



(12)发明专利申请

(10)申请公布号 CN 109480833 A

(43)申请公布日 2019. 03. 19

(21)申请号 201811569263.9

(22)申请日 2018.12.21

(66)本国优先权数据

201811001414.0 2018.08.30 CN

(71)申请人 北京航空航天大学

地址 100191 北京市海淀区学院路37号

(72)发明人 龚光红 王夏爽 李妮

(74)专利代理机构 北京永创新实专利事务所

11121

代理人 姜荣丽

(51)Int.Cl.

A61B 5/0476(2006.01)

A61B 5/00(2006.01)

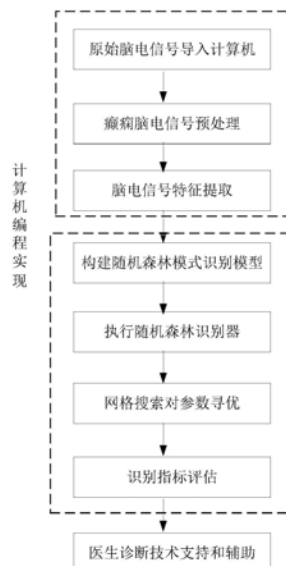
权利要求书2页 说明书8页 附图4页

(54)发明名称

基于人工智能的癫痫患者脑电信号的预处理和识别方法

(57)摘要

本发明公开了一种基于人工智能的癫痫患者脑电信号的预处理和识别方法,涉及脑科学、癫痫发作临床数据识别技术领域。所述方法采集癫痫患者的脑电信号,并对所述的脑电信号进行预处理;对预处理后的信号进行时频分析提取时频特征,同时,提取无噪声脑电信号的统计学特征;然后构建随机森林识别模型,进行不同程度病况的癫痫患者脑电信号的模式识别测试。本发明提供的特征提取方法,可以简化计算过程,提高脑电信号识别的准确性和实效性。由于随机森林识别方法的高度并行化,提高了运行效率,实现了计算机自动识别癫痫患者脑电信号的功能,为医疗工作者耗时耗力的诊断提供技术支持。



1. 一种基于人工智能的癫痫患者脑电信号的预处理和识别方法,其特征在于:包括如下步骤,

步骤一:采集癫痫患者的脑电信号,并对所述的脑电信号进行预处理;

所述的预处理包括滤波处理、剔除坏道及去除无用电极处理、叠加平均处理、基线校正及重参考处理、独立成分分析去除眼电、心电和肌电处理,最终得到无噪声脑电信号;

步骤二、对所述的无噪声脑电信号进行时频分析提取时频特征,同时,提取无噪声脑电信号的统计学特征;

具体步骤如下:

步骤201、运用短时傅里叶变换的方法进行时频分析,提取无噪声脑电信号的时频特征,得到 γ (25~50Hz), β (12~25Hz), α (6~12Hz), θ (3~6Hz), δ (0~3Hz) 中每段数据的时频特征;

步骤202、提取无噪声脑电信号的能量均值及能量标准方差作为统计学特征;

步骤203、对所述的时频特征和统计学特征进行主成分分析降低特征维度的处理,从而降低计算机运行过程中的开销,为输入自动识别器做准备;

步骤三、构建随机森林识别模型,利用步骤二提取出来的脑电特征对随机森林识别模型进行训练,生成最优的随机森林识别模型;

步骤四、将优化后的随机森林模型在测试集上进行不同程度病况的癫痫患者脑电信号的模式识别测试。

2. 根据权利要求1所述的一种基于人工智能的癫痫患者脑电信号的预处理和识别方法,其特征在于:所述的滤波处理采用巴特沃斯高通滤波器和低通滤波器。

3. 根据权利要求1所述的一种基于人工智能的癫痫患者脑电信号的预处理和识别方法,其特征在于:步骤三具体步骤如下:

步骤301、对脑电特征进行归一化处理;

步骤302、将归一化后的脑电特征分成三组,分别为训练集、测试集以及用于防止随机森林识别模型过拟合的验证集;

步骤303、对不同程度病况的癫痫患者的脑电特征进行时间同步的分帧,得到多导联的脑电特征信号;

步骤304、对癫痫患者发作期、间歇期和健康期的脑电信号的进行标记:发作期“1”、间歇期“0”、健康期“-1”;

步骤305、从训练集中随机选取一个训练样本 M_i ,并且是有放回地进行 k 次抽取,从而得到一个随机生成的训练集 $M^* = \{M_1, M_2, \dots, M_k\}$, $i = 1, 2, \dots, k$;

步骤306、以上述305步骤中随机生成的训练集 M^* 为训练数据,建立 k 个决策树 T_j , $j = 1, 2, \dots, k$,从决策树的每一个节点的特征集合中随机选择一个包含 k 特征的子集;

步骤307、通过对决策树的每个终端节点递归地重复步骤306上述步骤,将随机森林决策树生长为增强绑定数据,直到该决策树能够准确地对训练集 M^* 进行识别,同时达到最小节点大小;

步骤308、集合所有决策树,对于一个输入样本 M_i , k 个决策树有 k 个识别结果,随机森林继承了所有的识别投票结果;

步骤309、在新的节点上进行预测,投票数目最多的识别是输出,进入步骤310;

