

利用小波分解和支持向量机的心理意识 真实性识别研究

赵敏¹, 郑崇勋¹, 赵春临^{1,2}

(1. 西安交通大学生物医学工程教育部重点实验室, 710049, 西安; 2. 武警工程学院通信工程系, 710086, 西安)

摘要: 采用小波分解和支持向量机(SVM)技术,提出了一种对说谎脑电(EEG)信号特征进行分类的方法,将其应用于心理意识真实性的检测,获得了满意的结果.以真伪已明确的有意义的个人信息(如姓名、生日)作为被测试的隐藏信息,应用隐藏信息(CIT)测试模式对15名受试者各进行两组测试,并记录其脑电(EEG)信号.提取了探测刺激和无关刺激诱发EEG信号的小波系数,并应用具有统计学意义的特征参数作为SVM分类器的输入进行识别分类.实验结果显示,应用leave-one-out交叉验证法对30组样本数据进行训练测试,获得平均正确识别率为88.3%.因此,该方法可以作为一种心理意识真实性检测的新方法,具有无创、较高正确检测率等优点.

关键词: 心理意识;小波分解;支持向量机;测谎

中图分类号: R318 **文献标志码:** A **文章编号:** 0253-987X(2010)04-0119-06

Identification of Mentality Facticity Based on Wavelet Decomposition and Support Vector Machines

ZHAO Min¹, ZHENG Chongxun¹, ZHAO Chunlin^{1,2}

(1. Key Laboratory of Biomedical Information Engineering of Education Ministry, Xi'an Jiaotong University, Xi'an 710049, China;

2. Communication Engineering Department, Army Police Engineering College, Xi'an 710086, China)

Abstract: A practical strategy for classifying the lying electroencephalograph (EEG) characters by wavelet decomposition and support vector machines(SVM) techniques is presented to get a satisfactory results in identifying the mentality facticity. Some significant personal information is ensured, such as name and birthday, and selected as the concealed information. 15 subjects participate in two groups of concealed information tests (CIT) and their EEGs are recorded. Applying wavelet decomposition, the wavelet coefficients corresponding to EEG evoked by probe information and by irrelevant information respectively are evaluated. Then the feature coefficients containing statistical significance are extracted as the input parameters of SVM. 30 samples are chosen to train and test the performance of classifier by leave-one-out cross-validation, 88.3% accuracy can be achieved in probe information detection.

Keywords: mentality; wavelet decomposition; support vector machines; lie detection

心理意识真实性是人脑对客观存在的反映,通常人们在社会交往中利用各种行为或语言来掩饰或否定自己的真实意识,以达到欺骗对方的目的,其多数含有刻意的欺骗,成为社会和经济发展的不安定

因素.因此,心理意识真实性检测,俗称测谎,成为刑侦、反恐、安全审查、认知损伤评估及人事评测等部门迫切需要而十分关注的技术.目前,普遍使用的测谎技术是多道心理测试,由于其测试机理主要依据

的是人在说谎时引发的由自主神经系统调节的生理反应值的异常,而这些反应值极易受到主观心理因素(比如焦虑、紧张等)的影响,其有效性和准确性不断受到质疑和挑战^[1].

脑电(EEG)信号的相关研究发现,其有效成分可以反映大脑的认知和注意机制,而欺骗性的决定是在对信息认知基础上做出的.因此,国内外学者逐渐转向从大脑的注意和认知机理方面来研究测谎技术,所依据的假设是,不管受试者做出的回答是“是”或者是“不是”,在其给出判断反应之前,受试者想隐藏的对自己认识并较熟悉的真实信息相对于其他同类不熟悉的非真的无关信息,会诱发出更大的事件相关电位值^[2].所以,基于 EEG 信号的心理意识真实性识别技术更具有认知决定性,不易受人为的主观因素的影响,相对于多道心理测谎技术在有效性和准确性方面更具有优势.

目前,说谎与否产生的事件相关电位的特征提取主要是 P300 幅值和潜伏期^[1-2],识别方法主要是基于重采样技术的自举幅值差异法(BAD)和自举相关系数差异法(BCD)的阈值分类模型^[1],该方法简单,信息单一,识别率不高.小波变换可以提取需要频段的小波系数,各小波系数实际上是信号时频特征的一种表现,适用于表达信号的多维信息和特征.因此,本研究期望在小波特征提取的基础上,选择适当的分类学习方法以更有效地识别受试者在隐藏信息(CIT)测试模式下的心理意识真实性.

