基于改进遗传模拟退火K-means的心电波形的分类研究*

何云斌†，张晓瑞，万 静，李 松
（哈尔滨理工大学 计算机科学与技术学院，哈尔滨 150080）

摘 要： 针对心电图自动诊断困难这一问题，提出了一种新的聚类算法：基于均方差属性加权的遗传模拟退火K-means改进聚类算法，用于改进心电图（ECG）信号的自动识别技术。利用小波变换的多分辨率和抗干扰能力好的特点，检测QRS波、P波、T波，提高了特征检测的准确性；利用聚类分析具有较好的鲁棒性和适合于大数据量分析的特点，对心电信号进行波形分类。采用MIT-BIH标准心电数据库中的部分数据对识别结果进行判断，改进后的K-means聚类算法的准确率高于传统的K-means聚类算法，实验表明该算法对心电信号可以进行有效分类。

关键词： 心电图信号；聚类；特征提取；K-means；遗传算法；模拟退火；属性权重；均方差；小波变换
中图分类号： TP391.4 **文献标志码：** A **文章编号：** 1001-3695（2014）11-3328-05
doi:10.3969/j.issn.1001-3695.2014.11.029

Research of ECG waveforms classification based on improved genetic simulated annealing K-means

HE Yun-bin†, ZHANG Xiao-rui, WAN Jing, LI Song
(*School of Computer Science & Technology, Harbin University of Science & Technology, Harbin* 150080, *China*)

Abstract: In view of the difficulties to recognize ECG signal automatically, this paper presented a new clustering algorithm, which was proposed based on the MSE attribute weights genetic simulated annealing to improve K-means clustering algorithm, in order to improve the ECG signal automatic identification technology. It used wavelet transform and multi-resolution and good anti-jamming capability to detect QRS complex, P wave, T wave, improved the accuracy of feature detection. Because of the cluster method had more robust and suitable for large data volume analysis, it classified the ECG signals by using this method to analyze large data volume. It adopted the parts of data from the MIT-BIH standard ECG database to judge the result of the identification. The improved K-means clustering algorithm is more accurate than the traditional K-means clustering algorithm, experiments indicate that this algorithm is effective and accurate to classify ECG signals.

Key words: ECG signal; clustering; feature extraction; K-means; genetic algorithms; simulated annealing; attribute weights; MSE; wavelet transform

0 引言

心电图（ECG）信号的计算机辅助诊断对严重心脏病患者的治疗起着重要作用，通过对ECG各个特征段和特征点进行统计分析，就可对受检测者的心脏活动状况进行诊断[1]。但是ECG受被检测者的心脏状态和被检测者的运动状况这两个因素影响。对于不同的人或同一个人不同时刻采集到的ECG通常存在很大的差异性，这种差异性大大增加了ECG的分类难度。

关于ECG的分类方法有基于硬件和基于软件两种。采用硬件的方法因其检测参数不容易调整，缺乏灵活性，不易处理复杂的情况已经很少使用。软件的分类方法有许多，如葛丁飞等人[2]提出的可变阈值检测方法，ECG经过数字带通滤波器滤波，提取特征之后用阈值进行检测，阈值的大小随信号的波动不断调整，这样可提高检测的可靠性。此外，该算法还利用双重阈值去重新检测。这种方法虽然对QRS波具有比较高的检测率，但对其他分量考虑较少，并且该方法基本是依靠经验判断，没有明确的数学模型。Wang等人[3]提出了利用不确定性推理的方法用于ECG诊断，牟善玲等人[4]提出的范围覆盖方法也是基于这个原理。该方法的缺点是表达的知识有限，知识库无法调整。结构模式识别大多是利用时域特征进行结构分析，Martis等人[5]提出用语音信号处理中常用的高斯混合模型分析诊断ECG，蒋德育等人[6]提出用隐马尔可夫模型完成正常和异常心电信号的分类。由于该方法需要借助领域知识，故表示结构单一。刘彤彤等人[7]提出用带纠错输出编码的SVM进行心电信号分类，马永杰等人[8]提出遗传算法分类

收稿日期：2013-10-23；修回日期：2013-12-03 基金项目：黑龙江省教育厅科学技术研究项目（12511100）；黑龙江省自然科学基金资助项目（F201014，F201134）

作者简介：何云斌（1972-），男（通信作者），教授，硕导，博士，主要研究方向为数据库理论与应用、时空数据库、嵌入式系统（hybha@163.com）；张晓瑞（1990-），女，硕士，主要研究方向为空间数据挖掘；万静（1972-），女，教授，硕导，博士，主要研究方向为数据库理论及应用；李松（1977-），男，副教授，博士，主要研究方向为空间数据库理论及应用.

