

## 基于自适应学习的心律失常心拍分类方法

王凯<sup>1</sup>, 杨枢<sup>1,2</sup>

1. 蚌埠医学院卫生管理系, 安徽 蚌埠 233030; 2. 合肥工业大学信息与计算机学院, 安徽 合肥 233009

**【摘要】**心律失常是因心脏疾病引起的心电活动中的异常症状,早期心室收缩(PVC)是由异位心跳引起的常见心律失常形式。通过心电图(ECG)信号检测PVC对于预测可能的心力衰竭具有重要意义。本文提出一种面向PVC心拍分类的心电信号分类算法,重点研究基于自适应学习的PVC异常心拍分类特征提取模型,通过计算心拍关联后验概率,结合领域专家标注信息训练分类器,提高整体分类效果。实验采用MIT-BIH心律失常数据库的ECG数据,研究结果表明所提方法针对非线性流形结构数据,能够有效提升小样本心拍自适应分类器的准确性。

**【关键词】**心电图;自适应分类器;特征提取;分类

**【中图分类号】**R318;TP391

**【文献标志码】**A

**【文章编号】**1005-202X(2019)01-0092-05

## Adaptive learning-based method for classification of arrhythmic heartbeats

WANG Kai<sup>1</sup>, YANG Shu<sup>1,2</sup>

1. Department of Health Management, Bengbu Medical College, Bengbu 233030, China; 2. School of Information and Computer, Hefei University of Technology, Hefei 233009, China

**Abstract:** Arrhythmia is a common electrocardiogram (ECG) abnormality in heart diseases, among which premature ventricular contraction (PVC) is a widespread arrhythmia caused by ectopic heartbeat. The detection of PVC by ECG signals is significant for predicting possible heart failure. Herein an ECG signal classification algorithm for PVC heartbeat classification is proposed. The feature extraction model of PVC abnormal heartbeat classification based on adaptive learning is mainly studied. The posterior probability of heartbeat correlation is calculated and then combined with information classifier to improve the classification performance. The ECG data adopted in this study are from MIT-BIH arrhythmia database. The research results show that the proposed algorithm can effectively improve the accuracy of adaptive classifier for small-sample heart beats of nonlinear manifold data.

**Keywords:** electrocardiogram; adaptive classifier; feature extraction; classification

### 前言

为降低心血管疾病引起的死亡率,心脏相关疾病的早期并发症检测与识别越来越受到医学领域专家的重视。心律失常是由心脏电脉冲引起的传导缓慢、阻滞或经异常通道传导的规律性心率异常<sup>[1]</sup>。在各种异常中,早期心室收缩(Premature Ventricular Contraction, PVC)是一种危害性较高的心律失常,通

常是由心室的异位节律点发出激动波使心室提早收缩,即QRS综合波在P波前出现,是成人中广泛存在的心律失常形式<sup>[2]</sup>。临床上对PVC的评估和治疗相对较复杂,针对心血管系统疾病的自动识别,领域专家进行了大量研究。PVC在心电图(Electrocardiogram, ECG)上的生物特征主要表现为具有异常时序的心跳序列,通过使用已建立的疾病类别关联关系,能够在ECG记录中标记心拍序列,从所得序列中确定ECG信号的节拍属性,达到相关疾病的诊断与识别目的。

基于ECG的PVC心律失常分类方法,研究多采用时频分析技术、统计测量等混合方法。文献[3]应用动态贝叶斯网络进行PVC分类,分类过程中使用简单的决策规则,适用于嵌入式分类监测。文献[4]使用基于神经网络的分类方案检测PVC,比较了

