

多参考信号的 ICA 方法及其在去除脑电伪差中应用

邱天爽*, 武建

(大连理工大学 电子与信息工程学院, 辽宁 大连 116024)

摘要: 为消除脑电信号中的心电、眼电等伪差,在已有的不动点算法和带参考信号的独立分量分析算法的基础上,提出了一种多参考信号的独立分量分析方法.该方法通过计算各伪差在各路观测信号中的比重,去除伪差对观测信号的影响,从而得到较为纯净的脑电信号.所提方法相对于传统的 fast ICA 算法具有更小的计算量,并且不需要对分离的独立源进行人工干预,同时也是对 ICA-R 算法的一种扩展,解决了其只能提取单路源信号的缺点.仿真实验证明该方法更切合实际情况,而且能够更加有效地去除脑电信号中的多个伪差.

关键词: 独立分量分析; 参考信号; 伪差去除; 脑电信号

中图分类号: R318.08 **文献标志码:** A

0 引言

脑电(electroencephalograph, EEG)是大脑神经元突触后的综合,是大脑活动产生的电场经容积导体传导后在头皮上的电位分布.通过对病人不同头皮位置测得的 EEG 信号进行观察分析,可以大致判断出大脑疾病类型以及病灶部位.但是,在 EEG 的多导测量中,往往会混入脑外来源所产生的电位信号,这些信号就是伪差.常见的伪差有心电、眼电、血管波、肌电、出汗以及工频干扰等^[1].伪差有时和脑电波非常相似,常常给 EEG 信号的阅读和分析造成极大困难.因此,伪差的消除是迫切需要解决的问题.

独立分量分析(independent component analysis, ICA)方法是近年来新发展起来的一种多维信号处理方法.与传统的多维信号分析方法不同的是,经 ICA 处理得到的各个分量不仅去除了相关性,而且还是相互统计独立的.因此,ICA 技术在信号处理领域受到广泛的关注.特别在脑电处理领域,可以认为各种伪差与 EEG 信号分别由相互独立的源产生,且是瞬时线性混合的,所以将 ICA 技术应用到 EEG 信号中伪差去除成为可能^[2].

本文在已有的不动点算法以及带参考信号的 ICA 算法的基础上,提出了一种新的多参考信号的 ICA 算法,并将其应用到脑电信号中多个伪差的去除.

1 算法原理

1.1 ICA 问题描述

设 $s = (s_1 \cdots s_n)^T$ 为 n 个源信号构成的 n 维向量, $x = (x_1 \cdots x_m)^T$ 为 m 维观测数据向量,其元素是各个传感器得到的输出,观测信号可用下面的方程描述:

$$x = As \quad (1)$$

其中混合矩阵 A 为 $m \times n$ 维矩阵.式(1)的含义是 n 个源信号通过混合得到 m 维观测数据向量.

ICA 方法的目的是:在混合矩阵 A 和源信号未知的情况下,假设各个源信号 s_i 是相互统计独立的,根据观测数据向量 x 确定分离矩阵 W ,得到

$$u = Wx \quad (2)$$

其中 $u = (u_1 \cdots u_n)^T$ 是对源信号向量 s 的估计.在实际应用中常常假设 $m = n$.

1.2 多参考信号的 ICA 算法

1997 年 Hyvarinen 等^[3~5]首先提出了基于

四阶累积量的不动点算法,其后又提出了基于负熵的不动点算法.之后,他们又做了进一步的简化,简化后的算法比批处理甚至比自适应方法具有更快的收敛速度,因此又被称为“快速ICA算法”(fast ICA).这类算法基于统计学上的投影追踪法(exploratory projection pursuit, EPP),能够实现独立分量的逐次提取.2002年,Barros等^[6]在fast ICA算法的基础上引入周期性的参考信号,提出了一种带参考信号的ICA算法,记为ICA-R.该算法通过维纳滤波对权值进行初始化,以提取源信号中的周期信号,该算法也能够应用于非周期信号的提取.

实际应用中感兴趣的信号可能仅仅为某几个信号,所以可以利用先验信息,避免为分离所有独立分量而做的大量的冗余运算.本文在fast ICA算法的串行逐次提取法以及ICA-R算法的基础上,提出一种多参考信号的ICA算法.

设已知的参考信号 $\mathbf{d} = (d_1 \cdots d_l)^T$ 为 l 维向量.由式(2),可以得到输出 $\mathbf{u} = \mathbf{W}\mathbf{x}$,其中 $\mathbf{u} = (u_1 \cdots u_l)^T$ 是对应参考信号 $\mathbf{d}$ 的 l 维输出, $\mathbf{W}$ 是 $l \times m$ 维的分离矩阵.式(2)也可以写成分量的形式

$$(u_i) = \mathbf{w}_i^T \mathbf{x}; \quad i = 1, \cdots, l \tag{3}$$

其中 $\mathbf{w}_i^T = (w_{i,1} \cdots w_{i,m})$ 是对应每个输出 u_i 的分离向量,且为 $\mathbf{W}$ 的第 i 个行向量.对于每个输出 u_i ,可以得到其计算误差如式(4),原理框图如图1所示.

