

基于近似熵的心肌猝死预警诊断

汤丽平^a, 刘 剑^b

(重庆医科大学附属第一医院 a. 设备处; b. 心内科, 重庆 400016)

摘 要: 为提高传统线性分析方法的准确性, 提出一种非线性动力学与时频分析相结合的心电信号分类方法。利用经验模式分解和阈值法抑制噪声, 从而更有效地分解心电信号得到内蕴模式函数, 分别计算其近似熵特征值, 利用支持向量机分类器验证特征值的分类效果。实验结果表明, 该方法能有效实现信号的自动分类识别, 简便快速地初步诊断心肌猝死疾病的发生, 正常和异常心电信号的分类识别准确率均达到 90% 以上。

关键词: 心电图; 心肌猝死; 非线性动力学; 经验模式分解; 近似熵; 支持向量机

Warning Diagnosis of Sudden Cardiac Death Based on Approximate Entropy

TANG Li-ping^a, LIU Jian^b

(a. Department of Equipment; b. Department of Cardiology,

The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China)

【Abstract】 In order to improve the accuracy of traditional linear dynamics method, an automated Empirical Mode Decomposition(EMD) signals recognition method is proposed. The method which combines the nonlinear dynamics with the time-frequency analysis is used to study signals. In order to effectively restrain noise and decompose the signals, EMD is employed to preprocess the sample data, and the approximate entropy of some important intrinsic mode functions can be obtained. Support Vector Machine(SVM) is adopted to achieve the optimal classification. Experimental results show the correct classification rate can be more than 90%, which proves the reliability and veracity of the method.

【Key words】 Electrocardiogram(ECG); Sudden Cardiac Death(SCD); nonlinear dynamics; Empirical Mode Decomposition(EMD); approximate entropy; Support Vector Machine(SVM)

DOI: 10.3969/j.issn.1000-3428.2012.09.061

1 概述

心电信号主要反映心脏的电活动过程, 在心脏病理研究方面具有重要的研究意义。研究表明心电信号某些特征参数在进行心肌缺血和猝死的诊断中具有极高的价值^[1]。由于心电图(Electrocardiogram, ECG)的简便无创和诊断可靠, 因此在临床上得到了广泛应用。心电信号仍旧是当前生物医学工程学重点研究的对象之一。

目前, 使用心电信号进行疾病分类的研究主要包括两部分: (1)对信号进行线性^[2-4]分析, 通过时域分析得到均值、方差、波形结构等特征值; 通过频域分析得到频谱、功率谱等特征值或者采用小波分解、小波包等分析工具获得相关时频特征值。(2)采用神经网络进行模式识别获得结果。然而, 文献[5]发现用非线性动力学的方法从心电信号中提取动力学特征, 能更好地为早期临床诊断疾病提供依据。因此, 本文选取心电信号的非线性特征参量作为模式识别特征值。

随着模式识别技术的发展, 支持向量机^[6](Support Vectors Machine, SVM)不仅能从有限的训练集样本中得到较小误差, 而且能避免神经网络中因为规定的合理结构而难以确定局部最优值的问题, 在小样本条件下仍然具有良好的泛化能力, 近年来被成功应用于故障预测^[7]和生物信息识别等领域^[8-9]。因此, 本文选择 SVM 来解决心电信号分类识别中因不具备大量的病理样本而制约正确率提高的问题。

2 实验数据

20 世纪 80 年代开始, 国际上出现的标准心电数据库主

要有 3 个: CSE, AHA 和 MIT-BIH。

心电信号的标准数据库指此数据库中的心电分类一般是经过权威专家认证的或是采用临床证据作为分类标准的心电数据集。由于 MIT-BIH 数据库包括心率失常、房颤、ST 段改变和 TWA 等多个数据库, 因此本文采用 MIT-BIH 心电数据库的数据验证算法的准确性。

