

doi: 10.13241/j.cnki.pmb.2018.18.011

## · 临床研究 ·

## 基于多尺度快速样本熵与随机森林的心电图分析\*

姜苗苗<sup>1</sup> 于波<sup>1</sup> 张烁<sup>1</sup> 陈寅生<sup>2</sup> 王祁<sup>3</sup>

(1 哈尔滨医科大学第二附属医院心血管内科 黑龙江 哈尔滨 150001;

2 哈尔滨理工大学测控技术与仪器重点实验室 黑龙江 哈尔滨 150001; 3 哈尔滨工业大学自动化测试系 黑龙江 哈尔滨 150001)

**摘要 目的:**探讨基于多尺度快速样本熵与随机森林的心电图分析方法对常见心律失常(房性早搏、室性早搏)的自动诊断的可行性和有效性。**方法:**利用不同心律失常疾病的心电信号存在复杂性差异的特点,通过多尺度熵计算心电信号在不同尺度下的样本熵值以组成特征向量;利用 kd 树提高多尺度熵的计算效率,增强算法的实时性。利用训练样本的特征向量构建随机森林分类器,再根据众多决策树的分类结果结合投票原则确定测试样本心律失常疾病的类型。**结果:**本文提出的心电图分析方法能够有效地识别正常心律、房性早搏(APB)及室性早搏(VPB),平均识别准确率达到 91.60%。**结论:**本文提出的心电图分析方法对常见心律失常(APB,VPB)具有较高的识别准确率及临床实用价值。

**关键词:**多尺度样本熵;随机森林;心电信号;心律失常

**中图分类号:**R541.7;R540.41 **文献标识码:**A **文章编号:**1673-6273(2018)18-3453-06

## Electrocardiogram Analysis Based on Multiscale Fast Sample Entropy and Random Forest\*

JIANG Miao-miao<sup>1</sup>, YU Bo<sup>1</sup>, ZHANG Shuo<sup>1</sup>, CHEN Yin-sheng<sup>2</sup>, WANG Qi<sup>3</sup>

(1 Department of Cardiology, The Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, 150001, China;

2 Key Laboratory of Measurement and Control Technology and Instrument, Harbin University of Science and Technology, Harbin, Heilongjiang, 150001, China; 3 Department of Automated Testing, Harbin, Heilongjiang, 150001, China)

**ABSTRACT Objective:** To explore the feasibility and effectiveness of ECG analysis method based on multiscale fast sample entropy and random forest for automatic diagnosis of common arrhythmia (atrial premature beat, ventricular premature beat). **Methods:** Different arrhythmia diseases having the features of complexity difference of ECG signals are adopted, and the sample entropy of ECG signal at different scales calculated by the multi-scale entropy forms eigenvectors; Using kd tree to improve the computation efficiency of multiscale entropy, the real-time performance of algorithm is enhanced. The random forest classifier is constructed by the eigenvectors of training samples, and then the type of arrhythmias is determined by the classification results of the numerous decision trees coupled with voting principle. **Results:** The proposed electrocardiogram analysis method can effectively identify the normal heart rhythm, atrial premature beats (APB) and ventricular premature beats (VPB), with an average identification accuracy of 91.60%. **Conclusions:** The ECG analysis method presented in this paper has high recognition accuracy and clinical value for common arrhythmia (APB, VPB).

**Key words:** Multiscale sample entropy; Random forest; Electrocardiogram; Arrhythmia

**Chinese Library Classification(CLC):** R541.7; R540.41 **Document code:** A

**Article ID:** 1673-6273(2018)18-3453-06

## 前言

心电图(Electrocardiograph, ECG)是通过皮肤电极对心脏电活动的图形化记录,能够反映心脏基本功能及疾病特征,对心血管临床诊断具有重要的参考价值<sup>[1]</sup>。正常的心电信号是一种低频周期性的非平稳信号,一旦心脏发生疾病或出现功能性退化,将直接导致心电信号的时-频变化。医生正是利用心电信号的变化特征,并结合丰富的临床经验对心律失常疾病患者进行诊断。随着医疗技术的不断发展与进步,临床医学辅助分析设备对心电图自动诊断结果的准确性和实时性的需求不断

增加,并引起相关领域专家的重点关注。鉴于此,作为心血管内科的重要临床辅助诊断工具,ECG 自动诊断方法的研究与应用具有重要的实际临床意义。

为进一步解决典型心律失常自动诊断的实时性和准确性问题,本文提出了基于多尺度快速样本熵与随机森林相结合的心电图分析方法。该方法利用多尺度熵对心电信号进行多尺度特征提取并组成特征向量描述心电信号的病理特征;通过 kd 树计算样本熵,显著提高多尺度熵的计算效率。再利用训练样本的特征向量集构建随机森林分类器,实现对常见心律失常(房性早搏、室性早搏)疾病的自动诊断。

\* 基金项目:黑龙江省教育厅科研规划项目(GBC1213121)

