

基于 ICA 和聚类的 EEG 脑源定位研究

文 念, 黄丽亚, 于 涵, 杨 晨, 杨俊宇
(南京邮电大学 电子科学与工程学院, 江苏 南京 210003)

摘 要:将独立成分分析(ICA)与聚类 Cluster 相结合应用到脑电的多偶极子源定位,先采用快速独立成分分析(fastICA)得到多个独立成分(ICs),然后通过聚类方法对得到的 ICs 进行特征提取和有效归类。该方法在去除脑电伪迹和噪声干扰的同时解决了 ICA 分解后独立成分的选取依赖于经验的局限性。以上消除了伪迹干扰和 ICs 的不确定性选择对源定位性能的影响,为源定位创造了有利条件。对 ICs 进行定位也使得整个定位过程像单偶极子定位一样稳定简单。仿真实验中源定位效果得到改善,表明了该方法的有效性。

关键词:脑电;独立成分分析;多偶极子源定位;聚类分析

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2015)05-0228-05

doi:10.3969/j.issn.1673-629X.2015.05.053

EEG Sources Localization Based on Independent Component Analysis and Clustering

WEN Nian, HUANG Li-ya, YU Han, YANG Chen, YANG Jun-yu

(School of Electronic Science and Engineering, Nanjing University of Posts and Telecommunications,
Nanjing 210003, China)

Abstract: A multiple-dipole localization procedure combined Independent Component Analysis (ICA) and cluster analysis for EEG (electroencephalograph) is suggested in this paper. Firstly, the fastICA is applied to obtain Independent Components (ICs) which are then utilized by cluster analysis for feature extraction and effective classification. Removing the brain artifacts and noise by cluster, this approach overcomes the constraint that the selection of the independent components depends on experiences. The work above eliminates the artifact and the uncertainty of ICs effects on the source localization performance, and creates favorable conditions for source localization. Locating the ICs also makes the multiple-dipole localization simple and stable as the single dipole localization. The simulation shows that the effect of the source localization has been significantly enhanced, indicating the effectiveness of this method.

Key words: EEG; ICA; multiple-dipole localization; clustering analysis

0 引 言

EEG (Electroencephalograph) 脑源定位是根据头表测量电势反演估计脑内神经活动源(偶极子)的位置、方向和强度信息,也即 EEG 逆向问题^[1]。EEG 单偶极子定位的解决方案已相对成熟,多偶极子定位仍然受到数据有限、噪声干扰的困扰,没有统一的解决方法,且大多数方法都存在一个共同问题,那就是多偶极子定位结果不如单偶极子稳定,对误差及模型的变化比较敏感。

近年来,采用独立成分分析(Independent Compo-

nent Analysis, ICA)实现盲源分离,对感兴趣的独立成分进行源定位成为研究热点^[2-4]。Subramaniam^[5]等将 ICA 与 dSPM (dynamic Statistical Parametric Mapping)相结合对癫痫数据进行源定位。他们采用 ICA 将癫痫信号从其他伪迹信号和背景噪声中分离出来,然后对重构的 EEG 信号使用 dSPM 方法进行源定位。该方法的缺点是不能保证癫痫信号提取的准确性,整个癫痫活动的源定位过程也比较繁琐。

文中利用聚类方法,根据 ICA 的各个独立成分之间的相关特性进行有效归类,剔除伪迹干扰,为感兴趣的独立成分的选取提供可靠依据,减少人工特征识别

收稿日期:2014-07-06

修回日期:2014-10-07

网络出版时间:2015-04-22

基金项目:国家自然科学基金资助项目(61003237);江苏省高校自然科学研究项目(10KJB510018)

作者简介:文 念(1990-),女,硕士研究生,研究方向为脑电信号的分析与处理;黄丽亚,教授,硕士研究生导师,研究方向为脑电信号的分析与处理。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20150422.1008.025.html>

和挑选中产生的误差,为后续的定位提供保障。该方法不仅在抑制伪迹干扰影响的同时使整个定位过程也得到简化,而且一定程度上改善了定位性能。通过MATLAB仿真对比了聚类前后源定位的性能,结果表明了ICA和聚类算法结合的有效性。

