

采用 S 变换特征选择方法的心律失常分类

吕 卫 邓为贤 褚晶辉 李 喆

(天津大学电子信息工程学院, 天津, 300072)

摘 要: 针对短时傅里叶变换与小波变换对心电图 (Electrocardiogram, ECG) 信号特征提取不足以及心律失常识别困难的问题, 提出了一种基于 S 变换特征选择的心律失常分类算法。首先对 ECG 信号进行 S 变换, 并从幅值和相位两个角度提取 ECG 信号的时频特征, 与形态特征和 RR 间隔组成原始特征向量。然后将遗传算法与支持向量机 (Support vector machine, SVM) 结合组成 Wrapper 式特征选择方法, 并在其中融入 ReliefF 算法, 即采用 ReliefF 算法计算特征权重, 并根据特征权重大小来指导遗传算法种群初始化, 遗传算法以 SVM 的分类性能作为适应度函数来搜索特征子集。最后使用“一对多” (One against all, OAA) SVM 对 MIT-BIH 心律失常数据库 8 种类型心拍进行分类。实验结果表明, 该算法达到了较好的分类效果, 灵敏度、特异性和准确率分别为 96.14%, 99.75% 和 99.81%。

关键词: 心律失常; S 变换; 遗传算法; ReliefF 算法; 支持向量机 (SVM)

中图分类号: TP391.4 **文献标志码:** A

Arrhythmia Classification Based on Feature Selection Method of S-transform

Lü Wei, Deng Weixian, Chu Jinghui, Li Zhe

(School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: Short time Fourier transform and wavelet transform are not effective in extracting features of electrocardiogram (ECG) signal for arrhythmia detection. Therefore, a novel algorithm based on the feature selection of S-transform is proposed for arrhythmia classification. First, ECG signals are processed by S-transform, and the time-frequency features are extracted from both the amplitude and the phase of ST results. Then, time-frequency features, morphological features, and RR interval are combined as the original feature vector. Second, the genetic algorithm (GA) and support vector machine (SVM) are combined as a Wrapper approach to search an optimal feature subset. The feature weights are computed by ReliefF algorithm, and the initialization of genetic population depends on the feature weights. Moreover, GA searches an optimal feature subset using classification performance as the fitness function. Finally, a multi-SVM model with one against all (OAA) strategy is built for the classification of eight types of ECG beats from the MIT-BIH arrhythmia database. Experimental results indicate that the proposed approach has the best performance among other state-of-the-art approaches, and the sensitivity, specificity, and accuracy reach 96.14%, 99.75%, and 99.81%, respectively.

Key words: arrhythmia; S-transform; genetic algorithm; ReliefF algorithm; support vector machine

引言

心脏搏动频率、节律以及 ECG 形态的改变被称为心律失常,其长期存在会引起严重的心脏病变甚至死亡,因此对心律失常的诊断十分重要^[1]。专家通过分析心电图(Electrocardiogram, ECG)检测心律失常,但是长时间分析心电图记录既耗时又枯燥,而且通过人眼很难精准辨别心电信号微小的形态变化,致使专家在诊断时可能会丢失或弄错重要信息,因此需要一个高效准确的计算机辅助诊断系统辅助医生检测心律失常。

目前,研究人员已提出一些分类方法。按照特征提取方法的不同可以将这些方法分为:时域方法、基于统计的方法和基于变换的方法。时域方法除了提取常用的 RR 间隔、QRS 复合波和 T 波持续时间作为特征外,还直接使用 ECG 信号采样值作为形态特征^[2]。时域方法简单、快捷,但时域特征无法揭示 ECG 信号细微变化和隐藏的信息。基于统计的方法是通过高阶累积量等提取特征来获取 ECG 信号隐藏的非线性信息^[3]。基于变换的方法是将采集到的心电信号变换到另一个域,如利用傅里叶变换获得信号的频域特征^[4]。然而通过分析发现采用傅里叶变换方法识别出的 5 类心拍中有 3 类心拍有相似的频谱,在一定程度上导致了误识别。这是由于生物医学信号是非平稳信号,傅里叶变换无法表示时间和频率的相互关系,对非平稳信号分析有局限性^[5]。因此,许多算法使用时频变换提取 ECG 信号特征,如短时傅里叶变换(Short time Fourier transform, STFT)^[6]、小波变换^[7]及 S 变换幅值信息^[8]等,这些变换能反映出非平稳信号频率随时间的变化,可以提取到时域或频域方法所不能提取到的 ECG 信号局部特征。然而,时频方法提取的特征往往维数较高,不仅增加了计算量,而且会给分类性能带来负面影响。特征选择是常用的降低维数的方法,可分为过滤式(Filter)和包裹式(Wrapper)。Filter 式独立于后续分类结果,而 Wrapper 法使用分类结果评估特征子集,其准确率高于 Filter 法,因此心律失常分类主要采用 Wrapper 法,如在使用最近邻分类器对心律失常分类时,通过对比发现基于准确性准则和顺序浮动前向选择法,即 Wrapper 式方法好于其他 3 种 Filter 式方法^[9]。在利用遗传算法(Genetic algorithm, GA)和支持向量机(Support vector machine, SVM)进行特征选择时也得到了较好的效果^[10]。除此之外,一种效果较好的 Filter 式特征选择算法 ReliefF^[11]比 Wrapper 法具有速度上的优势,所以也经常使用。