**【收稿日期】**2018-09-15

**【基金项目】**安徽省高校人文社会科学重点研究基金(SK2018A1064, SK2018A1072);安徽省高校自然科学基金重点研究基金(KJ2018A1007);蚌埠医学院科技发展基金(BYKF1717)

**【作者简介】**王凯,硕士,讲师,主要研究方向:医学信息决策、数据挖掘, E-mail: wangkai0552@126.com;杨枢,博士,教授,主要研究方向:医学信号分析, E-mail: 475870652@qq.com

PVC 聚类学习和分类技能,提取了 10 个 ECG 结构特征,设计了一种用于 PVC 识别的低复杂度数据自适应方法。文献[5]着重于 PVC 检测的多层次学习模型,提出了一种使用多支路学习和支持向量机(Support Vector Machine, SVM)的 PVC 识别方法。文献[6]提出了基于神经网络的 ECG 模式识别方法,通过 Nearest Neighbo(NN)分类器对 PVC 进行分类,并使用小波变换来提取心电图数据的形态学特征。文献[7]使用独立分量分析(Independent Component Analysis, ICA)进行特征提取,采用 k-means 和 Fuzzy C-Means(FCM)分类器来识别 PVC 节拍。文献[8]采用贝叶斯正则化反向传播训练模型,通过优化多特征最优解模型,提高临床 ECG 信号标注的准确性。ECG 心拍分类算法能够有效降低 PVC 心律失常的临床检测难度,简化记录和分析的工作量,逐渐成为领域专家研究的热点方向。目前基于 ECG 的心拍自动分类算法主要分为 3 大类:基于 ECG 信号特征识别的方法、分类器模型方法以及自适应模型分类过滤算法。

然而,上述方法在提取 ECG 信号特征时将信号波形的线性判别与聚类模型相结合,虽然在分类效果上有所提高,却降低了分类器的样本外点处理能力,无法有效处理变化的样本数据特征识别。本文提出一种面向 PVC 心拍分类的心电信号分类算法,重点研究基于自适应学习的 PVC 异常心拍分类特征提取模型,通过计算心拍关联后验概率,按估计置信度排序,获取极大似然概率估计。自适应分类器使用领域专家的标注信息,提高心跳标签系统的分类性能。根据 AAMI 指南<sup>[8]</sup>处理心拍分类,使用 MIT-BIH 心律失常数据库提供的心拍标签。训练分类器,动态调整分类器参数,提高整体分类效果。

## 1 方法

使用 MIT-BIH 心律失常数据库的非起搏器记录数据,每个记录包含两组 ECG 导联信号。采用 0.1~100.0 Hz 进行带通滤波,并以 360 Hz 进行采样。通过信号预处理,心拍分割获取规范化的心拍采样片段,通过特征提取以及概率权重排序,结合专家注释标注信息,构建自适应分类器,输出相关类的分类结果。

### 1.1 数据预处理

ECG 信号中的噪声主要产生于以下几个方面:EMG 噪声、电力线噪声、基线漂移和复合噪声<sup>[9]</sup>。不同患者的 ECG 信号显示出显著的差异,对计算出的信号振幅波动特征影响较大。执行归一化的预处理操作能够显

著降低 ECG 信号的差异。本文将信号的均值设置为零,零均值信号  $\{y(t) | 1 \leq t \leq L\}$  计算如式(1)所示:

$$y(t) = |x(t) - x| \quad (1)$$

其中,  $y(t)$  是待计算的信号,  $\{x(t) | 1 \leq t \leq L\}$  是原始 ECG,  $x$  是  $x(t)$  的算术平均值,  $L$  是信号长度。使用中值滤波器降低噪声。中值滤波器是一种简单的非线性平滑器,可以保持信号边缘光滑的同时抑制噪声。滤波信号  $\{y(t) | 1 \leq t \leq L\}$  计算如式(2)所示:

$$Y(t) = \text{Med}\{y(t-1), y(t), y(t+1)\} \quad (2)$$

其中,  $Y(t)$  是滤波信号,  $y(t)$  是输入信号。应用级联低通滤波器从信号中去除频率分量低于 0.5 Hz 的信号数据,消除基线漂移和电力线噪声。基线漂移的频率分量通常低于 0.5 Hz,将频率限制调整为 2 Hz<sup>[10]</sup>。通过从适当延迟的输入信号中减去低通滤波器的输出,实现信号从低通滤波器到高通滤波器的平稳过渡。根据 QRS 复合波中 R 点的位置(根据滤波后的 ECG 信号建立每个心拍的 R 点,从 R 点右侧的 100 点开始分割)。R 点的相关位置组成了 MIT-BIH 数据库的注释文件,所选择的心拍构成了一个 7 000×200 矩阵。