1 研究方法

1.1 FastICA

文中采用基于负熵最大的FastICA算法,以负熵最大为搜寻方向。FastICA具有收敛速度快、并行、分布式、占用内存少等特点。因为负熵需要引入复杂的概率密度计算,则通常对负熵进行近似估计以降低计算复杂度。文中推广了高阶累积量近似方法,引入非线性函数来合理地近似估计负熵^[6]:

$$J_g(Y) = \{E[g(Y)] - E[G(Y_{\text{Gauss}})]\} \quad (1)$$

其中, $E[\cdot]$ 为均值运算; $g(\cdot)$ 为非线性函数。FastICA寻找这样一个方向,使得 $\mathbf{W}^T \mathbf{X} (\mathbf{Y} = \mathbf{W}^T \mathbf{X})$ 具有最大非高斯性。基于负熵的FastICA算法的基本步骤如下:

- (1) 初始观测信号数据 x 去均值,并球化得到 z 。
 - (2) 选择一个初始随机权矢量 \mathbf{W} 。
 - (3) 权值更新为
- $$\mathbf{W} = E\{\mathbf{Z}g(\mathbf{W}^T \mathbf{Z})\} - E\{g'(\mathbf{W}^T \mathbf{Z})\} \mathbf{W}$$
- 这里选 $g = u^3$ 。
- (4) 标准化 \mathbf{W} , 令 $\mathbf{W} \leftarrow \mathbf{W} / \|\mathbf{W}\|$ 。
 - (5) 若未收敛,则返回第3步。

1.2 聚类分析

聚类分析可以帮助找寻各偶极子的信息,将特征相近的偶极子进行归类。文中采用的是两种典型基于距离的Linkage(分层聚类)和K-Means聚类算法。

Linkage通过创建用于聚类的树或是树图,在一定尺度形成聚类的方法,其算法步骤如下:

- (1) 确定样本间的相似性。可采用欧氏距离法定义各个样本之间的相似性。
- (2) 连接各个相似样本,形成二进制分层聚类树。若是已经成树的聚类样本,则将聚类再组合形成新的聚类,直到形成完整的等级树。
- (3) 评价聚类信息。检测实际的二叉聚类树中各样本间距离和计算得到的距离之间的相关性。
- (4) 创建聚类,决定在哪个层级划分等级树形成聚类。剪除等级树的低层枝叶,一个聚类是由一枝下的样本形成,从而进一步创建数据样本的分类。

与Linkage聚类不同,K-Means聚类在实际观察数据上操作,更适合大数据聚类,其具体的算法流程是:

- (1) 从 N 个数据样本中任意选择 k 个作为初始聚

类中心。

(2) 根据每个聚类样本的均值(中心样本),计算每个样本与这些中心样本的距离;并根据最小距离重新对相应样本进行划分。

(3) 重新计算每个(有变化的)聚类的均值(中心样本)。

(4) 循环步骤(2)到(3)直到每个聚类不再发生变化为止。

对于 k 值的选择会影响到聚类的结果,通过不断变换值来确定最佳聚类。由准则函数难以直接找到最佳聚类数 k_{opt} ,多数学者认为 $k_{\text{opt}} \leq \sqrt{n}$,其中 n 为需要聚类的成分个数^[7]。文中的独立成分个数为30,则最多可以有5个聚类。一个好的聚类应该使不同聚类中心的间距尽可能的大,而样本与其所属类中心尽可能的小。构造一个描写聚类质量函数:

$$S(k) = \frac{\sum_{i=1}^k \sum_{x \in C} |x - \bar{x}_i|^2}{\min_{i,j < k \text{ 且 } i \neq j} |x - \bar{x}_i|^2} = \frac{A(k)}{B(k)} \quad (2)$$

其中, \bar{x}_i 和 \bar{x}_j 是第 i 和第 j 类; $S(k)$ 的分子 $A(k)$ 为 $\sum_{i=1}^k \sum_{x \in C} |x - \bar{x}_i|^2$,表示所有类中样本与其中心间距的平方和;分母 $B(k)$ 为 $\min_{i,j < k \text{ 且 } i \neq j} |x - \bar{x}_i|^2$,表示两个不同类中心间距平方的最小值。所以 S 越小,选择对应 k 值表示将样本划分为 k 类时聚类效果越好^[8]。

为了评价Linkage和K-Means这两种聚类算法的性能,通常采用聚类有效性指标来选择较优的聚类方法。在目前提出的一些检验聚类有效性指标中,silhouette指标^[9]以其良好的检验性能而得到广泛应用。其工作原理是计算每个样本与同一聚类中样本的相似度以及与其他聚类中样本的不相似度,进而检验聚类的有效性。通常采用样本之间的平均欧氏距离来度量相似度,则第 i 个样本的silhouette指标值计算公式如下:

$$\text{Sil}(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

silhouette取值范围为 $[-1, 1]$ 。其中, $b(i)$ 表示样本 i 到其他每个类中样本的平均欧氏距离的最小值; $a(i)$ 表示样本 i 与其所在类的其他样本的平均欧氏距离。一般用整个数据集的所有样本的平均silhouette指标值来评价聚类结果的有效性,值越大表示聚类效果越好。