支持向量机(SVM)^[3]学习方法能有效地避免过学习、维数灾难、局部极小等传统分类方法(如神经网络、决策树)存在的问题,而且在小样本条件下仍然具有良好的泛化能力,近年来被广泛应用到生物信息识别领域^[4],并取得了很好的效果.因此,本研究选用 SVM 作为分类器对提取的小波特征进行分类识别.

1 实验设计与数据采集

实验对象为在校学生自愿者 15 名,包括本科生和研究生,没有任何精神疾病和脑病病史,男性 11 名,女性 4 名,年龄为 21~24 岁,正常或矫正视力.实验时间为上午 9:30~10:30 或下午 3:30~4:30,整个实验过程在正常照明、安静、尽量舒适的室内环境中进行.每个受试者完成两组实验,要求受试者在测试过程中尽量避免身体移动.

测试方法为采用受试者的真实姓名或无关姓名和真实生日或无关日期作为两组实验的被测试信

息,以“你是 XXX?”、“生日是 X 月 X 日?”的形式出现,测试序列的编排和出现依据典型的 CIT 范式^[2,4],包括以下 3 类刺激信息.

(1)探测刺激,测试者和被测试者都知晓的真实情况(真实姓名或生日),但受试者试图隐瞒并故意否认的项,回答“不是”,即说谎项.

(2)无关刺激,与探测刺激同类型的非真实情况(无关姓名或日期),要求受试者确实否认的项,回答“不是”,在本实验的两组测试中分别是 5 个陌生人的姓名和与生日无关的日期.

(3)靶刺激,由测试者专门指定的无关刺激,要求受试者当该刺激信息出现时完成与其他两类信息相反的反应任务,即做肯定回答“是”,其作用主要是要求受试者在测试过程中集中注意力并监督其是否配合实验.在本实验中,为了统一回答问题时“是”和“不是”的对应关系,我们对靶刺激信息做了修改,以我们指定一个受试者认识人的姓名,以“你认识 XX?”和“国庆是 10 月 1 日?”的提问方式作为靶刺激.

每组测试中共 160 次刺激随机出现,其中无关刺激 112 次(70%),探测刺激和靶刺激各 24 次(15%),刺激呈现的时间为 1.5 s,刺激之间间隔在 2.0~2.2 s.受试者按照实验要求,对随机出现的刺激问题判断后以尽快按下鼠标“左键”或“右键”的形式回答“是”或“不是”.一组测试完成后受试者休息 5 min 后进行下一组. EEG 数据采集用标准 10-20 导联头皮电极系统 32 导电极帽,记录 FP1、FP2、F7、F3、Fz、F4、F8、FT7、FC3、FCz、FC4、FT8、T7、C3、Cz、C4、T8、TP7、CP3、CPz、CP4、TP8、P7、P3、Pz、P4、P8、O1、Oz、O2 共 30 个通道的脑电信号,同时记录垂直眼电(EOG),参考电极 of 双侧乳突,前额接地,放大器的通道频带为 0.3~70 Hz,频率为 1 kHz,各导联阻抗均小于 5 k Ω .

采集的数据用 Neuroscan 软件,离线去除眼电伪迹后,取刺激出现前 500 ms 到后 1 500 ms 的数据分段,分段数据的伪迹去除标准是 70 μ V,平均所有去除伪迹后的分段数据得到各类刺激的诱发脑电信号.采用 SPSS16.0 和 Matlab7.0 软件,实现后续的统计分析和分类识别.

2 脑电信号特征提取

2.1 离散小波变换

小波变换是傅里叶变换的新发展,小波系数能反映信号在时域和频域的局部信息.因而,小波分析作为一种时频分析方法,在生物医学信号处理方面

有着广阔的应用前景,特别适合于脑电信号的去噪、降维和特征提取等处理^[5-6].

