

文章编号: 2095-2163(2023)03-0246-04

中图分类号: TP391

文献标志码: A

一种基于蜂群优化算法的 ICA-R 算法

张守成

(广西民族师范学院 数理与电子信息工程学院, 广西 崇左 532200)

摘要: 基于峭度的参考独立分量分析算法存在阈值参数难以确定的问题, 通过对该算法的理论分析, 以输出信号峭度的绝对值和接近性量度函数的乘积作为新的目标函数, 并采用蜂群优化算法处理目标函数, 避免了人为选择阈值参数和步长的问题。针对合成数据的仿真实验表明了所提算法具有良好的分离效果和稳定性。

关键词: 蜂群优化算法; 峭度; 独立分量分析

An ICA-R algorithm based on bee colony optimization algorithm

ZHANG Shoucheng

(College of Mathematics, Physics and Electronic Information Engineering, Guangxi Normal University for Nationalities, Chongzuo Guangxi 532200, China)

【Abstract】 To solve the problem that it is difficult to determine the threshold parameter of reference independent component analysis algorithm based on kurtosis, through the theoretical analysis of this algorithm, the product of the absolute value of output signal kurtosis and the measure function of proximity is taken as the new objective function, and the swarm optimization algorithm is adopted to process the objective function, avoiding the problem of artificial selection of threshold parameters and step size. The simulation results of synthetic data show that the proposed algorithm has good separation effect and stability.

【Key words】 artificial bee colony optimization algorithm; kurtosis; independent component analysis

0 引言

独立分量分析 (Independent Component Analysis, ICA) 是 20 世纪 90 年代后期发展起来的一项新的信号分解技术, 目前已成功应用在语音分析、特征提取、生物医学工程等领域^[1-2]。传统 ICA 方法以分离出所有源信号为目标, 往往同时分离出所有的源信号^[3-6], 而在很多实际应用中, 可能仅对其中某一个源信号感兴趣。

在此背景下, Lu 等学者^[7-8]提出了参考独立成分分析 (ICA-R) 算法, 利用期望提取信号的波形先验信息构造出参考信号, 将其引入到传统的 ICA 算法中, 通过优化算法提取出在一定量度下与参考信号最接近的输出作为期望源信号, 具有直接高效的优点, 并且参考信号的引入能提高期望源信号的分离质量和精度。但该方法计算相对繁琐。本文通过对 ICA-R 算法的原理进行分析, 利用峭度作为度

量, 巧妙调整目标函数, 并利用蜂群优化算法处理目标函数, 得到一种计算简单、分离性能好的 ICA-R 算法。最后通过仿真实验证明了该算法的有效性。

1 蜂群优化算法

德国生物学家 Frisch 发现在自然界中, 虽然各组织内不同职能的蜜蜂只能完成单一的任务, 但蜜蜂通过摇摆舞、气味等多种信息交流方式, 使得整个蜂群能够迅速地发现优良蜜源, 实现自组织行为^[9]。Seeley 在 1995 年最先提出了蜂群的自组织模拟模型; Teodorovic 于 2003 年进一步提出了蜂群优化算法 (Artificial Bee Colony, ABC)^[10-12]。ABC 算法是一种新兴的仿生智能优化算法, 具有全局收敛能力强、设置参数少、适用范围广等特点, ABC 已被应用于许多科学与工程领域。例如, 文献[11]将 ABC 算法应用于盲信号分离代替了传统的梯度法, 克服了梯度法收敛速度慢、易陷入局部极值点等缺

基金项目: 广西民族师范学院引进副教授科研启动项目 (2020FG003)。

作者简介: 张守成 (1980-), 男, 硕士, 副教授, 主要研究方向: 信号与信息处理。

通讯作者: 张守成 Email: 29245340@qq.com

收稿日期: 2022-05-01

点,提高了盲源分离算法的分离精度和稳定性。

2 基于蜂群优化算法的ICA-R算法

线性瞬时混合ICA模型通常表示为^[1]: $x(t) = \mathbf{A}s(t)$, 其中 $s(t) = [s_1(t), \dots, s_M(t)]^T$ 是 M 个假设为统计独立的未知源信号, \mathbf{A} 是一个未知的 $N \times M$ 混合矩阵 (一般假设 $N \geq M$), $\mathbf{x}(t) = [x_1(t), \dots, x_N(t)]^T$ 是源信号经 \mathbf{A} 线性混合后得到的 N 个观测信号。