1.3 源定位

1.3.1 正向问题

脑电正向问题研究是逆向问题研究的基础^[10]。正向问题是通过电流源的位置、极矩和导电率,计算头

表面电势分布,其中最主要的两个问题是构造源模型和头模型^[11]。

对于源模型,文中采用 Brazie 提出的电流偶极子模型来描述脑内激励源^[12],则脑电正问题也就转化为求解电流偶极子在头表产生电势的问题。应用麦克斯韦方程,头表电势在数学上可用泊松方程描述:

$$\nabla \cdot (\sigma \nabla \phi) = -\nabla \cdot J^p \tag{4}$$

头表的纽曼边界条件可表示为:

$$\sum \sigma \frac{\partial \phi}{\partial n} = 0 \tag{5}$$

其中, $\sigma(x,y,z)$ 为传导率分布; J^p 为当前电流源; ϕ 表示偶极子源在头表产生的电势; n 为外法线向量。

对于头模型,现有多种数值计算的方法用于求解泊松方程来构造真实头模型。文中采用边界元方法 (Boundary Elements Method, BEM), 也叫 BEM 头模型^[13]。BEM 头模型通常由一组各项同性的区域构成,每个区域具有相同的导电性,各个区域的边界问题通过 BEM 数值计算解决。应用格林公式可得

$$\phi_i(r') = \frac{1}{\sigma_i} \phi_i^\infty(r') + \frac{1}{4\pi\sigma_i} \sum_k \Delta\sigma_k \int_{\Gamma} \phi_k(r_0) d\Omega_k \tag{6}$$

式中, $d\Omega_k$ 为立体角微元,且 $d\Omega_k = \frac{(r_0' - r')}{\|r_0' - r'\|^3} dn_k$; t 表示各个区域; Γ 表示各区域边界; $\phi_i^\infty(r')$ 表示电流源在边界内产生的电势;当 J^p 采用电流偶极子模型时

$$\phi_i^\infty(r') = \frac{J^p \cdot (r' - r_0)}{4\pi \|r' - r_0\|^3} \tag{7}$$

其中, r_0 表示偶极子的位置; r_0' 表示边界上的等效偶极子源位置; r' 表示边界上待计算位置; $\bar{\sigma}_i$ 表示内部和外部区域的平均电导率; $\Delta\sigma_i$ 表示内外电导率之差。通常边界被划分为多个三角形网格,则公式(6)可转化为对这些三角形的积分。通过对三角形顶点值进行插值得到三角形上的电势值,如采用平均插值法,则公式(6)转化为线性方程组形式:

$$\bar{\sigma} \Phi = \Phi^\infty + A \Phi \tag{8}$$

其中, A 的值可由各三角形对场点的立体角计算得到。

1.3.2 逆向问题

文中采用 DIPFIT 方法实现源定位,其主要思想是调整偶极子的各个参数使正向模型中所计算得到的电势分布与所测脑电数据之间拟合一个非线性代价函数,并采用非线性优化方法求解。假设 $\bar{B} = (\bar{b}_1, \bar{b}_2, \dots, \bar{b}_N)$ 是大脑头皮上所记录到的电势分布, $B = (b_1, b_2,$

$\dots, b_N)$ 为正向模型得到的电势分布,则要求改变各偶极子的参数,包括位置、强度、方向,使代价函数达到极小,所对应的各参数就是最小二乘意义下的最优解。其中代价函数可设为:

$$E = \|B - \bar{B}\|_{2F} \tag{9}$$

其中, E 则为与 \bar{B} 之差的 Frobenius 范数。DIPFIT 方法采用相对误差 (Relative Variance, RV) 值来衡量所测得头皮电势分布与偶极子源在头皮上产生电位分布差别的程度。其中

$$RV = \frac{\sum_{i=1}^M [\bar{b}_i - b_i]^2}{\sum_{i=1}^M \bar{b}_i} \times 100\%$$