作者简介:姜苗苗(1988-),女,硕士,住院医师,主要研究方向:冠心病诊断与治疗,E-mail: lium65@163.com,电话:0451-86297647

(收稿日期:2018-05-22 接受日期:2018-06-16)

## 1 材料与方 法

### 1.1 材料

本文采用美国麻省理工学院 MIT-BIH 心律失常数据库<sup>[2]</sup>对本文提出的心电图分析方法的可行性和有效性进行实验验证。MIT-BIH 数据库由 Beth Israel 医院心律失常实验室于 1975 至 1979 年间获得的 4000 余个长期动态心电图记录组成,记录中大约 60%的心电图来自于住院的患者。MIT-BIH 心律失常数据库主要记录了 MLII 导联和 V5 导联的心电信号数据。目前,该数据库的应用广泛,是国际上公认的标准心电数据库之一。表 1 所示为验证本文提出的心电图分析方法的实验样本,包括正常心律、室性早搏和房性早搏的心律失常的样本。每种心律类型的实验样本都包括 50 个训练样本和 20 个测试样本。在每次实验中,训练样本与测试样本均为随机抽取。

表 1 实验样本

Table 1 Experimental samples

| Heart Rhythm Type                | Training Sample Number | Testing Sample Number |
|----------------------------------|------------------------|-----------------------|
| Normal Sinus Rhythm( Normal )    | 50                     | 20                    |
| Ventricular Premature Beat (VPB) | 50                     | 20                    |
| Atrial Premature Beats( APB )    | 50                     | 20                    |

### 1.2 方法

**1.2.1 样本熵** 样本熵的物理含义表示非线性动力学系统产生新信息的速率,是一种时间序列复杂性测度方法,与心律失常的心电信号产生方式的机理相一致,其具体计算过程如下所述<sup>[3-7]</sup>:

假设长度为  $N$  的时间序列  $x(n), n=1,2,\dots,N$ ,在嵌入维数(用来进行比较的时间序列长度)选择为  $m$  的条件下,可以获得一组  $m$  维矢量,表示为

$$X(i)=[x(i),x(i+1),\dots,x(i+m-1)] \quad (1)$$

其中,  $i=1,2,\dots,N-m+1$ 。

将式(1)中所示的一组  $m$  维矢量中的任意两个矢量  $X(i)$ 与  $X(j)$ 之间的距离定义为最大坐标差,表示为

$$d[X(i),X(j)]=\max [ |x(i+k)-x(j+k)| ] \quad (2)$$

式中,  $k=1,2,\dots,m-1, i,j=1,2,\dots,N-m+1$ 。

对于预先设置的相似容限  $r$ ,对式(1)中  $m$  维矢量中的第  $i$  个向量与其余  $N-m$  个向量之间的距离小于相似容限的个数进行计数,并与  $N-m$  的相除,得到

$$C_i^m = \begin{cases} \frac{\sum_{j=1, j \neq i}^{N-m} \Theta(r-d)[X(i),X(j)]}{(N-m)}, & i \leq N-m \\ 0, & i > N-m \end{cases} \quad (3)$$

式中,  $\Theta$  为 Heaviside 函数,定义为

$$\Theta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (4)$$

分别对式(1)中表示的  $N-m+1$  个向量分别计算  $C_i^m$ ,并计算在相似容限为  $r$  的条件下所有  $C_i^m(r)$ 的均值,记作

$$C^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} C_i^m(r) \quad (5)$$

再将嵌入维数设为  $m+1$ ,并重复以上计算过程得到  $C^{m+1}(r)$ 。综上所述,时间序列  $x(n)$ 的样本熵定义为

$$\text{SampEn}(m,r) = \lim_{N \rightarrow \infty} [-\ln \frac{C^{m+1}(r)}{C^m(r)}] \quad (6)$$

当时间序列的长度  $N$  取有限值时,式(6)将转化为

$$\text{SampEn}(m,r,N) = -\ln \frac{C^{m+1}(r)}{C^m(r)} \quad (7)$$

由式(7)可知,时间序列  $x(n)$ 的样本熵值与嵌入维数  $m$ 、相似容限  $r$  及时间序列的长度  $N$  有关。一般地,参考近似熵, $m$  取 1 或 2,  $r$  取 0.1~0.25 倍,  $N$  取  $10^m \sim 3^m$ , 其中  $\text{std}$  表示时间序列  $x(n)$ 的标准差。

**1.2.2 基于 kd 树的快速样本熵算法** 通过样本熵的计算过程

可知,该算法的时间复杂度主要由  $n_i^m = \sum_{j=1}^{N-m} \Theta(r-d[X(i),X(j)])$  的计算复杂度决定。文献<sup>[8]</sup>表明,可以将  $n_i^m$  的计算转化为正交范围搜索问题。正交范围搜索问题的简要介绍如下<sup>[9,10]</sup>:

将式(1)中  $m$  维矢量  $X(i)$  记作  $(X_i)_m$ ,对于每个  $i(i=1,2,\dots,N-m+1)$ ,  $(X_i)_m$  可以转化为一个  $m$  维点集  $P_i(x_i, y_i, z_i, \dots)$ , 即

$$x_i = X_i, y_i = X_{i+1}, z_i = X_{i+2}, \dots \quad (8)$$

那么,  $n_i^m$  等于边界框  $W_i$  内部包含的点集  $P_i$  中点的个数,  $W_i$  表示为

$$W_i = [(x_{LB})_i:(x_{UB})_i] \times [(y_{LB})_i:(y_{UB})_i] \times [(z_{LB})_i:(z_{UB})_i] \times \dots \quad (9)$$

其中,  $LB$  和  $UB$  分别表示边界框  $W_i$  的上边界和下边界,表示为

$$\begin{aligned} (x_{LB})_i &= x_i - r, (x_{UB})_i = x_i + r \\ (y_{LB})_i &= y_i - r, (y_{UB})_i = y_i + r \\ (z_{LB})_i &= z_i - r, (z_{UB})_i = z_i + r \end{aligned} \quad (10)$$

给定一个在  $d$  维空间的正交范围(边界框),怎样获取每个正交范围内部的空间点的个数为几何领域中的正交范围搜索问题<sup>[16]</sup>。因此,对于每个点集  $P_i$  与其对应的边界框,  $W_i, n_i^m$  (或  $n_i^{m+1}$ ) 的计算等价于解决  $m$  (或  $m+1$ ) 维正交范围计数问题。一旦  $n_i^m$  和  $n_i^{m+1}$  被计算出来,样本熵  $\text{SampEn}(m,r,N)$  可以通过式(7)直接计算出来。

kd 树能够用于解决正交范围搜索问题,是一种可以用于计算样本熵的高效算法。kd 树的基本原理是将需要进行索引的点集存储于特殊设计的二叉树形数据结构中。对于给定的边界框, kd 树对于给定的查询点和查询距离阈值,能够快速地从数据集中查询所有与查询点距离小于阈值的数据个数。

**1.2.3 多尺度快速样本熵** 多尺度分析是由 Costa 提出的在不同尺度下对原始时间序列进行复杂性估计的算法<sup>[11,12]</sup>。多尺度样本熵算法由以下两个步骤组成<sup>[13,14]</sup>。

(1) 对于给定的时间序列  $x(k), k=1,2,\dots,N$ , 尺度  $i$  的粗粒化

过程通过计算原始时间序列  $x(k)$  中连续而不重叠的  $i$  个元素的平均值获得的。粗粒化的时间序列表示为,

$$y(r)i = \frac{1}{\tau} \sum_{j=i\tau+1}^{(i+1)\tau} x_j, 1 \leq j \leq N/\tau \quad (11)$$

可以看出,当尺度  $i=1$  时,粗粒化的时间序列与原始时间序列  $x(k)$  相等。粗粒化后的时间序列  $\{y^{(i)}\}$  的长度为  $N/\tau$ 。如图 1 所示为粗粒化过程示意图。

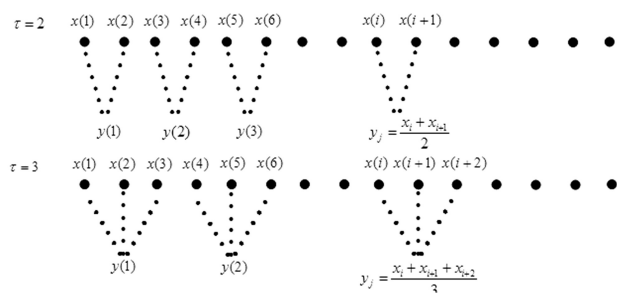


图 1 基于 kd 树的多尺度快速样本熵算法流程

Fig.1 Schematic illustration of the coarse-graining procedure

(2) 针对于步骤(1)中获得的每个不同尺度下的粗粒化时间序列,计算样本熵值。

为了提高多尺度熵的计算效率,将基于 kd 树的快速样本熵算法引入多尺度熵,得到基于 kd 树的多尺度快速样本熵算法,以提高心电图特征提取的实时性,其伪代码如图 2 所示。

可以看出,基于 kd 树的多尺度快速样本熵算法在多尺度熵的任意一尺度下,其计算时间复杂度为  $O(N N^{1-(1/d)})$ ,如表 2 所示。例如,当  $m=2$  时,定义中的样本熵计算复杂度为  $O(N^2)$ ,而基于 kd 树的多尺度快速样本熵算法的计算复杂度为  $O(N^{5/3})$ 。

```
Input: 时间序列  $X$ , 嵌入维数  $m$  与噪声容限  $r$ 
Output: 时间序列  $X$  的多尺度样本熵值  $SampEn(m, r, N)$ 
1 for scale = 1 : maxScale do
2    $X_r = \text{coarseGrainSeries}(\text{scale}, X)$ ; % 生成粗粒度时间序列
3    $\text{pointArray} = \text{transformToSpace}(m, X_r)$ ; % 将时间序列转化为空间点集
4   for k = m - 1 : m do
5      $d = m + 1$ ;
6   end
7    $\text{tree} = \text{buildKDTree}(d, \text{pointArray})$ ; % 建立 Kd 树结构
8   for i = 1 : N - m do
9      $W_i = \text{findBoundingBox}(m, r, \text{PointArray}[i])$ ; % 寻找边界框
10     $n_i^m = \text{KDsearch}(W_i, \text{tree})$ ;
11    if  $k == m - 1$  then
12       $n_n(i) = n_n(i) + n_i^m$ ;
13    end
14    else
15       $n_d(i) = n_d(i) + n_i^m$ ;
16    end
17  end
18   $SampEn(m, r, N) = -\ln(\frac{n_n}{n_d})$ ;
19 end
```

