

基于小波变换和粗糙集的早搏信号识别算法

唐 孝^{1,2} 舒 兰¹ 郑 伟³

(电子科技大学数学科学学院 成都 611731)¹ (四川师范大学数学与软件科学学院 成都 610068)²
(四川省医学科学院心内科 成都 610072)³

摘 要 心电特征参数的选择和提取是心电图(ECG)分析的基础,提升检测算法的识别率和特征分类的精度是自动分析技术的关键。提出了基于小波变换和属性约简的心电早搏信号识别算法。该算法首先依据心血管专家的诊断标准选择了 12 个心电特征参数;然后运用基于小波变换的特征检测算法进行了特征提取,并利用基于粒计算的属性约简算法对特征参数进行了属性约简;最后,将约简后的数据用于模式分类并通过 MIT-BIH 数据库对结果进行验证。实验表明,约简后的分类精度大大高于约简前的数据,特征参数的合理选择(约简)是提高识别效率的重要因素。

关键词 心电图,小波变换,特征提取,属性约简,粒计算

中图法分类号 TP274 文献标识码 A

Premature Beat Signal Recognition Algorithm Based on Wavelet Transform and Rough Set

TANG Xiao^{1,2} SHU Lan¹ ZHENG Wei³

(School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China)¹

(College of Mathematics and Software Science, Sichuan Normal University, Chengdu 610068, China)²

(Department of Cardiology, Sichuan Academy of Medical Sciences, Chengdu 610072, China)³

Abstract The selection and extraction of electrocardiogram feature parameter are the base of the analysis of electrocardiogram(ECG). To improve recognition rate of detection algorithm and classification accuracy is the key to automatic analysis technology. Thus, a hybrid algorithm based on wavelet transform(WT) and attribute reduction of granular computing(GC) to detect premature beat signal of electrocardiogram(ECG) was present. At first, 12 electrocardiogram feature parameters are chosen based on diagnostic criteria from cardiovascular experts. Then the feature detection algorithm based on wavelet transform is used for feature extraction, and an attribute reduction algorithm based on granular computing is also used for attribute reduction. Finally, the data are put into pattern classification and the result is verified by MIT-BIH database. As the experiment shows, the classification accuracy after reduction is much higher than it before reduction. Therefore, that reasonable selection of feature parameter is an important factor to improve the recognition efficiency was justified in this article.

Keywords ECG, WT, Feature extraction, Attribute reduction, Granular computing

1 引言

心电图特征参数的定义与选择是心电信号自动分析技术的重要环节。心电图特征参数的准确定义是客观、全面地反映心电信号中疾病信息的基础。从理论上讲,掌握的特征信息越多,越能够充分反映心电图中的疾病信息,从而诊断的可靠性就越高。

由于心电图的时频域特性非常符合医师们的思维习惯,因此 ECG 信号各波形特征:波形角度、面积、幅度值、间期等信息常被作为心电信号自动分析的特征参数。1990 年, P. Trahanian^[1] 等人选取正向峰、负向峰、曲线段和直线段 4 个特征定义了属性字符串来识别 QRS 波形;1995 年, A. Koski 等人^[2] 则利用近似直线段作为基元,重新定义了基元字符串。

而后威建中等人^[3] 利用两个特征量:QRS 波宽度和 RR 间期,定义了 12 种心律失常信号。由于 ECG 信号复杂多变,仅靠时频域特征来描述心电信号是显然不够的,要想尽可能准确、客观地描述心脏活动的相关信息,必须从多角度对 ECG 信号进行描述和提取。M. Engin^[4] 对心电信号进行了离散小波变换后,选取了离散小波变换各尺度参数、AR 模型参数、三阶累积量作为心电参数。季虎等人^[5] 定义和提取了心律和 QRS 波群形态的 26 个特征参数。M. M. Tantawi 等人^[6,7] 利用小波变换和高阶统计量定义了 P 波、T 波、R 波的幅度值, QRS 波群起止时间值、S-T 段和 RR、PQ、QT 间期等 28 个特征来描述心电图。在 M. M. Tantawi 的基础上,我们也曾试着在特征参数的定义上下功夫:拆分了间期时间特征,增加了波形面积特征,定义了 32 个特征参数^[8],以求尽可能完整地

