

基于改进 K -均值聚类的欠定混合矩阵盲估计

付卫红, 马丽芬, 李爱丽

(西安电子科技大学综合业务网理论及关键技术国家重点实验室, 陕西 西安 710071)

摘要: 在源信号个数未知条件下, 提出一种基于改进 K -均值聚类的欠定混合矩阵盲估计方法。该方法首先计算观测信号在单位半超球面上投影点的密度参数, 然后去掉低密度投影点, 并从高密度投影点中选取初始聚类中心, 最后对剩余投影点进行聚类, 根据 Davies-Bouldin 指标估计源信号个数, 并估计出混合矩阵。仿真结果表明, 该方法的复杂度低, 其运行时间仅为拉普拉斯势函数法的 1%~3%; 该方法的源信号个数估计正确率远高于鲁棒竞争聚类算法, 当信噪比高于 13 dB 时, 该方法源信号个数估计正确率大于 96.6%, 且混合矩阵估计误差较小。该方法在信噪比较高时, 可降低对源信号稀疏度的要求。

关键词: 混合矩阵估计; Davies-Bouldin 指标; 密度参数; 改进 K -均值聚类

中图分类号: TN 911.7

文献标志码: A

DOI:10.3969/j.issn.1001-506X.2014.11.06

Blind estimation of underdetermined mixing matrix based on improved K -means clustering

FU Wei-hong, MA Li-fen, LI Ai-li

(State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China)

Abstract: A method for blind estimation of underdetermined mixing matrix based on improved K -means clustering is proposed when the source number is unknown. First, the density parameter of the projection points of the mixing signals on half of the unit ultra sphere is calculated. Then, the projection points with low density are removed and the initial clustering centers are chosen from the projection points with high density. Finally, cluster the remaining points, use the Davies-Bouldin index to estimate the source number, and estimate the mixing matrix. The simulation results show that the proposed algorithm's complexity is lower and its running time is only about 1% to 3% of that of the Laplace mixed model potential function algorithm; its source number estimation accuracy is much higher than that of the robust competitive agglomeration algorithm; when the signal to noise ratio is greater than 13 dB, its accuracy is higher than 96.6% and its estimated mixing matrix error is small. When SNR is higher, it can relax the sparsity requirement of the sources.

Keywords: mixing matrix estimation; Davies-Bouldin (DB) index; density parameter; improved K -means clustering

0 引言

近年来, 观测信号个数少于源信号个数的欠定盲源分离问题 (underdetermined blind sources separation, UBSS) 逐渐成为一个研究热点, 其在机械故障检测、语音信号处理、图像信号处理、生物医学工程、雷达信号处理跳频信号分选等领域都得到应用^[1-6]。目前基于稀疏分量分析 (sparse component analysis, SCA) 的两步法是解决该问题的一种常用方法^[7]。两步法根据观测信号的线性聚类特

征, 首先采用某种聚类算法估计混合矩阵, 然后再根据混合矩阵和观测信号恢复源信号。因此, 在两步法中混合矩阵的估计十分关键, 其估计精度直接影响源信号的恢复效果。文献[8]采用 Hard-Lost 算法估计混合矩阵, 该方法需给定源信号个数, 且估计性能易受初始值和孤立点影响。文献[9]基于混合聚类和网格密度估计混合矩阵, 提高了混合矩阵的估计精度, 但也只适用于源信号个数已知情况。文献[10]提出的势函数法可以在源信号个数未知时估计出混合矩阵, 但该方法仅适用于二维空间。文献[11]提出的

收稿日期: 2013-12-18; 修回日期: 2014-04-22; 网络优先出版日期: 2014-06-13。

网络优先出版地址: <http://www.cnki.net/kcms/detail/11.2422.TN.20140613.1643.001.html>

基金项目: 国家自然科学基金 (61201134, 61201135); 中央高校基本科研业务费专项资金 (72124669); 高等学校学科创新引智计划 (B08038); 国家新一代宽带无线和移动通信重大专项 (2012ZX03001027-001); 中国航天科技集团公司卫星应用研究院创新基金 (2014_CXJJ-TX_06) 资助课题

