

基于互信息统计特性的桥接电极辨出方法^①

王冬翠¹, 陈真诚², 邓阳光³, 巫放明⁴, 张阳德⁴

¹(中南大学 湘雅医院肝胆肠外科研究中心, 长沙 410008)

²(桂林电子科技大学 生命与环境科学学院, 桂林 541004)

³(桂林电子科技大学 机电工程学院, 桂林 541004)

⁴(国家卫生计生委肝胆肠外科研究中心, 长沙 410008)

摘要: 电极桥接是一项重要但又极易被忽视的脑电噪声来源. 基于互信息的统计特性提出了一种筛查桥接电极的方法, 并将该方法应用到了4个被试不同任务下的数据中. 4个被试的脑电数据中分别检出3、4、4和0对桥接电极; 对采集条件或不同预处理步骤进行的单因素鲁棒性分析未从中发现任何影响因素; 进一步的仿真对比实验表明, 所提方法较电气距离法更为准确. 因此, 互信息的统计特性可有效用于检测脑电数据中的电极桥接, 进而对及时提醒修复脑电数据或正确解释所分析出的结果具有重要意义.

关键词: 脑电图; 桥接电极; 互信息

Method of Identifying Bridged Electrodes Based on Mutual Information Statistics

WANG Dong-Cui¹, CHEN Zhen-Cheng², DENG Yang-Guang³, MO Fong-Ming⁴, ZHANG Yang-De⁴

¹(Hepatobiliary and Enteric Surgery Research Center, Xiangya Hospital, Central South University, Changsha 410008, China)

²(School of Life and Environmental Science, Guilin University of Electronic Technology, Guilin 541004, China)

³(Electromechanical Engineering College, Guilin University of Electronic Technology, Guilin 541004, China)

⁴(Hepatobiliary and Enteric Surgery Research Center, National Health and Family Planning Commission, Changsha 410008, China)

Abstract: Electrode bridging is a common but easily ignored EEG artifact source. Based on the distinctive statistical characteristics of mutual information, a novel algorithm to automatically detect these bridges was developed and further applied to four EEG data sets acquired from different subjects. The applications identified four, four, three and zero pairs of bridged electrodes in these four data sets, respectively. No influencing factors were returned by One-way robustness analyses across different recording tasks and/or pre-processing procedures. And further comparison experiments performed on simulated data indicated that it outperformed the electrical distance method. All these findings suggest that the novel method is able to screen electrode bridges in a satisfying manner, making it of great significance in providing an indication to timely remedy the contaminated EEG data so as to avoid distortions to the resultant EEG topographies.

Key words: electroencephalogram; bridged electrodes; mutual information

脑电图是一种在颅内或头皮上将微弱电位变化放大并记录下来的检查手段, 因其优良的时间分辨率特性而成为认知研究、脑机接口、临床医学等诸多领域的重要研究工具. 而脑电信号的分析严重依赖于采集过程的准确性, 如对各种伪迹源的严格控制等. 常见的伪迹^[1]有来自仪器设备和环境的非生理性因素如放大器产生的固有高频噪声、电极与设备间的不安全连

接等, 以及来自被试生理性因素如眼电、肌电、心电、颅骨缺陷等的干扰. 生理性的伪迹通常可借助一些附加实验控制条件如要求被试减少眨眼、吞咽动作等或特殊算法如回归法、伪迹减法、主成分分析、独立成分分析和小波变换等来去除; 而非生理性的伪迹则被认为可通过精心设计记录系统、改造环境以及规范采集流程等措施来尽可能避免. 但如此多的伪迹来源在

① 基金项目: 广西信息科学实验中心项目(20130106); 广西研究生教育创新计划(YCSZ2012064)

收稿时间: 2014-01-22; 收到修改稿时间: 2014-03-24

实际操作过程中很难面面俱到、无一疏漏,其中源于电极与设备间不安全连接的电极桥接就是其中通常被忽视的一种。

电极桥接可由电极导联过密、注入了过多的导电膏或导电液、被试出汗或采集设备自身的物理性损害等引起^[2],以两个或多个电极的阻值基本相同且同步变化、所在通道的信号近似等同为主要表现。若未及时地发现桥接电极以采取相应补救措施将导致一些后续信号处理方法的不再适用,如独立成分分析前提假设的失效,以及研究结果的曲解等。

