

基于 Holter 的心肌梗死猝死预警分析技术的研究

李明达

院（系）：计算机科学与技术

专 业：计算机科学与技术

学 号：1080310614

指导教师：黄剑华

2012 年 7 月

哈尔滨工业大学

毕业设计（论文）

题 目 基于 Holter 的心肌梗死猝死预警
分析技术的研究

专 业 计算机科学与技术

学 号 计算机科学与技术

学 生 李明达

指 导 教 师 黄剑华

答 辩 日 期 2012/7/2

哈尔滨工业大学毕业设计（论文）评语

姓名： 李明达 学号： 1080310614 专业： 计算机科学与技术
毕业设计（论文）题目： 基于 Holter 的心肌梗死猝死预警分析技术的研究

工作起止日期： 2012 年 3 月 14 日起 2012 年 7 月 2 日止

指导教师对毕业设计（论文）进行情况，完成质量及评分意见：

指导教师签字： _____ 指导教师职称： _____

评阅人评阅意见：

评阅教师签字： _____ 评阅教师职称： _____

答辩委员会评语：

根据毕业设计（论文）的材料和学生的答辩情况，答辩委员会作出如下评定：

学生 _____ 毕业设计（论文）答辩成绩评定为： _____

对毕业设计（论文）的特殊评语：

答辩委员会主任（签字）： _____ 职称： _____

答辩委员会副主任（签字）： _____

答辩委员会委员（签字）： _____

年 月 日

--

--	--

--	--

MATLAB R2011b SYM 工具箱

主题词：方位角、方位角以、方位角（北）、法上、方位角、方位角其一种分析

进度安排：开题 - 2012.3.31：使用 Matlab 软件，完成特征提取；
2012.4.1 - 2012.4.15：尝试使用分类器进行模式分类；
2012.4.16 - 2012.4.30：挑选最为接近随访情况的分类器模型；
2012.5.1 - 2012.5.15：将实验结果与医大医生交流，并加以修正；
2012.5.16 - 2012.6.15：论文撰写；
2012.6.16 - 结题：准备答辩

同组设计者及分工：无 独立完成

指导教师签字_____

年 月 日

教研室主任意见：

教研室主任签字_____

年 月 日

摘 要

心脏性猝死这一恶性疾病与自主神经功能的强弱有着密切的联系。而自主神经的功能可以在一些动态心电图（DCG or Holter）的某些特征中得以体现。为达到联合这些特征从而提高心脏性猝死预测能力的目的，提出了一种基于 Holter 的心肌梗死猝死预警系统。为构造该预警系统，首先创新性地提出了一种基于 DCG 心电信号的 R 波检测新算法，提高了 R 波波峰检测的准确率；然后利用得到的 RR 间期序列提取了心率变异性分析、心率减速力分析以及窦性心律震荡分析的多项特征，经过对这些特征以及哈尔滨医科大学第一附属医院心内科提供的 Holter 样本的筛选，最终确定了用于分类的特征和样本；最后应用构造出的朴素贝叶斯分类器以及 SVM 分类器，对样本进行了模式分类，将心脏性猝死高危患者从普通患者中筛选出来。通过与病人的回访记录（死亡或存活）进行的比对，绘制了能够评价分类系统性能的 ROC 曲线，计算了 AUC 的值，通过对各种分类结果的对比和分析，得到了 3 个分类表现良好（ $AUC > 0.8$ ）的分类模型。

关键词：Holter；猝死预警；R 波检测；朴素贝叶斯；SVM；ROC

Abstract

It is known that sudden cardiac death is in connection with the autonomic nervous. There are some electrocardiogram related to the autonomic nervous. To combine these to improve the prediction ability of sudden cardiac death, we propose a system base on Holter signal to predict the probability of myocardial infarction sudden death. To build this system, firstly, a novel R-wave detection algorithm of DCG signal is proposed, so the precise detection rate is increased; then in advantage of RR interval series, the characteristics of the heart rate variability analysis, deceleration capacity of heart rate analysis and heart-rate turbulence analysis are proposed. After sifting, the final characters and classifying instances is chosen. Finally, Naive Bayes classifier and Support Vector Machine classifier are used for pattern classification and patients at high risk of sudden cardiac death are picked up from all the cardiopath. Compared with the record of return visit (dead or alive), the receiver operating characteristic curve is drawn and the area under the roc curve is calculated. By comparison and analysis, three good classification models ($AUC > 0.8$) are obtained.

Keywords: Holter; prediction of sudden death; R-wave detection ; naive bayes; SVM; ROC

目 录

摘 要	I
ABSTRACT	II
第 1 章 绪 论	1
1.1 引言	1
1.2 题目背景和研究意义	1
1.2.1 课题的来源及背景	1
1.2.2 课题的研究意义	1
1.3 课题研究现状及分析	2
1.3.1 心率变异性分析	2
1.3.2 心率减（加）速力分析	2
1.3.3 窦性心律震荡分析	3
1.4 本文的主要研究内容	3
第 2 章 R 波波峰的定位与 RR 间期的提取	4
2.1 引言	4
2.2 R 波检测原理	4
2.2.1 斜率步长的确定	5
2.2.2 平均双向斜率	6
2.2.3 第一个 R 波波峰的确定	7
2.2.4 相对高度	7
2.3 R 波检测流程	8
2.3.1 扫描定位	9
2.3.2 伪 R 波的排除与 R 波波峰疑似点的定位	10
2.4 实验结果分析	10
2.5 异常 RR 间期的剔除	13
2.6 本章小结	13

第 3 章 心电分析的特征提取	14
3.1 引言	14
3.2 心率变异性分析的特征提取	14
3.2.1 时域分析	14
3.2.2 频域分析	15
3.3 心率减（加）速力分析的特征提取	16
3.4 窦性心律震荡分析中的特征提取	19
3.4.1 室性早搏的检测	19
3.4.2 震荡初始	20
3.4.3 震荡斜率	20
3.5 本章小结	20
第 4 章 心电样本的模式分类	22
4.1 引言	22
4.2 分类前的特征选择	22
4.3 样本来源及筛选	23
4.4 分类器原理	23
4.4.1 朴素贝叶斯分类器	24
4.4.2 SVM 分类器	25
4.5 实验结果及分析	26
4.6 本章小结	29
结 论	30
参考文献	31
致 谢	33

第1章 绪 论

1.1 引言

作为世界最致命的十大疾病之一的心脏病每年可以带走数以百万计的生命。据世界卫生组织统计,每年全球约有 1700 万人死于心脑血管疾病,其中心脏性猝死(sudden cardiac death, SCD)占 40%~50%^[1]。卫生部心血管病防治研究中心披露的信息显示,我国每年有 50 余万人死于 SCD,居世界各国之首^[2]。

世界卫生组织关于猝死的定义,是指一个人在 6 小时之内自然发生出乎意料的突然死亡,称之为“猝死”^[3]。导致猝死的因素很多,但以心脏性猝死最为常见,占 80% 以上^[4]。心脏性猝死救治成功率低,我国抢救成功率不到 1%^[5],绝大部分患者还没来得及送到医院就已死亡。因此,对这类疾病的高危人群进行心肌梗死猝死预警分析技术的研究是很有意义的一个课题。

1.2 题目背景和研究意义

1.2.1 课题的来源及背景

心脏性猝死在全世界范围内都是心血管防治中的重要问题。现代医疗已经有一些防治 SCD 的有效措施,如溶栓和抗溶栓药物治疗、机械性血管重建技术、安置可植入性心律转复除颤器(implantable cardioverter defibrillator, ICD)。然而,SCD 的突发性仍然使得抢救治疗后的存活率不到 2%^[6]。因而对心肌梗死后患者进行 SCD 危险预测,从中识别出高危人群并及时施以预防性治疗显得非常重要。

SCD 的危险预测可分为有创和无创两类:有创方法是电生理检查属于介入方法;无创方法是通过心电图变量等无创指标进行危险预测。前者风险大,费用高,耗时长,而且对于许多心肌梗死后患者的检测结果往往是阴性,导致了较低的阴性预测正确值。因此,如何采用无创方法来实现较好的心肌梗死 SCD 危险预测,是近年来心血管防治研究热点之一^[6]。

针对这一热点研究方向,模式识别研究中心和哈尔滨医科大学第一附属医院心内科合作开展了“基于 Holter 的心肌梗死猝死预警分析技术的研究”。

1.2.2 课题的研究意义

动态心电图(Dynamic Electrocardiography, DCG)是一种可以长时间(24 小时)连

续记录并编集分析人体心脏在活动和安静状态下心电图变化的方法^[7]。1957 年美国实验物理学家 Nor-man J•Holter 发明了动态心电图，故动态心电图简称 Holter。

在医学上，预防猝死的研究是一个受到广泛关注的课题。心性猝死最常见的原因是室速或室颤，发生前常有心电活动不稳的室性心律失常，从 Holter 心电信号当中能够发现其发生规律，进而对心律失常进行定性和定量的分析^[8]。

心肌梗死急性期的自主神经失衡与其预后密切相关^[9]，因而在猝死的研究中很重视自主神经的作用。研究中发现诸如心率变异性^[10]、心率减(加)速力^[11]、窦性心律震荡^[12]、T 波电交替^[13]等因素均能检测自主神经功能。

本项目的研究能及时并比较全面地进行 SCD 的危险预警，提高心梗患者猝死及全因死亡预警的敏感性，有助于临床上及时采取有效的预防和治疗措施，具有相当高的应用价值。

1.3 课题研究现状及分析

随着计算机和信号处理技术的发展，通过分析 ECG、Holter 等大量信息获得具有潜在临床价值的新知识已经成为可能。近几年来对心肌梗死后 SCD 危险预测指标的研究也已取得很大进展，其中对反映自主神经作用的指标对 SCD 的危险预测而言是有极大意义的预测指标。

1.3.1 心率变异性分析

心率变异性(heart rate variability, HRV)是指连续心跳间 *RR* 间期的微小涨落^[10]。HRV 反映了心脏交感神经和迷走神经活动的紧张性与均衡性，是一种能够检测自主神经性活动的非侵入性指标。

我国近几年在 HRV 分析方面，主要开展了时域分析算法和频域分析算法的研究。该项研究肯定了心率变异性分析的无创性、敏感度高和可定量等优点，同时也指出了时域频域分析的特异性较低，不能进一步区别交感或迷走神经的作用及均衡性等缺点。

1.3.2 心率减（加）速力分析

DC(心率减速力)指标、*AC*(心率加速力)指标和自主神经密切相关。交感神经兴奋性增加、心率变快，*AC* 增加；迷走神经兴奋性增加，心率变慢，*DC* 增加。*DC* 和 *AC* 的检测能定量地评估受检者迷走神经张力的高低，是预警猝死高危患者的一种新的无创心电技术^[11]。

心率减速力检测技术是德国慕尼黑心脏中心 Georg Schmidt 教授近年发现并提出的一种检测自主神经张力的新技术。这项分析技术在国外应用较早也较为广

泛，而这项新技术有着较强的优势，敏感性高，特异性强^[14]。因此该技术的推广和应用在我国临床的 SCD 预警方面意义重大。

1.3.3 窦性心律震荡分析

窦性心律震荡(heart rate turbulence, HRT)是指一次伴有代偿间期的室性期前收缩后出现的窦性心律先加速后减速的现象^[15, 16]。窦性心律震荡主要反映了迷走神经(与交感神经之间平衡)的反射功能，迷走神经活性的增加对心脏具有保护的作用，而交感神经活性增强会促使心律失常的发生^[15, 16]。

研究表明，虽然它具有测量简单、预测精度高、可充分反映心脏自主神经功能的特点，然而 HRT 现象并不是在所有的心血管病患者中都能够观察到。这使得其应用受到一定程度的限制^[15, 16]。

1.4 本文的主要研究内容

通过上文可以看到，研究人员提出了很多 SCD 危险预测的方法和指标，对临床心电的研究产生了巨大的影响。同时，我们也看到每种分析方法都有各自的局限性：HRV 分析的特异性较低；心率减速力分析未在我国临床的 SCD 预警方面进行推广。