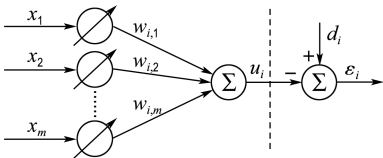


图1 维纳最优权估计原理框图

Fig. 1 The principle diagram of Wiener optimal power estimate

$$\epsilon_i = d_i - u_i; \quad i = 1, \cdots, l \tag{4}$$

由式(4)可以得到其均方误差

$$E\{\epsilon_i^2\} = E\{d_i^2\} + \mathbf{w}_i^T \mathbf{R} \mathbf{w}_i - 2 \mathbf{w}_i^T \mathbf{P}_i \tag{5}$$

其中 $\mathbf{R} = E\{\mathbf{x}\mathbf{x}^T\}$, $\mathbf{P}_i = E\{d_i \mathbf{x}\}$, 则其梯度公式

$$\partial E(\epsilon_i^2) / \partial \mathbf{w}_i = 2 \mathbf{R} \mathbf{w}_i - 2 \mathbf{P}_i \tag{6}$$

从而得到在最小均方误差(MSE)准则下的最优权 $\mathbf{w}_i^* = \mathbf{R}^{-1} \mathbf{P}_i$.在实际算法中,首先要对数据 $\mathbf{x}$ 进

行白化预处理,即 $\mathbf{v} = \mathbf{M}\mathbf{x}$,其中 $\mathbf{v}$ 满足 $E\{\mathbf{v}\mathbf{v}^T\} = \mathbf{I}$. 所以用 $\mathbf{v}$ 替换 $\mathbf{R}$ 和 $\mathbf{P}_i$ 中的 $\mathbf{x}$,可以得到维纳最优权,即

$$\mathbf{w}_i^* = \mathbf{R}^{-1} \mathbf{P}_i = E\{d_i \mathbf{v}\} \tag{7}$$

在盲分离算法中,常常采用基于峰态的代价函数

$$kurt(\mathbf{w}_i^T \mathbf{v}) = E\{(\mathbf{w}_i^T \mathbf{v})^4\} - 3[E\{(\mathbf{w}_i^T \mathbf{v})^2\}]^2 \tag{8}$$

经推导得到权向量的迭代公式^[3]

$$\mathbf{w}_i(k) = E\{\mathbf{v}(\mathbf{w}_i^T(k-1)\mathbf{v})^3\} - 3\mathbf{w}_i(k-1) \tag{9}$$

另外,假设对权向量进行初始化的维纳最优权在空间上接近所求的结果,所以在 n 维超平面上寻找结果需要满足下面的条件:

$$\|\mathbf{w}_i(k) - \mathbf{w}_i^*\| < \zeta_i \tag{10}$$

其中 ζ_i 是 $\mathbf{w}_i^*$ 所张成的超球的半径.

综合以上内容,给出多参考信号的ICA算法的基本步骤如下:

- (1) 将观测向量 $\mathbf{x}$ 去均值,然后加以白化预处理得 $\mathbf{v} = \mathbf{M}\mathbf{x}$,满足 $E\{\mathbf{v}\mathbf{v}^T\} = \mathbf{I}$, $\mathbf{M}$ 为白化矩阵.
- (2) 获取多路的参考信号 $\mathbf{d} = (d_1 \cdots d_l)^T$,令 $i = 1$.
- (3) 用维纳最优值初始化权向量 $\mathbf{w}_i(0) = E\{d_i \mathbf{v}\} / \|E\{d_i \mathbf{v}\}\|$,令 $k = 1$.
- (4) 计算权向量 $\mathbf{w}_i(k) = E\{\mathbf{v}(\mathbf{w}_i^T(k-1)\mathbf{v})^3\} - 3\mathbf{w}_i(k)$.
- (5) 对 $\mathbf{w}_i(k)$ 归一化, $\mathbf{w}_i(k) = \mathbf{w}_i(k) / \|\mathbf{w}_i(k)\|$,并令 $k = k + 1$.
- (6) 判断是否满足条件 $\|\mathbf{w}_i(k) - \mathbf{w}_i(0)\| < \zeta_i$.如果不满足,用维纳最优权加上一个小的随机误差更换现在的权值.
- (7) 重复(4)到(6)步,直至 $|\mathbf{w}_i^T(k)\mathbf{w}_i(k-1)|$ 足够接近于1.
- (8) 令 $i = i + 1$,回到第(3)步,循环直至 $i = l + 1$,即提取出所有的目标信号为止.