而对电极桥接的检测,最直观的方法是针对电极桥接的上述两个表现特征提出的,如在采集准备阶段往电极里注入导电膏或导电液时观察是否出现电极阻值联动变化情况,在数据的预处理阶段将任意两通道的信号叠加在同一幅图中进行对比等。显然,目测阻值联动变化或信号形态相似性的方法均需要实验员非常细心且比较费时。因此,发展自动化的判别方法将实验员们解放出来显得尤为必要。Junghfer 等人曾提出了一种基于阈值标准的统计方法用于控制伪迹^[3],但其标准的建立依赖于参考电极的选取,而若参考电极处的信号也是电极桥接污染的,该方法则不能够做出正确判断。此外,还有一种无需参考电极信息的自动识别方法,其以电极对之间的电气距离(Electrical Distance, ED)^[2,4]或电气距离统计特性^[5,6]为依据能够有效检出数据中存在的电极桥接噪声。而本文受后者的启发考虑从信息论的角度利用互信息(Mutual Information, MI)的统计特性实现电极桥接的自动检测,而进一步的对比实验表明该方法相较于电气距离统计特性法对电极桥接的标准要求更高、检出结果也更为准确。

1 材料和方法

1.1 互信息与桥接检测

互信息 MI 是一种基于信息熵的度量。信息熵可用于描述时间序列的随机性程度,可以量化随机变量之间概率分布的差异;而互信息 MI 则可反应两个时间序列的统计上的独立性或相关性。

两个离散随机变量 X 和 Y 的互信息,可由边缘概率密度 $p(x)$ 和 $p(y)$ 以及联合概率密度 $p(x,y)$ 得到,具体如下:

$$MI(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

其中, $p(x)$ 、 $p(y)$ 和 $p(x,y)$ 可通过直方图估计法得到,即将 X 、 Y 在一定取值范围内分为 N_x 、 N_y 个区域,然后分别考察离散序列的值落在每个区间的个数,进而获得 $p(x)$ 和 $p(y)$ 。类似地, $p(x,y)$ 可通过计算离散随机变量位于 $N_x \times N_y$ 个区域中的概率获得。

显然,MI 可以作为衡量一个变量依赖另一个变量的测度,其值越大,表明随机变量综合相关程度或相似程度越强。

1.2 脑电数据

本文共选用了四个被试的脑电数据:第一个被试的具体个人信息未知,其脑电数据来源于网络(<http://psychophysiology.cpmc.columbia.edu/eBridge>),且采样率为 128Hz、采样时间为 99s、通道数为 32。第二至四个被试均为在校研究生,每位被试的两组数据分别采自静息(睁-闭-睁-闭,共四个试次)和新异声音刺激任务下。在采集准备阶段通过观察电极阻值的变化判断嫌疑桥接,因采集所采用的是 Neuroscan 配套的尼龙电极帽,取下擦除被试头部的多余导电膏后再重新带上这一补救措施并不方便采用,故只记录下桥接嫌疑电极而继续对以上三种状态下的脑电信号进行了采集。采集的具体参数为:采样率 1000Hz、通道数 32、以左右乳突均值为参考,其中睁、闭眼静息态采样时间均为 50s;而新异声音刺激任务中新异刺激(即目标刺激)与标准刺激分别为 2000Hz 和 1000Hz 的 75dB 单纯音,30 次目标刺激与 160 次标准刺激以伪随机顺序(保证无连续出现的目标刺激)依次呈现,且刺激间隔为 950ms-1050ms 间的随机值。

1.3 预处理

主要使用 EEGLAB v12.0.2.0b 工具箱^[7]在 MATLAB R2010b^[8]环境下实现。对于被试一的脑电数据滤波仅做了分段处理,即利用自写的 MATLAB 代码将其分成长 0.5s(64 个数据点)的连续片段(共 198 段)。对被试二至四的静息态数据,先利用 EEGLAB 函数取睁、闭眼提示事件后 5-55s 区间数据得到四个试次的连续数据 O1、C1、O2 和 C2,然后再分别利用自写代码将其分为长 0.5s(500 个数据点)的连续片段(共 100 段)。而对于被试二至四的新异声音刺激任务下的脑电数据,则分别经过了以下四种预处理方式:1)直接经自写代码分段,2)利用 EEGLAB 分段函数取目标刺激前 0.1s 至目标刺激后 0.9s 得到 30 段,3)利用 EEGLAB 分段函数取标准刺激前 0.1s 至标准刺激后 0.9s 得到 160 段,4)利用 EEGLAB 分段函数取刺激前 0.1s 至刺激后 0.9s 得到 190 段。