本课题旨在综合以上各 SCD 危险预测指标（包括心率变异性、心率减（加）速力和窦性心律震荡三种分析当中的一些特征），对心肌梗死患者进行模式分类，进而将 SCD 高危患者从普通患者中筛选出来，最终达到 SCD 预警的目的。

为解决这一课题，本文首先提出一种基于 DCG 心电信号的 R 波检测新算法；然后利用提取的 RR 间期序列针对心率变异性分析、心率减速力分析以及窦性心律震荡分析进行特征提取；之后再对已经提取的特征和获得的样本进行筛选，利用朴素贝叶斯分类器和 SVM 分类器进行模式分类；最后针对分类结果进行分析，挑选出分类性能较为优秀的几个分类器模型。

第 2 章 R 波波峰的定位与 RR 间期的提取*

2.1 引言

心率变异性分析和心率减（加）速力分析是两种基于 DCG 心电信号的有效预警心律失常的方法，而两种分析都依赖于准确的 R 波检测。因此，R 波的准确定位对于 SCD 的预警诊断有着积极的临床意义。

文献^[17]采用的是最大值双重搜索技术，文献^[18]采用的是差分运算的方法。两种算法的共同点是利用了心电信号中 R 波斜率大的特点，以相邻点之间的斜率为主要特征对 R 波进行检测。但上述算法对噪声干扰比较敏感，对于采样频率较低的 DCG 心电信号的处理效果不够理想。

本文创新地提出了 DCG 心电信号处理中的一种 R 波检测新算法，利用了 R 波波峰斜率大、幅度高的特点，提取了平均双向斜率和相对高度两项特征，对 MIT-BIH Long-Term ECG 数据库中的全部 7 个样本实现了 R 波波峰点坐标的定位。以每个样本的前 50 个 R 波波峰的实际位置为标准，本文算法的正确率高达 98.3%，高于最大值双重搜索技术的 95.2% 以及差分运算法的 90.7%。

2.2 R 波检测原理

一个正常心电图的每个心动周期由 P 波、Q 波、R 波、S 波、T 波以及 U 波组成（如图 2-1 所示）。其中特征最为显著的就是 R 波，它充分反映了心室收缩时心脏的电行为。因此解决心动周期提取问题的关键就是 R 波的准确定位。

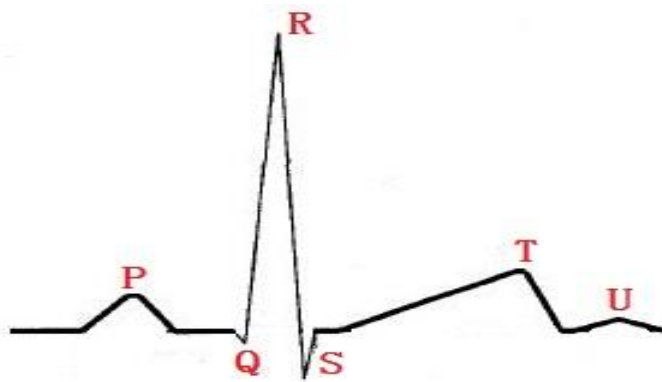


图 2-1 正常心电图的一个心动周期

观察发现，QRS 波群中的 QR 波和 RS 波斜率高，幅度大。利用这两个特点，

* 本章关于“R 波波峰定位”的内容取自本文作者编写的《一种基于 DCG 心电信号的 R 波检测新算法》一文，该文章已经投递到《天津大学学报（自然科学版）》

对心电信号这两方面的特征进行提取可以突显 R 波波峰的位置，从而达到准确定位 R 波的目的。

2.2.1 斜率步长的确定

在本文对心电信号的处理中，定义斜率步长为用于求取斜率的两采样点间的采样点间隔。

设 $N(i)$ 为心电图中采样点 i 的纵坐标（ i 点对应的电压值）， f 为心电信号的采样频率。采样点 i 和采样点 j 之间的斜率，见式（2-1）：

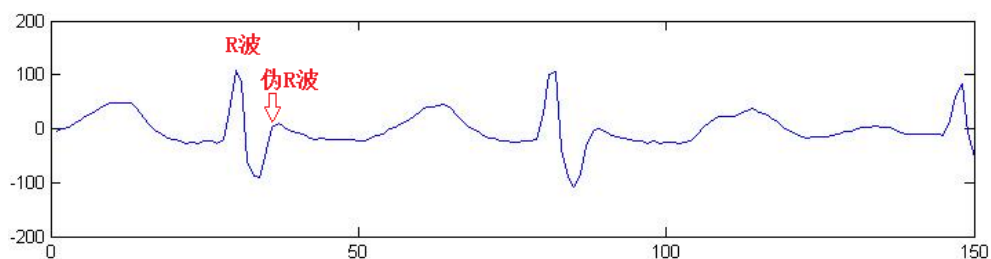
$$K = \frac{N(i) - N(j)}{(i - j) \times \frac{1}{f}} \quad (2-1)$$

由于采样频率 f 是一个定值，因此相邻采样点间的时间间隔相等，于是式（2-1）可以简化为：

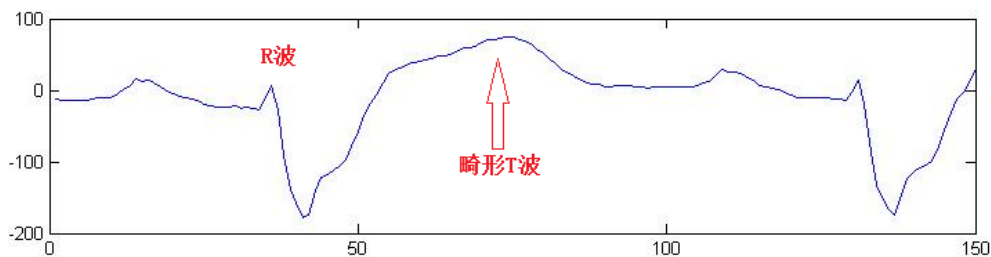
$$k = \frac{N(i) - N(j)}{i - j} \quad (2-2)$$

由公式（2-2），采样点 i 和 j 的斜率步长就是 $i - j$ 。

因为 DCG 心电信号的采样频率较小，那么如果斜率步长设置的不够合理，容易误将伪 R 波（如图 2-2 a）所示）或畸形 T 波（如图 2-2 b）所示）误识为 R 波。为了排除这两种情况的发生，需要对斜率步长进行合理的设定。



a) 14149号记录中的伪R波波形



b) 15814号记录中的畸形T波波形

图2-2 MIT-BIH Long-Term ECG数据库中容易误识的波形

本文所用的 MIT-BIH Long-Term ECG 数据库所有心电信号的采样频率均为

128Hz, 资料表明, 一般 QRS 波的时间是 0.06s 左右, 也就是说 QR 斜率步长和 RS 斜率的步长和大约在 7 个采样点左右 ($0.06 \div (1 \div 128) = 7.68$), 为突显 QR 斜率和 RS 斜率高的同时保持 R 波的对称性, QR 坡的斜率步长和 RS 坡的斜率步长统一取 3。

2.2.2 平均双向斜率

为突显出 R 波两侧斜率高的特性, 本文定义了前向斜率、后向斜率以及平均双向斜率三个概念:

设 $tstep$ 为斜率步长。对采样点 i 而言, 点 i 与点 $i - tstep$ 之间的斜率 $k_1(i)$ 为点 i 的前向斜率, 见式 (2-3):

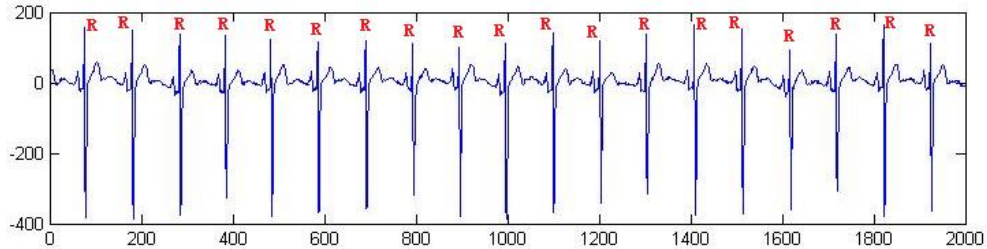
$$k_1(i) = \frac{N(i) - N(i - tstep)}{tstep} \quad (2-3)$$

而点 i 与点 $i + tstep$ 之间的斜率的相反数 $k_2(i)$ 为点 i 的后向斜率, 见式 (2-4):

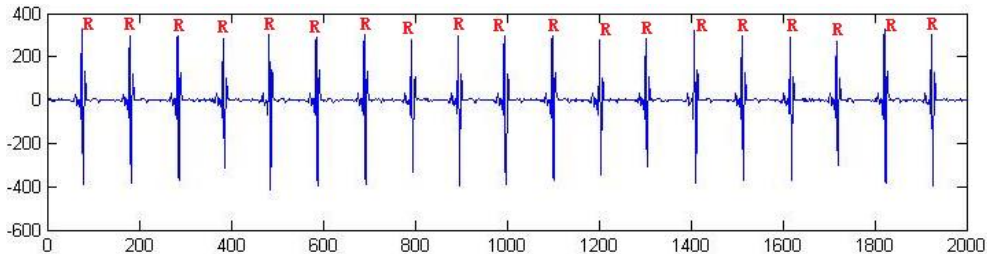
$$k_2(i) = -\frac{N(i + tstep) - N(i)}{tstep} \quad (2-4)$$

点 i 的前向斜率 $k_1(i)$ 和后向斜率 $k_2(i)$ 的平均值 $m(i)$ 为点 i 的平均双向斜率, 见式 (2-5):

$$m(i) = \frac{k_1(i) + k_2(i)}{2} \quad (2-5)$$



a) 心电信号序列



b) 平均双向斜率序列

图 2-3 MIT-BIH Long-Term ECG 数据库中 14046 号心电记录

将式 (2-3) 和 (2-4) 代入式 (2-5) 得:

$$m(i) = \frac{2N(i) - N(i - tstep) - N(i + tstep)}{2 \times tstep} \quad (2-6)$$

由于对每个点 i 而言, $tstep$ 都相等。因此在计算平均双向斜率时, 式 (2-6) 可简化为:

$$M(i) = \frac{2N(i) - N(i - tstep) - N(i + tstep)}{2} \quad (2-7)$$

式中 $i = tstep + 1, tstep + 2, \dots$

如图 2-3 a) 和 b) 所示, 在平均双向斜率序列中, R 波波峰的位置更加突出, 识别起来更加容易。

2.2.3 第一个 R 波波峰的确定

定位第一个 R 波波峰是定位全部 R 波前需要解决的首要问题。只有确定了第一个 R 波波峰的位置才能获得 R 波形态特征的阈值, 进而迭代地把 R 波检测进行下去。

第一个 R 波波峰的定位依赖于初始平均双向斜率阈值 ($Mx \cdot kran$) 的确定: 首先在信号开头的至少含有两个 R 波的信号段中求取最大平均双向斜率 Mx 。然后选择一个合适的阈值 $kran$ (如 0.6), 以 $Mx \cdot kran$ 为平均双向斜率的阈值重新搜索这一段信号, 可以定位第一个最大平均双向斜率的位置 t 。在该点附近 ($t - 5 \sim t + 5$) 搜索到的最大峰值 top , 见式 2-8, 就是第一个 R 波波峰的位置。

$$top = \arg \max_i N(i) \quad (2-8)$$

式中 $i \in (t - 5, t + 5)$

2.2.4 相对高度

为突显 R 波幅度特性的同时又能克服某些具有基线漂移 (如图 2-4 a) 和 b) 所示) 的信号在 R 波检测中的干扰, 本文定义前向相对高度、后向相对高度以及相对高度的概念:

对采样点 i 而言, 点 i 与点 $i - tstep$ 之间的纵坐标之差 $h_1(i)$ 为点 i 的前向相对高度, 见式 (2-9):

$$h_1(i) = N(i) - N(i - tstep) \quad (2-9)$$

点 i 与点 $i + tstep$ 之间的纵坐标之差 $h_2(i)$ 为点 i 的后向相对高度, 见式 (2-10):

$$h_2(i) = N(i) - N(i + tstep) \quad (2-10)$$

点 i 的前向相对高度 $h_1(i)$ 与后向相对高度 $h_2(i)$ 的平均值 $H(i)$ 即为点 i 的相对高度, 见式 (2-11):