步骤310、采用网格搜索的优化方法对随机森林模型进行优化；

步骤311、得出对随机森林进行网格搜索优化后的参数，同过计算机采用10重交叉验证去识别该随机森林识别模型的准确性；

步骤312、改变节点数，再随机打乱数据，选择一个全新的验证集进行随机森林识别模型计算机识别；

步骤313、重复步骤312，直至交叉验证准确性最高；

步骤314、此时的节点数认为是最优节点数；训练出的参数为随机森林识别模型的最优参数，模型确定。

4. 根据权利要求3所述的一种癫痫患者脑电信号的计算机模式识别方法，其特征在于：所述的网格搜索包括粗搜索和精确搜索两部分；

(A) 所述的粗搜索超参数具体为：

首先，用大步长划分网格，设置搜索步长为10，进行粗搜索选择出最优参数；在一次最优参数附近利用小步长划分网格，使网格划分更加密集，再次进行搜索选择出二次最优参数；在二次最优参数附近再次进行网格划分和搜索，……，直至网格间距或目标函数变化量小于给定值；如果有多组参数使得随机森林的方法达到最优，则从这些组参数中选出使得惩罚函数最小的那组参数为最优参数；

(B) 精确搜索具体为：设置步长为0.1，根据随机森林模型的实际情况，调整步长大小，使得惩罚参数最小化是最佳参数组；将最佳参数组中的超参数作为最终的超参数。

5. 根据权利要求1或2所述的一种癫痫患者脑电信号的计算机模式识别方法，其特征在于：所述的超参数包括决策树数量 k 、分裂特征数 m 、叶子节点最少样本数以及惩罚参数。

基于人工智能的癫痫患者脑电信号的预处理和识别方法

技术领域

[0001] 本发明涉及脑科学、癫痫发作临床数据识别技术领域，具体是一种基于人工智能的癫痫患者脑电信号的预处理和识别方法。

背景技术

[0002] 癫痫病的发作是脑内神经元阵发性异常超同步化电活动的临床表现，具有反复性、突发性和暂时性等特点。脑电信号作为研究癫痫疾病的重要工具，它所实时反映的发作信息是其它生理学方法所不能提供的。目前，在癫痫患者的脑电信号的分析研究中，机器学习为癫痫脑电信号识别的强有力工具，但是多数机器学习方法识别脑电信号具有较复杂的计算过程，无法保证识别方法的准确性和实效性。基于计算机的模式识别方法有很多，例如K近邻方法、支持向量机方法、神经网络方法等。这些方法在不同脑疾病的数据集上识别的准确性不同，对于计算机的脑电模式识别方法本身仍然存在着不适用性。此外对于识别方法的参数选择依靠于具有丰富经验研究人员进行人工调参，不能完全保证生成的参数为识别方法的模型最优参数。

发明内容

[0003] 本发明为了克服现有方法的不足，针对癫痫病患者的脑电信号识别问题，提出一种基于人工智能的癫痫患者脑电信号的预处理和识别方法。

[0004] 本发明提供的一种基于人工智能的癫痫患者脑电信号的预处理和识别方法，包括如下几个步骤：

[0005] 步骤一：采集癫痫患者的脑电信号，并对所述的脑电信号进行预处理。

[0006] 所述的脑电信号采用64导联的脑电信号采集装置，采集的脑电信号包括癫痫病患的发作期、间歇期和健康期的脑电信号。

[0007] 所述的预处理包括滤波处理、剔除坏道及去除无用电极处理、叠加平均处理、基线校正及重参考处理、独立成分分析去除眼电、心电、肌电等多种伪迹处理，最终得到无噪声脑电信号。

[0008] 所述的滤波处理采用巴特沃斯高通滤波器和低通滤波器；

[0009] 对所述的无噪声脑电信号进行存储，存储为txt格式的数据。

[0010] 步骤二、对所述的无噪声脑电信号进行时频分析提取时频特征，同时，提取无噪声脑电信号的统计学特征；

[0011] 具体步骤如下：

[0012] 步骤201、运用短时傅里叶变换的方法进行时频分析，提取无噪声脑电信号的时频特征，得到 γ (25~50Hz) , β (12~25Hz) , α (6~12Hz) , θ (3~6Hz) , δ (0~3Hz) 中每段数据的时频特征；

[0013] 步骤202、提取无噪声脑电信号的能量均值及能量标准方差作为统计学特征；

[0014] 步骤203、对所述的时频特征和统计学特征进行降低维度的处理，从而降低方法的

开销,为输入自动识别器做准备。优选的,采用主成分分析法进行降维处理。

[0015] 步骤三、构建随机森林识别模型,利用步骤二提取出来的脑电特征对随机森林识别模型进行训练,生成最优的随机森林识别模型;

[0016] 步骤四、将优化后的随机森林模型在测试集上进行不同程度病况的癫痫患者脑电信号的模式识别测试。

[0017] 本发明的优点在于:

[0018] (1) 本发明提供的特征提取方法,可以简化计算过程,提高脑电信号识别的准确性和实效性。

[0019] (2) 本发明一种癫痫患者脑电信号的计算机模式识别方法,设计了一种ICU病房中不同程度病况的癫痫患者脑电信号识别方法。由于随机森林识别方法的高度并行化,提高了运行效率,实现了计算机自动识别癫痫患者脑电信号的功能,为医疗工作者耗时耗力的诊断提供技术支持。