RV 越小,表示头皮所测电势与偶极子源在该处产生电势之间的差别越小,拟合程度越好。

由于是针对每个独立成分进行定位操作,整个定位过程像单偶极子定位一样稳定简单^[14]。

2 仿真实验

2.1 数据描述

仿真实验基于 Matlab 环境下的 EEGLAB^[15] 工具包。EEG 数据来自加州大学圣迭戈分校 (UCSD) 的 SCCN 研究中心。数据类型是稳态视觉诱发电位 (Steady State Visual Evoked Potential, SSVEP), 被试者注视着前方液晶屏幕中央的一个 3×3 cm 大小、以固定频率进行黑白闪烁的方块 (闪烁频率在 8 ~ 13 Hz 之间随机选择), 整个过程持续 30 s, 数据采集设备为 Bio-semi。电极数量为 264 个, 采样频率为 2 048, 电极系统为 10 ~ 20 国际标准导联系统^[16]。

2.2 数据降维和 FastICA

数据首先经过 [1 80] Hz 的带通滤波器,以滤除低频和低频噪声。考虑数据的电极有 264 个,如果直接进行 ICA,程序运行需要较长的时间,且得到的 264 个独立成分中起决定作用的集中在前 30 个,因此文中在 ICA 分析之前,先通过主成分分析 (Principal Component Analysis, PCA) 将原始的 264 个电极的数据投影到低维子空间,形成 30 个主成分,在最小二乘意义上保留最大信息,在一定程度上也达到降噪的性能。对 PCA 的 30 个主成分运行 FastICA 算法后得到的 ICA 独立成分通过分析初步判断 IC1、IC2 为心电伪迹, IC17 为眼动伪迹。图 1 显示了这三个伪迹信号在 5 ~ 10 s 期间的波形图和头图,可以看到 IC1、IC2 波动频率与心率一致,头图中信号强度集中在后颈处,靠近颈部动脉,因此可以判断为心电干扰。IC17 的波形在 9 s 附近出现明显振荡,且与未处理前信号出现振荡时间一致,头图的信号强度集中在前额眼睛附近,可判断为

眼动伪迹。

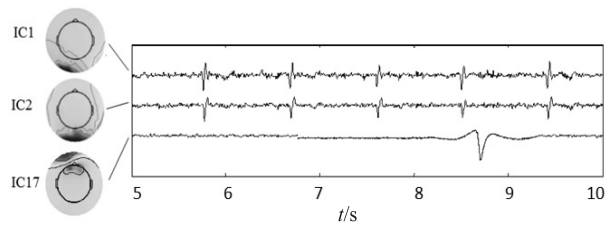


图1 伪迹成分波形图及其头图

2.3 Cluster 聚类

对ICA独立成分通过经验性判断伪迹后,分别采用K-Means和Linkage两种聚类方法进行研究,并对两者的性能进行比较。

(1) K-Means 聚类分析。
根据前文中所提到的K-Means算法,必须先确定最优k值。表1为不同k值下的A(k)、B(k)和S(k)。

表1 K-Means的K值选择表

K 值	各类距离 总和 A(k)	聚类中心间 最小距离 B(k)	S(k)
2	5.020 9	0.421 1	11.923
3	4.645 7	0.071 5	64.975
4	4.298 4	0.065 2	65.926
5	3.983 6	0.061 7	64.564

可以看到,当k=2时对应聚类质量函数S(k)最小,但分类数目过少,因此选择了k=5时的聚类结果,其S(k)仅次k=2时的值。在确定k值之后,采用silhouette值来衡量聚类结果,其结果如图2所示。

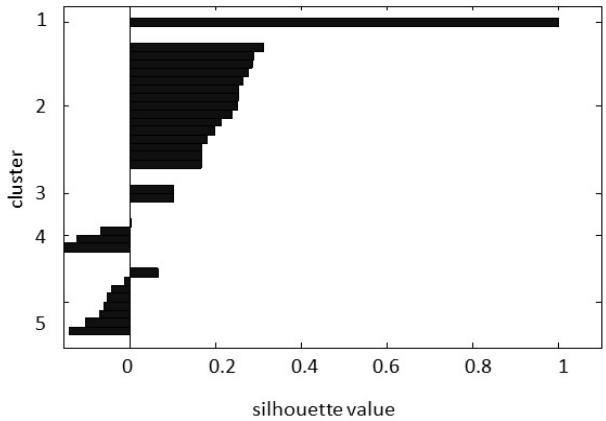


图2 K-Means Cluster 相关系数

图2中显示出K-Means聚类的silhouette值分布,30个独立分量中,一部分silhouette值小于0,其平均silhouette值也较小,表示该方法的聚类效果差,不能对FastICA的独立分量有效归类。表2给出了FastICA的K-Means聚类后所包含的独立成分编号,由于知道IC1和IC2是属于心电伪迹一类,而聚类分析结果将其分到不同的类,因此在该实验中,K-Means算法的聚类效果不太理想,不能实现有效的聚类。