从 TWA Challenge Database 中选取 40 例心肌猝死(Sudden Cardiac Death, SCD)病人, 取 37 例非 SCD 病人分别组成 2 组: 学习组和验证组, 其中, 学习组包含 20 例 SCD 病人和 20 例非 SCD 病人; 剩余数据组成验证组, 其中, SCD 已经由临床医师确诊。在非 SCD 疾病组中有 5 例是其他心脏疾病患者, 其余为正常人。本文随机选择学习组和验证组的数据。

3 心电信号分析

3.1 心电信号的预处理

ECG 信号的频率范围为 0.05 Hz~100 Hz, 但 90% 的 ECG 频能量集中在 0.25 Hz~45 Hz。因此, 信号检测分析的首要问题是进行预处理。主要目的是增强检测目标波的能量、抑制噪声和消除基线漂移。

本文采用经验模式分解(Empirical Mode Decomposition,

基金项目: 重庆市卫生局基金资助项目[渝医(2006)27 号文]

作者简介: 汤丽平(1986—), 女, 硕士, 主研方向: 模式识别; 刘 剑, 副教授

收稿日期: 2011-08-29 **E-mail:** tang0733@126.com

EMD)结合阈值的方法对信号进行预处理,方法不但简单易行,同时能获得较好的去噪效果。EMD是一种经验筛选法,本质是将任意信号中真实存在的不同尺度波动或趋势逐级分解,产生一系列具有不同特征尺度的内蕴模式函数(Intrinsic Mode Function, IMF)和一个残余分量。每个IMF分量必须满足^[10]: (1)在整个数据段内,极值点的个数和零交叉点的个数必须相等或相差最多不超过1个; (2)在任意时刻,由局部极大值点形成的包络线和由局部极小值点形成的包络线的平均值为0。

在实际应用中,如果严格按照上述约束条件进行,EMD处理心电信号的耗时很长。因此,根据心电信号特点,本文采用标准差(SD)作为终止EMD“筛分”的标准,认为当 $SD \leq 0.2$ 时,形成了第1个IMF分量和第1个残余分量^[11]。

由图1可知,随着阶数的增加,每个IMF的频率逐渐衰减。心电信号的高频成分和肌电、工频等高频噪声主要分布在低阶的IMF1上,而低频成分和基线漂移等低频噪声则分布在高阶的IMFs上。根据心电信号的频率分布特点,本文摒弃IMFs的最高阶和最低阶,完成初步噪声的滤除,然后将选择的IMF2、IMF3通过一个截止频率为150 Hz的4阶切比雪夫低通数字滤波器,减少剩余的毛刺等高频信号对波形的干扰,利于特征参数的提取。