图 2 基于 kd 树的多尺度快速样本熵算法的伪代码

Fig.2 Pseudocode of multiscale fast sample entropy algorithm based on kd tree

**1.2.4 随机森林分类器** 美国科学院院士 Breiman 详细地对随机森林进行了数学推导与证明,并给出了随机森林的定义:由多个决策树  $\{h(x, \Theta_k), k=1, 2, \dots, n\}$  组成的分类器,其中  $\{\Theta_k\}$  是相互独立且同分布的随机向量,分类结果通过所有决策树的分类

结果,结合投票原则确定<sup>[15]</sup>。随机森林分类原理如图 3 所示,针对原始训练样本,利用 Bootstrap 方法获得  $n$  个 Bootstrap 子样本;对于每个 Bootstrap 子样本生成与之对应的  $n$  个 CART 决策树;在测试过程中,利用  $n$  个决策树的分类结果进行投票以确定测试样本的类型。本文提出的方法利用随机森林分类器对常见心律失常信号进行识别。

表 2 快速样本熵各部分的计算时间复杂度

Table 2 The computation time complexity of each part of the fast sample entropy

| Computational Process | Time Complexity     |
|-----------------------|---------------------|
| Storage               | $O(N)$              |
| Construct Tree Time   | $O(N \log N)$       |
| Search Time           | $O(N^{1-(1/d)})$    |
| Overall Process       | $O(N N_{-1/(1/d)})$ |

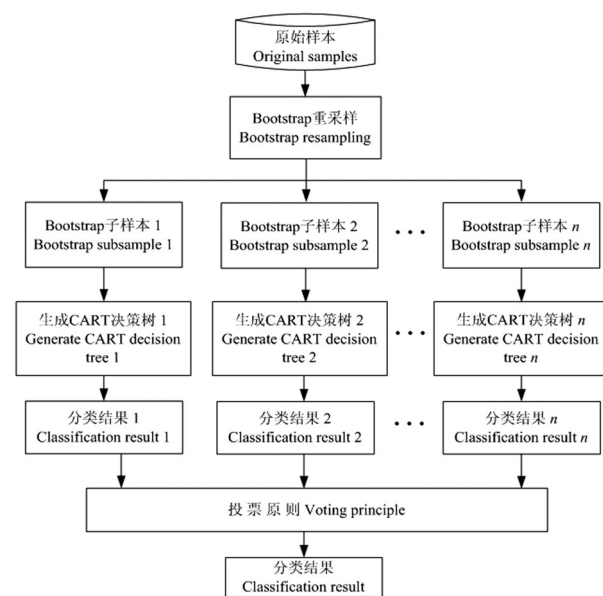


图 3 随机森林分类原理

Fig.3 Random forest classification principle

## 2 结果

### 2.1 基于 kd 树的心电信号快速样本熵计算

图 4 所示为 MIT-BIH 心律失常数据库中的一个正常心律的 ECG 样本。图 5 所示为  $m=1, r=0.25\text{std}$  时,基于 kd 树对给定点  $P_i$  确定  $n_i^1$  的过程。可见,基于 kd 树的快速样本熵计算方法将样本熵定义中的复杂迭代运算转化为正交范围搜索问题,通过范围内计数结果就能够获得样本熵值,提高计算效率。特别地,对于多尺度熵中需要对不同尺度下的心电信号进行分析时,基于 kd 树的快速样本熵能够显著提高计算速度。

### 2.2 多尺度样本熵特征提取

表 3 给出了部分正常、室性早搏、房性早搏的 ECG 信号经多尺度样本熵提取的特征向量(样本序号 10, 11)。特征值  $d_1, d_2, \dots, d_8$  为 ECG 信号在不同尺度下经过样本熵计算后获得的样本熵值。可见,对于正常、室性早搏、房性早搏经过多尺度样本熵提取的特征向量  $[d_1, d_2, \dots, d_8]$  具有比较明显的差异。因此,多尺度样