表2 FastICA的K-Means和Linkage聚类分析结果

Cluster	ICs (K-Means)	ICs (Linkage)
1	7	1,2,12,14,19
2	2,3,5,6,8,10,12,13,14,15,16,18,19,21	3,4,5,6,8,10,11,13,15,16,20,21,22,23,24,25,26,28,29,30
3	5,26,29,30	9,18
4	17,27	7
5	9,18,22,28,4,11,20,23,24,26	17,27

(2) Linkage 聚类分析。

采用Linkage方法对FastICA的30个独立分量进行Linkage聚类分析,结果如图3所示。

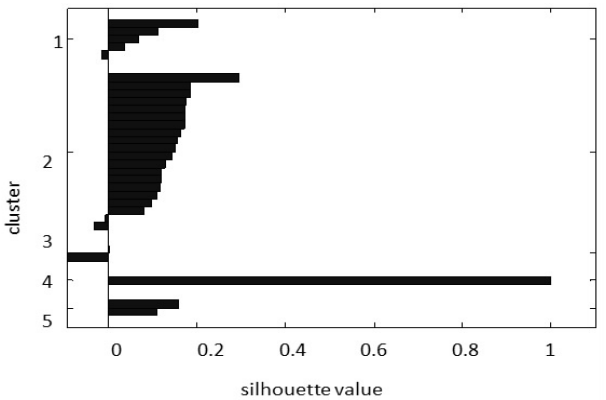


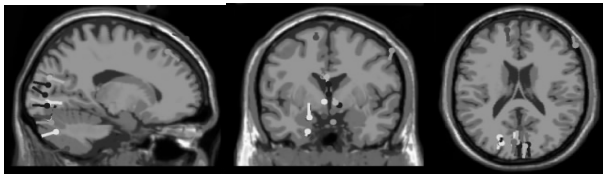
图3 Linkage Cluster 相关系数

图中为30个样本的silhouette值分布,可以看到silhouette值基本集中大于0的范围内,其平均silhouette值也相对较大,说明Linkage对于独立分量归类效果良好。表2中也显示了聚类数目为5时Linkage聚类的分类结果,IC12、IC14、IC19与上文中所判断出的心电伪迹成分IC1、IC2归为一类,IC27与眨眼伪迹IC17归为一类,说明这些独立成分与明显的心电和眨眼伪迹的特性相似,从而也将会是影响源定位的成分,故要将此排除。

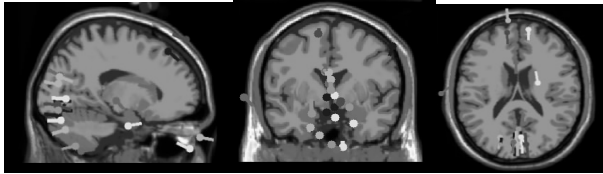
通过以上对FastICA独立成分的K-Means和Linkage聚类分析,结果表明两者都能将IC27和IC17,IC9和IC18,IC7各自分为一类,而总的来说Linkage聚类的效果更好。

2.4 源定位

源定位中的源模型采用电流偶极子模型,头模型采用BEM头模型,将FastICA和聚类效果较好的Linkage Cluster相结合,去除了表2中Cluster1、3、4、5这4类中的10个独立成分,只保留了含大部分感兴趣独立成分的Cluster2进行源定位。整个定位过程在EEGLAB中采用DIPFIT方法进行,结果如图4(a)所示。为了说明Linkage聚类对源定位的影响,对聚类前FastICA(仅去除IC1、IC2、IC17)的独立成分进行源定位,结果如图4(b)所示。



(a) FastICA 与聚类结合后各独立成分源定位结果



(b) 聚类前 FastICA 各独立成分源定位结果

图4 聚类前后独立成分源定位结果比较

图4中分别从三个视角展现了源定位结果。由于实验数据是稳态视觉诱发信号,视觉中枢产生刺激源是在后枕区,在图4(a)、(b)中,都可以明显看到独立成分源定位后的偶极子大多是集中在后脑,同时偶极子位置坐标参数表明约71%的偶极子是处于BEM头模型中后枕区视觉中枢的范围。而较图4(a),图4(b)中仅约53%的偶极子是处于视觉中枢范围,而更多的偶极子是分散在后枕区以外的其他区域,这是因为其中除感兴趣成分以外的其他伪迹成分也被定位;图4(a)中正是根据聚类结果而成功去掉了大部分伪迹,将感兴趣的成分定位出来。两者相比,FastICA与聚类结合对源定位更有利,一定程度上改善了定位效果。