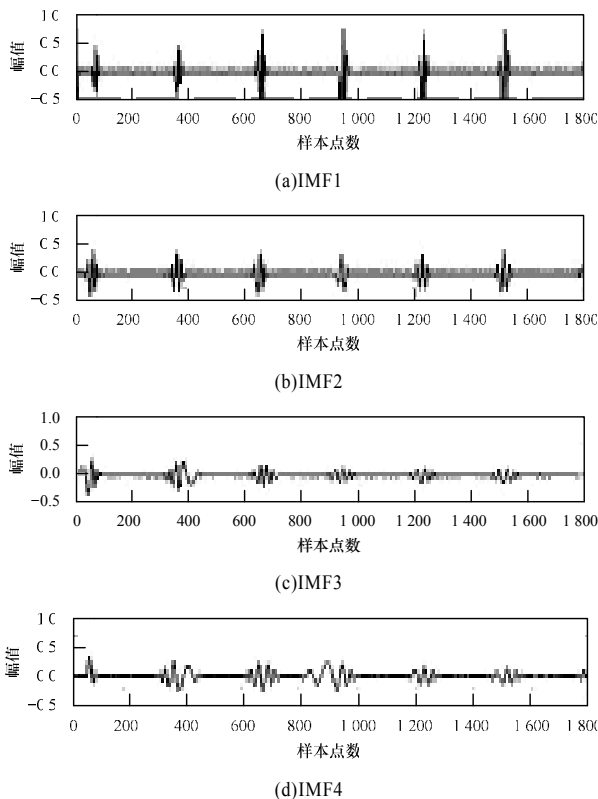


图1 EMD分解心电信号得到的IMF分量

3.2 心电信号特征值的提取

心脏是最复杂的非线性动力系统之一,即使在静息的状态,人的心率都会出现较大的变化。采用非线性系统的数值分析方法分析心电信号,理论上能更好地描绘出心脏的各种动力特征。

近似熵是文献[12]提出的一种度量时间序列复杂性和统计量的规则。由于该算法简单,具有数据点少、抗噪能力强等优点,因此已经成功地用于量化泄露信号的“不可重复”性,用于分辨不同生理、病理状态下心率动力学在时间序列

上数据集的微小差异^[13]。

本文定义一个离散信号时间序列 $\{x(1), x(2), \dots, x(N-1), x(N)\}$,按照以下步骤计算近似熵 $ApEn$:

(1)从序列中抽取连续的 m 个数据构造一组 m 维矢量 $X(i)$:

$$X(i) = \{x(i), x(i+1), \dots, x(i+m-2), x(i+m-1)\} \quad (1)$$

(2) m 维空间中 $X(i)$ 与 $X(j)$ 间的距离为:

$$d[X(i), X(j)] = \max \{ \text{abs}(x(i+k) - x(j+k)) \} \quad (2)$$

$$k = 1, 2, \dots, m-1$$

(3)给出阈值 r ,对于每个 $i \leq N-m+1$,可得:

$$C_i^m(r) = \frac{\sum_{j=1}^{N-m+1} \{d[X(i), X(j)] \leq r\}}{N-m} \quad (3)$$

$$\Phi^m(r) = \frac{\sum_{i=1}^{N-m+1} \ln C_i^m(r)}{N-m} \quad (4)$$

(4)将维数 m 加1,重复以上步骤,最后得到:

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r) \quad (5)$$

每个实验样本经过信号预处理形成了不同频段部分,即多个节点,每个节点是一个数据序列,其复杂程度都不同。本文使用近似熵特征值的方法表示复杂度,首先必须确定维数 m 和阈值 r 的参数值。文献[14]的研究表明, m 的取值通常为1和2。为了保证运算能在序列的联合概率下进行动态重构,选择 $m=2$ 。 r 是衡量时间序列相似性大小的阈值。如果 r 的取值太小,估计出的统计概率不理想;如果太大,则会丢失很多信息。文献[6,12]的研究表明, $r=0.1SD \sim 0.2SD$ 时得到的近似熵具有较合理的统计特征,本文选用 $r=0.2SD$ 。最后计算不同节点的近似熵的平均值,保证只用1个特征值就能表征每例样本,减少了运算数据量,满足了系统实时性的要求。

3.3 心电信号识别算法

SVM的核心内容是Vapnik提出的结构风险最小化原则,具有比非线性函数逼近方法更强的泛化能力,达到最小经营风险和置信范围的目的^[15]。

算法最终转化为一个关于二次型寻优的问题,首先采用内积函数定义的非线性变换将输入空间变换为一个高维空间,再在该空间中求解最优超平面。

通过应用拉格朗日乘子和KKT条件(Karush-Kuhn-Tucker最优化条件),上述二次优化问题的最优分类函数为:

$$f(x) = \text{sgn} \{w^* \cdot x + b^*\} = \text{sgn} \left\{ \sum_{k=1}^d \alpha_k^* y_k (x_k \cdot x) + b^* \right\} \quad (6)$$