在特征提取方面,时域分析虽然不能提取信号隐藏的特征,但专家在判断心律失常类型时首先关注 ECG 信号的形态特征^[12],因此时域特征是一种识别心律失常重要且有效的特征。STFT 窗函数的高度与宽度固定不变,所以不能很好地同时提取高频和低频特征;小波变换可实现多尺度聚焦,揭示了 ECG 信号的时频特性,但不能很好地提取心电信号的局部特征,可能丢失一些关键特征,影响心律失常的分类效果。S 变换是一种良好的局部时频分析技术,相比 STFT 和小波变换,其对信号局部解析能力更强。S 变换中含有的相位因子保留了每个频率的绝对相位特征,这是小波变换所不具有的特性^[13],但 S 变换的相位信息并没有得到使用。

本文提出一种新的心率失常分类方法:首先,利用 S 变换从幅值和相位上提取 ECG 信号的时频特征,使用时间窗截取 ECG 信号的形态特征。由于 GA 具有全局寻优能力强且不易误入局部最优解等特点,利用 GA 和 SVM 结合组成 Wrapper 进行特征选择,同时为了加快特征选择的速度,使用 ReliefF 计算出特征权重大小指导遗传算法种群初始化。最后,使用“一对多”SVM 进行心律失常分类。

1 特征提取与分类器理论

1.1 特征提取——S 变换

S 变换是由 Stockwell 提出,可以由 STFT 和小波变换导出^[14]。从 STFT 角度导出 S 变换,连续信号 $x(t)$ 的 S 变换为 $S(t, f)$, 如式(1)所示。

$$S(\tau, f) = \int_{-\infty}^{\infty} x(t) \omega(t - \tau) e^{-i2\pi f t} dt \quad (1)$$

$$\omega(t - \tau) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(t-\tau)^2}{2\sigma^2}} \quad (2)$$

式中 $\omega(t - \tau)$ 为高斯窗口, 窗宽 $\sigma = 1/|f|$ 是关于频率倒数的函数, 这使得 S 变换窗口形状可以随频率的变化而变化, 在低频部分时可获得较高的频率分辨率; 而高频段部分可获得很高的时间分辨率。因此, S 变换可以被看作一种频变的 STFT, 克服了 STFT 分辨率不能调节的缺陷。

另外, 对于信号 $x(t)$ 的 S 变换也能用一个特殊的母小波的连续小波变换乘以一个相位因子得到, 如式(3)所示。

$$S(\tau, f) = e^{i2\pi f \tau} W(\tau, d) \quad (3)$$

式中 $W(\tau, d)$ 为信号 $x(t)$ 的小波变换, $d = 1/|f|$ 。式(4)为特殊的母小波。

$$\omega(t, f) = \frac{|f|}{\sqrt{2\pi}} e^{-\frac{t^2 f^2}{2}} e^{-i2\pi f t} \quad (4)$$

式中 $e^{-i2\pi f t}$ 为小波变换所不具有的相位因子, 所以 S 变换也可以看作是相位修正的连续小波变换, 可以解决小波的相位局部化问题。同时 S 变换保持了每个频率成分的绝对参考相位不变。

连续信号 $x(t)$ 的离散形式为 $x[kT]$, S 变换的离散形式为

$$\begin{cases} S[jT, 0] = \frac{1}{N} \sum_{k=0}^{N-1} x[kT] & n = 0 \\ S[jT, \frac{n}{NT}] = \sum_{m=0}^{N-1} X[\frac{m+n}{NT}] e^{-\frac{2\pi^2 m^2}{n^2}} e^{\frac{i2\pi m j}{N}} & n \neq 0 \end{cases} \quad (5)$$

式中: N 为总采样点数; T 为时域采样间隔; $j, m, n = 0, 1, \dots, N-1$; $X[\frac{m+n}{NT}]$ 为 $x[kT]$ 的 N 点离散傅里叶变换。信号经 S 变换后得到一个 $N \times M$ 的复时频矩阵: S 矩阵(S-matrix), 其行为频率值, 列为采样点; 矩阵中元素为复数, 其模为幅值。

1.2 分类器——SVM

SVM 是 Vapnik 等基于统计学习理论提出的一种机器学习方法^[15], 其主要思想是寻找能够成功分开两类样本并且具有最大分类间隔的最优分类超平面。

不失一般性, 给定训练样本集 $W = \{(x_i, y_i), i = 1, 2, \dots, l\}$, 其中输入样本 $x_i \in \mathbf{R}^d$, 类标签 $y_i \in \{+1, -1\}$ 。对于线性二分类问题, 超平面方程一般为: $\mathbf{x} \cdot \mathbf{w} + b = 0$, \mathbf{w} 为权值向量, b 为偏差项。将超平面规范化, 得到分类间隔等于 $2/\|\mathbf{w}\|$ 。则求解最优分类超平面等效于最小化 $\|\mathbf{w}\|$, 即求解式(6)的条件极值问题, 通过拉格朗日乘数求解法, 最后得到最优分类函数, 如式(7)所示。

$$\min \Phi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 \quad (6)$$

$$\text{s. t. } y_i [(\mathbf{x}_i \cdot \mathbf{w}) + b] \geq 1 \quad i = 1, 2, \dots, l$$