拉普拉斯势函数法(Laplace mixed model potential function algorithm, LMMPF)可用于高维空间,但为估计出所有的局部最大值,该方法将所有观测数据作为初始聚类中心,致使计算复杂度太高,不具有实用性。另外文献[11]中的鲁棒竞争聚类算法(robust competitive agglomeration, RCA)减小了计算复杂度,但混合矩阵估计精度较低。针对原始 K -均值聚类算法需给定聚类个数 K ,聚类效果依赖于初始聚类中心的选取^[12],文献[12-13]均基于密度和距离优化初始聚类中心,减小算法对初始聚类中心的敏感性。本文为解决欠定盲源分离中源信号个数未知时混合矩阵估计问题,提出一种改进 K -均值聚类算法,该算法对文献[13]中数据对象的邻域半径进行重新定义,然后基于密度和距离选取初始聚类中心,并采用文献[14]中的 Davies-Bouldin (DB)指标确定最佳聚类个数。

1 欠定混合盲源分离系统模型

通常情况下,线性瞬时混合欠定盲源分离系统的基本模型可以表示为

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t), t = 1, 2, \dots, T_0 \quad (1)$$

式中, $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$ 为 t 采样时刻观测信号向量; \mathbf{A} 为未知的 $M \times N$ 混合矩阵; $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$ 为 t 采样时刻未知的源信号向量; $\mathbf{v}(t) = [v_1(t), v_2(t), \dots, v_M(t)]^T$ 为 t 采样时刻 M 维加性高斯白噪声向量; M 为观测信号个数(传感器个数); N 为源信号个数; T_0 为采样数据长度。

稀疏分量分析假设源信号在时域或变换域为稀疏信号,其在大多数采样时刻的取值为零或接近于零,只有少数采样时刻的取值远离零。当源信号均为稀疏信号时,在大多数采样时刻仅有一个源信号起主导作用。假设 t 采样时刻仅有 $s_n(t)$ 起主导作用,则式(1)可以简化为

$$\mathbf{x}(t) = \mathbf{a}_n s_n(t) + \mathbf{v}(t) \quad (2)$$

式中, \mathbf{a}_n 为混合矩阵 \mathbf{A} 的第 n 列。由式(2)可知,如果没有噪声,采样数据分布在 M 维观测信号空间中一条经过原点的直线上;如果有噪声,采样数据也分布在这条直线附近,直线的方向取决于混合矩阵的列向量,直线条数就是稀疏源信号的个数。因此在充分稀疏源信号欠定混合情形下,可以利用某种聚类算法根据观测信号估计出直线的个数和方向。对于时域并不稀疏的源信号可以利用稀疏变换工具,如短时傅里叶变换(short-time Fourier transform, STFT),WVD (Wigner-Vile distribution) 等对信号进行稀疏表示,在变换域进行盲源分离^[15-16]。

2 改进的 K -均值聚类算法

原始 K -均值聚类算法首先从 n 个数据对象中任意选取 K 个作为初始聚类中心;然后根据最小距离原则将其余数据对象划分到各类;重新计算各类的聚类中心,再次根据最小距离原则将数据对象划分到各类,如此循环迭代,直到聚类中心不再发生变化^[12]。针对原始 K -均值算法需要给定聚类个数 K 且性能易受初始值影响等缺点,本文从两个方面进行改进,一是基于密度和距离选取初始聚类中心,二

是采用 DB 指标确定聚类个数。

2.1 基于密度和距离选取初始聚类中心

设待聚类的数据集合为 $\tilde{\mathbf{X}}$, 表示为

$$\tilde{\mathbf{X}} = \{\tilde{\mathbf{x}}_i | \tilde{\mathbf{x}}_i \in \mathbf{R}^M, i = 1, 2, \dots, n\}$$

定义 1 数据对象的邻域半径 r

数据对象的邻域半径 r 为所有对象之间最小距离的平均值,具体计算公式为

$$r = \frac{1}{n} \times \sum_{j=1}^n \min(d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)), j = 1, 2, \dots, n, j \neq i \quad (3)$$

定义 2 数据对象的密度参数

以任意数据对象 $\tilde{\mathbf{x}}_i$ 为中心, r 为半径的邻域内数据对象的个数称为 $\tilde{\mathbf{x}}_i$ 基于距离 R 的密度参数 $B_r(\tilde{\mathbf{x}}_i)$, 具体计算公式为^[13]

$$\begin{cases} B_r(\tilde{\mathbf{x}}_i) = \sum u(r - d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)) \\ u(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}, i, j = 1, 2, \dots, n \end{cases} \quad (4)$$

