



(12)发明专利申请

(10)申请公布号 CN 111557659 A

(43)申请公布日 2020.08.21

(21)申请号 202010445892.1

(22)申请日 2020.05.22

(71)申请人 郑州大学

地址 450000 河南省郑州市高新技术开发
区科学大道100号

(72)发明人 李润川 冀沙沙 申圣亚 王宗敏
周兵

(74)专利代理机构 郑州裕晟知识产权代理事务
所(特殊普通合伙) 41142

代理人 徐志威

(51)Int.Cl.

A61B 5/0402(2006.01)

A61B 5/0452(2006.01)

权利要求书1页 说明书7页

(54)发明名称

基于多特征融合与Stacking-DWKNN的心律
失常分类方法

(57)摘要

本发明涉及一种基于多特征融合与Stacking-DWKNN的心律失常分类方法,包括以下步骤:S1、采用连续小波变化去除心电信号中的噪声;S2、对经过步骤S1处理后的心电信号进行分割截取出完整的心搏,然后从截取出的心搏中进行特征提取,并将提取到的特征按类别建立以下数据集:集合A={235单心搏形态特征},集合B={P-QRS-T波},集合C={PR间期},集合D={QT间期},集合E={ST段},集合F={RR间期},集合G={R幅值},集合H={T幅值};S3、将步骤S2中的数据集中的任一个集合或任意多个集合的组合输入到采用Stacking集成多个通过权值改进的KNN算法中进行心搏分类;本发明提供的心搏分类方法能够有效提高心搏分类的结果准确性。

1. 一种基于多特征融合与Stacking-DWKNN的心律失常分类方法,其特征在于,包括以下步骤:

S1、采用连续小波变化去除心电信号中的噪声;

S2、对经过步骤S1处理后的心电信号进行分割截取出完整的心搏,然后从截取出的心搏中进行特征提取,并将提取到的特征按类别建立以下数据集:

集合A = {235单心搏形态特征},

集合B = {P-QRS-T波},

集合C = {PR间期},

集合D = {QT间期},

集合E = {ST段},

集合F = {RR间期},

集合G = {R幅值},

集合H = {T幅值};

S3、将步骤S2中的数据集中的任一个集合或任意多个集合的组合输入到采用Stacking集成多个通过权值改进的KNN算法中进行心搏分类。

2. 根据权利要求1所述的基于多特征融合与Stacking-DWKNN的心律失常分类方法,其特征在于:所述235单心搏形态特征的提取方法为,利用注释文件中确定的位置,从心电图记录单个导联中提去除235个R峰附近的点,其中,R峰前有90个采样点,R峰后有144个采样点。

3. 根据权利要求1所述的基于多特征融合与Stacking-DWKNN的心律失常分类方法,其特征在于:将步骤S2中所述数据集中的集合B、集合C、集合D、集合E、集合F组合后定义为新的集合I = {P-QRS-T波、PR间期、QT间期、RR间期、ST段}。

4. 根据权利要求1所述的基于多特征融合与Stacking-DWKNN的心律失常分类方法,其特征在于:将数据集中集合A、集合G以及集合I中的一个集合或任意多个集合的组合输入到采用Stacking集成多个通过权值改进的KNN算法模型中进行心搏分类。

5. 根据权利要求4所述的基于多特征融合与Stacking-DWKNN的心律失常分类方法,其特征在于,所述步骤S3中Stacking-DWKNN算法的训练过程如下:以DWKNN算法作为Stacking的基分类器,将所述数据集输入到Stacking的第一层基分类器DWKNN算法中,产生的新的数据集输入到Stacking的第二层基分类器DWKNN算法中输出最终分类结果。

6. 根据权利要求5所述的基于多特征融合与Stacking-DWKNN的心律失常分类方法,其特征在于,所述DWKNN算法的实现方式为:

S61、计算测试数据与个训练数据之间的距离;

S62、根据计算得出的距离按递增顺序进行排列,选取距离最小的K个点;

S63、确定前K个点所在类别的出现频率;