1.3 伪差去除

本文给出一个多参考信号的ICA算法,但最终的目的还是为了将该算法应用到脑电信号的伪差去除中.伪差去除的思想是,先计算出各个伪差在各路观测信号 $\mathbf{x}$ (去均值后)中的比重,然后将各路观测信号减去加权的各个伪差分量,得到较为纯净的脑电信号.具体的算法步骤如下:

- (1) 根据上面所求得的 W , 计算出经过 ICA 算法的伪差分量输出 $u = Wx$.
- (2) 计算出各个伪差输出分量分别在各路观测信号中所占比重的权值, $B = E\{xu^T\}$.
- (3) 最后消除伪差后的输出为 $y = x - Bu$.

2 实验结果

本文给出了一个实例, 将多参考信号的 ICA 算法应用到脑电信号的伪差去除中. 文中用到的一些数据取自 MIT-BIH 的睡眠脑电数据库^[7] (MIT-BIH Polysomnographic Database). 数据的采样率为 250 Hz, 单位为 mV. 根据采集的脑电数据的特点, 将自发脑电、心电 (electrocardiogram, ECG)、眼电 (electrooculogram, EOG) 以及一些高斯白噪声随机混合, 模拟出 6 导实际的头皮记录的 EEG 信号, 如图 2 所示. 图 3 中的 (a)、(b) 两导信号分别是同步记录的 ECG 和 EOG. 对比图 2 与图 3 中的 (a)、(b) 两导信号可以发现, 6 导头皮 EEG 信号受到了 ECG 和 EOG 信号的严重干扰.

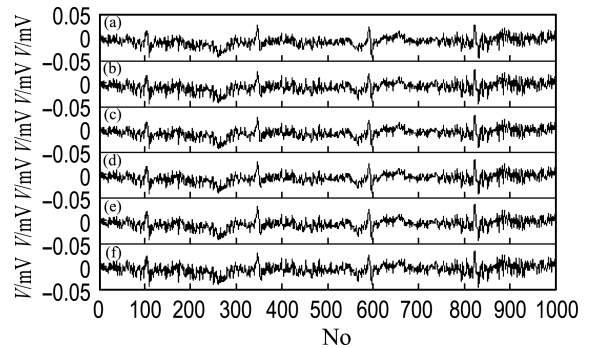


图 2 模拟头皮记录的 6 导脑电信号

Fig. 2 The six simulated EEGs recorded from scalp skin

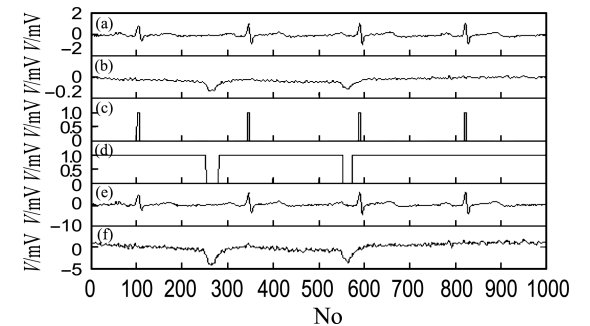


图 3 加入伪差后的脑电信号

Fig. 3 EEG added ECG and EOG

图 3 中的 (c)、(d) 为两导参考输入, 它们分别是对同步记录的 ECG、EOG 做阈值处理得到的

矩形波. (e) 和 (f) 分别是对应两导参考输入的算法输出. 对比图 3 中的几个波形, 可以发现 (e) 与 (a)、(f) 与 (b) 的波形非常接近. (a) 与 (e) 的相关系数达到了 0.999 2, 而 (b) 和 (f) 的相关系数也达到了 0.990 4, 说明本算法基本上成功地分离出了 ECG 和 EOG 伪差.

图 4 是去除伪差后的多导头皮 EEG 信号. 与图 2 的观测信号相比, 基本上滤除了 ECG 和 EOG 伪差干扰. 为了更加清晰地描述去除脑电伪差的效果, 将图 4 中去除伪差后的脑电输出与仿真时未加入伪差的观测信号进行了比较. 主要从二者的相关系数、误差均值、误差功率的角度进行考察, 如表 1. 实验结果表明, 本算法不但能够有效地去除脑电伪差, 而且原有的脑电信息几乎没有损失.