式中, $d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) = \|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|$ 为集合 $\tilde{\mathbf{X}}$ 中任意两个数据对象之间的欧式距离, $\|\cdot\|$ 表示 2 范数。为防止选择的初始聚类中心聚集在采样数据密度最高的区域,第一个初始聚类中心选择密度参数最大的点,剩余聚类中心基于最大最小距离选取,具体选取参见第 2.3 节步骤 3。

2.2 采用 DB 指标确定最佳聚类个数

DB 指标以类间的离散程度和类内的紧密程度为依据来评估聚类效果。同一类内的数据对象相似程度越高,不同类间的相似程度越低, DB 指标值就越小,分类效果越好,数据间的相似度采用欧式距离来度量。

定义 3 类间距离^[14]表示为

$$M_{cd} = \|\tilde{\mathbf{z}}_c - \tilde{\mathbf{z}}_d\| \quad (5)$$

式中, $\tilde{\mathbf{z}}_c$ 和 $\tilde{\mathbf{z}}_d$ 分别表示第 c 和第 d 个聚类中心。

定义 4 类内距离^[14]表示为

$$S_c = \frac{1}{T_c} \sum_{q=1}^{T_c} \|\tilde{\mathbf{x}}_q - \tilde{\mathbf{z}}_c\| \quad (6)$$

式中, T_c 表示第 c 类中数据对象的个数; $\tilde{\mathbf{x}}_q (q=1, 2, \dots, T_c)$ 为第 c 类中的数据对象。

定义 5 DB 指标^[14]表示为

$$I_K = \frac{1}{K} \sum_{c=1}^K \max_d \left(\frac{S_c + S_d}{M_{cd}} \right) \quad (7)$$

$$c = 1, 2, \dots, K, d \neq c$$

当类间距离 M_{cd} 增大,类内数据对象与该类聚类中心的平均距离 S_c 和 S_d 变小时, DB 指标随之变小。当 DB 指标最小时,代表该聚类数目下的聚类结果最佳。因此,可以根据 DB 指标值的大小确定最佳聚类数目。欠定条件下观测信号个数小于源信号个数,故选取初始聚类中心个数与观测信号个数相同,即 $K=M$ 。

2.3 改进 K -均值聚类算法实现步骤

改进 K -均值聚类算法的具体实现步骤如下:

步骤 1 选取初始聚类中心个数 $K=M$;

步骤 2 计算集合 $\tilde{\mathbf{X}}$ 中任意数据对象 $\tilde{\mathbf{x}}_i$ 的密度参数

$B_r(\tilde{x}_i)$, 设定门限 η_1 和门限 η_2 , 将密度参数 $B_r(\tilde{x}_i) < \eta_1$ 的数据对象删除, 得到剩余数据对象集合 \tilde{X} ; 保存密度参数 $B_r(\tilde{x}_i) > \eta_2$ 的数据对象, 得到高密度数据对象集合 Π ; 门限 η_1 和门限 η_2 与源信号的稀疏度和信噪比有关, 可通过实验选取;

步骤 3 基于密度和距离选取初始聚类中心^[13]:

步骤 3.1 从高密度集合 Π 中选择密度参数最大的数据对象作为第一个初始聚类中心 z_1 , 取距离 z_1 最远的数据对象作为第 2 个初始聚类中心 z_2 ;

步骤 3.2 对于 $l=3, 4, \dots, K$, 计算高密度集合 Π 内其余数据对象 \tilde{x}_i 到已选 $l-1$ 个初始聚类中心 z_c 的欧式距离 $d(\tilde{x}_i, z_c) (c=1, 2, \dots, l-1, i \in [1, p])$, p 为高密度集合 Π 中数据对象个数; 选取到 $l-1$ 个初始聚类中心的最小欧式距离最大的数据对象作为第 l 个初始聚类中心 z_l

$$z_l = \arg \max_{\substack{\tilde{x}_i \in \Pi \\ \tilde{x}_i \neq z_c}} \left\{ \min_{\substack{c=1, 2, \dots, l-1 \\ i=1, 2, \dots, p}} d(\tilde{x}_i, z_c) \right\} \quad (8)$$

步骤 4 计算集合 \tilde{X} 内所有数据对象到所有聚类中心的欧式距离, 并将其分配给距离最小的类; 计算每个类内所有数据对象的平均值, 作为重新分类后的聚类中心 $\tilde{z}_1 \dots \tilde{z}_K$, 重复这一过程, 直到聚类中心不再发生变化为止;

