

基于复合特征和 FOAGRNN 的心电图分类

郭 庆, 吴汝琴, 徐翠锋

(桂林电子科技大学 电子工程与自动化学院, 广西 桂林 541004)

摘 要: 为提高心电图分类的准确度, 提出一种基于复合特征和 FOAGRNN 的心电图分类方法. 该方法首先用核独立主元分析(KICA)对心电信号进行非线性特征提取得到特征向量 A, 其次采用小波包变换对心电信号进行多尺度分解, 提取小波包节点系数重构后的归一化能量组成特征向量 B, A 和 B 组合成复合特征向量 C 作为心电信号特征, 再者利用果蝇算法(FOA)优化广义回归神经网络(GRNN)参数构建出 FOAGRNN 模型, 最后利用优化后的分类模型对心电特征进行识别分类. 仿真实验结果表明, FOAGRNN 分类方法较其它方法具有很高的分类准确度, 分类正确率可达到 99.0%.

关键词: 果蝇算法; 广义回归神经网络; 核独立主元分析; 小波包; 心电图分类; 特征提取

中图分类号: R318.04

文献标识码: A

文章编号: 1000-7180(2018)06-0031-05

The Electrocardiogram Classification Approach Based on Wavelet Transform and FOAGRNN

GUO Qing, WU Ru-qin, XU Cui-feng

(School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: In order to improve the accuracy of ECG classification, an ECG classification method based on composite characteristics and FOAGRNN is proposed. First of all, this method begins by using the Kernel Independent Component Analysis (KICA) to get the characteristic vector A from extracting the nonlinear feature of the ECG signals, then, wavelet packet transform is used to analyze multi-scale decomposition of ECG signals, extracting the normalized energy which is reconstructed from wavelet packet node coefficient to compose the characteristic vector B, acting the compound eigenvector C which is composed by A and B as the ECG characteristics. Meanwhile, the fruit fly algorithm (FOA) was used to optimize the generalized regression neural network parameter to construct the FOAGRNN model. Finally, identifying and classifying the ECG characteristics by the optimized classification model. The simulation results show that the classification accuracy of FOAGRNN classification method is so higher compared with other methods, the classification accuracy can reach 99.0%.

Key words: fruit fly algorithm; GRNN; KICA; wavelet packet; ECG classification; feature extraction

1 引言

心电图分类是诊断心血管疾病的重要依据, 近年来国内外对心电图的分类技术进行了广泛的研究, 但效果往往不是十分理想.

心电信号的分类过程实际上是一个模式识别过

程, 即对心电信号进行预处理后利用特定的技术手段进行形态特征提取, 其次建立特征判决模型对特征进行学习, 最后利用学习后的判决模型对心电信号进行分类. 其中最为关键的是特征提取与判决模型的规划和设计^[1]. 目前的特征提取方法局限于单一空间的特征提取, 比如采用小波变换、PCA 等降

收稿日期: 2017-09-08; 修回日期: 2017-10-10

基金项目: 广西自然科学基金(2016GXNSFBA380117); 桂林市科学研究与技术开发项目(2016010404-3); 广西自动检测技术与仪器重点实验室基金(YQ17116, YQ16114, YQ15116)

维优化方法,提取到的特征并不能最大反映心电波形包含的信息.此外,目前多数采用的分类模型大多是 BP 神经网络、概率神经网络或支持向量机等,模型的泛化能力并不突出.

本文提出了一种基于复合特征和 FOAGRNN 的心电分类方法.

2 心电复合特征提取

2.1 心电预处理

心电信号是一种微弱且易受噪声干扰的信号,噪声的干扰会影响所提取特征的质量,从而会降低分类模型的准确度,因此对波形进行预处理去噪很有意义.

小波去噪即是利用小波变换特性,将原始心电信号进行多层分解,并根据噪声信号所在的频段对相应的频段置零,最后对有效信号进行重构即可得到消除噪声的心电信号.利用小波变换用小波基“sym8”函数对心电图进行 8 层小波分解可知,肌电干扰、工频干扰主要集中在第 1 第 2 尺度,而第 8 尺度下的逼近信号充分体现基线漂移特性,故我们把 1,2,8 尺度下的小波系数 D1,D2,A8 置零即可消除心电信号中的大部分干扰^[2].以 MIT-BIH 心率失常数据库中的 105 号部分心电图波形为例,如图 1 所示,经过小波变换去噪后的心电图波形表面更加光滑,特征更为明显.

