



(12)发明专利申请

(10)申请公布号 CN 110192851 A

(43)申请公布日 2019.09.03

(21)申请号 201910538631.1

(22)申请日 2019.06.20

(71)申请人 鲁东大学

地址 264025 山东省烟台市芝罘区红旗中路184号

(72)发明人 刘通 李春涛 臧睦君 邹海林
柳婵娟 周树森 赵玲玲

(74)专利代理机构 长春市东师专利事务所
22202

代理人 张铁生 刘延军

(51)Int.Cl.

A61B 5/0245(2006.01)

A61B 5/00(2006.01)

权利要求书2页 说明书4页 附图1页

(54)发明名称

一种基于机器学习技术的心律失常分析系统

(57)摘要

本发明提供了一种基于机器学习技术的心律失常分析系统,它包括:一.读入数据;二.对单通道心电图样本的第一导联进行特征提取,三.对单通道心电图样本的第二导联进行特征提取,四.输入分类器进行识别。相比于目前已公开的技术方案,本发明的主要区别技术特征是通过采样生成单通道心电图样本,并利用相似度特征进行特征提取,最后通过KNN分类器对提取的特征进行识别。这种方法表现出了与现有特征提取方法相同的准确率,为实现此系统提供了新的技术方案。

1. 一种基于机器学习技术的心律失常分析系统,它包括:

一. 读入数据

外部获取一组两导联心电信号输入到MATLAB中;

设置一个空数组A,m行 $2*n$ 列;

根据信号标注的R波点的位置按照其采样频率每个导联依次截取相应个点,两个导联截取出的数据依次存入数组A;

数组A中前m行n列的数据为第一导联的数据,m行n+1到 $2*n$ 列为第二导联的数据;

二. 对单通道心电图样本的第一导联进行特征提取的步骤:

(1) 设置i,将A中前m行n列的数据进行i次采样取A的1到m行 $1+2*(i-1)$ 到 $3+2*(i-1)$ 列的采样方法进行采样变为 $m*x$ 的样本信号B,B中包含m个 $1*x$ 的数组,设置空数组Y1;

(2) 对B进行聚类,聚类数为x,这里假设x为3,聚类后得到大小为 $x*x$ 的信号,并命名为C;

(3) 设置a1到a3为C的第1到x行数组,对这三个数组进行求距离:先设定数组a1的距离值 $q1$ 为 $2*i$ (i与步骤一中的i相同),求a1 a2之间的欧式距离 $s1$;a1 a3之间的欧式距离 $s2$;

然后求得a2的绝对距离 $q2=2*i+s1$;a3的绝对距离 $q3=2*i+s2$;

将 $q1$ 的值放入C第1行第4列; $q2$ 的值放入C第2行第4列; $q3$ 的值放入C第3行第4列;

(4) 将数组B对a1 a2 a3求相似性:

求B的第j行的数组 ($j=1,2,\dots,m$) 与a1 a2 a3的欧式距离 $d1$ $d2$ $d3$,B的第j行的数组与a1的相似性 $m1=1/(d1-1)$,相应的与a2 a3的相似性为 $m2$ $m3$;

比较 $m1$ $m2$ $m3$ 三个值,取最大值 m_k ($k=1,2,3$),将C的第k行第4列的值赋给Y1的第j行第i列;

(5) 输出Y1;

三. 对单通道心电图样本的第二导联进行特征提取的步骤:

(1) 设置i,将A中m行n+1到 $2*n$ 列的数据进行i次采样取A的1到m行 $1+2*(i-1)$ 到 $3+2*(i-1)$ 列的采样方法进行采样变为 $m*x$ 的样本信号B1,B1中包含m个 $1*x$ 的数组,设置空数组Y2;

(2) 对B1进行聚类,聚类数为 $x1$,这里假设 $x1$ 为3,聚类后得到大小为 $x1*x1$ 的信号,并命名为C1;

(3) 设置a11到a31为C1的第1到 $x1$ 行数组,对这三个数组进行求距离:先设定数组a11的距离值 $q11$ 为 $2*i$ (i与步骤一中的i相同),求a11 a21之间的欧式距离 $s11$;a11 a31之间的欧式距离 $s21$;

