

基于 P300 和机器学习的测谎方法研究

高军峰, 王沛, 郑崇勋

(西安交通大学生命科学与技术学院, 710049, 西安)

摘要: 为了克服传统测谎方法没有考虑到相同刺激下受试者思维状态变化的缺点,提出了基于 P300 和机器学习的测谎方法. 该方法使用典型的 3 刺激测谎范式,首先记录 30 名随机划分的撒谎者和无辜者的 12 导脑电(EEG)信号,使用独立成分分析方法(ICA)分解由探针刺激产生的脑电信号,利用在 Pz 电极上分布强度大的独立分量重建 Pz 波形,将每名受试者的若干个 Pz 波形进行平均,提取两步降噪后的每个 Pz 波形的时域和小波特征. 最后,使用分类器区分 P300 和非 P300 波形,进一步计算出个体测谎诊断率. 实验结果表明,支持向量机(SVM)适合于说谎意识状态的分类,提出的方法可以有效地改进单次刺激记录上的信噪比,提高 P300 成分的识别率,进而提高测谎诊断率.

关键词: 测谎;独立成分分析;脑电;P300;两步降噪;支持向量机

中图分类号: R318 **文献标志码:** A **文章编号:** 0253-987X(2010)10-0120-05

Lie Detection Method Based on P300 and Machine Learning

GAO Junfeng, WANG Pei, ZHENG Chongxun

(School of Life Sciences and Technology, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: For solving the unvariability of subject cognitive states under the same kind of stimulus in the conventional machine learning methods, the method based on P300 and machine learning was proposed. The standard three-stimuli protocol was chosen. Thirty guilty and innocent subjects were randomly divided into two groups and their EEG signals were first recoded. Independent component analysis (ICA) was carried out to decompose the datasets in the probe stimuli. The ICs with the largest projection strength at Pz were selected to reconstruct the Pz waveforms. Then small number of Pz waveforms within each subject is further averaged. Afterwards, the time-domain and wavelet features were extracted from each denoised Pz waveforms. In terms of the classifier to identify the P300 and non-P300 waveforms, the individual diagnostic rate was evaluated. The experimental results show that the SVM classifier is suitable to identify the sense states of lying, and the proposed method enables to improve the SNR in single trails, enhancing the accuracy of identifying the P300 and of individual diagnostic rate.

Keywords: lie detection; independent component analysis; electroencephalo graph; P300; two-step denoising; support vector machine

传统的多道图测谎方法已经受到来自各方面的挑战^[1],很多研究人员开始使用诱发脑电(ERP)信号的 P300 成分进行测谎. P300 成分是对小概率新异事件的响应^[2]. 尽管潜伏期和波峰幅度是检测

P300 波形的两个典型特征,但是由于在单次响应下的诱发脑电信号的信噪比(SNR)极低(除了 P300 成分,其余均为噪声信号,包括自发脑电),所以直接使用 P300 波形的上述两个特征进行测谎仍然非常

困难.

目前,使用较多的测谎方法有引导带幅度区分(BAD)^[3]、引导带相关区分(BCD)^[4]和模式识别 3 种方法,它们基本上使用标准的 3 刺激模式^[5]. 对于 BAD、BCD 方法,虽然使用若干个单次波形平均的方法来降噪,但是缺少对单次波形的丰富变化信息的分析. 另外,它们与传统的模式识别方法有两个共同的缺点:①P300 成分与噪声信号在时间和空间上是重叠的,因此普通方法很难提取 P300 成分;②实际测试使用时,仍需要大量的刺激提交给受试者. 实际测谎的对象往往是嫌疑犯,而且关于犯罪场景和犯罪者信息往往很少,嫌疑犯更易于对少量信息的反复刺激使用反测谎策略. 很多研究人员已经发现了“非靶刺激出现 P300”和“靶刺激出现非 P300”现象^[5-6],我们称它们为“错误的 P300 响应”. 这些方法都没有考虑这种大脑思维状态的变化引起的“错误的 P300 响应”. 这就使得训练和测试都是基于任务的,而不是基于 P300 的,显然这样会影响基于 P300 的测谎的准确率.

本文首先利用 P300 成分与其他噪声源基本独立的原理,使用独立成分分析(ICA),提取与噪声源独立的 P300 成分,并重建 Pz 电极上的波形,然后将少量的相同刺激下的 Pz 电极上的波形进行平均以进一步提高 SNR. 进一步地,从这些降噪后的波形上提取若干个选择的特征,最后分类器在特征向量上进行训练并测试.

