



(12)发明专利申请

(10)申请公布号 CN 109674468 A

(43)申请公布日 2019.04.26

(21)申请号 201910088652.8

(22)申请日 2019.01.30

(71)申请人 大连理工大学

地址 116024 辽宁省大连市高新园区凌工
路2号

(72)发明人 刘蓉 梁洪宇 王永轩

(74)专利代理机构 大连星海专利事务所有限公
司 21208

代理人 王树本 徐雪莲

(51)Int.Cl.

A61B 5/0476(2006.01)

A61B 5/00(2006.01)

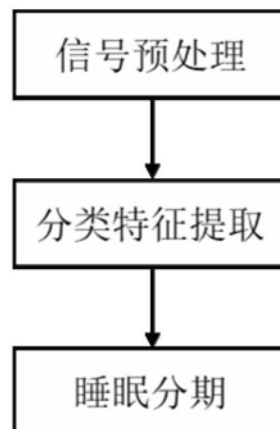
权利要求书4页 说明书11页 附图3页

(54)发明名称

一种单导脑电自动睡眠分期方法

(57)摘要

本发明属于睡眠监测技术领域,一种单导脑电自动睡眠分期方法,包括以下步骤:(1)信号预处理,(2)分类特征提取,(3)睡眠分期。本发明具有以下优点:一是设计了预处理算法,获得质量较好的单导脑电信号;二是本发明从时域、频域、非线性领域提取出多个特征,筛选出具有代表性的特征;三是本发明采用随机森林模型,所述方法不必担心过度拟合、具有很好的抗噪声能力,并且能够得出每一棵决策树的分期结果,从而得到随机森林对每一个睡眠分期的置信概率;四是本发明结合D-S证据理论进一步提高睡眠分期的准确率。



1. 一种单导脑电自动睡眠分期方法,其特征在于包括以下步骤:

步骤1、信号预处理,考虑到原始脑电信号含有各种噪声干扰,所以首先要对原始脑电信号进行预处理,具体包括以下子步骤:

(A) 去除基线漂移,基线漂移属于低频噪声,是一种非平稳随机信号,基于小波变换的多分辨率性质采用小波分解重构的方法去除基线漂移;

(B) 去除工频干扰,工频干扰强度大,而且是在非屏蔽环境内无处不在的,必须通过专用的陷波器滤除工频干扰,采用4阶IIR陷波器就能达到很好的效果,陷波频率设置为50Hz;

(C) 去除非脑电信号成分,睡眠脑电信号包含以下几种脑电特征波: α 波频率范围为8~13Hz、 β 波频率范围为13~30Hz、 δ 波频率范围为1~4Hz、 θ 波频率范围为4~8Hz、锯齿波频率范围为2~6Hz、纺锤波频率范围为12~14Hz、 β_1 波频率范围为13~17Hz、 β_2 波频率范围为17~30Hz;对于在这些频率之外的信号成分,是由人体其他电生理活动产生的,需要滤除;采用10阶IIR巴特沃斯带通滤波器,考虑到脑电信号的有效成分,将带通滤波频率设定为0.2~40Hz;

(D) 提取特征波形,为下一步提取时频域分类特征做准备,采用小波包分解wavelet packets decomposition即WPD的方法,将脑电信号分解为 α 波、 β 波、 δ 波、 θ 波、锯齿波、纺锤波、 β_1 波和 β_2 波各种节律成分,即将脑电信号在时频平面上细致划分,以提高脑电信号的分析能力;

步骤2、分类特征提取,包括时域特征提取、频域特征提取及非线性特征提取,将所有特征提取完毕后,进行特征筛选,筛选出质量好、具有代表性的特征形成分类特征向量;考虑每个睡眠分期持续时长及分析数据信息量的关系,设定以每 $\tau=30s$ 的数据进行一次睡眠分期计算,每次处理的脑电数据点总个数为 $N=f_s \cdot \tau$,其中 f_s 表示待分析脑电数据的采样频率,分类特征提取具体包括以下子步骤:

(A) 时域特征提取,包括方差var,过零率,Hjorth Mobility,Hjorth Complexity,75百分位数, α 波能量、 β 波能量、 δ 波能量、 θ 波能量、锯齿波能量、纺锤波能量、 β_1 波能量及 β_2 波能量;时域特征中的方差var和75百分位数常见,这里不再叙述;过零率是信号的符号变化比率,即脑电数据从正数变成负数及从负数变为正数的总个数与脑电数据点总个数的比值,过零率ZCR通过公式(1)进行描述,

$$ZCR = \frac{1}{2N} \sum_{in=2}^N |\text{sgn}(x_{in}) - \text{sgn}(x_{in-1})| \quad (1)$$

其中, x_{in} ($1 \leq i \leq N$) 表示第 in 个脑电数据的信号幅值, $\text{sgn}()$ 为符号函数,通过公式(2)进行描述,

$$\text{sgn}(x_{in}) = \begin{cases} 1 & x_{in} > 0 \\ 0 & x_{in} = 0 \\ -1 & x_{in} < 0 \end{cases} \quad (2)$$

Hjorth Mobility表示平均频率,由公式(3)计算得到,

$$Hjorth\ Mobility = \sqrt{\frac{m_2(x)}{m_0(x)}} \quad (3)$$

其中, m_0 表示脑电数据 x 的方差, m_2 表示脑电数据 x 一阶导数的方差;

Hjorth Complexity表示频率变化,由公式(4)计算得到,

$$Hjorth\ Complexity = \sqrt{\frac{m_4(x)/m_2(x)}{m_2(x)/m_0(x)}} \quad (4)$$

其中 m_4 表示脑电数据 x 二阶导数的方差;

各类特征波能量,即 α 波能量、 β 波能量、 δ 波能量、 θ 波能量、锯齿波能量、纺锤波能量、 β_1 波能量及 β_2 波能量Energy分别由公式(5)计算得到,

$$Energy = \sum_{in=1}^N x_{in}^2 \quad (5)$$

(B) 频域特征提取,包括 α 波、 β 波、 δ 波、 θ 波绝对功率谱, α 波、 β 波、 δ 波、 θ 波相对功率谱,脑电特征波功率比,是其绝对功率谱的比值,即 δ/α 、 δ/β 、 δ/θ 、 θ/α 、 θ/β 、 α/β 、 $(\delta+\theta)/(\alpha+\beta)$ 、 $\alpha \times \beta_1/\delta$ 、 $\theta \times \beta_2/\delta$,相对功率谱,是其绝对功率谱与总功率谱的比值,功率谱计算采用Welch法,将脑电数据 x 分成 L 段,每段的长度为 M ,分别求出每一段的功率谱,然后平均,进而计算功率谱;Welch方法允许每一段都有部分数据重叠,采用汉明窗,50%重叠进行周期图谱估计,设第 l 段的功率谱为 $\tilde{P}_{PER}^l(f)$,通过公式(6)进行描述,

$$\tilde{P}_{PER}^l(f) = \frac{1}{M \cdot U} \left| \sum_{m=1}^M x_m^l \cdot w_m \cdot e^{-j2\pi \cdot f \cdot m} \right|^2 \quad (6)$$

其中, $1 \leq m \leq M$, $1 \leq l \leq L$, x_m^l 表示第 l 段数据中第 m 个脑电数据点, w_m 表示所加的窗函数, f 为频率,窗函数平均能量 U 通过公式(7)计算得到,

$$U = \frac{1}{M} \sum_{m=1}^M w_m^2 \quad (7)$$

平均后的功率谱通过公式(8)进行描述,

$$\tilde{P}_{PER}(f) = \frac{1}{L} \sum_{l=1}^L \tilde{P}_{PER}^l(f) = \frac{1}{M \cdot U \cdot L} \sum_{l=1}^L \left| \sum_{m=1}^M x_m^l \cdot w_m \cdot e^{-j2\pi \cdot f \cdot m} \right|^2 \quad (8)$$

计算得到各类特征波的功率谱密度和频率,并计算其曲线下面积,得到绝对功率谱、相对功率谱、功率比;

(C) 非线性特征提取,采用的非线性特征为样本熵,样本熵的值越大,样本序列就越复杂,样本熵SampleEnt的计算具体包括以下子步骤:

(a) 导入脑电数据 x ,按照顺序将 x 组成一组嵌入维数为 dim 的向量组 $X_1, X_2, \dots, X_i, \dots, X_{N-dim+1}$,其中 X_i 通过公式(9)进行描述, x_i 表示脑电数据 x 第 i 个脑电数据的信号幅值;

$$X_i = [x_i, x_{i+1}, \dots, x_{i+dim-1}] \quad (9)$$

(b) X_i 与 X_j 表示向量组中第 i 个和第 j 个向量,将 X_i 与 X_j 间的距离 $d[X_i, X_j]$ 定义为两个向量中元素差值最大的一个,通过公式(10)进行描述, x_j 表示脑电数据 x 第 j 个脑电数据的信号幅值;

$$d[X_i, X_j] = \max |x_{i+k} - x_{j+k}| \quad \begin{cases} 1 \leq k \leq dim-1 \\ 1 \leq i \leq N-dim+1 \\ 1 \leq j \leq N-dim+1 \end{cases} \quad (10)$$

(c) 对于满足公式(10)的*i*和*j*,统计 $d[X_i, X_j]$ 小于“相似度”的度量值*r*的数目即模板匹配数,计算此数目与矢量总个数的比值,记为平均模板匹配数 $A_i^{dim}(r)$,通过公式(11)进行描述,

$$A_i^{dim}(r) = \frac{1}{N-dim} \left\{ d[X_i, X_j] < r \text{ 的数目} \right\} \begin{cases} 1 \leq i \leq N-dim+1 \\ 1 \leq j \leq N-dim+1 \\ i \neq j \end{cases} \quad (11)$$

(d) $A_i^{dim}(r)$ 的平均值 $A^{dim}(r)$,通过公式(12)计算得到,

$$A^{dim}(r) = \frac{1}{N-M+1} \sum_{i=1}^{N-M+1} A_i^{dim}(r) \quad (12)$$

(e) 嵌入维数 $dim+1$,组成 $dim+1$ 维矢量,重复上述子步骤(C)中的子步骤(a)~(d),得到 $A^{dim+1}(r)$;

(f) 定义样本熵,通过公式(13)计算得到,

$$\text{SampleEnt}(dim, r) = \lim_{N \rightarrow \infty} \left\{ -\ln \frac{A^{dim+1}(r)}{A^{dim}(r)} \right\} \quad (13)$$

(g) 当*N*为有限值时,样本熵通过公式(14)计算得到,

$$\text{SampleEnt}(dim, r, N) = -\ln \frac{A^{dim+1}(r)}{A^{dim}(r)} \quad (14)$$

(D) 特征筛选,将提取到的特征与人工睡眠分期结果组成特征向量组,基于过滤法中的最小冗余最大相关性进行特征筛选,按照相关性对各个特征进行评分,所述特征中,样本熵与分类标记之间相关性大,因此保留该特征,其它特征主要从时域与频域特征中进行筛选;设定拟选择的特征数量,最大化特征与分类变量之间的相关性,最小化特征与特征之间的相关性,从而筛选出具有代表性的特征;

步骤3、睡眠分期,将经过特征提取与特征筛选后的脑电特征重新组成训练集与测试集,采用随机森林的方法得到睡眠分期的模型,并进行睡眠分期验证,得到各个睡眠分期的置信概率,最后将置信概率进行D-S证据理论融合,得到单导脑电睡眠分期的结果,睡眠分期具体包括以下子步骤:

(A) 划分数据集,将时域特征,非线性特征以及睡眠分期结果组成第一个数据集Data1,将频域特征和睡眠分期结果组成第二个数据集Data2;

(B) 模型训练,将上述两个数据集,分别随机抽取80%数据作为训练集,20%数据作为测试集,训练随机森林模型model1和model2,并通过测试集进行验证,得到各个睡眠分期的置信概率,具体包括以下子步骤:

(a) 随机抽样,应用bootstrap方法有放回的从训练集中抽取TN次,构成TN个样本,每次未被抽到的样本组成TN个袋外数据Out-of-bag, OBB用来估计模型的预测准确度;

(b) 构建TN棵决策树,设有*nf*个特征,在每一棵树的每个节点处随机抽取FN个特征($FN = \lfloor nf \rfloor$),通过计算每个特征蕴含的信息量,在FN个特征中选择一个最具有分类能力的特征进行节点分裂,并且每棵树最大限度地生长,不做任何剪裁,从而构建出TN棵决策树;

(c) 决策树棵树个数寻优,通常TN的范围是100~800,在取值范围内利用子步骤(B)子

步骤(a)中未被抽到的袋外数据来估计决策树模型准确度,当准确度最高时,对应最优决策树个数 TN^* ;

(d) 构建随机森林,计算置信概率,将 TN^* 棵决策树组合在一起构成随机森林模型,应用数据集Data1得到model1,重复上述方法用数据集Data2训练随机森林模型得到model2,将测试数据输入到所述模型,得到对W、N1、N2、N3、REM五个分期的概率 P_{tw} ,其中 $t=1,2,w=1,2\cdots 5$, P_{tw} 表示第t个随机森林模型中睡眠分期为第w个分期的数目与总决策树数目的比值;

(C) D-S证据理论数据融合,D-S证据理论数据融合方法是指Dempster组合规则,也称为证据组合公式,设 Θ 为识别框架,由睡眠分期W、N1、N2、N3、REM、不确定分期和空集 Φ 组成单元素命题识别框架,A代表识别框架中任一命题,在识别框架上定义两个随机森林模型的基本概率分配函数 $mass_t(A) \in (0,1)$, $t=1,2$ 满足:

$$mass_t(\Phi) = 0 \quad (15)$$

$$\sum_{A \subseteq \Theta} mass_t(A) = 1 \quad (16)$$

对于 $\forall A \subseteq \Theta$,识别框架 Θ 上的 $mass_1, mass_2$ 函数的Dempster合成规则通过公式(17)进行描述,

$$(mass_1 \oplus mass_2)(A) = \frac{1}{K} \sum_{A_p \cap A_q = A} mass_1(A_p) \cdot mass_2(A_q) \quad (17)$$

其中, $1 \leq p \leq 7, 1 \leq q \leq 7$, A_1 代表识别框架中的W, A_2 代表识别框架中的N1, A_3 代表识别框架中的N2, A_4 代表识别框架中的N3, A_5 代表识别框架中的REM, A_6 代表识别框架中的不确定分期, A_7 代表识别框架中的 Φ ;K为归一化常数:

$$K = \sum_{A_p \cap A_q \neq \emptyset} mass_1(A_p) \cdot mass_2(A_q) = 1 - \sum_{A_p \cap A_q = \emptyset} mass_1(A_p) \cdot mass_2(A_q) \quad (18)$$

对于命题A的可信度分配 $Bel(A)$ 为,

$$Bel(A) = \sum_{B \subseteq A} mass_t(B) \quad (19)$$

表示命题A中所有子集B的基本概率分配之和,即对A的总的信任度,当A是单元素命题时,

$$Bel(A) = mass_t(A) \quad (20)$$

两个随机森林模型对每个睡眠分期的概率作为基本概率赋值, $mass_t(A_1)$ 即 P_{t1} , $mass_t(A_2)$ 即 P_{t2} , $mass_t(A_3)$ 即 P_{t3} , $mass_t(A_4)$ 即 P_{t4} , $mass_t(A_5)$ 即 P_{t5} , $mass_t(A_6)$ 和 $mass_t(A_7)$ 为0,运用上述Dempster合成规则,将各个证据合成一个新的证据体,将不同证据体的基本可信度分配合并产生一个总体的可信度分配 $Bel(A)$,并根据最大信任值法,计算各命题的信任函数值,选择具有最大信任值的结果作为最终识别的睡眠分期结果。

一种单导脑电自动睡眠分期方法

技术领域

[0001] 本发明涉及一种单导脑电自动睡眠分期方法,属于睡眠监测技术领域。

背景技术

[0002] 睡眠是生命的需要,是人重要的生理活动之一。根据最新美国睡眠医学会AASM标准,睡眠可分为清醒期W,非快速眼动期N1、N2、N3,快速眼动期REM,其中N1和N2为浅睡眠期,N3为深度睡眠期。足够的深度睡眠时长对于缓解疲劳程度、提高工作效率有很大的促进作用;心脑血管疾病发病概率也与不同的睡眠分期之间存在着一定的关系;有一些大脑方面的疾病在清醒状态下不易被发觉,但随着睡眠深度的增加,大脑活动程度降低,这些病灶就开始显现出来,并且在不同的睡眠阶段显现出来的症状也不同。因此,准确的睡眠分期显得尤为重要。