设 $x(n)$ 为实验采集的 EEG 离散信号,则 $x(n)$ 的离散小波变换定义^[7]为

$$C_{j,k}(n,\varphi_{j,k})=2^{-j/2}\sum_{n=-\infty}^{\infty}x(n)\overline{\varphi}(2^{-j}n-k),j,k\in Z$$

(1)

$x(n)$ 的离散小波逆变换定义为

$$x(n)=\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}C_{j,k}\varphi_{j,k}(n)=\sum_{j=-N}^{\infty}x_j(n),j,k\in Z$$

(2)

式中: $\varphi_{j,k}(n)=2^{-j/2}\varphi(2^{-j}n-k)$ 为小波序列; $\varphi(n)$ 是满足一定条件的小波基函数; j,k 分别代表频率分辨率和时间平移量; $x_j(n)$ 表示信号 $x(n)$ 在某一尺度 (2^j) 的分量.采用 Mallat 算法,对信号进行有限层分解,即

$$x(n)=A_L+\sum_{j=1}^LD_j$$

(3)

式中: L 为分解层数; A_L 为低通逼近分量; D_j 为不同尺度下的细节分量.信号的整个频带划分为多个子频带,设信号 $x(n)$ 的采样频率为 f_s ,则 $A_L,D_L,D_{L-1},\cdots,D_1$ 各分量所对应的子频带依次为

$$\left[0,\frac{f_s}{2^{L+1}}\right],\left[\frac{f_s}{2^{L+1}},\frac{f_s}{2^L}\right],\left[\frac{f_s}{2^L},\frac{f_s}{2^{L-1}}\right],\cdots,\left[\frac{f_s}{2^2},\frac{f_s}{2}\right]$$

2.2 脑电信号的小波分解

脑电信号的采样频率是 1 kHz,已有研究显示^[1-2],与认知相关的意识真实性事件相关脑电在

Fz、Cz 和 Pz 电极最为明显,本文主要分析 Fz、Cz 和 Pz 3 个电极的脑电信号.应用 Bior1.3 小波分别对无关刺激和探测刺激诱发的 0~1 024 ms 的脑电信号进行 7 层分解,即 $x(n)=A_7+D_7+\cdots+D_1$,各分量对应的子频带依次为 0~3.9、3.9~7.8、7.8~15.6、15.6~31.2、31.2~62.5、62.5~125、125~250、250~500 Hz.由于事件相关脑电信号的有意义成分主要分布在低频带,本文只对 A_7,D_7,D_6 频带的小波系数进行分析,为了方面描述,下文分别标记为 A、B、C 频带.

2.3 特征提取

图 1 为两类刺激信息在 Fz、Cz、Pz 电极诱发的事件相关脑电信号,从中可以看出,两类刺激信息诱发的脑电信号在刺激出现后呈现明显的不同.

对于分解得到的 A、B、C 频带共 32 个小波系数(A 子带 8 个,B 子带 8 个,C 子带 16 个),用独立样本 t 检验进行统计分析,其中刺激类型(无关刺激/探测刺激)作为组变量,差异性显著的小波系数结果(t,P)如表 1 所示. Fz、Cz 和 Pz 电极信号按 $A(a_1,a_2,\cdots,a_8)$ 、 $B(b_1,b_2,\cdots,b_8)$ 、 $C(c_1,c_2,\cdots,c_{16})$ 子带依次排列的 32 个归一化后的小波系数均值如图 2 所示.结果显示,在 Fz、Cz 和 Pz 3 个电极共有 26 个小波系数在无关刺激和探测刺激之间有显著性差异,其中差异性最为显著的系数为 a_4,a_5 (对应时间为 380~512 ms),与传统的事件相关电位 P300 一致.显然,我们可以选择具有显著差异性的 26 个小波系数作为分类器的特征向量.