本熵能够有效提取心律失常状态下的 ECG 信号特征, 并可以利用多尺度样本熵特征向量进一步实现对于心电信号的分析。

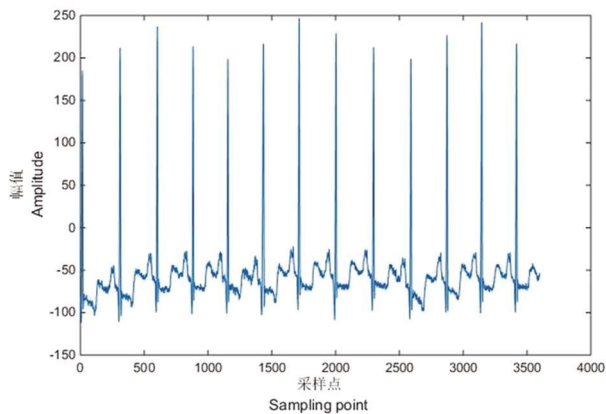


图 4 正常心律的 ECG 样本(MLII 导联)

Fig.4 ECG sample of normal heart rhythm (MLII lead)

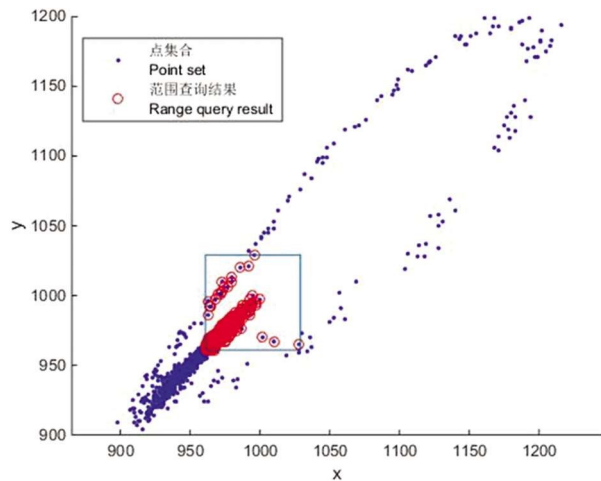


图 5 基于 kd 树确定样本熵的过程

Fig.5 The process of determining of sample entropy based on tree

表 3 部分 ECG 样本的特征向量

Table 3 The eigenvectors of some ECG samples

| Heart Rhythm Type | Sample Number | Feature Vector $[d_1, d_2, \dots, d_8]$   |
|-------------------|---------------|-------------------------------------------|
| Normal            | 10            | [0.45 0.42 0.41 0.39 0.34 0.30 0.29 0.27] |
|                   | 11            | [0.47 0.44 0.43 4.00 0.36 0.33 0.31 0.29] |
| VPB               | 10            | [0.24 0.19 0.19 0.18 0.17 0.16 0.13 0.12] |
|                   | 11            | [0.22 0.17 0.17 0.16 0.15 0.15 0.13 0.11] |
| APB               | 10            | [0.18 0.14 0.14 0.13 0.12 0.12 0.11 0.11] |
|                   | 11            | [0.19 0.15 0.15 0.13 0.13 0.12 0.11 0.11] |

### 2.3 决策树数量确定

图 6 所示,为通过本文实验样本获得的随机森林分类器中生成的决策树的数量与袋外(out-of-bag, OOB)分类误差之间的关系。可见,在生成的决策树数量大约为 250 时,袋外分类误差趋于平稳,因此为了保证分类的准确率较为稳定,随机森林分类器中的设置决策树的数量应大于 250。如图 7 所示,为本文生成的随机森林分类器中的一棵决策树,可以看出每棵决策树都是利用二叉树对正常心律、室性早搏(VPB)与房性早搏

(APB)进行分类的。随机森林分类器正是利用多个决策树的分类结果并结合投票原理实现多分类。

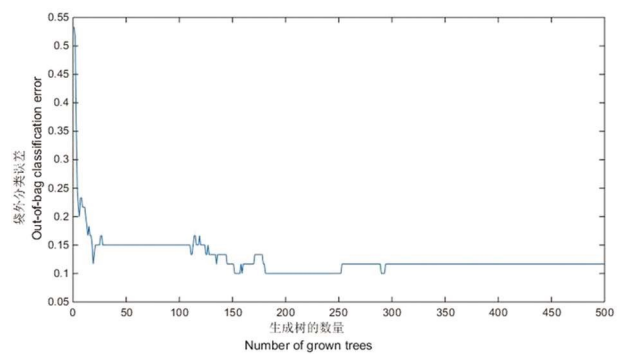


图 6 决策树数量与袋外分类误差的关系

Fig.6 The relationship between the number of decision tree and out-of-bag classification error

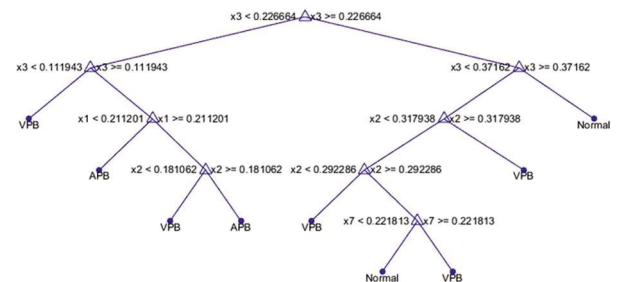


图 7 文中随机森林生成的决策树

Fig.7 The decision tree generated by random forest in this paper

### 2.4 不同尺度特征对分类器决策的贡献

本文中的特征向量是由 ECG 信号在 8 个尺度下的样本熵值组成,每个特征编号对应一个尺度。如图 8 所示为特征编号对于随机森林分类的重要性,即 ECG 信号经过多尺度样本熵特征提取后的特征向量中各元素对分类器决策的贡献大小。