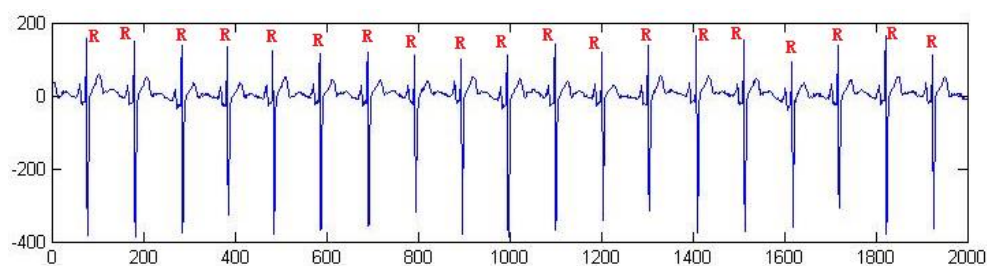
$$H(i) = \frac{h_1(i) + h_2(i)}{2} \quad (2-11)$$

将式（2-9）和（2-10）代入式（2-11）得：

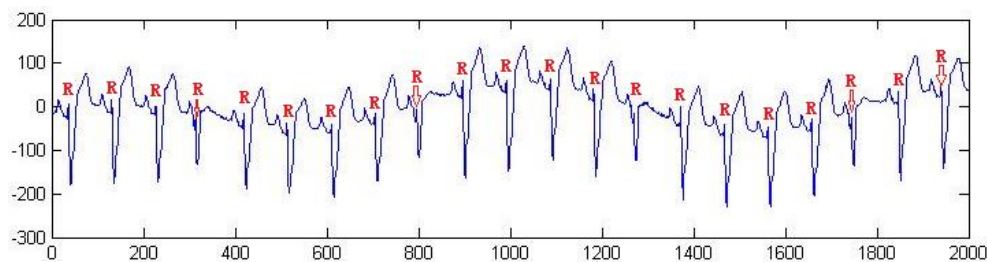
$$H(i) = \frac{2N(i) - N(i - tstep) - N(i + tstep)}{2} \quad (2-12)$$

式中 $i = tstep + 1, tstep + 2, \dots$

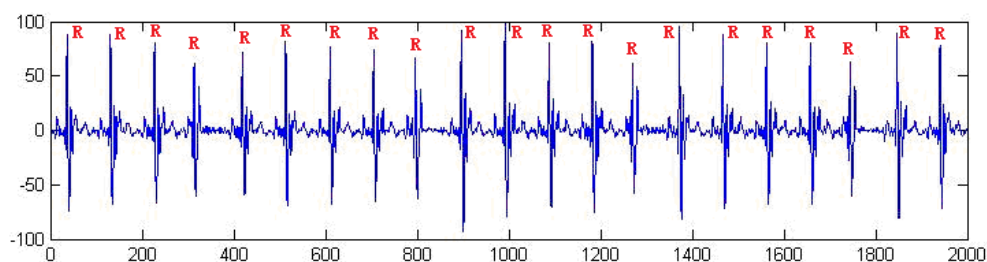
如图 2-4 b) 和 c) 所示，相对高度序列不仅突出了 R 波波峰的位置，而且有效解决了基线漂移的干扰。



a) 没有基线漂移的 14046 号心电记录



b) 有基线漂移的 15814 号心电记录



c) 15814 号心电记录的相对高度序列

图 2-4 MIT-BIH Long-Term ECG 数据库中 14046 号和 15814 号心电记录

2.3 R 波检测流程

基于平均双向斜率和相对高度的 R 波检测算法流程图，如图 2-5 所示。

本文算法的流程大致可分为三部分：第一部分为双向斜率和相对高度的计算；第二部分为第一个 R 波波峰的定位；第三部分为迭代地依次定位 R 波。其中前两部分在算法原理中已经做了说明，以下介绍第三部分。

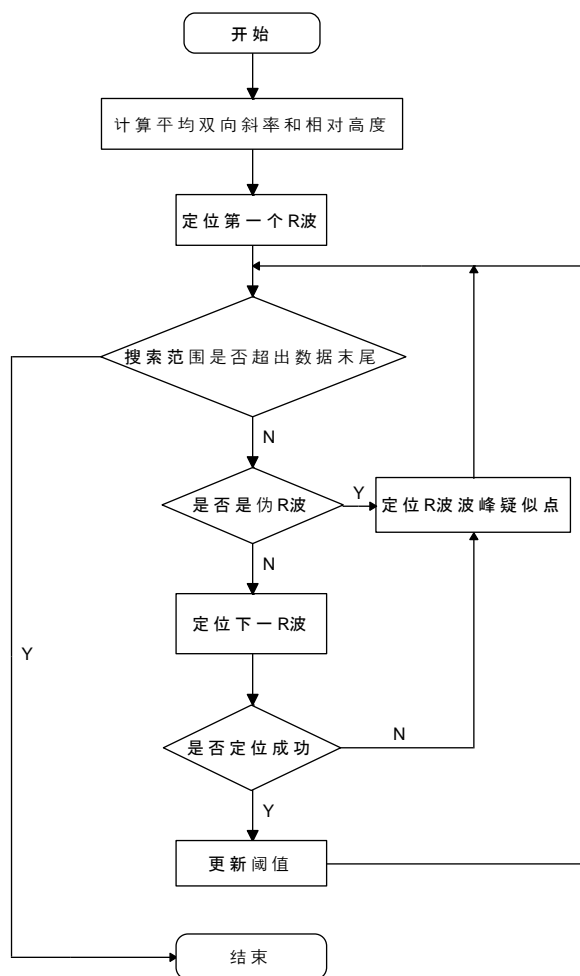


图 2-5 R 波检测程序流程图

2.3.1 扫描定位

以当前已经检测出的 R 波波峰 a 的平均双向斜率 $M(a)$ 和相对高度 $H(a)$ 为标准，选择合适的阈值 dis 、 $downrange$ 和 $uprange$ ，在一定的范围 ($min \sim max$) 内搜索，满足下列条件的采样点 i 即为下一个 R 波波峰：

1) 控制平均双向斜率的条件：

$$M(i) \geq dis * M(a) \quad (2-13)$$

2) 控制相对高度的条件：

$$H(i) \geq downrange * H(a) \ \& \ H(i) \leq uprange * H(a) \quad (2-14)$$

3) 判断是否是波峰的条件：

$$N(i) > N(i-1) \ \& \ N(i) \geq N(i+1) \quad (2-15)$$

经实验整理，可以参考的阈值为 $dis = 0.5$ ， $downrange = 0.5$ ， $uprange = 3$ ， $min = 40$ ， $max = 640$ 。

2.3.2 伪 R 波的排除与 R 波波峰疑似点的定位

寻找下一个 R 波波峰疑似点是在 R 波定位中需要解决的另一问题。当定位的“R 波”是由噪声等干扰产生的伪 R 波（如图 2-6 所示）或在搜索范围内未能定位出 R 波时，这时就需要识别并排除伪 R 波，继而定位后续信号中的 R 波波峰疑似点，以便于再次迭代搜索时进行筛选和判断。

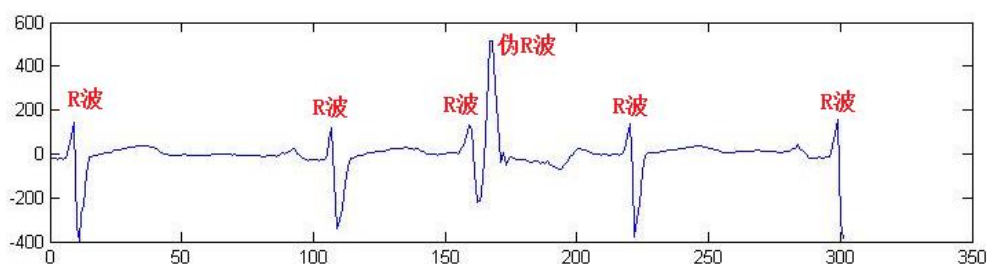


图2-6 MIT-BIH Long-Term ECG数据库中14046号心电记录

本文的算法分为两种情况进行处理：

1) 搜索时出现异常波峰（如图 2-6 所示）：

若 $H(i) > uprange * H(a)$ ，需立即排除该点，并将 $i + 1 \sim i + max$ 范围内的平均双向斜率最大值点 s_1 ，见式（2-16），作为 R 波波峰疑似点，并以该点为起点开始下一轮迭代。

$$s_1 = \arg \max_j M(j) \quad (2-16)$$

式中 $j \in (i + 1, i + max)$

2) 在搜索范围内未能搜索到符合阈值的波峰：

在 $a + 1 \sim a + max$ 范围内未能搜索到满足（2-16）的 s_1 作为 R 波波峰疑似点，应将 $a + max + 1 \sim a + 2 * max$ 范围内的平均双向斜率最大值点 s_2 ，见式（2-17），作为 R 波波峰疑似点，并以该点为起点开始下一轮迭代。

$$s_2 = \arg \max_i M(i) \quad (2-17)$$

式中 $i \in (a + max + 1, a + 2 * max)$

在每轮迭代之前需要对该疑似点进行式（2-13）、（2-14）、（2-15）三项条件的审核以避免误识的情况发生，若未通过审核，则判定该疑似点为伪 R 波波峰，并再次搜索，直到搜索到通过审核的疑似点为止。

2.4 实验结果分析

目前国际上公认的可作为标准的心电数据库有三个——美国麻省理工学院所提供的 MIT-BIH 数据库、美国心脏学会的 AHA 数据库以及欧洲 ST-T 心电数据库。

其中 MIT-BIH 数据库近年来应用最为广泛。

本文针对 MIT-BIH Long-Term ECG 数据库中的全部 7 个心电信号记录——10446、14134、14149、14157、14172、14184 和 15814，以每个心电信号第一导的前 50 个 R 波波峰坐标的实际位置为标准，运用本文算法以及最大值双重搜索技术和差分运算法算法进行测试，检测结果如表 2-1 所示。

正确率 AR 的计算见式（2-18）：

$$AR = \frac{TP - FP - FN}{TPB} \times 100\% \quad (2-18)$$

式中 TPB ——为心拍总数；

FP ——为漏检个数；

NP ——为误检个数。

表 2-1 本文算法与最大值双重搜索技术和差分运算法的 R 波检测结果

	本文算法				最大值双重搜索技术				差分运算法			
	FP	NP	TPB	AR	FP	NP	TPB	AR	FP	NP	TPB	AR
10446	0	3	50	94.0%	3	0	50	94.0%	1	0	50	98.0%
14134	0	0	50	100.0%	0	0	50	100.0%	0	2	51	96.1%
14149	0	0	50	100.0%	0	0	50	100.0%	7	0	50	86.0%
14157	0	0	50	100.0%	0	0	50	100.0%	0	1	50	98.0%
14172	0	3	50	94.0%	0	8	52	84.6%	1	7	52	84.6%
14184	0	0	50	100.0%	1	2	50	94.0%	4	4	50	84.0%
15814	0	0	50	100.0%	0	3	50	94.0%	1	5	50	88.0%
合计	0	6	350	98.3%	4	13	352	95.2%	14	19	353	90.7%

