

基于 AdaBoost 算法的癫痫脑电信号识别

张涛 陈万忠 李明阳

Recognition of epilepsy electroencephalography based on AdaBoost algorithm

Zhang Tao Chen Wan-Zhong Li Ming-Yang

引用信息 Citation: *Acta Physica Sinica*, 64, 128701 (2015) DOI: 10.7498/aps.64.128701

在线阅读 View online: <http://dx.doi.org/10.7498/aps.64.128701>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2015/V64/I12>

您可能感兴趣的其他文章

Articles you may be interested in

基于自适应模板法的脑电信号转移熵分析

Transfer entropy analysis of electroencephalogram based on adaptive template method

物理学报.2015 64(8): 088701 <http://dx.doi.org/10.7498/aps.64.088701>

改进的相对转移熵的癫痫脑电分析

Analysis on relative transfer of entropy based on improved epileptic EEG

物理学报.2014 63(21): 218701 <http://dx.doi.org/10.7498/aps.63.218701>

基于 Kendall 改进的同步算法癫痫脑网络分析

An improved synchronous algorithm based on Kendall for analyzing epileptic brain network

物理学报.2014 63(20): 208705 <http://dx.doi.org/10.7498/aps.63.208705>

颠倒睡眠状态调制心率变异性信号的功率谱和基本尺度熵分析

Analysis on power spectrum and base-scale entropy for heart rate variability signals modulated by reversed sleep state

物理学报.2014 63(19): 198703 <http://dx.doi.org/10.7498/aps.63.198703>

基于改进功率谱熵的抑郁症脑电信号活跃性研究

Activity analysis of depression electroencephalogram based on modified power spectral entropy

物理学报.2014 63(17): 178701 <http://dx.doi.org/10.7498/aps.63.178701>

基于 AdaBoost 算法的癫痫脑电信号识别*

张涛 陈万忠[†] 李明阳

(吉林大学通信工程学院, 长春 130012)

(2014年11月24日收到; 2015年1月10日收到修改稿)

AdaBoost 算法作为 Boosting 算法的经典算法之一, 在人脸检测和目标跟踪等领域得到了广泛应用, 但该算法也有一个缺点——退化问题. 为了解决这个问题, 通过对弱分类器进行筛选、引入平滑因子和权值修正函数三个措施对算法进行优化, 并将优化后的算法与小波包分解相结合应用到癫痫脑电信号的识别上. 结果表明, 本文算法对癫痫脑电信号的识别率为 96.11%, 对正常脑电信号的识别率为 99.51%, 具有较高的识别率, 为癫痫的正确诊断提供了一种可能有效的解决方案.

关键词: AdaBoost 算法, 退化问题, 小波包分解, 癫痫脑电信号

PACS: 87.85.-d, 87.85.Ng, 87.15.A-

DOI: 10.7498/aps.64.128701

1 引言

脑电信号中包含了大量的生理与病理信息, 在临床医学与脑科学的研究中具有非常重要的作用, 是进行神经系统疾病和症状特别是癫痫病诊断的主要依据^[1]. 癫痫是由于脑部神经细胞群异常放电引起的过性脑功能紊乱综合征, 是一种严重且较顽固的慢性疑难性脑部疾病, 其发作时表现为大脑神经元群兴奋性增高以及过度同步化放电, 并导致短暂性中枢神经系统功能失常, 是神经系统疾病中仅次于脑血管疾病的第二大顽症, 人群患病率为 0.5%—1%, 我国约有 600 万癫痫患者, 每年新发病例近 65 万—70 万^[2], 癫痫脑电信号的自动检测有助于减轻医疗工作者的工作量, 进一步推进癫痫脑电信号自动检测的研究, 在临床应用中具有现实且重大的意义^[3].

庞春颖等^[4]用改进的经验模态分解对癫痫脑电信号进行识别, 识别率为 94%. 张振等^[5]采用近似熵与支持向量机相结合的方法对癫痫发作间歇期脑电信号和发作期脑电信号进行自动识别分类, 平均分类准确率达到 93.33%. Tibdewal 等^[6]

使用多参数量化和信号功率谱相结合的方法对癫痫脑电信号进行自动诊断, 识别率为 91%. Sakkalis 等^[7]分别应用近似熵、非线性替代和线性方差分析三种方法对癫痫脑电信号进行分类, 分类准确率分别为 90.12%, 84.18% 和 82.28%.