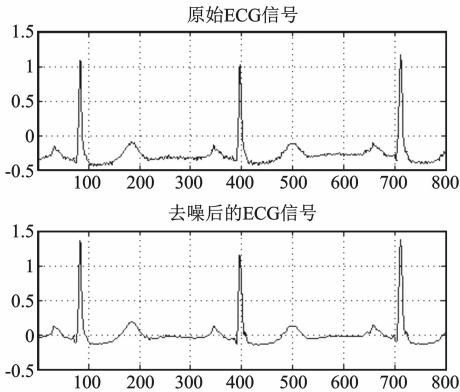


图 1 原始心电信号 VS 去噪后的心电信号

2.2 基于 KICA 的特征提取

KICA(Kernel Independent Component Analysis)是在 ICA 的基础上进行的,其基本思想是将原始输入数据映射到高维特征空间进行 ICA 分析,使特征空间尽可能呈现非 Gauss 性.对于任意一个周期的心电信号,其组成可以看作是一组相互独立的统计基信号的混叠, m 个周期信号的观测信号

矩阵为: $\mathbf{X} = (x_1, x_2, \dots, x_m)T$,KICA 对观测矩阵 $\mathbf{X}$ 进行分解以后得到一组统计独立的基信号 s_j ,在这组基信号矢量构造得到特征子空间,并将每一个心电信号样本 x_p 投影到这个空间,也就是说用独立基信号的线性混合表示心电信号.求伪逆的方法用以得到待识别信号的投影系数矢量,公式如下^[3]:

$$\mathbf{A}_p = f_p \times S^+ \quad (1)$$

式中, $\mathbf{A}_p = (a_{p1}, a_{p2}, \dots, a_{pm})T$, S^+ 代表独立分量矩阵 S 的伪逆,待识别心电信号 f_p 的特征矢量为投影系数矢量 $\mathbf{A}_p$.这里我们用 KICA 算法对每一个心电信号样本进行特征提取获取 20 维特征向量 $\mathbf{A}$.

2.3 基于小波包变换的特征提取

小波包变换的任意多尺度分解特性,是分析非平稳信号更为有效的方法,将其用于心电信号特征的提取,相比小波分析能获取更优质的特征,其用于心电特征提取的基本思想是将心电信号分解到多个子带频域上^[4],并对小波包节点的系数进行重构,获取节点能量.

小波包对心电样本信号进行 n 层分解,那么分解后获取的心电能量向量为 $\mathbf{E} = [\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_{2n}]$,为减少后面网络训练的计算量,对 $\mathbf{E}$ 进行归一化处理,即可得到心电特征向量 $\mathbf{ET} = [\mathbf{ET}_1, \mathbf{ET}_2, \dots, \mathbf{ET}_{2n}]$.这里我们选择对心电样本信号进行 3 层分解获取 8 维特征向量 $\mathbf{B}$,将其作为心电特征的辅助特征,最后把向量 $\mathbf{A}$ 和向量 $\mathbf{B}$ 组合形成我们需要的复合特征向量 $\mathbf{C}$,即 28 维特征向量.

3 FOAGRNN 算法

3.1 广义回归神经网络

GRNN 网络是建立在数理统计基础上的径向基函数网络,其理论基础是非线性回归分析^[5].GRNN 具有很强的非线性映射能力和学习速度,相比于 RBF 网络具有更大的优势.它的结构由输入层、模式层、求和层、输出层组成,如图 2 所示.输入层负责接收样本数据的输入,模式层的单元个数与训练样本个数相同,各单元分别对应相应的样本,求和层使用两种不同的神经元进行求和,输出层中的神经元数目等于学习样本中输出向量的维

数 k ,各个神经元将求和层中的两个输出进行相除,即可得到对应的估计结果^[6].GRNN 具备的优点如下:

(1)网络训练简单,学习样本经过模式层和求和层及完成训练,不需要像前馈神经网络一样需要反复迭代权值和阈值.