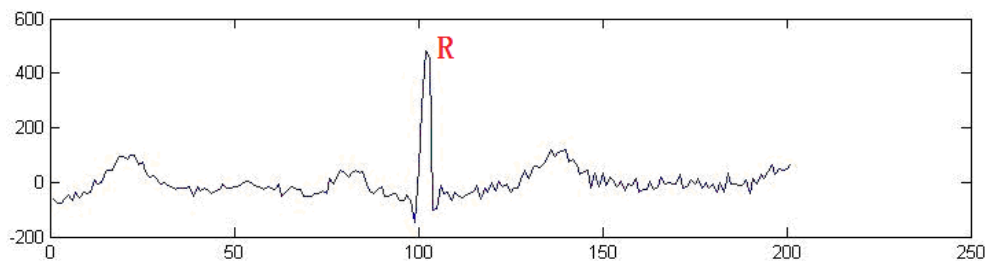


图 2-7 MIT-BIH Long-Term ECG 数据库中 14134 号心电记录

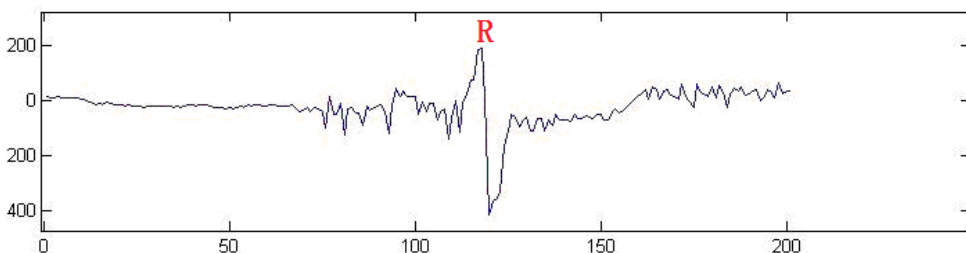


图 2-8 MIT-BIH Long-Term ECG 数据库中 14172 号心电记录

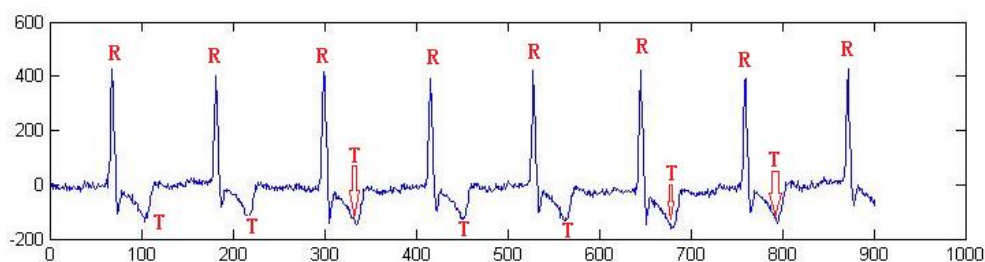
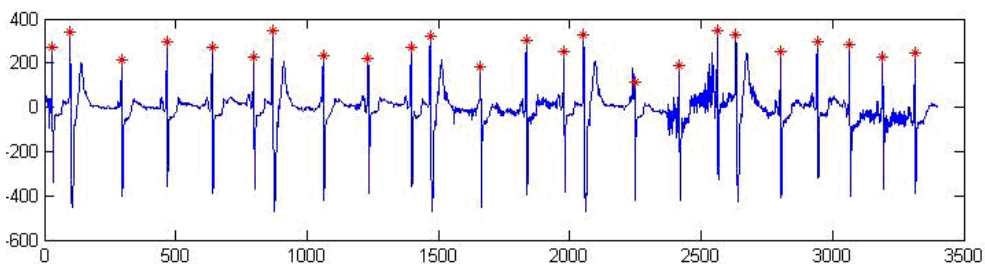
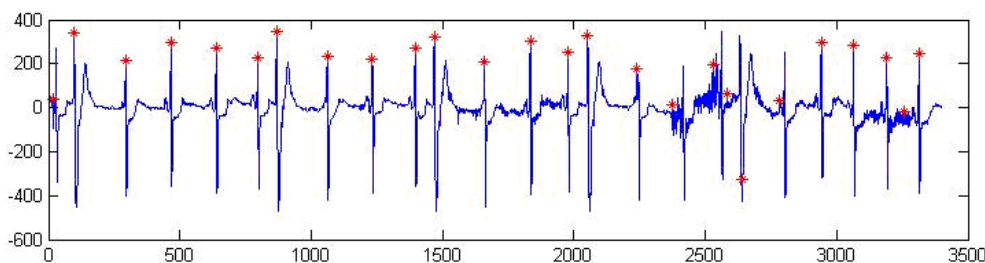


图 2-9 MIT-BIH Long-Term ECG 数据库中 14157 号心电记录

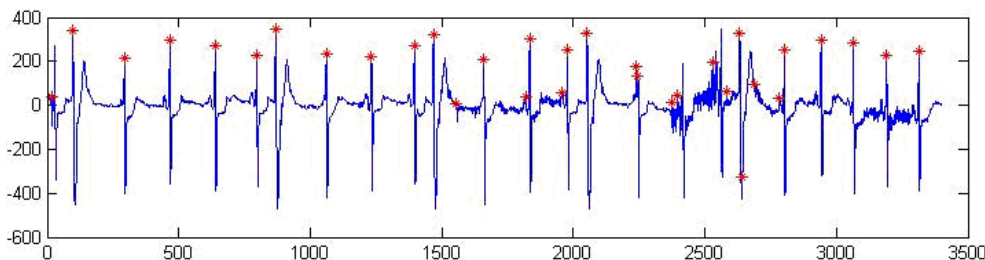
实验中所采用的 MIT-BIH Long-Term ECG 数据库的全部 7 个样本涵盖了 DCG 心电信号中可能出现的各种波形: 14046 号样本存在伪 R 波(如图 2-6 所示); 15814 号样本存在基线漂移和畸形 T 波(如图 2-2 和图 2-4 所示); 14134 和 14172 号样本存在一定的噪声干扰(如图 2-7 和图 2-8 所示); 14157 号样本存在倒置的 T 波(如图 2-9 所示)。



a) 本文检测结果



b) 最大值双重搜索技术检测结果



c) 差分运算法检测结果

图 2-10 MIT-BIH Long-Term ECG 数据库中 14172 号心电记录的检测结果

在这些波形的影响下, 最大值双重搜索技术的误检个数较多, 而差分运算法的漏检和误检个数都较多(如图 2-10 所示, 其中的星状点为各算法检测出的 R 波

波峰点)。这主要是由于两种算法在求取斜率时只考虑了相邻采样点间的单向斜率，而本文算法求取的是平均双向斜率，加上相对高度的协调控制，能够排除各种波形的干扰，达到准确定位的目的。

在检测精度方面，将本文算法检测出的 R 波位置与心电信号中 R 波的实际位置相对照，其误差不大于 1 个采样点（其中约 99.7% 能准确定位），具有相当高的定位精度。

2.5 异常 RR 间期的剔除

提取出相邻 R 波波峰之间的采样点距离即可获得 RR 间期序列。但其中必然会有一些离群点，资料^[19]显示，窗口滤波法可以有效剔除这些异常的 RR 间期值：

设 RR_i 为第 i 个 RR 间期，其中 $i \geq 21$ 。在以 i 点为中心，20 为窗口半径的窗口中计算 RR 间期平均值 $mean$ ，见式 (2-19)。当 i 点的心动周期 RR_i 满足式 (2-20) 或 (2-21) 时，点 i 被剔除。

$$mean = \frac{\sum_{j=i-20}^{i-1} RR_j + \sum_{j=i+1}^{i+20} RR_j}{40} \quad (2-19)$$

$$RR_i > mean * 120\% \quad (2-20)$$

$$RR_i < mean * 80\% \quad (2-21)$$

2.6 本章小结

本章提出了一种基于 DCG 心电信号的 R 波检测新算法。该算法有如下特点：

1) 利用平均双向斜率和相对高度双重特征进行定位的策略，使得该算法对信号噪声的敏感性较低，可以自动排除因信号噪声产生的伪 R 波波峰。因此在检测 R 波之前不用处理基线漂移和消噪，较好地保持了原信号波形的细节，提高了心电信号处理的效率；

2) 能够准确识别每个 R 波波峰的位置，克服了最大值双重搜索技术和差分运算算法中难以避免的漏识和误识的情况；

3) 算法易于重现，为心率变异性分析、心率减速力分析提供了准确的数据特征；

利用该算法和窗口滤波法可以将原始的心电信号转化为 RR 间期序列，为心率变异性分析、心率减（加）速力分析以及窦性心律震荡分析的开展做好了准备。

第3章 心电分析的特征提取

3.1 引言

针对上一章提取的 RR 间期序列，可以进行心率变异性分析、心率减(加)速力分析以及窦性心律震荡分析，从中提取出一些具有可分性的特征，为最终的模式分类提供候选特征集。

3.2 心率变异性分析的特征提取

目前 HRV 的分析方法主要有基于线性分析的时域分析、频域分析和非线性分析这三类。其中时域分析法和频域分析法具有理论成熟、算法简单、各项指标意义明确等优点，因此较广泛地应用于临床和医学实验中^[10]。本文主要选取 HRV 分析方法中的时域分析和频域分析进行特征提取。

3.2.1 时域分析

时域分析是通过统计学中的离散趋势分析法来计算 RR 间期的变化，该分析主要包括统计学分析和几何图形分析。统计学分析是通过计算一系列有关 RR 间期的数理统计指标，来评价 HRV 的临床价值^[10]。其主要指标有均值 ($MEAN$)、总体标准差 ($SDNN$)、均值标准差 ($SDANN$)、标准差均值 ($ASDNN$) 等。几何图形分析是对 RR 间期的分布进行分析，给出 RR 间期的变异度。常用的几何图形分析指标包括 HRV 三角指数 (Ti) 等。本文选取的时域分析特征为均值 ($MEAN$)、总体标准差 ($SDNN$)、均值标准差 ($SDANN$)、标准差均值 ($ASDNN$) 以及 HRV 三角指数 (Ti) 共五项特征。

下面给出的是这几项特征的计算方法：

将 RR 间期序列中的数据代入下列公式，进行指标计算。

$$MEAN = \overline{RR} = \sum_{i=1}^N RR_i / N \quad (3-1)$$

$$SDNN = \sqrt{\frac{1}{N} \sum_{i=1}^N (RR_i - \overline{RR})^2} \quad (3-2)$$

$$SDANN = \sqrt{\sum_{i=1}^{288} (\overline{RR}_i - \overline{RR}_{5min})^2 / 288} \quad (3-3)$$

$$ASDNN = \sum_{i=1}^{288} SDNN_{index_i} / 288 \quad (3-4)$$

$$Ti = RRnum / RRmaxnum \quad (3-5)$$

式中 RR_i ——为 RR 间期序列中第 i 个 RR 间期；

$\overline{RR_i}$ ——为 RR 间期序列中第 i 个 5 分钟内 RR 间期的均值；

$\overline{RR_{5min}}$ ——为 288 个 $\overline{RR_i}$ 的均值；

$SDNNindex_i$ ——为第 i 个 5 分钟内的 $SDNN$ ；

$RRnum$ ——为 RR 间期序列中 RR 间期总数；

$RRmaxnum$ ——为 RR 间期序列中所占比例最大的 RR 间期个数。

3.2.2 频域分析

将连续正常的 RR 间期进行基于快速傅里叶变换（FFT）的经典谱估计，可以作为定量的指标来描述 HRV 信号的能量分布情况。它将各种生理因素作适当分离后进行分析，因而有较大的临床应用价值^[10]。常用的谱参数包括极低频段功率（VLF）、低频段功率（LF）、高频段（HF）、信号总功率（TP）、LF 与 HF 的比值（LF_HF）等。本文选取的频域分析特征就是以上的五项特征。

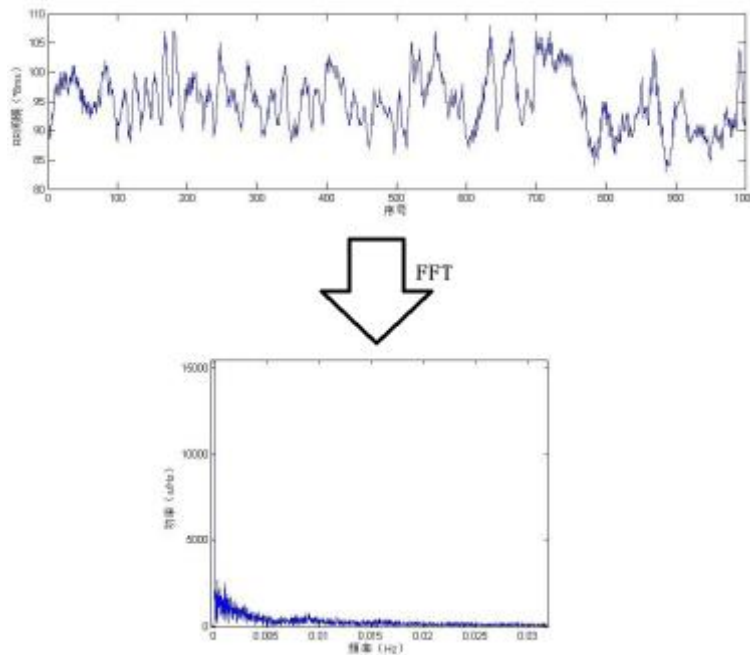


图 3-1 RR 间期序列图到频谱图的转换

下面给出的是这几项特征的计算方法：

步骤 1：对 RR 间期序列进行 FFT^[17, 20]，获得频谱图(如图 3-1 所示)；

步骤 2：频谱图中下列各指标的频率段的面积即为对应指标的值。

VLF.....极低频段(0.0033~0.04Hz)的功率

LF	低频段(0.04~0.15Hz)的功率
HF	高频段(0.15~0.4Hz)的功率
TP	信号总功率 (0.0033~0.4Hz) 的功率
LF_HF	LF 与 HF 的比值