[0020] (3) 本发明一种癫痫患者脑电信号的计算机模式识别方法,随机森林识别方法在训练过程中会产生大量的超参数,仅依靠程序员长期调节参数的经验进行人工调试这些生成的参数是很难计算出识别模型的最佳参数的。本发明引入网格搜索优化方法可以通过所述计算机以可变步长的形式重复过滤参数来加速搜索参数的最佳组合。加快随机森林模型的运行效率,使得训练出来的随机森林识别模型达到最优效果。本发明提高了基于优化随机森林的计算机模式识别方法来处理ICU病房中癫痫患者脑电信号的准确率,对于与三种不同病况的癫痫病情的模式识别准确率能达到96%以上。

[0021] (4) 本发明一种癫痫患者脑电信号的计算机模式识别方法,在使用网格搜索优化方法对随机森林进行优化的过程中,同时利用计算机对实验脑电信号数据进行10重交叉验证,避免了惩罚函数过高及过学习状态的发生,实现了网格搜索优化方法对随机森林识别模型优化的同时,提升了该模式识别方法的性能,为实现后续有新的脑电信号输入到该模型中也能精确的进行计算机识别,对医生在医疗诊断过程中进行辅助支持与决策。

附图说明

[0022] 图1是本发明的癫痫患者脑电信号的计算机模式识别方法流程图;

[0023] 图2是本发明基于优化随机森林模型的构建过程示意图;

[0024] 图3是本发明基于改进的网格搜索方法的随机森林模型的超参数寻优流程图;

[0025] 图4是本发明基于网格搜索优化方法的随机森林模型与原始随机森林算法的准确性指标对比曲线;

[0026] 图5是本发明利用计算机进行模型训练过程中10重交叉验证产生的准确性结果;

[0027] 图6是本发明基于改进的网格搜索优化方法的随机森林的ROC曲线和AUC值指标的评价结果。

具体实施方式

[0028] 下面结合附图和实施例对本发明进行详细说明。

[0029] 本发明提供一种癫痫患者脑电信号的计算机模式识别方法,所述方法首先通过对ICU病房中不同程度病况的癫痫病患者(癫痫发作期、癫痫间歇期、健康期)进行长时间段、

多通道脑电信号的采集取样,并对不同程度病况癫痫患者的脑电信号进行标记(打标签);对脑电信号进行预处理操作和脑电特征提取等操作。构建基于机器学习技术的随机森林识别模型,通过网格搜索优化方法对随机森林识别模型生成的参数进行优化处理,同时,将预处理后的脑电信号导入到构建并优化好的随机森林识别模型中,进行识别处理。本发明提供的优化的随机森林识别模型基于机器学习技术,利用层层决策树相关运算得到脑电信号不同病况模式的深度抽象,实现对不同程度病况癫痫患者脑电信号的模式进行自动识别,为ICU病房中医生对脑电的后续诊断提供技术支持。

[0030] 本发明提供一种癫痫患者脑电信号的计算机模式识别方法,该方法用于临床,通过识别癫痫患者发作期、间歇期和健康期的脑电信号,建立优化的随机森林识别模型,将所述的优化的随机森林识别模型应用于ICU病房中实时记录到的癫痫原始脑电信号中,将识别中生成的参数指标辅助及支持医生及其他医疗工作者对不同病况癫痫患者的识别工作,避免在ICU病房中医疗工作者的长时间留守病房的低效率工作。

[0031] 本发明提供的一种癫痫患者脑电信号的计算机模式识别方法,如图1所示流程,具体包括如下步骤:

[0032] 步骤一、采集癫痫患者的脑电信号,并对采集的脑电信号进行预处理,得到无噪声脑电信号;

[0033] 采用64导联的脑电信号的采集装置,采用美国临床神经生理学会提供的10-20头皮电极位点分布图,所述的脑电信号采集对象为癫痫病患,采集时间贯穿癫痫病患发作期、间歇期和健康期,对于癫痫病患发作期的脑电信号通常采自ICU病房中的癫痫病患。

[0034] 为了达到脑电信号大数据,可以对各大医院的神经科的发作期的癫痫脑电信号进行收集,同时采集大量健康人的脑电信号进行对比分析。一般情况下,采集的脑电信号主要包括三种类型的脑电信号:癫痫病患发作期、间歇期和健康期。

[0035] 采集装置主要包括两台工作站、液晶显示器、放大器、Ag-Cl电极脑电帽、导电膏、插座、与插座对应的插头及接地线。插头包括有若干个对应插接在插接口的插脚;进行脑电信号采集时,放大器的接线要与电脑连接好,同时脑电帽的接地线连接在任意一个插脚上,以将电阻所在的电路串联形成闭合回路。这样,传感器仪表的脑电信号经过电阻后则可以转换成电压值。其中两台工作站的系统配置信息主要包括双核CPU、内存4GB、显卡512M和WinXP professional系统。放大器的配置信息和参数包括24Bit的A/D转换分辨率,输入阻抗大于10Gohms,放大器具有AC和DC两种采集方式,AC模式的灵敏度为3nV/bit,DC模式的灵敏度为24nV/bit。

[0036] 对于采集的癫痫病患的脑电信号用eeglab这个matlab的工具箱进行预处理,预处理的方法包括:滤波、剔除坏道、叠加平均、基线校正、重参考,独立成分分析去除多种类型的伪迹;最后将得到的预处理后的脑电信号数据保存。