其中, $x_k \in R^n$; $y \in \{+1, -1\}$; a 和 b 是用于确定最优超平面的参数。

根据泛函的有关理论,选择满足Mercert条件的核函数构造输入空间中不同非线性决策面的学习机,从而构成不同的支持向量算法,成功地实现了非线性变换后的线性分类。

主要的核函数包括多项式核函数、径向基核函数和Sigmoid核函数3类。其中,径向基函数^[16]定义的是空间2点之间欧氏距离的单调函数。这使它更加符合特征提取的要求,即不同类别的样本在特征空间的距离应尽可能大,同一类别样本之间的距离应尽可能小。因此,本文选用径向基核函数进行实验。

径向基核函数有 $\Gamma(g)$ 和惩罚因子 $C(c)$ 2个参数,其大小决定了模型的复杂性和训练误差。如图2所示,本文使用交叉验证法,经过多次实验,最终确定参数 $g=2, c=2$,分类准确率达到93.33%。

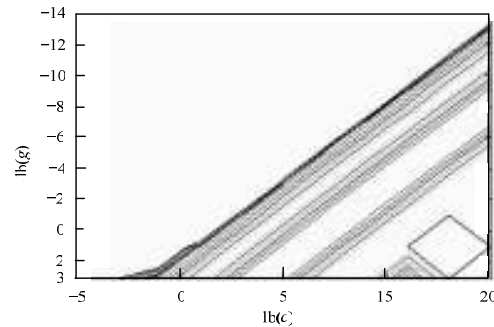
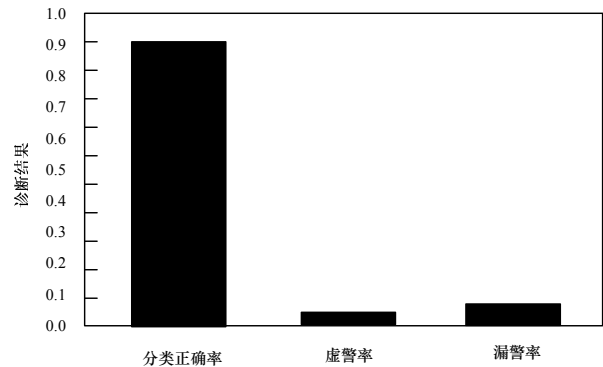


图 2 lb(c)和 lb(g)的关系

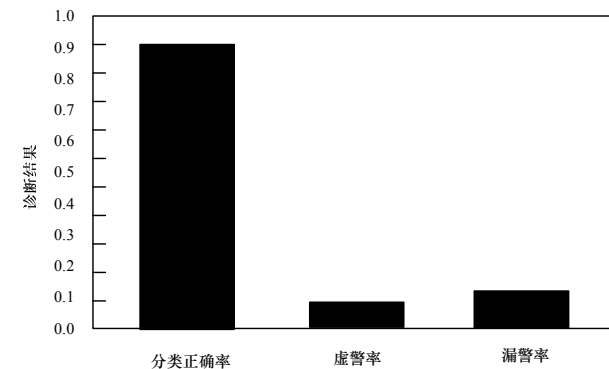
4 实验结果与分析

4.1 实验结果

将提取的近似熵特征值(训练样本集：测试样本集)按 2：1 和 1：1 这 2 种方案在数据集中随机抽取构成训练样本集和测试样本集，实验重复 20 遍，最后对结果取平均值。诊断结果如图 3 所示。从图 3(a)明显可知，采用前一种方案训练的 SVM 能更好地识别非 SCD 患者和 SCD 患者，分类正确率达到 90%以上。