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^l y_i \alpha_i (\mathbf{x}_i \cdot \mathbf{x}) + b\right) \quad (7)$$

式中拉格朗日乘子 $\alpha_i \geq 0, i = 1, 2, \dots, l$, 只有少数 $\alpha_i > 0$ 并且其对应的样本 x_i 被称为支持向量。

对于解决实际应用中的非线性分类问题, SVM 引入核函数 $K(\mathbf{x}_i, \mathbf{x}_j)$, 其中 $K(\mathbf{x}_i, \mathbf{x}_j)$ 需满足 Mercer 条件, 最后得到 SVM 基本形式(式(8)), 常用核函数有线性核函数、Sigmoid 核函数、径向基核函数和多项式核函数。

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^l \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b\right) \quad (8)$$

2 采用S变换特征选择方法的心律失常分类

采用S变换特征选择方法的心律失常分类包括:预处理、特征提取、特征选择和多分类SVM,流程框图如图1所示。在训练阶段,使用特征选择与多分类SVM得到最优特征向量和训练模型;测试阶段,根据训练阶段得到的最优特征选择测试样本的特征和模型实现多种心律失常分类。

2.1 ECG数据集

MIT-BIH心律失常数据库^[6](以下简称MIT数据库)包含48条ECG记录,每条记录由两种不同类型导联(记为A、B导联)记录了长约30 min的数据,采样率为360 Hz;其中45条记录的A导联采用MLII导联,其余采用V5导联;40条记录的B导联采用V1导联,其余采用II、V2、V4和V5导联。本文使用的ECG数据集来自MIT数据库中MLII导联记录的以下8种心拍类型:正常搏动(Normal beat, NORM)、左束支传导阻滞(Left bundle branch block beat, LBBB)、右束支传导阻滞(Right bundle branch block beat, RBBB)、室性早搏(Premature ventricular contraction, PVC)、房性早搏(Atrial premature beat, APB)、起搏心跳(Paced beat, PB)、室性扑动波(Ventricular flutter wave, VFW)和室性逸搏(Ventricular escape beat, VEB)。

2.2 预处理

预处理首先是消除噪声对心电信号的干扰,其次是检测R波。关于预处理的算法已得到了深入研究并且效果很好。本文研究的主要问题是心律失常的分类。因此,在噪声去除方面,使用中值滤波法去除基线漂移,利用低通滤波器去除电力线干扰和低频噪声,此方法简单、快捷^[17]。R波位置使用MIT数据库中注释文件提供的基准点位置。

2.3 特征提取

2.3.1 RR间隔特征

本文提取4个RR间隔特征:preRR指给定心拍和其前一心拍的RR间隔;postRR指给定心拍和其前一心拍的RR间隔;localavgRR指给定心拍前后各5个心拍,共10个RR间隔的平均值;avgRR指一个记录文件中所有RR间隔的平均值。

2.3.2 形态特征

本文使用3个时间窗截取QRS复合波、T波和P波,QRS复合波窗口从R-50 ms到R+100 ms(R表示R波位置),以60 Hz采样率提取10个特征点;T波窗口从R+150 ms到R+500 ms,以20 Hz采样率提取8个特征点;P波窗口从R-200 ms到R-100 ms,以60 Hz采样率提取7个特征点。

2.3.3 时频特征

在R波前后各取278 ms,形成一个包含200点的ECG信号样本,如图2(a)所示。然后对该信号进行S变换。图2(b)为样本信号的S变换的幅值等高线图,横坐标表示心电信号某个采样点的幅值在不同频率上的分布,纵坐标表示心电信号某一频率处的幅值在不同采样点处的分布。S变换结果为 200×45 的复数矩阵,如果直接取这些数据作为特征,必然会增加计算难度。为此本文从幅值和角度方面做了分析,结果如图3所示。其中图3(a)为频率对应最大幅值曲线,即S-matrix矩阵每行中的最大幅值。从图3(a)中可看出LBBB、RBBB、PVC、VFW及VEB幅度最大的频率成分主要集中在20 Hz范围内;NORM、PB和APB幅度最大的频率成分主要集中在60 Hz范围内;而且只有NORM和APB幅度最大