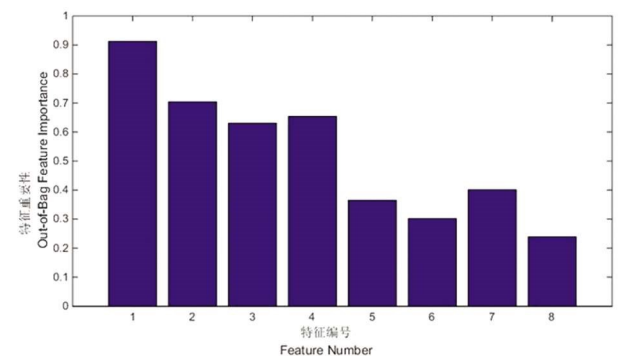


图 8 特征对于随机森林分类的重要性

Fig.8 The importance of feature numbers for random forest classification

### 2.5 提出方法的平均识别准确率

表 4 所示为本文提出的心电信号分析方法对心律失常的 ECG 信号样本的分析结果,平均准确率为经过 50 次独立实验的统计结果。可见,本文提出的基于多尺度快速样本熵与随机森林的心电信号分析方法能够有效地对正常心律与典型的心律失常(VPB、APB)进行识别,具有较高的识别准确率,平均识别准确率达到 91.60%。

表 4 本文方法的心电信号分析结果

Table 4 The results of electrocardiogram analysis method in this paper

| Heart Rhythm Type | Average Accuracy(%) |
|-------------------|---------------------|
| Normal            | 94.0                |
| VPB               | 85.0                |
| APB               | 94.0                |

### 3 讨论

目前,对于心律失常疾病的诊断依然是通过医生的临床经验对心电图进行分析得到的。随着医疗辅助诊断技术的发展,利用计算机技术对心电图进行自动诊断已经成为了生物医学领域的研究趋势和热点问题。

近年来,相关学者研究发现健康人与心律失常患者的心电信号的复杂性存在差异,并证明了心电信号的复杂程度能够反映心脏的健康状况<sup>[3,4]</sup>。基于复杂性量化的心电信号分析得到了研究人员的广泛关注,并取得了一定的研究成果<sup>[5]</sup>。Pincus 于 1991 年提出了近似熵 (Approximate Entropy, ApEn) 实现对生理信号复杂性的量化估计<sup>[6]</sup>。2002 年 Richman 在近似熵的基础上提出了样本熵 (Sample Entropy, SampEn), 其物理意义表示系统新信息产生的速率,适用于对生理信号复杂性的分析<sup>[7]</sup>。由于基于单一尺度的传统方法不能解释蕴含在生理系统中的多时间尺度特征, Costa 等人于 2002 年率先提出并采用多尺度熵 (Multiscale Entropy, MSE) 进行生物医学信号分析<sup>[11,12]</sup>。王俊等人将多尺度熵应用于心电图分析, 研究表明 MSE 的波动范围情况可以较为有效地揭示心脏的健康状况<sup>[13]</sup>。同时,王俊等人利用多尺度熵重点研究了 ECG 的 ST 段, 并指出 MSE 能够用于区分健康人与冠心病人的心电图, 适用于早期临床诊断<sup>[14]</sup>。Wang 等利用多尺度熵曲线对心电信号进行模式识别, 研究表明心电信号的复杂性将随着年龄的增长出现下降的趋势<sup>[17]</sup>。尽管现有研究揭示了 ECG 多尺度熵的变化能够反映心脏的健康状况, ECG 多尺度熵的均值随着心脏健康的恶化呈下降趋势, 但是, 利用 ECG 多尺度熵作为特征实现心律失常疾病的自动诊断研究相对较少。同时, 多尺度熵由于需要对多个尺度下的心电信号进行样本熵计算, 计算量大, 导致自动诊断实时性较差。特别地, 24 h 动态心电图的临床分析对心电图分析方法的实时性要求迫切<sup>[22,23]</sup>。鉴于此, 本文提出了基于多尺度快速样本熵与随机森林相结合的心电信号分析方法。该方法利用多尺度熵对心电信号进行多尺度特征提取, 描述心电信号的病理特征, 再利用随机森林分类器对心脏疾病进行有效识别, 实现常见心律失常 (房性早搏、室性早搏) 的自动诊断。