1 研究方法

1.1 ICA 原理

ICA 是近年来发展起来的一种多通道盲源分离算法,其基本思想是将观测信号按统计独立原则分解为若干独立信号源成分^[7]. 在 ICA 方法中,观察 $\mathbf{X}(t)=[x_1(t), x_2(t), \dots, x_K(t)]^T, t=1, 2, \dots, N$, 其中 N 是抽样点数, K 是感知器个数. 对于 ICA 方法而言, $\mathbf{X}(t)$ 被看作是未知的 K 个独立源 $\mathbf{S}(t)=[s_1(t), s_2(t), \dots, s_K(t)]^T$ 的未知的线性混合, $\mathbf{X}(t)=\mathbf{AS}(t)$. ICA 方法试图通过引入解混矩阵 $\mathbf{W}$ 来计算未知的 $\mathbf{S}(t)$ 信号源. 在理想情况下, $\mathbf{W}$ 是混合矩阵 $\mathbf{A}$ 的逆,所以 $\mathbf{W}^{-1}$ 也被称为混合矩阵. 近年来,很多研究人员都论证了 ICA 方法用于提取 P300 成分的有效性^[8],但是它们的计算都较为复杂. 根据 ICA 分解的混合矩阵 $\mathbf{W}^{-1}$,本文提出一个简单合理的方法提取 P300 成分,然后利用 ICA 的重建原始波形的的方法,重建在 Pz 电极上的电压波形,该步骤在提

出的方法中称为第 1 步降噪.

1.2 小波变换

ERPs 信号是典型的非平稳信号,包含 δ 、 θ 、 α 、 β 、 γ 等不同的频率成分^[9]. 时域特征不能反映时域波形的频率成分,而傅里叶变换只适合分析时不变信号,变换得到的功率谱无法看出在某个时间段内的时域波形的频率成分. 短时傅里叶变换(SFFT)是将时域波形划分为相同长度的时间窗,在每个时间窗内运行傅里叶变换. 该变换对高频成分不能提供很好的时域定位,对低频成分的频域分辨率较差. 小波变换^[10]对低频分析采用较长的时间窗,对高频分析采用较短的时间窗,因此该变换也被称为多尺度分析,它能同时对低频和低频成分提供比较好的时域和频域分辨率. 由于以上的特点,使得小波变换特别适合于 ERPs 信号的分析.

1.3 测谎协议

使用与文献[11]类似的测试协议和实验方法,受试者被分为说谎组和无辜组. 实验前准备好 6 个首饰和它们对应的图片,一个保险箱内放置 2 个首饰,让说谎组的受试者偷取其中 1 个,并记住 2 个首饰的外貌特征. 类似地,保险箱内只放置 1 个首饰并让无辜者记住该首饰的外貌特征. 6 个图片将以 1.6 s 的间隔随机地显示在屏幕上,每个图片持续 1.1 s. 对于说谎者而言,被他们偷取的首饰图片为探针(P)刺激. 对于无辜者而言,从 5 个未见过的首饰图片中随机选择一个作为 P 刺激. 对两类受试者,见过而没有偷取的首饰图片为靶(T)刺激,其余为无关(I)刺激. 对说谎受试者,当屏幕上显示 P 刺激时,要求它们按手柄上的“no”按钮表达他们并未偷取首饰的说谎意图. 对于其余刺激响应,两类受试者一律“说实话”,通过按手柄的“yes”按钮表达他们确实见过或者没有见过该首饰.

2 数据采集与处理

2.1 大脑 EEG 信号采集

实验挑选了 30 个健康的本校本科生和研究生作为受试者,平均年龄 24 岁. 电极放置在如下 12 个位置: Fp1, Fp2, F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, Oz. 采样频率为 500 Hz. 将 Neuroscan 公司的 Synamps 作为信号放大器,在线滤波器为 0.3~30 Hz. 垂直眼电极放置在右眼上下 2.5 cm 处,水平眼电极放置在两只眼睛的外眼角处,右耳电极作为参考电极. 每个图片在每次任务中显示 30 次,因此 P、T 和 I 刺激的概率分别为 0.17、0.17 和 0.66. 每人

做 5 次任务,每次任务时间为 5 min. 使用 EEGLAB 工具^[12]将连续记录的信号分解为 1.6 s/幅的数据集,即每个数据集为 14 800 的矩阵,然后将每人的 P 响应的数据集挑出,则撒谎组和无辜组各产生 2 250 个脑电(EEG)数据集. 使用 Neuroscan 工具的自带程序去除数据集中的垂直和水平眼电伪差,最后撒谎组和无辜组的数据集分别组成集合 R_1 和 R_2 .