ECG，但泛化能力不强。神经网络法解决了某些经典人工智能的难题，Ubeyli[9] 提出用粒子群优化的神经网络解决 ECG 分类问题，Yu 等人[10] 用基于块的神经网络分析插值系数和时域特征。由于难以确立最优的网络结构，所以该类方法的环境适应力不高，要通过临床上由经验丰富的医生作回顾性分析来处理。

针对上述问题，为了提高 ECG 分类效率，本文利用小波变换来提取心电信号特征值，并根据心电信号特征值的特点提出了一种基于均方差属性加权的改进遗传模拟退火 K-means 算法。应用此方法对 ECG 特征值进行聚类，在一定程度上可以改善心电分析算法，为心电实时监护的临床应用提供参考。

1　提取特征值

心电图信号是由 R 波、QRS 波群、P 波、T 波、ST 段等组成，R 波、QRS 波群对诊断有十分重要的意义[11]。所以 R 波等的特征提取是整个算法中重要的一环，ECG 聚类算法需要提供特征值输入，本文采用小波变换进行特征提取，从数据点中提取到特征值，然后求方差，组成特征矢量矩阵，提供给聚类算法。

为了模拟大样本、高维特征、数据高度分布不均衡的数据集环境，并要求提供准确的数据划分注释，以检验算法的准确性，本文选用 MIT-BIH 心电数据库作为分类对象，特征提取就是要在 MIT-BIH 数据库中提取心电信息的基本特征。

1.1　特征提取——R 波的检测

心电图具有同类间的个体差异性和不同类间的个体差异性，因此很难用统一的维度定义 R-R 间期，也不能用确定的阈值定义不同人的 R 波峰高度。为了提高 R 波的检测精度，本研究采用小波变换将 ECG 信号 $f(t)$，经过线性变换分解到不同尺度，为了使小波变换具有更好的时频局域化特性，采用二进离散小波变换：$W_f(2^j,\tau)=\frac{1}{\sqrt{2^j}}\int_{-\infty}^{\infty}f(t)\psi\left(\frac{t-\tau}{2^j}\right)\mathrm{d}t$。其中：ECG 信号 $f(t)$，对于给定的母小波 $\psi(t)$ 在时域和频域上进行一系列的伸缩和平移，构成了一组小波函数 $\psi_{s,b}(t)=\frac{1}{\sqrt{s}}\times\psi\left(\frac{t-\tau}{s}\right)\mathrm{d}t$，令 $s=2^j$，形成二进小波变换。

小波变换中信息的奇异点与其小波变换的一个正模极大值和负模极大值对相对应，其位置是正、负模极大值的过零点，基于小波变换的模极大值 $\max\{W_f(2^j,\tau)\}$，当计算的信号模大于这一阈值时，就判定为 QRS 波群。同时，这个阈值也是随着计算结果自适应更新的。在判定为 QRS 波群后，再检测过零点则可判定具体的 R 波位置。

1.2　特征提取——QRS 波群的检测

R 波峰值位置可以为其他波的检测提供重要依据[12]，所以完成 R 波检测后，查找距 R 波峰值前 250 ms 到 150 ms 处信号上的局部极大值点，即 P 波。T 波可以通过查找距 R 波峰值后 250 ms 到 500 ms 的局部极大值点确定。依照此方法对 R-R 间期波形进行特征提取，用于心电波形聚类分析。

本研究在此进行了一个小小的改进，可以减少高频噪声的误判：首先在 2^4 尺度寻找模最大值的位置，然后在该位置附近寻找 2^3 尺度上的模最大值，如果不是该尺度上的模最大值，该部分就不是 QRS 波群；依此类推，一直寻找到 2^1 尺度。这样，只有在 $2^1\sim2^4$ 四个尺度上都是模最大值的，才是 QRS 波群。

2　心电信号特征参数基于遗传模拟退火的 K-means 聚类

2.1　传统 K-means 聚类算法

K-means 算法是聚类算法中常用的算法之一[13]。其基本思想是通过不断迭代对样本数据进行划分，直至划分的结果不再发生变化，即误差平方和准则函数 E 的值达到最优。误差平方和准则函数定义为

$$E=\sum_{i=1}^{k}\sum_{x\in C_i}\|x_i-c_i\|^2 \tag{1}$$

E 越大说明簇内的相似度越低；E 越小说明簇内的相似度越高。但是 K-means 算法存在如下缺点：a）聚类结果易受到聚类质心选择问题干扰；b）分析方向单一，常常会出现局部最优解，造成孤立点等问题。

