



(12)发明专利申请

(10)申请公布号 CN 110558977 A

(43)申请公布日 2019.12.13

(21)申请号 201910848404.9

(22)申请日 2019.09.09

(71)申请人 西北大学

地址 710069 陕西省西安市碑林区太白北路229号

(72)发明人 王海 张侃 高岭 郭红波 郑勇
赵悦蓉 郑杰 杨旭东

(74)专利代理机构 西安西达专利代理有限责任公司 61202

代理人 刘华

(51)Int.Cl.

A61B 5/0476(2006.01)

A61B 5/00(2006.01)

权利要求书2页 说明书9页 附图2页

(54)发明名称

基于机器学习模糊特征选择的癫痫发作脑电信号分类方法

(57)摘要

一种基于机器学习模糊特征选择的癫痫发作脑电信号分类方法,步骤一、对原始的EEG数据进行预处理,先进行SWT+ICA再进行带通滤波器处理;步骤二、对预处理后的数据依照已有研究支撑进行特定通道选择;步骤三、对所选出的通道上的数据进行多种特征提取;步骤四、计算每个实验对于不同类的隶属度,依次通过FDM算法和特征子集的约简算法进行模糊特征选择;步骤五、采用优化的SVM进行癫痫信号分类。通过本发明,用户可以获得较高的癫痫分类准确率,另一方面,本发明提出的模糊特征选择思想也可以应用到脑电信号分类的其他领域,比如想象运动脑电信号分类,为基于EEG信号的分类和识别提供了借鉴。



1. 一种基于机器学习模糊特征选择的癫痫发作脑电信号分类方法,不同的脑电信号预处理方式、特定通道选择、多种特征提取、模糊特征选择、SVM支持向量机分类,其特征在于,包括以下步骤:

(1) 对原始数据预处理;使用MATLAB EEG LAB工具包将所有物理生物数据转换为MATLAB数据:

1) 采用平稳小波变换SWT和独立分量分析ICA相结合的方法SWT+ICA对脑电信号和EOG信号进行预处理,去除脑电信号中一些不需要的分量;

2) 利用EMD技术,从预处理的脑电图信号中提取并检测人眼眨眼;

3) 使用带通滤波器0.5-30Hz对数据进行滤波来完成,去除人工伪影使用的滤波器是用5阶巴特沃斯设计生成的,并使用正反向方法实现以生成零相位响应;

4) 去除伪影后所有纯通道数据都使用有限脉冲响应滤波器FIR;

(2) 对预处理后的数据进行特定通道选择;

1) 为了降低系统复杂度,选择了准确度高的四个通道,分别是(F3-C3), (C3-P3), (F4-C4), (C4-P4)这四个通道;

2) 对该4个通道进行平均,使用式(1)对4个通道平均值 $Averged_{EEG}[n]$ 的值进行特征提取;

$$Averged_{EEG}[n] = \frac{1}{c} \left(\sum_{i=1}^c x_i[n] \right) \quad (1);$$

其中, x, c 分别为输入的信号,通道数量;

(3) 对所选出的通道上的数据进行多种特征提取;

1) 采用自回归建模AR技术,使用Burg AR估计器,最大限度地减小了正向和反向预测误差,产生一个稳定的模型,直接计算反射系数,Burg方法可以估计PSD的数据记录,复制原始数据值,使用公式(2)实现产生一个稳定的模型,计算反射系数 $P_{xx}^{BU}(f)$;

$$P_{xx}^{BU}(f) = \frac{\tilde{E}_p}{\left| 1 + \sum_{k=1}^p \tilde{a}_p(k) e^{-j2\pi fk} \right|^2} \quad (2);$$

其中, \tilde{E}_p, k, f 分别表示前向预测误差功率,模型阶次,频率轴参数;