2.2 第 1 步降噪

很多作者已经证实了 Pz 电极附近的脑区域会产生较大的 P300 成分. 基于此,一个合理的假设,即经过 ICA 分解后的独立成分中,在 Pz 电极上强度最大的独立成分是 P300 成分. 设 Z_c 代表第 C 个独立成分, W_i^{-1} 为 W^{-1} 的第 i 行所有元素, X_{Pz} 是 R_1 或者 R_2 中某个数据集的 Pz 电极上的脑电波形,因此第 1 步降噪的算法描述如下.

(1) 将 W_{Pz}^{-1} 按照降序排列,令 W_{Pz}^{-1} 为排序后的向量,并令 L 是排序的索引向量,即 L 中的第 i 个元素的值代表 W_{Pz}^{-1} 在 W_{Pz}^{-1} 中的位置;

(2) 设 m 代表重建 Pz 波形的 ICs 的个数,则重建后的 Pz 波形如下

$$X'_{Pz} = \sum_{j=1}^m Z_{L_j} W_{Pz_j}$$

(1)

令 m 从 1~8 变化,采用交叉验证(CV)程序进行训练和测试,当测试的准确率最大时, m 的值即为最佳值.

2.3 第 2 步降噪

上文指出,对于测谎实验的受试者,即使在相同类型的刺激下,大脑思维状态仍然会产生变化. 为了降低错误的 P300 响应对测谎结果的影响,在第 1 次降噪 Pz 波形的基础上,将每个受试者的少数 Pz 波形平均. 设 X'_{Pzq} 代表第 q 个 X'_{Pz} , k 代表平均的 Pz 波形个数,则第 2 步降噪过程可表达为

$$X''_h = \frac{1}{k} \sum_{q=(h-1)*k+1}^{h*k} X'_{Pzq}, \quad h = 1, 2, \dots, 2\,250/k$$

(2)

X''_h 即为降噪后的撒谎组或者无辜组中的第 h 个 Pz 波形,最后它们组成撒谎组和无辜组的两个新集合 R_d 和 I_d . 类似地, k 的最佳值可以用训练的 CV 程序决定.

2.4 特征提取

对两步降噪后的每个 Pz 波形,抽取以下特征: ①最大信号值;②最小信号值;③潜伏期;④小波特征. 使用二次 B 样条函数作为母小波^[13],小波变换

后,产生了 7 个波段 0.3~3.9、3.9~7.8、7.8~15.6、15.6~31.2、31.2~62.5、62.5~125、125~250 Hz 对应的 7 个小波系数集. 由于 P300 成分的频率范围已经被证实基本在 delta 波段,因此 0.3~3.9 Hz 对应的 27 个小波系数被作为最后的小波特征.

2.5 分类器选择

为了选择较佳的分类器,采用线性分类器(FDA)、BP 神经网络(BPNN)和支持向量机(SVM)^[14]进行分类,并将它们的分类准确率进行比较. 在 BPNN 中,隐层节点数为 5~15,学习率为 0.01~0.05. 控制精度为 0.001~0.005. 在 SVM 中,取不同的核函数和相应的参数训练. 训练中,使用 10 则交叉验证,当 10 次测试准确率的平均值达到最大值时,可以决定 m 、 k 和各个分类器参数的最佳值.

3 结 果

3.1 预处理结果

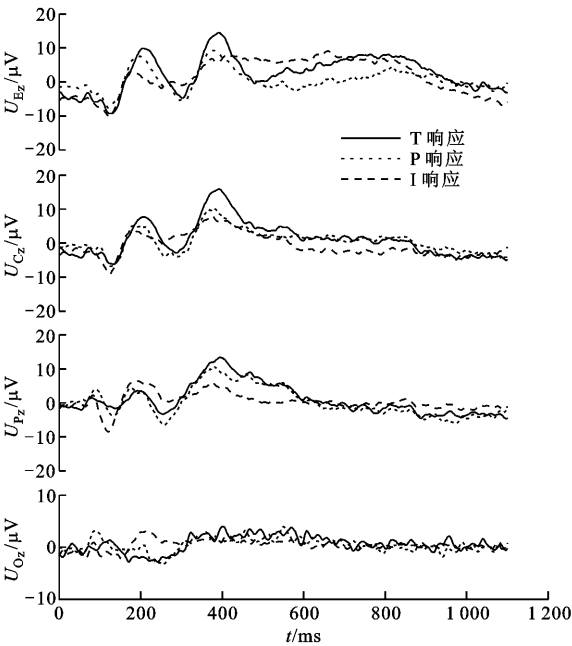
选择说谎者和无辜者各一名,挑出他们在 3 种刺激下的响应,即 EEG 数据集,对 Fz、Cz、Pz 和 Oz 电极上的 EEG 信号以刺激类型分组,然后对它们进行简单的 10 次平均,结果显示在图 1 中. 从该图可以看出,说谎者在 Pz 电极上 P、T 响应的 P300 特征比较相似,而且他们与 I 响应的 P300 特征差别比较大. 相反对于无辜者,在每个电极上, P 响应均和 I 响应比较相似. 以上结果表明,将 Pz 作为特征提取的电极,区分 P300 与非 P300 成分,进而对说谎者和无辜者进行判断是合理的.

