色体设计

在遗传算法中,传统的编码方法如二进制编码、浮点数编码、符号数编码,是把问题的解用一串数字表示出来,只能表达一维结构。这里采用大系统递阶优化的思想,设计递阶染色体编码方法。递阶结构染色体包括控制基因(操纵基因)和参数基因(结构基因)两个部分。控制基因用二进制编码方式。参数基因根据不同的问题对象,采用不同的编码方式。在控制基因层用 1 表示下层基因被激活,0 表示下层基因处于休眠状态。见图 1。

与一般的编码方式相比较,递阶结构编码具有以下特点:

1) 能够将问题的共同特征抽象出来。如求解 0-1 规划和运输混合问题,可以用控制层来表示问题的约束条件,用参数层存储具体的运输或调度方案,节省存储空间和计算量。

2) 能够完整地表达待解决问题的解。如对 BP 网络的训练,可以将网络的拓扑结构用控制层表示,参数层存储每个有

收稿日期:2008-12-17;修回日期:2009-02-26。

作者简介:陈卫东(1967-),男,河南鹿邑人,副教授,博士,主要研究方向:系统分析与决策;舒柏晔(1982-),男,贵州晴隆人,硕士,主要研究方向:高频金融数据分析。

效神经元的权值和阈值信息,实现在一次训练中同时确定网络的拓扑结构、权值和阈值信息。

3)能够发挥遗传算法的自适应性。如一般的盲信号分离算法,需要在分离之前已知源信号的个数。对源信号个数未知条件下的盲信号的分离,采用递阶染色体对分离阵进行编码,分离信号的个数等于控制层基因中“1”的个数,分离过程中能够自动识别出最佳的分离个数。

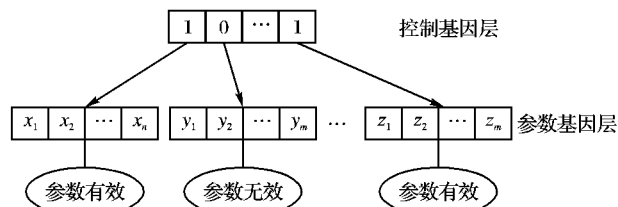


图1 递阶染色体结构示意图

2 盲源分离过程和分离目标函数

2.1 盲源分离过程

假设对源信号的混合为线性变换,那么可用数学模型表示如下:

$$\begin{cases} X(t) = MS(t) + N(t), & \text{混合过程} \\ Y(t) = AX(t), & \text{分离过程} \end{cases} \quad (1)$$

其中: $S(t)$ 为 $m \times 1$ 阶未知源信号向量, $X(t)$ 为 $n \times 1$ 阶已观测到的混合信号, $N(t)$ 为混合过程中接受到的噪声,其 n 个分量满足白色,高斯并且统计独立三个条件, M 为 $n \times m$ 阶未知混合阵, A 为 $m \times n$ 阶待求分离阵。

一般地,要解决盲信号分离问题,需要满足以下假设:

- 1) 未知源信号向量 $S(t)$ 中各分量为相互独立,零均值平稳过程。
- 2) 混合阵 M 至少是列满秩的,一般假定为 $M = N$,要求混合阵为可逆矩阵。
- 3) 未知源信号向量 $S(t)$ 最多有一个分量服从高斯分布。

假设条件1是解决盲分离问题的关键,虽然这是一个严格的统计假设,但却是一个物理意义上非常合理的假设,因为通常源信号是从不同的物理系统出发,一般都能够满足相互独立条件,对于零均值的要求可以通过中心化来实现,对于要求单位方差的算法通过白化来实现。假设条件2是为了保证观察信号的数目大于未知的源信号的数目,一般情况下,假定混合系数是不变的。对于假设条件3,在源信号都是高斯分布情况下,它们的线性混合仍然是高斯分布,有人已经证明这是不可分离的。在这个意义上,可以将盲源分离理解为图2所示的过程。

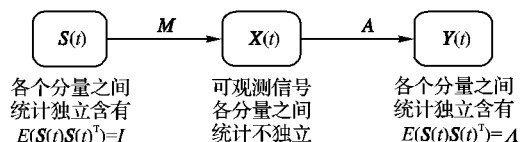


图2 盲源分离过程

2.2 分离目标函数——峭度

对零均值的随机信号,峭度即四阶累积量。零均值的高斯随机变量的峭度为0。峭度可正可负。峭度大于0的随机信号称为超高斯信号(super-gauss signals),声音信号就是超高斯信号。峭度小于0的随机信号称为亚高斯信号(sub-gauss signals),图像信号、通信信号都属于亚高斯信号。