由上述研究进展可知, 目前对癫痫脑电信号的识别精度并不高. 鉴于此, 本文试图将 AdaBoost 算法引入到癫痫脑电信号识别中, 在对 AdaBoost 算法进行优化的基础上, 提出先对脑电信号进行低通滤波, 再对滤波后的信号进行四层小波包分解, 提取相应的特征向量, 将特征向量输入到优化的 AdaBoost 集成神经网络, 实现了对癫痫脑电信号的有效识别.

2 AdaBoost 算法及其优化

2.1 AdaBoost 算法

AdaBoost 算法^[8]是一种迭代算法, 其核心思想是针对同一训练集训练若干个不同的弱分类器, 然后将这些弱分类器(弱学习算法指的是算法获得的分类器的预测能力只比随机预测的准

* 高等学校博士学科点专项科研基金(批准号: 20100061110029)、吉林省科技发展计划重点项目(批准号: 20090350)和吉林大学研究生创新研究计划(批准号: 20121107)资助的课题.

[†] 通信作者. E-mail: chenwz@jlu.edu.cn

准确率50%高一点)的分类结果通过某种算法合并成一种强分类算法. 该算法不要求事先知道弱学习算法预测精度的下限, 因而非常适用于实际问题, 由该算法构成的神经网络是集成神经网络的重要一员. 以Boosting算法^[9]为指导思想, AdaBoost, AdaBoost.M1和AdaBoost.M2以及AdaBoost.R^[10,11]四种算法在实际研究中得到了广泛的应用.

2.2 AdaBoost算法的优化

对AdaBoost算法生成的子分类器多样性进行分析发现: 在前几轮迭代中, 由于每一个新的子分类器都关注于上一个子分类器难以正确分类的样本, 因此有着较高的多样性. 然而几轮迭代之后, 难以被正确分类的样本很可能始终被错分, 导致后生成的子分类器越来越关注同一批不能被正确分类的样本, 多样性急剧下降, 这就是所谓的“退化问题”^[12], 使用平均或加权平均进行集成都有可能导导致过配(overfitting). 另外, AdaBoost算法对分类效果差的弱分类器只是通过降低相应的权值来减弱其在整个强分类器中的作用, 而没有选择舍弃该弱分类器. 在训练阶段, 若某个弱分类器的分类准确率为1时, 其他弱分类器就不再起作用, 并且AdaBoost系列算法一般都假设类分布是平衡的, 但不平衡问题是普遍存在的^[13], 癫痫和正常个体之间的关系就属于不平衡问题的一员. 上述3个因素都会降低强分类器的分类准确率, 造成分类准确率的强烈振荡.

考虑一种极端情况: 若有 m 个训练样本, 在第一次迭代时每个样本的权值均为 $1/m$, 若其中一个样本在之后迭代中都被错分而其他样本均正确分类, 则该样本在进行有限次迭代之后的权值会无限接近0.5. 当训练样本中有两个样本出现上述情况时, 后生成的子分类器就会把几乎所有的注意力都放在这两个样本上而忽略了其他样本, 最终的结果就是生成的强分类器的分类准确率急剧下降, 甚至出现准确率为0的情况.

为了能够克服上述缺点, 使AdaBoost算法适用于癫痫脑电信号的识别, 在原算法的基础上, 本文提出一种AdaBoost优化算法, 算法的具体步骤如下.

1) 从所给定的样本数据中随机选择 m 个样本作为训练数据, 每个训练数据的权值 $W_t(i)$ 为 $1/m$, 选择弱分类器种类及弱分类器的个数 T , 将训练数据输入到第一个弱分类器中.

2) 在训练第 t ($t = 1, 2, \dots, T$)个弱分类器时, 设训练数据的预测输出分类结果为 $g(t)$, y 为训练数据实际类别, 训练数据的分类误差和为 e_t , 待分类对象中训练个数较少的错误个数为 $errnum_t$. 若 $errnum_t > n$ ($n \in N^*$), 则返回第二步重新训练弱分类器; 其他情况下误差和 e_t 的计算公式为

$$e_t = \sum_t W_t(i), \quad i = 1, 2, \dots, m(g(t) \neq y). \quad (1)$$

3) 设 $a(t)$ 为第 t 个弱分类器的权重, 令 $e'_t = e_t + \varepsilon$ 则权重的计算公式为

$$a(t) = \frac{1}{2} \ln \left(\frac{1 - e'_t}{e'_t} \right), \quad (2)$$

其中, ε 为一个非常小的常数, 其作用是防止分母为0, 避免 T 个弱分类器中只有第 k 个弱分类器起作用的情况.