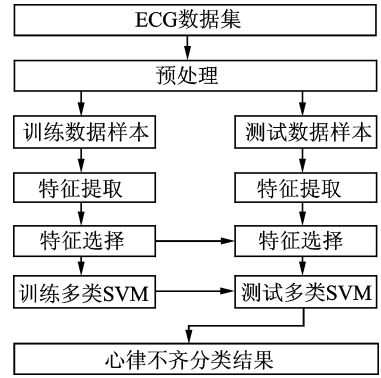


图1 本文提出的心律不齐分类算法流程图

Fig. 1 Block diagram of the proposed algorithm

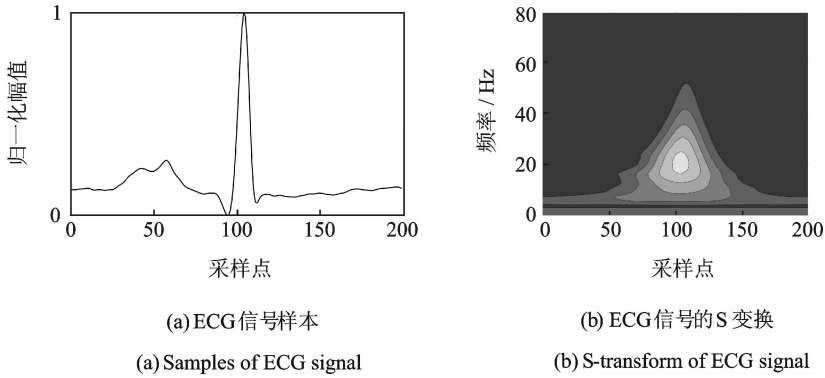


图2 ECG信号样本及其S变换

Fig. 2 Samples of ECG signal and its S-transform

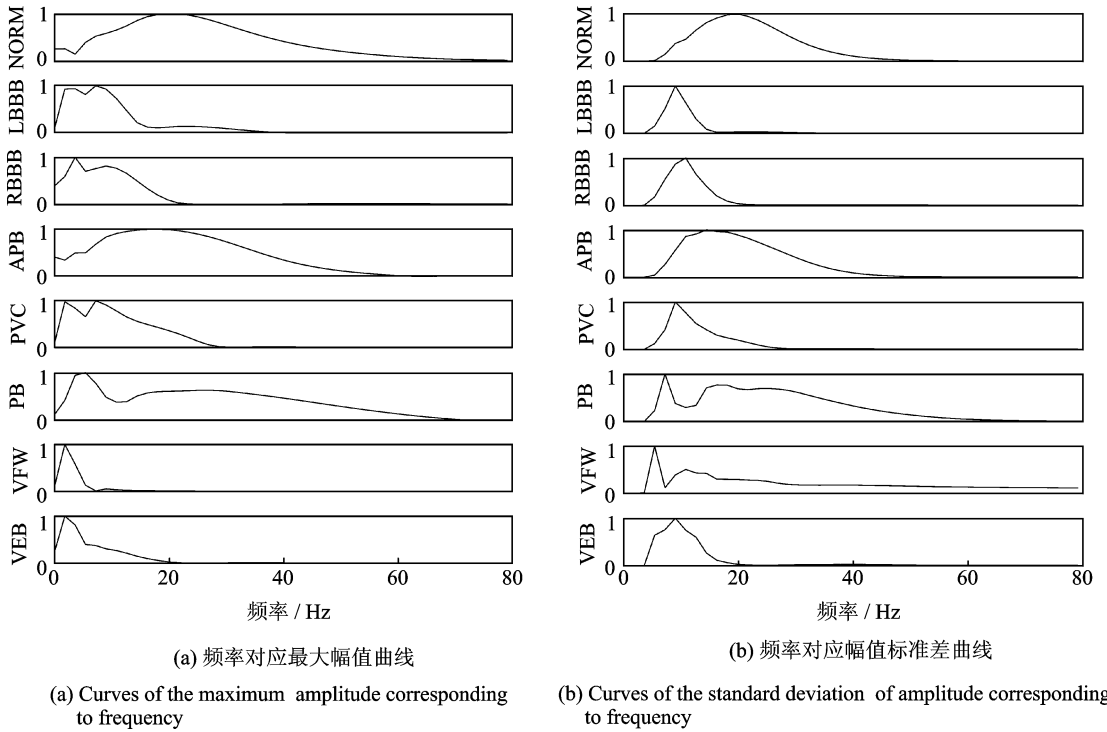
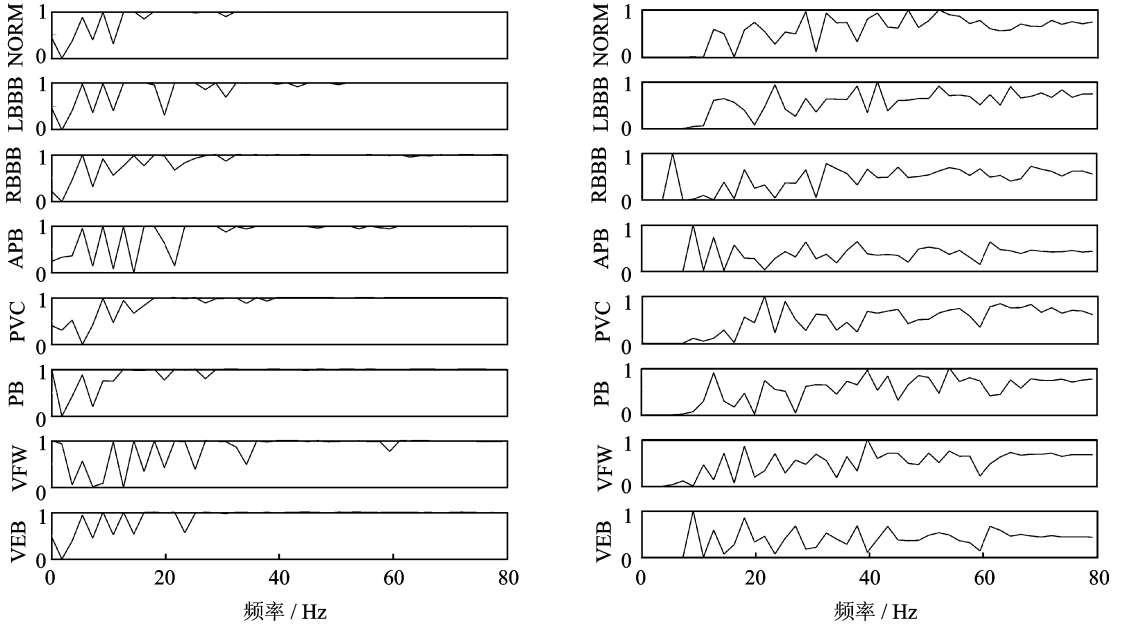