本文受四川师范大学科研项目(13KYL15),国家自然科学基金(61203285)资助。

唐 孝(1981—),男,博士生,讲师,主要研究方向为不确定性分析、数据挖掘、模式识别;舒 兰(1965—),女,教授,博士生导师,主要研究方向为不确定性分析、数据挖掘、信号处理。

刻画出所有心电信号信息。

然而在临床诊断和计算机自动分析中,却不是如此。对于临床诊断来讲,参数越多越容易影响医生的判断,反而会掩盖心脏病变信息;对于计算机自动分析来讲,参数越多分类器的设计就会越复杂,反而会影影响分类效率。换言之,特征选择的目的是从众多的参数信息中找出能准确代表分类信息且数量较少的特征参数集,从而提高分类识别的时间和准确度。粗糙集理论是进行特征选择的有效工具,它在分析和处理数据时,具有无需提供任何先验信息的优点。本文首先根据心电专家的建议选用了临床诊断中常用的12个特征作为参考指标,即P波宽度、P-R间期、P-R段、QRS宽度、S-T段、Q-T间期、T波宽度、P波振幅、Q波振幅、R波振幅、S波振幅和T波振幅;然后利用基于小波变换的特征检测算法对信号进行了特征提取,再利用基于粒计算的属性约简方法对参数特征进行了选择。分类实验表明,该算法具有较好的分类效果和训练速度。

2 基本原理

2.1 小波变换原理

对于非平稳信号 $f(x)$ 的连续小波变换有如下定义:

$$W_s f(x) = \langle f(x), \psi_s(x) \rangle = \frac{1}{s} \int_{-\infty}^{+\infty} f(t) \psi\left(\frac{x-t}{s}\right) dt$$

式中, s 为小波尺度, $\psi_s(x) = \frac{1}{s} \psi\left(\frac{x}{s}\right)$ 为母小波 $\psi(x)$ 在尺度 s 上的伸缩。在实际应用中,若令 $s=2^j$ ($j \in \mathbb{Z}$), 则可以将连续小波二进制离散化, 这样的小波变换 $W_{2^j} f(x)$ 称为信号 $f(x)$ 的二进小波变换。数字信号 $f(n)$ 的二进小波变换可用 Mallat 算法^[5,9]求得:

$$S_{2^j} f(n) = \sum_{k \in \mathbb{Z}} h_k S_{2^{j-1}} f(n-2^{j-1}k)$$

$$W_{2^j} f(n) = \sum_{k \in \mathbb{Z}} g_k S_{2^{j-1}} f(n-2^{j-1}k)$$

其中, S_{2^j} 为平滑算子, $S_{2^0} f(n) = d_n$ 即为待处理的数字信号。 $W_{2^j} f(x)$ 为数字信号 $f(x)$ 在尺度 j 下的二进小波变换。 $\{h_k | k \in \mathbb{Z}\}$ 和 $\{g_k | k \in \mathbb{Z}\}$ 分别为低通滤波器 $H(\omega)$ 和高通滤波器 $G(\omega)$ 的系数, 即

$$\begin{cases} H(\omega) = \sum_{k \in \mathbb{Z}} h_k e^{-ik\omega} \\ G(\omega) = \sum_{k \in \mathbb{Z}} g_k e^{-ik\omega} \end{cases}$$

2.2 基于小波变换的波形检测算法

采用二次样条小波作为基函数将去除了基线漂移的心电信号进行 Mallat 多尺度分解^[5,9], 根据小波变换的模极大值原理, 在 QRS 波能量大多聚集的且受噪声影响较小的 2^3 和 2^4 尺度上进行 R 波的检测与定位。类似地, P 波、T 波能量主要集中在 2^4 和 2^5 这两个尺度的小波子带信号上, 考虑到在 2^5 尺度上受机电噪声、基线漂移等噪声的影响较大, 选择在 2^4 尺度上进行检测。而 Q 波和 S 波都是低幅、高频信号, 其能量主要集中在 2^1 和 2^2 两个尺度上, 这里选用在 2^1 上检测 Q 波和 S 波。