1.4 算法流程

预处理后的脑电数据在 MATLAB 中以三维矩阵(通道数×数据点数×段数)的形式存储,进一步计算互信息存入三维矩阵(通道数×通道数×段数).然后对全体互信息数据进行分组频数分析,在 0-2 间共分 40 组,每组宽为 0.05.发现,很多频数分布图在 $MI \in (1, 2)$ 区间呈现有一个局部峰值,并且在沿 MI 值递减方向后继的局部最小值也均出现在 $MI \geq 0.9$ 区间.通过与电气距离法联合印证,发现该局部峰值基本上由被桥接电极污染的数据引起.该结论也就是判断依据,即在 $MI \in (1, 2)$ 区间局部峰值的存在与否是桥接电极有无的标志.若峰值存在,为得到桥接电极的详细信息,则自动提取出对应于局部最小值的 MI 值作为门限值,如果任意一对通道在 50%及以上的段上的 MI 值均大于或等于这个门限值,则该通道对被认为是桥接的.具体的算法流程如图 1 所示.

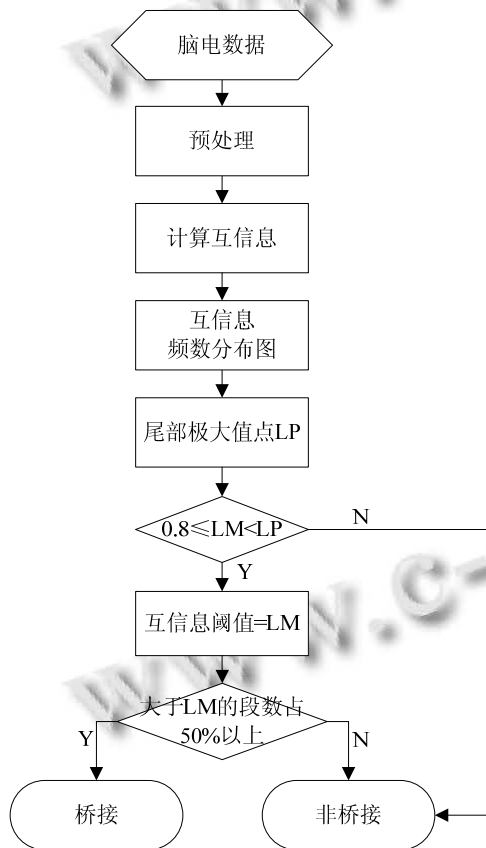


图 1 算法流程图

1.5 影响因素分析

为检验本文算法的鲁棒性,我们对任务状态、采样率、滤波、段数、段长度等各种可能的影响因素进行了对比分析,并采用各种不同因素下所辨出的桥接电极的 Sørensen 相似性指数^[9]为衡量指标,该指标在 0

至 1 范围内取值,若值为 0 则无相似性,而值为 1 则具有完全相似性.

数据任务状态的不同体现在睁眼静息 O、闭眼静息 C、新异声音刺激任务 T 等的区分上;通过重采样可实现采样率的变化,如由 1000Hz 降采样到 500Hz;滤波可能带来的影响则通过比较在预处理阶段进行带通(Band-Pass, BP)滤波与无任何滤波处理(Full-Pass, FP)的情况进行比较;段数的区分则是在原来段数的基础上随机抽取部分段数,如对静息态脑电数据分别选择 30、60、90 段,对新异声音刺激任务分别选择 30、90、180 段等;对段长度的设置则为 0.5、1s 两种情况,同时控制段数固定在 50 段.

1.6 桥接电极仿真

目测法、电气距离法以及本文互信息法三方共同证实,被试四的数据中无桥接电极噪声.遂在本数据的基础上展开了仿真实验,具体设计为:在被试四第一次睁眼静息 50s 数据中,任意取一对邻近电极 Chan1 和 Chan2 处脑电数据,如 T7 和 C3,然后在原来数据的基础上按如下公式重构数据:

$$NewChan1 = a * Chan1 + (1-a) * Chan2$$

$$NewChan2 = (1-a) * Chan1 + a * Chan2$$

其中,混合因子 a 在 [0, 1] 之间取值并影响重构后数据间的相似性;当 a=0.5 时重构后的两数据完全等同,随着它的递减或递增,相似性均逐渐递减.因此,本仿真实验中只考虑了 0.5 后的递增情况,并采用了 0.1 和 0.01 两种不同尺度,即综合了 a=0.5、a=0.6:0.1:1 和 a=0.51:0.01:0.59 三种情况.