4) 依据第三步得到的 $a(t)$ 对下一次迭代时训练样本的权值进行修正, 其修正公式为

$$\begin{cases} W_{t+1}(i) = W_t(i) \exp(-a(t)y_i g_t(x_i)), \\ W_{t+1}(i) \leq \beta, \\ W_{t+1}(i) = \log_\alpha [1 + W_t(i) \exp(-a(t)y_i g_t(x_i))], \\ W_{t+1}(i) > \beta, \end{cases} \quad (3)$$

其中, $i = 1, 2, 3, \dots, m$; $W_t(i)$ 和 $W_{t+1}(i)$ 分别是第 i 个训练样本在第 t 个和第 $t + 1$ 个弱分类器在进行训练时的权值; α 是对数函数的底数; β 是阈值; α 和 β 的值与训练样本个数 m 有关.

5) 对权值 $W_{t+1}(i)$ 进行归一化, 使全部权值和为1. 当 $t \leq T$ 时, 令 $t = t + 1$ 并返回第二步进行下一个弱分类器训练, 否则执行第六步.

6) 强分类器函数. 将训练数据进行 T 次训练后可以得到 T 个相应的弱分类函数 $f(g_t, a_t)$, 把这些函数进行相应处理得到强分类函数 $F(x)$, 强分类函数计算公式为

$$F(x) = \text{sign} \left(\sum_{t=1}^T a(t) f(g_t, a_t) \right). \quad (4)$$

为了验证优化算法的有效性, 图1给出了分别运用原始AdaBoost算法和优化的AdaBoost算法构造相应BP_AdaBoost集成神经网络对癫痫脑电信号进行识别的情况. 由图1可以看出, 原始AdaBoost算法在第11、第15和第17次均出现了“退化问题”, 而优化的AdaBoost算法则没有出现这种情况, 由此说明优化的AdaBoost算法克服了原算法的缺点, 适用于识别癫痫脑电信号.

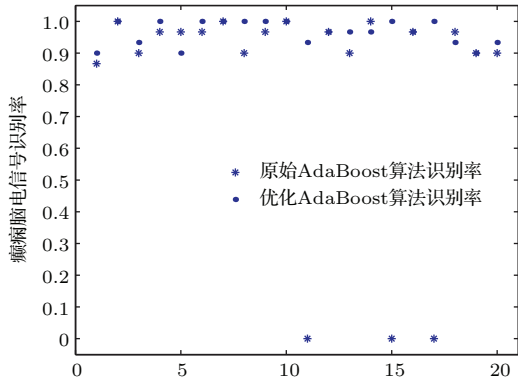


图1 癫痫脑电信号的识别率

Fig. 1. The recognition rate of epilepsy electroencephalography.

3 实验与结果

脑电信号处理可以分为3个阶段: 预处理、特征提取、识别. 在预处理阶段的主要工作是对所采集的脑电信号进行高频滤波、工频陷波等处理, 消除噪声干扰; 在特征提取阶段的主要工作是对进行预处理后的信号利用时域、频域和时频域等方法提取信号在相关域下的特征; 识别阶段的主要工作是将所选择的信号特征输入相应的分类器, 得到识别结果.

3.1 实验数据及环境

在本文中需要用到五组脑电信号数据集, 每组数据集里有100个脑电信号, 它们是由Andrzejak教授^[14]从德国波恩大学的一所癫痫病中心所采集的, 采样频率为173.61 Hz. 五组数据分别以SET A, SET B, SET C, SET D和SET E表示, 其中A, B, C, D是正常状态, E是癫痫发作期. 本文所使用的仿真软件为MATLAB2013a, 运行环境为Windows 8.1企业版, 处理器为Core(TM) i3 CPU, 工作频率为2.53 Hz, 运行内存为4.87 GB.

3.2 预处理

由于脑电信号主要成分在0.5—30.5 Hz之间^[15], 而本文所使用数据的采集频率是173.61 Hz, 根据Nyquist采样定律可知信号的最大频率为86.805 Hz, 在这么大的频率范围内并不都含有脑电信号频率分量, 所以在预处理阶段需利用数字低通滤波器对采集的脑电信号进行去噪处理. 将脑电信号通过一个通带截止频率为37 Hz, 阻带截止

频率为40 Hz的巴特沃斯数字低通滤波器, 滤波器特性如图2所示.