### 1.2 选取信号特征集

使用独立记录数据对系统进行测试,通过多个记录中的心拍子类估计其分类性能<sup>[11]</sup>。分类系统意图如图 1 所示。系统选取的信号特征集主要包括表 1 的信号数据<sup>[12]</sup>。首先使用初始分类器处理输入 ECG 记录,产生初始的心拍子标签集;然后将标记的节拍提交给领域专家,根据分类需要更正标签信息,使用校正的标签计算分类器参数集合;最后组合新分类器的参数集合,生成自适应的分类器参数集合,系统使用已修改的参数处理剩余未注释的心拍数据。

### 1.3 心拍分类

采用基于线性判别的分类系统,输出后验条件概率,在单次迭代中训练分类器。基于自适应学习的心拍分类算法(Adaptive Learning of the heart Beat Classification, ALHBC)通过后验概率分析与计算、构造特征矩阵,以分类器参数的加权平均值为基础,动态调整最佳分类器参数阈值,输出具有异常时序的心跳分类集合。ALHBC 算法的核心步骤描述如下。

**1.3.1 计算后验概率** 线性判别分类器参数(类均值  $\mu_k$  和协方差  $\Sigma$ )计算方法如公式(3)、(4)所示。式中  $c$  表示分类总数,  $N_k$  代表第  $k$  个类的训练数,  $\mathbf{x}_{kn}$  表示训练中属于第  $k$  个类的  $n$  个特征向量,  $\sum_k$  条件协方差矩阵,用于训练的节拍总数为  $N = \sum_{n=1}^c N_k$ ;  $\mu_k$ 's 和  $\Sigma$  由

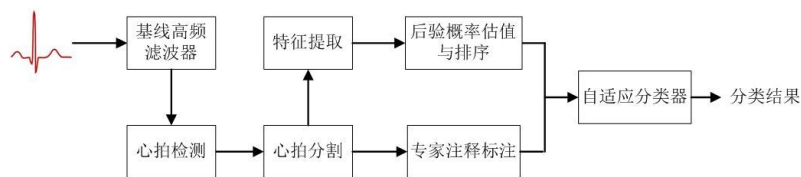


图1 基于自适应学习的心拍分类系统原理图

Fig.1 Schematics of heartbeat classification system based on adaptive learning

表1 自适应分类器特征集

Tab.1 Adaptive classifier feature set

ECG 特征集
RR 前后间隔、平均 RR 间隔
导联 A 和 B 的 QRS 持续时间 (QRS 偏移-QRS 起始)
导联 A 和 B 的 T 波持续时间 (T 波偏移-QRS 偏移)
导联 A 和 B 的 P 波标志
导联 A 和 B 之间 QRS 起始与偏移之间的 ECG 形状
导联 A 和 B 的 QRS 偏移和 T 波偏移之间的 ECG 形状

训练集的分类数决定,特征向量  $\mathbf{x}$  依据计算估计的后验概率进行特征分类,  $P(k|\mathbf{x})$  使用公式(5)计算第  $k$  个分类集,公式中  $y_k = -\frac{1}{2}\mu_k^T \Sigma^{-1} \mu_k + \mu_k^T \Sigma^{-1} \mathbf{x}$ 。系统的最终分类由计算出的后验概率估计中数值最大的类决定。在处理输入记录参数后,将选择的心拍提交领域专家进行标签注释。

$$\mu_k = \sum_{n=1}^{N_k} \mathbf{x}_{kn} / N_k \quad (3)$$

$$\sum_k = \sum_{n=1}^{N_k} (\mathbf{x}_{kn} - \mu_k)(\mathbf{x}_{kn} - \mu_k) \quad (4)$$