S64、返回前K个点中出现频率最高的类别作为测试数据的预测分类;

所述K的近邻算法中的度量函数采用曼哈顿距离,赋予样本点不同的权重作为KNN算法的度量函数。

7. 根据权利要求5所述的基于多特征融合与Stacking-DWKNN的心律失常分类方法,其特征在于:所述K的取值范围为1-10。

基于多特征融合与Stacking-DWKNN的心律失常分类方法

技术领域

[0001] 本发明属于心律不齐分类方法技术领域,基于多特征融合与Stacking-DWKNN的心律失常分类方法。

背景技术

[0002] 心电图 (ECG) 是一种非侵入性,廉价且成熟的诊断工具,广泛应用于多种应用。它代表心脏电活动随时间的变化,并包含广泛用于分析心脏功能的基本生理信息,这对于心律不齐的检测很重要。大多数心律不齐是无害的,但有些可能会立即危及人们的生命。因此,准确检测患者的心律不齐在预防心血管疾病中发挥着至关重要的作用。心电图 (ECG) 具有易于获取和设备成本低的优点,可用于判断心律不齐是窦性还是异位性,是初步诊断心律不齐的重要依据。通过逐一分析心电图上波形的变化,可以检测出不同类型的心律不齐。然而,在不同条件下,ECG信号的形态和时间特征有着显著的差异,部分心律不齐仅在患者的日常生活中偶尔发生,需要使用动态心电图记录长时间的ECG活动。传统的动态心电图分析是手动完成的,但逐拍手动分析长期心电图是耗时且不切实际的。因此智能诊断在医学中是至关重要的,它可以帮助医生更好地诊断心律不齐。

发明内容

[0003] 本发明的目的是克服现有技术的不足而提供一种基于多特征融合与Stacking-DWKNN的心律失常分类方法。

[0004] 本发明的技术方案如下:

[0005] 一种基于多特征融合与Stacking-DWKNN的心律失常分类方法,包括以下步骤:

[0006] S1、采用连续小波变化去除心电信号中的噪声;

[0007] S2、对经过步骤S1处理后的心电信号进行分割截取出完整的心搏,然后从截取出的心搏中进行特征提取,并将提取到的特征按类别建立以下数据集:

[0008] 集合A = {235单心搏形态特征},

[0009] 集合B = {P-QRS-T波},

[0010] 集合C = {PR间期},

[0011] 集合D = {QT间期},

[0012] 集合E = {ST段},

[0013] 集合F = {RR间期},

[0014] 集合G = {R幅值},

[0015] 集合H = {T幅值};

[0016] S3、将步骤S2中的数据集集中的任一个集合或任意多个集合的组合输入到采用Stacking集成多个通过权值改进的KNN算法中进行心搏分类。

[0017] 进一步的,所述235单心搏形态特征的提取方法为,利用注释文件中确定的位置,从心电图记录单个导联中提去除235个R峰附近的点,其中,R峰前有90个采样点,R峰后有

144个采样点。

[0018] 进一步的,将步骤S2中所述数据集中的集合B、集合C、集合D、集合E、集合F组合后定义为新的集合I={P-QRS-T波、PR间期、QT间期、RR间期、ST段}。

[0019] 进一步的,将数据集中集合A、集合G以及集合I中的一个集合或任意多个集合的组合输入到采用Stacking集成多个通过权值改进的KNN算法中进行心搏分类。

[0020] 进一步的,所述步骤S3中Stacking-DWKNN算法的训练过程如下:以DWKNN算法作为Stacking的基分类器,将所述数据集输入到Stacking的第一层基分类器DWKNN算法中,产生的新的数据集输入到Stacking的第二层基分类器DWKNN算法中输出最终分类结果。

[0021] 所述DWKNN算法的实现方式为:

[0022] S61、计算测试数据与个训练数据之间的距离;

[0023] S62、根据计算得出的距离按递增顺序进行排列,选取距离最小的K个点;

[0024] S63、确定前K个点所在类别的出现频率;