ICA的目标是估计一个最佳的 $M \times N$ 解混矩阵 \mathbf{W} , 使得观测信号经过 \mathbf{W} 变换后得到的 M 个输出信号 $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$ 是独立的, 从而作为源信号 $s(t)$ 的估计。一般情况下不可能直接求得解混矩阵的解, 需要通过基于目标函数的自适应方法求解, 具体过程是: 首先对实际问题的源信号和信道传输过程进行分析, 然后根据源信号的随机特征确定度量输出源分离结果的测度, 如非高斯性、稀疏性等测度方法, 根据测度理论构造目标函数, 并选择合适的优化方法。常用的优化算法有基于梯度的随机梯度算法和牛顿方法等。在进行优化后, 最终获得解混矩阵, 输出估计的源信号。因此ICA的核心问题可以表示为: ICA方法 = 目标函数 + 优化算法^[12]。

而ICA-R算法一般利用期望源信号的波形先验信息构造出其参考信号 $r(t)$, 随后将输出信号 $y(t)$ 与 $r(t)$ 的某个接近性量度 $\varepsilon(y, r)$ 作为先验约束引入到传统ICA的学习中, 算法收敛后得到一个最优权向量 \mathbf{w}^* (解混矩阵 \mathbf{W} 的一行), 将 $\mathbf{y} = \mathbf{w}^{*T}\mathbf{x}(t)$ 作为期望源信号的估计信号。

以上就是ICA-R算法的基本原理, 文献[7]选择负熵作为信号非高斯性的度量, 并据此构造目标函数, 利用拟牛顿方法得到了一种ICA-R算法。但由于涉及复杂的二阶导数计算, 效率较低。

峭度作为信号非高斯性的度量也被广泛使用, 基于ICA-R算法的原理, 本文选取输出信号峭度的绝对值 $K(y) = |E(y^4) - 3[E(y^2)]^2|$ 作为输出信号非高斯型的度量, 假定 $K(y)$ 有 M 个局部或全局最优解 $w_i (i = 1, \dots, M)$, 相应给出 M 个源信号的估计。则输出信号与参考信号的接近性量度 $\varepsilon(y, r)$ 当且仅当输出 $\mathbf{y} = \mathbf{w}^{*T}\mathbf{x}$ 是期望源信号时取得最小值, 即有:

$$\varepsilon(\mathbf{w}^{*T}\mathbf{x}, r) < \varepsilon(\mathbf{w}_1^T\mathbf{x}, r) \leq \dots \leq \varepsilon(\mathbf{w}_{M-1}^T\mathbf{x}, r) \quad (1)$$

由此可知, 一定存在一个阈值参数 ξ 满足以下不等式 $\xi \in [\varepsilon(\mathbf{w}^{*T}\mathbf{x}, r), \varepsilon(\mathbf{w}_1^T\mathbf{x}, r)]$, 当且仅当输

出信号 $\mathbf{y} = \mathbf{w}^{*T}\mathbf{x}$ 时使得 $g(w) = \varepsilon(y, r) - \xi \leq 0$ 成立。将 $g(w) \leq 0$ 作为一个对 $K(y)$ 的约束, 则算法的优化框架为:

(1) 最大化。研究推得的公式为:

$$K(y) = |kurt(y)| = |E(y^4) - 3[E(y^2)]^2| \quad (2)$$

(2) 约束。研究推得的公式为:

$$g(w) \leq 0, h(w) = E(y^2) - 1 = 0 \quad (3)$$

其中, 等式约束 $h(w) = 0$ 用于保证 $K(y)$ 和 w 有界。

将观测数据 \mathbf{x} 进行白化, 用 $\tilde{\mathbf{x}}$ 表示白化后数据, 并对权向量 \mathbf{w} 每次更新后归一化。则有 $E(y^2) = E(\mathbf{w}^T\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\mathbf{w}) = \mathbf{w}^T\mathbf{R}_{\tilde{\mathbf{x}}}\mathbf{w} = \|\mathbf{w}\|^2 = 1$, 其中 $\mathbf{R}_{\tilde{\mathbf{x}}} = \mathbf{I}$ 是 $\tilde{\mathbf{x}}$ 的协方差矩阵。上式保证了 $h(w) = E(y^2) - 1 = 0$ 恒成立, 从而算法的优化框架简化为:

(1) 最大化。研究推得的公式为:

$$K(y) = |kurt(y)| = |E(y^4) - 3| \quad (4)$$

(2) 约束。研究推得的公式为:

$$g(w) \leq 0 \quad (5)$$

由ICA-R算法设计的基本原理可知, 在约束 $g(w) \leq 0$ 下最大化 $K(y) = |E(y^4) - 3|$ 得到最优权向量 \mathbf{w}^* , 此时接近性量度 $\varepsilon(y, r)$ 也取得最小值, 则 $-\varepsilon(y, r)$ 取得最大值, 从而可将最大化的目标函数调整为 $K(y)$ 和 $-\varepsilon(y, r)$ 的乘积, 即:

$$M(y) = K(y) \cdot [-\varepsilon(y, r)] = -|E(y^4) - 3| \cdot \varepsilon(y, r) \quad (6)$$

其中, $\varepsilon(y, r) = E[(y - r)^2]$ 。

通过最大化 $M(y)$ 得到最优权向量 \mathbf{w}^* 的优化算法选择较多, 鉴于ABC算法具有全局收敛性强、收敛速度快的特点, 采用ABC算法对目标函数 $M(y)$ 进行优化得到最优权向量 \mathbf{w}^* , 就可以实现目标源信号的提取。

3 仿真实验与性能分析

为了验证本文改进算法的有效性, 对合成数据进行仿真实验。取4个零均值单位方差的独立源信号: S_1, S_2 为确定性信号, S_3, S_4 为随机信号。样本数据为2000个, 经白化后的源信号如图1所示 (为方便观察, 只绘出前300个样本点)。用一个随机生成的高斯矩阵对源信号进行混合, 得到的混合信号如图2所示。用各源信号相应的符号函数序列作为提取的对应参考信号, 如图3所示。采用本文所提算法对各源信号进行抽取, 收敛后抽取的信号如图4所示。直观上比较图4与图1, 波形基本一致, 表明算法很好地实现了对各个信号的盲抽取。