图3 S变换频率对应的幅值特性

Fig. 3 Amplitude-frequency characteristics of S-transform

值点对应频率在 20 Hz 左右,其他类型心拍对应频率都远小于 20 Hz。图 3(b)是频率对应幅值标准差曲线,可以看出,除 NORM、APB、PB 和 VFW 各频率对应幅值在 3~40 Hz 范围外,其余 4 种类型心拍各频率对应幅值变化均在 3~20 Hz 范围内。

S 变换保留了各频率分量的绝对相位信息,而小波变换不具有此特性,因此本文提取 S 变换的相位作为特征。图 4(a)为频率对应最大相位曲线,即图 3(a)中每个点对应的相位,可以看出每种心拍频率对应最大幅值的相位变换趋势不同,而且在大约 45 Hz 以后不再有太大变化;图 4(b)是频率对应相位标准差曲线,即 S-matrix 矩阵的每行所有点的相位标准差,可以看出在 10 Hz 以内只有 RBBB、APB 和



(a) 频率对应最大相位值曲线
(a) Curves of the maximum phase corresponding to frequency

(b) 频率对应相位标准差曲线
(b) Curves of the standard deviation of phase corresponding to frequency

图 4 S 变换频率对应的相位特性
Fig. 4 Phase-frequency characteristics of S-transform

VEB 相位存在变化,其他类型心拍相位标准差为 0。

研究表明,ECG 的能量主要集中在 0.5~45 Hz^[18],因此选取图 3(a)中 45 Hz 范围内的频率对应最大幅值作为特征,选取图 3(b)中 3~40 Hz 范围内的频率对应幅值标准差作为特征,选取图 4 中 45 Hz 范围内的频率对应最大相位值和频率对应相位标准差作为特征。

综上所述,本文将提取到的 4 个 RR 间隔、25 个形态特征和 101 个时频特征组成 130 维原始特征向量。

2.4 特征选择

ReliefF 是一种计算特征权重的 Filter 式特征选择算法,其思想为:通过计算特征对同类近邻样本与异类近邻样本之间的差异来赋予特征权重^[19]。

遗传算法是一种借鉴生物界适者生存、优胜劣汰进化思想形成的寻优算法。在使用遗传算法做特征选择时,用一串长度为 d (d 为原始特征向量维数)的二进制串结构(称为染色体或个体)表示 1 个特征组合,1 表示对应的特征被选中,0 表示对应的特征未被选中。以 N 个个体组成的初始种群为基础,根据个体适应度好坏进行选择、交叉和变异操作得到下一代,如此反复循环,直到满足终止条件,然后输出适应度最好的个体作为最终的特征。

本文将遗传算法和 SVM 结合组成 Wrapper 法进行特征选择。由于好的初始种群可以为遗传算法提供良好的搜索起点^[20],本文采用 ReliefF 算法先计算出特征权重并根据其大小指导种群初始化,为搜索提供一个好的起点,加快特征选择的速度。特征选择的流程图如图 5 所示,具体实现步骤如下:

步骤 1 计算特征权重,使用 ReliefF 算法计算原始特征向量中每个特征的权重。

步骤 2 初始化种群,首先设置种群个体数量 N 。根据步骤 1 计算出的特征权重指导种群初始化,

即让特征权重大小决定该特征在染色体个体中对应位为 1 的概率, 设权重最大的特征被选中的概率为 p_1 , 权重最小的特征被选中的概率为 p_2 , 然后按等差数列求出其余特征被选中的概率。

步骤 3 计算个体适应度, 使用分类器的分类准确率 r_i 作为第 i 个个体的适应度。

步骤 4 判断是否满足收敛条件(最大迭代次数 G 设置为 100), 若满足, 经个体解码得到最优特征向量, 不满足则重复步骤 4~7。

步骤 5 选择操作, 利用公式 $p_i = \frac{r_i}{\sum_j r_j}$ 计算第 i 个个体的被选中的概率。

步骤 6 交叉操作, 根据交叉概率 p_c 交叉种群中个体产生新个体。

步骤 7 变异操作, 根据变异概率 p_m 改变个体中某位的值产生新个体。

遗传算法初始化种群中个体数量太大, 势必增加计算量, 降低运行效率; 如果数量太小, 虽然提高了运行速度, 但种群的多样性被减弱, 降低了遗传算法的优化质量, 种群个体数量一般为 20~100; 交叉操作是产生新个体主要途径, 所以交叉概率取值较大, 一般取 0.4~0.99; 变异操作在一定程度上增加了遗传算法的局部搜索能力, 如果变异概率太大, 会使遗传算法变成随机搜索, 一般取 0.000 1~0.1; 本文遗传算法参数根据经验值^[21, 22], 并兼顾种群多样性和运行效率设定: 初始化种群 $N=40$ 、交叉概率 $p_c=0.9$ 和变异概率 $p_m=0.01$ 。在根据特征权重大小初始化种群时, $p_1 > 0.5 > p_2$, p_1 越接近 1 表示权重最大的特征出现的概率大, p_2 越接近 0 表示权重最小的特征出现的概率越小。实际中为了保持各个特征出现概率不会相差太大并保持适当梯度, p_1 不宜太大、 p_2 不宜太小^[23], 本文选 $p_1=0.8$, $p_2=0.3$ 。