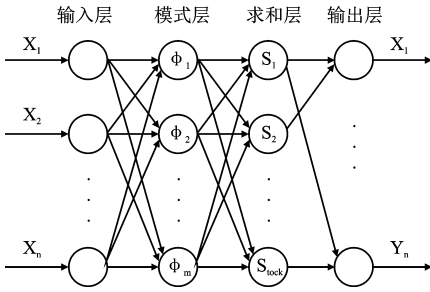


图 2 GRNN 结构图

(2) 网络的输出性能只受一个自由参数即径向向基的平滑因子影响。

(3) 相比于前馈神经网络容易停滞在局部收敛, GRNN 计算结果具有全局收敛性^[7]。

3.2 FOAGRNN 模型

FOA 算法^[8]是基于果蝇觅食的一种全局寻优算法。果蝇的嗅觉与视觉相当敏锐,能很好地搜集漂浮在空气中的各种气味,甚至能嗅到 40km 以外的食物源。果蝇根据嗅觉飞近食物位置后亦可使用敏锐的视觉发现食物与同伴聚集的位置,并且往该方向飞去^[9]。此算法正是基于果蝇这一特性提出并得以应用到各个领域。相关研究发现 FOA 算法比目前主流的群体优化算法:遗传算法、蚁群算法、鱼群算法和粒子群算法等具有算法简单、参数少、易调节、计算量小、寻优精度高的优点。

GRNN 网络的输出性能只受平滑因子的影响,使用果蝇算法优化 GRNN 网络即是利用 FOA 算法的收敛速度快与收敛精度高的性能快速寻找最优的 SPREAD 得到最优的 GRNN 网络。FOA 优化 GRNN 的流程图如图 3,详细步骤如下^[10]:

(1) 初始果蝇群体位置,设置最大迭代次数为 100,设置果蝇种群规模为 10。

$$X_axis = \text{rand}() \quad (2)$$

$$Y_axis = \text{rand}() \quad (3)$$

(2) 随机赋予果蝇飞行方向,增加果蝇个体的飞行距离。

$$X(i) = X_axis + 2 * \text{rand}() - 20 \quad (4)$$

$$Y(i) = Y_axis + 2 * \text{rand}() - 20 \quad (5)$$

(3) 求出果蝇个体与原点的距离;求出味道浓度判定值 — 距离的倒数。

$$D(i) = \sqrt{X(i)^2 + Y(i)^2} \quad (6)$$

$$S(i) = 1/D(i) \quad (7)$$

(4) 利用味道浓度判定函数求出味道浓度,这里的味道浓度表示为 GRNN 的 SPREAD 值。

$$\text{SPREAD} = S(i) \quad (8)$$

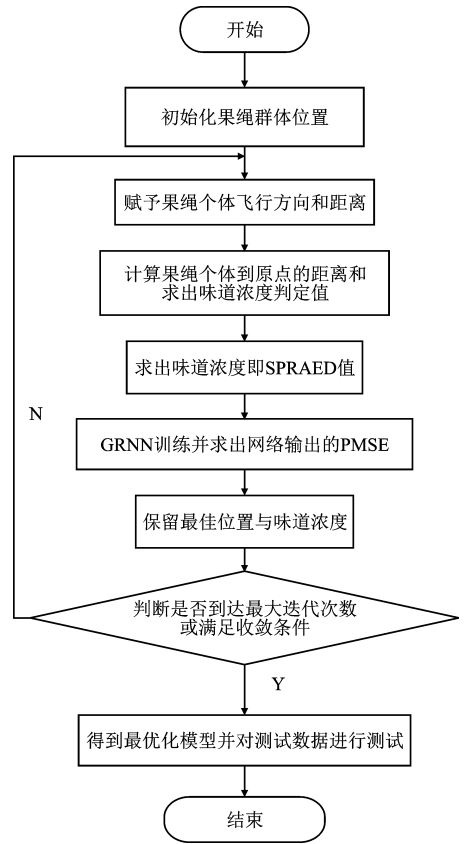


图 3 FOAGRNN 流程图

(5) 利用味道浓度带入 GRNN 进行训练计算,得到每一个果蝇的味道浓度对应的网络输出的 RMSE。

$$\text{Smell}(i) = \text{RMSE} \quad (9)$$

(6) 根据各个 RMSE 保存最佳的果蝇个体对应的最佳位置。并与网络期望输出和最大迭代次数进行比较。如果收敛于输出条件或达到最大迭代次数,则迭代停止,反之,重复步骤(2) ~ (6)。

$$[\text{bestSmell} \text{ bestindex}] = \min(\text{Smell}) \quad (10)$$

$$X_axis = X(\text{bestindex}) \quad (11)$$

$$Y_axis = Y(\text{bestindex}) \quad (12)$$

(7) 得到最优的 FOAGRNN 分类模型,并使用该模型对测试数据进行测试。

4 实例仿真与结果分析

采用 MIT-BIH 标准心电图数据库中的 5 类典型心电图数据进行特征提取,并将其分为 5 类:正常的心电信号(Normal)、左束支传导阻滞(LBBB)、右束支传导阻滞(RBBB)、心室早期收缩(PVC)、心房早期收缩(APC)。这些数据的来源组号与对应的分类模型输出向量见表 1。