[0003] 目前的临床睡眠监测技术多采用多导睡眠监测设备(Polysomnography, PSG)。PSG一般要求在睡眠实验室中进行,同步采集脑电信号(Electroencephalogram, EEG)、心电信号(Electrocardiograph, ECG)、眼电信号(Electrooculogram, EOG)、肌电信号(Electromyography, EMG)等多导电生理信号,综合以上电生理信号实现睡眠分期。其不足之处在于:(1)使用大量的电极穿戴在被试者头部,使被试者感到不适,影响其正常的睡眠活动;(2)PSG价格较为昂贵,且需要专业的采集设备及合适的实验环境;(3)需要有经验的相关专家来进行睡眠分期,耗费时间长,增加专家的工作量;(4)分期结果含有一定的主观因素。因此,基于较少导联或单导脑电信号的自动睡眠分期技术成为睡眠监测领域的重要发展方向之一。

发明内容

[0004] 针对现有技术中存在的不足,本发明目的是提供一种单导脑电自动睡眠分期方法。该方法通过对单导脑电信号进行处理,提取并筛选具有代表性的脑电特征,建立睡眠分期模型,结合有限状态机理论和D-S证据理论实现自动分期,可广泛应用于睡眠监测技术领域。

[0005] 为了实现上述发明目的,本发明采取的技术方案是:一种单导脑电自动睡眠分期方法,包括以下步骤:

[0006] 步骤1、信号预处理,考虑到原始脑电信号含有各种噪声干扰,所以首先要对原始脑电信号进行预处理,具体包括以下子步骤:

[0007] (A) 去除基线漂移,基线漂移属于低频噪声,是一种非平稳随机信号,基于小波变换的多分辨率性质采用小波分解重构的方法去除基线漂移;

[0008] (B) 去除工频干扰,工频干扰强度大,而且是在非屏蔽环境内无处不在的,必须通过专用的陷波器滤除工频干扰,采用4阶IIR陷波器就能达到很好的效果,陷波频率设置为50Hz;

[0009] (C) 去除非脑电信号成分,睡眠脑电信号包含以下几种脑电特征波: α 波频率范围

为8~13Hz、β波频率范围为13~30Hz、δ波频率范围为1~4Hz、θ波频率范围为4~8Hz、锯齿波频率范围为2~6Hz、纺锤波频率范围为12~14Hz、β1波频率范围为13~17Hz、β2波频率范围为17~30Hz；对于在这些频率之外的信号成分，是由人体其他电生理活动产生的，需要滤除；采用10阶IIR巴特沃斯带通滤波器，考虑到脑电信号的有效成分，将带通滤波频率设定为0.2~40Hz；

[0010] (D) 提取特征波形，为下一步提取时频域分类特征做准备，采用小波包分解wavelet packets decomposition即WPD的方法，将脑电信号分解为α波、β波、δ波、θ波、锯齿波、纺锤波、β1波和β2波各种节律成分，即将脑电信号在时频平面上细致划分；在小波分析理论的基础上，引入最优基选择的概念，将频带经过多层次的划分之后，根据被分析脑电信号的特征，自适应地选取最佳基函数，使之与脑电信号相匹配，以提高脑电信号的分析能力；

[0011] 步骤2、分类特征提取，包括时域特征提取、频域特征提取及非线性特征提取，将所有特征提取完毕后，进行特征筛选，筛选出质量好、具有代表性的特征形成分类特征向量；考虑每个睡眠分期持续时长及分析数据信息量的关系，设定以每 $\tau=30s$ 的数据进行一次睡眠分期计算，每次处理的脑电数据点总个数为 $N=f_s \cdot \tau$ ，其中 f_s 表示待分析脑电数据的采样频率，分类特征提取具体包括以下子步骤：

[0012] (A) 时域特征提取，包括方差var，过零率，Hjorth Mobility，Hjorth Complexity，75百分位数，α波能量、β波能量、δ波能量、θ波能量、锯齿波能量、纺锤波能量、β1波能量及β2波能量；时域特征中的方差var和75百分位数常见，这里不再叙述；过零率是信号的符号变化比率，即脑电数据从正数变成负数及从负数变为正数的总个数与脑电数据点总个数的比值，过零率ZCR通过公式(1)进行描述，

$$[0013] \quad ZCR = \frac{1}{2N} \sum_{in=2}^N |\text{sgn}(x_{in}) - \text{sgn}(x_{in-1})| \quad (1)$$

[0014] 其中， x_{in} ($1 \leq i \leq N$) 表示第in个脑电数据的信号幅值， $\text{sgn}()$ 为符号函数，通过公式(2)进行描述，

$$[0015] \quad \text{sgn}(x_{in}) = \begin{cases} 1 & x_{in} > 0 \\ 0 & x_{in} = 0 \\ -1 & x_{in} < 0 \end{cases} \quad (2)$$

[0016] Hjorth Mobility表示平均频率，由公式(3)计算得到，

$$[0017] \quad \text{Hjorth Mobility} = \sqrt{\frac{m_2(x)}{m_0(x)}} \quad (3)$$

[0018] 其中， m_0 表示脑电数据x的方差， m_2 表示脑电数据x一阶导数的方差；

[0019] Hjorth Complexity表示频率变化，由公式(4)计算得到，

$$[0020] \quad \text{Hjorth Complexity} = \sqrt{\frac{m_4(x)/m_2(x)}{m_2(x)/m_0(x)}} \quad (4)$$

[0021] 其中 m_4 表示脑电数据x二阶导数的方差；

[0022] 各类特征波能量，即α波能量、β波能量、δ波能量、θ波能量、锯齿波能量、纺锤波能量、β1波能量及β2波能量Energy分别由公式(5)计算得到，

$$[0023] \quad Energy = \sum_{in=1}^N x_{in}^2 \quad (5)$$

[0024] (B) 频域特征提取, 包括 α 波、 β 波、 δ 波、 θ 波绝对功率谱, α 波、 β 波、 δ 波、 θ 波相对功率谱, 脑电特征波功率比, 是其绝对功率谱的比值, 即 δ/α 、 δ/β 、 δ/θ 、 θ/α 、 θ/β 、 α/β 、 $(\delta+\theta)/(\alpha+\beta)$ 、 $\alpha \times \beta_1/\delta$ 、 $\theta \times \beta_2/\delta$, 相对功率谱, 是其绝对功率谱与总功率谱的比值, 功率谱计算采用Welch法, 将脑电数据 x 分成 L 段, 每段的长度为 M , 分别求出每一段的功率谱, 然后平均, 进而计算功率谱; Welch方法允许每一段都有部分数据重叠, 采用汉明窗, 50%重叠进行周期图谱估计, 设第 l 段的功率谱为 $\tilde{P}_{PER}^l(f)$, 通过公式 (6) 进行描述,

$$[0025] \quad \tilde{P}_{PER}^l(f) = \frac{1}{M \cdot U} \left| \sum_{m=1}^M x_m^l \cdot w_m \cdot e^{-j2\pi \cdot f \cdot m} \right|^2 \quad (6)$$

[0026] 其中, $1 \leq m \leq M$, $1 \leq l \leq L$, x_m^l 表示第 l 段数据中第 m 个脑电数据点, w_m 表示所加的窗函数, f 为频率, 窗函数平均能量 U 通过公式 (7) 计算得到,

$$[0027] \quad U = \frac{1}{M} \sum_{m=1}^M w_m^2 \quad (7)$$

[0028] 平均后的功率谱通过公式 (8) 进行描述,

$$[0029] \quad \tilde{P}_{PER}(f) = \frac{1}{L} \sum_{l=1}^L \tilde{P}_{PER}^l(f) = \frac{1}{M \cdot U \cdot L} \sum_{l=1}^L \left| \sum_{m=1}^M x_m^l \cdot w_m \cdot e^{-j2\pi \cdot f \cdot m} \right|^2 \quad (8)$$