2) 使用AR建模方法从预处理的脑电图数据中提取的一些特征包括:与PSD相关的四个特征:PSD的峰值频率、PSD的峰值、一阶谱矩、二阶谱矩,以下公式(3)和公式(4)分别为计算一阶谱矩 m_{01} 和二阶谱矩 m_{02} ;

$$m_{01} = \sum_{f=1}^N f \cdot PSD_f \quad (3);$$

$$m_{02} = \sum_{f=1}^N (f - m_{01})^2 \cdot PSD_f \quad (4);$$

3) 计算基于连通度的特征提取和熵的特征,其中基于连通度的特征提取要依照公式(5)和公式(6)计算左半球通道信息 $Left_{hemisphericEEG}[n]$,右半球通道信息 $Right_{hemisphericEEG}$

[n];

$$Left_{hemisphericEEG}[n] = \frac{1}{L} \left(\sum_{i=1}^L x_i[n] \right) \quad (5);$$

$$Right_{hemisphericEEG}[n] = \frac{1}{R} \left(\sum_{i=1}^R x_i[n] \right) \quad (6);$$

其中,L=2为左半脑通道数和L ∈ (F3-C3), (C3-P3);

R=2是右半脑通道的数量和R ∈ (F4-C4), (C4-P4);

4) 通过计算,得出从时域、频域、小波域提取的特征;

(4) 进行模糊特征选择;

1) 使用高斯隶属函数对数据集进行模糊化,对提取的特征数据集输入 x_t 的每一列计算基于类的标准偏差(σ_1^2 & σ_2^2)和均值(μ_1^2 & μ_2^2), t 是给定数据集中特征的总数;

2) 把得到的值带入式(7)中获取类标签保持清晰的模糊数据集,新的模糊数据集包含模糊特征(ActualFeatureNum*classNum) ActualFeatureNum表示之前模糊化数据集中特征的实际数量,classNum表示决策类的数量,其中数据集中标签1代表癫痫,2代表非癫痫;把得到的值带入式(7)中获取类标签保持清晰的模糊数据集;

$$f(\vec{X}; \sigma_i, c_i) = e^{\left(\frac{-(x_j - c_i)^2}{2\sigma_i^2} \right)} \quad (7);$$

其中, $\vec{X} = \{x_j\}$ 中 $j = \{1, 2, \dots, \text{实例的数量}\}$, σ_i 和 c_i 是离散的特征;

3) 获取类标签保持清晰的模糊数据集,使用FDM算法生成模糊可辨识矩阵,依据决策相关原理进行计算,模糊可分辨矩阵是对称矩阵,只需计算矩阵的一半,不包括对角元素;

4) 进行特征子集的约简计算,模糊可辨矩阵的每个单元FDM(i, j)是一个不同值的向量称为“可辨向量”,对所有的可分辨向量进行求和,并将得到的可分辨向量除以 $(m^2 - m) / 2$,按降序对可分辨向量进行排序,并使用用户输入选择前K个信息量最大的特性作为约简;

(5) 使用贝叶斯优化算法对支持向量机的超参数进行优化后分类;

1) 采用贝叶斯优化算法对支持向量机的超参数进行优化,给定一组实例标签对的训练集 (x_i, y_i) , $i = 1, 2, \dots, l$ 其中 $x_i \in R^n, y \in (1, -1)^1$;

$$\min_{\omega, b, \epsilon} \frac{1}{2} W^T W + C \sum_{i=1}^l \epsilon_i \quad (8);$$

2) $y_i (W^T \phi(x_i) + b) \geq 1 - \epsilon_i, \epsilon_i \geq 0$,训练向量 x_i 通过函数 ϕ 映射到一个更高的维空间, $C > 0$ 为误差项的惩罚参数;

3) $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$,核函数将不可分离的特征转化为可分离的特征空间,从而提高了识别的精度,径向基函数(RBF)核函数为式(9);

$$K(x_i, x_j) = \exp \left(\frac{-\|x_i - x_j\|^2}{2\sigma^2} \right), \sigma > 0 \quad (9)。$$