2.2　基于模拟退火 K-means 的改进算法

模拟退火算法是一种通过概率演算法在一个大的搜寻空间内找寻命题的最优解[14]。求解全局最优问题与物理固体退火具有相似性，该算法参数设计上采用的是物理领域的参数，假设温度 T 时粒子趋于平衡的概率是 $e^{-\frac{\Delta E}{(kT)}}$，其中，$k$ 是 Boltzmann 常数，ΔE 表示内能改变量，E 代表温度 T 时刻的内能，该算法基于 Metropolis 准则[15]，可控制温度由高到低的过程。模拟退火算法采用式（2）表示的退火方式。

$$T(t)=T_0\times\alpha^t \tag{2}$$

其中：α 为退火速度，控制温度下降的快慢，取 $\alpha'=0.99$。基于模拟退火算法具有描述简单、使用灵活、运用广泛、运行效率高和较少受到初始条件约束等优点，针对传统的 K-means 算法的缺点，本文采用模拟退火算法对其进行优化，提出基于模拟退火算法的 K-means 聚类算法。

经过分析，笔者认为模拟退火其实也是一种贪心算法，它的搜索过程引入了随机因素。模拟退火算法以一定的概率来接受一个比当前解要差的解，因此有可能会跳出这个局部的最优解，达到全局的最优解。模拟退火算法的局部搜索能力很强，可以使搜索过程避免陷入局部最优解，但是模拟退火方法对整个搜索空间的状况了解不是很多，导致全局的搜索能力较弱。这里，一定概率的计算参考了金属冶炼的退火过程[17]，很难得到满意的搜索效率。因此，本文提出基于遗传模拟退火算法的 K-means 的 ECG 改进聚类方法。

2.3　基于均方差属性加权的遗传模拟退火算法的 K-means ECG 改进聚类方法

遗传模拟退火算法是将传统的遗传算法和模拟退火算法相结合的一种优化算法。由于遗传算法的局部搜索能力较差，但是把握全局搜索的能力较强，模拟退火算法正好相反，它的局部搜索能力很强，可以使搜索过程避免陷入局部最优解，但是模拟退火算法却对整个搜索空间的状况了解得不是很多，导致全局的搜索能力较弱，很难得到满意的搜索效率。

为了改进 ECG 的分类，本文将遗传算法与模拟退火算法融合，产生一种混合遗传算法。遗传算法对编码串进行操作[18]，从中找到高适应度的个体。在解空间中从多点寻找，在很多区域中进行采样，大大减少了陷入局部解的可能性。模拟退火算法求得 K-means 算法初始的聚类数，把基本 K-means 聚类算法的聚类结果作为模拟算法的初始解，改进算法的局限性，从而使算法跳出局部最优解，达到全局最优解。这两种算

法吸取彼此的优点，弥补自身的不足，产生一种更优良的搜索算法，并引入了属性权重改进 K-means 算法。

2.3.1 权值计算

采用变量的均方差（RMSD）作为权值，RMSD 是反映随机变量离散程度常用的指标，对无法观察的参数进行估计，能够良好地反映 ECG 的数字特征。RMSD 是用来衡量 ECG 样本波动大小的量，RMSD 越大，数据的波动就越大。所以将 RMSD 作为 ECG 分类的特征权重。对于任意两个属性 c_i、c_j，如果 $\mathrm{RMSD}(c_i) > \mathrm{RMSD}(c_j)$，则 c_i 样本的波动较大，说明 c_i 属性在聚类中应占有更重要的地位。通过对 c_i 属性赋予较大权值来调整特征空间，可以更准确地反映类内相似度并得到更佳的聚类结果。本研究对小波变换提取的特征值计算属性均方差后归一化，作为权重，再进行后续算法。

2.3.2 ECG 染色体编码

为了简化译码工作并使算法具有直观性，本文采用整数编码。对于 $n \times n$ 的 ECG 特征矩阵 W，将其第 i 行的行向量 w_i（$1 \leqslant i \leqslant n$）作为遗传算法中的个体，其中元素 $-1 \leqslant w_{ij} \leqslant 1$（$1 \leqslant j \leqslant n$）。产生 n 个 $[-1,1]$ 的随机数作为初始种群的一个个体。相应的 ECG 基因和染色体具有如下形式：基因为 w_{ij}，染色体 w_i $[w_{i1}, w_{i2}, \cdots, w_{in}]$。

