

基于P300和极限学习机的脑电测谎研究

高军峰^{1,2}, 张文佳¹, 杨勇³, 胡佳佳¹, 陶春毅¹, 官金安¹

(1. 中南民族大学认知科学国家民委重点实验室 武汉 430074; 2. 电子科技大学生命科学与技术学院 成都 610054;
3. 江西财经大学信息技术学院 南昌 330013)

【摘要】极限学习机基于一种典型的单隐层前馈神经网络(SLFNs),其有效性在模式识别很多领域得到证实。该文针对当前的测谎方法的准确率不够高及训练时间较长的缺点,将ELM算法应用到测谎研究领域,作为分类器,对说谎者和诚实者的两类脑电信号进行分类识别,并将实验结果和三类典型的分类器:支持向量机(SVM)、人工神经网络(ANN)和线性分类器(FDA)的分类结果进行比较。实验结果表明,该方法不仅获得最高的训练和测试准确率,而且训练时间也大为缩短,证明了该方法的测谎有效性。

关键词 脑电; 极限学习机; 测谎; 神经网络; P300; 支持向量机

中图分类号 R318

文献标志码 A

doi:10.3969/j.issn.1001-0548.2014.02.028

Lie Detection Study Based on P300 and Extreme Learning Machine

GAO Jun-feng^{1,2}, ZHANG Wen-jia¹, YANG Yong³, HU Jia-jia¹, TAO Chun-yi¹ and GUAN Jin-an¹

(1. Key Laboratory of Cognitive Science, South-Central University for Nationalities, State Ethnic Affairs Commission Wuhan 430074;

2. School of Life Science and Technology, University of Electronic Science and Technology of China Chengdu 610054;

3. School of Information Technology, Jiangxi University of Finance and Economics Nanchang 330013)

Abstract Extreme learning machine (ELM) is a typical SLFN (single layer feedback network) and its efficiency has been proved by many literatures for pattern recognitions. In this paper, ELM is applied in lie detection for the first time in order to overcome the disadvantages of the current lie detection methods such as lower accuracy and slower training speed. ELM is used as a classifier to classify the guilty and innocent subjects. The experimental result is compared with support vector machine (SVM), artificial neural network (ANN) and fisher discrimination analysis (FDA). The comparison results show that the proposed method obtains the highest training and testing accuracy with the fastest training speed.

Key words EEG; extreme learning machine; lie detection; neural network; P300; support vector machine

近年来,基于生理心理学的测谎技术日趋成熟。脑电信号事件相关电位P300成分是对小概率新异刺激响应^[1]。很多研究人员发现,基于P300的测谎方法能有效地规避反测谎的影响,具有更高的准确性^[2]。

在分类器的选择上,针对支持向量机(support vector machine, SVM)^[3]和人工神经网络(artificial neural network, ANN)训练速度慢、分类器参数选择难、易产生局部最优解等问题,本文提出了基于极限学习机(extreme learning machine, ELM)作为分类器的一种新的测谎方法。文献[4]提出了ELM算法,是一种简单有效的基于单隐层前馈神经网络(single

layer feedback networks, SLFNs)的学习算法。在算法的执行过程中,无需调整网络的输入权重以及隐层偏置,与传统的ANN不同的是,计算过程并不是基于迭代的计算方式。

本文采集了30个说谎和诚实受试者的EEG信号,将两类受试者在Pz电极上的探针刺激响应的EEG信号作为P300和非P300信号做进一步分析,然后提取基于频域和时-频域的多维特征,最后利用ELM对特征向量进行训练和分类,并通过训练时间、测试时间、训练准确率、测试准确率及网络复杂度等指标分别与反向传播神经网络(back-propagation network, BPNN)、支持向量机(SVM)及线性分类器

收稿日期: 2012-11-29; 修回日期: 2013-12-02

基金项目: 国家自然科学基金(81271659, 61262034, 91120017); 江西省自然科学基金(20114BAB211020, 20132BAB201025); 江西省教育厅科技项目(GJJ13302)。