(a)2：1 方案



(b)1：1 方案

图 3 分类识别结果

4.2 与同类算法的对比

利用仿真实验比较不同算法对正常和异常心电信号的分类正确率识别，见表 1。算法 a 使用本文介绍的非线性动力学方法提取近似熵特征值，采用 BP 神经网络进行分类识别；算法 b 按照文献中的方法，运用常规线性分析的方法得到心电频谱和形态学上的特征值，再使用 SVM 进行分类。算法 c 按照本文提出的 EMD 和非线性分析结合的算法提取特征值，然后输入 SVM 进行分类识别。训练与测试数据集选用 2：1 的方案。正确识别率为 ρ ，其中， ρ_1 是正常心电信号的识别率， ρ_2 是异常心电信号的识别率。训练时间为 t ，使用 Matlab

中的 tic、toc 指令计算。

表 1 3 种算法的识别正确率比较

算法	ρ_1 /(%)	ρ_2 /(%)	t/s
算法 a	89.9	90.5	1 335
算法 b	89.2	90.8	213
算法 c	92.0	91.0	141

4.3 结果分析

从图 3 可知，采用比例为 2：1 的识别结果优于比例为 1：1 的实验，说明通过多次的训练和学习，模式识别的分类器能更好地建立模型，从而获得更好的分类效果。

从表 1 的识别结果对比可知，非线性特征提取的方法比传统特征提取方法能更准确地反映信号的特异性，因此，能获得更高的正确识别率。另外，由于 SVM 在小样本问题中表现的特有优势，使用 SVM 分类器进行分类识别的算法 c 能取得比算法 a 更高的识别率；从处理速度来看，SVM 分类器具有明显的优势，对同一批心电信号进行训练，SVM 只需 BP 神经网络模型耗时的 1/10。由于该实验暂在 PC 上仿真实施，数据处理并没有很高的实时性要求，运用 SVM 分类器的实验时间足够满足心电信号的诊断需要。综上所述，将非线性特征提取的方法与 SVM 结合，不仅减少了处理时间，而且提高了对信号的分类识别能力。

5 结束语

本文从非线性动力学的角度研究心电信号，采用 EMD 和近似熵相结合的方法提取心电信号的特征值，然后使用 SVM 分类器进行测试验证。通过标准数据库的心电数据实验表明，该方法能客观地识别非 SCD 患者与 SCD 患者，对进一步实现其他疾病诊断具有重要的参考价值。但本文实验的心电信号仅限于标准心电数据库，因此，下一步将结合医院临床心电数据提高算法的实用性。

参考文献

[1] Davis D. 快速准确解读十二导联心电图[M]. 李立志, 史大可, 译. 北京: 科学技术文献出版社, 2004: 156-159.

[2] Novak P. Time-frequency Mapping of the QRS Complex in Normal Subjects and in Postmyocardial Infarction Patients[J]. Journal of Electrocardial, 1994, 27(1): 48-49.

[3] Xue Qiuzhen, Hu Yuhon, Tompkins W. Neural-network-based Adaptive Matched Filtering For QRS Detection[J]. IEEE Trans. on Biomedical Engineering, 1992, 39(4): 317-329.

[4] Sun P, Wu Q H, Weindling A M. An Improved Morphological Approach to Background Normalization of ECG Signals[J]. IEEE Trans. on Biomedical Engineering, 2003, 50(1): 117-121.

[5] 杨明静, 张 辉, 葛霁光. 心电图数据的非线性动力学研究[J]. 内蒙古民族师范学院学报: 自然科学版, 1998, 13(1): 82-85.

[6] Vapnik V. The Nature of Statistical Learning Theory[M]. New York, USA: Springer-verlag, 1995.

[7] 杨宏晖, 侯 宏, 曾向阳. 基于声信号人耳听觉谱特征的风机故障诊断[J]. 仪器仪表学报, 2009, 30(1): 175-179.

[8] 吴建宁, 王 珏. 采用支持向量机评估老年人步态对称性的研究[J]. 西安交通大学学报, 2007, 41(8): 995-997.

[9] 赵 敏, 郑崇勋, 赵春临. 利用小波分解和支持向量机的心理意识真实性识别研究[J]. 西安交通大学学报, 2010, 44(4): 119-123.