2.3.3 适应度函数设计

为了提高 ECG 分类算法效率，本研究采用轮盘赌选择方法对个体进行优胜劣汰操作，适应度高的 ECG 个体遗传到下一代群体的概率更大。采用与 K-means 算法相同的方式进行 ECG 聚类的划分并重新计算各聚类的中心，每次得到聚类划分后，替换原来的 ECG 聚类中心。然后以 E 作为准则函数，E 的数学表示式如式（1）。由式（1）可以看出，ECG 初始中心越精确，E 越小，w_{ij} 是染色体中的一位，K 个聚类中心就是染色体的 K 位，那么设适应度为 E，每条染色体长度为 Len（Chr）。因此染色体的适应度函数为

$$\mathrm{Fit}(\mathrm{Chr}) = \sum_{k=1}^{\mathrm{Len}(\mathrm{Chr})} \sum_{w_{ij} \in c_j} \frac{1}{E \times \exp(T_0)} \quad (T_0 \text{ 为初始温度}) \qquad (3)$$

遗传算法在处理 ECG 基因和染色体过程中依据适应度函数，利用高效的 ECG 种群的适应度值进行搜索，可以较大程度地提高遗传算法的收敛速度。

2.3.4 交叉操作

ECG 的染色体交叉操作通过交换两个父 ECG 个体的一部分来产生新的 ECG 子个体。本文采用单点交叉，交叉操作按照一定的交叉概率 p_c 进行，确定交叉点位置后，两个 ECG 染色体对应基因互换。ECG 染色体交叉操作如下：两个 ECG 个体 $w_i\{w_{i1}, w_{i2}, w_{i3}, \cdots, w_{in}, w_{is}\}$，$w_j\{w_{j1}, w_{j2}, w_{j3}, \cdots, w_{jn}, w_{js}\}$。假设它第 n 位 ECG 基因交叉，生成新的 ECG 染色体 $w_i'\{w_{i1}, w_{i2}, w_{i3}, \cdots, w_{jn}, w_{is}\}$，$w_j'\{w_{j1}, w_{j2}, w_{j3}, \cdots, w_{jn}, w_{js}\}$。

2.3.5 变异操作

跳出局部最优关键是变异操作，它是一种局部随机搜索，与交叉重组 ECG 染色体相结合可以保证遗传算法的有效性，使其既有局部随机搜索能力，又保持种群的多样性。

ECG 染色体变异操作如下：有一个 ECG 体 $w_i\{w_{i1}, w_{i2}, w_{i3}, \cdots, w_{in}, w_{is}\}$，假设它第 n 位变异，$N = 10$，随机产生值 w_0，则变异后 ECG 染色体变成 $w_i\{w_{i1}, w_{i2}, w_{i3}, \cdots, w_0, w_{is}\}$。

K-means 算法对初始聚类中心有很高的要求，初始聚类中心越精确，聚类效果越好。遗传模拟退火算法可以有效并精确地找到 ECG 初始聚类中心。查找初始聚类中心可以看做是遗传模拟退火算法在查找 ECG 种群内的最优个体，这个最优个体就是 K-means 聚类算法需要的初始聚类中心。本文提出将模拟退火思想引入遗传算法中，可以有效地缓解遗传算法的选择压力，避免陷入局部最优解；并对基因变异产生的新 ECG 个体，依照 Boltzmann 机制按照概率选择接受。

当算法执行到一定代数后，遗传算法容易产生一个适应度远大于其他个体的 ECG 优良个体 n，此 ECG 个体被选中的概率远远高于其他个体，这样就会造成子代 ECG 个体很多都是来源于 n，不可避免地造成子代 ECG 个体相似，也就是几乎所有 ECG 个体都集中在一起，过早陷入局部最优，很难跳出局部走向全局最优，应避免这种情况。虽然传统的变异操作可以防止这种情况，但是变异概率的值通常很小，需要很多代才能产生一个不同于其他个体的新 ECG 个体，而如果新 ECG 个体的适应度又非常小，那么它是很难被选中进行交叉操作的，也就很难产生新子代个体。所以当某一代出现群内 ECG 个体过分集中现象时，就大幅度提高个体的变异概率 p_m，使它有一个很大的值，让 ECG 个体变异的可能性增大。

设 $\Delta F_a = (\mathrm{Fit}(\mathrm{chr}_a') - \mathrm{Fit}(\mathrm{chr}_a))/T_0$，$\Delta\mathrm{Fit}(\mathrm{chr}) = \mathrm{Fit}(\mathrm{chr}') - \mathrm{Fit}(\mathrm{chr})$，任取三条染色体分别为