3.3 心率减（加）速力分析的特征提取

心率减（加）速力的检测是通过 24 小时中心率的整体趋向性分析以及心脏减速能力的测定，定量评估受检者迷走神经张力的高低，进而筛选和预警猝死高危患者的一种新的无创心电技术。该技术的设想以及方法学已经过临床循证医学的验证，证实具有检出与预测猝死高危者的较强能力^[11, 14]。

心率减(加)速力分析分成以下 6 个步骤^[11, 14]：

步骤 1：提取 RR 间期序列图：将 24 小时的动态心电图(如图 3-1 所示)转化为以心搏序号为横坐标，以心动周期即 RR 间期的值为纵坐标的序列图(如图 3-2 所示)。

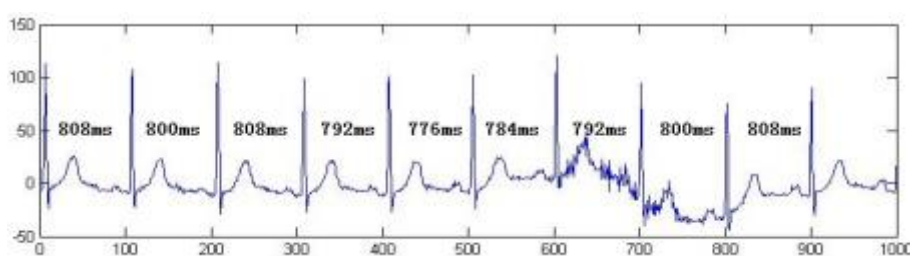


图 3-1 动态心电图

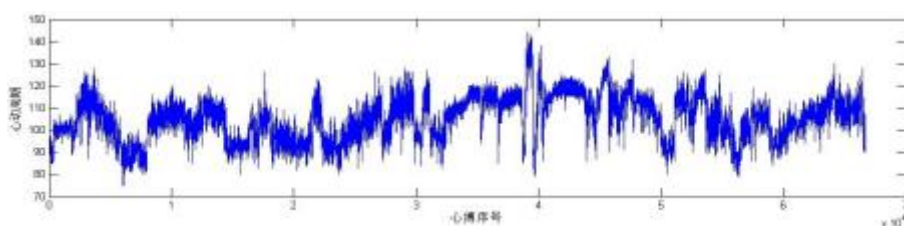


图 3-2 RR 间期序列图

步骤 2：检测并标记加（减）速周期：将每个心动周期值 (RR_i) 同该周期的前一个心动周期 (RR_{i-1}) 进行比较，确定该周期属于心率减速心动周期还是心率加速心动周期，再用不同的符号加以标记。如图 3-3 所示，比前一个心动周期大的心动周期被定义为减速周期，本文将其标注为“*”；比前一个心动周期小的心动周期被定义为加速周期，本文将其标注为“o”。

步骤 3：确定心率段：进行“位相整序”时所用的心率段是指以每一个减速点或加速点为心率段中心时，位于其两侧的心动周期各取多少个心动周期，而左右各

取多少心动周期的具体数值需要参考最低心率。例如，本文在进行心率减速力分析时，心率段数值定为 30 个间期，那么这就意味着以选定的减速点为中心时，其左右依次各取 15 个心动周期组成一个心率段(如图 3-4 所示)。

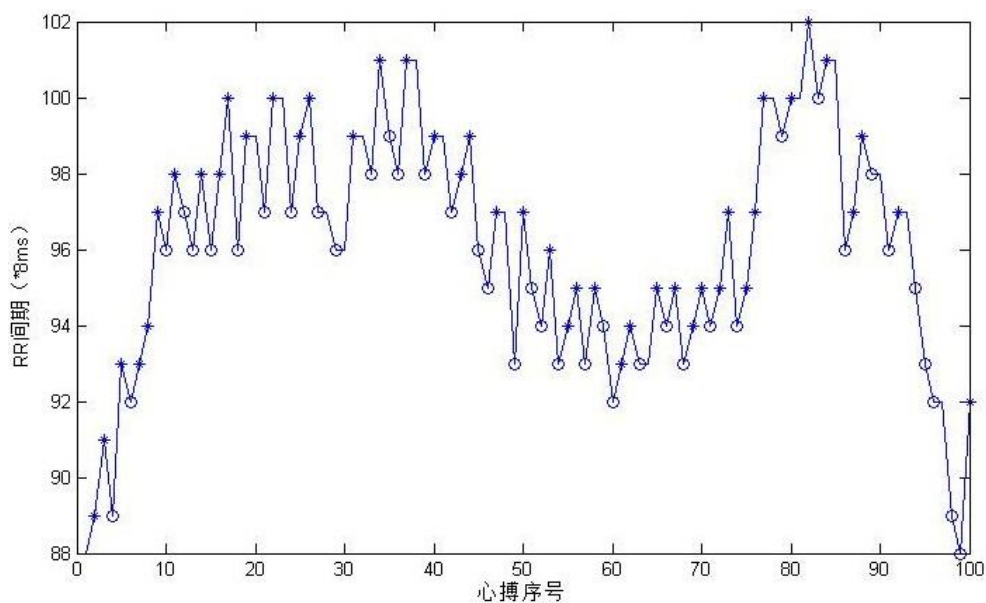


图 3-3 确定减速与加速周期并分别标注

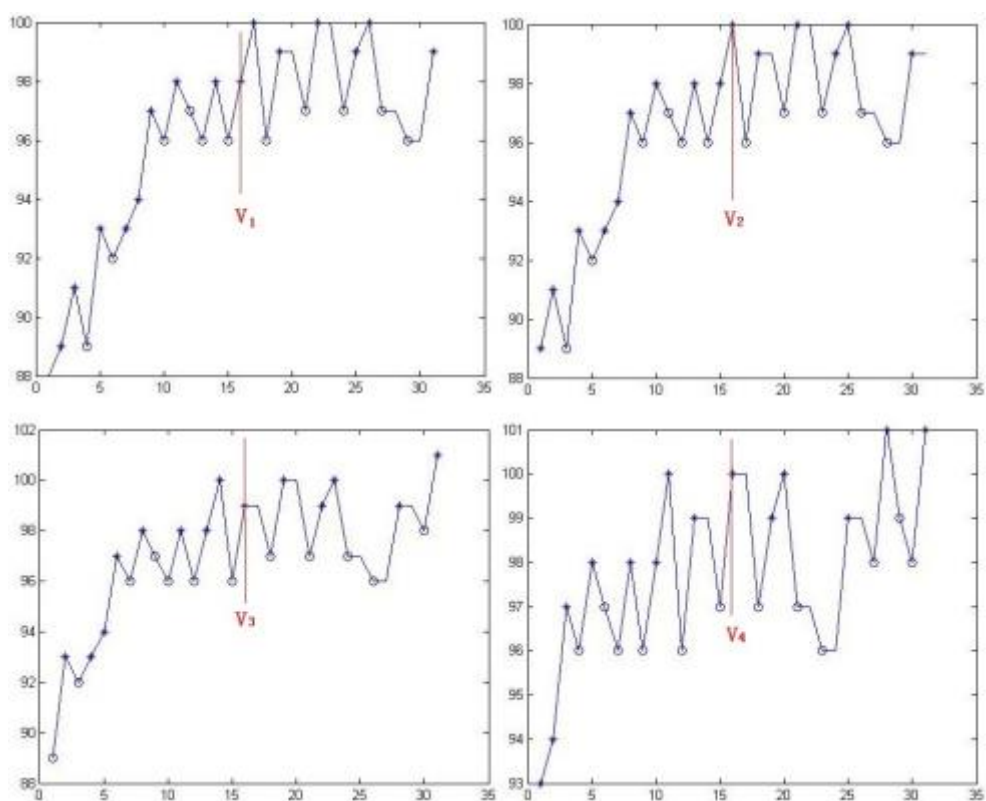


图 3-4 连续的四个心率段 V_1 、 V_2 、 V_3 、 V_4

步骤 4：位相整序：以入选的减速点（“*”点）为中心，进行不同心率段的叠加(如图 3-5 所示)。

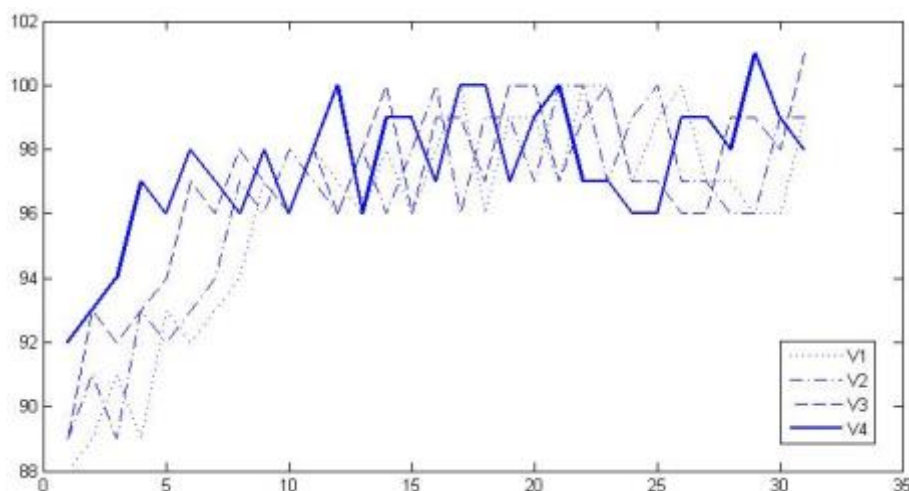


图 3-5 位相整序

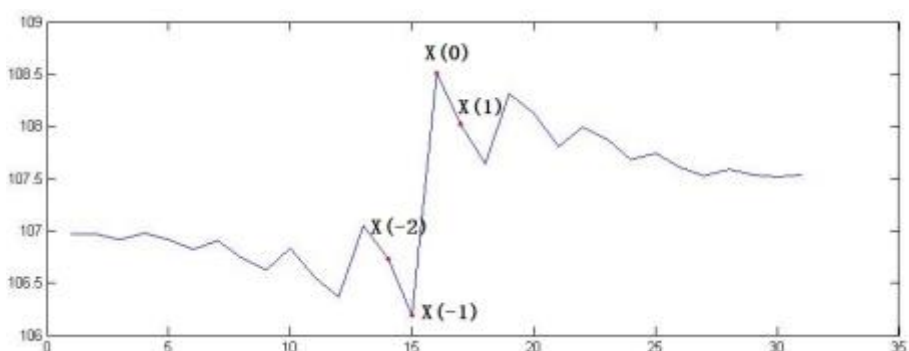


图 3-6 计算对应周期的平均值

步骤 5：计算对应序号的平均周期：经“位相整序”后，分别计算对应周期的平均值：① $X(0)$ ：所有中心点位置的 RR 间期的平均值；② $X(1)$ ：中心点右侧的第一个心动周期的平均值；③ $X(-1)$ ：中心点左侧的第一个心动周期的平均值；④ $X(-2)$ ：中心点左侧的第二个所有心动周期的平均值(图 2-9)。