[0037] 所述的滤波处理采用巴特沃斯高通滤波器和低通滤波器对连续的脑电信号进行滤波处理;巴特沃斯高通滤波器和低通滤波器的阈值分别为0.1-55Hz。

[0038] 对预处理后的无噪声脑电信号进行存储,存储为txt格式的数据。本发明采用波恩大学公开的癫痫病患的脑电信号作为样本数据集进行以下步骤的详细说明。

[0039] 第二步、对于无噪声脑电信号,运用短时傅里叶变换的方法对无噪声脑电信号进行时频特征提取,同时,提取无噪声脑电信号的统计学特征。

[0040] 采用短时傅里叶变换方法,分别获得有效频率为4~7Hz的 θ 波、8~13Hz的 α 波、14~17Hz的中间快波、18~30Hz的 β 波的时频特征;根据脑电信号的平均能量、能量方差得到脑电信号的统计学特征。

[0041] 一些研究已经解决了癫痫的自动监测和检测的非平稳脑电活动问题。最经典的是将韦尔奇谱分析方法引入癫痫发作特征分析中。在癫痫发作过程中,通过采用时域方法对癫痫患者的脑电信号进行分析。此外,流行的基于傅立叶的频谱分析技术通常用于分析脑电信号在频域中。傅立叶变换不能很好地解决癫痫患者的脑电分析问题。其次,这种方法是非常耗时的。本发明运用短时傅里叶变换的方法对脑电信号进行时频分析,通过调整不同的时间窗,避免了傅立叶变换的缺点。本发明提出的短时傅里叶变换方法既可以得到脑电信号的时域特征又能得到信号的频域特征,更大程度的挖掘了脑电信号数据的本质特征。具体地:

[0042] 步骤201、运用短时傅里叶变换的方法进行时频分析,提取脑电信号的频域特征,得到 α 波、 β 波、 δ 波、 θ 波和 γ 波每段的数据特征;并根据脑电信号的波段对频谱进行划分,生成脑电信号的时频图。

[0043] 步骤202、根据脑电信号的时频图,提取脑电信号的能量均值及能量标准方差作为统计学特征;

[0044] 基于计算机程序对脑电信号提取其能量均值及能量标准方差,其具体统计学特征信息如下:

	S	F/N	O/Z
[0045] 能量均值	-4.74	-5.94	-6.31
采样点数	4097	4097	4097
能量标准方差	38.55	13.10	4.56

[0046] 其中,数据集S、F/N和O/Z分别表示癫痫患者发作期、间歇期和健康期。

[0047] 上述提取的时频特征和统计学特征都是高维特征的脑电信号,因此需要对癫痫病患的脑电信号的时频特征和统计学特征进行降低维度处理。本发明通过主成分分析方法进行降低维度的处理,从而降低方法的开销,构造特征集为输入自动识别器做准备。

[0048] 所述降维处理原理如下:因为数据的最大方差给出了数据的最重要的信息。转换坐标系时,以方差最大的方向作为坐标轴方向,第一个新坐标轴选择的是原始数据中方差最大的方法,第二个新坐标轴选择的是与第一个新坐标轴正交且方差次大的方向。重复该过程,重复次数为原始数据的特征维数。

[0049] 步骤三、构建随机森林识别模型,利用计算机对提取出来的脑电特征进行大量训练,生成最优的随机森林识别模型。

[0050] 随机森林识别模型是由决策树组合的一种有效的集成机器学习方法,随机森林识别模型适用于高维数据并且运行速度快。但是运行过程中会产生大量的超参数,为了得到更高的识别癫痫脑电信号的精度,需要对随机森林识别模型的参数进行优化。目前,对随机森林识别模型中的参数进行优化选择的方法相对较少,通常是依靠经验进行人工参数选择。尤其是随机森林识别模型中决策树数量 k 对随机森林识别模型性能的影响比较大,并且对于不同的类别的数据而言,当随机森林识别模型的性能达到最优时的 k 值是不一样的。仅仅通过经验进行选定随机森林识别模型的参数,通常无法得到性能最优的随机森林识别模

型。本发明利用改进的网格搜索方法,对随机森林识别模型中的决策树数量 k 、分裂特征数 m 、叶子节点最少样本数和惩罚参数进行参数优化,同时利用机器学习中10重交叉验证的方法更加有效地避免了训练出的随机森林识别模型的过拟合问题。通过ICU病房中的原始癫痫脑电信号的仿真结果表明,本发明利用网格搜索方法优化得到的参数能够使随机森林识别模型的识别效果得到很大程度的提高。

[0051] 具体的,所述的最优的随机森林识别模型的生成步骤如下:

[0052] 步骤301、对步骤一中提取的脑电特征进行归一化处理;

[0053] 步骤302、将归一化后的脑电特征分成三组,分别为训练集、测试集以及用于防止随机森林识别模型过拟合的验证集;

[0054] 步骤303、对不同程度病况的癫痫患者的脑电特征进行时间同步的分帧,得到多导联的脑电特征信号;

[0055] 步骤304、对癫痫患者发作期、间歇期和健康期的脑电信号的进行标记:发作期“1”、间歇期“0”、健康期“-1”;

[0056] 步骤305、如图2所示流程,从训练集的数据 M 中随机选取一个bootstrap训练样本 M_i ,并且是有放回地进行 k 次抽取,从而得到一个随机生成的训练集 $M^* = \{M_1, M_2, \dots, M_k\}$, $i = 1, 2, \dots, k$;