2.5 多分类 SVM

为了使 SVM 能够解决多分类问题, 通过一定策略使用多个二分类 SVM 构造多分类 SVM, 常见策略有: 一对一策略(One against one, OAO)和一对多策略(One against all, OAA)。由于两种策略在分类准确性上接近且 OAA 在构建多分类 SVM 时需要的二分类 SVM 少, 本文采用“一对多”策略。

3 实验结果与分析

从 MIT 数据库存在 MLII 导联记录的 45 个记录文件中抽取待分类的 8 种类型心拍, 具体实验数据集的组成如表 1 所示。采用灵敏度(Sensitivity, ST)、特异性(Specificity, SF)和准确率(Accuracy, AC)评估分类性能, 计算公式如式(9~11)所示, 其中 TP、TN、FP 和 FN 分别表示真阳性、真阴性、假阳性和假阴性。

$$ST = TP / (TP + FN) \times 100 \quad (9)$$

$$SF = TN / (TN + FP) \times 100 \quad (10)$$

$$AC = (TP + TN) / (TP + FN + FP + FN) \times 100 \quad (11)$$

为了考查本文所使用的特征选择方法(ReliefF-GA-SVM)相对于其他特征降维方法所得到的特征子集对分类性能的影响, 增加了原始特征向量、特征抽取法 SNGA-KPCA^[24]、Filter 法 FM-Relief^[25] 和 Wrapper 法 RF-MI^[26] 对比实验。

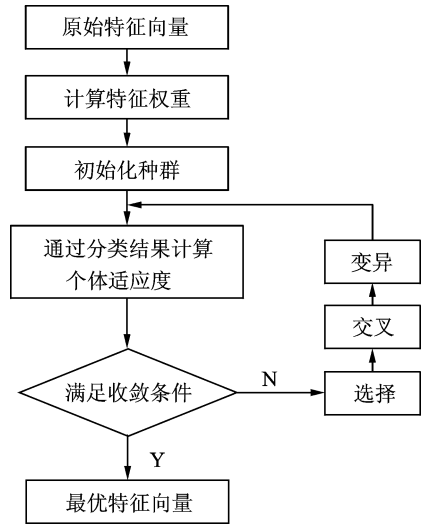


图 5 特征选择流程图

Fig. 5 Flow chart of feature selection process

表 1 实验数据集组成
Tab. 1 Composition of the data set

类型	训练数据集	测试数据集
NORM	34 135	34 544
LBBB	3 838	3 785
RBBB	3 437	3 309
PVC	3 352	3 369
APB	1 225	1 190
PB	1 733	1 646
VFW	225	195
VEB	44	56

表 2 给出了使用不同特征选择方法进行特征降维后的分类性能,可见 SNGA-KPCA 大大减少了特征维数,但其准确率比原始特征向量差,准确率为 95.72%。FM-Relief 和 RF-MI 的准确率比原始特征向量好,但降低的维数较少。本文使用的特征选择方法准确率最高,为 99.81%,明显高于原始特征向量;并且降维后的特征数量为 71,在所有方法中最少,其中 RR 间隔特征在降维后仍为 4 个,形态特征由 25 个被降维到 16 个,时频特征由 101 被降维到 51 个。

表 2 不同特征选择方法下的分类性能

Tab. 2 Classification performance with different feature selection methods

方法	准确率/%	特征维数
原始特征向量	96.14	130
SNGA-KPCA	95.72	73
FM-Relief	97.34	113
RF-MI	98.13	104
本文特征选择方法	99.81	71

以上结果说明,SNGA-KPCA 通过非线性映射将原始特征向量所在空间变换到线性可分的高维空间进行主成分分析,在大大减少特征维数的同时也丢失掉了一些信息,因而降低了分类的性能。在特征选择方面,由于本文使用的特征选择方法和 RF-MI 都是 Wrapper 法在特征选择过程中依赖分类结果,所以相对于使用 FM-Relief 算法的 Filter 法,能选出更重要的特征,获得更高的准确率。使用 FM-Relief 算法的 Filter 式特征选择方法的分类准确率提高不多,降维的效果也有限,主要因为 FM-Relief 算法在计算特征权重时依据特征对近距离样本的区分能力而非分类结果,只能粗略地得出特征权重。本文使用的特征选择方法优于 RF-MI,原因在于 RF-MI 通过 ReliefF 和互信息结合分类性能分两次寻找阈值只能达到局部最优,不能达到全局最优。从本文特征选择方法降维后的效果看,作为最常用的特征 RR 间隔最为重要,形态特征次之,同时也验证了这两种特征是专家诊断时需要的重要特征,时频特征虽有较多无关特征但也是一种计算机辅助诊断的重要特征。