$\mathrm{chr}_x = [\mathrm{chr}_x(1), \mathrm{chr}_x(2), \cdots, \mathrm{chr}_x(k)]$

$\mathrm{chr}_y = [\mathrm{chr}_y(1), \mathrm{chr}_y(2), \cdots, \mathrm{chr}_y(k)]$

$\mathrm{chr}_z = [\mathrm{chr}_z(1), \mathrm{chr}_z(2), \cdots, \mathrm{chr}_z(\mathrm{add}_m), \cdots, \mathrm{chr}_z(k)]$

令 add_m 作为染色体变异点，$0 < \mathrm{add}_m < k$，令 num 为变异点个数，$0 < \mathrm{num} < k$，令 $p = \mathrm{e}^{\frac{\Delta\mathrm{Fit}(\mathrm{chr})}{t}}$。

基于均方差属性加权的遗传模拟退火算法的 K-means 聚类方法的整体过程如下：

算法 1 ECG（x_i，k）

输入：样本 x_i（$i = 1, \cdots, n$），初始聚类中心个数 k，种群规模 M，染色体 $\mathrm{chr}_1, \mathrm{chr}_2, \cdots, \mathrm{chr}_n$，$\mathrm{chr}_i = [\underbrace{x_1 \cdots x_j}_{k}]$。染色体长度终止进化代数 T，交叉概率 p_c，变异概率 p_m，$p_0, p_0' \in [0,1]$，$\mathrm{random} \in [1,n]$，模拟退火输入参数 T_0，α。

输出：当前最优聚类划分 C_{opt}。

```
begin
1 根据式(3)计算每条染色体的适应度 Fit(chr_i)，用适应度参数 fit_0
2 令 0 < add_c < k，add_c 作为染色体交叉点
if add_c ≠ add_c' then
{add_c = add_c';
  if p_c > p_0 then
  chr_x(a) ⇔ chr_y(a);
          计算染色体 chr_x'、chr_y' 适应度 Fit(chr_x')、Fit(chr_y')。
   if min{1, exp(ΔF_x)} > p_0 then
     chr_x = chr_x';
   else if min{1, exp(ΔF_y)} > p_0 then
     chr_y = chr_y';}
3 if add_m ≠ add_m' then
{add_m = add_m';
  if i < num then
     { if p_m > p'_0 then
          chr_x(add_m) ⇔ random;
          计算染色体 chr' 的适应度 Fit(chr_z')
            if min{1, exp(ΔF_z)} > p'_0 then
              chr = chr';
             else  i ++;
     }
  }
4 for i: = 1 to T do
  {if ΔFit(chr_z) < 0 then
     Fit(chr_z') 为最优目标函数值，保存当前染色体 chr_z';
   else 以概率 p 接受新解;
   if 满足终止条件 then
     根据当前温度 T 和 α，根据式(2)计算衰减温度 T'，until 衰减温度 T' 至 0;
```

}

5 用此聚类中心对 ECG 特征集作 K-means 聚类，得到最终聚类结果；

end

K-means 算法的初始聚类中心越精准，聚类的效果才越好。遗传模拟退火算法可以有效并且很精确地找到初始聚类中心。遗传模拟退火算法利用模拟退火算法的 Boltzmann 机制来减轻遗传算法的 ECG 染色体选择操作的压力，不仅有利于优良 ECG 染色体的保留，而且防止算法早熟造成子代 ECG 个体相似。其增强了算法全局收敛性的同时也使算法后期的爬山性能增强，加快了进化后期的收敛速度，提高了算法效率。遗传算法的群体操作思想，使算法在 ECG 解空间进行多处局部搜索，不仅加快了算法的搜索速度，还提高了模拟退火的局部收敛能力。随着遗传过程的执行，既达到冷却状态，最优解也将产生，这样就得到了准确的 ECG 初始聚类中心。基于均方差属性加权的遗传模拟退火算法的 K-means 聚类方法可以看做是遗传模拟退火算法在查找 ECG 种群内的最优个体，这个最优个体就是 K-means 聚类算法需要的初始聚类中心。以遗传算法控制寻找最优 ECG 染色体方向，可以加快搜索，以模拟退火解决局部收敛问题，可以提高搜索精度。