表 1 小波系数特征统计差异显著性结果

Fz 电极			Cz 电极			Pz 电极		
特征系数	t	P	特征系数	t	P	特征系数	t	P
a_3^*	-2.646	0.010	a_4^{**}	-6.500	0.000	a_4^{**}	-4.169	0.000
a_4^{**}	-7.263	0.000	a_5^{**}	-4.678	0.000	a_5^{**}	-4.858	0.000
a_5^{**}	-3.861	0.000	a_8^{**}	2.927	0.005	a_6^*	-2.178	0.033
a_8^{**}	3.226	0.002	b_3^*	2.521	0.014	a_8^{**}	3.410	0.001
b_3^*	2.352	0.022	b_5^{**}	-2.841	0.007	b_3^{**}	2.821	0.007
b_5^{**}	-3.249	0.002	b_7^{**}	-2.801	0.008	b_5^*	-2.547	0.015
c_6^{**}	3.630	0.001	c_6^{**}	2.812	0.007	b_6^{**}	-3.000	0.004
c_{10}^*	-2.043	0.047	c_{11}^{**}	-4.286	0.000	b_7^{**}	-2.936	0.005
c_{11}^*	-2.599	0.012				c_{11}^{**}	-3.922	0.000

注:带 * 数据表示 $P<0.05$;带**数据表示 $P<0.01$.