基于机器学习模糊特征选择的癫痫发作脑电信号分类方法

技术领域

[0001] 本发明涉及脑电信号分类,具体涉及一种基于机器学习模糊特征选择的癫痫发作脑电信号分类方法。

背景技术

[0002] 癫痫是一种脑神经疾病,常见于发作期和发作期间。它是一种常见的神经系统疾病,影响所有年龄段的人。癫痫是影响全世界约1%人口的最常见的神经疾病之一,其中会有约0.2%的人丧生。一份报告显示,全世界约有5000万人患有癫痫症,每200名新生婴儿中就有1名面临某种形式的癫痫发作。不同的诊断工具,如磁共振成像(MRI)、计算机断层扫描(CT)、正电子发射断层扫描(PET)、超声波、脑电图(EEG)。其中MRI,CT和超声检查费用昂贵,不能用于长期评估。脑电图是一种低成本的工具和非侵入性工具,可用于长期评估。因此,脑电图是诊断癫痫最有用的工具。癫痫发作其特征在于脑结构中神经元活动的异常同步放电。可以通过放置在大脑头皮上的多通道脑电图(EEG)信号记录这些异常电活动。通常,从这个EEG设备记录的这些信号由神经学家解释,这需要它们的可用性并且特别是对于长持续时间的信号而言非常耗时。

[0003] 传统的机器学习进行的癫痫脑电信号分类方法,特征选择过程在保持较高的分类精度的同时,又要在处理较低的计算时间之间权衡。目前实现了多种特征提取技术,并进行了比较。我们注意到,使用所有特性可能并不总是能够确保高精度,反而增加了总体计算开销。可辨识矩阵(DM)在脑电图信号特征选择和比较的结果与其他广泛使用的方法,如主成分分析(PCA)相比较,我们观察到DM优于PCA,甚至在一些案例中提供了一个比实际的特性更好的精度。然而在决策分类系统中,传统的可辨识矩阵(DM)基于决策相对原理工作。当且仅当两个实例的决策类不同时,它才会根据特征比较每对实例,所以影响了分类的准确度和效率。

发明内容

[0004] 为了克服上述现有技术的不足,本发明的目的是提供一种基于机器学习模糊特征选择的癫痫发作脑电信号分类方法,采用SWT+ICA和带通滤波器相结合的方式对数据进行预处理,并选择特定通道,结合自回归建模(AR)技术提取多种特征,进行模糊特征选择、分类,最终形成一种精度高、低计算负荷的癫痫分类方法。

[0005] 为了实现上述目的,本发明采用的技术方案是:

[0006] 一种基于机器学习模糊特征选择的癫痫发作脑电信号分类方法,不同的脑电信号预处理方式、特定通道选择、多种特征提取、模糊特征选择、SVM支持向量机分类,包括以下步骤:

[0007] (1) 对原始数据预处理;使用MATLAB EEG LAB工具包将所有物理生物数据转换为MATLAB数据:

[0008] 1) 采用平稳小波变换SWT和独立分量分析ICA相结合的方法SWT+ICA对脑电信号和

EOG信号进行预处理,去除脑电信号中一些不需要的分量;

[0009] 2) 利用EMD技术,从预处理的脑电图信号中提取并检测人眼眨眼;

[0010] 3) 使用带通滤波器0.5-30Hz对数据进行滤波来完成,去除人工伪影使用的滤波器是用5阶巴特沃斯设计生成的,并使用正反向方法实现以生成零相位响应;

[0011] 4) 去除伪影后所有纯通道数据都使用有限脉冲响应滤波器FIR;

[0012] (2) 对预处理后的数据进行特定通道选择;

[0013] 1) 为了降低系统复杂度,选择了准确度高的四个通道,分别是(F3-C3), (C3-P3), (F4-C4), (C4-P4) 这四个通道;

[0014] 2) 对该4个通道进行平均,使用式(1)对4个通道平均值Avergared_{EEG}[n]的值进行特征提取;

$$[0015] \quad \text{Avergared}_{EEG}[n] = \frac{1}{c} \left(\sum_{i=1}^c x_i[n] \right) \quad (1);$$

[0016] 其中,x,c分别为输入的信号,通道数量;

[0017] (3) 对所选出的通道上的数据进行多种特征提取;