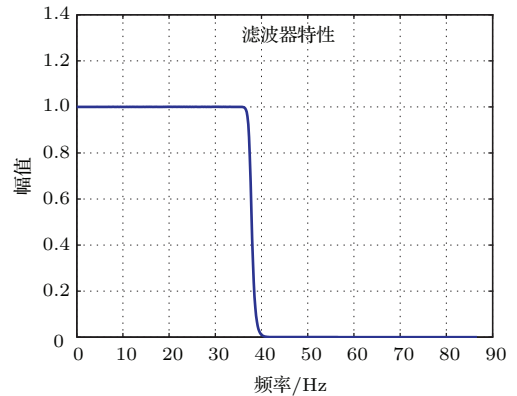


图2 巴特沃斯数字低通滤波器特性

Fig. 2. The characteristics of Butterworth digital low-pass filter.

随机选择脑电数据F051, 其滤波前和滤波后的对比如图3所示, 其中图3(a)为滤波前的波形, 图3(b)为滤波后的波形. 由图3可以看出, 经过低通滤波之后, 脑电信号中的噪声信号得到了有效滤除.

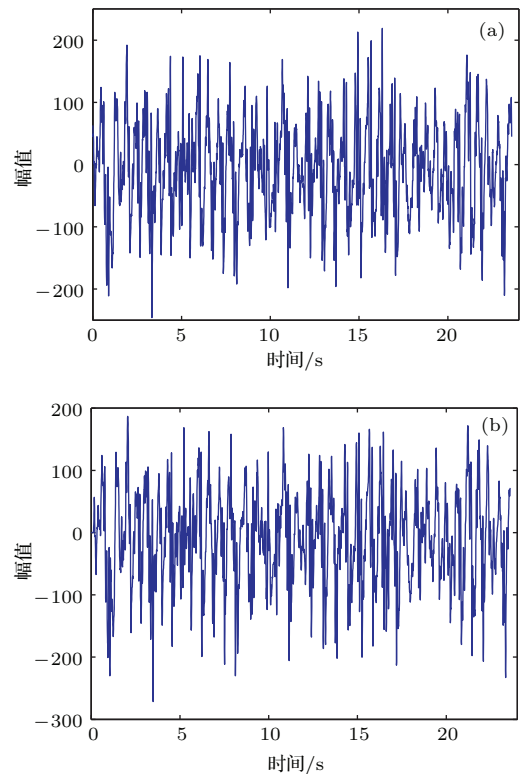


图3 滤波前和滤波后的波形对比 (a) 滤波前波形; (b) 滤波后波形

Fig. 3. Comparison of (a) pre-filtering waveform and (b) post-filtering waveform.

3.3 特征提取

小波分析^[16]是一种时频域窗口面积恒定但窗口形状可变(即时间和频率窗都可改变)的时频局部化分析方法. 在进行小波分解的过程中, 我们只对低频信号进行再分解, 对高频信号不再实施分解, 这种分解方法使得频率分辨率随频率的增大而减小, 当对高频成分进行研究时, 其低频率分辨率会影响分析的准确性. 为了解决这个问题, 小波包分解^[17]应运而生. 它不仅对低频部分进行分解, 对高频部分同样也实施相应的分解, 而且小波包分解能根据信号特性和分析要求自适应地选择相应频带与信号频谱相匹配, 是一种比小波分解更为精

细的分解方法.

鉴于小波包分解可以将信号从单一的时域转换成具有时频域特性的小波包系数, 而小波包系数在某种程度上体现了信号的特征, 本文提出将脑电信号进行四层小波包分解得到系数矩阵, 对每个不同频段的系数矩阵求绝对值和的千分之一及小波包熵并求出信号幅度平方和的百分之一作为信号的特征量. 从五组脑电信号中分别随机选取一个脑电信号, 其相应的特征向量如表1所列, 其中, Q1—Q16是小波包分解系数绝对值和的千分之一, Q17是幅度平方和的百分之一, Q18是小波包熵, 所有数据保留四位小数, 特征提取的时间是5.044 s.

表1 采用小波包提取癫痫脑电信号特征向量

Table 1. The feature vector of epilepsy electroencephalography extracted by wavelet packet.