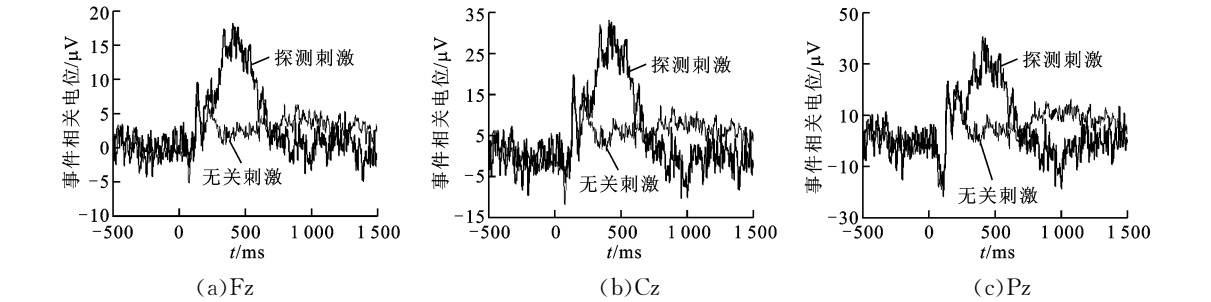


图 1 探测刺激和无关刺激在电极诱发的事件相关脑电信号

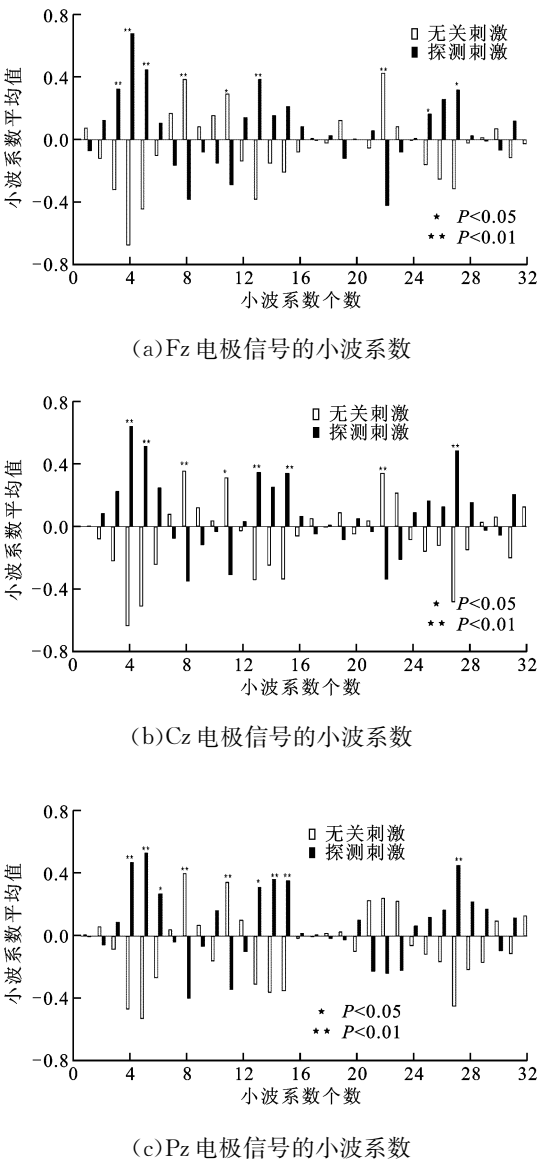


图 2 脑电信号的 32 个小波系数均值及统计学意义

3 支持向量机分类方法

SVM 是一种基于统计学习理论的模式识别方法^[3],是近年来在模式识别和机器学习领域中出现的新工具,以经验分类误差最小化和分类几何边缘

最大化为学习目的,而其他统计学习方法(如神经网络和决策树)仅仅以经验分类误差最小化为学习目的. 因此,SVM 学习算法能有效地避免过学习、维数灾难、局部极小等传统分类方法存在的问题,而且在小样本条件下仍然具有良好的泛化能力.

SVM 是通过构造最优分类超平面实现样本分类的,其二分类算法^[3]为:设训练样本集 $D=\{(x_i, y_i) \mid i=1, \cdots, n\}$, $x \in R^d$, $y \in \{-1, +1\}$, n 为训练样本的个数, d 为每个训练样本的维数, y 为类别标号(其值等于+1 的为一类,等于-1 的为另一类). 变换 $\varphi: x \rightarrow \Omega$ 将训练样本从输入空间映射到高维特征空间,核函数 $K(x_i, x_j) = \varphi(x_i) \varphi(x_j)$ 为对应高维特征空间 Ω 中的内积. 设计基于 SVM 的二分类器,就是寻找当训练样本集在 Ω 中线性可分时,使分类间隔最大的最优分类超平面 $W\varphi(x_i) + b = 0$,即在

$$y_i[W\varphi(x_i) + b] - 1 \geq 0, \quad i = 1, \cdots, n \quad (4)$$
的约束下求解

$$\min \varphi(W) = \frac{1}{2} \|W\|^2 \quad (5)$$

当训练样本集在 Ω 中线性不可分时,为保证分类的正确,引入松弛变量 $\xi_i \geq 0, i=1, \cdots, n$,即在

$$y_i[(W\varphi(x_i) + b) - 1 + \xi_i] \geq 0 \quad (6)$$
的约束下求解

$$\min \varphi(W, \xi_i) = \frac{1}{2} \|W\|^2 + C \sum_{i=1}^n \xi_i \quad (7)$$

式中: $C > 0$ 为惩罚因子,用于实现经验风险与置信范围的折衷.

由于特征空间的维数可能很高,甚至是无穷的,且变换 φ 并未直接给出,故不易直接求解. 该问题可转化为其对偶问题,即在

$$\sum_{i=1}^n y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \cdots, n \quad (8)$$
的约束下求解

$$\max Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i x_j) \quad (9)$$

式中: α_i 为约束条件式(6)对应的拉格朗日因子.

求得上述两个系数 α_i 、 b 对应的最优解 α_i^* 、 b^* 后,得到最优分类函数为

$$f(x) = \text{sgn}\left[\sum_{i=1}^n \alpha_i^* y_i K(x, x_i) + b^*\right] \quad (10)$$

核函数的主要作用是把样本空间映射到特征空间,不同的核函数对于算法具有一定的影响.核函数能够反映模型选择的复杂度,相对其他核函数,高斯核函数的核参数较少,可适用于任意分布的样本,也是目前在 SVM 中应用最为广泛的一种.因此,本研究选用高斯核函数

$$K(x, x_i) = \exp\left[-\frac{\|x - x_i\|^2}{2\sigma^2}\right] \quad (11)$$

按照上述方法,将训练集中对应无关刺激的样本归为负类,对应探测刺激的样本归为正类,并训练 SVM 分类器,然后对测试样本进行分类.本文为了克服小样本数据的缺陷,采用 leave-one-out 交叉验证法^[8]对分类器进行训练测试,即从所有样本中逐次选择 1 组样本留做测试,其余样本用来训练.

4 结果与讨论

根据上述方法,提取了在 CIT 测试模型下的事件相关脑电信号的小波系数特征,并建立基于小波特征的 SVM 意识真实性识别模型.由具有统计学意义的 26 个小波系数构成特征向量,共对 30 组样本数据采用 leave-one-out 交叉验证法训练及测试,获得了 88.3% 的正确识别率,测试结果见表 2.另外,为了评估本文提出方法的有效性,我们从研究方法和分类器性能两方面与其他方法进行了比较.