详细步骤如下:

Step 1(多尺度分解) 利用二阶样条小波和 Mallat 算法对 ECG 信号进行多尺度的分解, 并求得 2^3 和 2^4 尺度下的小波模极大值 $W_{2^j} f(n)$ ($j=2,3$)。

Step 2(极值点检测) 分别在 2^3 和 2^4 尺度下的小波子带信号上取平均模极大值的 α 倍为阈值 ϵ 。若幅值的绝对值超过 ϵ , 则该点是符合条件的极值点, 记为 $\{E^k | k=1, 2, \dots, n\}$; 若有多个符合条件的点, 则选取幅值最大的点; 若没有找到符合条件的点, 则调整阈值参数 α , 重复该步骤。

Step 3(QRS 波定位) 保留成对的正负模极大值对, 消除不成对的极值点。在一个心动周期内, 找出两个最相邻的正负模极大值点构成的极值对, 将两点中心位置记为 QRS 波位置, 记作 $\{E_{QRS}^k | k=1, 2, \dots, m, m \leq n\}$ 。

Step 4(R 波检测) 在一个心动周期内, 找出 E_{QRS}^k 点附近两侧幅值绝对值最大的点(波峰或者波谷), 记为 R 波峰值点, 记作 E_r 。计算所检测 ECG 信号中极大幅值的平均值, 取该平均值的 β 倍为阈值 ξ 。若 E_r 点幅值的绝对值大于阈值 ξ , 则该点为 R 波峰值点; 若 E_r 点幅值的绝对值小于阈值 ξ , 则该点不是 R 波峰值点, 消去, 调整阈值参数 β , 重复该步骤。

Step 5(Q 波、S 波检测) 在 R 波峰值点前、后分别找出与之最近的一个模极小值点, 这两个点分别对应 Q 波和 S 波的峰值点位置。那么 Q 波极值点所在波的起点就是 QRS 波群的起点, S 波极值点所在波的终点就是 QRS 波群的终点。

Step 6(P 波、T 波检测) 在 2^4 尺度上, P 波在 QRS 波群之前也会产生一对模极大值点, 以 R 波峰点为基准, 在其前的一小段时间内便可找到 P 波的起点和终点。与之类似, 可在 QRS 波群之后找出 T 波的起点和终点。

Step 7(计算间期、幅值) 一个典型心电信号的心动周期包含 QRS 波群、P 波、T 波、U 波等一系列的波形, 如图 1 所示, P 波的宽度为 $P_{off} - P_{on}$, P-R 间期值为 $Q_{on} - P_{on}$, P-R 段的长度为 $Q_{on} - P_{off}$, QRS 波宽度为 $S_{off} - Q_{on}$, S-T 段的长度为 $T_{on} - S_{off}$, Q-T 间期为 $T_{off} - Q_{on}$, T 波宽度为 $T_{off} - T_{on}$; 各波形的幅度值等于其峰值点幅值。

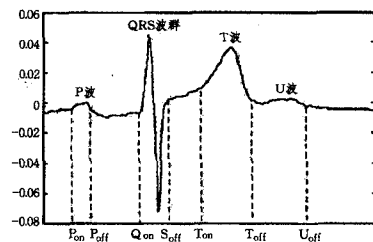


图 1 波形特征点

2.3 粗糙集理论

2.3.1 粒计算基本概念

定义 1^[10] 设 $K = \langle U, R \rangle$ 是一个知识库, $P \in R$ 是论域 U 上的等价关系, 称为知识。我们把知识 $P \in R$ 的粒度记作 $GD(P)$:

$$GD(P) = \frac{|P|}{|U \times U|} = \frac{|P|}{|U|^2}$$

其中, $|P|$ 表示 $P \subseteq U \times U$ 的基数。

定义 2^[11] 设 $S = \langle U, A, V, f \rangle$ 是一个信息系统, 属性集 A 中的属性 a 的重要度表示为: $Sig_{A-\{a\}}(a)$, 定义为:

$$Sig_{A-\{a\}}(a) = GD(A - \{a\}) - GD(A)$$

定义 3^[11] 设 $S = \langle U, A, V, f \rangle$ 是一个信息系统, C 是属性集 A 中的子集, $C \subseteq A$, $\forall a \in A - C$ 对于属性集 C 的重要度