[0030] 计算得到各类特征波的功率谱密度和频率, 并计算其曲线下面积, 得到绝对功率谱、相对功率谱、功率比;

[0031] (C) 非线性特征提取, 采用的非线性特征为样本熵, 样本熵的值越大, 样本序列就越复杂, 样本熵SampleEnt的计算具体包括以下子步骤:

[0032] (a) 导入脑电数据 x , 按照顺序将 x 组成一组嵌入维数为 dim 的向量组 $X_1, X_2, \dots, X_i, \dots, X_{N-dim+1}$, 其中 X_i 通过公式 (9) 进行描述, x_i 表示脑电数据 x 第 i 个脑电数据的信号幅值;

$$[0033] \quad X_i = [x_i, x_{i+1}, \dots, x_{i+dim-1}] \quad (9)$$

[0034] (b) X_i 与 X_j 表示向量组中第 i 个和第 j 个向量, 将 X_i 与 X_j 间的距离 $d[X_i, X_j]$ 定义为两个矢量中元素差值最大的一个, 通过公式 (10) 进行描述, x_j 表示脑电数据 x 第 j 个脑电数据的信号幅值;

$$[0035] \quad d[X_i, X_j] = \max |x_{i+k} - x_{j+k}| \quad \begin{cases} 1 \leq k \leq dim-1 \\ 1 \leq i \leq N-dim+1 \\ 1 \leq j \leq N-dim+1 \end{cases} \quad (10)$$

[0036] (c) 对于满足公式 (10) 的 i 和 j , 统计 $d[X_i, X_j]$ 小于“相似度”的度量值 r 的数目即模板匹配数, 计算此数目与矢量总个数的比值, 记为平均模板匹配数 $A_i^{dim}(r)$, 通过公式 (11) 进行描述,

$$[0037] \quad A_i^{dim}(r) = \frac{1}{N-dim} \{d[X_i, X_j] < r \text{ 的数目}\} \quad \begin{cases} 1 \leq i \leq N-dim+1 \\ 1 \leq j \leq N-dim+1 \\ i \neq j \end{cases} \quad (11)$$

[0038] (d) $A_i^{dim}(r)$ 的平均值 $A^{dim}(r)$, 通过公式 (12) 计算得到,

$$[0039] \quad A^{dim}(r) = \frac{1}{N-M+1} \sum_{i=1}^{N-M+1} A_i^{dim}(r) \quad (12)$$

[0040] (e) 嵌入维数 $dim+1$, 组成 $dim+1$ 维矢量, 重复上述子步骤 (C) 中的子步骤 (a) ~ (d), 得到 $A^{dim+1}(r)$;

[0041] (f) 定义样本熵, 通过公式 (13) 计算得到,

$$[0042] \quad \text{SampleEnt}(dim, r) = \lim_{N \rightarrow \infty} \left\{ -\ln \frac{A^{dim+1}(r)}{A^{dim}(r)} \right\} \quad (13);$$

[0043] (g) 当 N 为有限值时, 样本熵通过公式 (14) 计算得到,

$$[0044] \quad \text{SampleEnt}(dim, r, N) = -\ln \frac{A^{dim+1}(r)}{A^{dim}(r)} \quad (14);$$

[0045] (D) 特征筛选, 将提取到的特征与人工睡眠分期结果组成特征向量组, 基于过滤法中的最小冗余最大相关性进行特征筛选, 按照相关性对各个特征进行评分, 所述特征中, 样本熵与分类标记之间相关性大, 因此保留该特征, 其它特征主要从时域与频域特征中进行筛选; 设定拟选择的特征数量, 最大化特征与分类变量之间的相关性, 最小化特征与特征之间的相关性, 从而筛选出具有代表性的特征;

[0046] 步骤3、睡眠分期, 将经过特征提取与特征筛选后的脑电特征重新组成训练集与测试集, 采用随机森林的方法得到睡眠分期的模型, 并进行睡眠分期验证, 得到各个睡眠分期的置信概率, 最后将置信概率进行D-S证据理论融合, 得到单导脑电睡眠分期的结果, 睡眠分期具体包括以下子步骤:

[0047] (A) 划分数据集, 将时域特征, 非线性特征以及睡眠分期结果组成第一个数据集 Data1, 将频域特征和睡眠分期结果组成第二个数据集 Data2;

[0048] (B) 模型训练, 将上述两个数据集, 分别随机抽取80%数据作为训练集, 20%数据作为测试集, 训练随机森林模型 model1 和 model2, 并通过测试集进行验证, 得到各个睡眠分期的置信概率, 具体包括以下子步骤:

[0049] (a) 随机抽样, 应用bootstrap方法有放回的从训练集中抽取 TN 次, 构成 TN 个样本, 每次未被抽到的样本组成 TN 个袋外数据 Out-of-bag, OBB 用来估计模型的预测准确度;

[0050] (b) 构建 TN 棵决策树, 设有 nf 个特征, 在每一棵树的每个节点处随机抽取 FN 个特征 ($FN = \lfloor nf \rfloor$), 通过计算每个特征蕴含的信息量, 在 FN 个特征中选择一个最具有分类能力的特征进行节点分裂, 并且每棵树最大限度地生长, 不做任何剪裁, 从而构建出 TN 棵决策树;

[0051] (c) 决策树棵树个数寻优, 通常 TN 的范围是 $100 \sim 800$, 在取值范围内利用子步骤 (B) 子步骤 (a) 中未被抽到的袋外数据来估计决策树模型准确度, 当准确度最高时, 对应最优决策树个数 TN^* ;

[0052] (d) 构建随机森林, 计算置信概率, 将 TN^* 棵决策树组合在一起构成随机森林模型, 应用数据集 Data1 得到 model1, 重复上述方法用数据集 Data2 训练随机森林模型得到 model2, 将测试数据输入到所述模型, 得到对 W、N1、N2、N3、REM 五个分期的概率 P_{tw} , 其中 $t = 1, 2, w = 1, 2 \cdots 5$, P_{tw} 表示第 t 个随机森林模型中睡眠分期为第 w 个分期的数目与总决策树数

目的比值；

[0053] (C) D-S证据理论数据融合, D-S证据理论数据融合方法是指Dempster组合规则, 也称为证据组合公式, 设 Θ 为识别框架, 由睡眠分期W、N1、N2、N3、REM、不确定分期和空集 Φ 组成单元素命题识别框架, A代表识别框架中任一命题, 在识别框架上定义两个随机森林模型的基本概率分配函数 $mass_t(A) \in (0, 1)$, $t=1, 2$ 满足:

$$[0054] \quad mass_t(\Phi) = 0 \quad (15)$$

$$[0055] \quad \sum_{A \subseteq \Theta} mass_t(A) = 1 \quad (16)$$

[0056] 对于 $\forall A \subseteq \Theta$, 识别框架 Θ 上的 $mass_1, mass_2$ 函数的Dempster合成规则通过公式(17)进行描述,

$$[0057] \quad (mass_1 \oplus mass_2)(A) = \frac{1}{K} \sum_{A_p \cap A_q = A} mass_1(A_p) \cdot mass_2(A_q) \quad (17)$$

[0058] 其中, $1 \leq p \leq 7, 1 \leq q \leq 7$, A_1 代表识别框架中的W, A_2 代表识别框架中的N1, A_3 代表识别框架中的N2, A_4 代表识别框架中的N3, A_5 代表识别框架中的REM, A_6 代表识别框架中的不确定分期, A_7 代表识别框架中的 Φ ; K为归一化常数:

$$[0059] \quad K = \sum_{A_p \cap A_q \neq \emptyset} mass_1(A_p) \cdot mass_2(A_q) = 1 - \sum_{A_p \cap A_q = \emptyset} mass_1(A_p) \cdot mass_2(A_q) \quad (18)$$