[0018] 1) 采用自回归建模AR技术,使用Burg AR估计器,最大限度地减小了正向和反向预测误差,产生一个稳定的模型,直接计算反射系数,Burg方法可以估计PSD的数据记录,复制原始数据值,使用公式(2)实现产生一个稳定的模型,计算反射系数 $P_{xx}^{BU}(f)$;

$$[0019] \quad P_{xx}^{BU}(f) = \frac{\tilde{E}_p}{\left| 1 + \sum_{k=1}^p \tilde{a}_p(k) e^{-j2\pi fk} \right|^2} \quad (2);$$

[0020] 其中, \tilde{E}_p ,k,f分别表示前向预测误差功率,模型阶次,频率轴参数;

[0021] 2) 使用AR建模方法从预处理的脑电图数据中提取的一些特征包括:与PSD相关的四个特征:PSD的峰值频率、PSD的峰值、一阶谱矩、二阶谱矩,以下公式(3)和公式(4)分别为计算一阶谱矩 m_{01} 和二阶谱矩 m_{02} ;

$$[0022] \quad m_{01} = \sum_{f=1}^N f \cdot PSD_f \quad (3);$$

$$[0023] \quad m_{02} = \sum_{f=1}^N (f - m_{01})^2 \cdot PSD_f \quad (4);$$

[0024] 3) 计算基于连通度的特征提取和熵的特征,其中基于连通度的特征提取要依照公式(5)和公式(6)计算左半球通道信息Left_{hemisphericEEG}[n],右半球通道信息Right_{hemisphericEEG}[n];

$$[0025] \quad \text{Left}_{hemisphericEEG}[n] = \frac{1}{L} \left(\sum_{i=1}^L x_i[n] \right) \quad (5);$$

$$[0026] \quad \text{Right}_{\text{hemisphericEEG}}[n] = \frac{1}{R} \left(\sum_{i=1}^R x_i[n] \right) \quad (6);$$

[0027] 其中, L=2为左半脑通道数和L∈(F3-C3), (C3-P3);

[0028] R=2是右半脑通道的数量和R∈(F4-C4), (C4-P4);

[0029] 4) 通过计算, 得出从时域、频域、小波域提取的特征;

[0030] (4) 进行模糊特征选择;

[0031] 1) 使用高斯隶属函数对数据集进行模糊化, 对提取的特征数据集输入 x_t 的每一列计算基于类的标准偏差(σ_1^1 & σ_1^2)和均值(μ_1^1 & μ_1^2), t 是给定数据集中特征的总数;

[0032] 2) 把得到的值带入式(7)中获取类标签保持清晰的模糊数据集, 新的模糊数据集包含模糊特征(ActualFeatureNum*classNum) ActualFeatureNum表示之前模糊化数据集中特征的实际数量, classNum表示决策类的数量, 其中数据集中标签1代表癫痫, 2代表非癫痫; 把得到的值带入式(7)中获取类标签保持清晰的模糊数据集;

$$[0033] \quad f(\vec{X}; \sigma_i, c_i) = e^{\left(\frac{-(x_j - c_i)^2}{2\sigma_i^2} \right)} \quad (7);$$

[0034] 其中, $\vec{X} = \{x_j\}$ 中 $j = \{1, 2, \dots, \text{实例的数量}\}$, σ_i 和 c_i 是离散的特征;

[0035] 3) 获取类标签保持清晰的模糊数据集, 使用FDM算法生成模糊可辨识矩阵, 依据决策相关原理进行计算, 模糊可分辨矩阵是对称矩阵, 只需计算矩阵的一半, 不包括对角元素;

[0036] 4) 进行特征子集的约简计算, 模糊可辨矩阵的每个单元FDM(i, j)是一个不同值的向量称为“可辨向量”, 对所有的可分辨向量进行求和, 并将得到的可分辨向量除以 $(m^2 - m) / 2$, 按降序对可分辨向量进行排序, 并使用用户输入选择前K个信息量最大的特性作为约简;