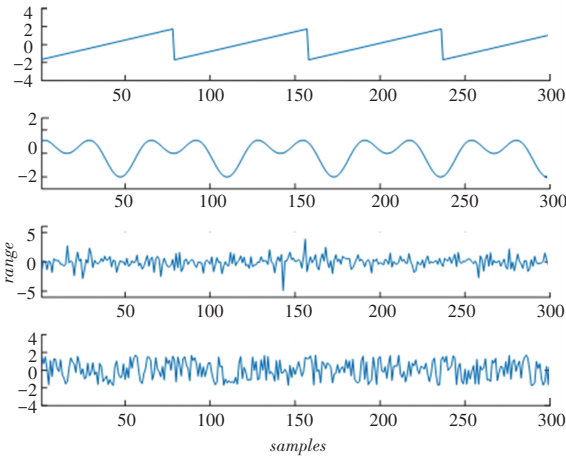


图1 源信号
Fig. 1 Source signal

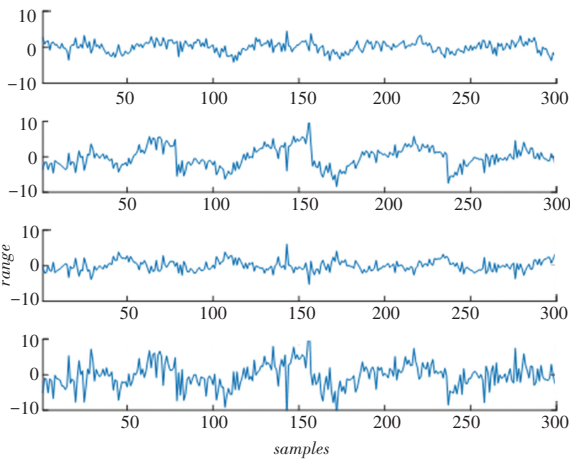


图2 混合信号
Fig. 2 Mixed signal

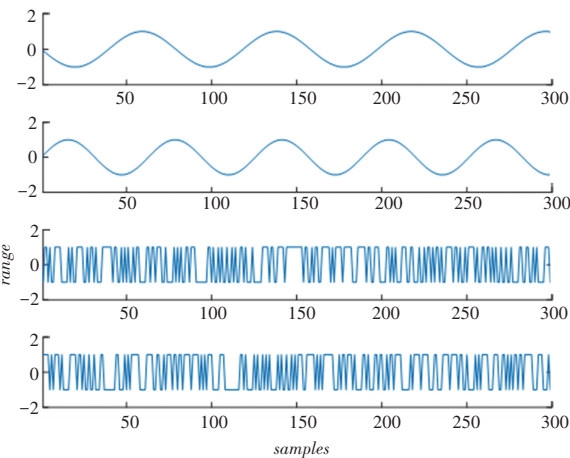


图3 参考信号
Fig. 3 Reference signal

为进一步说明分离效果,采用量化指标 *IPI* (Individual Performance Index) 对分离性能进行评价^[1]:

$$IPI = \sum_{j=1}^M |p_j| / \max_k |p_k| - 1 \quad k = 1, \dots, M \quad (7)$$

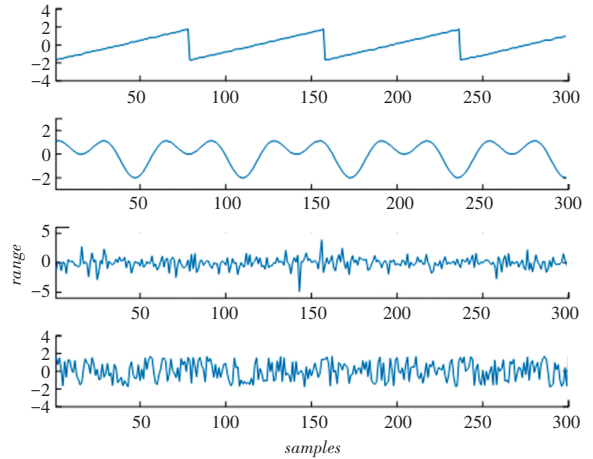


图4 抽取信号
Fig. 4 Extracted signal

其中, p_j 表示全局分离变量 $p = w^T A$ 的第 j 个元素。*IPI* 越接近于零,表明算法分离性能越好。

对上述信号采用本文算法进行 150 次蒙特卡洛实验,每次都随机产生混合矩阵。实验结果显示,算法每次都能抽取成功,表明算法在不需要阈值参数选择的情况下收敛性能稳定。

150 次提取信号后的平均 *IPI* 值见表 1。

表 1 算法收敛的平均 *IPI* 值

Tab. 1 Average *IPI* of the algorithm

源信号	S_1	S_2	S_3	S_4
抽取 <i>IPI</i>	0.016	0.021	0.032	0.028

表 1 显示在成功抽取时采用本文算法有低 *IPI* 值,说明算法对各源信号的抽取性能优异。

4 结束语

通过巧妙调整目标函数并采用蜂群优化算法有效避免了阈值选择问题,简化了计算。计算机仿真实验表明了本文算法收敛性能稳定,可有效实现对信号的盲提取。当然,对实际应用中不同的提取信号,如何构造合适的参考信号是需要进一步探讨的课题。

参考文献

[1] HYVÄRINEN A, KARHUNEN J, OJA E. Independent component analysis[M]. USA: John Wiley&Sons, Inc., 2001.

[2] HYVÄRINEN A, OJA E. A fast fixed-point algorithm for independent component analysis[J]. Neural Computation, 1997, 9(7):1483-1492.

[3] HYVÄRINEN A. Fast and robust fixed-point algorithms for independent component analysis[J]. IEEE Transactions on Neural Networks, 1999, 10(3):626-634.

[4] BELL A J, SEJNOWSKI T J. An information-maximization approach to blind separation and blind deconvolution[J]. Neural

Computation, 1995, 7(6): 1129-1159.