[0060] 对于命题A的可信度分配 $Bel(A)$ 为,

$$[0061] \quad Bel(A) = \sum_{B \subseteq A} mass_t(B) \quad (19)$$

[0062] 表示命题A中所有子集B的基本概率分配之和, 即对A的总的信任度, 当A是单元素命题时,

$$[0063] \quad Bel(A) = mass_t(A) \quad (20)$$

[0064] 两个随机森林模型对每个睡眠分期的概率作为基本概率赋值, $mass_t(A_1)$ 即 P_{t1} , $mass_t(A_2)$ 即 P_{t2} , $mass_t(A_3)$ 即 P_{t3} , $mass_t(A_4)$ 即 P_{t4} , $mass_t(A_5)$ 即 P_{t5} , $mass_t(A_6)$ 和 $mass_t(A_7)$ 为0, 运用上述Dempster合成规则, 将各个证据合成一个新的证据体, 将不同证据体的基本可信度分配合并产生一个总体的可信度分配 $Bel(A)$, 并根据最大信任值法, 计算各命题的信任函数值, 选择具有最大信任值的结果作为最终识别的睡眠分期结果。

[0065] 本发明有益效果是: 一种单导脑电自动睡眠分期方法, 包括以下步骤: (1) 信号预处理, (2) 分类特征提取, (3) 睡眠分期。与已有技术相比, 本发明具有以下优点: 一是设计了预处理算法, 获得质量较好的单导脑电信号; 二是本发明从时域、频域、非线性领域提取出三十多个特征, 并设计特征筛选算法, 筛选出具有代表性的特征; 三是本发明采用随机森林模型, 所述方法不必担心过度拟合、具有很好的抗噪声能力, 并且能够得出每一棵决策树的分期结果, 从而得到随机森林对每一个睡眠分期的置信概率; 四是本发明结合D-S证据理论进一步提高睡眠分期的准确率。总之, 本发明设计的一种单导脑电自动睡眠分期方法具有算法容易理解、解释性强、泛化能力强等优点。

附图说明

[0066] 图1是本发明方法步骤流程图。

- [0067] 图2是信号预处理方法步骤流程图。
 [0068] 图3是3层小波包分解结构示意图。
 [0069] 图4是提取脑电特征波流程图。
 [0070] 图5是分类特征提取步骤流程图。
 [0071] 图6是睡眠分期详细步骤流程图。

具体实施方式

[0072] 下面结合附图对本发明作进一步说明。

[0073] 如图1所示,一种单导脑电自动睡眠分期方法,包括以下步骤:

[0074] 步骤1、信号预处理,考虑到原始脑电信号含有各种噪声干扰,所以首先要对原始脑电信号进行预处理,具体包括以下子步骤,如图2所示。

[0075] (A) 去除基线漂移,基线漂移属于低频噪声,是一种非平稳随机信号,基于小波变换的多分辨率性质采用小波分解重构的方法去除基线漂移;导入脑电数据 x ,对 x 进行离散小波分解,第一层分解得到近似系数 C_1 和细节系数 D_1 ,其中近似系数主要为低频部分,细节系数主要为高频部分,对 C_1 再次进行小波分解,得到近似系数 C_2 和细节系数 D_2 ,以此类推;本发明采用‘db4’小波基对 x 进行8层离散小波分解,得到近似系数 C_8 和细节系数 D_1 、 D_2 、 D_3 、 D_4 、 D_5 、 D_6 、 D_7 、 D_8 ,其中 C_8 的频带范围为 $\left[0, \frac{f_s}{2^9}\right]$,在采样频率 f_s 为100Hz的条件下可得 C_8 的频带范

围为 $0 \sim 0.1953\text{Hz}$,这个频带范围对应于基线漂移;忽略 C_8 仅根据 $D_1 \sim D_8$ 进行重构,即可得到去除基线漂移后的脑电信号;

[0076] (B) 去除工频干扰,工频干扰强度高,而且是在非屏蔽环境内无处不在的,必须通过陷波器滤除工频干扰,采用4阶IIR陷波器就能达到很好的效果,陷波频率设置为50Hz;

[0077] (C) 去除非脑电信号成分,睡眠脑电信号包含以下几种脑电特征波: α 波频率范围为 $8 \sim 13\text{Hz}$ 、 β 波频率范围为 $13 \sim 30\text{Hz}$ 、 δ 波频率范围为 $1 \sim 4\text{Hz}$ 、 θ 波频率范围为 $4 \sim 8\text{Hz}$ 、锯齿波频率范围为 $2 \sim 6\text{Hz}$ 、纺锤波频率范围为 $12 \sim 14\text{Hz}$ 、 β_1 波频率范围为 $13 \sim 17\text{Hz}$ 、 β_2 波频率范围为 $17 \sim 30\text{Hz}$;对于在这些频率之外的信号成分,是由人体其他电生理活动产生的,需要滤除;采用10阶IIR巴特沃斯带通滤波器,考虑到脑电信号的有效成分,将带通滤波频率设定为 $0.2 \sim 40\text{Hz}$;

[0078] (D) 提取特征波形,为下一步提取时频域分类特征做准备,采用小波包分解 wavelet packets decomposition即WPD的方法,将脑电信号分解为 α 波、 β 波、 δ 波、 θ 波、锯齿波、纺锤波、 β_1 波和 β_2 波等各种节律成分,将脑电信号在时频平面上细致划分,提高脑电信号的分析能力;如图3所示,这是三层小波包分解结构示意图。对脑电数据 x 进行小波包分解,第一层分解得到近似系数 $(1, 0)$ 和细节系数 $(1, 1)$,其中近似系数主要为低频部分,细节系数主要为高频部分,对 $(1, 0)$ 再次进行小波包分解,得到近似系数 $(2, 0)$ 和细节系数 $(2, 1)$,同样 $(1, 1)$ 也可以进行分解,以此类推,对 x 进行6层小波包分解,共有 $2^6 = 64$ 个节点。每个节点可用一个坐标 (u, v) 表示,其中 u 代表小波包分解层数, v 代表 u 层的第 $v+1$ 个节点,图3所示的小波包分解树是一个满二叉树,为了计算方便, v 从0开始,这样小波包分解树任意非叶子节点的左子树坐标为 $(u+1, 2v)$ 。

[0079] 设信号的采样频率为 f_s ,所有第 z 层小波包分解系数的频带范围如下:

$$[0080] \quad \left[0, \frac{f_s}{2^{z+1}}\right], \left[\frac{f_s}{2^{z+1}}, \frac{2f_s}{2^{z+1}}\right], \left[\frac{2f_s}{2^{z+1}}, \frac{3f_s}{2^{z+1}}\right], \dots, \left[\frac{(2^z-1)f_s}{2^{z+1}}, \frac{f_s}{2}\right]$$

[0081] 当采样频率为100Hz时,对应第6层频带范围为:

[0082] $[0, 0.78125], [0.78125, 1.5625], [1.5625, 2.34375], \dots, [49.21875, 50]$

[0083] EEG信号包含以下几种脑电特征波,α波频率范围8~13Hz、β波频率范围13~30Hz、δ波频率范围1~4Hz、θ波频率范围4~8Hz、纺锤波频率范围12~14Hz、锯齿波频率范围2~6Hz、β1波频率范围12~18、β1波频率范围18~30,而人在睡眠过程中,不同的睡眠分期对应不同的特征波起主导作用,当人处于清醒期W时,α波、β波占据主导地位,快速眼动期REM也具有α波、β波特征,还具有锯齿波的特征,深睡期和浅睡期不具有α波、β波特征,因此可根据这些特征来区分W期和REM期;浅睡期为N1、N2,深睡期为N3,δ波只出现在深睡期,浅睡期可出现θ波、纺锤波、K复合波等一系列特征波,可通过这些特征来区分深度睡眠期和浅度睡眠期。因此,要对获取的脑电信号进行睡眠分期,需要获取脑电信号对应的特征波,本项目采取6层小波包分解的方法,以'db4'为小波基,获取α波、β波、θ波、δ波等一系列脑电特征波。如图4所示,这是提取脑电特征波流程图。首先将输入脑电信号进行6层小波包分解,得到64个节点,再根据要提取的脑电特征波,找出频带范围,选择对应节点;以α波为例,α波的频率为8~13Hz,选取频率范围为7.8125~12.5Hz的节点,提取α波特征波,得到的节点为(6,8)、(6,9)、(6,10)、(6,11)、(6,14)、(6,15),由于小波包分解树是一个满二叉树,任意非叶子节点与该节点左子树的纵坐标成2倍关系,因此可对上述节点进行一些合并,如(6,8)和(6,9)合并成(5,4), (6,10)和(6,11)合并成(5,5), (6,14)和(6,15)合并成(5,7), (5,4)和(5,5)合并成(4,2),最后得到α波对应小波包分解节点为(4,2)+(5,7),将这些节点进行小波包分解重构,即可得到重构后的α波脑电特征波。同理,重构出其余脑电特征波,为下一步特征提取做准备。