3　实验数据分析

为了研究 ECG 信号的自动识别，实现计算机自动诊断心电信号，提出了基于均方差属性加权的遗传模拟退火 K-means 改进聚类算法。本章进一步对本算法进行实验，并作了实验分析。在聚类算法设计过程中，需要采用权威性标准数据进行检测，以便对所设计的算法进行评估，优化设计算法。为此，在本文中采用麻省理工学院的 MIT-BIH 标准心电数据库数据，该库的数据共有 48 个病例，每例数据各长 30 min，总计有 11 613 个心搏，包含有两通道的正常心搏和各种稀有异常心搏的数据，分别由心电专家对每个心搏作了识别并加以注释。

采用 MIT-BIH 标准心电数据库，表 1 详细列出了实验中所使用的波形名称和波形总数。

表 1　实验数据中的心电波形类别组成

样本标号	原文件中所含的波形类别	原文件波形数	实验所用波形类别	所用波形数目
100_I	NORMAL，APC，PVC	2 273	NORMAL，PVC	2 240
102_I	NORMAL，PVC，PFUS，PACE	2 187	NORMAL，PVC，PFUS，PACE	2 187
104_I	NORMAL，PVC，FUSION，PACE，	2 229	NORMAL，PVC，FUSION，PACE	2 211
118_I	RBBB，APC，PVC	2 278	RBBB，PVC	2 182
119_I	NORMAL，PVC	1 987	NORMAL，PVC	1 987
121_I	NORMAL，APC，PVC	1 863	NORMAL，PVC	1 862

NORMAL、APC、PVC 等是 MIT-BIH 数据样本库中，注释文件的 atr 对应的标注表中关于 R 波的标注，是实验分析结果优劣的主要衡量标准[3,11]，如表 2 所示。

表 2　针对 R 波的注释信息对照信息

类别缩写	atr 文件中对应标号	类别英文名	中文名
NORMAL	1	normal beat	正常搏动
RBBB	3	right bundle branch block beat	右束支传导阻滞
PVC	5	premature ventricular contraction	室性早搏
FUSION	6	fusion of ventricular and normal beat	心室融合心跳
PACE	12	paced beat	起搏心跳

下面要提取 R 波，本文采用小波变换提取特征值，该方法一个很重要的特点就是可以使 ECG 减少工频干扰、肌肉收缩、病人移动以及呼吸带来的基线漂移等噪声对判断的影响[10]。

经过分析 ECG 功率谱密度的特点、小波变换的尺度和信号频率间的关系发现，QRS 波群的能量大多集中在 2^3 尺度上，在大于 2^4 尺度上则大大减小，而运动伪迹、基线漂移等能量大都集中在大于 2^5 的尺度上。因此本文在对心电信号处理时，取 $2^1 \sim 2^4$ 尺度上的小波变换结果来进行分析。

本文采用 MATLAB 进行编程实现 ECG 的小波变换，从图 1 可以看到该方法的处理结果。对于输入有基线漂移的 ECG，从 $2^1 \sim 2^4$ 尺度都消除了基线漂移的影响。其中在 2^3、2^4 尺度上 QRS 复波的特点显示得较为明显，具有最大的小波变换幅度，并且高频噪声、基线漂移、高尖 P 波、高尖 T 波的干扰得到了有效的抑制。所以本设计中采用 2^3、2^4 作为心电检测的特征尺度，在这两层的细节信号中检测对应 R 波的极值对，进一步进行实验。图 2 为经过小波变换后 R 波峰值检测效果，在 10 s 所获取的波形中，被检测人共检测出 14 个 R 波。如图 2 所示，经过小波变换后，检测 R 波峰值效果很好。提取出 R 波后按照上文所述，确定 QRS 波群得到 ECG 特征值进行聚类。