步骤 6：计算心率减速力：分别计算 $X(0)$ 、 $X(1)$ 、 $X(-1)$ 、 $X(-2)$ 的均值后，再将结果代入式 (3-6) 进行计算即可得到心率减速力 DC ：

$$DC = \frac{X(0) + X(1) - X(-1) - X(-2)}{4} \quad (3-6)$$

注：步骤 4 和 5 中对上述信号处理过程时，已经同时标出了心率加速点（“o”点），并用同样流程，代入加速点和加速周期就可以计算出心率加速力 AC 相关的 4 个均值($Z(0)$ 、 $Z(1)$ 、 $Z(-1)$ 、 $Z(-2)$)，代入式 (3-7)，计算出受检者的心率加速力 AC 的值：

$$AC = \frac{Z(0) + Z(1) - Z(-1) - Z(-2)}{4} \quad (3-7)$$

3.4 窦性心律震荡分析中的特征提取

一次室性早搏的出现对于随后出现的窦性心律存在两种不同的影响^[16]：一种是在室性早搏后，窦性心律出现心跳先加速后减速的现象，这种典型的变化常见于正常人以及心肌梗死后猝死的低危患者。另一种是室性早搏后，窦性心律的 RR 间期无明显变化，这种现象常见于心肌梗死后猝死的高危患者。

窦性心律震荡可进行定量计算，目前常用的检测特征包括：震荡初始 (turbulence onset, TO)，震荡斜率 (turbulence slope, TS)^[15, 16]。

3.4.1 室性早搏的检测

室性早搏（或室性期前收缩），简称室早，是临床上很常见的心律失常现象，其发生人群相当广泛，包括心脏健康的人群和各种心脏病患者。室早的检测作为窦性心律震荡分析的基础环节，是正确预警心律失常发生的关键。本文针对室早在 DCG 中的四大特征进行了以下的检测：

1) 早搏所在的“ RR 间期”过长（见式 2-20）；

2) QRS 波群时限超过 $0.12s$ ：

$$(i_s - i_Q) \times \frac{1}{f} > 0.12s \quad (3-8)$$

式中 i_s ——为 QRS 波群 S 波的采样点位置；

i_Q ——为 QRS 波群 Q 波的采样点位置。

3) T 波倒置：

$$N(t_{top}) < \frac{N(t_{start}) + N(t_{end})}{2} \quad (3-9)$$

式中 t_{top} ——为 T 波波顶采样点；

t_{start} ——为 T 波起始采样点；

t_{end} ——为 T 波截至采样点（如图 3-7 所示）。

4) 代偿间歇（见图 3-7）小于等于前一个 RR 间期的二倍：

$$RR_{instead} \leq R_{-1} \times 2 \quad (3-10)$$

式中 $RR_{instead}$ ——为代偿间歇；

R_{-1} ——为前间期（如图 3-7 所示）。

3.4.2 震荡初始

TO 描述的是室性早搏后窦性心律是否存在加速的现象^[15, 16]。其计算公式是用室性早搏代偿间期后的前 2 个窦性心律的 RR 间期（也称后间期, R_1, R_2 ）的和，减去室性早搏偶联间期前的 2 个窦性心律的 RR 间期（也称前间期, R_{-1}, R_{-2} ）的和，两者之差再除以后者，所得的结果即为 TO ^[15, 16]，见式 3-8。

$$TO = \frac{(R_1 + R_2) - (R_{-1} + R_{-2})}{R_{-1} + R_{-2}} \times 100\% \quad (3-11)$$

式中 R_{-1} 、 R_{-2} 、 R_1 、 R_2 ——如图 3-7 所示。

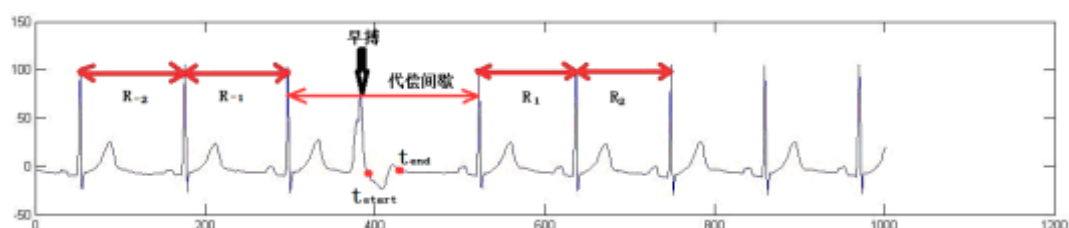


图 3-7 早搏，代偿间歇，前间期（ R_{-1} ， R_{-2} ）和后间期（ R_1 ， R_2 ）等

3.4.3 震荡斜率

震荡斜率（ TS ）是用于定量分析室性早搏后是否存在窦性心律减速现象^[15, 16]。

TS 的检测步骤如下：

步骤 1：检测室早出现后的前 20 个窦性心律的 RR 间期值，并把这些 RR 间期的值作为纵坐标，把 RR 间期的心搏序号作为横坐标，绘制 RR 间期值的序列图；

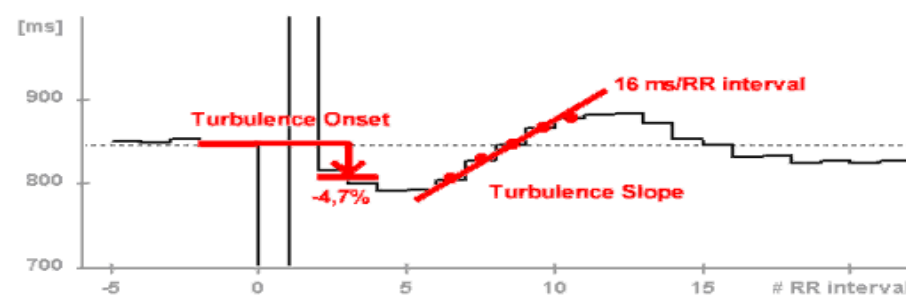


图 3-8 T_s 的计算示意图

步骤 2：在 RR 间期值的序列图中，对每 5 个连续的心搏序号的窦性心律点做出回归线，其中正向最大斜率即为 TS （如图 3-8 所示）

3.5 本章小结

本章针对心率变异性分析、心率减（加）速力分析以及窦性心律震荡分析中的一些特征进行了提取，提取了 $MEAN$ 、 $SDNN$ 、 $SDANN$ 、 $ASDNN$ 、 Ti 、 VLF 、 LF 、

HF、*TP*、*LF_HF*、*DC*、*AC*、*TO* 和 *TS* 共 14 项特征，为之后用于筛选出高危患者的模式分类提供了充足的备选特征。

第 4 章 心电样本的模式分类

4.1 引言

本章应用朴素贝叶斯分类器和 SVM 分类器，在有监督学习的模式下，对哈尔滨医科大学第一附属医院提供的心肌梗死患者的 Holter 数据样本进行分类实验，根据筛选出的每一项特征上每个识别对象所呈现出来的观察值将患者分为高危和低危两类。针对实验结果与回访情况（死亡或存活）的比对情况，绘制了接受者操作特性曲线，并通过一系列的分析，最终筛选出了分类效果较为优秀的三个分类模型。

4.2 分类前的特征选择

特征的选择是非常重要的，它强烈地影响到分类器的设计及其分类的性能。若对于不同类别样品而言，特征的差别非常大，那么就比较容易设计出具有较高性能的分类器。因此特征选择是模式识别分类器设计其中的一个至关重要的问题。那么，对于已经提取的 14 项特征所组成的样本特征库，本文首先进行的就是样本特征库的初步分析，即考察特征是否合理，能否达到模式分类的目的。

有关资料^[21]表明，心率变异性分析中的时域分析特征和频域分析特征有着很强的相关性，相关系数如表 4-1 所示。这必然会影响到某些基于统计概率的分类器（朴素贝叶斯）的分类性能。另有资料^[17]表明，R 波波峰的漏识会使心电信号中的高频分量明显增加进而造成心率变异性频域分析产生较大的误差。因此在模式分类之前，剔除了心率变异性分析中的频域特征。

表 4-1 时域特征与频域特征的相关系数

时域 特征	频域特征			
	TP	VLF	LF	HF
SDNN	0.96	0.87	0.72	0.67
SDANN	0.94	0.68	0.61	0.57
ASDNN	0.79	0.90	0.89	0.82

窦性心率震荡分析的两项特征 TO 和 TS 的计算依赖于室性早搏的发生，然而，在对样本的观察以及与医大医生的沟通过程中发现，并不是所有样本都会在心电图中出现室早的波形。因此在分类时，选择暂时不使用 TO 和 TS 这两项特征。

经过对特征的初步筛选，最终用于分类的 8 个指标为 $MEAN$ （RR 间期均值）、 $SDNN$ （RR 间期总体标准差）、 $SDANN$ （RR 间期均值标准差）、 $ASDNN$ （RR 间期

标准差均值)、 Ti (HRV 三角指数)、 DC (心率减速力)、 AC (心率加速力) 以及 MHR (平均心率, 由医大一院心内科提供)。

4.3 样本来源及筛选

本文实验所用的实验样本数据全部由哈尔滨医科大学附属第一医院心内科提供。所用 Holter 的采样频率为 128Hz, 采样时长为 24 小时。该样本集是从 2009 年三月到 2009 年十二月期间住院的 226 位心肌梗死患者的 Holter 记录。在这些患者当中, 男患者占 162 人, 女患者占 64 人。年龄分布如图 4-1 所示。

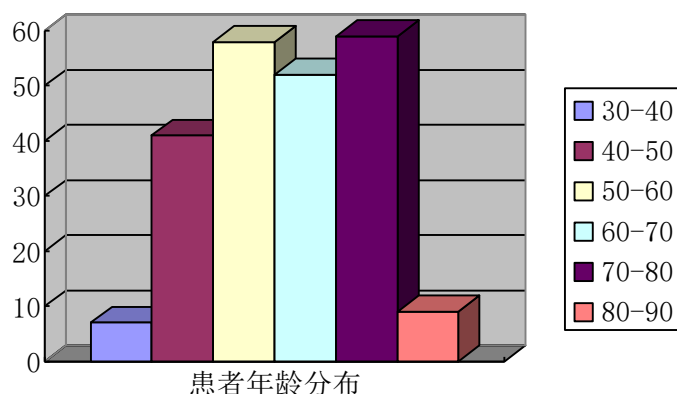


图 4-1 226 位心肌梗死患者的年龄分布图

在医大提供的 226 个病例样本当中, 6 例伴有房颤, 9 例与医大心内科心电图计算机 (含人机对话成分) 测定的数据偏差较大。根据医大医生的建议 (计算房颤患者的各项指标没有意义), 除去了这 6 个伴有房颤的样本; 根据指导老师的建议, 除去了误差较大的 9 例样本。另外, ID 为 102870 的患者死因不明, ID 为 103428 号患者失访, ID 为 104027 号患者死于非心源性疾病, 为使分类结果更为准确, 除去了这 3 例样本。最终使用的样本总数为 208。该样本集包括 12 个死亡病例, 196 个存活病例。

4.4 分类器原理

产生式模型和判别式模型作为两种主流的分类器模型, 被广泛应用在模式分类的实际应用当中。产生式模型 (Generative Model) 估计的是联合概率分布, 产生所有变量上的一个概率密度模型, 具有通用性、灵活性及清晰的层次结构等优点^[22]。判别式模型 (Discriminative Model) 可以看作是学习一个映射或函数的过程, 即对系统输入特征和输出变量进行建模, 并使映射或函数最优化^[22]。本文分别选择了产生式模型中和判别式模型中的两种典型分类器——朴素贝叶斯分类器

和 SVM 分类器，进行了分类器的设计与实现。

4.4.1 朴素贝叶斯分类器

朴素贝叶斯是一种非常流行的产生式分类方法，它巧妙利用了概率统计知识中的贝叶斯理论来解决模式识别中的分类问题。在许多问题中，朴素贝叶斯(Naive Bayes, NB)分类算法的分类效果甚至可以与决策树和神经网络等分类算法相媲美，该算法能够运用于针对大型数据库的海量数据处理的分类问题中，而且具有方法简单、分类准确率高、速度快等特点。