然后求得a21的绝对距离 $q21=2*i+s11$;a31的绝对距离 $q31=2*i+s21$;

将 $q11$ 的值放入C1第1行第4列; $q21$ 的值放入C1第2行第4列; $q31$ 的值放入C1第3行第4列;

(4) 将数组B1对a11 a21 a31求相似性:

求B1的第j行的数组 ($j=1,2,\dots,m$) 与a11 a21 a31的欧式距离 $d11$ $d21$ $d31$,B1的第j行的数组与a11的相似性 $m11=1/(d11-1)$,相应的与a21 a31的相似性为 $m21$ $m31$;

比较 $m11$ $m21$ $m31$ 三个值,取最大值 m_{k1} ($k=1,2,3$),将C1的第k行第4列的值赋给Y2的第j行第i列;

(5) 输出Y2;

四.输入分类器进行识别

label为标签矩阵;

(1)分出测试集训练集

设置变量end_sample;train;train_label;test;test_label;

end_sample为Y1 Y2 label三个数组按行拼接;取end_sample数组的1到m-m/5行1到2*n列,取出的这部分命名为train;

train_label是取end_sample的1到m-m/5行2*n+1列;test是取end_sample的m-m/5+1到m行1到2*n列;test_label是取end_sample的m-m/5+1到m行2*n+1列;

(2)输入分类器

将train train_label test 输入MATLAB自带的KNN分类器中(用函数knnclassify);

得到预测的标签并保存在变量predict_label中;

(3)计算准确率

将predict_label与test_label进行比较,在MATLAB中用find函数得到一个数组pre,用length函数计算pre的行数,并用这个数值除以数组test_label的行数得到准确率。

一种基于机器学习技术的心律失常分析系统

技术领域

[0001] 本发明是医学信号处理领域,具体涉及一种基于机器学习技术的心律失常分析系统。

背景技术

[0002] 心血管疾病是威胁人类健康的疾病之首,全球五分之二的人口死于心血管疾病。心电图(electrocardiogram, ECG)是心脏电活动过程中电位变化在体表的综合反映, ECG的各个部分代表着特定的生理意义。

[0003] 由于心血管疾病种类多样、差异巨大。如何根据心电图对病人心血管状态准确的评估和诊断,这要求临床医生具有专业的知识、丰富的经验。但医生在长期大量心电图诊断时、极易疲劳、极易出现诊断错误等问题。针对这些问题,心电图的辅助诊断设备发展迅速,随着信息领域的进步,特别是随着模式识别技术的进展,心电图设备的功能不再是仅仅获取心电信号、打印心电图,而是向着挖掘心电图中的有效数据以及自动识别、统计心拍信息方向发展。带自动识别心拍功能的分析设备能够为医生提供更直观有效的心电图信息,有效节省诊断时间,提高医生的诊断效率,是重要的辅助医疗设备之一。

[0004] 这些设备的核心是自动心律失常分析系统。本发明实现此系统的技术途径是通过提取心电图特征向量并输入分类器,得到分类结果。

[0005] 此系统的技术难点是特征提取,合理的特征提取将直接影响系统的准确率与稳定性。为实现此系统,本发明提出一种新的特征提取方法,这种新的特征提取方法可以达到与现有方法相同的准确率,为此系统的实现提供了一种新的解决方案。

发明内容

[0006] 本发明发明的目的是提供一种基于机器学习技术的心律失常分析系统。

[0007] 一种基于机器学习技术的心律失常分析系统,它包括:

一. 读入数据

外部获取一组两导联心电信号输入到MATLAB中

设置一个空数组A, m行2*n列。根据信号标注的R波点的位置按照其采样频率每个导联依次截取相应个点,两个导联截取出的数据依次存入数组A。数组A中前m行n列的数据为第一导联的数据, m行n+1到2*n列为第二导联的数据。

[0008] 二. 对单通道心电图样本的第一导联进行特征提取的步骤:

(1) 设置i, 将A中前m行n列的数据进行i次采样取A的1到m行 $1+2*(i-1)$ 到 $3+2*(i-1)$ 列的采样方法进行采样变为m*x的样本信号B, B中包含m个1*x的数组, 设置空数组Y1;

(2) 对B进行聚类, 聚类数为x, 这里假设x为3, 聚类后得到大小为x*x的信号, 并命名为C;

(3) 设置a1到a3为C的第1到x行数组, 对这三个数组进行求距离: 先设定数组a1的距离值q1为2*i (i与步骤一中的i相同), 求a1 a2之间的欧式距离s1; a1 a3之间的欧式距离s2。

然后求得 a_2 的绝对距离 $q_2=2*i+s_1$; a_3 的绝对距离 $q_3=2*i+s_2$ 。将 q_1 的值放入C第1行第4列; q_2 的值放入C第2行第4列; q_3 的值放入C第3行第4列;

(4) 将数组B对 a_1 a_2 a_3 求相似性:

求B的第j行的数组($j=1, 2, \dots, m$)与 a_1 a_2 a_3 的欧式距离 d_1 d_2 d_3 , B的第j行的数组与 a_1 的相似性 $m_1=1/(d_1-1)$, 相应的与 a_2 a_3 的相似性为 m_2 m_3 。比较 m_1 m_2 m_3 三个值, 取最大值 m_k ($k=1, 2, 3$), 将C的第k行第4列的值赋给Y1的第j行第i列;

(5) 输出Y1;

三. 对单通道心电图样本的第二导联进行特征提取的步骤:

(1) 设置i, 将A中m行n+1到 $2*n$ 列的数据进行i次采样取A的1到m行 $1+2*(i-1)$ 到 $3+2*(i-1)$ 列的采样方法进行采样变为 $m*x$ 的样本信号B1, B1中包含, m个 $1*x$ 的数组, 设置空数组Y2;

(2) 对B1进行聚类, 聚类数为 x_1 , 这里假设 x_1 为3, 聚类后得到大小为 x_1*x_1 的信号, 并命名为C1;

(3) 设置 a_{11} 到 a_{31} 为C1的第1到 x_1 行数组, 对这三个数组进行求距离: 先设定数组 a_{11} 的距离值 q_{11} 为 $2*i$ (i 与步骤一中的 i 相同), 求 a_{11} a_{21} 之间的欧式距离 s_{11} ; a_{11} a_{31} 之间的欧式距离 s_{21} 。然后求得 a_{21} 的绝对距离 $q_{21}=2*i+s_{11}$; a_{31} 的绝对距离 $q_{31}=2*i+s_{21}$ 。将 q_{11} 的值放入C1第1行第4列; q_{21} 的值放入C1第2行第4列; q_{31} 的值放入C1第3行第4列;

(4) 将数组B1对 a_{11} a_{21} a_{31} 求相似性:

求B的第j行的数组($j=1, 2, \dots, m$)与 a_1 a_2 a_3 的欧式距离 d_{11} d_{21} d_{31} , B1的第j行的数组与 a_{11} 的相似性 $m_{11}=1/(d_{11}-1)$, 相应的与 a_{21} a_{31} 的相似性为 m_{21} m_{31} 。比较 m_{11} m_{21} m_{31} 三个值, 取最大值 m_{k1} ($k=1, 2, 3$), 将C的第k行第4列的值赋给Y2的第j行第i列;

(5) 输出Y2;

四. 输入分类器进行识别

label为标签矩阵;

(1) 分出测试集训练集

设置变量end_sample; train; train_label; test; test_label。end_sample为Y1 Y2 label三个数组按行拼接; 取end_sample数组的1到 $m-m/5$ 行1到 $2*n$ 列, 取出的这部分命名为train。train_label是取end_sample的1到 $m-m/5$ 行 $2*n+1$ 列; test是取end_sample的 $m-m/5+1$ 到m行1到 $2*n$ 列; test_label是取end_sample的 $m-m/5+1$ 到m行 $2*n+1$ 列;