实验表明, 本文采用基于 kd 树的多尺度快速样本熵方法, 将样本熵计算中的反复迭代过程转化为正交范围搜索问题, 显著降低了样本熵的计算复杂度 (当  $m=2$  时, 样本熵的计算复杂度由  $O(N^2)$  降低为  $O(N^3)$ ), 解决 MSE 在心电信号特征提取过程中的计算效率偏低的问题。通过对 MIT-BIH 心律失常数据库中的心电信号样本的特征提取, 房性早搏和室性早搏的多尺度熵特征向量较正常心律的特征向量出现下降趋势, 这与相关研究获得的结论一致<sup>[5,12]</sup>。另外, 根据心律失常疾病类型的不同,

多尺度熵特征向量存在一定的差异性, 本方法正是利用此间的差异实现对正常心律、房性早搏和室性早搏进行有效地识别。

利用多尺度快速样本熵提取的信号特征构建随机森林分类器, 随机森林分类器通过多个决策树的分类结果并结合投票原理实现 ECG 自动诊断。通过实验可以看出, 本文提出的心电图分析方法能够实现对正常心律、房性早搏和室性早搏的有效识别, 平均识别准确率达到 91.60%, 其中对于正常心律和房性早搏的识别准确率都在 94.0% 左右, 而室性早搏的准确率为 85.0%。由于实验样本存在干扰噪声和局部范围内的多种心律失常信号的混叠, 导致室性早搏的识别准确率相对较低。相比于人工神经网络, 随机森林分类器不需考虑收敛速度过慢及过拟合等问题<sup>[18]</sup>。相对于 SVM 分类器, 随机森林分类器不用考虑二分类组合及多个参数优化调整<sup>[19,24,25]</sup>, 就能获得理想的多分类结果。另外, 随机森林分类器对于特征向量的维数没有限制, 不需要预先对样本进行降维处理, 能够对大量特征进行准确分类<sup>[20,21]</sup>。

综上所述, 本文提出的多尺度快速样本熵与随机森林相结合的心电图分析方法是一种有效的 ECG 自动诊断方法, 为相关研究提供了可行性的思路, 一定程度上解决了当前自动诊断方法的实时性和准确性问题。

目前, 本研究的实验样本仅为 MIT-BIH 心律失常数据库中的室性早搏和房性早搏样本, 下一步将扩充心律失常的疾病种类, 实现更多心律失常的自动诊断, 并研究心电信号去噪方法, 进一步提升算法性能及实用性。

### 参考文献 (References)

- [1] Mitra M, Samanta, R K. Cardiac Arrhythmia Classification Using Neural Networks with Selected Features [J]. Procedia Technology, 2013: 76-84
- [2] Moody G B, Mark R G. The impact of the MIT-BIH Arrhythmia Database [J]. IEEE Engineering in Medicine & Biology Magazine, 2002, 20(3): 45-50
- [3] Pincus S M. Assessing serial irregularity and its implications for health [J]. Annals of the New York Academy of Sciences, 2001, 954(1): 245
- [4] Lake D E, Richman J S, Griffin M P, et al. Sample entropy analysis of neonatal heart rate variability [J]. Am J Physiol Regul Integr Comp Physiol, 2002, 283(3): R789
- [5] Costa M, Healey J A. Multiscale entropy analysis of complex heart rate dynamics: discrimination of age and heart failure effects[C]//Computers in Cardiology. Thessaloniki-Chalkidiki, Greece: IEEE, 2003: 705-708
- [6] Pincus S M. Approximate entropy as a measure of system complexity [J]. Proceedings of the National Academy of Sciences of the United States of America, 1991, 88(6): 2297
- [7] Richman J S, Moorman J R. Physiological time-series analysis using approximate entropy and sample entropy [J]. American Journal of Physiology Heart & Circulatory Physiology, 2000, 278(3): 2039-2049
- [8] Pan Y H, Lin W Y, Wang Y H, et al. Computing Multiscale Entropy with Orthogonal Range Search[J]. Journal of Marine Science & Technology, 2011, 19(1): 107-113
- [9] Pan Y H, Wang Y H, Liang S F, et al. Fast computation of sample entropy and approximate entropy in biomedicine[J]. Computer Methods & Programs in Biomedicine, 2011, 104(3): 382-396