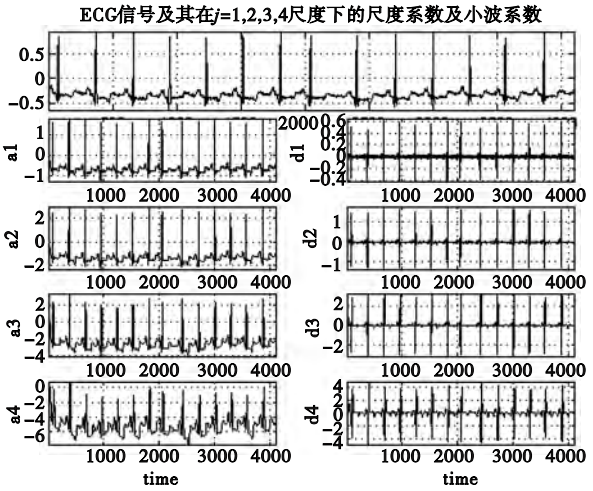

图1　ECG信号四级小波尺度系数及小波系数

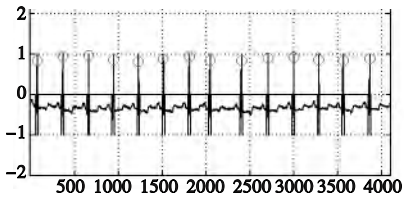

图2　ECG信号的R波峰值及QRS波波段

在这部分的实验中，本文就前面提到 K-means、基于模拟退火 K-means 和基于均方差属性加权的遗传模拟退火的改良 K-means 算法进行实验分析。

实验参数设置为种群规模 $M = 200$，最大进化代数 $G = 100$，交叉概率 $p_c = 0.84$，变异概率 $p_m = 0.03$，初始温度 $T0 = 1200$，冷却参数 $\alpha = 0.8$。以提取 ECG 信号的特征值作为测试数据，对 K-means、基于模拟退火 K-means 和基于均方差属性加权的遗传模拟退火的改良 K-means 算法的实验结果进行分析。

表 3、4 中记录了这三种聚类算法使用 ECG 信号的特征值进行聚类分析的结果。表 3 中记录了三种聚类算法的正确聚类波形数目，表 4 中记录了三种聚类算法的准确率。每种算法

执行 6 次，k 值根据表 1 选择。

表 3 三种聚类算法的正确聚类波形数目

样本标号	K-means 算法	模拟退火 K-means 聚类算法	改进的 K-means 聚类算法	样本标号	K-means 算法	模拟退火 K-means 聚类算法	改进的 K-means 聚类算法
100_I	2 060	2 071	2 117	118_I	1 894	1 953	1 987
102_I	2 029	2 034	2 050	119_I	1 649	1 660	1 714
104_I	1 914	1 937	1 960	121_I	1 705	1 722	1 763

表 4 三种聚类算法的准确率

样本标号	K-means 算法/%	模拟退火 K-means 聚类算法/%	改进的 K-means 聚类算法/%	样本标号	K-means 算法/%	模拟退火 K-means 聚类算法/%	改进的 K-means 聚类算法/%
100_I	91.964 2	92.455 3	94.508 9	118_I	86.801 1	89.505 0	91.063 2
102_I	92.775 4	93.004 1	93.735 7	119_I	82.989 4	83.543 0	86.260 7
104_I	86.567 1	87.607 4	88.647 7	121_I	91.567 2	92.481 2	94.683 1

从表 3、4 中可以看出，基于均方差属性加权的遗传模拟退火的 K-means 聚类算法的准确率不仅高于传统的 K-means 聚类算法，而且高于基于模拟退火 K-means 聚类算法。通过以上的实验数据可知，基于均方差属性加权的遗传模拟退火的 K-means 聚类算法的正确率比基于模拟退火 K-means 聚类算法高约 1.72%。由此可得出结论：基于均方差属性加权的遗传模拟退火的 K-means 聚类算法，具有较强的全局收敛能力，并且兼顾了局部收敛和全局收敛性能。

4 结束语

为了改进心电聚类算法，针对 K-means 聚类算法收敛时易陷入局部极值问题和对初始选值敏感的缺点，本文提出了一种基于均方差属性加权的遗传模拟退火的 K-means 聚类算法，引入特征权重以提高聚类结果的类内相似度，使用遗传方法在一个大的范围内搜索，并融合模拟退火算法局部搜索能力强的优点找到一个 K 值，充分发挥了遗传算法的快速全局搜索性能和模拟退火算法的局部搜索能力。能够寻找到最优聚类质心，使聚类的结果达到类内距最小、类间距最大，具有较高的效率和广泛的适用性。通过对 MIT-BIH 标准心电数据库的实验表明这种方法是比较有效的，聚类准确率相比传统 K-means 聚类算法和普通的模拟退火 K-means 聚类算法有明显提高。但是仍然存在一些问题有待改进：数据预处理过程中，由于心电波形 P 波和 T 波是通过阈值检测的。但是有时由于心律失常，QRS 波幅度、频率突然变得很小，P 波峰高度可能达不到阈值，导致漏检或误检，这对后面的聚类分析和分类分析也都会造成一定的影响。在今后的研究中，如何改进 P 波阈值和 R-R 间期的确定还有待研究，从根本上为波形的筛选提供准确性的保证。

参考文献：

[1] 王丽苹，董军. 心电图模式分类方法研究进展与分析 [J]. 中国生物医学工程学报，2010，29(6)：916-926.