设第 i 个病例样本可表示为一个 8 维的列向量： $\mathbf{X}_i=(x_1, x_2, \dots, x_8)^T$ 。其中， $x_1=MEAN$ ， $x_2=SDNN$ ， $x_3=SDANN$ ， $x_4=ASDNN$ ， $x_5=Ti$ ， $x_6=DC$ ， $x_7=AC$ ， $x_8=MHR$ 根据贝叶斯公式有式 (4-1) 和式 (4-2)。

$$P(\omega_{\text{高危}} | \mathbf{X}_i) = \frac{P(\mathbf{X}_i | \omega_{\text{高危}}) \times P(\omega_{\text{高危}})}{P(\mathbf{X}_i)} \quad (4-1)$$

$$P(\omega_{\text{低危}} | \mathbf{X}_i) = \frac{P(\mathbf{X}_i | \omega_{\text{低危}}) \times P(\omega_{\text{低危}})}{P(\mathbf{X}_i)} \quad (4-2)$$

根据朴素贝叶斯条件独立假设有式 (4-3) 和式 (4-4)。

$$P(\omega_{\text{高危}} | \mathbf{X}_i) = \prod_{j=1}^8 P(x_j | \omega_{\text{高危}}) \quad (4-3)$$

$$P(\omega_{\text{低危}} | \mathbf{X}_i) = \prod_{j=1}^8 P(x_j | \omega_{\text{低危}}) \quad (4-4)$$

由式 (4-1)，(4-2) 和 (4-3) 有式 (4-5)。

$$\frac{P(\omega_{\text{高危}} | \mathbf{X}_i)}{P(\omega_{\text{低危}} | \mathbf{X}_i)} = \frac{P(\mathbf{X}_i | \omega_{\text{高危}}) \times P(\omega_{\text{高危}})}{P(\mathbf{X}_i | \omega_{\text{低危}}) \times P(\omega_{\text{低危}})} = \frac{\prod_{j=1}^8 P(x_j | \omega_{\text{高危}})}{\prod_{j=1}^8 P(x_j | \omega_{\text{低危}})} \times \frac{P(\omega_{\text{高危}})}{P(\omega_{\text{低危}})} \quad (4-5)$$

其中 $P(\omega_{\text{高危}})$ 和 $P(\omega_{\text{低危}})$ 可由训练样本估计出来。由于 x_1, x_2, \dots, x_8 均为连续特征，因此 $P(x_j | \omega_{\text{高危}})$ 和 $P(x_j | \omega_{\text{低危}})$ 可由高斯分布估计得出，见式 (4-6) 和式 (4-7)。

$$P(x_j | \omega_{\text{高危}}) = \frac{1}{\sigma_{j\text{高危}} \sqrt{2\pi}} \exp\left(-\frac{(x_j - \mu_{j\text{高危}})^2}{2\sigma_{j\text{高危}}^2}\right) \quad (4-6)$$

$$P(x_j | \omega_{\text{低危}}) = \frac{1}{\sigma_{j\text{低危}} \sqrt{2\pi}} \exp\left(-\frac{(x_j - \mu_{j\text{低危}})^2}{2\sigma_{j\text{低危}}^2}\right) \quad (4-7)$$

令 $J(\mathbf{X}_i) = \frac{P(\omega_{\text{高危}} | \mathbf{X}_i)}{P(\omega_{\text{低危}} | \mathbf{X}_i)}$ 。若 $J(\mathbf{X}_i)$ 大于或等于 1，那么第 i 个病例样本所对应的

患者为高危患者，应对其及时采取预防治疗措施，若 $J(\mathbf{X}_i)$ 小于 1，那么第 i 个病例样本所对应的患者为低危患者，见式（4-8）。

$$J(\mathbf{X}_i) \begin{cases} \geq 1 \\ < 1 \end{cases} \Rightarrow \mathbf{X}_i \in \begin{cases} \omega_{\text{高危}} \\ \omega_{\text{低危}} \end{cases} \quad (4-8)$$

4.4.2 SVM 分类器

支持向量机(SVM)是建立在统计学习理论的基础之上的一种较新的判别式分类方法，有相对较好的泛化能力，是目前最热门的分类方法之一。SVM 能够较好地解决样本数量较少、非线性、维数较高以及局部极小点等实际问题，并非常成功地应用于模式分类、函数逼近和时间序列预测等方面。

设 \mathbf{X} 为病例样本的 8 维心电图特征向量，而在 8 维心电图特征空间中能将样本集分为高危和低危两类的超平面的方程式为：

$$\mathbf{W}^T \mathbf{X} + b = 0 \quad (4-9)$$

式中 \mathbf{W}^T 为超平面的法向量。

设 y 为 \mathbf{X} 所属的类别，式（4-9）表示的超平面一边样本的 y 值为-1，超平面另一边的 y 值为 1。令分类函数为：

$$f(\mathbf{X}) = \mathbf{W}^T \mathbf{X} + b \quad (4-10)$$

不妨设 $f(\mathbf{X}) < 0$ 的 \mathbf{X} 对应的 y 为-1，为低危患者，而 $f(\mathbf{X}) > 0$ 的 \mathbf{X} 对应的 y 为 1，为高危患者。那么对于训练样本集 $\{(\mathbf{X}_1, y_1), (\mathbf{X}_2, y_2), \dots, (\mathbf{X}_N, y_N)\}$ ，归一化超平面方程需要满足：

$$y_i(\mathbf{W}^T \mathbf{X}_i + b) \geq 1 \quad (4-11)$$

在线性可分情况下，在结构风险最小化准则下的最优超平面问题，可以表示为如下的约束优化问题：

$$\begin{cases} \min & \frac{1}{2} \|\mathbf{W}\|^2 \\ \text{s.t.} & y_i(\mathbf{W}^T \mathbf{X}_i + b) \geq 1 \quad i = 1, 2, \dots, N \end{cases} \quad (4-12)$$

式（4-12）问题的最优解可以通过求解拉格朗日函数的鞍点得到，定义如下的 Lagrange 函数：

$$L(\mathbf{W}, b, \alpha) = \frac{1}{2} \|\mathbf{W}\|^2 - \sum_{i=1}^N \alpha_i [y_i(\mathbf{W}^T \mathbf{X}_i + b) - 1] \quad (4-13)$$

式中 α_i ——为各样本对应的 Lagrange 系数且 $\alpha_i \geq 0$ 。

求解式（4-13）的最小值，可以对其求 \mathbf{W} 和 b 的偏导，并令它们等于 0，这样就可以把上述求最优分类面的问题转化为较简单的对偶问题，其对偶问题见式

(4-14)。

$$\begin{cases} Q(\alpha) = \sum_{k=1}^N \alpha_k - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j < \mathbf{X}_i, \mathbf{X}_j > \\ s.t. \quad \sum_{i=1}^N \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad i = 1, 2, \dots, N \end{cases} \quad (4-14)$$

对于这个不等式约束下二次函数寻优的问题，存在唯一解。以上优化问题的最优解为： $\alpha^* = (\alpha_1^*, \dots, \alpha_N^*)^T$ 这样，计算得到：

$$\mathbf{W}^* = \sum_{i=1}^N \alpha_i y_i \mathbf{X}_i \quad b^* = \frac{1}{2} < \mathbf{W}^*, \mathbf{X}_r + \mathbf{X}_s > \quad (4-15)$$

式中 \mathbf{X}_r 和 \mathbf{X}_s ——是两类中任意的支持向量。

对于测试样本 \mathbf{X}_t 代入式 (4-10) 得出 $f(\mathbf{X}_t)$ ，若 $f(\mathbf{X}_t) < 0$ ，那么该病例样本所对应的患者为低危患者；若 $f(\mathbf{X}_t) > 0$ ，那么该病例样本所对应的患者为高危患者，见式 (4-16)。

$$f(\mathbf{X}_t) \begin{cases} > 0 \\ < 0 \end{cases} \Rightarrow \mathbf{X}_t \in \begin{cases} \omega_{\text{高危}} \\ \omega_{\text{低危}} \end{cases} \quad (4-16)$$

4.5 实验结果及分析

为克服训练样本数量较小的缺点，这里采用了留一法 (leave-one-out) 对分类效果进行训练和评估。由于所有样本均为心肌梗死患者，因此将这些患者的回访记录 (死亡或存活) 作为本次实验的参考答案。为科学准确地评价猝死预警系统的分类性能，采用精确性 (AR)、敏感性 (TPR)、特殊性 (SPE) 这 3 项指标来进行评价，见式 (5-1)，(5-2)，(5-3)。

$$AR = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (4-17)$$

$$TPR = \frac{TP}{TP + FN} \times 100\% \quad (4-18)$$

$$SPE = \frac{TN}{TN + FP} \times 100\% \quad (4-19)$$

式中 TP ——为被正确分类的死亡样本的数量；

TN ——为被正确分类的存活样本的数量；

FP ——为被错误分类的存活样本的数量；

FN ——为被错误分类的死亡样本的数量。

AR 、 TPR 、 SPE 这三个评价指标的值越高，说明该分类系统的分类效果越好。当 AR 、 TPR 、 SPE 都接近 100% 时，说明这个分类系统接近完美。

通过观察和分析发现 MHR 和 $MEAN$ 两项指标的可分性较差，于是决定用去掉

MHR 和 *MEAN* 两项指标后的 6 维心电特征做了对比实验，同时针对心率变异性分析和心率减速力分析，加入单独用心率变异性特征以及单独用心率减速力特征进行的分类结果的对比数据。实验结果如表 4-2 所示。

表 4-2 分类结果

分类器	分类结果	<i>AR</i> /%	<i>TPR</i> /%	<i>SPE</i> /%
朴素贝叶斯 分类结果	8 维	84.13	66.67	85.20
	6 维	84.13	75.00	84.69
	HRV	83.65	75.00	84.18
	心率减速力	88.46	16.67	92.86
SVM 分类结果	8 维	79.33	83.33	79.08
	6 维	79.33	91.67	78.57
	HRV	79.81	91.67	79.08
	心率减速力	67.79	75.00	67.35

表中“8 维”代表用 *MEAN*、*SDNN*、*SDANN*、*ASDNN*、*Ti*、*DC*、*AC* 以及 *MHR* 这 8 项特征分类的分类结果；“6 维”代表用去掉 8 维心电特征中的 *MHR* 和 *MEAN* 两项的 6 项特征分类的分类结果；“HRV”代表用 *MEAN*、*SDNN*、*SDANN*、*ASDNN*、*Ti* 这 5 项心率变异性分析特征分类的分类结果；“心率减速力”代表用 *DC* 和 *AC* 两项特征分类的分类结果。

接受者操作特性曲线（ROC 曲线）是一种利用构图法揭示敏感性和特异性的相互关系的方法，它能够反映敏感性和特异性的连续变化规律。它通过将连续变量设定出多个不同的临界值，从而计算出一系列敏感性和特异性，再以敏感性为纵坐标、负正类率（1-特异性）为横坐标绘制成曲线。猝死预警分类系统的性能可以通过 ROC 曲线下的面积（Area Under the ROC Curve, *AUC*）来衡量：面积为 1（*AUC* = 1）代表一个完美的分类系统；面积为 0.5（*AUC* = 0.5）代表一个没有价值的分类系统。对一个分类系统的粗略评价如下：

