

心肌梗死诊断中的 SDEC 可分性准则研究

李小梅^a, 葛丁飞^b

(浙江科技学院 a. 理学院; b. 信息与电子工程学院, 杭州 310023)

摘要: 多数现存的可分性准则在高维情况下往往存在实现上的困难。针对心肌梗死心电信号(ECG)特征提取过程中的基于标准差和欧氏中心距(SDEC)的可分性准则进行研究,研究数据取自 PTB 诊断数据库,包括:正常状态病人、早期心肌梗死、急性期心肌梗死。实验结果表明,基于 SDEC 的可分性准则应用于心肌梗死心电信号特征提取和分类能有效地克服现存可分性准则实现上的困难,并与心肌梗死实际演变过程中的可分性相一致。

关键词: 可分性准则;心肌梗死;特征提取;标准差

中图分类号: R540.41; TN911.7

文献标识码: A

文章编号: 1671-8798(2006)04-0254-04

Study of Separability Criterion Based on SDEC in Myocardial Infraction Diagnosis

LI Xiao-mei^a, GE Ding-fei^b

(a. School of Science; b. School of Information and Electronic Engineering,
Zhejiang University of Science and Technology, Hangzhou 310023, China)

Abstract: Most of existing separability criteria is difficult to be implemented, especially in the hyperdimensional cases. In this investigation, the myocardial infraction features were extracted from electrocardiogram (ECG) for the classification, and the data was collected from PTB diagnostic database including normal condition, early infraction and acute infraction. The separability based on standard deviation and Euclidean center distance (SDEC) were analysed during the feature extraction. The experimental results show that the separability criterion based on the SDEC is able to overcome implementary difficulties effectively in myocardial infraction ECG feature extraction and classification, and consistent with the evolution process of myocardial infraction.

Key words: separability criterion; myocardial infraction; feature extraction; standard deviation

常用的可分性准则在特征提取的有效性等方面有着各自的优点,但都不能保证取得满意的结果^[1-4]。特别是在高维的情况下往往会遇到实现和

计算上的困难。为此,文献[5]提出了一种计算简单实现容易的基于标准差和欧氏中心距(SDEC)的可分性准则,并有效地应用于 MIT-BIH 数据库中心

收稿日期: 2006-10-15

基金项目: 浙江省教育厅科研计划项目(20050606)

作者简介: 李小梅(1956—),女,黑龙江宝清人,副教授,主要从事应用数学的教学与研究。

率失常的特征提取、特征有效性的评测和决策树的设计。同时该文还对常用的可分性准则的优缺点进行较为详细的论述。本文以心肌梗死 ECG 信号的特征提取和分类为例,对 SDEC 的有效性进行进一步的验证和研究。

从心电生理学角度看,12 个常规心电导联是心电向量空间分布的一种表现,不仅可以反映心脏疾病的不同特征,而且可以反映不同部位的心脏疾病^[6]。因此,12 导联 ECG 信号可以比较完整地反映心脏疾病的综合特征和全貌,并由此组成了超高维的时间序列。直接提取 ECG 信号的临床参数如 P 波和 T 波等方法虽然可以降低特征空间的维数,但对于包括心肌梗死在内的冠心病计算机智能诊断来说是不足够的,是难以取得令人满意的效果的^[7],所以,目前大多利用如 K-L 变换的方法来消除时间序列数据的冗余,再结合神经网络等方法提取特征和分类诊断^[8-10]。