[0057] 步骤306、以上述305步骤中随机生成的训练集 M^* 为训练数据,建立 k 个决策树 T_j , $j = 1, 2, \dots, k$,从决策树的每一个节点的特征集合中随机选择一个包含 k 个特征的子集,从所述子集中随机选取 m 个最优特征的分化作为识别特征,这里面的 k 控制了随机程度的大小;进入步骤307;

[0058] 步骤307、通过对决策树的每个终端节点递归地重复步骤306上述步骤,将随机森林决策树生长为增强绑定数据,直到该决策树能够准确地对训练集 M^* 进行识别,同时达到最小节点大小。

[0059] 在模型训练的过程中,采用识别回归树CART方法分裂节点,基尼指数GINI值作为分裂节点的依据。训练集 M^* 含有 k 个不同的特征,则此训练集 M^* 的基尼指数为:

$$[0060] \quad GINI(k) = 1 - \sum_{i=1}^k p_i$$

[0061] 其中 p_i 为第 i 个特征的出现概率。 $\{M_1, M_2, \dots, M_k\}$ 对应的特征数分别为 $\{n_1, n_2, \dots, n_k\}$,则分裂的基尼指数为:

$$[0062] \quad GINI(M^*) = \frac{n_1}{n} GINI(M_1) + \frac{n_2}{n} GINI(M_2) + \dots + \frac{n_k}{n} GINI(M_k)$$

[0063] 其中, n 表示 $n_1 + n_2 + \dots + n_k$, $i = 1, 2, \dots, k$ 。 $GINI(M_i)$ 表示样本 M_i 的GINI指数。

[0064] 步骤308、集合所有决策树,对于一个输入样本 M_i , k 个决策树有 k 个识别结果,随机森林继承了所有的识别投票结果;

[0065] 步骤309、在新的节点上进行预测,投票数目最多的识别是输出,进入步骤310;

[0066] 随机森林方法如下:

[0067] 步骤310、采用网格搜索的优化方法对随机森林模型进行优化;

[0068] 网格搜索优化方法是指将变量区域进行网格化,遍历所有的网格点,求解满足约束条件的目标函数值,选出最优数值。遍历网格上所有参数需要大量训练时间,为了提高训练速度,本发明提出了一种基于改进的网格搜索的随机森林参数优化方法。

[0069] 步骤311、首先粗搜索超参数：决策树数量 k 、分裂特征数 m 、叶子节点最少样本数以及惩罚参数；进入步骤312；

[0070] 首先，在较大范围内用大步长划分网格，设置搜索步长为10，进行粗搜索选择出一次最优参数。

[0071] 步骤312、然后在一次最优参数附近利用小步长划分网格，使网格划分更加密集，再次进行搜索选择出二次最优参数；所述小步长（当次小步长指的是上一次步长的一半）是相对于大步长而言的相对量，略小即可，一般选取5~10范围的步长。

[0072] 步骤313、重复步骤312，在二次最优参数附近利用小步长划分网格和搜索参数，直至网格间距或目标函数变化量小于给定值；

[0073] 为了提高计算机程序的模式识别性能，需要同时考虑单个决策树的识别正确率和决策树的多样性，然而两者之间也存在着一定关系。本发明针对随机森林方法中决策树的个数 k 和候选分裂特征数 m 为离散值的特点，采用网格搜索方法进行参数优化。本发明提出的基于改进的网格搜索的随机森林参数优化的目标函数值选用未被抽中的训练数据即袋外数据估计的识别准确性。由于随机森林在构建过程中的随机性，识别准确性可能会在一定范围内波动，因此为减小不确定性对参数选择产生的影响，本发明在求识别准确性时选用多个随机森林模型识别准确性的平均值。

[0074] 步骤314、如果有多组参数使得随机森林的方法达到最优，则从这些组参数中选出使得惩罚函数最小的那组参数为最优参数；

[0075] 步骤315、对随机森林方法得到的参数进行精确搜索，减小搜索范围，即减小网格搜索优化方法的搜索步长，此时步长一般设置为0.1，也可以根据随机森林识别模型的实际情况，调整步长大小，使得惩罚参数最小化是最佳参数组；具体步骤如下：

[0076] 确定决策树数量 k 和分裂特征数 m 的范围，设定步长，在决策树数量 k 和分裂特征数 m 坐标系上建立二维网格，网格节点就是相应的决策树数量 k 和分裂特征数 m 的参数对；

[0077] 对网格节点上的每一组参数构建随机森林识别模型，并利用袋外数据评估识别的准确性；

[0078] 选择识别准确性最高的参数 k 和 m ，若识别准确性或者步长满足要求，则输出最优参数和识别准确率；否则，缩小步长，继续精确搜索。上述基于改进的网格搜索方法的随机森林参数寻优流程图由图3所示。

[0079] 步骤316、得出对随机森林进行网格搜索优化后的参数，同过计算机采用10重交叉验证去识别该随机森林识别模型的准确性；对于给定的训练集进行样本采样，产生出10个不同的子集，再从每个子集中训练出来一个属于该子集的随机森林识别子模型，这样通过不同子集训练出来的随机森林识别子模型具有比较大的差异，从而有效地避免随机森林模型的过拟合问题。同时为获得更优的随机森林识别模型，还希望每个随机森林识别子模型不能太差，如果分化出的子集过于多，只能通过小部分数据进行训练，这样的模型不足以成为好的随机森林识别子模型，为此本发明考虑使用以一种相互交叠的方式对训练集进行采样。