作者简介: 高军峰(1973-), 男, 博士, 主要从事脑电信号处理与模式识别方面的研究。

(fisher discrimination analysis, FDA)进行比较。

1 极限学习机的理论基础与算法

1.1 SLFN的统一模型

给定 N 个训练样本 $(\mathbf{x}_i, \mathbf{t}_i)$, $i=1,2,\dots,N$, 其中 $\mathbf{x}_i=[x_{i1}, x_{i2}, \dots, x_{im}]^T \in R^m$, $\mathbf{t}_i=[t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$ 为样本导师。激活函数为 $g(x)$ 的标准单隐层前馈神经网络可表述为:

$$\sum_{i=1}^{\tilde{N}} \beta_i g(\mathbf{w}_i \mathbf{x}_j + b_i) = \mathbf{o}_j \quad j=1,2,\dots,N \quad (1)$$

式中, \tilde{N} 为隐层节点个数; $\mathbf{w}_i=[w_{i1}, w_{i2}, \dots, w_{im}]^T$ 为连接第 i 个隐层节点与输入神经元的权重向量;

$$\mathbf{H}(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{\tilde{N}}, b_1, b_2, \dots, b_{\tilde{N}}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & g(\mathbf{w}_2 \cdot \mathbf{x}_1 + b_2) & \dots & g(\mathbf{w}_{\tilde{N}} \cdot \mathbf{x}_1 + b_{\tilde{N}}) \\ g(\mathbf{w}_1 \cdot \mathbf{x}_2 + b_1) & g(\mathbf{w}_2 \cdot \mathbf{x}_2 + b_2) & \dots & g(\mathbf{w}_{\tilde{N}} \cdot \mathbf{x}_2 + b_{\tilde{N}}) \\ \vdots & \vdots & \dots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & g(\mathbf{w}_2 \cdot \mathbf{x}_N + b_2) & \dots & g(\mathbf{w}_{\tilde{N}} \cdot \mathbf{x}_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}} \quad (4)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \quad \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \mathbf{t}_2^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m} \quad (5)$$

式中, \mathbf{H} 为该神经网络的隐层输出矩阵, \mathbf{H} 中的第 i 列是第 i 个隐层节点对应于 N 个输入样本 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ 的输出向量。

1.2 ELM算法

在传统的函数逼近理论中, 需要人为设置大量的网络训练参数, 且很容易产生局部最优解。与传统的函数逼近理论不同, 在极限学习机算法中, 若激活函数无限可微, 则输入权重与隐层偏置能够随机选择。对于前馈神经网络, 无需人为调整输入权重 \mathbf{w}_i 和隐层偏置 b_i , 一旦这些参数被随机分配后, 在网络开始训练时, 隐层输出矩阵 \mathbf{H} 便保持不变。

换言之, 单隐层前馈神经网络的训练过程, 等价于寻找线性系统 $\mathbf{H}\beta = \mathbf{T}$ 的最小二乘解 $\hat{\beta}$, 即有:

$$\begin{aligned} & \|\mathbf{H}(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{\tilde{N}}, b_1, b_2, \dots, b_{\tilde{N}})\hat{\beta} - \mathbf{T}\| = \\ & \min_{\beta} \|\mathbf{H}(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{\tilde{N}}, b_1, b_2, \dots, b_{\tilde{N}})\beta - \mathbf{T}\| \end{aligned} \quad (6)$$

如果隐层节点数 \tilde{N} 与网络的输入样本数 N 相等, 即 $\tilde{N} = N$, 当随机选择输入权重 \mathbf{w}_i 和隐层偏置 b_i 时, 矩阵 \mathbf{H} 是可逆方阵, 且该单隐层前馈神经网络能够零误差地逼近训练样本。

然而, 在大多数情况下, 隐层节点数 \tilde{N} 远小于网络的输入样本数 N , 即 $\tilde{N} \ll N$, 此时, 矩阵 \mathbf{H} 不

$\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ 为连接第 i 个隐层节点和输出神经元的权重向量; b_i 是第 i 个隐层节点的偏置; \mathbf{o}_j 为第 j 个输出神经元的实际输出值。

含有 \tilde{N} 个隐层节点, 且激活函数为 $g(x)$ 的标准单隐层前馈神经网络能够无误差地逼近 N 个训练样本, 即 $\sum_{j=1}^N \|\mathbf{o}_j - \mathbf{t}_j\| = 0$, 因此存在 β_i 、 \mathbf{w}_i 和 b_i , 使:

$$\sum_{i=1}^{\tilde{N}} \beta_i g(\mathbf{w}_i \mathbf{x}_j + b_i) = \mathbf{t}_j \quad j=1,2,\dots,N \quad (2)$$

