基于改进源信号数目估计算法的欠定盲分离*

薄祥雷[1]，何怡刚[1,2]，尹柏强[1]，方葛丰[1,3]，樊晓腾[3]
(1. 湖南大学 电气与信息工程学院，长沙 410082；2. 合肥工业大学 电气与自动化工程学院，合肥 230009；
3. 中国电子科技集团公司第 41 研究所 电子测试技术国防科技重点实验室，山东 青岛 266555)

摘 要：两步法是解决稀疏信号欠定盲分离的一种常用方法，通常首先利用 K-means 聚类算法估计混叠矩阵，然后利用最短路径法恢复源信号。在使用 K-means 聚类算法时要求知道源信号的数目，而现实中往往不知道源信号的数目，需要对其进行估计。因此研究了聚类有效性评价指标——BWP 指标，结合粒子群算法，提出了一种改进的确定源信号数目的算法，并将这种算法引入到欠定盲分离。实验表明，提出的算法在保证分离精度的同时能缩短分离时间，并可节省一定的内存，在观测信号数据量大时，这种优势更加明显。

关键词：粒子群算法；K-means 聚类；最短路径法；欠定盲分离
中图分类号：TN911.7 **文献标志码**：A **文章编号**：1001-3695(2014)05-1358-04
doi:10.3969/j.issn.1001-3695.2014.05.017

Underdetermined BSS based on improved estimate algorithm for source signal's number

BO Xiang-lei[1], HE Yi-gang[1,2], YIN Bai-qiang[1], FANG Ge-feng[1,3], FAN Xiao-teng[3]
(1. *College of Electrical & Information Engineering, Hunan University, Changsha* 410082, *China*; 2. *School of Electrical Engineering & Automation, Hefei University of Technology, Hefei* 230009, *China*; 3. *National Key Laboratory of Science & Technology on Electronic Test & Measurement, The* 41*st Research Institute of China Electronics Technology Group Corporation, Qingdao Shandong* 266555, *China*)

Abstract: Two-step method is a commonly used solution to the underdetermined blind source separation of sparse source signals, K-means clustering algorithm is usually utilized to estimate mixing matrix firstly, and then, the shortest path algorithm is used to recover source signals. The source signal's number needed in the K-means clustering algorithm is usually unknown in fact, so have to estimate it. This paper studied clustering validity index: BWP index, combined with particle swarm optimization, proposed a new algorithm to determine the source signal's number. Then it introduced the proposed algorithm into the underdetermined blind source separation. The results of experiment show that the proposed algorithm has good separation precision, meanwhile it can shorten the separation time and save some memory, this advantage is more obvious when the observation data is huge.

Key words: particle swarm optimization; K-means clustering algorithm; shortest path algorithm; underdetermined BSS

0 引言

盲源分离问题，就是在未知混合参数的情况下，仅仅根据观测到的混合信号恢复出源信号。盲源分离的数学模型为

$$\boldsymbol{X}(t)=\boldsymbol{A}\boldsymbol{S}(t)+N(t)\quad t=1,\cdots,T \tag{1}$$

其中：$\boldsymbol{X}(t)=[x_1(t),x_2(t),\cdots,x_m(t)]^{\mathrm{T}}$ 为观测信号向量；$\boldsymbol{A}$ 为混叠矩阵；$\boldsymbol{S}(t)=[s_1(t),s_2(t),\cdots,s_n(t)]^{\mathrm{T}}$ 为源信号向量；$N(t)=[n_1(t),n_2(t),\cdots,n_m(t)]^{\mathrm{T}}$ 为噪声。盲分离的目的就是通过仅仅已知的观测信号向量 $\boldsymbol{X}(t)$ 恢复源信号的波形，因此恢复出来的源信号可以有顺序和幅度的不确定性，而且通常在盲分离中暂时不考虑噪声的影响。因此上述模型也可写做：

$$\boldsymbol{X}(t)=\boldsymbol{A}\boldsymbol{S}(t)\quad t=1,2,\cdots,T \tag{2}$$