步骤 5 按式(5)~式(7)计算当前聚类个数下的 DB 指标值 I_K ;

步骤 6 令聚类个数 $K=K+1$, 重复步骤 2~步骤 5, 直到某次计算出的指标值满足迭代停止条件: $I_K > I_{K-1}$ 且 $I_{K-1} < I_{K-2}$, 则最佳聚类个数为 K_{opt} , 对应的聚类中心 $\tilde{z}_1 \dots \tilde{z}_{K_{opt}}$ 为最优聚类中心, 迭代结束。

3 基于改进 K -均值聚类的欠定混合矩阵盲估计

对于充分稀疏混合信号, 在有噪声的背景下近似为直线分布, 为便于估计高维混合信号空间中混合矩阵各个列矢量所构成的直线方向, 将混合信号采样数据 $x(t)$ 投影到单位半球面上^[9], 即将采样数据进行正则化

$$\tilde{x}(t) = \text{sign}(x_i(t)) \times x(t) / \|x(t)\| \quad (9)$$

式中, $t=1, 2, \dots, T_0$; $\text{sign}(\cdot)$ 表示求变量的符号; $x_i(t)$ 为 $x(t)$ 的第 i 个分量。正则化使混合矩阵各个列矢量所构成的直线在超球面上只有一点, 这些点构成了各个聚类中心, 采样数据集中分布在这些点的周围。计算所有投影点的密度参数, 由于噪声和异常值偏离混合矩阵列矢量所对应的直线, 其对应的投影点密度参数较小, 可以删除密度参数较小的点, 来减小噪声和异常值的影响; 选取密度参数较大的点, 同时基于最大最小距离选取初始聚类中心, 减小算法对初始值的敏感性。

算法步骤如下:

步骤 1 为了减小噪声的影响和计算量, 将 $\|x(t)\| < \epsilon_1$ 的观测数据删除不计, 按式(9)将剩余观测数据投影到单位半球面, 将所有投影点记为集合 \tilde{X} ; 本文实验选取 $\epsilon_1 = 0.2 \times \bar{E}$, 其中, \bar{E} 为观测信号能量的均值;

步骤 2 采用第 2.3 节提出的改进 K -均值聚类算法将集合 \tilde{X} 中的数据对象进行聚类, 确定最佳聚类个数 K_{opt} , 每

个类的聚类中心 \tilde{z}_k 以及每个类包含的数据对象个数 C_k , 其中, $k=1, 2, \dots, K_{opt}$;

步骤 3 为保证得到的聚类不是由少数异常值和噪声构成的, 计算每个类包含的数据对象个数与个数最大值 C 的比值 $p_k = C_k / C$, 其中 $C = \max(C_1, C_2, \dots, C_K)$; 设定门限 ϵ_2 , 如果 $p_k < \epsilon_2$, 则去掉该类, 最后保留下来的聚类个数即为源信号个数的估计值, 对应的聚类中心矢量即为混合矩阵的各列矢量的估计值^[11], 由此得到混合矩阵 A 的估计值 \hat{A} ; 由于通过去低密度投影点去除了部分噪声和异常值, ϵ_2 的取值不易太大, 一般在 0.1~0.3 左右。

4 算法复杂度分析对比

LMMPF 法需要估计两个参数, 并将所有观测数据作为初始聚类中心, 计算复杂度为 $O(T_0 + a\tau_0 T_0^2 + \tau r T_0^2)$, 其中, T_0 为数据长度, τ_0 为参数估计时的迭代次数, τ 为估计聚类中心时的迭代次数, a 为常数, r 为算法估计出的参数值; RCA 法将初始聚类中心个数减少为 C , 计算复杂度为 $O(a_1 \tau C T_0)$, $C \ll T_0$, 其中, a_1 为常数; 本文方法的复杂度高于原始 K -均值算法, 为 $O((K-m+2)\tau k L + a_3 T_0)$, 其中, L 为去低密度后剩余的点数, K 为聚类中心个数, m 为观测信号个数, k 为每次选取的聚类中心数 ($m \leq k \leq K+1$)。可以看出 LMMPF 法的复杂度远高于 RCA 算法和本文方法, 在后续的仿真实验中, 复杂度采用算法每执行一次的运行时间来表示。