国际上,对冠心病的研究主要集中在早期冠心病心肌缺血的检测上,以达到预防的目的^[8,9]。对心肌梗死的研究则主要集中在心肌梗死发生部位的诊断上,而对早期心肌梗死的检测相对较少。心肌梗死演变过程依次可分为几个阶段:正常→很早期(数分钟或数小时)→急性期等。随着病情的加剧,ECG 异变程度逐渐加剧^[6]。相对于正常的 ECG 而言,很早期心肌梗死 ECG 相似性应该高于急性期的相似性。一个好的可分性准则应该反映出这一病理过程,并与之保持一致。文中的数据取自 PTB 诊断数据库,包括心肌梗死演变过程中不同阶段的 12 导 ECG 信号:正常状态病人(NORM)、很早期心肌梗死(EARI)和急性期心肌梗死(ACTI)。急性期心肌梗死一旦发生,对心肌造成损伤几乎是不可逆转的。在心肌梗死发生后的很早期,如果心肌能得到再灌注治疗,濒临死亡的心肌可能得以存活。于是,能否及时诊断出此类病人是预后改善的积极治疗措施成败的关键,所以本文同时对心肌梗死很早期的检测分类进行探讨研究。

1 方 法

1.1 预处理

在 PTB 诊断数据库中,数据的采样频率是 1 000 Hz。本文中选取了如表 1 所示的数据进行测试和分类,并经过两位心血冠疾病主任医师的独立确认。用 0.92 s 的固定数据窗对信号进行分割,即 ECG 峰值 R 以前取 0.31 s,ECG 峰值 R 以后取 0.61 s。

为了满足诊断的需要,ECG 信号的带宽应保持在 1~100 Hz 之间^[2,11]。为此,采用一个 60 Hz 陷滤波器、一个 0.4 Hz 陷滤波器和一个截止频率为 100 Hz 的低通滤波器先后对 ECG 信号进行滤波,以分别消除 60 Hz 电源噪音、基线漂移和 100 Hz 以上的高频噪音。为了降低运算量和特征空间的维数,在特征提取以前先对信号进行再采样,使其采样频率下降为 125 Hz。

表 1 取自 PTB 诊断数据库中的数据 个

项目	NORM	EARI	ACTI
样本总数	1 259	2 004	748
训练样本数	500	500	300
测试样本数	759	1 504	448

1.2 基于 SDEC 的可分性评价准则

为了叙述的方便,重新引入葛丁飞等^[5]一文中基于欧氏中心距和标准差可分性评价准则 J ,其计算公式如下:

$$J = \frac{\|D\|}{3\left(\frac{1}{d}\sum_{i=1}^d\sigma_{1ii} + \frac{1}{d}\sum_{i=1}^d\sigma_{2ii}\right)} \quad (1)$$

式(1)中, σ_{1ii} 、 σ_{2ii} , $i=1,2,3\cdots d$,分别是特征向量各分量的标准差。其中,分子为两类之间的欧氏中心距,其值越大,越有利于分类;分母包含了概率分布的信息,其值越小,样本的收敛性越好。

1.3 特征提取和分类

12 导联 ECG 信号经预处理后按如图 1 所示的顺序进行重新排列,即 I、II、III、avr、avl、avf、v1、v2、v3、v4、v5、v6。由此构成一个 1 380 维的特征矢量。利用 K-L 变换对 1 380 维的特征空间进行降维以消除冗余的数据。为了使 K-L 变换以后的特征矢量尽可能地保留原有的信息,用两种方法选取 K-L 变换的特征向量:

1)选前 k 个最大特征值的特征向量,使得样本在前 k 个轴上的能量占整个能量的 99% 以上(EP 准则);

2)基于欧氏中心距和标准差的不可分性评价准则 J ,选前 k 个最大特征值的特征向量,使得 $J \geq 1$ 或 J 值没有明显增加为止。

为了使样本接近球形分布以利于后续的特征提取,对上述 K-L 变换以后的特征向量 x 进行白化处理,得到相应的白化特征向量 y 。白化算法为:计算样本协方差,并分别计算其特征向量和特征值 p_k 和 λ_k ($k=1,2,3\cdots m$), p_k 为 $m \times 1$ 的列向量。然后对样