$$P(k|\mathbf{x}) = \frac{e^{y_k}}{\sum_{i=1}^c e^{y_i}} \quad (5)$$

**1.3.2 计算组合后验概率** 为减少提交给领域专家的心拍数量,选择合适的自适应分类器心拍。使用公式(5)计算每个类中心拍的后验概率,并按降序排列,获取置信度最高的后验值  $R_i$ ,计算公式见公式(6)。高置信度下的非自适应分类器产生的心拍,具有相对较高的  $R_i$ 。为获取最佳自适应训练心拍,将  $R_i$  按升序排列,将数值较低的心拍交专家重新评估,最终输出组合后验概率最高的类。

$$R_i = \max_{k=L_c} (P(k|\mathbf{x}_i)) \quad (6)$$

设  $X$  为随机变量,ECG 信号波形分布服从概率密度函数  $f(x, \theta)$ ,利用公式(6)计算的最高后验概率,构造关于样本观测值的似然函数,极大似然概率估计表示如公式(7)所示:

$$L(\theta) = \prod_{i=1}^n f(x_i, \theta) = f(x_1, \theta) f(x_2, \theta), \dots, f(x_n, \theta) \quad (7)$$

**1.3.3 计算特征矩阵** ECG 特征提取与分类是一种涉及基线降噪、心拍分割及特征抽取等的复杂相关性过程。通过构建核心矩阵  $\mathbf{Z}$ ,定义并存储表1中的特征集合,该矩阵会随着输入向量的增加而进行动态更新。若  $X_1, X_2, \dots, X_N \in R_N$  与  $Y_1, Y_2, \dots, Y_M \in R_N$  分别来自两个不同的训练组,核心矩阵  $\mathbf{Z} = \mathbf{X} \cup \mathbf{Y}$  定义如公式(8)所示,对应的协变矩阵  $\mathbf{CM}_z$  定义如公式(9)所示:

$$\mathbf{Z} = \{X_1, X_2, \dots, X_N, Y_1, Y_2, \dots, Y_M\} \in R^N \quad (8)$$

$$\mathbf{CM}_z = \frac{(\sum_{i=1}^N X_i X_i^T + \sum_{j=1}^M Y_j Y_j^T)}{N + M} \quad (9)$$

针对核心矩阵,可进一步求解其特征值  $\lambda$  和特征向量  $\mathbf{y}$ ,分别定义如公式(10)、(11)所示,上式中  $\alpha$ 、 $\beta$  分别表示  $X_i$ 、 $Y_i$  的特征向量。若  $\mathbf{Z}$  的元素总数为  $k$ ,为便于将数据映射到约简的特征向量空间,分别将上式改写成公式(12)、(13)所示:

$$\lambda \mathbf{y} = \mathbf{CM}_z \mathbf{y} \quad (10)$$

$$\mathbf{y} = \sum_{i=1}^{N+M} \mathbf{Z}_i = \sum_{i=1}^N \alpha_i X_i + \sum_{i=1}^M \beta_i Y_i \quad (11)$$

$$\lambda \mathbf{Z}_k \mathbf{y} = \mathbf{Z}_k \mathbf{CM}_z \mathbf{y} \quad (12)$$

$$\lambda \begin{bmatrix} \mathbf{X} \mathbf{X}^T & \mathbf{X} \mathbf{Y}^T \\ \mathbf{Y} \mathbf{X}^T & \mathbf{Y} \mathbf{Y}^T \end{bmatrix} \xi = \frac{1}{N+M} \begin{bmatrix} \mathbf{X} \mathbf{X}^T & \mathbf{X} \mathbf{Y}^T \\ \mathbf{Y} \mathbf{X}^T & \mathbf{Y} \mathbf{Y}^T \end{bmatrix}^2 \xi \quad (13)$$