在实际信号环境中，由于潜在的源信号数目未知，而接收阵元数目有限，往往导致接收信号中源信号数目 n 大于阵元数目 m，这种混合信号的盲分离被称为欠定盲源分离(underdetermined blind source separation，UBSS)[1]。目前欠定盲分离是一个国际上非常棘手的问题，但是在实际应用中，因为有很多信号具备稀疏特性，或者可以通过适当的数学变换如 Fourier 变换、小波变换等相应变换使其稀疏化，故可以通过对信号的稀疏表示(sparse representation)[2]，对盲分离中的一些棘手问题进行探讨。到目前为止，两步法是解决稀疏信号盲源分离的一种常用方法[3,4]：a) K-means 聚类，估计混叠矩阵；b) 利用最短路径法恢复源信号。而在两步法中，K-means 聚类又显得格外重要，由于在 K-means 聚类时要求知道源信号的数目，在实际应用中往往不知道源信号的数目，就需要估计源信号的数目。有关阵列信号处理中的源信号数目的估计，人们对此作过

收稿日期：2013-07-08；**修回日期**：2013-08-27 **基金项目**：国家杰出青年科学基金资助项目(50925727)；国家自然科学基金资助项目(60876022,61102039,51107034)；湖南省科技计划资助项目(2011J4,2011JK2023)；湖南省自然科学基金资助项目(12JJA004)；国防计划资助项目(C1120110004,9140A27020211DZ5102)；国家教育部科学技术研究重大项目(313018)；中央高校基本科研业务费计划资助项目

作者简介：薄祥雷(1987-)，男，山东济宁人，硕士研究生，主要研究方向为盲信号处理、复杂电磁场模拟与监测(boxl@hnu.edu.cn)；何怡刚(1966-)，男，湖南邵阳人，教授，博士，主要研究方向为自动测试与诊断、高速低压低功耗集成电路与系统；尹柏强(1983-)，男，湖南邵阳人，博士研究生，主要研究方向为信号处理；方葛丰(1964-)，男，山东青岛人，高级工程师，硕士，主要研究方向为无线电电子学；樊晓腾(1971-)，男，山东青岛人，高级工程师，硕士，主要研究方向为微波毫米波测试信号发生技术.

一些专门的研究,但是在以往的研究中都是针对过定盲分离中源信号数目的估计,而且具体估计过程比较复杂,对于欠定盲分离的源信号数目的估计相关研究尚不多见,所以源信号数目估计对盲分离技术的发展具有重要的意义,也是目前必须予以解决的问题。文献[5]基于样本的几何结构,设计了一种叫类间类内划分(between-within proportion,BWP)有效性指标,并在其基础上提出了一种确定样本最佳聚类数的方法,用来评估 K-means 算法的聚类结果和确定样本的最佳聚类数,本文称其为 K-means 遍历算法。经验证,这种方法能够非常准确地得到最佳聚类数,但存在运算量大、消耗时间长的缺点。

本文采用粒子群算法对 K-means 遍历算法进行改进,并将改进的算法用于确定欠定盲分离的源信号数。最后给出了六路源信号两路观测信号的欠定盲分离实验,并与不同算法对比了本文所提出算法的分离精度和所耗时间与平均内存。

1　基于 BWP 指标的 K-means 聚类数确定算法

周世兵等人[5]借鉴 Rousseeuw 的指标说明方法,提出了类间类内划分(between-within proportion,BWP)指标,在寻找最佳聚类数方面进行了尝试,并取得了一定的成效。

1.1　BWP 指标含义

定义 1　令 $K=\{X,R\}$ 为聚类空间,$X=\{x_1,x_2,\cdots,x_n\}$,假设 n 个样本对象被聚类为 c 类,定义第 j 类的第 i 个样本的最小类间距离 $b(j,i)$ 为该样本到其他每个类中样本平均距离的最小值,即

$$b(j,i)=\min_{1<<k<<c,k\neq j}\left(\frac{1}{n_k}\sum_{p=1}^{n_k}\|x_p^{(k)}-x_i^{(j)}\|^2\right) \tag{3}$$

其中:k 和 j 表示类标,$x_i^{(j)}$ 表示第 j 类的第 i 个样本;$x_p^{(k)}$ 表示第 k 类的第 p 个样本;n_k 表示第 k 类中的样本个数;$\|\cdot\|^2$ 表示平方欧氏距离。

定义 2　定义第 j 类的第 i 个样本的类内距离 $w(j,i)$ 为该样本到第 j 类中其他所有样本的平均距离,即

$$w(j,i)=\frac{1}{n_j-1}\sum_{l,q=1,q\neq i}^{n_j}\|x_q^{(j)}-x_i^{(j)}\|^2 \tag{4}$$