2 结果

2.1 桥接电极检出情况

对所涉及数据进行处理的过程中,本文中算法自动确定的 MI 门限值均经过主观观测验证,验证结果表明所有门限值均位于相应的局部最小值附近;而且对比发现本文所提出的方法与电气距离法取得的结果具有良好的一致性.而且互信息法的原理有效性可通过图 2 直观看出,其中,所有电极对所有数据段 MI 的频数分布图近 2 端的峰值主要由桥接电极(如 T7 和 C3)贡献,而非桥接电极(如 T7 和 Cz)则共同决定了总频数分布图近 0 端的分布.

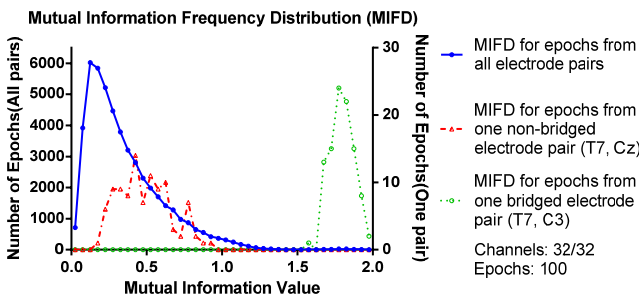


图 2 互信息频数分布图.

本图是基于被试四 O1 数据而得(a=0.5), 其中红(右 Y 坐标)、绿(右 Y 坐标)、蓝(左 Y 坐标)线条分别为非桥接电极对(图中取 T7 和 Cz)、仿真桥接电极对(T7 和 C3)、所有电极对(C₃₂²=496 对)间的互信息分布情况.

而对来源于四个被试的脑电数据中的桥接电极的实际检出情况为: 因被试一数据来源于网络, 对采集过程细节未知, 因此无先验知识, 但 MI 与 ED 方法均检出 PO7 和 PO3、Iz 和 I2 以及 I1 和 Po9 三对桥接电极; 在采集被试二数据之前的准备过程中怀疑 Ft8 和 T4、Cp4 和 P4、Tp8 和 T6 以及 Tp7 和 T5 四对连通电极, 对采集到的数据分段分组后, MIS 与 EDS 法对 O1、C1、O2、C2、ERP 目标刺激、ERP 标准刺激以及 ERP 全部刺激共七个片段数据的分析结果均与怀疑吻合; 而对被试三数据则初始怀疑 F8 和 Ft8、O1 和 O2 以及 F7 和 Ft4 三对桥接电极, 而 MI 与 ED 法对七个子分组数据的辨出结果也均证实了怀疑的正确性, 此外也共同认为 F4 和 Fc4 是桥接在一起的; 至于被试四的数据, 采集准备阶段的阻值目测、MI 法以及 ED 法均未发现桥接电极嫌疑. 四个数据集中桥接电极的大致辨出情况参见表 1.

2.2 影响因素分析结果

对各因素影响情况的探讨以被试三的脑电数据为研究对象, 结果(见表 2)表明: 变换不同采集条件或处理步骤中的细节, 如被试任务状态、采样率、滤波、段长度、段数目等, 最终检测到的桥接电极的情况, 包

括桥接电极对数目以及所含电极, 均完全一致.

表 1 各数据集中的电极桥接情况

	桥接电极数目		桥接电极对数目	
	MI 法	ED 法	MI 法	ED 法
Sub1	6	6	3	3
Sub2	8	8	4	4
Sub3	8	8	4	4
Sub4	0	0	0	0

表 2 不同变换采集或预处理条件下所测出桥接电极的 Sørensen 相似性指数对比

Sørensen 相似性指数	
被试任务状态	
睁眼静息 O	
闭眼静息 C	1
单纯音刺激 T	1
采样率	
1000(默认)	
500	1
滤波	
未滤波(默认)	
带通滤波	1
段长度	
500(默认)	
1000	1
段数目	
睁眼静息	
100(默认)	
30	1
60	1
90	1
单纯音刺激	
190(默认)	
30	1
90	1
180	1