5 仿真实验

实验 1 源信号个数 $N=4$, 观测信号个数 $M=3$, 源信号为语音信号, 采样长度 $T_0=50\,000$, 如图 1 所示。

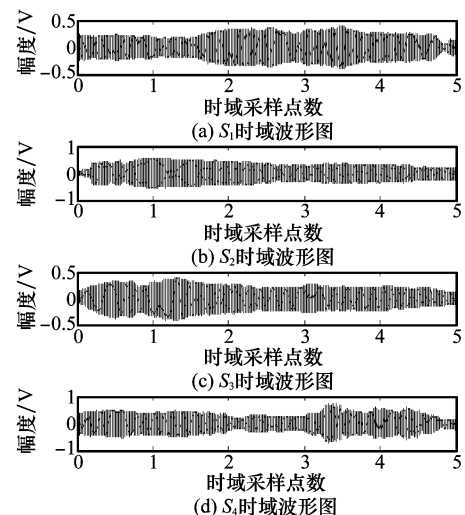


图 1 源信号

随机选取混合矩阵为

$$A = \begin{bmatrix} 0.478\,7 & -0.511\,5 & 0.986\,3 & 0.159\,5 \\ 0.456\,9 & 0.414\,4 & -0.120\,4 & 0.839\,9 \\ 0.749\,8 & 0.752\,8 & -0.112\,4 & -0.518\,8 \end{bmatrix}$$

得到的观测信号在时域不具有稀疏性;采用短时傅里叶变换 STFT 将其变换到时频域,取变换长度 $N_{fft} = 4\ 096$,加矩形窗,窗长为与变换长度相同,移动步长为 256;取 STFT 变换后观测信号的实部,如图 2 所示,可以看出语音信号在时频域具有很好的稀疏性。仿真参数 $\epsilon_1 = 0.2 \times \bar{E}$, $\epsilon_2 = 0.2$, $\eta_1 = 0.1 \times B_{\max}$, $\eta_2 = 0.2 \times B_{\max}$, B_{\max} 为密度参数的最大值。

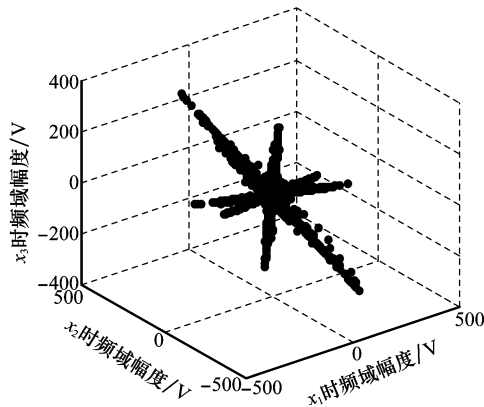


图 2 观测信号时频域散点图

用本文方法估计混合矩阵,得到混合矩阵估计值为

$$\hat{\mathbf{A}} = \begin{bmatrix} -0.986\ 2 & 0.478\ 1 & -0.159\ 4 & -0.511\ 4 \\ 0.120\ 6 & 0.458\ 5 & -0.840\ 1 & 0.413\ 1 \\ 0.113\ 6 & 0.749\ 1 & 0.518\ 4 & 0.753\ 5 \end{bmatrix}$$

估计误差 $\min \|\hat{\mathbf{A}}\mathbf{P} - \mathbf{A}\|_F = 0.002\ 7$, 式中, \mathbf{P} 为一置换矩阵, $\|\cdot\|_F$ 为矩阵的 F-范数^[9], 不同聚类个数下 DB 指标的值如表 1 所示。

表 1 不同聚类个数下 DB 指标值

聚类个数 k	3	4	5
DB 指标值	0.318 4	0.017 3	0.494 5

由仿真结果可以看出,改进 K-均值聚类算法可以正确估计出语音信号的个数,且混合矩阵估计精度较高。在估计出的混合矩阵基础上采用线性规划法^[17]恢复出语音源信号,并按式(10)计算信干比,结果如表 2 所示。

$$SIR_{ni} = 10 \lg \left(\frac{\|\hat{s}_i\|}{\|\hat{s}_i - s_i\|_2^2} \right) \quad (10)$$