其中:$x_q^{(j)}$ 表示第 j 类中的第 q 个样本,且 $q\neq i$;n_j 表示第 j 类中的样本个数。

定义 3　定义第 j 类的第 i 个样本的聚类距离 baw(j,i) 为该样本的最小类间距离和类内距离之和,即

$$\text{baw}(j,i)=b(j,i)+w(j,i) \tag{5}$$

定义 4　定义第 j 类的第 i 个样本的聚类离差距离 bsw(j,i) 为该样本的最小类间距离和类内距离之差,即

$$\text{bsw}(j,i)=b(j,i)-w(j,i) \tag{6}$$

定义 5　定义第 j 类的第 i 个样本的类间类内划分(BWP)指标 BWP(j,i) 为该样本的聚类离差距离和聚类距离的比值,即

$$\text{BWP}(j,i)=\frac{\text{bsw}(j,i)}{\text{baw}(j,i)} \tag{7}$$

1.2　基于 BWP 指标的 K-means 最佳聚类数确定方法

基于 BWP 指标,周世兵等人提出了 K-means 最佳聚类数的确定方法——K-means 遍历算法,如式(8)(9)所示。

$$\text{avg}_{\text{BWP}}(k)=\frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\text{BWP}(j,i) \tag{8}$$

$$k_{\text{opt}}=\arg\max_{2\leqslant k\leqslant k_{\max}}\{\text{avg}_{\text{BWP}}(k)\} \tag{9}$$

其中:$\text{avg}_{\text{BWP}}(k)$ 表示数据集聚成 k 类时的平均 BWP 指标值;k_{opt} 表示最佳聚类数;$k_{\max}$ 表示聚类算法搜索的上限。

图 1 显示了对 SM2 数据集进行 K-means 聚类,BWP 指标的变化曲线。SM2 数据集由中心分别为(0,0),(5,5),(10,10),(15,15)的二维四高斯分布数据组成,每个类有 225 个样本,每个类的协方差矩阵为 $2I_2$,该数据集的结构特征为某些类的类间距离很近且存在大量重叠的聚类结构。实验中聚类数的搜索范围为$[2,k_{\max}]$,根据普遍使用的经验规则 $k_{\max}\leqslant\sqrt{n}$,取 $k_{\max}=\text{Int}(\sqrt{n})$。由图 1 可知 BWP 指标在聚类数为 4 时达到最大值,且有很高的区分度,因此利用 BWP 指标可以较好地判断聚类数。图 2 显示了聚类数为 4 时的聚类效果。

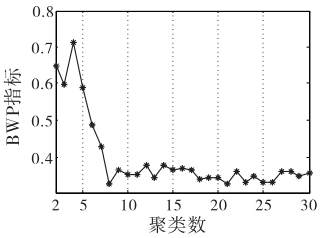

图1　SM2数据集聚类数—BWP指标关系

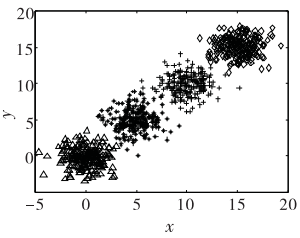

图2　k=4时数据集SM2的聚类结果

由图 1 和 2 可以看出,这种方法能较好地确定最佳聚类数,但需要遍历区间$[2,k_{\max}]$的所有值,当观测信号样本较大时,其计算量将非常大。

2　改进的最佳聚类数确定算法

2.1　粒子群算法原理

粒子群算法(PSO)是模拟鸟群行为规律而提出的一种基于群体智能优化算法。通过群体中粒子个体间的合作与竞争产生的群体智能优化搜索,每代种群的解都具有学习自身最优和群体最优的特点,每个粒子用其位置和速度向量表示。在迭代寻优过程中,每个粒子参考自己既定飞行方向、所经历的最优方向和整个鸟群的最优方向确定自己的飞行,更新速度和位置为

$$v_i^{t+1}=wv_i^t+c_1r_1(\text{pbest}_i^t-x_i^t)+c_2r_2(\text{gbest}_i^t-x_i^t) \tag{10}$$