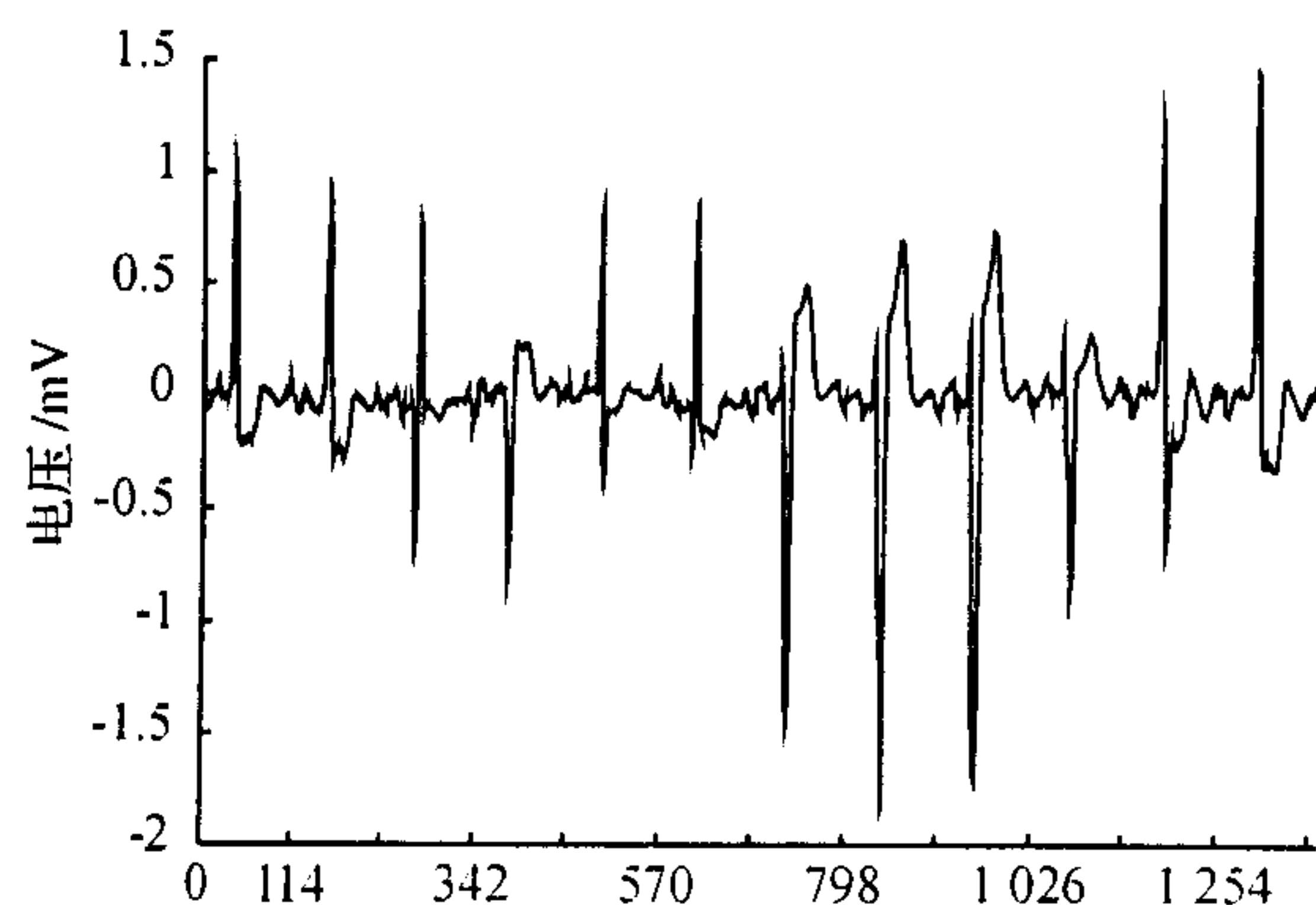


图 1 由 12 导联 ECG 重新排列成的序列

本做如下的变换^[4]:

$$y = P^T x \quad (2)$$

式(2)中, $P = (p_1 / \sqrt{\lambda_1}, p_2 / \sqrt{\lambda_2}, \dots, p_m / \sqrt{\lambda_m})$ 为一 $m \times m$ 的矩阵, y_m 为 $m \times 1$ 的列向量。