进而对于 N 个样本, 式(2)可统一表述为矩阵形式的等式:

$$\mathbf{H}\beta = \mathbf{T} \quad (3)$$

其中:

再是方阵, 从而不存在 \mathbf{w}_i 、 b_i 、 $\beta_i (i=1,2,\dots,\tilde{N})$, 使 $\mathbf{H}\beta = \mathbf{T}$, 但可以求解上述线性系统的最小模的最小二乘解, 即:

$$\hat{\beta} = \mathbf{H}^\dagger \mathbf{T} \quad (7)$$

式中, \mathbf{H}^\dagger 为矩阵 \mathbf{H} 的穆耳-潘洛斯(Moore-Penrose)一般逆。

综上所述, 给定一个训练样本集 $\mathcal{S} = \{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in R^m, \mathbf{t}_i \in R^m, i=1,2,\dots,N\}$, 激活函数 $g(x)$ 和隐层节点数 \tilde{N} , ELM算法的计算步骤如下:

- 1) 随机设定输入权重 \mathbf{w}_i 和隐层偏置 b_i , $i=1,2,\dots,\tilde{N}$ 。
- 2) 计算隐层输出矩阵 \mathbf{H} 。
- 3) 计算输出权重 β : $\beta = \mathbf{H}^\dagger \mathbf{T}$, 其中, $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_N]^T$ 。

2 实验过程与数据处理

2.1 测谎协议

实验挑选了30个身体健康的本科生作为实验对象, 平均年龄21岁, 随机分为无辜和说谎两组, 其中无辜者15名和说谎者15名。采用标准的三刺激模式的测谎协议^[5]。实验前准备6个物品(人造仿真珠宝)及其相对应的图片。针对说谎者, 在事先准备好的保险箱中放入任意两个物品, 并要求受试者看清两个物品后, 拿走其中的一个物品。被受试者拿走的物品作为探针刺激(Probe, P), 箱子中剩下的物品作为靶刺激(Target, T), 其余的物品作为无关刺激

(Irrelevant, I); 针对无辜者, 保险箱中放入任意一个物品, 要求受试者看清物品即可, 该物品作为T刺激, 任选其他一个物品作为P刺激, 剩下4个全部作为I刺激。脑电采集过程中, 屏幕上随机的显示6张图片, 出现一张图片相当于对受试者提出问题: “你有没有见过这样物品?” 要求受试者对看到的每一张图片都要做出反应(按手柄中的按键)。要求说谎组看见P刺激的时候做出欺骗行为, 即回答没有看过(按No键), 出现其余两类刺激的时候要求说真话(T刺激按Yes键, I刺激按No键), 无辜组全说真话。

2.2 数据预处理

受试者的脑部电极安放在如下的位置: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz, O2, 垂直眼电极放在左眼上下2 cm处, 水平眼电极放在外眼角处, 前额接地, 以右耳电极为参考电极M1。采用Neuroscan公司的Synamps作为放大器, 通带频带为0.3~30 Hz, 采样频率500 Hz, 各导联阻抗均小于3 k Ω ; E-Prime2.0作为刺激软件。实验时每张图片随机出现50次, 持续1.2 s, 一次实验的时间为6 min。每次实验中, I刺激、P刺激和T刺激出现的频率依次为66.7%、16.7%和16.7%。休息5 min后, 再次重复上述实验, 每名受试者共做上述实验3次。最后, 使用Scan4.0自带的工具滤除脑电波形的垂直和水平眼电伪差。

很多学者已经证明, 在Pz电极附近的P300是最明显的^[6-7], 进一步的统计分析也显示了Pz电极上的脑电波形对两类受试者具有更显著的区别意义(由于篇幅有限, 此统计结果未给出), 因此本文将Pz电极上的P300波形作为下一步的分析对象。使用EEGlab工具箱^[8]对连续的EEG波形进行分割, 将刺激前200 ms到刺激后1 000 ms分割为一个Epoch(称为一个P刺激响应)。然后将每名受试者在Pz电极上的每5个P刺激响应进行平均, 得到两类受试者的ERP波形。最后, 两类受试者分别得到450 (15人 \times 30 single-trials=450)个待分析的ERP波形(共计900个 single-trials)。