式中, \hat{s}_i 为源信号 s_i 的估计值。 SIR_{ni} 越大说明 \hat{s}_i 与 s_i 越匹配,分离效果越好。

表 2 恢复信号的 SIR

信号	1	2	3	4
SIR	20.674 7	19.923 4	25.646 4	25.542 3

由仿真结果可以看出,改进 K-均值聚类算法可以应用在实际语音信号的盲源分离中,且混合矩阵的估计误差较小。

实验 2 源信号个数 $N=5$, 观测信号个数 $M=3$, 稀疏源信号为

$$\begin{aligned} \mathbf{S} &= -\lg(\text{rand}(N, T_0)) \cdot \\ &\max(0, \text{sign}(\text{rand}(N, T_0) - q)), \\ 0 &\leq q \leq 1 \end{aligned} \quad (11)$$

式中, \cdot 表示点乘; T_0 为采样数据长度; 参数 q 的值越大信号的稀疏度越高,反之越小。取 $T_0 = 3\ 000$, $q = 0.9$, 随机选取混合矩阵为

$$\mathbf{A} = \begin{bmatrix} 0.552\ 5 & -0.190\ 4 & 0.570\ 7 & 0.724\ 0 & -0.075\ 9 \\ 0.686\ 3 & 0.514\ 8 & -0.516\ 6 & 0.408\ 8 & -0.829\ 3 \\ 0.473\ 0 & 0.845\ 9 & 0.638\ 3 & -0.555\ 6 & 0.553\ 6 \end{bmatrix}$$

参数 $\epsilon_1 = 0.2 \times \bar{E}$, $\epsilon_2 = 0.2$, $\eta_1 = 0.25 \times B_{\max}$, $\eta_2 = 0.33 \times B_{\max}$, 信噪比为 $[10:1:20]$ dB。由于没有对信噪比进行估计,所以对于不同的信噪比都是采用相同的门限,故门限值 η_1 和 η_2 的取值较大。将本文方法和 LMMPF 法以及 RCA 法在计算复杂度、源信号个数估计正确率和混合矩阵估计误差 3 个方面的性能进行比较。为验证算法的有效性,仿真中混合矩阵保持不变,改变源信号,重复 500 次。需要说明,混合矩阵估计误差是在源信号个数估计正确的前提下计算。仿真结果如图 3~图 5 所示。

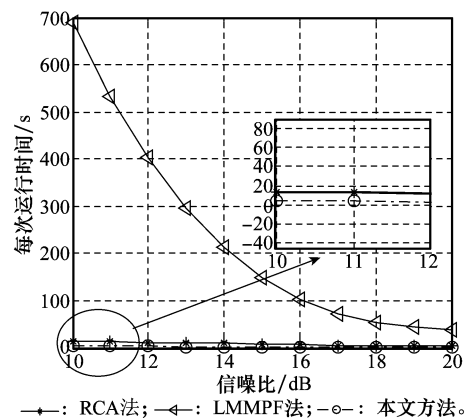


图 3 计算复杂度随信噪比变化曲线

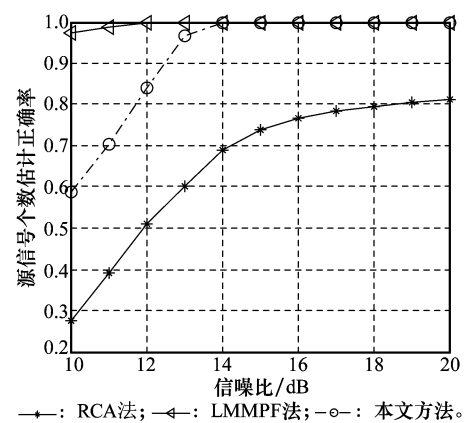


图 4 源信号个数估计正确率随信噪比变化曲线

由图 3 可以看出,本文所提算法大大降低了计算复杂度,其运行时间为 RCA 法的 10%~50%, LMMPF 算法的 1%~3%;由图 4 可以看出,本文文算法的源信号个数估计

正确率远高于 RCA 算法的正确率;当信噪比为 13 dB 时,本文算法的正确率达到 96.6%,当信噪比大于 13 dB 时,本文算法的正确率与 LMMPF 法相当,几乎可以完全正确地估计出源信号个数;由图 5 可以看出,本文算法的混合矩阵估计误差在 10~13 dB 之间迅速减小,当信噪比高于 13 dB 时,本文算法的混合估计误差最小,精确度最高。