表示为 $Sig_C(a)$, 定义为:

$$Sig_C(a) = GD(C) - GD(C \cup \{a\})$$

定义 3 表明, 属性 a 对于属性集 C 的重要性可以通过在 C 中添加 a 后所引起的知识粒度的变化来度量, 变化越大, 认为 a 对于 C 越重要。

定义 4^[11] 设 $S = (U, A, V, f)$ 是一个信息系统, $a \in A$, 如果 $GD(A - \{a\}) = GD(A)$, 则称 a 在 A 中是不必要的(可省去的); 否则 a 在 A 中是必要的(不可省去的)。如果每个 $a \in A$ 都为 A 中必要的, 则称 A 为独立的, 否则称 A 为依赖的。

定义 5^[11] 设 $P \subseteq A$, 若 P 是独立的, 且 $GD(P) = GD(A)$, 则称 P 是 A 中的一个约简, A 的所有约简记为 $red(A)$; A 中所有必要属性组成的集合称为 A 的核, 记为 $Core(A)$, $Core(A) = \bigcap red(A)$ 。

性质 1 属性 a 在 A 中是必要的(不可省去的), 当且仅当 $Sig_{A-\{a\}}(a) > 0$ 。

性质 2 $Core(A) = \bigcup \{a \in A \mid Sig_{A-\{a\}}(a) > 0\}$ 。

基于粒计算的属性约简方法的思路是首先计算出系统的约简核, 然后计算剩余属性对于核的重要度 $Sig_{Core(A)}(a)$, 取重要度中最大值与核组成系统的最终约简 $red(A)$ ^[12]。然而在实际应用中, 系统的约简往往不是唯一的, 而且有些系统是没有约简核的。对于没有约简核的系统, 基于核的属性重要度的约简方法就无能为力了^[13]。

2.3.2 基于粒计算的属性约简改进算法

在一个决策信息系统 $S = (U, A, V, f)$ 中, 若 $Sig_{A-\{a\}}(a) = 0$, 则表示属性 a 在 A 中是不必要的(可省去的)。若对于 $\forall a \in A$, 都有 $Sig_{A-\{a\}}(a) = 0$, 则系统不存在约简核。

定义 6 设 $S = (U, A, V, f)$ 是一个没有约简核的信息系统, 计算 A 中属性子集 P 的重要度, 定义 $red'(A) = \{P \mid \max Sig_{A-P}(P), P \subseteq A\}$ 为系统的次优约简。

性质 3 若 $GD(red'(A))$ 与 $GD(A)$ 相等, 则 $red(A) = red'(A)$ 。

在性质 3 中, 若 $GD(red'(A))$ 与 $GD(A)$ 不等, 则需要计算 $B = A - red'(A)$ 中每一个属性 b 对于 $red'(A)$ 的重要度, 取 $\max_{b \in B} Sig_{red'(A)}(b)$ 的属性与 $red'(A)$ 合并, 然后再次讨论其粒度的变化。由此, 可以得到下列算法。

算法描述:

输入: 信息系统 $S = (U, A, V, f)$, 其中 $U = \{x_1, x_2, \dots, x_n\}$, $A = \{a_1, a_2, \dots, a_m\}$;

输出: 该信息系统的最小约简 $red(A)$ 和核 $Core(A)$ 。

Step 1 计算属性集 A 的知识粒度 $GD(A)$ 。

Step 2 计算属性集 A 中每一个属性 a 的重要度 $Sig_{A-\{a\}}(a)$:

若 $Sig_{A-\{a\}}(a) \neq 0$, 表示系统有约简核, 转 Step 8;

若 $Sig_{A-\{a\}}(a) = 0$, 表示系统没有约简核, 转下一步。

Step 3 计算属性的两两组合对于 A 的重要度 $Sig_{A-\{a_i, a_j\}}(a_i, a_j)$, $1 \leq i \neq j \leq m$ 。