在研究方法方面,与 BAD 和 BCD 方法进行了比较,结果如表 3.从表中可以看出,基于小波特征的 SVM 方法(SVM-Wavelet)的正确识别率明显高于基于事件相关电位 P300 的 BAD(BAD-P300)和 BCD(BCD-P300)方法.这是因为事件相关脑电信号的小波系数除了能有效地表征事件相关电位 P300 (如系数 a_4 、 a_5) 的信息外,还可以提供比 P300 更丰富的多维可以分辨真实和非真实的有效信息.

在分类方法选择方面,将 SVM 分类器与反向传输(BP)神经网络方法进行了比较,也采用 leave-one-out 交叉验证法,分类结果见表 3.从表中可以看出,在相同训练测试样本的条件下,SVM 的分类性能(SVM-Wavelet)优于 BP 神经网络(BP-Wavelet)方法.

表 2 基于小波特征的 SVM 分类结果

	识别数	测试数	识别率/%
无关刺激	27	30	90.0
探测刺激	26	30	86.6
平均	53	60	88.3

表 3 基于小波特征的 SVM 方法与其他方法识别结果的比较

方法	识别率/ %		
	无关刺激	探测刺激	平均
SVM-Wavelet	90.0	86.6	88.3
BP-Wavelet	86.6	73.3	80.0
BAD-P300	86.6	80.0	83.3
BCD-P300	80.0	73.3	76.6

5 结 论

应用小波特征和支持向量机模型,对执行心理意识测试任务中记录的 EEG 信号模式进行分类识别.以受试者自己的姓名和生日作为实验探测刺激信息,尽管受试者对真实的探测刺激问题和无关刺激问题做了同样的否定回答,但是受试者对探测刺激信息较无关刺激信息在头皮记录的事件相关脑电信号中会表现出明显不同的反应.多层小波系数可以从时间和频率的角度反映刺激信息诱发脑电信号的多个特征,通过独立样本 t 检验统计分析显示,26 个小波系数在探测刺激和无关刺激之间存在显著性差异 $P<0.05$,并以此作为支持向量机分类模型的输入特征参数.与基于 P300 的 BAD 和 BCD 方法及 BP 神经网络模型的分类效果相比,本文提出的方法取得了更好的识别效果.

参考文献:

[1] ABOUTALEBI V, MORADI M H, KHALILZADEH M A. A comparison of methods for ERP assessment in a P300-based GKT [J]. International Journal of Psychophysiology, 2006,62(2): 309-320.

[2] ROSENFELD J P, ELLWANGER J W, NOLAN K, et al. P300 scalp amplitude distribution as an index of deception in a simulated cognitive deficit model[J]. International Journal of Psychophysiology, 1999,33(1): 3-19.

[3] VAPNIK V N. Statistical learning theory[M]. New York,USA: John Wiley and Sons Inc, 1998.

- [4] LOTTE F, CONGEDO M, LECUYER A, et al. A review of classification algorithms for EEG-based brain-computer interfaces[J]. Journal of Neural Engineering, 2007, 4(2): R1-R13.
- [5] MERZAGORA A C, BUNCE S, IZZETOGLU M et al. Wavelet analysis for EEG feature extraction in deception detection[C]//Proceedings of the 28th IEEE EMBS Annual International Conference. New York, USA: IEEE Press, 2006: 2434-2437.
- [6] 杨立才, 李金亮, 姚玉翠, 等. 基于小波分解和支持向量机的 P300 识别算法[J]. 中国生物医学工程学报, 2007, 26(6): 804-809.
- YANG Licai, LI Jinliang, YAO Yucui, et al. P300 detection algorithm based on wavelet decomposition and support vector machines[J]. Chinese Journal of Biomedical Engineering, 2007, 26(6): 804-809.
- [7] 杨福生. 小波变换的工程分析与应用[M]. 北京: 科学出版社, 2001.
- [8] MARTENS H A, DARDENNE P. Validation and verification of regression in small data sets [J]. Chemometrics and Intelligent Laboratory Systems, 1998, 44(1/2): 99-121.

(编辑 杜秀杰)