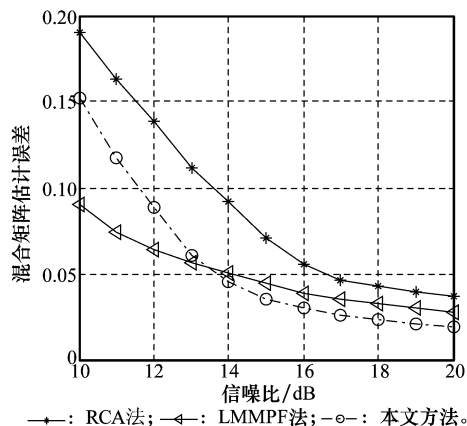


图 5 混合矩阵估计误差随信噪比变化曲线

为分析源信号稀疏度对本文算法性能的影响,设信噪比为 20 dB,稀疏度参数 $q=0.6:0.05:0.95$,其余参数保持不变。仿真结果如图 6~图 8 所示。

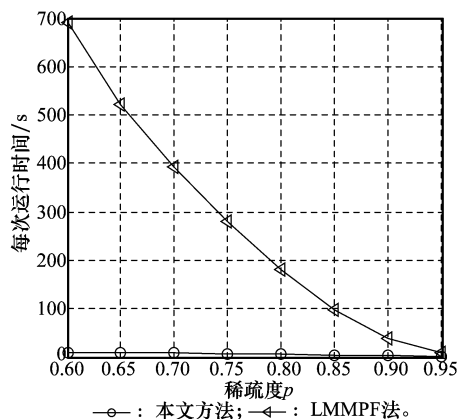


图 6 运算时间随稀疏度变化曲线

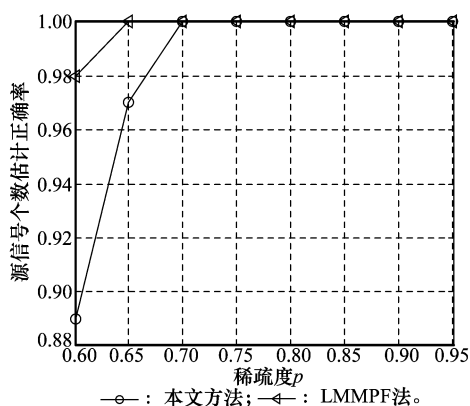


图 7 源信号个数估计正确率随稀疏度变化曲线

由图 6 可以看出,在相同稀疏度下,本文方法的计算复杂度远远低于 LMMPF 算法,具有更强的实用性。由图 7 可以看出,在稀疏度大于 0.7 时,本文算法的源信号个数估计正确率与 LMMPF 算法相当;由图 8 可以看出,在相同稀疏度下,本文方法混合矩阵估计误差小于 LMMPF 算法,为源信号的恢复提供了更高的精确度保证。

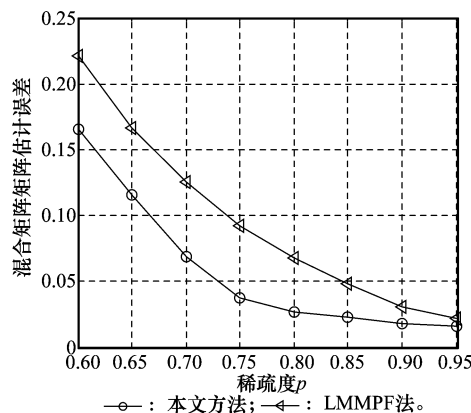


图 8 混合矩阵估计误差随稀疏度变化曲线

6 结 论

本文提出的基于改进 K-均值聚类的欠定混合矩阵盲估计算法,可以很好地解决欠定盲源分离中源信号个数未知时混合矩阵估计问题。从仿真实验结果可以看出,该方法具有简单、快速的优点,较 LMMPF 法和 RCA 法,算法复杂度大大降低;当信噪比不低于 13 dB 时,该方法可以正确估计出源信号个数且混合矩阵估计误差较小;当信噪比较高时可以降低对源信号稀疏度的要求。

参考文献:

- [1] Li Z N, Liu W B, Yi X B. Underdetermined blind source separation method of machine faults based on local mean decomposition[J]. *Journal of Mechanical Engineering*, 2011, 47(7): 97-102. (李志农, 刘卫兵, 易小兵. 基于局域均值分解的机械故障欠定盲源分离方法研究[J]. *机械工程学报*, 2011, 47(7): 97-102.)
- [2] Thiagarajan J J, Ramamurthy N R, Spanias A. Mixing matrix estimation using discriminative clustering for blind source separation[J]. *Digital Signal Processing*, 2013, 23(1): 9-18.
- [3] Koldovsky Z, Phan A H, Tichavsky P, et al. A treatment of EEG data by underdetermined blind source separation for motor imagery classification[C]// *Proc. of the 20th European Signal Processing Conference*, 2012: 1484-1488.
- [4] Gai K, Shi Z, Zhang C. Blind separation of superimposed moving images using image statistics[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2012, 34(1): 19-32.
- [5] Ma J, Huang G, Zhou D, et al. Underdetermined blind sorting of radar signals based on sparse component analysis[C]// *Proc. of the 14th IEEE International Conference on Communication Technology*, 2012: 1296-1300.

- [6] Sha Z C, Huang Z T, Zhou Y Y, et al. Frequency-hopping signals sorting based on underdetermined blind source separation[J]. *IET Communications*, 2013, 7(14): 1456-1464.
- [7] Koldovsky Z, Tichavsky P, Phan A H, et al. A two-stage MMSE beamformer for underdetermined signal separation[J]. *IEEE Signal Processing Letters*, 2013, 20(12): 1227-1230.
- [8] Gray P D O, Pearlmutter B A. Hard-LOST: modified k-means for oriented lines[C]// *Proc. of the Irish Signals and Systems Conference*, 2004: 247-252.
- [9] Bi X J, Gong R J. Underdetermined blind mixing matrix estimation algorithm based on mixing clustering and mesh density[J]. *Systems Engineering and Electronics*, 2012, 34(3): 614-618. (毕晓君, 宫汝江. 基于混合聚类和网格密度的欠定盲矩阵估计[J]. 系统工程与电子技术, 2012, 34(3): 614-618.)
- [10] Bofill P, Zibulevsky M. Underdetermined blind source separation using sparse representations[J]. *Signal Processing*, 2001, 81(11): 2353-2362.
- [11] Zhang Y. Blind separation of underdetermined mixtures of sources[D]. Shanghai: Shanghai University, 2009 (张烨. 欠定混合信号的盲分离[D]. 上海: 上海大学, 2009.)
- [12] Xie J Y, Guo W J, Xie W X, et al. K-means clustering algorithm based on optimal initial centers related to pattern distribution of samples in space[J]. *Application Research of Computers*, 2012, 9(3): 888-892. (谢娟英, 郭文娟, 谢维信, 等. 基于样本空间分布密度的初始聚类中心优化 K-均值算法[J]. 计算机应用研究, 2012, 9(3): 888-892.)
- [13] Yuan F, Zhou Z Y, Song X. K-means clustering algorithm with meliorated initial center[J]. *Computer Engineering*, 2007, 33(3): 65-66. (袁方, 周志勇, 宋鑫. 初始聚类中心优化的 k-means 算法[J]. 计算机工程, 2007, 33(3): 65-66.)
- [14] Davies D L, Bouldin D W. A cluster separation measure[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1979(2): 224-227.
- [15] Xie S, Yang L, Yang J M, et al. Time-frequency approach to underdetermined blind source separation[J]. *IEEE Trans. on Neural Networks and Learning Systems*, 2012, 23(2): 306-316.
- [16] Alshabrawy O S, Ghoneim M E, Salama A A, et al. Underdetermined blind separation of an unknown number of sources based on Fourier transform and matrix factorization[C]// *Proc. of the Federated Conference on Computer Science and Information Systems*, 2013: 19-25.
- [17] Gong J. The study on underdetermined blind source separation of mixed speech separation[D]. Dalian: Dalian University of Technology, 2010. (巩俊. 语音信号欠定盲源分离技术的研究[D]. 大连: 大连理工大学, 2010.)

作者简介:

付卫红(1979-), 女, 副教授, 博士, 主要研究方向为通信信号处理、无线通信。

E-mail: whfu@mail. xidian. edu. cn

马丽芬(1989-), 女, 硕士研究生, 主要研究方向为欠定盲源分离。

E-mail: lfmahit@163. com

李爱丽(1989-), 女, 硕士研究生, 主要研究方向为欠定盲源分离。

E-mail: 1211817343@qq. com