[0037] (5) 使用贝叶斯优化算法对支持向量机的超参数进行优化后分类;

[0038] 1) 采用贝叶斯优化算法对支持向量机的超参数进行优化, 给定一组实例标签对的训练集 (x_i, y_i) , $i = 1, 2, \dots, l$ 其中 $x_i \in \mathbb{R}^n, y \in (1, -1)^1$;

$$[0039] \quad \min_{\theta, b, \epsilon} \frac{1}{2} W^T W + C \sum_{i=1}^l \epsilon_i \quad (8);$$

[0040] 2) $y_i (W^T \phi(x_i) + b) \geq 1 - \epsilon_i, \epsilon_i \geq 0$, 训练向量 x_i 通过函数 ϕ 映射到一个更高的维空间, $C > 0$ 为误差项的惩罚参数;

[0041] 3) $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$, 核函数将不可分离的特征转化为可分离的特征空间, 从而提高了识别的精度, 径向基函数(RBF)核函数为式(9);

$$[0042] \quad K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2} \right), \sigma > 0 \quad (9)。$$

附图说明

[0043] 图1为本发明的脑电信号采集系统;

[0044] 图2为本发明的流程图；

[0045] 图3为本发明的左右半脑的通道信息。

具体实施方式

[0046] 以下结合附图对本发明进一步叙述。

[0047] 如图1、2、3所示，一种基于机器学习模糊特征选择的癫痫发作脑电信号分类方法，不同的脑电信号预处理方式、特定通道选择、多种特征提取、模糊特征选择、SVM支持向量机分类，包括以下步骤：

[0048] (1) 对原始数据预处理；使用MATLAB EEG LAB工具包将所有物理生物数据转换为MATLAB数据：

[0049] 1) 采用平稳小波变换SWT和独立分量分析ICA相结合的方法SWT+ICA对脑电信号和EOG信号进行预处理，去除脑电信号中一些不需要的分量；

[0050] 2) 利用EMD技术，从预处理的脑电图信号中提取并检测人眼眨眼；

[0051] 3) 使用带通滤波器0.5-30Hz对数据进行滤波来完成，去除人工伪影使用的滤波器是用5阶巴特沃斯设计生成的，并使用正反向方法实现以生成零相位响应；

[0052] 4) 去除伪影后所有纯通道数据都使用有限脉冲响应滤波器FIR；

[0053] (2) 对预处理后的数据进行特定通道选择；

[0054] 1) 为了降低系统复杂度，选择了准确度高的四个通道，分别是(F3-C3)，(C3-P3)，(F4-C4)，(C4-P4)这四个通道；

[0055] 2) 对该4个通道进行平均，使用式(1)对4个通道平均值Avergared_{EEG}[n]的值进行特征提取；

$$[0056] \quad \text{Avergared}_{EEG}[n] = \frac{1}{c} \left(\sum_{i=1}^c x_i[n] \right) \quad (1);$$

[0057] (3) 对所选出的通道上的数据进行多种特征提取；

[0058] 1) 采用自回归建模AR技术，使用Burg AR估计器，最大限度地减小了正向和反向预测误差，产生一个稳定的模型，直接计算反射系数，Burg方法可以估计PSD的数据记录，复制原始数据值，使用公式(2)实现产生一个稳定的模型，计算反射系数 $P_{xx}^{BU}(f)$ ；

$$[0059] \quad P_{xx}^{BU}(f) = \frac{\tilde{E}_p}{\left| 1 + \sum_{k=1}^p \tilde{a}_p(k) e^{-j2\pi fk} \right|^2} \quad (2);$$