**1.3.4 计算分类器参数加权平均值** 获取自适应分类器的最佳参数  $\mu_k$  和协方差  $\Sigma$ ,选择合适的心拍,并将其提交给领域专家进行标签注释。在训练初始数据参数记录后,计算分类器参数加权平均值,使用公式(14)、(15)确定  $\mu$  和  $\Sigma_k$  的加权平均值。

$$\mu_k = \alpha \mu_k + (1 - \alpha) \mu_k^T \quad (14)$$

$$\sum_k = \beta \Sigma_k + (1 - \beta) \Sigma_k^{-1} \quad (15)$$

## 2 实验与分析

在获取后验概率估值后,使用交叉验证过程估计分类器的综合性能。为了获得性能测量的无偏估计,测试数据中包括不属于测试心拍分类信号的数据。将所有先验概率设为 1/11,分类条件加权值设置为 0.5,由领域专家标注的最佳分类自适应参数在 1 到 500 次心拍之间随机变动,以适应不同的分类信号采样。使用



MIT-BIH 心律失常数据库验证 PVC 心跳分类,不平衡比设置为 1:100,其中选取 200 次正常心跳(NOR),200 次右束支传导阻滞(Right Bundle Branch Block, RBBB),200 次左束支传导阻滞(Left Bundle Branch Block, LBBB),200 次心房早搏(Atrial Premature, AP)以及 200 次 PVC 作为信号的原始特征数据。

本实验包含两个子实验,实验一将所有心律失常心拍无区别地输入未知类,重点测试 ALHBC 算法对异常心拍的分类能力;实验二重点关注 PVC 的分类效果,在测试阶段添加未知心跳类型,检测算法对实验一中 PVC 类数据的分类能力。

实验一对样本中的原始数据进行特征分解,将特征集分别输入 SVM<sup>[13]</sup>、Back Propagation Neural Network(BPNN)<sup>[14]</sup>以及 Learning Vector Quantization(LVQ)<sup>[15]</sup>分类模型,并将分类结果与 ALHBC 算法结果进行横向比较,实验结果如图 2 所示。在本实验中,特征提取阶段输出 60 个特征集合,其中 LVQ 的准确率为 85.31%,SVM 准确率为 87.44%,BPNN 准确率为 91.72%,ALHBC 算法准确率为 92.18%。结果表明通过集成特征提取阶段的 ALHBC 算法能够解决特征提取与分类数据的不一致问题。在非平衡数据条件下,对非线性流形结构数据具有相对较高的分类准确性。

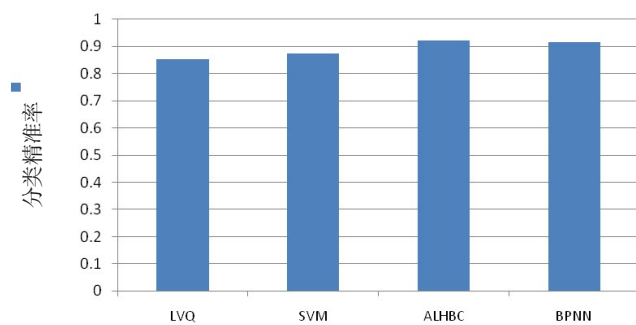


图2 心律失常心拍分类结果

Fig.2 Arrhythmic heartbeat classification results

LVQ: Learning vector quantization; SVM: Support vector machine; ALHBC: Adaptive learning of the heart beat classification; BPNN: Back propagation neural network

实验二重点分析 PVC 的分类效果,在测试阶段添加未知心跳类型,选取实验一中分类效果较好的两种分类模型(ALHBC 和 BPNN),构建如表 2 所示的混淆矩阵,其中列数据表示实验预测类,每列总数为预测该类数据总数;每行表示数据的真实归类,每行总数为该类别数据实例总数。结果表明 ALHBC 算法针对的 PVC 异常心拍分类特征提取较其他分类更加接近真实数据分类。

表2 特征分类混淆矩阵

Tab.2 Confusion matrix of feature classification

		PVC	NOR	RBBB	LBBB	AP
PVC	ALHBC	98	0	0	0	1
	BPNN	75	0	0	0	0
NOR	ALHBC	0	100	0	0	0
	BPNN	0	100	0	0	0
RBBB	ALHBC	0	1	36	1	5
	BPNN	1	10	40	6	12
LBBB	ALHBC	0	0	0	21	0
	BPNN	2	0	0	13	0
AP	ALHBC	0	0	0	0	56
	BPNN	11	5	3	6	37