由中心极限定理可知,一组独立随机变量的分布比任何源信号更接近于高斯分布。因此可以将非高斯性作为随机信

号相互独立的度量。根据峭度的特性,可通过寻求观测信号 X 的一个线性组合 AX ,使其峭度最大化(超高斯信号)或最小化(亚高斯信号)来分离独立分量。一般而言,要使用峭度作为分离目标函数,必须满足 $E((AX)^2) = I$ 的假设。从 $E((AX)^2) = I$,又有 $Y = AX$,所以 $E(Y Y^T) = I$,令 $Z = AM$;对于分量相互独立的源信号 S ,其峭度为: $k(AX) = k(AMS) = k(ZS)$ 。根据峭度的性质,则有:

$$k(ZS) = \sum_{i=1}^N k(z_i s_i) = \sum_{i=1}^N z_i^4 k(s_i) \quad (2)$$

在 $E(Y Y^T) = I$ 的约束下,上式有局部极大值。这样通过最大化给定限制条件下的峭度计算式(2),就可以将峭度作为分离的目标函数,得到一个独立分量。

3 基于递阶遗传算法的盲信号分离

3.1 遗传编码

这里给出一种两层递阶结构的染色体算法。用控制层表示分离阵 A 的有效行数(可能潜在的源信号个数),参数层表示分离阵有效行的具体数值,利用遗传算法的自适应性,在分离的过程中自动识别出最优分离方式。编码结构见图3。

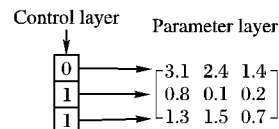


图3 用于盲信号分离的两层递阶结构染色体

在控制层使用二进制编码,控制层编码的位数对应潜在的源信号的最大可能个数,当控制层中某一基因位等于0,表示参数层对应的编码无效,也就是说虽然在参数层中有对应的参数,在该染色体对应的分离矩阵中不出现该基因位所对应的数值;反之,某基因位等于1表示参数层编码有效。在参数层依然使用浮点数编码,描述可能的分离阵对应的具体数值。为了满足观测信号不少于源信号个数,这里控制层只取三个控制基因。上述染色体实际对应的分离阵为:

$$A = \begin{bmatrix} 0.8 & 0.1 & 0.2 \\ 1.3 & 1.5 & 0.7 \end{bmatrix} \quad (3)$$

3.2 适值函数

考虑到处理信号既有超高斯信号(峭度大于零),又有亚高斯信号(峭度小于零),单纯使用峭度无法对其准确分离,因此,选择峭度绝对值的和作为独立性的判据。仿真结果表明,采用峭度的绝对值之和作为适应度函数可准确从混合信号中分离出亚高斯信号和超高斯信号。适值函数表示为:

$$\begin{aligned} J(Y) &= |k(Y)| = \\ &= |E(Y^4) - 3(E(Y^2))^2| = \\ &= \sum_{i=1}^N |k(y_i)| = \\ &= \sum_{i=1}^N |E(y_i^4) - 3(E(y_i^2))^2| \end{aligned} \quad (4)$$

在计算峭度时,式(4)中的 N 可以根据控制层中“1”的个数确定。算法开始时,不需要任何关于源信号个数的信息。在 $E(Y Y^T) = I$ 的条件下,对于迭代过程中的分离阵 A 而言, $J(Y)$ 越大表明各个估计信号之间的独立性越强。

应用峭度作为非高斯性度量,要求随机变量必须是零均值,并且满足 $E(Y Y^T) = I$ 。因此,使用信号计算峭度时必须先作白化和中心化处理。中心化处理就是用信号向量 X 减去其均值,这样得到的信号能够保证是零均值。白化处理是将信号向量 X 进行线性变换 Q ,得到 $A = QX$,其中, A 的各分量 v 互

不相关,且具有单位方差(即 $E(\mathbf{A}\mathbf{A}^T) = \mathbf{I}$)。信号白化处理可用主成分分析(Principal Component Analysis, PCA)来实现,操作步骤如下:

1) 求观测信号的相关矩阵:

$$\mathbf{R}_X = E(\mathbf{X}\mathbf{X}^T) \quad (5)$$

2) 对相关阵进行特征值分解,得到 N 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_N$, 和相应的特征向量 $\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_N$;

3) 计算白化信号,得到白化处理后的信号向量:

$$\mathbf{V} = \left([\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_N] \mathbf{D}^{\frac{1}{2}} \right)^T \mathbf{X} \quad (6)$$

其中 $\mathbf{D} = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_N]$ 。

3.3 遗传操作和参数

对参数层基因使用非均匀变异和浮点数交叉,在控制层采用均匀变异和单点交叉策略,使用“轮盘赌”方法进行选择。遗传参数可设定为:种群数为 50,交叉概率 0.45,变异概率 0.001,使用保优策略。

4 实例研究

实例一 考查观测信号与同源信号个数相等的盲信号分离情况。三种相互独立的信号分别为正弦信号 $\sin(8\pi t)$ 、Matlab 提供的锯齿波信号 $\text{sawtooth}(16\pi t)$ 、方波信号 $\text{square}(100\pi t)$ 。采样频率为 1000 Hz,取 1000 个数据。三种信号的峭度分别为 1.137、-0.11、-2,前一个为超高斯信号,后两个均为亚高斯信号。信号图像见图 4。

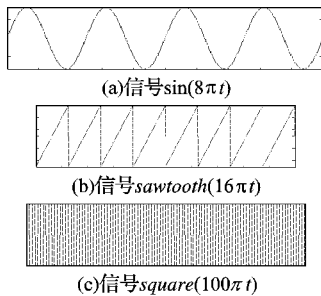


图4 $\sin(8\pi t)$ 、 $\text{sawtooth}(16\pi t)$ 、 $\text{square}(100\pi t)$ 源信号图像

混合矩阵使用随机方式产生:

$$\mathbf{M} = \begin{bmatrix} 0.9501 & 0.468 & 0.4565 \\ 0.2311 & 0.8913 & 0.0185 \\ 0.6068 & 0.7621 & 0.8214 \end{bmatrix} \quad (7)$$

得到观测信号,三组信号图像见图 5。

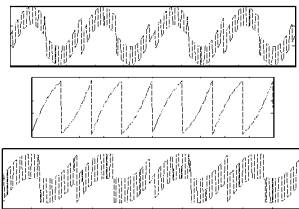


图5 随机方式产生的三组混合观测信号

染色体控制层有三个基因,参数层为一个 3×3 矩阵。经过 300 代进化,得到最优个体:

$$\mathbf{A} = \begin{bmatrix} -0.397 & 0.98 & 0.175 \\ 1.4231 & -0.073 & -0.8964 \\ -1.578 & -2.23 & 3.574 \end{bmatrix} \quad (8)$$

得到分离信号见图 6。

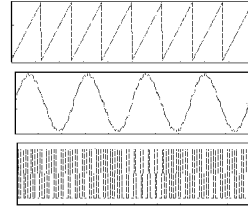


图6 三组观测信号分离结果

实例二 采用例一的 $\sin(8\pi t)$ 、 $\text{sawtooth}(16\pi t)$ 、 $\text{square}(100\pi t)$ 源信号图像,选用 5 组观测信号进行分离。随机产生的混合矩阵如下:

$$\mathbf{M} = \begin{bmatrix} 0.5893 & 1.487 & -0.6873 \\ -4.3786 & 0.6753 & 1.2484 \\ 5.0023 & -3.1675 & 1.0184 \\ 0.4103 & 0.0099 & 0.2722 \\ 0.8936 & 0.1389 & 0.1988 \end{bmatrix} \quad (9)$$

混合观测信号见图 7。

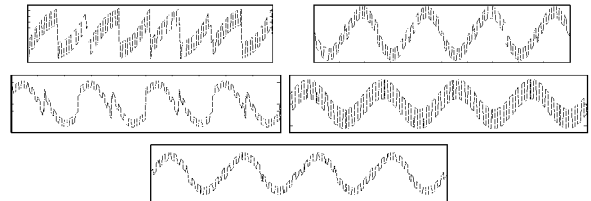


图7 随机方式产生的五组混合观测信号

因为有 5 组观测信号,所以染色体的控制层应为 5 个基因,参数层为 5×5 阶矩阵。经过 500 代进化得到最优个体如下:

$$\mathbf{A} = \begin{bmatrix} 1 & 0.4057 & 0.0579 & 0.2028 & 0.017 & -0.218 \\ 1 & 0.9823 & 0.37 & 0.1987 & 0.76 & 0.0013 \\ 0 & 0.8936 & 0.1389 & 0.1988 & 3.765 & -0.879 \\ 0 & 0.4103 & 0.0099 & 0.2722 & 0.794 & -0.9398 \\ 1 & 1.1200 & 0.9132 & 0.5908 & -0.241 & 0.1212 \end{bmatrix}$$