$$x_i^{t+1}=x_i^t+v_i^{t+1} \tag{11}$$

其中:惯性权重系数 $w=w_{\max}-\dfrac{w_{\max}-w_{\min}}{m}\times k$ 为随时间线性减少函数,它可使得 PSO 在开始时探索较大的区域,较快地定位最优聚类的大致位置,随着 w 逐渐减小,粒子速度减慢,开始精细地局部搜索[6];$w_{\max}$、$w_{\min}$ 分别为初始权重和最终权重;m 为最大迭代次数;k 为当前迭代次数;v_i^t 是粒子的速度向量;x_i^t 是粒子当前位置;r_1、r_2 是[0,1]间的随机数,以保持群体的多样性;c_1、c_2 为学习因子,具有自我总结和向群体最优个体学习的能力;pbest_i^t 是粒子自身所找到的最好解的位置;gbest_i^t 是整个种群找到的最优解的位置。

2.2　粒子群算法改进最佳聚类数确定算法

由第 1 章的分析可知采用 K-means 遍历法存在计算量大、消耗时间长的缺点。而 PSO 算法不用遍历所有可能解,使其具备减少计算量、快速获得最优解的能力。下面使用 PSO 算法改进 K-means 遍历算法。

首先,对粒子群作如下设定:

a)每个粒子是一维的,其位置对应聚类数 k。

b)数据集聚成 k 类时的平均 BWP 指标对应粒子的适应度。

c)粒子的移动范围为$[2,k_{\max}]$。由于采用式(11)得到的位置是非整数的,而 K-means 聚类数 k 是整数,因此要将每次更新得到的新位置按式(12)作近似处理。

$$x_i^{t+1} = \text{round}(x_i^t + v_i^{t+1}) \tag{12}$$

d)粒子群的大小可视观测信号的数据量大小设定,一般设为 10 即可。

其次,对于同样的聚类数,K-means 算法多次聚类得到的平均 BWP 指标是相近的[5],而粒子群算法中的不同粒子可能会多次更新到同一个位置,这也就意味着对同一个聚类数要进行多次聚类,这将消耗大量的时间。对此作如下改进:将每个粒子更新过的位置及此位置的适应度保存至一个全局变量,在计算每个粒子的适应度之前,先查询全局变量中是否已包含此次粒子当前位置的数据。如果包含,则直接取此变量中对应的值;反之则继续计算。改进的求取最佳聚类数算法的伪代码如下:

```
初始化粒子群;
初始化全局变量 ListGB
do
    for 每个粒子
        if( ListGB 中包含粒子的位置 x_i^t )
            取 ListGB 中此位置的适应度赋给粒子;
        else
            计算其适应度;
        将粒子的位置和适应度保存至 ListGB;
      end
      if( 适应度优于粒子历史最佳值)
          用 x_i^t 更新历史个体最佳 pbest_i^t;
        end
    end
    选取当前粒子群中最佳粒子;
    if( 当前群最佳粒子优于群历史最佳粒子)
        用当前最佳粒子更新 gbest_i^t;
      end
      for 每个粒子
        按式(10)更新粒子速度;
        按式(11)更新粒子位置,并对其取整;
    end
  while 最大迭代数未达到或最小误差未达到
```

以上伪代码最终获得的群最佳粒子的位置即为最佳聚类数。本文称改进的最佳聚类数求取算法为 PSO-Kmeans 算法。

3 基于 PSO-Kmeans 算法的 UBSS

3.1 观测信号初始化

当源信号为稀疏信号时,混合信号具有线性聚类特性,如果忽略噪声的影响,将式(1)展开为

$$\begin{pmatrix} x_1(t) \\ \vdots \\ x_m(t) \end{pmatrix} = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} s_1(t) + \cdots + \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} s_n(t) \tag{13}$$

当某一时刻只有一个原信号(如只有 $s_i(t)$)起作用时,式(13)可转换为

$$\begin{pmatrix} x_1(t) \\ \vdots \\ x_m(t) \end{pmatrix} = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix} s_1(t) \Rightarrow \frac{x_1(t)}{a_{1i}} = \cdots = \frac{x_m(t)}{a_{mi}} = s_i(t) \tag{14}$$