集	编号	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
A	Z069	31.4519	15.4106	8.2374	17.0206	1.1776	4.0495	6.4188	6.1892	0.3058
B	O027	53.1410	36.2644	9.8811	33.6503	1.2929	8.6045	7.1675	8.0567	0.3482
C	N052	29.4241	11.7759	3.5610	5.5027	0.7583	1.5341	1.9668	1.7119	0.1948
D	F087	30.7315	7.3649	2.3488	3.2715	0.5339	1.0226	1.4666	1.0472	0.1382
E	S064	85.3511	52.0179	20.3212	36.8394	3.5676	9.3189	12.3908	13.4400	0.8615
集	编号	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18
A	Z069	0.9793	1.5332	1.6090	0.3425	0.9986	1.4791	1.2419	1.0272	1.9074
B	O027	2.1472	1.9329	2.2189	0.2409	0.8289	1.5115	1.1742	3.0199	1.9205
C	N052	0.3661	0.5104	0.4276	0.1277	0.3400	0.4854	0.4341	0.7226	0.9534
D	F087	0.2395	0.3473	0.2363	0.0776	0.2192	0.3806	0.2819	0.5468	0.5500
E	S064	2.2435	3.1209	3.4006	0.6993	2.0552	3.3308	2.5708	7.7833	1.8422

3.4 识别

在用平均数方差法对数据进行归一化处理时, 线性判别式分析中最大特征值所对应的特征向量的分布会发生改变, 并且由训练数据映射成的一维分类平面中的两类值之间的距离也会比较接近, 这样会减弱分类阈值的判别作用, 影响分类的准确性, 所以本文采用最大最小法对分类数据进行归一化处理, 该归一化函数如下:

$$x_k = (x_k - x_{\min}) / (x_{\max} - x_{\min}), \quad (5)$$

其中, x_k 为样本 x 的第 k 个元素, x_{\max} 和 x_{\min} 分别为样本 x 的最大值和最小值.

BP (back propagation) 神经网络^[18]是一种多层前馈神经网络, 属于一种基本的神经网络. 该网

络的特点是信号前向传递, 误差反向传播. 在前向传递中, 输入信号从输入层经隐含层逐层处理, 直至输出层. 每一层的神经元状态只影响下一层神经元状态. 如果输出层得不到期望输出, 则转入反向传播, 根据预测误差调整网络权值和阈值, 从而使BP神经网络预测输出不断逼近期望输出^[19], 并且双隐层比单隐层具有更好的非线性逼近效果. 基于此, 本文选择双隐层BP神经网络作为AdaBoost优化算法的弱分类器.

在进行了大量实验的基础上, 本文最终将优化的AdaBoost算法中的弱分类器个数定为6, 每个弱分类器的隐含层为两层, 隐含层节点个数分别为8和9, 弱分类器训练次数为20次, 学习率为0.08, ϵ 定为 10^{-15} , α 定为3.28, β 定为0.14.

从400个正常脑电的特征向量和100个癫痫脑

电的特征向量中分别随机选择 250 个正常脑电的特征向量和 70 个癫痫脑电的特征向量作为强分类器的训练数据, 剩下 180 个特征向量作为测试数据. 若强分类器的输出为“1”, 则表示所分类的脑电对应的是正常状态; 相反, 若强分类器的输出为“-1”, 则表示所分类的脑电对应的是癫痫发作期. 最终的结果如图 4 所示, 其中前 150 个是正常脑电信号, 后 30 个是癫痫脑电信号, “*”代表强分类器的预测输出分类, “.”代表实际的脑电信号类别. 为了能够清楚地区分出预测输出和实际脑电类别, 在预测输出的基础上对每个结果增加 0.2 的偏移量, 即原预测输出为 1 的变为 1.2, 原预测输出为 -1 的变为 -0.8. 此次运行结果中癫痫的识别率为 96.67%, 正常状态识别率为 100.00%.

为了验证本文算法的准确性, 在使用相同的 BP 神经网络结构和相同个数的弱分类器的情况

下, 分别将优化算法和原始算法的处理程序运行 1 次、5 次、20 次、100 次和 200 次, 运行结果如表 2 和表 3 所列.