这里有两个基因位为零,说明有两组参数层基因无效,这意味着得到的估计信号只有三组。根据最优个体,可以得到分离矩阵:

$$\mathbf{A} = \begin{bmatrix} 0.4057 & 0.0579 & 0.2028 & 0.017 & -0.218 \\ 0.9823 & 0.37 & 0.1987 & 0.76 & 0.0013 \\ 1.12 & 0.9132 & 0.5908 & -0.241 & 0.1212 \end{bmatrix} \quad (10)$$

分离以后得到分离图像见图 8。

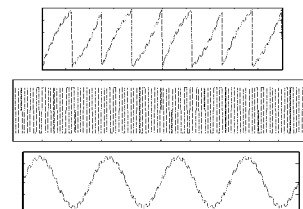


图8 五组观测信号分离结果

从图 8 可以看出信号已经得到分离,达到预期效果。

5 结语

研究表明,利用遗传算法的自适应性,采用峭度的绝对值之和作为适应度函数,可以在没有源信号个数信息的条件下分离盲源信号。结论对语音数字信号处理、多用户通信、阵列处理、生物医学、图像分析等领域中的信息处理具有一定的应用价值。研究的不足之处是仅选用规则波形作为盲源信

(下转第 1513 页)

可抢占区等任务属性均由系统随机产生,且系统始终在满负荷状态下运行。实验产生结果以任务执行的次数做统计,如图4所示。从图上可以看出,在系统满负荷条件下,多个软实时和硬实时任务各执行了150次,硬实时任务无一错过截止期(截止期错失率曲线与X轴重合),而软实时任务截止期错失率的峰值也仅有5%(前30个任务截止期错失率为0),仿真结果表明改进框架及算法达到了预期目标。

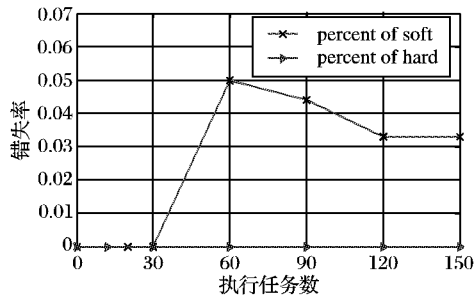


图4 硬、软实时任务执行结果对比

与原方案相比,改进方案虽然不能保证软实时任务都在截止期内完成,然而在统计意义上,系统为软实时任务提供了更好的并发性,只要保证软实时任务中的不可抢占区不影响硬实时任务可调度性,系统就可以接受更多的软实时任务并发执行,使系统始终处于满负荷状态,充分利用系统资源。

4 结语

本文深入研究了开放式实时系统双层调度策略,指出了其仅适应硬实时环境的局限性,在保证硬实时可调度的基础上,加入了软实时应用的处理,提出了一种改进方案,解决了原方案将硬实时与软实时应用作为同一类实时应用进行无区分调度问题,使之能适应更复杂的应用环境。通过证明和实验仿真,改进方案能达到前面阐述的预期研究目标。

应当指出,提出的改进方案中,对于更复杂的调度环境,如不可预测的应用和应用中含有突发任务等情况,还没有作详细探讨,这将是下一步的研究工作。

参考文献:

- [1] DENG Z, LIU JW-S. Scheduling real-time applications in an open Environment[C]// RTSS'97: Proceedings of the 18th IEEE Real-Time Systems Symposium. Washington, DC: IEEE Computer Society, 1997: 308 - 319.
- [2] WANG YC, LIN KJ. Implementing a general real-time scheduling framework in the RED-Linux real-time kernel[C] // RTSS'99: Proceeding of the 20th IEEE Real-Time Systems Symposium. Washing-

ton, D C: IEEE Computer Society, 1999: 246 - 255.