(2) 输入分类器

将train train_label test 输入MATLAB自带的KNN分类器中(用函数knnclassify)。得到预测的标签并保存在变量predict_label中;

(3) 计算准确率

将predict_label与test_label进行比较, 在MATLAB中用find函数得到一个数组pre, 用length函数计算pre的行数, 并用这个数值除以数组test_label的行数得到准确率。

[0009] 相比于目前已公开的技术方案, 本发明的主要区别技术特征是通过采样生成单通道心电图样本, 并利用相似度特征进行特征提取, 最后通过KNN分类器对提取的特征进行识别。这种方法表现出了与现有特征提取方法相同的准确率, 为实现此系统提供了新的技术方案。

附图说明

[0010] 图1特征提取步骤图。

[0011] 实施例1

下面通过具体的实施方式对本发明作进一步说明。

[0012] 具体实例为国际通行心电图数据库MIT-BIH Arrhythmia Database (mitdb), 该数据库的数据及使用说明公开于行业内周知的physionet.org网站。数据库包含47位患者两导联方式的半小时360Hz心电图记录, 并已通过心脏医生手动标注。从所述数据集中选出四种按照AAMI标准划分的心拍类别组合作为效果评估依据, 包括N类(正常心拍或束支传导阻滞心拍)、S类(室上性异常心拍)、V类(心室异常心拍)、F类(融合心拍)。这四种类别的标签以及与mitdb数据集中类别的对应关系。在本实例中, 通过工作在计算机上的软件系统和行业内所周知的Matlab仿真环境进行实现。

[0013] 一. 读入数据

设置一个空数组sample, 该数组变量是30136*1301维的数组。

[0014] (1.1) 输入mitdb数据集中中心电信号, 此心电信号有两个导联并且是用360Hz的速率采样得到的而且已标注R波顶点位置和R波对应的标签。对于第n个R波的第一个导联从R波顶点向前取300个点向后取349个点, 将截取的这650个点, 放入sample的第n行第1到650列, 对于第n个R波的第二个导联从R波顶点向前取300个点向后取349个点, 将截取的这650个点, 放入sample的第n行第651到1300列, 将第n个R波的标签放入sample第n行第1301列。令n循环赋值为1, 2, 3...30136, 每次循环执行步骤(1.1)。

[0015] (1.2) 设置数组变量sample1; sample2, 使sample1为sample的1到30136行1到650列; sample2为sample的1到30136行651到1300列。

[0016] 二. 对单通道心电图样本的第一导联进行特征提取的步骤:

将sample1赋值给X1, 设置大小为30136*323的数组Y1。

[0017] 循环执行步骤(2.1) (2.2) (2.3) (2.4) 执行323次。i由1增加到323, 每次增加1。

[0018] (2.1) 在MATLAB中对30136*650维的单通道心电图样本X进行采样, 对于第i次采样($i=1, 2, \dots, 323$), 取X的1到30136行 $1+2*(i-1)$ 到 $3+2*(i-1)$ 列的采样方法进行采样, 并得到30136*3的样本信号, 并命名为signal。则signal中包含30136个1*3的数组。

[0019] (2.2) 对signal (亦即signal中包含30136个1*3的向量) 用MATLAB中自带的K-means函数进行聚类, 聚类数为3, 聚类后得到大小为3*3的信号, 并命名为feature。feature在MATLAB中是一个数组变量。

[0020] (2.3) 设置a1为feature的第1行数组, a2为feature的第2行数组, a3为feature的第3行数组, 对这三个数组进行求距离: 先设定数组a1的距离值q1为 $2*i$ (i与步骤一中的i相同), 然后用MATLAB自带的函数pdist2并设置为欧氏距离, 求得a1 a2之间的相对距离s1; a1 a3之间的相对距离s2。然后求得a2的绝对距离 $q2=2*i+s1$; a3的绝对距离 $q3=2*i+s2$ 。将q1的值放入feature第1行第4列; q2的值放入feature第2行第4列; q3的值放入feature第3行第4列;