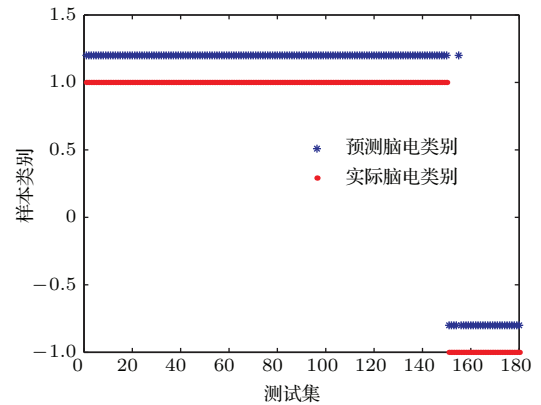


图 4 (网刊彩色) 识别结果

Fig. 4. (color online) The identification results.

表 2 优化算法在不同次数下的运行结果

Table 2. The results of optimized algorithm at different times.

	运行次数				
	1	5	20	100	200
癫痫识别率 /%	96.67	96.67	96.17	96.09	96.10
正常识别率 /%	100.00	99.73	99.50	99.51	99.50
总的运行时间 /s	15.540	111.811	347.572	2012.502	4567.607
平均运行时间 /s	15.540	22.362	17.378	20.125	22.838

表 3 原始算法在不同次数下的运行结果

Table 3. The results of original algorithm at different times.

	运行次数				
	1	5	20	100	200
癫痫识别率 /%	93.33	76.00	76.17	71.50	66.37
正常识别率 /%	100.00	79.33	79.67	74.61	68.68
总的运行时间 /s	8.899	34.594	133.790	681.561	1452.624
平均运行时间 /s	8.899	6.919	6.690	6.816	7.263

由表 2 和表 3 可以看出, 利用小波包分解和原始 AdaBoost 算法相结合的处理方法在运行次数不断增大的情况下, “退化问题”会越来越明显, 并且由于缺少平滑因子, 随着运行次数的增大, 强分类器的识别效果会发生比较剧烈的振荡, 其识别率会越来越低. 而利用小波包分解和优化的 AdaBoost 算法相结合的处理方法则可以很好地对癫痫和正常脑电信号进行分类, 而且对癫痫脑电信号和正常脑电信号的识别率没有发生太大的波动, 有效消除了原始 AdaBoost 算法的“退化问题”. 但由于在训

练弱分类器的过程中, 优化算法需要对弱分类器进行筛选, 会舍弃分类效果不佳的弱分类器并进行再训练, 这是引起优化算法的运行时间比原始算法的运行时间更长的原因.

3.5 算法的健壮性

健壮性又叫鲁棒性, 它是算法容错能力的体现, 容错能力越强, 代表鲁棒性越好, 算法的可靠性就越高. 对于优化的 AdaBoost 算法来说, 弱分

类器的个数 T 、对数函数的底数 α 以及阈值 β 是影响其识别率的主要因素. 为了研究它们各自对优化的 AdaBoost 算法的影响, 本文以程序运行 20 次的平均识别率作为评价标准, 具体结果如图 5 所示. 其中, 图 5(a) 是在 $\alpha = 3.28$, $\beta = 0.14$ 的条件下不同 T 值对应的识别率, 图 5(b) 是在 $T = 6$, $\beta = 0.14$ 的条件下不同 α 值对应的识别率, 图 5(c) 是在 $T = 6$, $\alpha = 3.28$ 的条件下不同 β 值对应的识别率. 由图 5 可以看出, 在不同参数的条件下, 无论是癫痫还是正常脑电信号的识别率都比较高, 没有太大的波动, 这说明优化的 AdaBoost 算法具有较好的鲁棒性.