- [3] PAREKH A K. A generalized processor sharing approach to flow control in integrated services networks [D]. Cambridge: Massachusetts Institute of Technology, 1992.
- [4] KUO T W, YANG W R, LIN K J. EGPS: A class of real-time scheduling algorithms based on processor sharing[C]// Proceedings of the 10th Euromicro Workshop on Real Time Systems. Los Alamitos, CA: IEEE Computer Society Press, 1998: 27 - 34.
- [5] ABENI L, BUTTAZZO G. Integrating multimedia applications in hard real-time systems[C]// RTSS'98: Proceedings of the 19th IEEE Real-Time Systems Symposium. Washington, D C: IEEE Computer Society, 1998: 4 - 13.
- [6] ABENI L, BUTTAZZO G. Resource reservation in dynamic real-time systems[J]. Real-Time Systems, 2004, 27(2): 123 - 167.
- [7] LIPARI G, BARUAH S. A hierarchical extension to the constant bandwidth server framework[C]// Proceedings of the 7th IEEE Real-Time Technology and Applications Symposium. Washington, DC: IEEE Computer Society, 2001: 26 - 35.
- [8] MARCHAND A, SILLY-CHESTO M. Dynamic real-time scheduling of firm periodic tasks with hard and soft aperiodic tasks[J]. Real-Time Systems, 2006, 32(1/2): 21 - 47.
- [9] LI WEN-MING, KAVI K, AKL R. A non-preemptive scheduling algorithm for soft real-time systems[J]. Computers and Electrical Engineering, 2007, 33(1): 12 - 29.
- [10] DEVI U C, ANDERSON J H. Tardiness bounds under global EDF scheduling on a multiprocessor[J]. Real-Time System, 2008, 38(2): 133 - 189.
- [11] LIPAI G, CARPENTER J, BARUAH S. A framework for achieving inter-application isolation in multiprogrammed, hard real-time environments[C]// RTSS'00: Proceedings of the 21 st IEEE Real-Time Systems Symposium. Washington, DC: IEEE Computer Society, 2000: 217 - 226.
- [12] 龚育昌,王立刚,陈香兰,等.一种严格按比例派发服务的混合实时调度算法[J].软件学报,2006,17(3): 611 - 619.
- [13] 淮晓永,邹勇,李明树.一种开放混合实时系统的开放自适应调度算法[J].软件学报,2004,15(4): 487 - 496.
- [14] 谭朋柳,金海,张明虎.用于开放式系统的二维优先级实时调度[J].电子学报,2006,34(10): 1773 - 1777.
- [15] 邹勇,李明树,王青.开放式实时系统的调度理论与方法分析[J].软件学报,2003,14(1): 83 - 90.
- [16] 金永贤,黄经州,王建国.基于双层调度框架的开放式实时系统非实时应用调度[J].计算机应用,2008,28(6): 1608 - 1611.
- [17] LIU C L, LAYLAND J W. Scheduling algorithms for multiprogramming in a hard real-time environment[J]. Journal of the ACM, 1973, 20(1): 46 - 61.

(上接第1501页)

号,将超高斯信号和亚高斯信号进行了分离,没有考虑更复杂的混合信号,得出的结论尚难以解释或解决更复杂的问题。对不规则波形多组盲源信号的分离有待进一步研究。

参考文献:

- [1] HERAULT J, JUTTEN C. Blind separation of sources — 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(3): 287 - 314.
- [3] BELL A J, SEJNOWSKI T J. An information-maximization approach to blind separation and blind deconvolution [J]. Neural Computation, 1995, 7(6): 1129 - 1159.
- [4] AMARI S, CICHOCKI A, YANG H H. A new learning algorithm

for blind signal separation [EB/OL]. [2008 - 10 - 10]. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.1433>.

- [5] HYVARINEN A, OJA E. A fast fixed point algorithm for independent component analysis[J]. Neural Computation, 1997, 9(7): 1483 - 1492.
- [6] 朱孝龙,张贤达,冶继民.基于自然梯度的递归最小二乘盲信号分离[J].中国科学: E 辑, 2003, 33(8): 741 - 748.
- [7] 郑丕涛,马艳华. RBF 神经网络的递归遗传训练新方法[J].控制与决策, 2000, 2(3): 165 - 168.
- [8] 冶继民,张贤达,朱孝龙.信源数目未知和动态变化时的盲信号分离[J].中国科学: F 辑, 2005, 35(12): 1277 - 1287.
- [9] 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): 693 - 703.