最后进行按欧氏距离的特征提取,其算法为:

1) 计算白化后样本的类内散度矩阵 S_w 和类间散度矩阵 S_b ;

2) 计算 $S_w^{-1} S_b$ 的特征向量;

3) 将白化后样本投影到 $S_w^{-1} S_b$ 的特征向量上得到 $C-1$ 维样本(本文 $C=3$)。

采用线性分类器对样本进行分类。采用 SDEC 可分性评价准则 J 构造如图 2 所示的二叉树进行分类,其间将类间 J 值小的分为一组。

2 结 果

为了选择有效的 K-L 变换坐标向量,预选 k 个

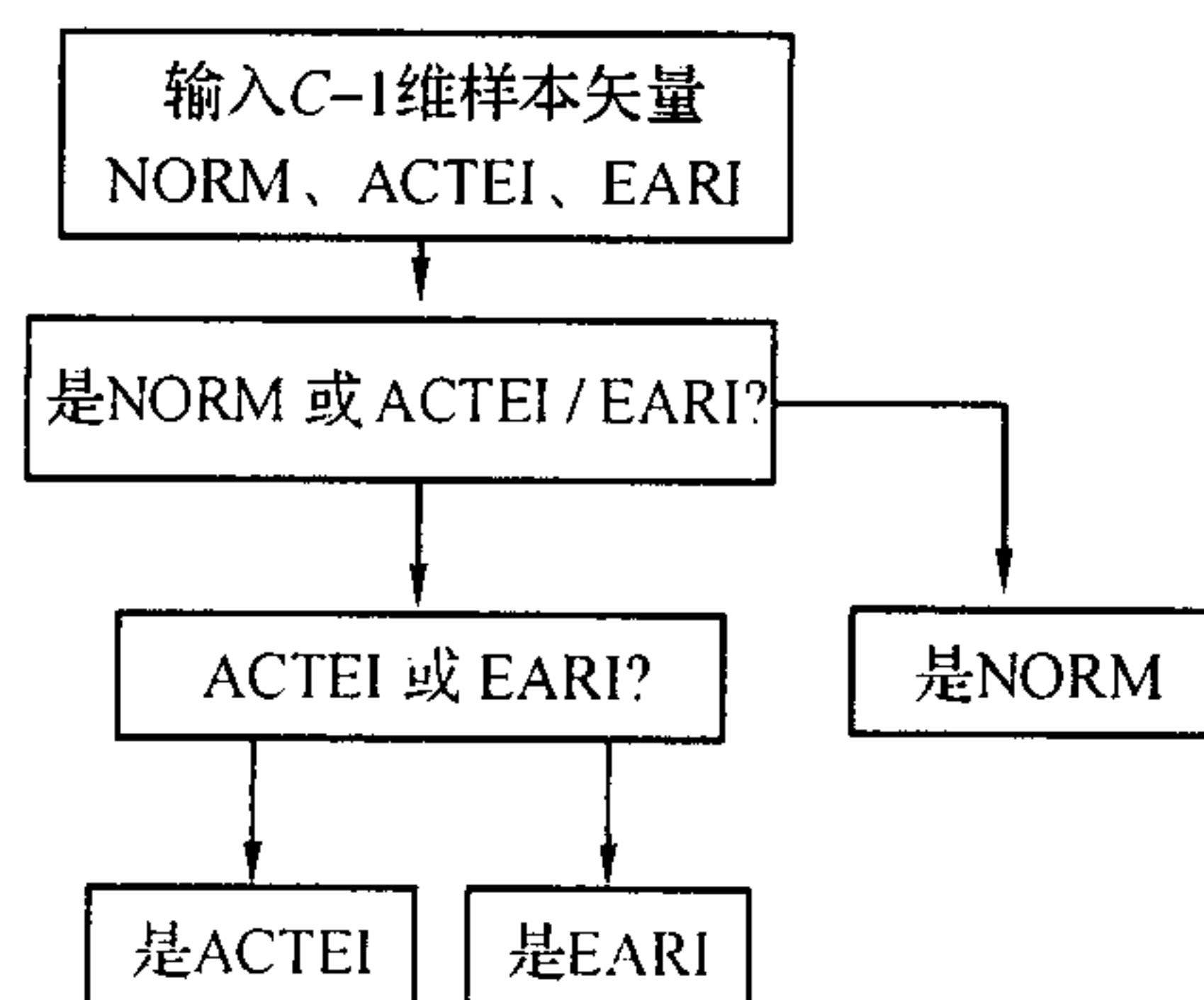


图 2 用于分类的二叉树

最大特征值和相应的特征向量,并且将样本投影到该特征向量上,然后计算出上述 EP 值和各类间基于 SDEC 的可分性准则 J ,其结果如表 2 所示。结果表明,为了同时使各类间的 $J \geq 1$ 和 $EP \geq 0.99$, k 的取值最少应为 70。所以, K-L 变换后样本的维数可由 1 380 降到了 70。70 维样本经白化后各类样本间的 J 值如表 3 所示。为了分析可分性准则 J 的有效性,在表 3 中还给出了 1 280 维原始样本和基于欧氏距离特征提取法^[1]得到的 2 维样本的 J 值。

由表 3 可见, ACTEI/EARI 间的 J 值相对较小, NORM 与 EARI/ACTEI 间的 J 值相对较大,因此,在二叉树构建中,将类 EARI/ACTEI 分为一组,而将 NORM 单独列为另一组, NORM 首先被分离开来。利用线性分类器的平均分类结果如表 4 所示。

表 2 不同 k 值的实验结果

k	10	20	30	40	50	70	80	90
EP%	0.819 0	0.918 4	0.959 1	0.976 8	0.990 3	0.999 7	0.999 7	0.999 9
J	NORM / ACTEI	0.574 1	0.928 5	1.074 6	1.236 6	1.434 7	1.719 2	1.824 5
	NORM/EARI	0.565 8	0.695 1	1.178 6	1.093 2	1.063 8	1.281 2	1.357 6
	ACTEI/EARI	0.203 8	0.420 1	0.535 5	0.665 9	0.851 6	1.039 4	1.243 2

表 3 特征提取过程中类间 J 值

	NORM/ EARI	NORM/ ACTEI	EARI/ ACTEI
1 280 维原始样本	5.269 6	6.751 2	3.860 6
K-L 变换后 70 维样本	1.281 2	1.719 2	1.039 4
白化后 70 维样本	1.233 4	1.716 1	1.132 5
2 维样本	0.637 8	0.639 2	0.367 7

表 4 分类结果

类别	NORM	ACTEI	EARI
$k=70$ 时的分类精度	99.84	97.34	91.80
$k=50$ 时的分类精度	95.70	91.70	81.15