Step 4 取 $\max_{1 \leq i \neq j \leq m} Sig_{A-\{a_i, a_j\}}(a_i, a_j)$ 的属性集作为系统的次优约简 $red'(A)$:

$$red'(A) = \{(a_i, a_j) \mid \max_{1 \leq i \neq j \leq m} Sig_{A-\{a_i, a_j\}}(a_i, a_j)\}$$

Step 5 判断 $GD(red'(A))$ 与 $GD(A)$ 是否相等, 若 $GD(red'(A)) = GD(A)$, 输出最小约简 $red(A) = red'(A)$; 若 $GD(red'(A)) > GD(A)$, 转下一步。

Step 6 计算 $B = A - red'(A)$ 中每一个属性对于 $red'(A)$ 的重要度, 取 $\max_{b \in B} Sig_{red'(A)}(b)$ 。

Step 7 判断 $GD(red'(A) \cup b)$ 与 $GD(A)$ 是否相等, 若 $GD(red'(A) \cup b) = GD(A)$, 输出 $red(A) = red'(A) \cup b$; 若 $GD(red'(A) \cup b) > GD(A)$, 重复 Step 6 的过程, 直到与 $GD(A)$ 相等, 终止算法, 输出 $red(A)$ 。

Step 8 计算核 $Core(A)$:

$$Core(A) = \{a \in A \mid Sig_{A-\{a\}}(a) > 0\}$$

Step 9 判断 $GD(Core(A))$ 与 $GD(A)$ 是否相等, 若 $GD(Core(A)) = GD(A)$, 输出最小约简 $red(A) = Core(A)$; 若 $GD(Core(A)) > GD(A)$, 转下一步。

Step 10 计算 $B = A - Core(A)$ 中每一个属性对于 $Core(A)$ 的重要度, 取 $\max_{b \in B} Sig_{Core(A)}(b)$ 。

Step 11 判断 $GD(Core(A) \cup b)$ 与 $GD(A)$ 是否相等, 若 $GD(Core(A) \cup b) = GD(A)$, 输出 $red(A) = Core(A) \cup b$ 和 $Core(A)$; 若 $GD(Core(A) \cup b) > GD(A)$, 重复 Step 10 的过程。

Step 12 计算 $C = A - Core(A) \cup b$ 中每一个属性对于 $Core(A) \cup b$ 的重要度, 取重要度中最大值 $\max_{c \in C} Sig_{Core(A) \cup b}(c)$ 的那个属性与 $Core(A) \cup b$ 合并, 直到与 $GD(A)$ 相等, 终止算法。

2.3.3 信息系统的建立

选择 MIT-BIH 数据库中 $T_{100}, T_{102}, T_{103}, T_{104}, T_{114}, T_{116}, T_{213}, T_{221}, T_{230}$ 和 T_{233} 共计 10 条记录来做分析。在这 10 条记录中选取了 300 个心动周期波形, 并随机选择 100 个波形作为训练样本, 来区分 3 类心电信号: 正常信号 N、房性早搏 APC、室性早搏 PVC。

由此, 可以得到一个信息系统 $S_1 = (U, A, f, V)$, 其中:

(1) U 是全体对象组成的集合, $U = \{x_1, x_2, \dots, x_{300}\}$ 即信号样本集。

(2) $A = C \cup \{d\}$ 是全体属性构成的集合, 包含条件属性 $C = \{a_1, a_2, \dots, a_{12}\}$ 和决策属性 $\{d\} = \{1, 2, 3\}$ 。

(3) $f: U \times A \rightarrow V$ 是一个映射, 论域中对象的每个属性通过映射 f 都有一个信息值与其对应, 即对于 $\forall a \in A, x \in U, f(x, a) \in V_a$ 。