它在 m 维空间中是一条经过原点的直线,直线方向取决于混叠矩阵的第 i 个列矢量 $(a_{1i}, \cdots, a_{mi})^{\mathrm{T}}$。基于此,$n$ 个源信号 $s_i(t)$,$i \in (1,2,\cdots,n)$将确定 n 条直线,因此,A 的估计就转换为观测空间中直线方向的估计。采用文献[7]的四路长笛音信号,左乘以随机产生的 2×4 维混叠矩阵得到两路观测信号。由于长笛音信号在时域稀疏性不明显,而频域稀疏性比较明显,故将两路观测信号进行傅里叶变换。变换后的混叠信号散点图如图 3 所示。由图 3 可见,四路源信号确定了四条直线。

3.2 估计混叠矩阵

由于 n 路源信号经混叠后的散点图表现为经过原点的 n 条直线,同时各直线的方向对应于混叠矩阵的各列向量。通过聚类算法将观测信号按散点图中各直线所在方向进行分类,每一类对应一条直线,每个聚类中心表示的方向近似散点图中一条直线的方向,即对应混叠矩阵的某个列向量,这样通过聚类就可以获得对混叠矩阵的估计。

本文考虑首先将直线聚类转变成致密聚类,对观测信号进行标准化处理,包括尺度归一化、方向镜像[8]和计算到点(1,0)的弧线距离。对观测信号 $X(t)$ 的尺度归一化为

$$X'(t) = X(t) / \| X(t) \| \tag{15}$$

其中:$\| X(t) \|$ 表示 $X(t)$ 的欧几里德范数。方向镜像就是将下半平面向量映射到上半平面。进行标准化时,由于位于原点附近的数据点存在重叠现象或为噪声点,可去掉这部分数据点,通过去除这部分数据可以有效降低计算量和提高估计精度。经过尺度归一化、方向镜像后的观测信号均位于上半圆周上,如图 4 所示。

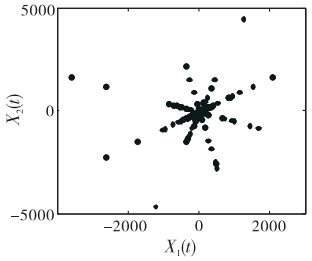

图3 两路观测信号散点图

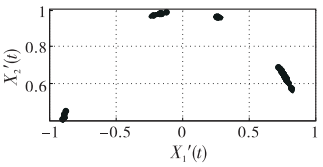

图4 尺度归一化和方向镜像处理后的观测信号

然后计算尺度归一化和方向镜像处理后的观测信号到点(1,0)的弧线距离 $d(t)$,如式(16)所示。

$$d(t) = \begin{cases} \arctan\left(\dfrac{X'_2(t)}{X'_1(t)}\right) & X'_1(t) > 0 \\ \arctan\left(\dfrac{X'_2(t)}{X'_1(t)}\right) + \pi & X'_1(t) < 0, t = 1,2,\cdots,T \\ \pi/2 & X'_1(t) = 0 \end{cases} \tag{16}$$

因为属于同一直线的点属于同一类,因此属于同一类的观测信号到(1,0)的弧线距离相差不大。图 5 显示了进行尺度归一化和方向镜像处理后的观测信号到(1,0)弧线距离分布。

最后对标准化处理后的观测信号采用 PSO-Kmeans 算法获得最佳聚类数 k 和对应的聚类中心集合 $C = \{c_1, c_2, \cdots, c_k\}$。因为单位圆中弧线的长度等于弧度,所以每个聚类中心与混叠矩阵 A 的某一列共线,故 A 的某一列可由式(17)求得。

$$a_i = [\cos(c_i), \sin(c_i)]^{\mathrm{T}} \quad i = 1,2,\cdots,k \tag{17}$$

3.3 源信号的恢复

估计出混叠矩阵后,可通过最短路径法进行源信号的恢复。稀疏信号盲源分离归结为以下优化问题:

$$\min_{A,S} \frac{1}{2\sigma^2} \| AS - X \|^2 + \sum_{i,t} |s_i(t)| \tag{18}$$

其中:σ^2 为噪声的方差,第一项为重构误差平方和,第二项为非稀疏项。在不考虑噪声的前提下,由于混叠矩阵 A 已估计出来,则式(18)等效为

$$\begin{cases} \min\limits_{s(t)} \sum\limits_{i}^{n} |s_i(t)| \\ As(t) = x(t) \quad t = 1,2,\cdots,T \end{cases} \tag{19}$$