[0084] 步骤2、分类特征提取。如图5所示,对经过带通滤波处理后的信号进行时域特征和非线性特征提取,对提取的脑电特征波进行时域和频域特征提取,将所有特征提取完毕后,进行特征筛选,筛选出质量好、具有代表性的特征形成分类特征向量。综合考虑每个睡眠分期持续时长及分析数据信息量的关系,本发明设定以每 $\tau=30s$ 的数据进行一次睡眠分期计算。在待分析脑电数据的采样频率为100Hz的情况下,每次处理的脑电数据点总个数为 $N=f_s \cdot \tau=3000$,以下说明即以3000个数据点为例。所述分类特征具体计算包括以下子步骤:

[0085] (A) 时域特征提取,包括方差var,过零率,Hjorth Mobility,Hjorth Complexity,75百分位数,α波能量、β波能量、δ波能量、θ波能量、锯齿波能量、纺锤波能量、β1波能量及β2波能量;时域特征中的方差var和75百分位数常见,这里不再叙述;过零率是信号的符号变化比率,即脑电数据从正数变成负数及从负数变为正数的总个数与脑电数据点总个数的比值,过零率ZCR通过公式(1)进行描述,

$$[0086] \quad ZCR = \frac{1}{2N} \sum_{in=2}^N |\text{sgn}(x_{in}) - \text{sgn}(x_{in-1})| \quad (1)$$

[0087] 其中, $x_{in} (1 \leq i \leq N)$ 表示第in个脑电数据的信号幅值,sgn()为符号函数,通过公式(2)进行描述,

$$[0088] \quad \text{sgn}(x_{in}) = \begin{cases} 1 & x_{in} > 0 \\ 0 & x_{in} = 0 \\ -1 & x_{in} < 0 \end{cases} \quad (2)$$

[0089] Hjorth Mobility表示平均频率,由公式(3)计算得到,

$$[0090] \quad \text{Hjorth Mobility} = \sqrt{\frac{m_2(x)}{m_0(x)}} \quad (3)$$

[0091] 其中, m_0 表示脑电数据 x 的方差, m_2 表示脑电数据 x 一阶导数的方差;

[0092] Hjorth Complexity表示频率变化,由公式(4)计算得到,

$$[0093] \quad \text{Hjorth Complexity} = \sqrt{\frac{m_4(x)/m_2(x)}{m_2(x)/m_0(x)}} \quad (4)$$

[0094] 其中 m_4 表示脑电数据 x 二阶导数的方差;

[0095] 各类特征波能量,即 α 波能量、 β 波能量、 δ 波能量、 θ 波能量、锯齿波能量、纺锤波能量、 β_1 波能量及 β_2 波能量Energy分别由公式(5)计算得到,

$$[0096] \quad \text{Energy} = \sum_{in=1}^N x_{in}^2 \quad (5)$$

[0097] (B) 频域特征提取,包括 α 波、 β 波、 δ 波、 θ 波绝对功率谱, α 波、 β 波、 δ 波、 θ 波相对功率谱,脑电特征波功率比,是其绝对功率谱的比值,即 δ/α 、 δ/β 、 δ/θ 、 θ/α 、 θ/β 、 α/β 、 $(\delta+\theta)/(\alpha+\beta)$ 、 $\alpha \times \beta_1/\delta$ 、 $\theta \times \beta_2/\delta$,相对功率谱,是其绝对功率谱与总功率谱的比值,功率谱计算采用Welch法,将脑电数据 x 分成 L 段,每段的长度为 M ,分别求出每一段的功率谱,然后平均,进而计算功率谱;Welch方法允许每一段都有部分数据重叠,采用汉明窗,50%重叠进行周期图谱估计,设第 l 段的功率谱为 $\tilde{P}_{PER}^l(f)$,通过公式(6)进行描述,

$$[0098] \quad \tilde{P}_{PER}^l(f) = \frac{1}{M \cdot U} \left| \sum_{m=1}^M x_m^l \cdot w_m \cdot e^{-j2\pi \cdot f \cdot m} \right|^2 \quad (6)$$

[0099] 其中, $1 \leq m \leq M$, $1 \leq l \leq L$, x_m^l 表示第 l 段数据中第 m 个脑电数据点, w_m 表示所加的窗函数, f 为频率,窗函数平均能量 U 通过公式(7)计算得到,

$$[0100] \quad U = \frac{1}{M} \sum_{m=1}^M w_m^2 \quad (7)$$

[0101] 平均后的功率谱通过公式(8)进行描述,

$$[0102] \quad \tilde{P}_{PER}(f) = \frac{1}{L} \sum_{l=1}^L \tilde{P}_{PER}^l(f) = \frac{1}{M \cdot U \cdot L} \sum_{l=1}^L \left| \sum_{m=1}^M x_m^l \cdot w_m \cdot e^{-j2\pi \cdot f \cdot m} \right|^2 \quad (8)$$

[0103] 本发明采用汉明窗,窗宽为512,即为分割出每段信号的长度,每段之间有50%的重叠,通过以上公式计算得到功率谱密度和频率,频率的范围为0~50Hz;对频率范围为8~13Hz的 α 波而言,将经过小波包分解提取出的 α 脑电特征波进行周期图谱估计,得到功率谱密度和频率,并截取出8~13Hz功率谱密度部分,计算其曲线下面积,得到绝对功率谱;将原始脑电信号进行周期图谱估计,并计算其曲线下面积得到的总功率谱,而相对功率谱,是其绝对功率谱与总功率谱的比值;所述脑电特征波功率比,是其绝对功率谱的比值。

[0104] (C) 非线性特征提取,采用的非线性特征为样本熵,样本熵的值越大,样本序列就越复杂,样本熵SampleEnt的计算具体包括以下子步骤:

[0105] (a) 导入脑电数据 x ,按照顺序将 x 组成一组嵌入维数为 dim 的向量组 $X_1, X_2, \dots, X_i, \dots, X_{N-dim+1}$,其中 X_i 通过公式(9)进行描述, x_i 表示脑电数据 x 第 i 个脑电数据的信号幅值;

$$[0106] \quad X_i = [x_i, x_{i+1}, \dots, x_{i+dim-1}] \quad (9)$$

[0107] (b) X_i 与 X_j 表示向量组中第 i 个和第 j 个向量,将 X_i 与 X_j 间的距离 $d[X_i, X_j]$ 定义为两个矢量中元素差值最大的一个,通过公式(10)进行描述, x_j 表示脑电数据 x 第 j 个脑电数据的信号幅值;

$$[0108] \quad d[X_i, X_j] = \max |x_{i+k} - x_{j+k}| \quad \begin{cases} 1 \leq k \leq dim-1 \\ 1 \leq i \leq N-dim+1 \\ 1 \leq j \leq N-dim+1 \end{cases} \quad (10)$$

[0109] (c) 对于满足公式(10)的 i 和 j ,统计 $d[X_i, X_j]$ 小于“相似度”的度量值 r 的数目即模板匹配数,计算此数目与矢量总个数的比值,记为平均模板匹配数 $A_i^{dim}(r)$,通过公式(11)进行描述,

$$[0110] \quad A_i^{dim}(r) = \frac{1}{N-dim} \{d[X_i, X_j] < r \text{ 的数目} \} \quad \begin{cases} 1 \leq i \leq N-dim+1 \\ 1 \leq j \leq N-dim+1 \\ i \neq j \end{cases} \quad (11)$$

[0111] (d) $A_i^{dim}(r)$ 的平均值 $A^{dim}(r)$,通过公式(12)计算得到,

$$[0112] \quad A^{dim}(r) = \frac{1}{N-M+1} \sum_{i=1}^{N-M+1} A_i^{dim}(r) \quad (12)$$

[0113] (e) 嵌入维数 $dim+1$,组成 $dim+1$ 维矢量,重复上述子步骤(C)中的子步骤(a)~(d),得到 $A^{dim+1}(r)$;