3 结束语

文中通过将聚类 Cluster 方法与 ICA 源定位相结合,将 ICA 分解的独立成分进一步通过聚类进行筛选,在剔除伪迹干扰的同时,也为感兴趣的独立成分的稳健选取提供可靠依据,一定程度上也降低了定位复杂度。仿真对比结果显示该方法对源定位效果有一定的改善。

参考文献:

- [1] Grech R, Cassar T, Muscat J, et al. Review on solving the inverse problem in EEG source analysis[J]. Journal of Neuro-Engineering and Rehabilitation, 2008, 5(1): 1-33.
- [2] Zhukov L, Weinstein D, Johnson C. Independent component

analysis for EEG source localization[J]. Engineering in Medicine and Biology Magazine, 2000, 19(3): 87-96.

- [3] Kobayashi K, James C J, Nakahori T, et al. Isolation of epileptiform discharges from unaveraged EEG by independent component analysis [J]. Clinical Neurophysiology, 1999, 110(10): 1755-1763.
- [4] Hild K E, Nagarajan S S. Source localization of EEG/MEG data by correlating columns of ICA and lead field matrices[J]. IEEE Transactions on Biomedical Engineering, 2009, 56(11): 2619-2626.
- [5] Subramaniam N P, Peltola J, Tanskanen J M A, et al. Source localization based on ictal electroencephalographic recordings [C]//Proc of international symposium on biomedical engineering and medical physics. Berlin: Springer, 2013: 26-30.
- [6] Comon P. Independent component analysis, a new concept? [J]. Signal Processing, 1994, 36(3): 287-314.
- [7] 杨善林, 李永森, 胡笑旋, 等. K-means 算法中的 k 值优化问题研究[J]. 系统工程理论与实践, 2006, 26(2): 97-101.
- [8] 王 芳. 传统聚类方法的分析及改进[D]. 长沙: 中南大学, 2007.
- [9] Dudoit S, Fridlyand J. A prediction-based resampling method for estimating the number of clusters in a dataset[J]. Genome Biology, 2002, 3(7): 1-21.
- [10] Moshier J C, Leahy R M, Lewis P S. EEG and MEG: forward solutions for inverse methods[J]. IEEE Transactions on Biomedical Engineering, 1999, 46(3): 245-259.
- [11] Acar Z A, Makeig S. Effects of forward model errors on EEG source localization[J]. Brain Topography, 2013, 26(3): 378-396.
- [12] Brazier M A B. A study of the electrical fields at the surface of the head[J]. Electroenceph Clin Neurophysiol, 1949, 2(S): 38-52.
- [13] Hämäläinen M S, Sarvas J. Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data[J]. IEEE Trans on Biomedical Engineering, 1989, 36(2): 165-171.
- [14] Delorme A, Palmer J, Onton J, et al. Independent EEG sources are dipolar[J]. Plos One, 2012, 7(2): e30135.
- [15] EEGLAB Tutorial [EB/OL]. [2013-12-13]. http://scn.ucsd.edu/wiki/EEGLAB#EEGLAB_Tutorial.
- [16] Wang Y, Wang Y T, Jung T P. Visual stimulus design for high-rate SSVEP BCI [J]. Electronics Letters, 2010, 46(15): 1057-1058.

基于ICA和聚类的EEG脑源定位研究

作者:

[文念](#), [黄丽亚](#), [于涵](#), [杨晨](#), [杨俊宇](#), [WEN Nian](#), [HUANG Li-ya](#), [YU Han](#), [YANG Chen](#), [YANG Jun-yu](#)

作者单位:

[南京邮电大学 电子科学与工程学院, 江苏 南京, 210003](#)

刊名:

[计算机技术与发展](#) 

英文刊名:

[Computer Technology and Development](#)

年, 卷(期):

[2015\(5\)](#)

引用本文格式: [文念](#). [黄丽亚](#). [于涵](#). [杨晨](#). [杨俊宇](#). [WEN Nian](#). [HUANG Li-ya](#). [YU Han](#). [YANG Chen](#). [YANG Jun-yu](#) 基于ICA和聚类的EEG脑源定位研究[期刊论文]-[计算机技术与发展](#) 2015(5)