由式(19)可知,每个时刻 t 确定一个优化问题,从而可估计出源信号。

3.4　欠定盲分离算法步骤

通过上述分析,本文的欠定盲分离算法步骤如下:a)将观测信号标准化; b)采用 PSO-Kmeans 算法求得标准化处理后的观测信号的最佳聚类数即源信号数;c)取最佳聚类数对应的中心利用式(17)计算分离矩阵 A;d)利用最短路径法恢复源信号。

4　实验分析

实验仿真均在 MATLAB 7.12 环境下进行,采用六路长笛信号作为源信号[7],产生两路混叠信号,即 $m=2$,$n=6$,选取混叠矩阵为

$$A=\begin{bmatrix} 0.7660 & 0.5000 & 0.2588 & -0.1736 & -0.7071 & -0.9063 \\ 0.6428 & 0.8660 & 0.9659 & 0.9848 & 0.7071 & 0.4226 \end{bmatrix}$$

使用 3.2 节的方法对观测数据进行标准化处理,得到标准化处理后的观测信号如图 6 所示。

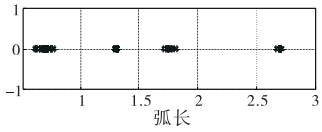

图5　标准化处理后的观测信号

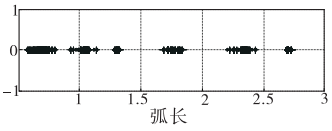

图6　标准化处理后的观测信号

标准化处理后的观测信号共有 6 405 个数据点,则聚类上限 $k_{\max}=\mathrm{Int}(\sqrt{6405})\approx 80$。设粒子数为 10,每个粒子搜索范围为[2,80],最大迭代次数为 50,初始权重 $w_{\max}=0.9$,最终权重 $w_{\min}=0.4$,学习因子 $c_1=1.5$,$c_2=2$。然后采用 PSO-Kmeans 聚类法求得最佳聚类数为 6,对应的 $\mathrm{avg}_{\mathrm{BWP}}$ 为 0.958 4,得到聚类结果如图 7 所示。

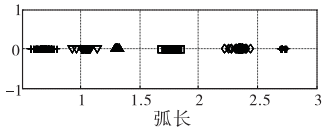

图7　标准化处理后的观测信号的聚类结果

然后计算估计混叠矩阵,得到

$$\hat{A}=\begin{bmatrix} 0.7703 & 0.5050 & 0.2601 & -0.1893 & -0.7016 & -0.9073 \\ 0.6377 & 0.8630 & 0.9656 & 0.9819 & 0.7126 & 0.4205 \end{bmatrix}$$

为了说明本文算法在估计混叠矩阵的优势,下面给出聚类个数,即假设源信号数目已知的条件下,利用常规 K-means 聚类算法估计得到混叠矩阵为

$$\tilde{A}=\begin{bmatrix} 0.7956 & 0.5317 & 0.3477 & -0.0782 & -0.6078 & -0.9113 \\ 0.6052 & 0.8643 & 0.7636 & 0.9726 & 0.7059 & 0.4624 \end{bmatrix}$$

通过式(20)分别计算 $\hat{A}$、$\tilde{A}$ 与 A 对应列向量的夹角,以此来检验估计混叠矩阵的精度[9]。

$$\mathrm{ang}(a,\hat{a})=\frac{180}{\pi}\arccos\left\{\frac{\langle a,\hat{a}\rangle}{\|a\|\cdot\|\hat{a}\|}\right\} \tag{20}$$

其中:a 代表原混叠矩阵 A 中的列向量,$\hat{a}$ 代表通过上述算法估计得到的混叠矩阵 $\hat{A}$ 中对应的列向量。估计混叠矩阵列向量与原混叠矩阵列向量之间的夹角如表 1 所示,可以看出本文算法相比常规 K-means 聚类有较大优势。

表 1　混叠矩阵与估计混叠矩阵对应列向量之间的夹角　/°

算法	混叠矩阵与估计混叠矩阵对应列向量之间的夹角					
	$\langle a_1,\hat{a}_1\rangle$	$\langle a_2,\hat{a}_2\rangle$	$\langle a_3,\hat{a}_3\rangle$	$\langle a_4,\hat{a}_4\rangle$	$\langle a_5,\hat{a}_5\rangle$	$\langle a_6,\hat{a}_6\rangle$
算法	0.3797	0.0748	0.9154	0.1333	0.3422	0.4447
常规 K-means 聚类	2.7439	1.5991	9.4873	5.4032	4.2728	1.9053