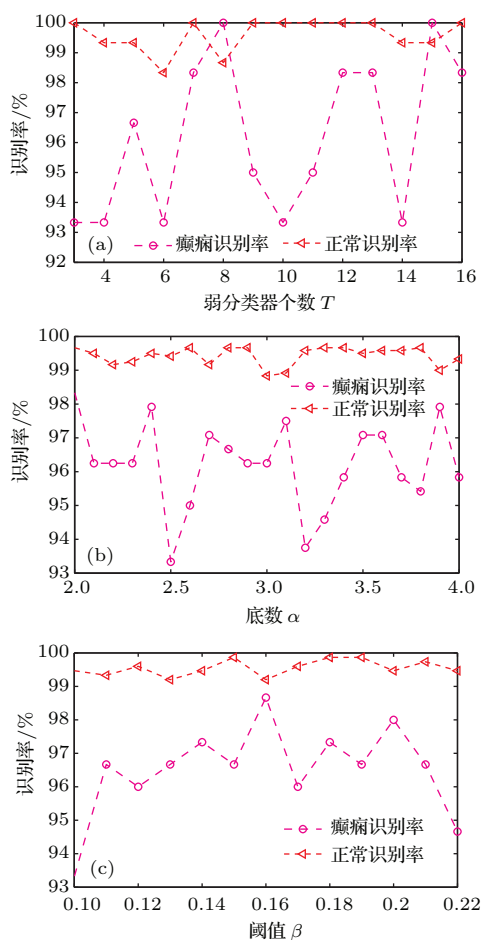


图 5 (网刊彩色) 不同参数条件下的识别率 (a) T ; (b) α ; (c) β
 Fig. 5. (color online) The recognition rates with different parameters: (a) T ; (b) α ; (c) β .

4 结 论

为了解决 AdaBoost 算法的“退化问题”, 本文提出了一种 AdaBoost 优化算法, 并将优化后的算法和小波包分解相结合构成癫痫脑电信号的处理方法. 实验结果表明, 本文的算法可以很好地对癫

痫脑电信号和正常脑电信号进行识别, 对癫痫脑电信号的识别准确率可以达到 96.11%, 对正常脑电信号的识别准确率可以达到 99.51%, 在癫痫脑电识别过程中可以有效地消除原 AdaBoost 算法的“退化问题”, 该算法为癫痫的正确诊断提供了一种可能的解决方案. 同时也应注意到, 由于训练过程中增加了筛选环节, 所以优化后的算法的训练时间也会变长. 下一步的工作重心是尝试将本文算法应用到脑机接口中, 实现对不同思维活动的正确分类.

参考文献

- [1] Meng Q F, Zhou W D, Chen Y H, Peng Y H 2010 *Acta Phys. Sin.* **59** 123 (in Chinese) [孟庆芳, 周卫东, 陈月辉, 彭玉华 2010 物理学报 **59** 123]
- [2] Wang Y, Hou F Z, Dai J F, Liu X F, Li J, Wang J 2014 *Acta Phys. Sin.* **63** 218701 (in Chinese) [王莹, 侯凤贞, 戴加飞, 刘新峰, 李锦, 王俊 2014 物理学报 **63** 218701]
- [3] Cai D M, Zhou W D, Li S F, Wang J W, Jia G J, Liu X W 2011 *Acta Biophys. Sin.* **27** 175 (in Chinese) [蔡冬梅, 周卫东, 李淑芳, 王纪文, 贾桂娟, 刘学伍 2011 生物物理学报 **27** 175]
- [4] Pang C Y, Wang X T, Sun X L 2013 *Chin. J. Biomed. Eng.* **32** 663 (in Chinese) [庞春颖, 王小甜, 孙晓琳 2013 中国生物医学工程学报 **32** 663]
- [5] Zhang Z, Du S H, Chen Z Y, Tian X H, Zhou Y, Zhang Y 2013 *J. Biomed. Eng. Res.* **32** 74 (in Chinese) [张振, 杜守洪, 陈子怡, 田翔华, 周毅, 张洋 2013 生物医学工程研究 **32** 74]
- [6] Tibdewal M N, Manjunatha M, Ray A K, Malokar M 2013 *1st IEEE-EMBS Special Topic Conference on Point-of-Care (POCT) Healthcare Technologies (PHT)* Bangalore, India, January 16–18, 2013 p72
- [7] Sakkalis V, Giannakakis G, Farmaki C, Mousas A, Pedititis M 2013 *35th Annual International Conference of the IEEE EMBS* Osaka, Japan, July 3–7, 2013 p6333
- [8] Fu Z L 2014 *Acta Autom. Sin.* **40** 1075 (in Chinese) [付忠良 2014 自动化学报 **40** 1075]
- [9] Fu Z L 2013 *J. Comput. Res. Dev.* **50** 861 (in Chinese) [付忠良 2013 计算机研究与发展 **50** 861]
- [10] Schapire R E, Singer Y 1999 *11th Annual Conference on Computational Learning Theory* Madison, Wisconsin, July 24–26, 1998 p297
- [11] Islam M M, Yao X, Murase K 2003 *IEEE Trans. Neural Networks* **14** 820
- [12] Cao Y, Miao Q G, Liu J C 2013 *Acta Autom. Sin.* **39** 745 (in Chinese) [曹莹, 苗启广, 刘家辰 2013 自动化学报 **39** 745]
- [13] Fu Z L 2011 *J. Comput. Res. Dev.* **48** 2326 (in Chinese) [付忠良 2011 计算机研究与发展 **48** 2326]
- [14] Andrzejak R G, Lehnertz K, Rieke C, Mormann F, David P, Elger C E 2001 *Phys. Rev. E* **64** 061907
- [15] Ji Z 2003 *Ph. D. Dissertation* (Chongqing: Chongqing University) (in Chinese) [季忠 2003 博士学位论文 (重庆: 重庆大学)]