$$AUC \in (0.9, 1) \Rightarrow \text{excellent}(A) \quad (4-20)$$

$$AUC \in (0.8, 0.9) \Rightarrow \text{good}(B) \quad (4-21)$$

$$AUC \in (0.7, 0.8) \Rightarrow \text{fair}(C) \quad (4-22)$$

$$AUC \in (0.6, 0.7) \Rightarrow \text{poor}(D) \quad (4-23)$$

$$AUC \in (0.5, 0.6) \Rightarrow \text{fail}(F) \quad (4-24)$$

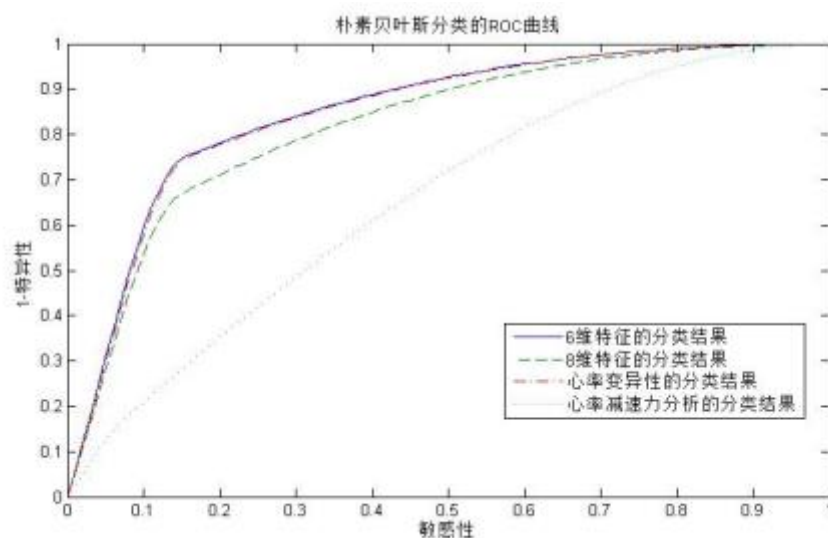


图 4-2 朴素贝叶斯分类的 ROC 曲线

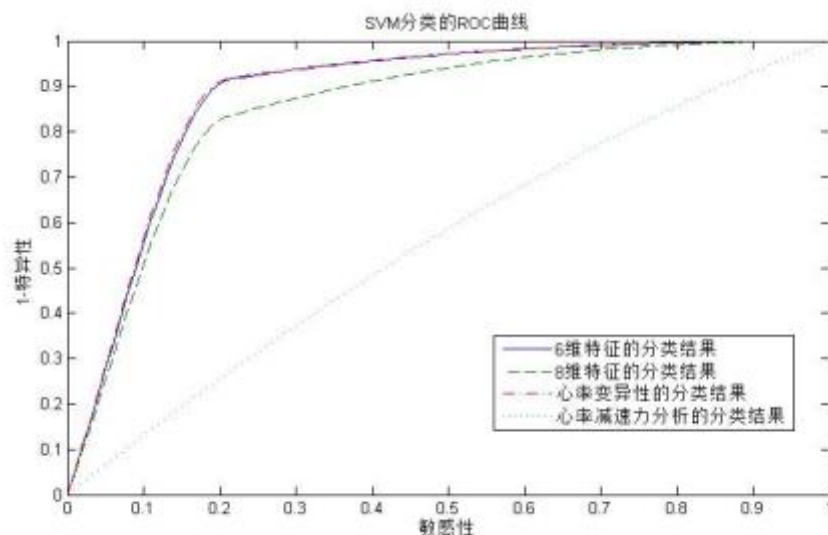


图 4-3 SVM 分类的 ROC 曲线

朴素贝叶斯和 SVM 分类的 ROC 曲线如图 4-2 和 4-3 所示。

各分类的 AUC 值如表 4-3 所示。

通过对表 4-2 和表 4-3 中分类结果的分析可以得出以下结论：

- 1) 朴素贝叶斯分类器分类的精确性要好于 SVM 分类器；
- 2) SVM 分类器分类的敏感性要好于朴素贝叶斯分类器；
- 3) 朴素贝叶斯分类器分类的特异性要好于 SVM 分类器；
- 4) 对朴素贝叶斯分类器而言，6 维心电图特征的 AUC 值最高，而对 SVM 分类器而言，用 HRV 分析特征的 AUC 值最高。

表 4-3 各种分类的 AUC 值

分类器	分类结果	AUC
朴素贝叶斯 分类结果	8 维	0.8262
	6 维	0.8548
	HRV	0.8521
	心率减速力	0.6598
SVM 分类结果	8 维	0.8579
	6 维	0.8880
	HRV	0.8902
	心率减速力	0.5627

经过分析，造成这种结果的原因如下：

1) 样本中的死亡病例过少，极大地影响了超平面的分类效果，以至于 SVM 的特异性较差，从而影响了 SVM 的精确性；

2) 用于分类的 8 维心电特征中，某些特征之间存在隐藏的相关性，影响了朴素贝叶斯的敏感性；

3) 某些心电特征如 *MHR* 和 *MEAN* 等对于高危患者和低危患者的可分性较差，造成低维（6 维和 5 维）心电特征的分类效果优于全部心电特征（8 维）的分类效果。

经过综合比较发现，应用 6 维心电特征的朴素贝叶斯分类器（ $AUC = 0.8548$ ）、应用 6 维的 SVM 分类器（ $AUC = 0.8880$ ）以及应用 HRV 分析特征的 SVM 分类器（ $AUC = 0.8902$ ）的分类效果比较优秀，可以作为扩充大量样本后的备选分类模型。

4.6 本章小结

本章针对已经提取的 14 项特征在相关性和可行性方面进行了初步的筛选，选取了 *MEAN*（RR 间期均值）、*SDNN*（RR 间期总体标准差）、*SDANN*（RR 间期均值标准差）、*ASDNN*（RR 间期标准差均值）、*Ti*（HRV 三角指数）、*DC*（心率减速力）、*AC*（心率加速力）以及 *MHR*（平均心率）这 8 维心电特征作为猝死预警模式分类的基本特征。针对这 8 维心电特征，分别选择了产生式模型和判别式模型中两种典型的分类器——朴素贝叶斯分类器和 SVM 分类器，完成对 208 个病例样本的留一法的分类实验。将分类的结果同回访记录（死亡或存活）进行了比对，绘制了用于评价分类系统的 ROC 曲线，计算了 AUC 的值。经对比和分析，筛选出了三个分类表现较好的三个分类系统——应用 6 维心电特征的朴素贝叶斯分类器、应用 6 维的 SVM 分类器和应用 HRV 分析特征的 SVM 分类器。

结 论

本文提出了一种基于 Holter 的心肌梗死猝死预警系统，该系统在 R 波波峰的识别部分创新性地提出并应用了一种基于 DCG 心电信号的 R 波检测新算法，该算法对 R 波的识别率高于最大值双重搜索法以及差分运算法；在特征提取方面，该系统利用得到的 *RR* 间期序列提取了心率变异性分析、心率减速力分析以及窦性心律震荡分析的多项特征，经过对特征和样本的筛选，确定了可用于进行模式分类实验的特征和样本；最终应用朴素贝叶斯和 SVM 分类器对 208 个病例样本进行了留一法的分类实验，心肌梗死病例样本被分为了高危和低危两类。

在分析分类结果时，通过与病人的回访记录（死亡或存活）进行的比对，将 SCD 高危患者从普通患者中筛选出来。绘制了能够评价分类系统性能的 ROC 曲线，计算了 *AUC* 的值，通过对各种分类结果的对比和分析，得到了 3 个分类表现良好（ $AUC > 0.8$ ）的分类模型——应用 6 维心电特征的朴素贝叶斯分类器、应用 6 维的 SVM 分类器和应用 HRV 分析特征的 SVM 分类器。

经过实验和分析得出的这 3 个猝死预警模式分类模型，为以后的样本扩充后的分类模型提供了备选方案。同时也为融合朴素贝叶斯和 SVM 分类器的融合提供了参考。

参考文献

- [1] 曹克将. 室性心律失常与心脏性猝死[J]. 中华心律失常杂志, 2009, 13 (2): 85-86.
- [2] 秦培强, 王凤秀, 贾邢倩, 马伟. 预测心脏性猝死的临床心电综合征[J]. 心血管病防治知识 (学术版), 2012 (1): 26-28.
- [3] 常俊. 远离心脏性猝死[J]. 健康指南: 中老年, 2012 (2): 12-13.
- [4] 石武献. 心源性猝死的预防[J]. 中华临床医学研究杂志, 2007, 13 (17): 2579-2581.
- [5] 高修仁, 王景峰, 陈泗林, 林宇. 猝死日记[J]. 家庭药师, 2010 (5): 22-30.
- [6] 王茜, 张正国. 心肌梗死后心脏性猝死的无创危险预测[J]. 国际生物医学工程杂志, 2006, 29 (5): 272-276.
- [7] 宋华静, 李金国, 周长勇, 李洁, 马丽华. 378 例心律失常患者动态心电图分析[J]. 社区医学杂志, 2006, 4 (3): 86-87.
- [8] 闻怀臣, 吴利平, 韩桂兰, 卢喜烈. 动态心电图 1500 例心律失常结果分析[J]. 实用心电学杂志, 2006, 15(1): <http://journal.9med.net/qikan/article.php?id=3095> 28.
- [9] 曲秀芬, 刘丽晓, 王晓峰, 刘莹, 纪树彦, 郭晖, 苏亚芬, 朴晶燕, 张树生, 黄永麟. 美托洛尔和培哚普利对心肌梗死患者自主神经失衡和预后的影响[J]. 中华心血管病杂志, 2001, 29 (6): 344-347.
- [10] 王步青, 王卫东. 心率变异性分析方法的研究进展[J]. 北京生物医学工程, 2007, 26(5): 552-554.
- [11] Axel Bauer, Jan W Kantelhardt, Petra Barthel, Raphael Schneider, Timo Mikallio, Kurt Ulm, Katerina Hnatkova, Albert Schig, Heikki Huikuri, Armin Bunde, Marek Malik, Georg Schmidt. Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: cohort study[J]. Lancet, 2006, 367: 1674-1681.
- [12] Georg Schmidt, Marek Malik, Petra Barthel, Raphael Schneider, Kurt Ulm, Linda Rolnitzky, A John Camm, J Thomas Bigger Jr, Albert Schömig. Heart-rate turbulence after ventricular premature beats as a predictor of mortality after acute myocardial infarction[J]. Lancet, 1999, 353 (9162): 1390-1396.

- [13] Richard L. VeRRier, Thomas Klengenheben, Marek Malik, Nabil El-Sherif, Derek V. Exner, Stefan H. Hohnloser, Takanori Ikeda, Juan Pablo Martı´nez, Sanjiv M. Narayan, Tuomo Nieminen, David S. Rosenbaum. Microvolt T-Wave Alternans[J]. Journal of the American College of Cardiology, 2011, 58 (13) : 1309-1324.
- [14] 郭继鸿, 心率减速力检测[J]. 临床心电学杂志, 2009, 18 (1): 60-68.
- [15] 刘福, 赵瑞平. 窦性心律震荡的临床研究进展[J]. 心血管病学进展, 2011, 32 (2): 260-262.
- [16] 张德强, 吴永全, 黄捷英. 窦性心率振荡现象[J]. 中华心血管病杂志, 2005, 33 (3): 290-292.
- [17] 张开滋等. 临床心电信息学[M]. 长沙: 湖南科学技术出版社, 2002: 54-63.
- [18] Yeh Y, Wang W. QRS complexes detection for ECG signal: the difference operation method[J]. Comput Meth Prog Bio, 2008, 91 (3): 245-254.
- [19] Joseph E.Mietus B.S., Ary L.Golberger M.D.. Heart Rate Variability Analysis with the HRV Toolkit[EB/OL]. (2008-6-21) [2012-6-6]
<http://www.physionet.org/tutorials/hrv-toolkit/>.
- [20] 周建兴. MATLAB 从入门到精通[M]. 北京: 人民邮电出版社, 2008: 200-201
- [21] 郭继鸿, 张萍, 程贤德, 雷斌, 屈建石. 动态心电图对心律失常的分析[M]. 动态心电学, 第4章第1节
<http://www.ceheart.com/Article/ShowArticle.asp?ArticleID=203&Page=3>.
- [22] 孔环环, 石洪波. 产生式与判别式分类方法比较研究[J]. 科技情报开发与经济, 2010, 20 (8): 78-80.

致 谢

本文是在黄剑华教授和张英涛副教授的精心指导和大力支持下完成的。黄老师以渊博的知识、开阔的视野和敏锐的思维给了我深深的启迪。张老师以其严谨求实的治学态度、高度的敬业精神、兢兢业业、孜孜以求的工作作风和大胆创新的进取精神对我产生重要影响。他们的言传身教将使我终生受益。

另外，我还要特别感谢鲜敏学长和史健婷师姐对我实验以及论文写作的指导，他们为我完成这篇论文提供了巨大的帮助。

最后感谢哈尔滨医科大学第一附属医院心内科的宋涛博士给本文的实验提供的心电样本和资料。

本课题承蒙以下基金的资助，特此致谢：

（1）国家自然科学基金：《多模态乳腺超声影像自适应特征融合模型研究》批准号：61100097；

（2）哈尔滨工业大学理工医交叉学科基础研究基金；