- [10] Manis G. Fast computation of approximate entropy [J]. Computer Methods & Programs in Biomedicine, 2008, 91(1): 48
- [11] Costa M, Goldberger A L, Peng C K. Multiscale entropy analysis of complex physiologic time series[J]. Physical Review Letters, 2002, 89 (6): 0681021-0681024
- [12] Costa M, Goldberger A L, Peng C K. Multiscale entropy analysis of biological signals [J]. Physical review E: Statistical, Nonlinear, and Soft Matter Physics, 2005, 71(2): 021906
- [13] 王俊,宁新宝,马千里.基于多尺度熵的心电图分析[J].中国生物医学工程学报, 2008, 27(3): 331-334  
Wang Jun, Ning Xin-bao, Ma Qian-Li. Multiscale Entropy Based on Electrocardiogram Analysis [J]. Chinese Journal of Biomedical Engineering, 2008, 27(3): 331-334
- [14] 王俊,马千里.基于多尺度熵的心电图 ST 段研究[J].南京邮电大学学报(自然科学版), 2008, 28(3): 70-72  
Wang Jun, Ma Qian-Li. The Study of the Electrocardiogram's ST Segment Based on Multiscale Entropy [J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science), 2008, 28 (3): 70-72
- [15] Breiman L. Random Forests[J]. Machine Learning, 2001, 45(1): 5-32
- [16] 陈寅生. MOS 气体传感器阵列的自确认方法研究[D].哈尔滨: 哈尔滨工业大学, 2017  
Chen Yin-sheng. Research on Self-validating Methods for Metal Oxide Semiconductor Gas Sensor Arrays [J]. Harbin: Harbin Institute of Technology, 2017
- [17] Wang C C, Chang C D, Jiang B C. Pattern Recognition of Multiscale Entropy Curve for ECG Signal Analysis [J]. Applied Mechanics & Materials, 2011, 195-196: 603-607
- [18] Al-Nashash H. Cardiac arrhythmia classification using neural networks [J]. Technology & Health Care Official Journal of the European Society for Engineering & Medicine, 2000, 8(6): 363
- [19] Knyvs R, Dhuli R. Classification of ECG heartbeats using nonlinear decomposition methods and support vector machine[J]. Computers in Biology & Medicine, 2017, 87: 271
- [20] Zhang Y, Wei S, Zhang L, et al. Comparing the Performance of Random Forest, SVM and Their Variants for ECG Quality Assessment Combined with Nonlinear Features[J]. Journal of Medical & Biological Engineering: 1-12
- [21] Alickovic E, Subasi A. Medical Decision Support System for Diagnosis of Heart Arrhythmia using DWT and Random Forests Classifier [J]. Journal of Medical Systems, 2016, 40(4): 108
- [22] 谢福生,夏中华,刘琼,等.24 h 动态心电图对冠心病心律失常的临床监测价值[J].现代生物医学进展, 2016, 16(29): 5736-5738  
Xie Fu-sheng, Xia Zhong-hua, Liu Qiong. Clinical Value of 24h Dynamic Electrocardiogram on Monitoring Coronary Heart Disease with Arrhythmia [J]. Progress in Modern Biomedicine, 2016, 16 (29): 5736-5738
- [23] 肖姗姗,唐冰,李毅,等.动态心电图与常规心电图诊断冠心病患者心律失常的效果比较 [J]. 现代生物医学进展, 2017, 17(14): 2663-2665  
Xiao Shan-shan, Tang Bing, Li Yi. Comparison of Efficiency of Dynamic Electrocardiogram and Routine Electrocardiogram in the Diagnosis of Arrhythmia in Patients with Coronary Heart Disease [J]. Progress in Modern Biomedicine, 2017, 17(14): 2663-2665
- [24] Acır N. A support vector machine classifier algorithm based on a perturbation method and its application to ECG beat recognition systems [J]. Expert Systems with Applications, 2006, 31(1): 150-158
- [25] Asl B M, Setarehdan S K, Mohebbi M. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal[J]. Artificial Intelligence in Medicine, 2008, 44(1): 51-64

(上接第 3452 页)

- [24] Yuan Q, Wu J, Qin C, et al. Spin-coating synthesis and characterization of Zn-doped hydroxyapatite / polylactic acid composite coatings [J]. Surface & Coatings Technology, 2016, 307: 461-469
- [25] Nagata F, Miyajima T, Kato K. Preparation of phylloquinone-loaded poly(lactic acid)/hydroxyapatite core-shell particles and their drug release behavior [J]. Advanced Powder Technology, 2016, 27 (3): 903-907
- [26] Rogina A, Pribolšan L, Hanžek A, et al. Macroporous poly(lactic acid) construct supporting the osteoinductive porous chitosan-based hydrogel for bone tissue engineering[J]. Polymer, 2016, 98: 172-181
- [27] He S, Lin K F, Sun Z, et al. Effects of Nano-hydroxyapatite/Poly (DL-lactic-co-glycolic acid) Microsphere- Based Composite Scaffolds on Repair of Bone Defects: Evaluating the Role of Nano-hydroxyapatite Content[J]. Artificial Organs, 2016, 40(7): E128-E135
- [28] Liu Y, Jiang D. Effect of bone-like hydroxyapatite/poly amino acid loaded with rifapentine microspheres on bone and joint tuberculosis in vitro[J]. Cell Biology International, 2017, 41(4): 369
- [29] Ginjupalli K, Averineni R K, Shavi G V, et al. Biodegradable composite scaffolds of poly (lactic-co-glycolic acid) 85:15 and nano-hydroxyapatite with acidic microclimate controlling additive [J]. Polymer Composites, 2017, 38(6): 1175-1182
- [30] Liu C, Wong H, Yeung K, et al. Novel Electrospun Polylactic Acid Nanocomposite Fiber Mats with Hybrid Graphene Oxide and Nanohydroxyapatite Reinforcements Having Enhanced Biocompatibility[J]. Polymers, 2016, 8(8): 287