[0114] (f) 定义样本熵,通过公式(13)计算得到,

$$[0115] \quad \text{SampleEnt}(dim, r) = \lim_{N \rightarrow \infty} \left\{ -\ln \frac{A^{dim+1}(r)}{A^{dim}(r)} \right\} \quad (13);$$

[0116] (g) 当 N 为有限值时,样本熵通过公式(14)计算得到,

$$[0117] \quad \text{SampleEnt}(dim, r, N) = -\ln \frac{A^{dim+1}(r)}{A^{dim}(r)} \quad (14);$$

[0118] 上述式子中,对于脑电数据 x 的样本熵计算, N 取3000, m 取2, r 表示相似容限,如果 r 取值太大,会丢失很多详细信息,本发明取 $r=0.2 \times \text{std}(X)$,所述 $\text{std}(X)$ 为输入脑电数据的标准差。

[0119] (D) 特征筛选,将提取到的特征与人工睡眠分期结果组成特征向量组,基于过滤法中的最小冗余最大相关性进行特征筛选,按照相关性对各个特征进行评分,所述特征中,样本熵与分类标记之间相关性大,因此保留该特征,其它特征主要从时域与频域特征中进行筛选;设定拟选择的特征数量,最大化特征与分类变量之间的相关性,最小化特征与特征之间的相关性,从而筛选出具有代表性的特征;

[0120] 步骤3、睡眠分期,具体步骤如图6所示,将经过特征提取与特征筛选后的脑电特征

重新组成训练集与测试集,采用随机森林的方法得到睡眠分期的模型,并进行睡眠分期验证,得到各个睡眠分期的置信概率,最后将所述置信概率进行D-S证据理论融合,得到单导脑电睡眠分期的结果。所述睡眠分期具体计算包括以下子步骤:

[0121] (A) 划分数据集,将时域特征,非线性特征以及睡眠分期结果组成第一个数据集Data1,将频域特征和睡眠分期结果组成第二个数据集Data2;

[0122] (B) 模型训练,将上述两个数据集,分别随机抽取80%数据作为训练集,20%数据作为测试集,训练随机森林模型model1和model2,并通过测试集进行验证,得到各个睡眠分期的置信概率,具体包括以下子步骤:

[0123] (a) 随机抽样,应用bootstrap方法有放回的从训练集中抽取TN次,构成TN个样本,每次未被抽到的样本组成TN个袋外数据Out-of-bag,0BB用来估计模型的预测准确度;

[0124] (b) 构建TN棵决策树,设有nf个特征,在每一棵树的每个节点处随机抽取FN个特征 ($FN = \lfloor nf \rfloor$),通过计算每个特征蕴含的信息量,在FN个特征中选择一个最具有分类能力的特征进行节点分裂,并且每棵树最大限度地生长,不做任何剪裁,从而构建出TN棵决策树;

[0125] (c) 决策树棵树个数寻优,通常TN的范围是100~800,在取值范围内利用子步骤(B)子步骤(a)中未被抽到的袋外数据来估计决策树模型准确度,当准确度最高时,对应最优决策树个数TN*;

[0126] (d) 构建随机森林,计算置信概率,将TN*棵决策树组合在一起构成随机森林模型,应用数据集Data1得到model1,重复上述方法用数据集Data2训练随机森林模型得到model2,将测试数据输入到所述模型,得到对W、N1、N2、N3、REM五个分期的概率 P_{tw} ,其中 $t=1,2,w=1,2,\dots,5$, P_{tw} 表示第t个随机森林模型中睡眠分期为第w个分期的数目与总决策树数目的比值;

[0127] (C) D-S证据理论融合,D-S证据理论数据融合方法是指Dempster组合规则,也称为证据组合公式,设 Θ 为识别框架,由睡眠分期W、N1、N2、N3、REM、不确定分期和空集 Φ 组成单元命题识别框架,A代表识别框架中任一命题,在识别框架上定义两个随机森林模型的基本概率分配函数 $mass_t(A) \in (0,1)$, $t=1,2$ 满足:

$$[0128] \quad mass_t(\Phi) = 0 \quad (15)$$

$$[0129] \quad \sum_{A \subseteq \Theta} mass_t(A) = 1 \quad (16)$$

[0130] 对于 $\forall A \subseteq \Theta$,识别框架 Θ 上的 $mass_1, mass_2$ 函数的Dempster合成规则通过公式(17)进行描述,

$$[0131] \quad (mass_1 \oplus mass_2)(A) = \frac{1}{K} \sum_{A_p \cap A_q = A} mass_1(A_p) \cdot mass_2(A_q) \quad (17)$$

[0132] 其中, $1 \leq p \leq 7, 1 \leq q \leq 7$, A_1 代表识别框架中的W, A_2 代表识别框架中的N1, A_3 代表识别框架中的N2, A_4 代表识别框架中的N3, A_5 代表识别框架中的REM, A_6 代表识别框架中的不确定分期, A_7 代表识别框架中的 Φ ;K为归一化常数:

$$[0133] \quad K = \sum_{A_p \cap A_q \neq \emptyset} mass_1(A_p) \cdot mass_2(A_q) = 1 - \sum_{A_p \cap A_q = \emptyset} mass_1(A_p) \cdot mass_2(A_q) \quad (18)$$

[0134] 对于命题A的可信度分配Bel(A)为,

[0135]
$$Bel(A) = \sum_{B \subseteq A} mass_t(B) \quad (19)$$

[0136] 表示命题A中所有子集B的基本概率分配之和,即对A的总的信任度,当A是单元素命题时,

[0137]
$$Bel(A) = mass_t(A) \quad (20)$$

[0138] 两个随机森林模型对每个睡眠分期的概率作为基本概率赋值, $mass_t(A_1)$ 即 P_{t1} , $mass_t(A_2)$ 即 P_{t2} , $mass_t(A_3)$ 即 P_{t3} , $mass_t(A_4)$ 即 P_{t4} , $mass_t(A_5)$ 即 P_{t5} , $mass_t(A_6)$ 和 $mass_t(A_7)$ 为0, 运用上述Dempster合成规则,将各个证据合成一个新的证据体,将不同证据体的基本可信度分配合并产生一个总体的可信度分配 $Bel(A)$, 并根据最大信任值法,计算各命题的信任函数值,选择具有最大信任值的结果作为最终识别的睡眠分期结果。