表 3 列出了在相同实验环境下,本文使用 ReliefF-GA-SVM 算法与不采用 ReliefF 算法随机初始化种群的基本遗传算法在特征选择时间和预测准确率上的性能。可以看出本文使用 ReliefF-GA-SVM 算法具有更高的预测准确率和更短的特征选择时间。

表 4 为本文使用 ReliefF-GA-SVM 方法对不同类型心拍的分类性能。由表 4 可以看出,本文使用的 ReliefF-GA-SVM 方法特异性和准确率基本都达到 99%;灵敏度都达到 91%以上,其中 NORM、LBBB、RBBB、PB 达到了 99%。APB 的灵敏度低于 90%的原因在于 MIT 数据中 1 个病人同时存在心动

表 3 ReliefF-GA-SVM 与基本遗传算法性能对比

Tab. 3 Performance comparison of ReliefF-GA-SVM and the basic genetic algorithm

搜索算法	运行时间/min	预测准确率/%
ReliefF-GA-SVM	73	99.81
基本遗传算法	91	98.02

表 4 ReliefF-GA-SVM 方法对不同的心拍类型的分类性能

Tab. 4 Classification performance of ReliefF-GA-SVM approach for various heartbeats %

分类性能	NORM	LBBB	RBBB	APB	PVC	PB	VFW	VEB	平均
灵敏度	99.77	99.15	99.45	88.90	97.86	99.87	91.28	92.85	96.14
特异性	98.45	99.97	99.95	99.89	99.83	99.99	99.98	100.00	99.75
准确率	99.40	99.91	99.94	99.63	99.69	99.99	99.95	99.99	99.81

过缓和房性早搏(APB)症状,导致该病人的 APB 心拍与其他病人的 NORM 心拍在特征上相似,造成错检。

表 5 列出了本文分类算法和其他分类算法所支持的心拍种类以及分类的准确率。其他分类算法分别为文献[1]的 PCA-ELM 法、文献[3]的 HOC-NN 法、文献[5]的 CSD-SVM 法、文献[7]的 DTCWT-ANN 法和文献[9]的 Gabor-KNN 法,这些方法可识别的心拍种类都少于本文算法,并且准确率都低于本文算法;在同样区分出 8 种心拍的情况下,本文算法的准确率高出文献[9]方法。

表 5 不同方法的分类性能比较

Tab. 5 Comparison of classification accuracies of different methods

方法	心拍种类	准确率/%
PCA-ELM	7	98.72
HOC-NN	5	94.52
CSD-SVM	5	98.60
DTCWT-ANN	5	94.64
Gabor-KNN	8	98.65
本文算法	8	99.81

4 结束语

本文利用 S 变换、特征选择方法和 SVM 实现心律失常分类。通过分析 ECG 信号存在频率范围内的 S 变换频率对应的幅度和相位,发现了其对不同种类心拍有较好的区分能力,提取了 S 变换的幅度与相位作为时频特征,并与 RR 间隔和形态特征一起组成特征向量。使用遗传算法和 SVM 组成的 Wrapper 法对其进行特征选择,并利用 ReliefF 算法提供的特征权重先验知识初始化遗传算法种群。采用“一对多”SVM 对 MIT-BIH 心律失常数据库的 8 种心拍类型进行分类。实验结果表明,本文提出的 ReliefF-GA-SVM 特征选择算法在分类准确率和降维效果上优于其他 3 种方法,同时特征选择花费的时间少于基本遗传算法。然而,目前 S 变换在心律失常分类上研究较少,如何进一步利用 S 变换及其变形、减少 S 变换计算复杂度和提高特征选择速度是下一步需要深入研究的工作。

参考文献:

- [1] Kim J, Hang S S, Shin K, et al. Robust algorithm for arrhythmia classification in ECG using extreme learning machine [J].

Biomedical Engineering Online, 2009, 8(1): 31.