[0025] S64、返回前K个点中出现频率最高的类别作为测试数据的预测分类;

[0026] 所述K的近邻算法中的度量函数采用曼哈顿距离,赋予样本点不同的权重作为KNN算法的度量函数。

[0027] 优选的,所述K的取值范围为1-10。

[0028] 与现有技术相比,本发明的有益效果是:

[0029] 本发明将经过连续小波变换去除噪声后进行特征提取,将提取到的特征根据其类型建立八个集合的数据集,再将数据集中的一个或多个集合的组合输入到Stacking-DWKNN算法中进行心搏分类,有效提高心搏分类的结果准确性。

具体实施方式

[0030] 下面将结合本发明实施例,对本发明中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0031] 一种基于多特征融合与Stacking-DWKNN的心律失常分类方法,包括以下步骤:

[0032] S1、采用连续小波变化去除心电信号中的噪声;

[0033] S2、对经过步骤S1处理后的心电信号进行分割截取出完整的心搏,然后从截取出的心搏中进行特征提取,并将提取到的特征按类别建立以下数据集:

[0034] 集合A={235单心搏形态特征},

[0035] 集合B={P-QRS-T波},

[0036] 集合C={PR间期},

[0037] 集合D={QT间期},

[0038] 集合E={ST段},

[0039] 集合F={RR间期},

[0040] 集合G={R幅值},

[0041] 集合H={T幅值};

[0042] S3、将步骤S2中的数据集中的任一个集合或任意多个集合的组合输入到采用

Stacking集成多个通过权值改进的KNN算法中进行心搏分类。

[0043] 进一步的,所述235单心搏形态特征的提取方法为,利用注释文件中确定的位置,从心电图记录单个导联中提去除235个R峰附近的点,其中,R峰前有90个采样点,R峰后有144个采样点。

[0044] 进一步的,将步骤S2中所述数据集中的的集合B、集合C、集合D、集合E、集合F组合后定义为新的集合I={P-QRS-T波、PR间期、QT间期、RR间期、ST段}。

[0045] 进一步的,将数据集中集合A、集合G以及集合I中的一个集合或任意多个集合的组合输入到采用Stacking集成多个通过权值改进的KNN算法中进行心搏分类。

[0046] 进一步的,所述步骤S3中Stacking-DWKNN算法的训练过程如下:以DWKNN算法作为Stacking的基分类器,将所述数据集输入到Stacking的第一层基分类器DWKNN算法中,产生的新的数据集输入到Stacking的第二层基分类器DWKNN算法中输出最终分类结果。

[0047] 所述DWKNN算法的实现方式为:

[0048] S61、计算测试数据与个训练数据之间的距离;

[0049] S62、根据计算得出的距离按递增顺序进行排列,选取距离最小的K个点;

[0050] S63、确定前K个点所在类别的出现频率;

[0051] S64、返回前K个点中出现频率最高的类别作为测试数据的预测分类;

[0052] 所述K的近邻算法中的度量函数采用曼哈顿距离,赋予样本点不同的权重作为KNN算法的度量函数。

[0053] 本实施中,所述K的取值范围为1-10,优选K取值为4。

[0054] 本实施中,Stacking集成方法的核心思想是使用初始数据集训练第一层分类器,产生新的数据集训练第二层的分类器,并将第一层分类器的输出作为第二层分类器的输入特征,同时原始标签仍被作为新数据集的标签。Stacking就是进一步进行泛化来综合降低偏差和方差的方法。第一层分类器通常是使用不同的机器学习算法产生,因此,Stacking算法通常是异质集成,有时也会构造同质的stacked generation。

[0055] 所述Stacking集成方法的算法描述如下:

输入: 数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$;

一层学习算法 ζ_1, \dots, ζ_T ;

二层学习算法 ζ .