- [5] 陈莹,王法松,王忠勇,等. 基于改进ICA-R算法的多用户信号盲提取[J]. 通信技术, 2016, 49(03): 282-285.
- [6] 赵民全. 基于负熵最大化判据的ICA算法研究与应用[J]. 舰船电子工程, 2021, 41(02): 36-38, 68.
- [7] LU Wei, RAJAPAKSE J C. ICA with reference [J]. Neurocomputing, 2006, 69(16-18): 2244-2257.
- [8] LU Wei, RAJAPAKSE J C. Approach and applications of constrained ICA [J]. IEEE Transactions on Neural Networks, 2005, 16(1): 203-212.

- [9] 薛羽. 仿生智能优化算法及其应用研究[D]. 南京:南京航空航天大学, 2013.
- [10] KARABOGA D, GORKEMLI B, OZTURK C, et al. A comprehensive survey: Artificial bee colony (ABC) algorithm and applications [J]. Artificial Intelligence Review, 2012, 42(1): 21-57.
- [11] 张银雪, 田学民, 邓晓刚. 基于改进人工蜂群算法的盲源分离方法[J]. 电子学报, 2012, 40(10): 2026-2030.
- [12] 赵挺. 蜂群算法及其仿生策略研究[D]. 杭州:浙江大学, 2016.

(上接第245页)

在用户消费数据可视化界面中,涉及用户交易信息、交易量排名、商品分类占比、商品销量排行、各平台占比以及城市排行可视化功能板块,可从多个维度了解用户的消费数据。实现了对用户准确推送其感兴趣内容,到后来电商行业借助社交平台的画像能力展开营销推广,再转变为电商行业自行根据用户消费习惯、消费能力进行主动画像与营销。

电商可根据用户消费习惯、消费能力进行主动画像与营销。在了解客户需求和消费能力,以及客户信用额度的基础上,寻找潜在产品的目标客户,并利用画像信息为客户开发产品。只有对用户的各个行为指标进行有效的分析,才能设计出让用户满意的产品,最终创建出精准的用户画像。

5 结束语

当今的社会,随着信息技术的发展,数据呈现出指数级增长,利用用户画像技术来解决信息匹配问题不仅是在电商领域,还可以在越来越多的领域中得到应用。本文围绕着大数据技术构建用户画像系统,通过抓取网上客户数据进行抓取,然后进行数据去重、清洗和标准化等一系列数据处理后,根据实际商业业务需求构建系统框架。将系统各个组件整合

起来,完成搭建工作,并对数据进行处理、再计算出标签,最后根据相关算法模型构建出用户画像模型,从而为下一步的信息匹配、推荐等业务奠定良好基础。

参考文献

- [1] 王尚超. 基于大数据的企业用户画像系统设计研究[J]. 电子技术与软件工程, 2021(22): 157-158.
- [2] 刘海鸥,孙晶晶,苏妍妍,等. 国内外用户画像研究综述[J]. 情报理论与实践, 2018, 41(11): 155-160.
- [3] 刘茗. 大数据时代电子商务精准营销实现个性化推荐研究[J]. 现代营销(下旬刊), 2022(09): 164-166.
- [4] 刘姜. 基于大数据的用户画像构建方法与运用探究[J]. 科技创新与应用, 2022, 12(09): 139-142.
- [5] 于欢,孙青,易静,等. 基于大数据技术的高校图书馆智慧服务模式研究[J]. 江苏科技信息, 2022, 39(31): 45-48.
- [6] 刘学太,李阳,巴志超,等. 数据驱动环境下数据画像若干问题探析[J]. 情报理论与实践, 2022, 45(04): 87-94.
- [7] 刘相金,王梦菊. 大数据环境下基于用户画像的学术文献推荐系统研究[J]. 河南图书馆学刊, 2019, 39(12): 113-114, 119.
- [8] 孙秋莲. 基于大数据的农产品电商用户画像系统的设计和应用[J]. 山西农经, 2019(06): 96.
- [9] 许鹏程,毕强,张晗,等. 数据驱动下数字图书馆用户画像模型构建[J]. 图书情报工作, 2019, 63(03): 30-37.
- [10] 冯宪伟. 基于大数据分析的电商企业精准营销研究[J]. 济南职业学院学报, 2022(05): 89-92.