(4) V 是属性 a 的值域 V_a 的集合, $V = \bigcup_{a \in A} V_a$ 。

(5) $V_d = \{1, 2, 3\}$ 是决策属性的值域, 对于 $\forall x \in U, f(x, d) \in V_d$ 。若 $f(x, d) = 1$, 则表示对象 x 是正常信号(N); 若 $f(x, d) = 2$, 则表示对象 x 是房性早搏; 若 $f(x, d) = 3$, 则表示对象 x 是室性早搏, 如表 1 所列。

表 1 信号样本类别

信号类别	记号	$f(x_i, d)$
正常信号	N	1
房性早搏	APC	2
室性早搏	PVC	3

3 数据分析

3.1 特征提取

在进行了波形特征点的检测和定位之后, 通过一系列的运算便可得到间期和幅值等特征参数。所采集的 300 个样本数据中包含 100 个正常窦性信号(N), 88 个房性早搏(APC)、112 个室性早搏(PVC), 得到信息系统 $S_1 = (U, A, f, V)$ 。

由于心电特征参数的多样化, 得到的数据不统一会使得

分类器在处理具有较小和较大初始值域的特征属性时权重相差过大,导致误差增加。因此,在分类器数据输入之前需要先对条件属性进行离散化处理,本文选用等距离散法来离散条件属性。对于 $\forall a_i \in A (i=1, 2, \dots, 12)$ 设属性的值域 $V_a = [a_{\min}, a_{\max}]$,为了方便,将其记为 V_i ,离散原理就是产生一个划分 π ,这个划分能够完全覆盖 V_i 。根据每个属性的 a_i 实际取值,记 $V_i^t (t=1, 2, \dots, k)$ 为值域 V_i 的划分区间, k 为所划分区间的个数,即

$$\pi = \{[V_i^0, V_i^1], (V_i^1, V_i^2], \dots, (V_i^{k-1}, V_i^k]\}$$

这里 $V_i^0 = a_{\min}, V_i^k = a_{\max}, V_i^{t-1} < V_i^t$ 。

在本实验中,考虑到全部的属性取值,对于 $\forall V_i$,取 $k=5$,即对每个属性值 V_i 进行5个区间的划分。在对每一个条件属性 a 的划分上,划分区间都不相同,会根据 a_i 的实际取值 V_i 而定,从而将属于每个区间的属性值 V_i 重新赋值 V_i' 。离散属性值 $V_i' = \{k | (k=1, 2, 3, 4, 5)\}$,其中, $[V_i^0, V_i^1] = 1, \dots, (V_i^t, V_i^{t+1}] = 5$ 。

以属性Q-T间期 a_6 为例, $V_6 = [0.32, 0.44], [V_6^0, V_6^1] = [0.32, 0.344], [V_6^1, V_6^2] = (0.344, 0.368], [V_6^2, V_6^3] = (0.368, 0.392], [V_6^3, V_6^4] = (0.392, 0.416], [V_6^4, V_6^5] = [0.416, 0.44]$,因此 $f(x_1, a_6) = 1$,其他类似。其中,对象 $U = \{x_1, x_2, \dots, x_{300}\}$,条件属性 $C = \{a_1, a_2, \dots, a_{12}\}, V_i' = \{1, 2, 3, 4, 5\}$ 。

3.2 基于粗糙集的属性约简

通过前面所提出的约简方法,结合Rosetta软件对离散化的信息系统进行属性约简。通过属性约简,去掉了两个不必要属性,得到了5个约简集合,并且得到约简核 $Core(A) = \{a_1, a_2, a_4, a_6, a_{10}, a_{11}\}$,如表2所列。结合心电图医生的建议,选取约简集 C_5 中的属性构成新的信息系统来作为分类器的输入数据。由此,得到信息系统 $S_2 = (U, A, f, V')$ 。

表2 约简属性集

约简	属性	属性数
C_1	$\{a_1, a_2, a_4, a_6, a_{10}, a_{11}, a_{12}\}$	7
C_2	$\{a_1, a_2, a_4, a_6, a_8, a_9, a_{10}, a_{11}\}$	8
C_3	$\{a_1, a_2, a_4, a_6, a_7, a_9, a_{10}, a_{11}\}$	8
C_4	$\{a_1, a_2, a_4, a_5, a_6, a_7, a_8, a_{10}, a_{11}\}$	9
C_5	$\{a_1, a_2, a_4, a_5, a_6, a_8, a_{10}, a_{11}, a_{12}\}$	9