1: 步骤一: 通过训练数据 T 来训练一层学习算法来产生初级学习器

2: for $t=1, \dots, T$ do

3: $h_t = \zeta_t(D)$

4: end for

5: 步骤二: 产生一个新的训练集 D'

[0056]

6: $D' = \emptyset$

7: for $I = 1, \dots, m$ do

8: for $t=1, \dots, T$ do

9: $Z_{it} = h_t(x_i)$

10: end for

11: $D' = D' \cup ((Z_{i1}, \dots, Z_{iT}), y_i)$

12: end for

13: $h' = \zeta(D')$

输出: $H(x) = h'(h_1(x), \dots, h_T(x))$

[0057] 本实施例中, KNN作为一种高度灵活、通用性强的分类算法, 具有广阔的应用前景。在实际应用中, 由于其精度高, 对异常值不敏感, 无需数据输入假定, 对各种各样的样本分布都能进行分类, 并且对小样本数据KNN也能有很好的分类能力, 经常被应用于分类问题中, 所述KNN算法描述如下:

[0058]

算法: KNN 模型训练

输入：训练数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ ；

其中， $x_i \in X \in R^n$ 为实例的特征向量， $y_i \in Y = \{c_1, c_2, \dots, c_k\}$ 为

实例的类别， $i = 1, 2, \dots, N$ ；实例特征向量 x ；

输出：根据公式 (1) 的多数投票原则确定实例 x 所属的类 y 。

过程：

[0059]

1: for 每个测试样例 $z = (x', y')$ do

2: 根据给定的距离度量，在训练集 T 中找出与 x 最邻近的 K 个

点，涵盖这 k 个点的 x 的邻域记作 $N_k(x)$ ；

3: 在 $N_k(x)$ 中根据分类决策规则决定 x 的类别 y ；

4: end for

[0060] 其中，公式 (1) 为 $y = \arg \max_{c_j} \sum_{x_i \in N_k(x)} I(y_i, c_j)$ ， $i = 1, 2, \dots, N$ ； $j = 1, 2, \dots, K$ ；

[0061] 所述公式 (1) 中 I 的计算公式为公式 (2)： $I(y_i, c_j) = \begin{cases} 0, & y_i = c_j \\ 1, & y_i \neq c_j \end{cases}$ 。

[0062] 本实施例中，采用曼哈顿距离计算训练集和训练样本之间的距离，定义为：

[0063] $D(x, y) = |x_i - x_j| + |y_i - y_j|$ 公式 (3)；

[0064] 所述曼哈顿距离作为 K 的邻近算法中度量函数，将样本的不同特征量赋予相同的权重，但是不同特征量对分类结果准确性影响不同，因此本实施例根据距离远近，赋予样本点不同的权重作为 KNN 算法的度量函数，前 K 个点出现的概率工时如下：

[0065] $f(x_i) = (w_1 f(x_1) + w_2 f(x_2) + \dots + w_m f(x_m)) / (w_1 + \dots + w_m)$ 公式 (4)。

[0066] 实验及结果：

[0067] 以下实验均在 MIT-BIH 心律失常数据库上进行。

[0068] 根据 ANSI/AAMI EC57:1998 标准，心律失常的 16 种类别被分为 MIT-BIH 心律失常数据库中五个类别，即 N (正常或束支传导阻滞)， S (室上性异位搏动)， V (心室异位搏动)， F (融合搏动) 和 Q (未指定搏动)。

[0069] 本实施例实验基于 MIT-BIH 心律失常数据库中 101413 个心搏进行分类，包括 90,595 个 N 类别的心搏；2,781 个 S 类别的心搏；7,235 个 V 类别的心搏； F 类别的心搏数量仅有 802 个。本实施例实验数据从数据集中随机选取心搏数据的 90% 作为训练集，剩余的 10% 测试集。从各类别的心搏数量可以看出，本实验是在不平衡数据下进行的。数据的具体分布如下表。

[0070] 实验数据统计

	训练集	测试集	总计
N	81,560	9,035	90,595
[0071] S	2,528	253	2,781
V	6,450	785	7,235
F	723	79	802