PVC: Premature ventricular contraction; NOR: Normal heartbeat; RBBB: Right bundle branch block; LBBB: Left bundle branch block; AP: Atrial premature

### 3 总结与讨论

本文提出一种自适应心拍分类算法,研究解决 ECG 形态变异性的分类指证,构建基于自适应学习的 PVC 异常心拍分类特征提取模型,实现 ECG 的自动分类问题。通过计算心拍关联后验概率,结合领域专家标注信息训练分类器,提高整体分类效果。研究表明,所提方法能够处理非平衡类数据特征提取的分类问题,特别针对非线性流形结构数据,能够有效提升小样本心拍的自适应分类器的准确性。

结合临床环境中的实际应用,本文的后续研究将围绕分类器参数的动态自适应调整问题,改进输入参数的初始簇最优解模型,减少所需领域专家的标签总数,提升模型用于复杂条件下的自主学习能力。

### 【参考文献】

- [1] DE LANNON G, FRANCOIS D, DELBEKE J, et al. Weighted conditional random fields for supervise dinterpatient heartbeat classification[J]. IEEE Trans Biomed Eng, 2012, 59(1): 241-247.
- [2] SUFI F, KHALIL I. Diagnosis of cardiovascular abnormalities from compressed ECG: a data mining-based approach[J]. IEEE Trans Inf Technol Biomed, 2011, 15(1): 33-39.
- [3] GANDOMI A H, ALAVI A H. A new bio-inspired optimization algorithm[J]. Commun Nonlinear Sci Numer Simul, 2012, 17(12): 4831-4845.
- [4] JATMIKO W, NULAD W P, MATUL I, et al. Heart beat classification using wavelet feature based on neural network[J]. WSEAS Trans Syst, 2015, 10(1): 17-26.

- [5] GHORBANIAN P, GHAFARI A, JALALI A, et al. Heart arrhythmia detection using continuous wavelet transform and principal component analysis with neural network classifier[J]. IEEE Comput Cardiol, 2014, 12(4): 669-672.
- [6] FAEZIPOUR M, SAEED A, BULUSU S C, et al. A patient-adaptive profiling scheme for ECG beat classification[J]. IEEE Trans Inf Technol Biomed, 2010, 14(5): 1153-1165.
- [7] CAI X, FAN S, TAN Y. Light responsive curve selection for photosynthesis operator of APOA[J]. Int J Bio-inspir Com, 2016, 4(6): 373-379.
- [8] RAD A B, EFTESTOL T, ENGAN K, et al. ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis[J]. IEEE Trans Biomed Eng, 2017, 64(10): 2411-2418.
- [9] CHAZAL P. An adapting system for heartbeat classification minimizing user input[C]//36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2014: 82-85.
- [10] GANDOMI A H, TALATAHARI S, YANG X S, et al. Design optimization of truss structures using cuckoo search algorithm[J]. Struct Des Tall Spec, 2016, 22(17): 1330-1349.
- [11] KIRANYAZ S, INCE T, GABBOUJ M. Real-time patient-specific ECG classification by 1-D convolutional neural networks[J]. IEEE Trans Biomed Eng, 2016, 63(3): 664-675.
- [12] YANG S, HOSSEINI S, GANDOMI A H. Algorithm for solving non-convex economic dispatch problems with valve loading effect[J]. Appl Soft Comput, 2015, 12(3): 1180-1186.
- [13] ALMAZAYDEH L, ELLEITHY K, FAEZIPOUR M. Obstructive sleep apnea detection using SVM-based classification of ECG signal features[J]. Conf Proc IEEE Eng Med Biol Soc, 2012: 4938-4941.
- [14] QU X, JIAN C W, FEI G D. ECG signal classification based on BPNN[C]//International Conference on Electric Information and Control Engineering. IEEE, 2011: 1362-1364.
- [15] LIU X, DU H, WANG G, et al. Automatic diagnosis of premature ventricular contraction based on Lyapunov exponents and LVQ neural network[J]. Comput Methods Programs Biomed, 2015, 122(1): 47-55.

(编辑:薛泽玲)