2.3 特征提取

本文提取了两组特征值, 它们分别基于频域和时-频域。

2.3.1 频域特征

每个ERP波形通过功率谱的计算, 得到功率谱曲线, 在此基础上, 计算3个频域特征, 描述如下:

1) MAF(maximum frequency): 功率谱曲线的极值。

2) MEF(mean frequency): 功率谱均值。

3) PFBP3(the power of frequency band involving the P3): P300成分对应的功率谱总能量。

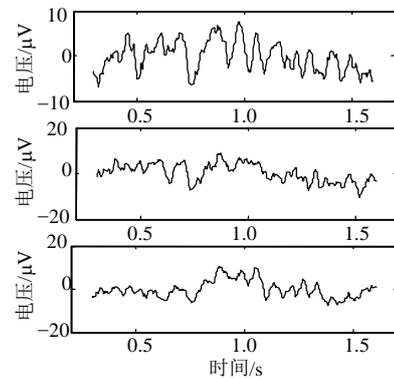
2.3.2 时-频域特征与小波变换

最近, 很多的研究人员使用小波变换对ERP信号进行分析^[9-10]。本文使用了一种快速DWT (discrete wavelet transform)算法^[11]对ERP波形进行小波变换, 得到以下7个频率段的小波系数: 0.3~3.9, 3.9~7.8, 7.8~15.6, 15.6~31.2 (采样率为500 Hz, 带限为0.3~30 Hz)。然后对它们分别进行小波重建, 由于P300成分的主要频率范围在0~4 Hz, 因此将0.3~3.9 Hz波段对应的小波重建系数取出做进一步分析。图1中显示了任取的3个说谎者和无辜者的5个Epoched ERP的平均波形及其对应的小波重建波形, 0.5 s处为刺激开始时间。可以看出相对原始的平均波形, 重建波形对说谎者具有更明显的P300成分; 而无辜者均没有明显的P300成分。其他受试者均有相似的结果。在前期报道的结果基础上^[5, 9], 为了压缩特征向量的维数, 在重建波形的基础上提取以下特征:

1) ARW(the positive area of reconstructed waveforms): 重建波形的正值区域的面积。

2) MRW(the maximum amplitude of reconstructed waveforms): 重建波形的最大波峰值。

3) LRW(the latency of reconstructed waveforms): 重建波形的最大波峰值的发生时间。



a. 3名说谎者的ERP

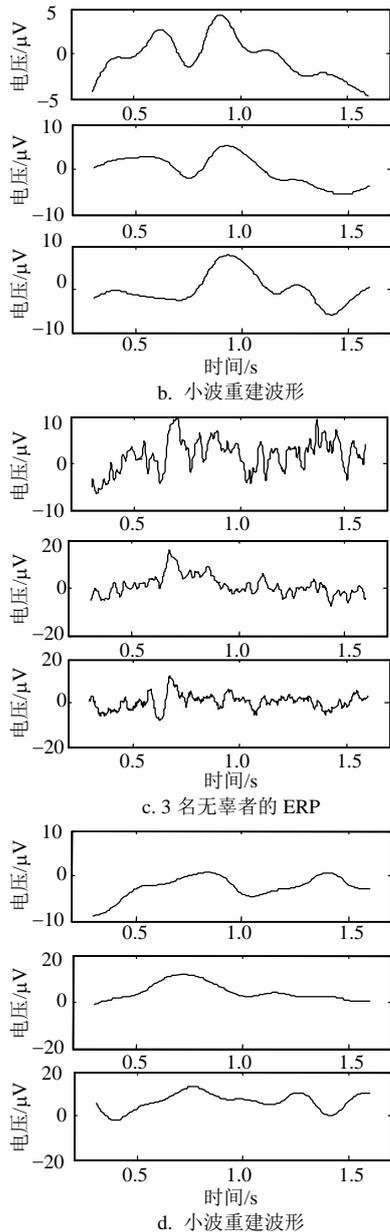


图1 3名说谎者和无辜者的ERP及小波重建波形

4) PTPRW: 重建波形的最大正值与最小负值之差。

5) LMR (LRW/MRW): 潜伏期与最大正波峰值的比。

通过以上特征提取, 每类受试者的ERP波形转换为8维的特征向量, 类别标识分别为1和-1。特征样本共计900个。

2.4 数据分类

本文的实验中, 将ELM与FDA、BPNN及SVM的分类准确率进行比较。为了增加对比性, ELM、BPNN及SVM分类器均使用的sigmoid激活函数($g(x) = 1/(1 + e^{-x})$)。在BPNN分类器中, 隐层节点数变化范围为8~40, 学习率变化范围为0.01~0.05, 控制精度变化范围为0.001~0.005。在SVM分类器中,

取惩罚因子 C 和RBF核函数($K(x, y) = e^{-1/2 * \|x-y\|^2 / \sigma^2}$, Burges, 1998)的参数 σ 通过下列网格 $C=[2^{-5}, 2^{-4}, \dots, 2^8]$ 和 $\sigma=[2^{-5}, 2^{-4}, \dots, 2^8]$ (步长为2)进行训练。

本文的方法训练和测试步骤如下:

1) 将训练样本送入到ELM分类器中, 并令ELM的隐层节点数 K 的初始值为8。训练过程使用SWCV(subject-wise cross validation)^[7]的交叉验证方法, 该方法类似于普通的留一法的交叉验证方式, 即留下一个无辜者和一个说谎者的样本作为测试样本, 然后将28个受试者的特征样本进行混合, 得到两类训练样本(P300和非P300, 分别为 $14 \times 30 = 420$ 个)。计算在特定的参数 K 值时, 10个不同的训练集各自的敏感度值(说谎者的P300被正确划分的准确率)和特异度的(无辜者的非P300被正确划分的准确率), 然后分别计算10个准确率的均值及对应的标准差(standard variance, SD), 分别记为 $TR_{sen} \pm SD$ 和 $TR_{spe} \pm SD$, 然后计算平衡准确率: $BA = (TR_{sen} + TR_{spe})/2$ 。

2) 令 $K=K+1$, 循环执行步骤1)和步骤2), 直至 $K=41$, 退出循环。

3) 比较不同 K 下的平衡准确率BA值, BA值最高时所对应的参数值即为最佳的参数值, 并将此时对应的 $TR_{sen} \pm SD$ 和 $TR_{spe} \pm SD$ 作为网络的最佳训练准确率。

4) 训练完成后, 将训练后得到的最佳参数 k 作为隐层节点数, 对应的网络输入及输出权值矩阵及隐层节点的偏置均代入到ELM分类器中, 将不同测试样本送入训练后的分类器中进行测试, 得到测试准确率, 按照敏感度和特异度记为: $TE_{sen} \pm SD$ 和 $TE_{spe} \pm SD$ 。

对于BPNN和SVM, 同样使用上述交叉验证的方法对分类器参数进行训练, 也得到了各个分类器的最优参数及上述最佳准确率。

3 实验结果

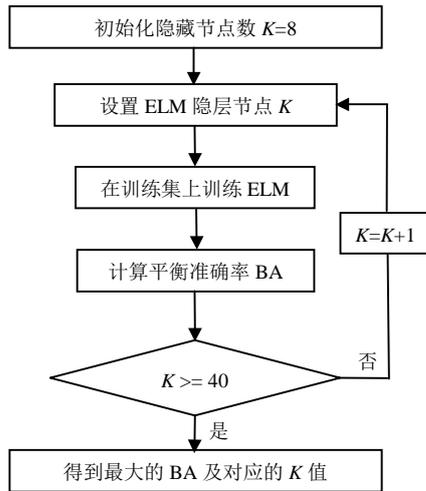


图2 提出方法的训练过程

表1 4种分类器的分类表现(准确率和时间)

分类器	时 间		准确率/(%)			
	训练 /h	测试 /s	训 练		测 试	
			TR _{sen} ±SD	TR _{spe} ±SD	TE _{sen} ±SD	TE _{spe} ±SD
BPNN	1.36	0.54	96.54±2.23	96.17±3.61	90.09±3.88	89.52±2.77
SVM	21.62	21.57	98.13±1.25	97.46±1.35	96.38±2.11	95.92±1.88
ELM	0.27	0.31	98.83±1.55	98.52±1.70	97.87±1.46	98.36±1.59
FDA	0.01	0.30	70.43±5.25	67.89±4.76	68.99±4.12	71.43±5.17