3.2 分类结果

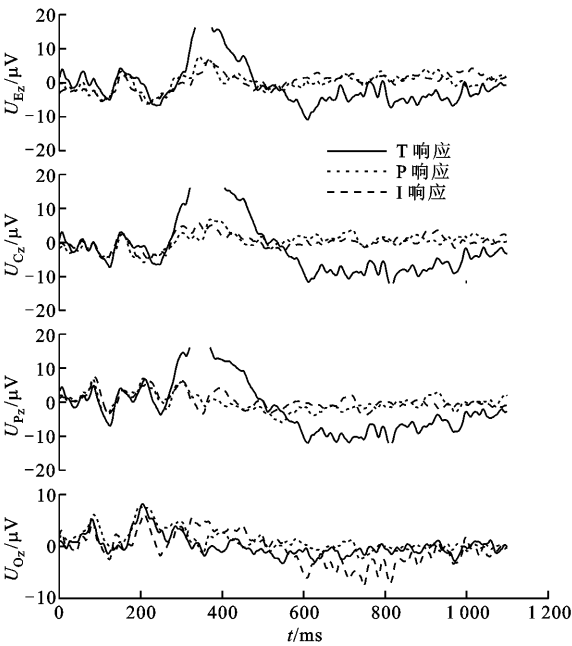
将各个分类器的准确率分为敏感度和特异度列在表 1 中,它们是基于 10 则交叉验证的 10 个准确率值的均值和标准差. 表中各分类器对应的各个参数的最佳值为: ①对于 BPNN,隐层节点数为 5,学习率为 0.02,控制精度为 0.002; ②对于 SVM,核函数为径向基(RBF),即 $K(x, y) = e^{-\|x-y\|^2/\gamma}$,函数参数 γ 值为 1 800; ③ $m=2$; ④ $k=5$.

表 1 各个分类器的训练和测试准确率

分类器	敏感度/%		特异度/%	
	训练	测试	训练	测试
FDA	61.45±3.42	61.99±2.90	63.87±4.17	60.61±3.88
BPNN	76.35±2.84	75.53±3.29	77.32±3.01	74.87±3.21
SVM	92.80±2.29	93.21±2.43	93.52±2.55	92.35±2.17



(a)说谎者



(b)无辜者

图 1 受试者对 3 种刺激 的响应

从表 1 可以看出,SVM 分类器的训练和测试的准确率均高于其他两个分类器.将 SVM 和 BPNN 两组数据的准确率进行了配对 t 检验,以 0.005 为显著差异的概率阈值,结果 $P<0.005$,即两组数据在统计上有显著差异. FDA 的分类效果并不理想,这表明与说谎有关的大脑意识形态线性不可分,较高的 SVM 分类结果表明用 SVM 作为分类器是合理的.如果选取 92%作为测试的阈值,提出的方法

将取得 100%的个体测谎诊断率,此值明显高于以往报道中的几种传统测谎方法的诊断率^[5].

4 讨 论

本文将测谎实验中相对于 P300 成分的噪声分为两类:①普通噪声,比如眼电伪差和自发脑电;②受试者对 P 刺激的错误响应.针对前者,我们使用 ICA 方法将与噪声对立的 P300 成分分离出来,然后重建时域波形.针对后者,我们使用少次平均的方法尽量将错误响应降到最低程度.两步降噪过程是本文提出的方法与传统方法的根本区别,也是保证系统具有灵活性、鲁棒性、高性能的关键.

BAD 和 BCD 测谎方法不能分析获得单次响应之间丰富的变化信息,传统的基于机器学习的测谎方法虽然是基于单次响应的分析,但是它们都没有考虑到相同刺激下受试者的大脑响应的变化.所以,这些方法都是基于任务的,而并非是基于 P300 的.