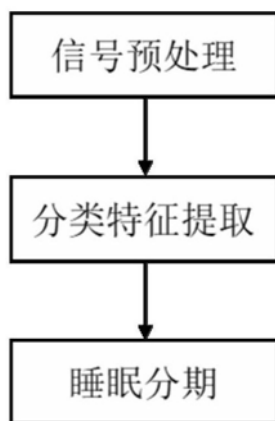


图1

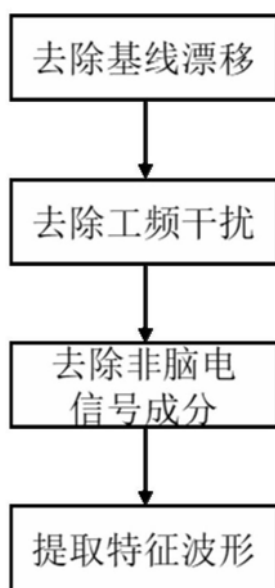


图2

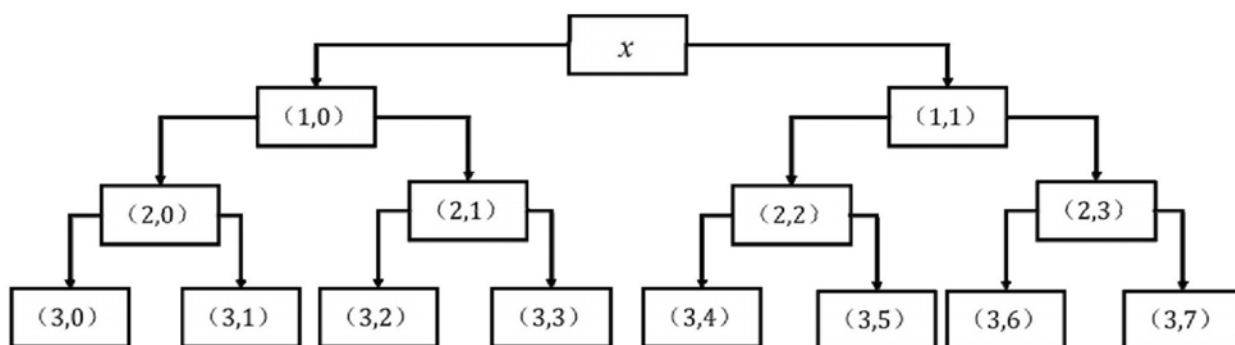


图3

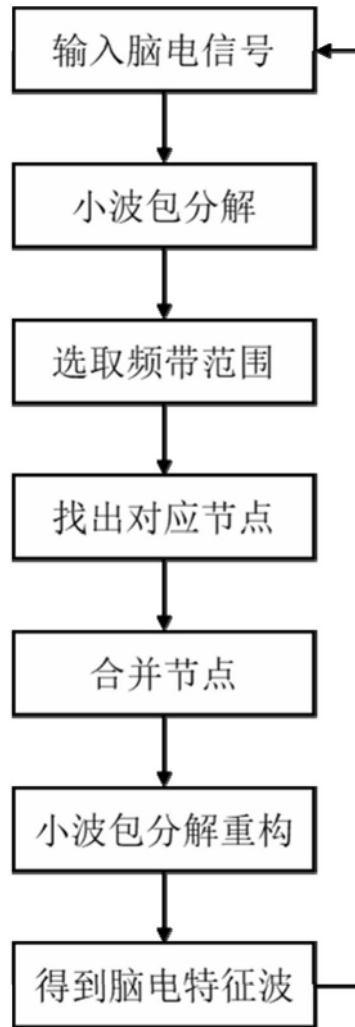


图4

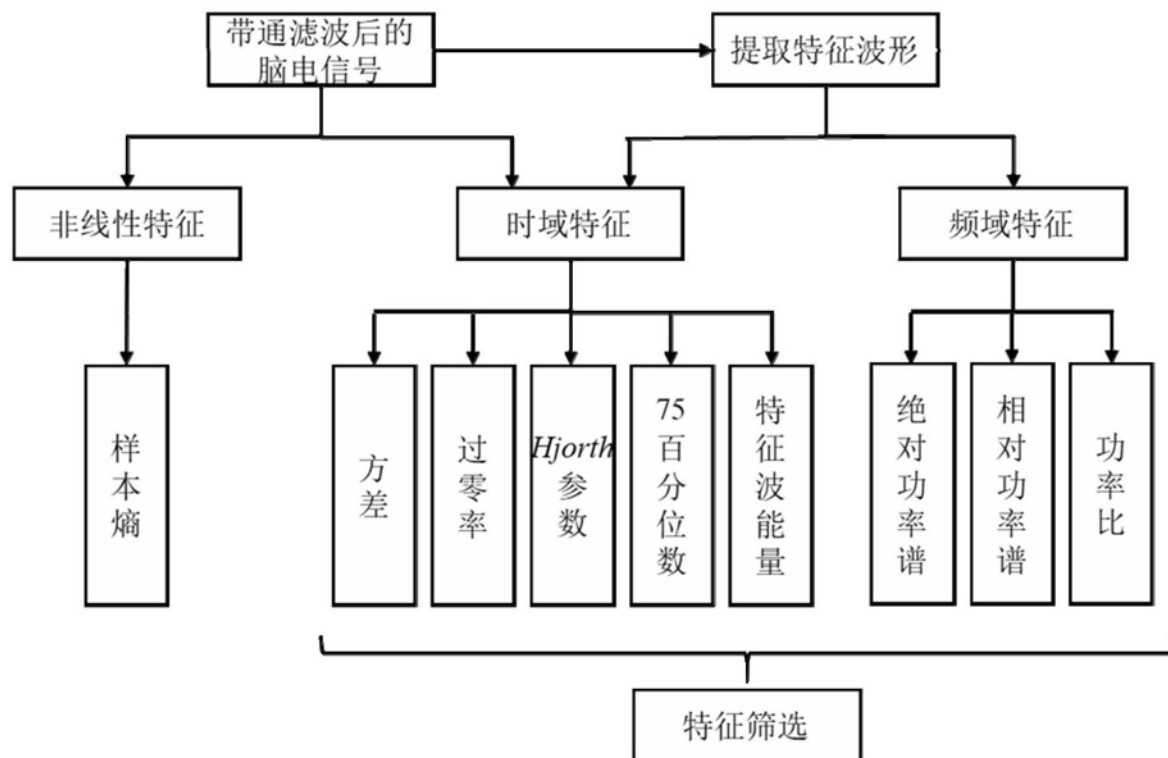


图5

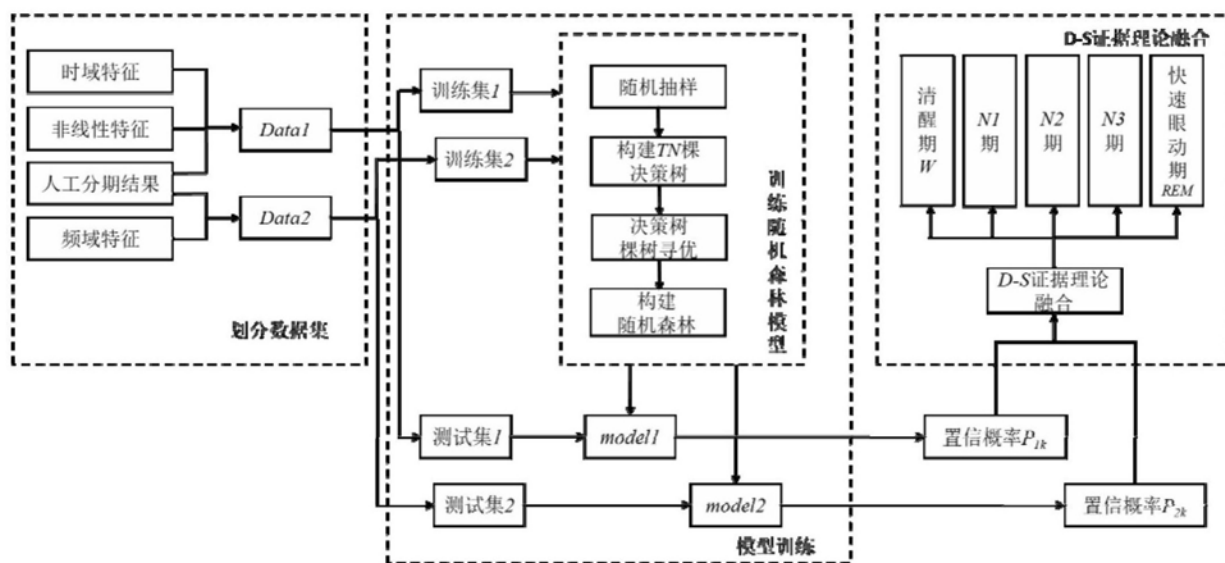


图6

专利名称(译)	一种单导脑电自动睡眠分期方法		
公开(公告)号	CN109674468A	公开(公告)日	2019-04-26
申请号	CN201910088652.8	申请日	2019-01-30
[标]申请(专利权)人(译)	大连理工大学		
申请(专利权)人(译)	大连理工大学		
当前申请(专利权)人(译)	大连理工大学		
[标]发明人	刘蓉 梁洪宇 王永轩		
发明人	刘蓉 梁洪宇 王永轩		
IPC分类号	A61B5/0476 A61B5/00		
CPC分类号	A61B5/0476 A61B5/4809 A61B5/4812 A61B5/4815 A61B5/7203 A61B5/7225 A61B5/725		
代理人(译)	徐雪莲		
外部链接	Espacenet SIPO		

摘要(译)

本发明属于睡眠监测技术领域，一种单导脑电自动睡眠分期方法，包括以下步骤：(1)信号预处理，(2)分类特征提取，(3)睡眠分期。本发明具有以下优点：一是设计了预处理算法，获得质量较好的单导脑电信号；二是本发明从时域、频域、非线性领域提取出多个特征，筛选出具有代表性的特征；三是本发明采用随机森林模型，所述方法不必担心过度拟合、具有很好的抗噪声能力，并且能够得出每一棵决策树的分期结果，从而得到随机森林对每一个睡眠分期的置信概率；四是本发明结合D-S证据理论进一步提高睡眠分期的准确率。