表 1 心电类型来源及输出信息

心电类型	MIT-BIH	输出向量
Normal	101. dat	[1 11 1 1]
LBBB	109. dat	[1 11 1 0]
RBBB	118. dat	[1 11 0 1]
PVC	106. dat	[1 1 0 11]
APC	209. dat	[1 0 11 1]

首先利用小波变换检测出 QRS 波群,确定 R 波位置,以每个心拍的 R 点为中心,取其前后各 150 个数据点、一共 300 个数据作为训练测试的一个样本,其次利用能够 KICA 和小波包变换提取心电信号每一个样本对应的复合特征. 对每一种类型的心电数据各采取 200 组样本,共采取 1 000 组样本作为训练测试. 其中将每一类样本的 120 组作为训练样本,剩余的 80 组作为测试样本.

建立 FOAGRNN 模型,设置网络输出的 RMSE 期望值是 $1.0\text{e-}08$,对训练样本进行训练,其训练过程的果蝇个体每一次迭代的最佳飞行路线如图 4,对应的网络输出 RMSE 变化曲线如图 5. 可以看到,使用 FOAGRNN 模型只需要迭代 8 次就可找到最优 SPREAD 值,网络输出 RMSE 为 $4.27\text{e-}10$,从而确定最优的 FOAGRNN 模型. 使用最优模型对测试样本进行测试,结果如表 2. 由表看出使用 FOAGRNN 模型对心电进行分类其分类的准确率可高达 99.0%.

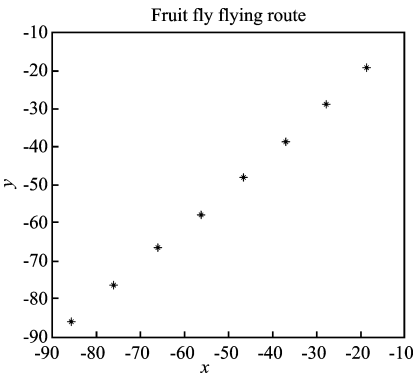


图 4 最佳果蝇个体飞行路线

表 2 FOAGRNN 测试结果

心电类型	样本数/个	正确数/个	正确率/%
Normal	80	80	100
LBBB	80	79	98.75
RBBB	80	79	98.75
PVC	80	80	100
APC	80	78	97.5

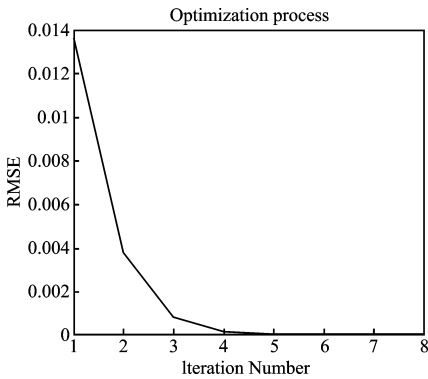


图 5 网络输出 RMSE 变化曲线

为验证 FOAGRNN 分类模型的优越性,建立目前心电分类领域多数采用的 BPNN、PNN 模型进行对比. 设置 BPNN 模型最大训练次数为 30 000,学习速率设置为 0.01,训练函数采用“trainlm”函数,收敛条件为 RMSE 小于等于 $1.0\text{e-}05$. 对 3 种不同模型分别进行 10 次训练测试,结果如表 3 所示.

表 3 三种模型分类结果

分类模型	FOAGRNN	BPNN	PNN
分类率/%	99.0	86.5	96.7

由表可知 FOAGRNN 模型的分类准确率明显高于其它两种模型,并保持在 99.0%.

5 结束语

提出了基于复合特征和 FOAGRNN 的心电图分类新方法,利用 KICA 的非线性分析能力提取 20 维特征,并采用小波包变换的时频域分析特性提取心电图信号 8 维能量特征作为辅助特征,两者组合成复合特征. 结合 FOA 算法和 GRNN 网络的优点建立快速寻优的 FOAGRNN 分类模型,并通过实际实验测试验证该方法的优劣性. 仿真结果表明,本文心电分类方法较其它方法具有相当大的优越性,平均分类正确率为 99.0%.

参考文献:

[1] 周珂,彭宏,胡劲松. 基于小波神经网络方法的心电图分类研究[J]. 微电子学与计算机, 2007, 24(5): 127-129.