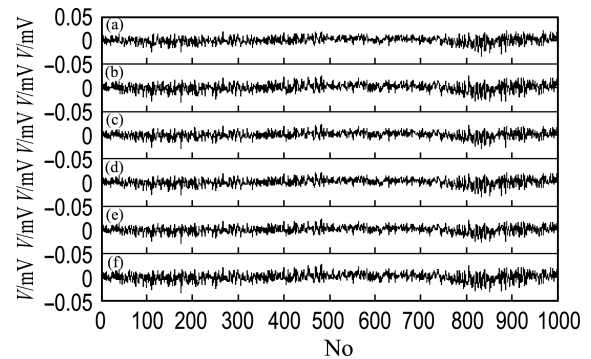


图 4 消除心电伪差和眼电伪差后的脑电信号

Fig. 4 The EEG removed ECG and EOG

表 1 去除伪差后的脑电输出与未加入伪差的脑电的比较

Tab. 1 EEGs comparison with and without ECG and EOG			
对应的导联	相关系数	误差均值/mV	误差功率
1	0.996 5	$0.512\ 5\times10^{-3}$	$0.446\ 8\times10^{-6}$
2	0.996 9	$0.586\ 3\times10^{-3}$	$0.563\ 4\times10^{-6}$
3	0.996 3	$0.539\ 5\times10^{-3}$	$0.492\ 9\times10^{-6}$
4	0.996 7	$0.515\ 9\times10^{-3}$	$0.462\ 5\times10^{-6}$
5	0.995 5	$0.588\ 3\times10^{-3}$	$0.604\ 8\times10^{-6}$
6	0.996 3	$0.617\ 7\times10^{-3}$	$0.682\ 0\times10^{-6}$

在实际处理的头皮 EEG 信号中, EOG 信号可以通过放置在眼睑上方的导联获取, 一般脑电图仪在记录 EEG 信号时都有此导联. 通常情况下, 脑电图仪不提供额外的电极记录同步的 ECG 信号. 此时, 可以构造一个参考信号. 具体做法是, 在头皮记录的未加工的 EEG 信号中, ECG 干扰

的幅度会较大,而且具有一定的准周期性,利用其特点,可以通过阈值法直接构造矩形波作为参考信号.

3 结 语

本文提出的算法是一种多参考信号的ICA方法,是一种介于fast ICA算法和ICA-R算法之间的算法.一方面,由于本算法只提取某几路源信号,相比于fast ICA算法分离出所有的源信号,具有更小的计算量,而且无需对分离的独立源进行人工判别.另一方面,本文提出的算法实际上对ICA-R算法进行了一定的扩展,解决了ICA-R算法只能提取某一路源信号的缺点,当然,当本算法中的参考信号为一导时,算法就退化为ICA-R算法.

实验表明,本文提出的算法更切合实际情况,而且能够更加有效地去除脑电中的多个伪差.但是,与ICA-R算法类似,本算法也存在一定的局限性.算法常常用来去除一些幅度较大、容易获取参考信号的伪差,如本文中的眼电伪差、心电伪差等.对于其他的伪差,如肌电干扰、血管波干扰,还需要进一步地研究获取参考信号的方法,才能应用本文多参考信号的算法.

参考文献：

[1] 周昌贵. 临床脑电图手册[M]. 成都:四川辞书出版社,1990

[2] 杨福生,洪 波. 独立分量分析的原理与应用[M]. 北京:清华大学出版社,2006

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

[4] HYVARINEN A. New approximations of differential entropy for independent component analysis and project pursuit [J]. **Advances in Neural Information Processing System**, 1998, **10**:273-279

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

[6] BARROS A K, VIGARIO R. Extraction of event-related signals from multichannel bioelectrical measurements [J]. **IEEE Transaction on Biomedical Engineering**, 2000, **47**(5): 583-588

[7] GOLDBERGER A L, AMARAL L, GLASS L, *et al.* PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals [J/OL]. (2000-01-13) **Circulation**, 2000, **101**(23):215-220

ICA method with multi-references and its application to removing artifacts from brain signals

QIU Tian-shuang*, WU Jian

(School of Electronic and Information Engineering, Dalian University of Technology, Dalian 116024,China)

Abstract: In order to eliminate the effect of artifacts in electroencephalograph (EEG) signals, such as electrocardiogram (ECG) and electrooculogram (EOG), etc., a new ICA (independent component analysis) method with multi-references is proposed based on the fixed-point algorithm and the ICA algorithm with references. After the proportions of artifacts in observation signals are calculated respectively, the comparatively pure EEGs are separated from the artifacts. Compared with fast ICA, the new method has less computational complexity and doesn't need to dispose the independent sources. In addition, it can overcome the defect of ICA-R algorithm which just extracts single signal. Computer simulation demonstrates that the new method can effectively remove several artifacts from brain signals.

Key words: independent component analysis (ICA); reference signals; artifacts removing; electroencephalograph (EEG) signal