[0080] 步骤317、改变节点数，再随机打乱数据，选择一个全新的验证集进行随机森林识别模型计算机识别；

[0081] 步骤318、重复步骤317，直至交叉验证准确性最高；

[0082] 步骤319、此时的节点数被认为是最优节点数；训练出的参数为随机森林识别模型的最优参数，模型确定；

[0083] 步骤四、将优化后的随机森林识别模型在测试集上进行不同程度病况的癫痫患者脑电信号的模式识别测试；根据多种测试指标来评判随机森林这个计算机自动识别方法的性能；评估随机森林模型的泛化能力。

[0084] 首先，确定随机森林计算机模式识别算法的所有参数及取值；

[0085] 然后，通过网格搜索优化算法对随机森林识别模型生成的所有参数进行粗搜索，限定参数的局部范围；在此局部范围的基础上进行小步长的二次寻优，得到随机森林模型的最优参数；

[0086] 最后，将最优参数输入到随机森林计算机模式识别模型中，生成最优的识别模型，用于对癫痫患者的脑电信号进行计算机识别，从而为医生进行癫痫病况诊断提供技术支持和辅助。

[0087] 实验结果

[0088] 采用的脑电数据来源于德国波恩癫痫实验室。该数据分为O、Z、F、N、S共5组脑电信号，每组数据中包含100个样本，受试者均为5人，每个样本包含4097个采样点，信号记录采用标准10—20系统，采样频率为173.61Hz，采样时间为23.6s。5组脑电数据集的具体信息如下。

[0089]

数据集类别	被试病况	电极位置	对应电极采集区域	受试者状态	样本数 /个
O/Z	健康期	头皮表面	所有脑区	所有区域	200
F/N	间歇期	颅内	致病灶外/内区域	间歇期未发作	200
S	发作期	颅内	致病灶内区域	发作期	100

[0090] 本发明分成O/Z、F/N、S这3组数据集对脑电信号进行识别。其中数据集O/Z为健康人处于清醒睁眼状态下头皮表面的脑电信号，数据集F/N为癫痫患者未发作处于间期致病灶内区域的颅内EEG信号，数据集S为癫痫患者处于发作期致病灶内区域的颅内脑电信号，分别记为健康期、发作间歇期和发作期脑电信号。

[0091] 对实验结果采用准确性ACC曲线指标进行评估。对实验结果采用真阳性率和假阳性率指标进行评估，生成的受试者工作特征曲线ROC和AUC值等评价价值指标根据得到的识别结果优化的随机森林方法的评估。

[0092] 本次实验一共采用四个评价指标，包括癫痫脑电信号模式识别方法的准确性、真阳性概率(True positive rate,TPR)和假阳性概率(False positive rate,FPR)，假阳性概率为横轴，真阳性概率为纵轴所组成的坐标图，从而生成的受试者工作特征曲线(ROC)和ROC区域下面积AUC值两个评估指标。

$$[0093] \quad ACC = \frac{TP + TN}{TN + FP + TP + FN} \cdot 100\%$$

$$[0094] \quad TPR = \frac{TP}{TP + FN} \cdot 100\%$$

$$[0095] \quad FPR = \frac{FP}{TN + FP} \cdot 100\%$$

[0096] 其中TP,FP,TN,FN分别表示真阳个数、假阳个数、真阴个数和假阴个数。

[0097] 改进前后的随机森林识别方法的准确性结果如图4,通过网格搜索优化后的随机森林模型的准确率达到96.7%,比单独使用随机森林方法准确率提升了近10个百分点。对训练集做10重交叉验证的模型精度变化如图5所示,应用改进的随机森林优化方法的另两个评价指标ROC和AUC的结果如图6所示,其中AUC取得了99%的高准确率。

[0098] 本发明将改进的随机森林计算机识别方法应用于癫痫患者的脑电信号的自动识别过程中,实现不同程度病况癫痫患者脑电信号的识别,同时设计了网格搜索的优化方法对随机森林识别方法进行优化,实现了计算机运行计算结果准确率的提升和程序运行过程的加速,辅助广大的医疗工作者在为癫痫病患诊断的过程中提供技术支持和服务,对有潜在发病的癫痫患者消除和控制发病几率。