[2] 范玉庆,王小华,曹志锋,等. 小波变换与概率神经网络的心电图分类[J]. 电子技术应用, 2013, 39(3): 136-137.

[3] 赵勇,洪文学,孙士博. 基于多特征和支持向量机的心律失常分类[J]. 生物医学工程学, 2011(2): 292-295.

[4] 张征凯,谷立臣,曹向辉,等. 基于最优小波包的电机定子电流故障特征提取方法[J]. 煤矿机械, 2016, 37(8):149-151.

[5] 王少福,张金磊,赵仕俊,等. 广义回归神经网络的改进及在预测控制中的应用[J]. 微电子学与计算机, 2009, 26(6):32-35.

[6] 郁磊. MATLAB 智能算法 30 个案例分析. 第 2 版[M]. 北京航空航天大学出版社, 2015.

[7] 洪波,刘龙,王涛. 修正型果蝇算法优化 GRNN 的大梁自动焊障碍预测[J]. 焊接学报, 2017, 38(1): 73-76.

[8] Pan W T. Pan, W. T. A new Fruit Fly opti- mization Algorithm; Taking the financial distress model as an example. Knowledge-Based Systems 26, 69-74 [J]. Knowledge-Based Systems, 2012, 26(2):69-74.

[9] 祁丽婉,梁庚,童国炜. 基于果蝇算法优化 BP 神经网络

的齿轮箱故障诊断[J]. 电网与清洁能源, 2014(9): 31-36.

[10] 田源,张彼德,刘代伟,等. 基于果蝇优化算法的 GRNN 水电机组状态趋势预测[J]. 水电能源科学, 2012(12):127-129.

作者简介:

郭庆 男,(1962-),教授,硕士生导师. 研究方向为嵌入式测控系统、微弱信号检测.

吴汝琴 女,(1992-),硕士研究生. 研究方向为心电检测与分类.

徐翠锋(通信作者) 女,(1977-),硕士,高级实验师. 研究方向为信号处理、测控技术等.

E-mail:343183324@qq. com.

(上接第 30 页)

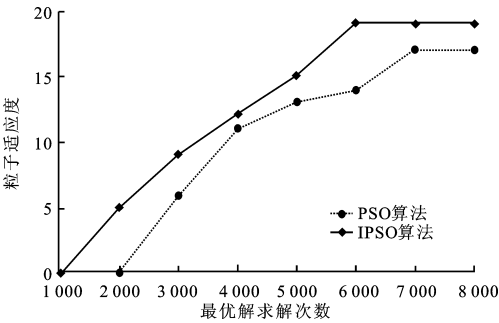


图 2 大任务量实验结果

6 结束语

改进的粒子群算法 IPSO,通过改进 PSO 算法中个体项和社会认知项公式,更能反映粒子位置、速度更新与迭代次数之间的关联性. 通过不同规模实验的对比,验证了 IPSO 算法的收敛性、适应性强于 PSO 算法,能有效解决云计算资源调度中易陷入局部最优解的问题. 改进的 IPSO 算法节约了云资源分配、完成任务的时间,是较理想的方案.

参考文献:

[1] Lee Y C, Zomaya A Y. Energy efficient utilization of

resources in cloud computing systems[J]. Journal of Supercomputing, 2012, 60(2):268-280.

[2] Iosup A, Ostermann S, Yigitbasi M N, et al. Performance analysis of cloud computing services for many-tasks scientific computing[J]. IEEE Transactions on Parallel & Distributed Systems, 2011, 22 (6): 931-945.

[3] 孙大为,常桂然,李风云,等. 一种基于免疫克隆的偏好多维 QoS 云资源调度优化算法[J]. 电子学报, 2011, 39(8):1824-1831.

[4] 孟令玺,李洪亮. 基于 CA-PSO 算法的云计算资源调度策略[J]. 计算机仿真, 2013, 30(10):406-410.

[5] 刘卫宁,靳洪兵,刘波. 基于改进量子遗传算法的云计算资源调度[J]. 计算机应用, 2013, 33(8): 2151-2153.

[6] 袁正午,李君琪. 基于改进粒子群算法的云资源调度[J]. 计算机工程与设计, 2016, 37(2):401-404.

作者简介:

蔡晓丽 女,(1981-),硕士,副教授. 研究方向为云计算、网络安全. E-mail:caixl@czu. cn.

钱 诚 男,(1981-),博士,讲师,研究方向为云计算、物联网和移动计算.