- [2] De C P, Reilly R B. A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features[J]. IEEE Transactions on Biomedical Engineering, 2006, 53(1):2535-2543.
- [3] Martis R J, Acharya U R, Lim C M, et al. Application of higher order cumulant features for cardiac health diagnosis using ECG signals [J]. International Journal of Neural Systems, 2013, 23(4):806-839.
- [4] Owis D M I, Youssef A B M, Kadah Y M. Characterisation of electrocardiogram signals based on blind source separation [J]. Medical & Biological Engineering & Computing, 2002, 40(5):557-564.
- [5] Khalaf A F, Owis M I, Yassine I A. A novel technique for cardiac arrhythmia classification using spectral correlation and support vector machines [J]. Expert Systems with Applications, 2015, 42(21):8361-8368.
- [6] Tspouras M G, Fotiadis D I. Automatic arrhythmia detection based on time and time-frequency analysis of heart rate variability [J]. Computer Methods & Programs in Biomedicine, 2004, 74(2):95-108.
- [7] Thomas M, Das M K, Ari S. Automatic ECG arrhythmia classification using dual tree complex wavelet based features [J]. AEU-International Journal of Electronics and Communications, 2015, 69(4):715-721.
- [8] Das M K, Ari S. Electrocardiogram beat classification using S-transform based feature set [J]. Journal of Mechanics in Medicine & Biology, 2014, 14(5):396-411.
- [9] 黄惠芳, 胡广书. 基于 Gabor 变换特征选择的 ECG 心拍分类 [J]. 清华大学学报:自然科学版, 2010(3):442-445.
Huang Huifang, Hu Guangshu. Electrocardiogram beat classification based on the feature selection of Gabor transform[J]. Journal of Tsinghua University (Science and Technology), 2010, 50(3):442-445.
- [10] Nasiri J A, Naghibzadeh M, Yazdi H S, et al. ECG arrhythmia classification with support vector machines and genetic algorithm[C]// EMS 2009, Third UKSim European Symposium on Computer Modeling and Simulation. Athens, Greece:[s. n.], 2009:187-192.
- [11] 刘全金, 赵志敏, 李颖新. 基于特征间距的二次规划特征选取算法[J]. 数据采集与处理, 2015, 30(1):126-136.
Liu Quanjin, Zhao Zhimin, Li Yingxin. Feature selection algorithm based on quadratic programming with margin between features [J]. Journal of Data Acquisition & Processing, 2015, 30(1):126-136.
- [12] 张嘉伟. 心电图形态特征的识别及其在分类中的作用研究 [D]. 上海:华东师范大学软件学院, 2011.
Zhang Jiawei. Research on ECG morphological features recognition and its effect to classification [D]. Shanghai: School of Computer Software, East China Normal University, 2011.
- [13] Stockwell R G. Why use the S-transform? [C]. ISAAC Workshop. Toronto, Canada:AMS, 2007, 52: 279-309.
- [14] Stockwell R G, Mansinha L, Lowe R P. Localization of the complex spectrum: The S transform [J]. Signal Processing IEEE Transactions on, 1996, 44(4):998-1001.
- [15] Cortes C, Vapnik V. Support-vector networks [J]. Machine Learning, 1995, 20(3):273-297.
- [16] Moody G B, Mark R G. The impact of the MIT-BIH arrhythmia database[J]. IEEE Engineering in Medicine & Biology Magazine, 2001, 20(3):45-50.
- [17] De C P, Dwyer M O, Reilly R B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features [J]. IEEE Transactions on Bio-medical Engineering, 2004, 51(7):1196-1206.
- [18] 胡广书, 汪梦蝶. 生物医学信号处理研究综述[J]. 数据采集与处理, 2015, 30(5):915-932.
Hu Guangshu, Wang Mengdie. Survey on biomedical signal processing [J]. Journal of Data Acquisition & Processing, 2015, 30(5):915-932.
- [19] Kononenko I. Estimating attributes: Analysis and extensions of RELIEF [J]. Lecture Notes in Computer Science, 1994, 784:356-361.
- [20] 刘奕晨, 王毅, 牛奕龙, 等. 基于标准差的自适应激素调节遗传算法[J]. 数据采集与处理, 2012, 27(3):333-339.
Liu Yichen, Wang Yi, Niu Yilong, et al. An adaptive genetic algorithm based on hormone regulation [J]. Journal of Data Acquisition & Processing, 2012, 27(3):333-339.
- [21] 李敏强, 寇纪淞. 遗传算法的基本理论与应用[M]. 北京:科学出版社, 2003:163-215.
Li Minqiang, Kou Jisong. The basic theory and application of genetic and algorithm [M]. Beijing: Science Press, 2003: 163-215.
- [22] 刘新平, 薛希文. 基于改进 LS-SVM 的随钻测量数据传输误码率预测[J]. 数据采集与处理, 2014, 29(5):790-794.

Liu Xinping, Xue Xiwen. Prediction of error rate in measurement-while-drilling data transmission based on improved LS-SVM method[J]. *Journal of Data Acquisition & Processing*, 2014, 29(5):790-794.

- [23] 周志红, 周新聪, 袁成清. 基于过滤器-封装器组合模型的故障特征选择算法 [J]. *中国机械工程*, 2007, 18(16):1988-1991.
Zhou Zhihong, Zhou Xincong, Yuan Chengqing. A novel hybrid filter-wrapper algorithm of feature selection for fault diagnosis[J]. *China Mechanical Engineering*, 2007, 18(16):1988-1991.
- [24] Feng Zhigang, Xu Tao. Fault feature extraction of ground-testing bed based on SNGA optimized KPCA [C]// *International Conference on Computer Science and Service System*. [S. l.]:IEEE, 2011:3244-3247.
- [25] Deng Z, Chung F L, Wang S. Robust relief-feature weighting, margin maximization, and fuzzy optimization[J]. *IEEE Transactions on Fuzzy Systems*, 2010, 18(4):726-744.
- [26] Yang F, Cheng W, Dou R, et al. An improved feature selection approach based on ReliefF and mutual information[C]// *International Conference on Information Science and Technology*. [S. l.]:IEEE, 2011:246-250.

作者简介:



吕卫(1976-),男,博士,副教授,硕士生导师,研究方向:数字信号处理、图像处理、模式识别、嵌入式系统设计, E-mail: luwei@tju.edu.cn.



邓为贤(1989-),男,硕士研究生,研究方向:模式识别、数字信号处理、嵌入式系统设计。



褚晶辉(1969-),女,博士,副教授,硕士生导师,研究方向:数字信号处理、图像处理、模式识别, E-mail: cjh@tju.edu.cn.



李喆(1992-),男,硕士研究生,研究方向:模式识别、数据挖掘、图像处理。

(编辑:张黄群)