2.3 电极桥接的仿真结果

在被试四 O1 数据 T7 与 C3 通道间进行的电极桥接仿真(即 a=0.5 的情况), 本文方法能正确检出. 而通过观察在 a 的另外两种取值方式下的检出情况则可对本方法的实际检测力进行判断, 结果表明只有在两个电极处的数据差别极小时才会将其判断为桥接, 且较 ED 法对检出条件的要求更为严苛、检出率也更低, 具体体现在本文方法仅在 0.50 ≤ a ≤ 0.53 时认为两电极是桥接的, 而 ED 法将该范围扩展到 0.50 ≤ a ≤ 0.56(见表 3).

表 3 对于仿真桥接电极的检出情况

	0.5	0.51	0.52	0.53	0.54	0.55	0.56	0.57	0.58	0.59	0.6	0.7	0.8	0.9	1.0
MI 法	√	√	√	√	×	×	×	×	×	×	×	×	×	×	×
ED 法	√	√	√	√	√	√	√	×	×	×	×	×	×	×	×

3 结语

本文所介绍的桥接电极检测方法主要应用了桥接

电极对之间所具备的互信息统计特性. 该方法不但操作简易, 检测结果亦准确有效. 首先, 它可以辨出在

采集准备阶段根据观测阻值同步变化情况所发现的桥接电极,却较目测法更为严谨,因而也使得所检出结果更为可信;其次,在实测数据中,它与基于电气距离统计特性的方法取得的结果具有良好的一致性,似乎两者方法相当,但仿真实验的结果却说明本文所提出的方法的准确性要优于电气距离法。此外,通过改变各种采集状态条件或数据处理步骤实现的影响因子筛查实验更是表明该方法具有良好的稳健性。

在实际应用本文方法的过程中,分别选用了来自于网络的公开数据、实测数据以及基于实测数据的仿真数据,使得本文方法的有效性以及广泛适用性进一步得到了体现。其实,造成电极桥接的因素主要有电极的高排布密度、被试的头发条件和操作者不娴熟的技巧(如注入太多导电膏)等。但本文中所分析的数据均采自于32通道,相比于64、128通道,电极的排布密度处于较低水平,且采集时室温恒定在不易出汗的22℃。但处理结果却表明,尽管是较低排布密度,但依然极易被桥接电极污染。因而,桥接电极在实际的脑电数据采集过程中是极易发生的。不过,所幸已发展了一些相应的检测分辨方法,而本文方法也在一定程度上为此做出了些许贡献。

总之,MI法可应用于数据的预处理阶段以分辨极易存在但同时也极易被忽视的桥接电极,及时提请采取一定措施规避该物理噪声对最终处理结果的扭曲以提高相应结论的可信度。

参考文献

- 1 Tatum WO, Dworetzky Ba, Schomer DL. Artifact and recording concepts in EEG. *J Clin Neurophysiol*, 2011, 28(3): 252-63.
- 2 Greischar LL, Burghy CA, van Reekum CM, Jackson DC, Pizzagalli DA, Mueller C, Davidson RJ. Effects of electrode density and electrolyte spreading in dense array electroencephalographic recording. *Clin Neurophysiol*, 2004, 115(3): 710-720.
- 3 Junghfer M, Elbert T, Tucker DM, Rockstroh B. Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, 2000, 37(4): 523-532.
- 4 Tenke CE, Kayser J. A convenient method for detecting electrolyte bridges in multichannel electroencephalogram and event-related potential recordings. *Clin Neurophysiol*, 2001, 112(3): 545-50.
- 5 Kayser J, Tenke C. Electrical distance as a reference-free measure for identifying artifacts in multichannel electroencephalogram (EEG) recordings. *Psychophysiology*, 2006, 43, S51.
- 6 Alschuler DM, Tenke CE, Bruder GE, Kayser J. Identifying electrode bridging from electrical distance distributions: A survey of publicly-available eeg data using a new method. *Clin Neurophysiol*, 2013.
- 7 Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *J Neurosci Methods*, 2004, 134(1): 9-21.
- 8 The MathWorks Inc. MATLAB and Statistics Toolbox Release 2010b. Natick, MA. The MathWorks, Inc. 2010.
- 9 Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biologiske Skrifter*, 1948, 5: 1-34.