另外,本文对比 K-means 遍历算法与 PSO-Kmeans 算法在确定最佳聚类数即源信号数时所耗时间与平均内存,分别如图 8 和 9 所示。由图 8 可以看出,随着聚类数上限的增加,K-means 遍历算法所耗时间呈直线上升趋势,而本文算法的增长比较缓慢,在聚类数上限较大即观测信号样本量较大时,本文算法可节约大量时间。由图 9 可以看出,在聚类数上限较小时,两种算法的平均内存消耗差不多,但聚类数较大时,K-means 遍历算法的平均消耗内存呈增长趋势。

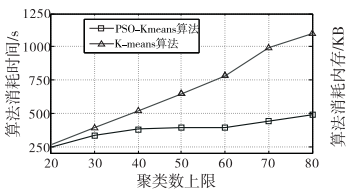

图8　消耗时间对比

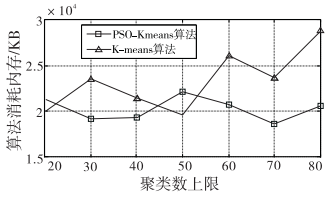

图9　消耗平均内存对比

图 10 ~ 12 分别显示了六路源信号图形、两路混合信号和采用本文算法恢复的六路信号波形。

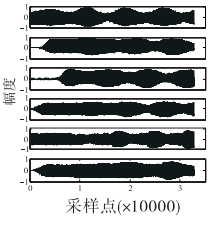

图10　六路源信号

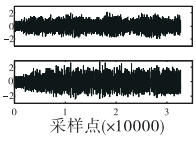

图11　两路混合信号

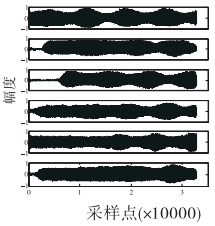

图12　六路恢复信号

5　结束语

本文研究了评价聚类结果优劣的 BWP 指标和 K-means 遍历算法,结合粒子群算法,提出了一种改进的源信号数目估计算法——PSO-Kmeans 算法,并将 PSO-Kmeans 算法引入欠定盲分离。文中详细介绍了采用本文算法的分离流程,最后进行了六路源信号两路观测信号的欠定盲分离实验,与常规 K-means 算法对比了分离精度,并与 K-means 遍历法对比了获得源信号数所用时间和内存。实验结果表明,本文算法在保证恢复精度的同时能缩短分离时间,并可节省一定的内存,在观测信号数据量大时,这种优势更加明显。

参考文献:

[1] XIE Sheng-li, YANG Liu, YANG Jun-mei, *et al*. Time-frequency approach to underdetermined blind source separation[J]. IEEE Trans on Neural Networks and Learning Systems, 2012, 23(2): 306-315.

[2] 谢胜利,孙功宪,肖明,等. 欠定和非完全稀疏性的盲信号提取[J]. 电子学报,2010,38(5):1028-1031.

[3] ZIBULLEVSKY M, PEARLMUTTER B A. Blind source separation by sparse decomposition in a signal dictionary[J]. Neural Computation, 2001, 13(4): 863-882.

[4] LI Yuan-qing, CICHOCKI A, AMARI S. Analysis of sparse representation and blind source separation[J]. Neural Computation, 2004, 16(6): 1193-1234.

[5] 周世兵,徐振源,唐旭清. K-means 算法最佳聚类数确定方法[J]. 计算机应用,2010,30(8):1995-1998.

[6] 李洪兵,余成波,闫俊辉,等. 基于改进粒子群聚类的无线传感器网络能量均衡分簇策略[J]. 计算机应用研究,2011,28(2):657-660.

[7] 谭北海,杨祖元,周郭许,等. 欠定盲分离中源的个数估计和分离算法[J]. 中国科学,F 辑:信息科学,2009,39(3):349-356.

[8] HE Zhao-shui, XIE Sheng-li, FU Yu-li. Sparse representation and blind source separation of ill-posed mixtures[J]. Science in China Series F: Information Sciences, 2006, 49(5): 639-652.

[9] BOFILL P, ZIBULEVSKY M. Underdetermined blind source separation using sparse representations[J]. Signal Process, 2001, 81(11): 2353-2362.