(2.4) 取signal的第j行的数组与a1 a2 a3求相似性, 求相似的方法如下(j由1开始增加到30136, 每次循环j加1.):

先分别求signal的第j行的数组与a1 a2 a3的欧式距离(用函数pdist2)为d1 d2 d3,

signal的第j行的数组与a1的相似性 $m1=1/(d1-1)$,相应的与a2 a3的相似性为 $m2\ m3$ 。比较 $m1\ m2\ m3$ 三个值,取最大值 $m_k(k=1,2,3)$,将feature的第k行第4列的值赋给Y的第j行第i列。

[0021] (2.5)输出Y1

三.对单通道心电图样本的第二导联进行特征提取的步骤:

将sample2赋值给X2,设置大小为30136*323的数组Y2。

[0022] 循环执行步骤(3.1)(3.2)(3.3)(3.4)执行323次。i由1增加到323,每次增加1。

[0023] (3.1)在MATLAB中对30136*650维的单通道心电图样本X1进行采样,对于第i1次采样($i1=1,2,\dots,323$),取X1的1到30136行 $1+2*(i1-1)$ 到 $3+2*(i1-1)$ 列的采样方法进行采样,并得到30136*3的样本信号,并命名为signal。则signal中包含30136个1*3的数组。

[0024] (3.2)对signal1(亦即signal1中包含30136个1*3的向量)用MATLAB中自带的K-means函数进行聚类,聚类数为3,聚类后得到大小为3*3的信号,并命名为feature1。feature1在MATLAB中是一个数组变量。

[0025] (3.3)设置a11为feature1的第1行数组,a12为feature1的第2行数组,a13为feature1的第3行数组,对这三个数组进行求距离:先设定数组a11的距离值 $q11$ 为 $2*i1$ (i与步骤一中的i相同),然后用MATLAB自带的函数pdist2并设置为欧氏距离,求得a11 a12之间的相对距离 $s11$;a11 a13之间的相对距离 $s12$ 。然后求得a12的绝对距离 $q12=2*i1+s11$;a13的绝对距离 $q13=2*i1+s12$ 。将 $q11$ 的值放入feature1第1行第4列; $q12$ 的值放入feature1第2行第4列; $q13$ 的值放入feature1第3行第4列;

(3.4)取signal1的第j1行的数组与a11 a12 a13求相似性,求相似的方法如下(j1由1开始增加到30136,每次循环j1加1。):

先分别求signal1的第j1行的数组与a11 a12 a13的欧式距离(用函数pdist2)为 $d11\ d12\ d13$,signal1的第j1行的数组与a11的相似性 $m11=1/(d11-1)$,相应的与a12 a13的相似性为 $m12\ m13$ 。比较 $m11\ m12\ m13$ 三个值,取最大值 $m1s(s=1,2,3)$,将feature1的第s行第4列的值赋给Y2的第j1行第i1列。

[0026] (3.5)输出Y2

四.输入分类器进行识别

label为sample的1到30136行1301列。

[0027] (4.1)分出测试集训练集

设置变量end_sample;train;train_label;test;test_label。end_sample为Y1 Y2 label三个数组按行拼接;取end_sample数组的1到28000行1到1300列,取出的这部分命名为train。train_label是取end_sample的1到28000行1301列;test是取end_sample的28001到30136行1到1300列;test_label是取end_sample的28001到30136行1301列。

[0028] (4.2)输入分类器

将train train_label test 输入MATLAB自带的KNN分类器中(用函数knnclassify)。得到预测的标签并保存在变量predict_label中。

[0029] (4.3)计算准确率

将predict_label与test_label进行比较,在MATLAB中用find函数得到一个数组pre,用length函数计算pre的行数,并用这个数值除以数组test_label的行数得到准确率。