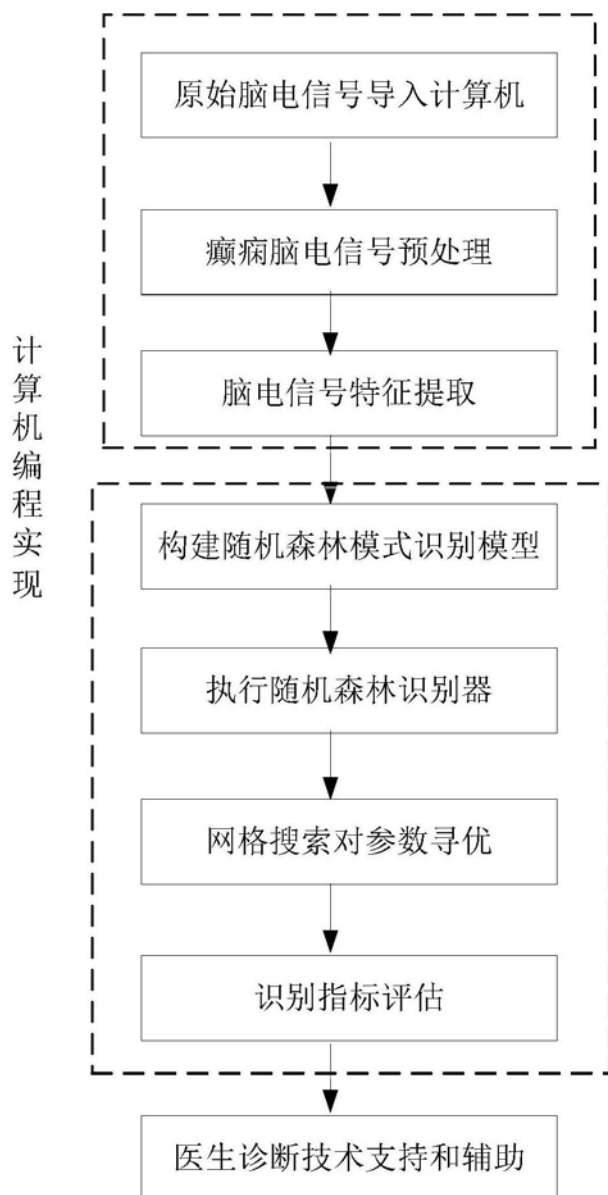


图1

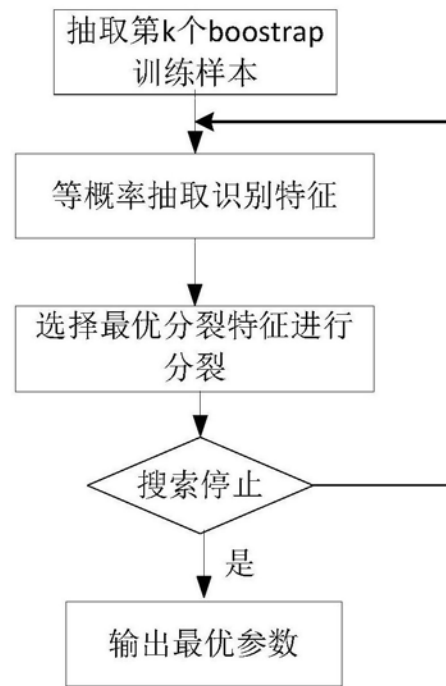


图2

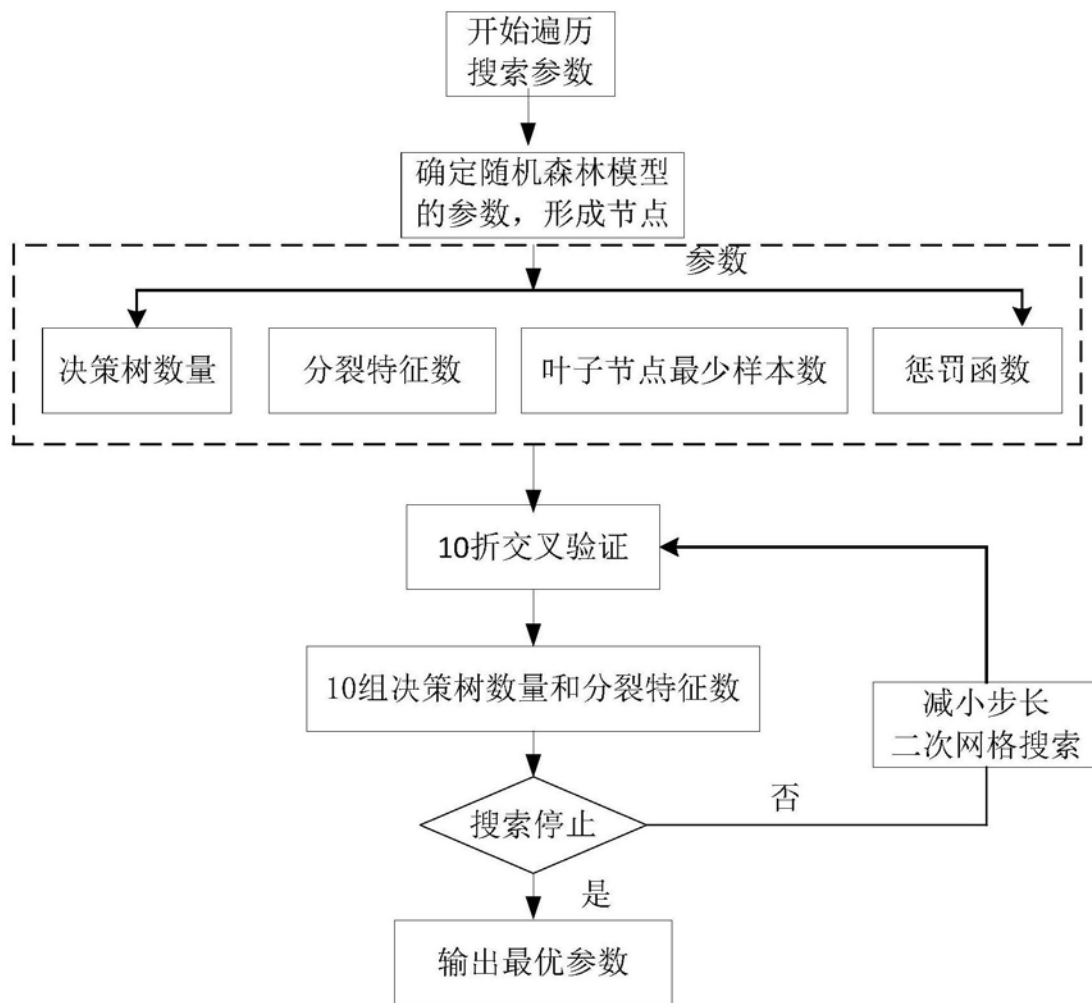


图3



图4

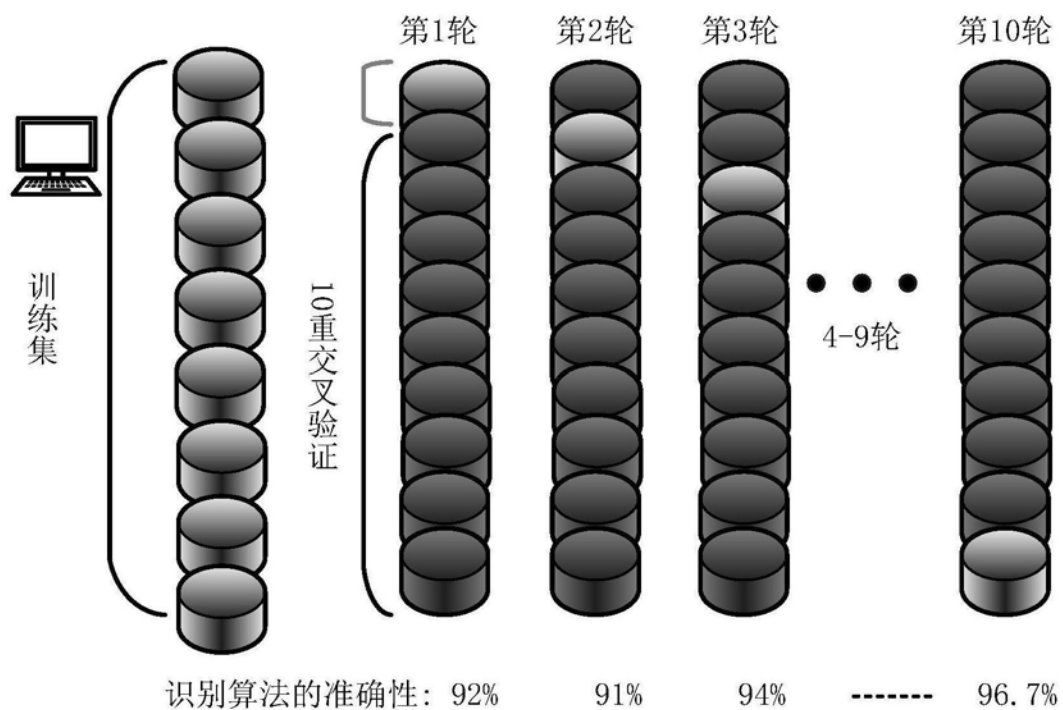


图5

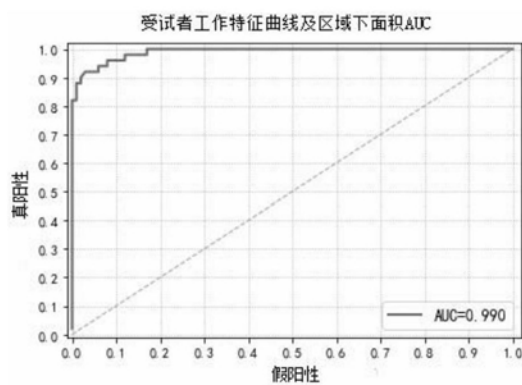


图6

专利名称(译)	基于人工智能的癫痫患者脑电信号的预处理和识别方法		
公开(公告)号	CN109480833A	公开(公告)日	2019-03-19
申请号	CN201811569263.9	申请日	2018-12-21
[标]申请(专利权)人(译)	北京航空航天大学		
申请(专利权)人(译)	北京航空航天大学		
当前申请(专利权)人(译)	北京航空航天大学		
[标]发明人	龚光红 王夏爽 李妮		
发明人	龚光红 王夏爽 李妮		
IPC分类号	A61B5/0476 A61B5/00		
CPC分类号	A61B5/04012 A61B5/0476 A61B5/7203 A61B5/7225 A61B5/7235 A61B5/725 A61B5/7257		
优先权	201811001414.0 2018-08-30 CN		
外部链接	Espacenet SIPO		

摘要(译)

本发明公开了一种基于人工智能的癫痫患者脑电信号的预处理和识别方法，涉及脑科学、癫痫发作临床数据识别技术领域。所述方法采集癫痫患者的脑电信号，并对所述的脑电信号进行预处理；对预处理后的信号进行时频分析提取时频特征，同时，提取无噪声脑电信号的统计学特征；然后构建随机森林识别模型，进行不同程度病况的癫痫患者脑电信号的模式识别测试。本发明提供的特征提取方法，可以简化计算过程，提高脑电信号识别的准确性和实效性。由于随机森林识别方法的高度并行化，提高了运行效率，实现了计算机自动识别癫痫患者脑电信号的功能，为医疗工作者耗时耗力的诊断提供技术支持。

计算机编程实现