3 讨 论

本文采用 K-L 变换以消除 12 导联间的数据冗余, K-L 变换坐标矢量的选择直接关系到保留原有

信息的程度。现有的研究表明,有关的多种准则都不能始终保证取得满意的结果。为此,本文采用了两种方法进行 K-L 坐标系的选择。由表 1 可见,在 k 值达到 50 时,EP 值为 0.99,也就是说按 EP 准则 k 取 50 就够了,但是表 4 的分类结果表明 EARI 的分类精度并不理想,这一结果与直接 ECG 参数检测法相类似^[7]: IEARI 为 78.3%; ACTEI 为 90.6%。在 $k=50$ 时,ACTEI/EARI 间的 J 值是 0.851 6,而当 k 增加到 70 时,ACTEI 和 EARI 间的 J 值增加到 1.039 4,此时各类间的 J 值都满足 $J \geq 1$ 的要求。由表 4 可见, k 增加到 70 时分类效果取得了明显的改进。其原因是 EP 能量准则是从保留总体样本信息观点出发的,并且不包含类间分类信息,而基于 SDEC 的可分性准则 J 却包含了类间的分类信息。利用 EP 能量准则选择 K-L 变换坐标轴和神经网络的分类方法得到的结果是^[11]: NORM 为 99.4%,ACTEI+IEARI 为 88.6%。增加 k 值至 120 重复以上的实验:在 $k > 70$ 时,分类结果的改善不再明显;在 $k \geq 100$ 时,分类结果几乎不再提高。

从心肌梗死的演变过程和电生理过程看,EARI/ACTEI 间的 ECG 信号相似性是最高的(J 值最小),并且 NORM/EARI 间的相似性应该大于(J 值小于)NPRM/ACTEI 间的相似性。表 3 的实验结果表明,特征提取过程中得到的类间 J 值始终与这一实际演变过程相一致。笔者用 Bhattacharyya 距离和散度的可分性准则重新计算上述特征提取过程中的可分性^[1],在对 1 380 维原始样本进行计算时遇到了近奇异矩阵求逆的困难。此外,基于 Bhattacharyya 距离和散度的可分性也不能始终完全与实际演变过程相一致,换句话说不能真实地反映其间的可分性。

正如许多现有的可分性准则有一定的局限性一样,基于 SDEC 的可分性评价准则 J 适用于单峰分布的数据。其最大的优点是计算简单而有效,并能真实地反映类间的可分性。就心电信号而言,基于 SDEC 的可分性准则 J 能与实际的电生理过程保持一致的特性。

4 结 语

基于 SDEC 的可分性准则 J 应用于心肌梗死特征提取和决策树设计可取得良好的结果, J 值能真实地反映类间的可分性。

参考文献:

- [1] 边肇祺,张学工. 模式识别[M]. 北京:清华大学出版社,2002:176-196.
- [2] 葛丁飞,夏顺仁. AR 模型在远程心电诊断中的应用[J]. 中国生物医学工程学报,2004,23(3):222-229.
- [3] OLMOS S, MILLAN M, GARCIA J. ECG data compression with the Karhunen-Loeve transformation [J]. Computer in Cardiology, 1996, 5: 253-256.
- [4] FUKUNAGA K. Introduction to statistical pattern recognition [M]. London: Academic Press Limited, 1990.
- [5] 葛丁飞,李时辉,瞿晓. 一种适于心电数据的可分性评价准则研究[J]. 浙江科技学院学报,2005,17(1):9-12.
- [6] 戚仁铎. 诊断学[M]. 3 版. 北京:人民卫生出版社, 1988.
- [7] BAXT F, SHOFER F, SITES J H. A neural computational aid to the diagnosis of acute myocardial infarction[J]. Ann Emerg Med, 2002 (34): 366-373.
- [8] TELEMACHOS S, KONSTANTINOS D, NICOS M. ECG analysis using nonlinear PCA neural networks for ischemia detection[J]. IEEE Transactions on Signal Processing, 1998, 46 (11): 3058-3067.
- [9] PAPADIMITRIOU S, MAVROUDI S, VLADUTU L. Ischemia Detection with a self-organizing map supplemented by supervised learning[J]. IEEE Transaction on Neural Networks, 2001, 12 (3): 503-515.
- [10] OHLSSON M, OHLIN H, WALLERSTEDT S, et al. Usefulness of serial electrocardiograms for diagnosis of acute myocardial infarction[J]. Am J Cardiol, 2001 (88): 478-481.
- [11] TOMPKINS W. Biomedical Digital Signal Processing [M]. Englewood Cliffs, New Jersey: Prentice Hall, 1993: 246-253.