传统的测谎方法都有一个缺点,即只有等到所有的刺激结束后,才能测试判断某人是否说谎.很少的犯罪相关信息的反复刺激除了容易使犯罪嫌疑人产生疲劳,也极易导致反测谎行为的产生.总之,这些测谎方法并不灵活也不实用.本文提出的方法在测试使用时,只需要提交 5 次 P 刺激,通过提出系统的分析,最后使用训练过的分类器进行分类,即可判定说谎行为.可见,提出的方法是基于少次响应的分析,相对于 BAD 和 BCD 方法,仍然能从一定程度上反映响应之间丰富的变化信息.另外,两步降噪的过程也使得提出的方法是基于 P300 的,对于一个基于 P300 的测谎系统而言,就能得到更准确的测谎诊断结果.

在实际操作时,我们建议将单次的测谎过程(即提交 5 次 P 刺激)重复运行几次,然后综合几次测试的结果再进行最后的判定,将会获得更为鲁棒的个体诊断结果.

传统的测谎方法中,都有一个阈值难于决定的问题^[5],该值的选取必须在敏感度和特异度之间取得较佳的平衡,而本文提出的方法并没有此类问题.实验结果表明,如果选取 92%作为测试的阈值,本文方法将取得 100%的个体测谎诊断率.

参考文献:

[1] ABOOTALABI V, MORADI M H, KHALILZADEH M A. A comparison of methods for ERP assessment in a P300-based GKT[J]. International Journal of Psy-

- chophysiology, 2006, 62 (2):309-320.
- [2] POLICH J. P300 in clinical applications: meaning, method, and measurement[J]. American Journal of EEG Technology, 1991, 31 (3):201-231.
- [3] ROSENFELD J P, CANTWELL B, NASMAN V T, et al. A modified event-related potential-based guilty knowledge test [J]. International Journal of Neuroscience, 1988, 42 (1/2):157-161.
- [4] FARWELL L A, DONCHIN E. The truth will out: interrogative polygraphy ("lie detection") with event-related potentials [J]. Psychophysiology, 1991, 28 (5):531-547.
- [5] ABOOTALEBI V, MORADI M H, KHALILZADEH M A. A new approach for EEG feature extraction in P300-based lie detection[J]. Computer Methods and Programs in Biomedicine, 2009, 94 (1):48-57.
- [6] RÖSCHKE J, MANN K, WAGNER P, et al. An approach to single trial analysis of event-related potentials based on signal detection theory[J]. International Journal of Psychophysiology, 1996, 22 (3):155-162.
- [7] JUNG T P, MAKEIG S, HUMPHRIES C, et al. Removing electroencephalographic artifacts by blind source separation [J]. Psychophysiology, 2000, 37 (2):163-178.
- [8] LIN C T, CHUNG I F, KO L W, et al. EEG-based assessment of driver cognitive responses in a dynamic virtual-reality driving environment[J]. IEEE Transactions on Biomedical Engineering, 2007, 54 (7):1349-1352.
- [9] HERRMANN C S, KNIGHT R T. Mechanisms of human attention: event related potentials and oscillations[J]. Neuroscience and Biobehavioral Reviews, 2001, 25 (6):465-476.
- [10] HSU W Y, LIN C C, JU M S, et al. Wavelet-based fractal features with active segment selection: application to single-trial EEG data[J]. Journal of Neuroscience Methods, 2007, 163 (1):145-160.
- [11] 周亮, 杨文俊, 廖四照, 等. 模拟盗窃者与熟悉现场者在测谎实验中事件相关电位的比较研究[J]. 中国临床心理学杂志, 2000, 8(2):86-88.
ZHOU Liang, YANG Wenjun, LIAO Sizhao, et al. Compared study of event-related potentials in lie detection between subjects for simulated crime condition and subjects for visiting condition [J]. Chinese Journal of Clinical Psychology, 2000, 8(2):86-88.
- [12] DELORME A, MAKEIG S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis [J]. Journal of Neuroscience Methods, 2004, 134 (1):9-21.
- [13] QUIROGA R Q, SAKOWITZ O W, BASAR E, et al. Wavelet transform in the analysis of the frequency composition of evoked potentials[J]. Brain Research Protocols, 2001, 8 (1):16-24.
- [14] SHOKER L, SANEI S, CHAMBERS J. Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm[J]. IEEE Signal Processing Letters, 2005, 12 (10):721-724.

(编辑 杜秀杰)