- [16] Ge Y Z, Xu M Y, Mi J C 2013 *Acta Phys. Sin.* **62** 104701 (in Chinese) [戈阳祯, 徐敏义, 米建春 2013 物理学报 **62** 104701]
- [17] Gao Q, Yi S H, Jiang Z F, Zhao Y X, Xie W K 2012 *Chin. Phys. B* **21** 064701
- [18] Wang D M, Wang L, Zhang G M 2012 *J. Zhejiang Univ. (Eng. Sci.)* **46** 837 (in Chinese) [王德明, 王莉, 张广明 2012 浙江大学学报 (工学版) **46** 837]
- [19] Wang X C, Shi F, Yu L, Li Y 2013 *43 Cases of MATLAB Neural Network Analysis* (Beijing: Beihang University Press) p1 (in Chinese) [王小川, 史峰, 郁磊, 李洋 2013 MATLAB 神经网络 43 个案例分析 (北京: 北京航空航天大学出版社) 第 1 页]

Recognition of epilepsy electroencephalography based on AdaBoost algorithm*

Zhang Tao Chen Wan-Zhong[†] Li Ming-Yang

(Department of Communication Engineerings, Jilin University, Changchun 130012, China)

(Received 24 November 2014; revised manuscript received 10 January 2015)

Abstract

Automatic recognition of epilepsy electroencephalography (EEG) signal has become a research focus because of its high efficiency, and many algorithms have been put forward to achieve it. As one of the classic algorithms of boosting algorithm, AdaBoost algorithm has been widely used in face detection and target tracking fields, but the algorithm also has a disadvantage that is its degradation. In order to solve this problem, this paper puts forward three measures to optimize the algorithm by filtering the weak classifiers whose recognition rates are low, introducing the smoothing factor and a weighted correction function. In order to verify the robustness of optimized algorithm, we choose three main parameters, i.e., the number of weak classifier, which is denoted by T ; the base of logarithmic function, which is denoted by α ; the threshold of weight, which is denoted by β . The experimental results of optimized AdaBoost show that it has good robustness and high recognition rate.

In this paper, we divide the whole process into three steps. The first step is to use the Butterworth digital low-pass filter in which the cutoff frequency of pass band is 40 Hz to filter noise whose frequency is above 40 Hz. The second step is to do feature extraction with the help of wavelet packet decomposition. The third step is to compute the sum of absolute value which are the wavelet packet coefficients of fourth layer, the wavelet package entropy and the sum of signal amplitude square and combine them together to form the feature vector of each EEG. Because the wavelet package entropy is far less than the sum of absolute value and the sum of signal amplitude square, in order to make sure that the entropy reacts in the third step, we use one thousandth of the sum of absolute wavelet packet coefficients, one hundredth of the sum of signal amplitude square and the wavelet package entropy as the weighted feature vector. Finally, we succeed in distinguishing EEGs between epilepsy and normal by using the optimized AdaBoost whose input is the weighted feature vector. The result shows that the presented method has a high recognition rate, it can identify 96.11% epilepsy EEGs and 99.51% normal EEGs, thus it provides an effective solution for the correct diagnosis of epilepsy.

Keywords: AdaBoost algorithm, degradation, wavelet packet decomposition, epilepsy electroencephalography

PACS: 87.85.-d, 87.85.Ng, 87.15.A-

DOI: 10.7498/aps.64.128701

* Project supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20100061110029), the Key Project of Jilin Province Science and Technology Development Plan, China (Grant No. 20090350), and the Graduate Innovation Research Program of Jilin University, China (Grant No. 20121107).

[†] Corresponding author. E-mail: chenwz@jlu.edu.cn