4 结论

本文选择支持向量机(SVM)作为分类器,选取径向基函数核函数(RBF核函数)作为SVM的核函数,选用CVSVM软件来进行ECG信号分类实验。分别对信息系统 $S_1 = (U, A, f, V)$ 和信息系统 $S_2 = (U, A, f, V')$ 进行分类,结果如表3、表4所列。

表3和表4分别展示了属性约简前、后SVM的分类效率。在对信息系统的分类试验中,正常窦性信号(N)、房性早搏(APC)、室性早搏(PVC)的识别率分别为92%、93%、90%。在对约简后的信息系统的分类试验中,正常窦性信号(N)、房性早搏(APC)、室性早搏(PVC)的误判和错判数都有了大幅的减少,使得它们的识别率分别达到了97%、97%和96%。实验表明,特征参数的合理选择是提高识别效率的重要因素,基于粗糙集理论的约简方法是特征选择的有效工具。

表3 SVM对信息系统 $S_1 = (U, A, f, V)$ 的识别率

	训练样本集			测试样本集		
	样本数	错检数 误判 漏判	识别率	样本数	错检数 误判 漏判	识别率
N	50	0 0	100%	100	5 3	92%
APC	50	0 0	100%	88	4 2	93%
PVC	50	0 0	100%	112	8 3	90%
小计	150	0 0	100%	300	17 7	92%

表4 SVM对信息系统 $S_2 = (U, A, f, V')$ 的识别率

	训练样本集			测试样本集		
	样本数	错检数 误判 漏判	识别率	样本数	错检数 误判 漏判	识别率
N	50	0 0	100%	100	2 1	97%
APC	50	0 0	100%	88	3 0	97%
PVC	50	0 0	100%	112	3 1	96%
小计	150	0 0	100%	300	8 2	97%

参考文献

- [1] Trahanias P, Skordalakis E. Syntactic pattern recognition of the ECG[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1990, 12(7): 648-657
- [2] Koski A, Juhola M, Meriste M. Syntactic. Recognition of ECG Signals by Attributed Finite Automata[J]. Pattern Recognition, 1995, 28(12): 1927-1940
- [3] 戚建新, 卞正中, 杨强. 基于微机的心电信号实时自动分析系统[J]. 北京生物医学工程, 1997, 16(3): 157-161
- [4] Engin M. ECG beat classification using neuro-fuzzy network[J]. Pattern Recognition, 2004, 25(15): 1715-1722
- [5] 季虎, 孙即祥. 心电信号自动分析关键技术研究[D]. 长沙: 国防科技大学, 2006
- [6] Tantawi M M, Revett K, Salem A, et al. An evaluation of the generalisability and applicability of the PhysioNet electrocardiogram(ECG) repository as test cases for ECG-based biometrics[J]. International Journal of Cognitive Biometrics, 2012, 1(1): 66-97
- [7] Tantawi M M, Revett K, Salem A, et al. Fiducial Feature Reduction Analysis for Electrocardiogram (ECG) Based Biometric Recognition[J]. Journal of Intelligent Information Systems, 2013, 40: 17-39
- [8] Tang X, Shu L. A Frame work of Automatic Analysis System of Electrocardiogram Signals[J]. International Journal of Signal Processing, Image Processing and Pattern Recognition, 2014, 7(2): 211-222
- [9] 陈文利, 莫智文, 郭文. 基于小波变换和黄金分割搜索法的QRS波检测算法[J]. 生物医学工程学杂志, 2009, 26(4): 748-751
- [10] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001
- [11] 梁吉业, 李德玉. 信息系统中的不确定性与知识获取[M]. 北京: 科学出版社, 2005
- [12] 赵敏, 罗可, 秦哲. 基于粒计算的属性约简算法[J]. 计算机工程与应用, 2008, 44(30): 157-159
- [13] 唐孝, 舒兰. 基于粒计算的属性约简改进算法[J]. 计算机科学, 2014, 41(11A): 313-315