[2] 葛丁飞，翁剑枫. 基于 2D-LDA 和高频心电信号的心肌梗死特征提取和分类 [J]. 航天医学与医学工程，2013，26(2)：125-130.

[3] WANG Wen-june，YE Yun-chi. Feature selection algorithm for ECG signals using range-overlaps method [J]. Expert Systems with Applications，2010，37(4)：2088-2096.

[4] 牟善玲，郑刚. 动态心电波形数据的聚类有效性评价方法研究 [J]. 计算机工程与应用，2011，47(32)：148-150.

[5] MARTIS R J，CHAKRABORTY C，RAY A J. A two-stage mechanism for registration and classification of ECG using Gaussian mixture model [J]. Pattern Recognition，2009，42(11)：2979-2988.

[6] 蒋德育，刘光远. 基于心电 P-QRS-T 波的特征提取及情感识别 [J]. 计算机工程与应用，2009，45(8)：213-215.

[7] 刘彤彤，戴敏，李忠义. 基于窗口斜率表示法的心电波形相似性分析 [J]. 计算机应用，2012，32(10)：2969-2972.

[8] 马永杰，云文霞. 遗传算法研究进展 [J]. 计算机应用研究，2012，29(4)：1201-1206.

[9] UBEYLI E D. ECG beats classification using multiclass support vector machines with error correcting output codes [J]. Digital Signal Processing，2006，17(3)：675-684.

[10] YU Sung-nien，CHOU Kuan-to. Selection of significant independent components for ECG beat classification [J]. Expert Systems with Applications，2009，36(2)：2088-2096.

[11] 练仕榴，郑刚，牟善玲. 用于心电波形分析的相似性度量策略 [J]. 计算机工程，2011，37(9)：263-265.

[12] EBRAHIMZADEH A A，KHAZAEE A. Detection of premature ventricular contractions using MLP neural networks：a comparative study [J]. Measurement，2010，43(1)：103-112.

[13] 张扬，杨松涛，张香芝. 一种模拟退火遗传算法的传感器网络数据融合技术研究 [J]. 计算机应用研究，2012，29(5)：1860-1862.

[14] KORUREK M，DOGAN B. ECG beat classification using particle swarm optimization and radial basis function neural network [J]. Expert Systems with Applications，2010，37(12)：7563-7569.

[15] WEI Jiang，KONG Song. Block based neural networks for personalized ECG signal classification [J]. IEEE Trans on Neural Networks，2007，18(6)：1750-1761.

[16] ZHAO Zhi-dong，YANG Lei，CHEN Dian-dian. Research of ECG identification based on FFT-matching pursuit algorithm [J]. 传感技术学报，2013，26(3)：307-314.

[17] 王庆荣，袁占亭，张秋余. 基于改进遗传—模拟退火算法的公交排班优化研究 [J]. 计算机应用研究，2012，29(7)：2461-2463.

[18] 景波，刘莹，黄兵. 基于遗传算法的 Job-Shop 调度问题研究 [J]. 计算机应用研究，2013，30(3)：688-691.

(上接第 3327 页)

[11] ZHANG Jian，YANG Yi-ming. Robustness of regularized linear classification methods in text categorization [C] //Proc of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York：ACM Press，2003：190-197.

[12] 陈玉芹. 多类别科技文献自动分类系统 [D]. 武汉：华中科技大学，2008.

[13] 樊兴华，孙茂松. 一种高性能的两类中文文本分类方法 [J]. 计算机学报，2006，29(1)：124-131.

[14] 贾宁. 使用概念基元特征进行自动文本分类 [J]. 计算机工程与应用，2007，43(1)：24-26.

[15] ZHENG Zhao-hui，WU Xiao-yun，SRIHARI R. Feature selection for text categorization on imbalanced data [J]. ACM SIGKDD Explorations Newsletter，2004，6(1)：80-89.

[16] GUPTA R，RATINOV L. Text categorization with knowledge transfer from heterogeneous data sources [C] //Proc of the 23rd AAAI Conference on Artificial Intelligence. [S. l.]：AAAI Press，2008：842-847.

[17] LEWIS D D. Reuters-21578 text categorization text collection [EB/OL]. [2013-08-22]. http://www.daviddlewis.com/resources/testcollections/reuters21578.

[18] 搜狗实验室—文本分类语料库 [EB/OL]. [2013-08-22]. http://www.sogou.com/labs/dl/t.html.

[19] 何琳，刘竟，侯汉清. 基于《中图法》的多层自动分类影响因素分析 [J]. 中国图书馆学报，2009，35(184)：49-55.