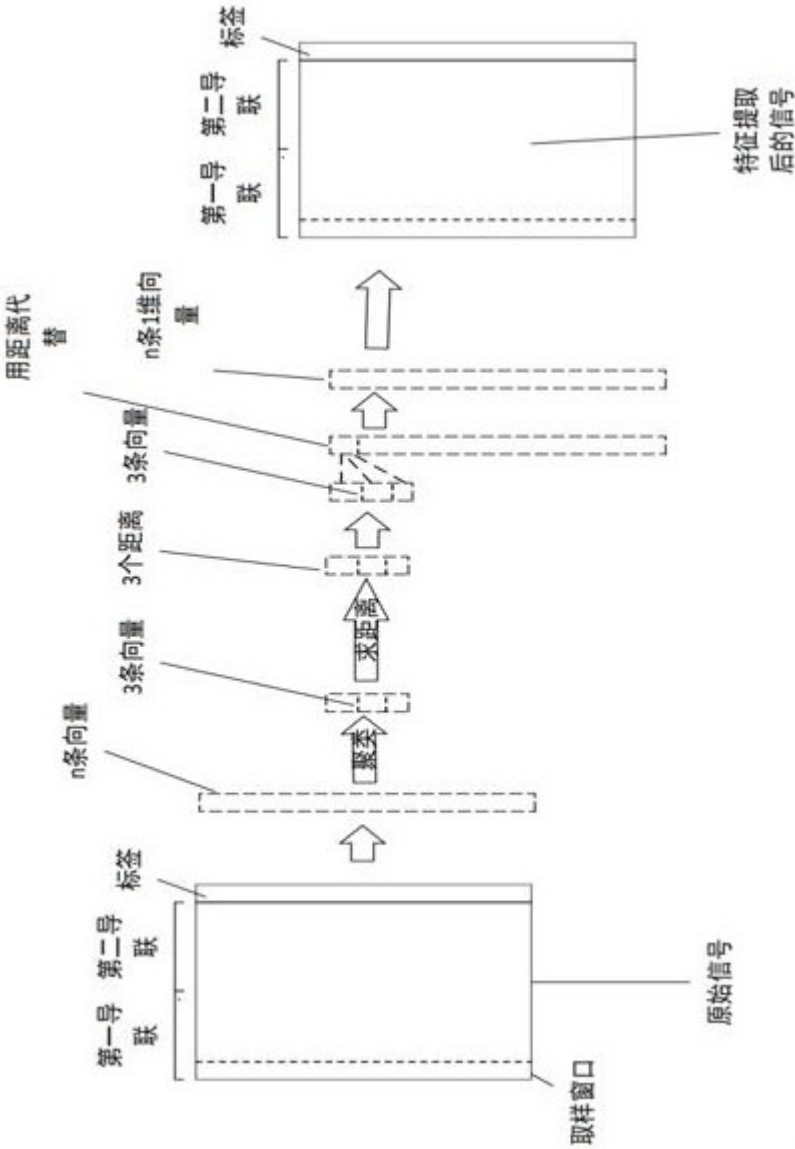


图1

专利名称(译)	一种基于机器学习技术的心律失常分析系统		
公开(公告)号	CN110192851A	公开(公告)日	2019-09-03
申请号	CN201910538631.1	申请日	2019-06-20
[标]申请(专利权)人(译)	鲁东大学		
申请(专利权)人(译)	鲁东大学		
当前申请(专利权)人(译)	鲁东大学		
[标]发明人	臧睦君 邹海林 柳婵娟		
发明人	刘通 李春涛 臧睦君 邹海林 柳婵娟 周树森 赵玲玲		
IPC分类号	A61B5/0245 A61B5/00		
CPC分类号	A61B5/02405 A61B5/0245 A61B5/72 A61B5/7267		
代理人(译)	张铁生 刘延军		
外部链接	Espacenet SIPO		

摘要(译)

本发明提供了一种基于机器学习技术的心律失常分析系统，它包括：
一．读入数据；二.对单通道心电图样本的第一导联进行特征提取，三.对单通道心电图样本的第二导联进行特征提取，四．输入分类器进行识别。相比于目前已公开的技术方案，本发明的主要区别技术特征是通过采样生成单通道心电图样本，并利用相似度特征进行特征提取，最后通过KNN分类器对提取的特征进行识别。这种方法表现出了与现有特征提取方法相同的准确率，为实现此系统提供了新的技术解决方案。