本文所有的实验结果均在MATLAB2010b环境下运行得到。实验PC机为P4 3.2 GHz CPU, 2 GB RAM。使用图2中的训练方法, 得到的ELM分类器的隐层节点数 $K=15$, BPNN网络的隐层节点数 $K=24$, SVM网络的支持向量为57。可见ELM分类器最为紧致, 即网络最为简洁。以训练时间(单位: h)、测试时间(单位: s)、训练和测试准确率作为分类器的评价指标, 各个分类器的分类表现如表1所示。从表中可以看出, ELM训练准确率均高于BPNN和SVM。FDA的分类效果最差, 表明与说谎有关的大脑认知状态线性不可分, 而BPNN的训练准确率比测试准确率高较多, 说明BPNN的泛化性不强, 也不适于测谎领域。SVM和ELM的训练和测试准确率都高于以往报道值^[12], 说明用SVM和ELM作为测谎数据的分类器是合理的。但相比SVM, ELM的准确率也比SVM的高, 更为重要的是, ELM无论在训练时间还是测试时间上都有极大的优势, 诸如SVM的训练时间约为21 h, 而ELM的训练时间仅仅约为18 min, 以上结果说明本文的方法尤其是ELM分类器, 在基于P300的测谎数据的分类上具有很强的有效性和鲁棒性。

4 讨 论

不同于以往其他文献提出的用于基于ERP的测谎研究的分类器, 本文采用了由文献[4]提出的极限学习机作为分类器。相比BPNN、SVM等传统的模式识别方法, ELM更适合应用于基于ERP的测谎分类。相对传统的ANN及SVM分类器, ELM分类器的训练时间大幅缩减, 不仅如此, ELM分类器的分类准确率也比BPNN及SVM高。而在实际的测谎中, 更快、更准确地识别方法是至关重要的。因此在训练时间和测试准确率上都表现出色的ELM非常适合应用于在线的脑电测谎, 对于需要频繁更新相关实验内容的测谎应用, ELM更显现出了其巨大的优势。虽然本文主要将ELM应用在基于ERP的脑电测谎中, 但是提出的方法可以推广应用到其他基于P300的模式识别相关领域。如果结合使用更好的ERP重建算法及特征提取或者特征选择算法, 测谎准确率及稳定性将进一步提高, 这是未来的研究方向之一。

参 考 文 献

- [1] POLICH J. P300 in clinical applications: Meaning, method, and measurement[J]. American Journal of Eeg Technology, 1991, 31(3): 201-231.
- [2] GAMER M, BERTI S. Task relevance and recognition of concealed information have different influences on electrodermal activity and event-related brain potentials[J]. Psychophysiology, 2010, 47(2): 355-364.
- [3] GAO J F, LU L, YANG Y, et al. A novel concealed information test method based on independent component analysis and support vector machine[J]. Clinical Eeg & Neuroscience, 2012, 43(1): 54-63.
- [4] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2006, 70(1-3): 489-501.
- [5] GAO J F, YAN X G, SUN J C, et al. Denoised P300 and machine learning-based concealed information test method[J]. Computer Methods and Programs in Biomedicine, 2011, 104(3): 410-417.
- [6] POLICH J. Updating P300: an integrative theory of P3a and P3b[J]. Clinical Neurophysiology, 2007, 118(10): 2128-2148.
- [7] XU N, GAO X R, HONG B, et al. BCI competition 2003 data set Iib: Enhancing P300 wave detection using ICA-based subspace projections for BCI applications[J]. IEEE Transaction on Biomedical Engineering, 2004, 51(6): 1067-1072.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] ADEMOGLU A, MICHELI-TZANAKOU E, ISTEFANOPULOS Y. Analysis of pattern reversal visual evoked potentials (PRVEP's) by spline wavelets[J]. IEEE Transactions on Biomedical Engineering, 1997, 44(9): 881-890.
- [12] ABOOTALEBIAL V, MORADI M H, KHALIZADEH 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.

编辑 黄 莘