[0072] 本实施实验中的评估指标：

[0073] 本实施例将测试结果分为四种，即，真阴性 (TN)：预测为负向，实际上也是负向的数量；假阴性 (FN)：漏报率，把正向归类为负向的数量；真阳性 (TP)：预测为正向，实际上也是正向的数量；假阳性 (FP)：误报率，把负向归类为正向的数量。其中，N类心搏的计算公式如 (5) - (8)，其它心搏类别的分类结果按照同样的方式计算。表1显示了分类结果的混淆矩阵。本实施例实验中N,S,V,F,Q表示心搏的真实类别；n,s,v,f,q表示预测结果。

[0074] 表1：分类结果的混淆矩阵

	n	s	v	f	to tal
[0075] N	N _n	N _s	N _v	N _f	\sum^N
S	S _n	S _s	S _v	S _f	\sum^S
V	V _n	V _s	V _v	V _f	\sum^V
[0076] F	F _n	F _s	F _v	F _f	\sum^F

[0077] $TP_N = N_n$ (5)

[0078] $FN_N = N_s + N_v + N_f$ (6)

[0079] $TN_N = S_s + S_v + S_f + V_s + V_v + V_f + F_s + F_v + F_f$ (7)

[0080] $FP_N = S_n + V_n + F_n$ (8)

[0081] 本实施使用灵敏度、特异性、阳性预测值和准确度评估分类器性能。灵敏度 (se) 指被判断为正例的样本占有所有正例的比例。灵敏度越高，正确预测的样本越多。特异性 (sp) 指被判断为负例的样本占有所有负例的比例。阳性预测值 (+p) 在文献中也称为精确度。准确度

是被正确分类的样本数占总样本数的比率,反映了测试结果与实际结果之间的一致性。上述四个评价指标的计算公式如下:

[0082] $Se = TP / (TP + FN)$ (9)

[0083] $Sp = TN / (TN + FP)$ (10)

[0084] $+p = TP / (TP + FP)$ (11)

[0085] $Acc = (TP + TN) / (TP + TN + FP + FN)$ (12)。

[0086] 对于采用前述集合A、B、C、D、E、F、G、H中的一个或多个组成的不同特征组合的实验结果进行分析,作为优选的,本实施例实验数据以集合A、由集合B、集合C、集合D、集合E、集合F重新定义的集合I、集合G为输入数据集,以GNB、LDA、LR、SVM、DT、GBDT、RF、KNN、DWKNN分类器作为参照实验组,以Stacking-DWKNN算法作为实验组,进行实验分析,分析过程如下:

[0087]

特征组合			分类器									
集合 A	集合 I	集合 G	GNB	LDA	LR	SVM	DT	GBDT	RF	KNN	DWKNN	Stacking -DWKNN
●			51.06	91.71	92.39	97.35	97.35	97.61	98.55	98.88	98.95	98.94
	●		86.84	89.06	93.69	94.45	93.47	94.45	95.73	95.41	95.36	95.95
		●	89.49	90.42	90.59	90.58	87.98	90.58	87.81	90.26	88.47	90.43
●	●		51.33	93.03	98.69	97.73	97.43	97.73	98.57	98.87	98.97	98.97
●		●	51.08	91.74	98.69	97.55	97.47	97.55	98.58	98.89	98.94	
	●	●	88.85	90.14	95.71	96.01	95.74	96.01	97.08	98.89	96.93	97.09
●	●	●	52.20	93.82	98.74	98.15	97.80	98.15	98.83	98.91	98.96	99.01

[0088]

[0089] 由上表可知,搞死朴素贝叶斯 (GNB) 是在集合G上的分类结果最好,其它分类器均是在集合A、集合I、集合G上的分类结果最好,且KNN的分类结果由于其它几种分类器,通过权值改进KNN算法 (DWKNN),分类结果得到了改进,但仅使用集合I或者集合G进行心搏分类时分类结果稍有下降,说明这两个特征不适合该算法。通过Stacking集成多个KNN模型改进KNN算法,分类结果都得到了提升,最好的分类结果为99.01%。

[0090] 尽管参照前述实施例对本发明进行了详细的说明,对于本领域的技术人员来说,其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。