[0060] 2) 使用AR建模方法从预处理的脑电图数据中提取的一些特征包括：与PSD相关的四个特征：PSD的峰值频率、PSD的峰值、一阶谱矩、二阶谱矩，以下公式(3)和公式(4)分别为计算一阶谱矩 m_{01} 和二阶谱矩 m_{02} ；

$$[0061] \quad m_{01} = \sum_{f=1}^N f \cdot PSD_f \quad (3);$$

$$[0062] \quad m_{02} = \sum_{f=1}^N (f - m_{01})^2 \cdot PSD_f \quad (4);$$

[0063] 3) 计算基于连通度的特征提取和熵的特征,其中基于连通度的特征提取要依照公式(5)和公式(6)计算左半球通道信息 $Left_{hemisphericEEG}[n]$,右半球通道信息 $Right_{hemisphericEEG}[n]$;

$$[0064] \quad Left_{hemisphericEEG}[n] = \frac{1}{L} \left(\sum_{i=1}^L x_i[n] \right) \quad (5);$$

$$[0065] \quad Right_{hemisphericEEG}[n] = \frac{1}{R} \left(\sum_{i=1}^R x_i[n] \right) \quad (6);$$

[0066] 其中,L=2为左半球通道数和 $L \in (F3-C3), (C3-P3)$;

[0067] R=2是右半球通道的数量和 $R \in (F4-C4), (C4-P4)$;

[0068] 4) 通过计算,得出从时域、频域、小波域提取的特征;

[0069] (4) 进行模糊特征选择;

[0070] 1) 使用高斯隶属函数对数据集进行模糊化,对提取的特征数据集输入 x_t 的每一列计算基于类的标准偏差(σ_i^1 & σ_i^2)和均值(μ_i^1 & μ_i^2),t是给定数据集中特征的总数;

[0071] 2) 把得到的值带入式(7)中获取类标签保持清晰的模糊数据集,新的模糊数据集包含模糊特征(ActualFeatureNum*classNum) ActualFeatureNum表示之前模糊化数据集中特征的实际数量,classNum表示决策类的数量,其中数据集中标签1代表癫痫,2代表非癫痫;把得到的值带入式(7)中获取类标签保持清晰的模糊数据集。

$$[0072] \quad f(\bar{X}; \sigma_i, c_i) = e^{\left(\frac{-(x_j - c_i)^2}{2\sigma_i^2} \right)} \quad (7);$$

[0073] 3) 获取类标签保持清晰的模糊数据集,使用FDM算法生成模糊可辨识矩阵,依据决策相关原理进行计算,模糊可分辨矩阵是对称矩阵,只需计算矩阵的一半,不包括对角元素;

[0074] 4) 进行特征子集的约简计算,模糊可辨矩阵的每个单元FDM(i, j)是一个不同值的向量称为“可辨向量”,对所有的可分辨向量进行求和,并将得到的可分辨向量除以 $(m^2-m)/2$,按降序对可分辨向量进行排序,并使用用户输入选择前K个信息量最大的特性作为约简;

[0075] (5) 使用贝叶斯优化算法对支持向量机的超参数进行优化后分类;

[0076] 1) 采用贝叶斯优化算法对支持向量机的超参数进行优化,给定一组实例标签对的训练集 $(x_i, y_i), i=1, 2, \dots, l$ 其中 $x_i \in R^n, y \in (1, -1)^1$;

$$[0077] \quad \min_{\omega, b, \epsilon} \frac{1}{2} W^T W + C \sum_{i=1}^l \epsilon_i \quad (8);$$

[0078] 2) $y_i (W^T \phi(x_i) + b) \geq 1 - \epsilon_i, \epsilon_i \geq 0$,训练向量 x_i 通过函数 ϕ 映射到一个更高的维空间,C>0为误差项的惩罚参数;

[0079] 3) $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$,核函数将不可分离的特征转化为可分离的特征

空间,从而提高了识别的精度,径向基函数(RBF)核函数为式(9);

$$[0080] \quad K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right), \sigma > 0 \quad (9)。$$

[0081] 实施例

[0082] 步骤1:对原始EEG数据进行预处理,具体包括以下步骤:

[0083] (1)在本实施例中采用两种不同类型的数据集:CHB-MIT数据集和德国Bonn大学癫痫研究中的公开数据集。CHB-MIT数据集它包括来自顽固性癫痫儿童受试者的多通道686头皮脑电图记录。该数据集包含23名癫痫患者的数据,其中18名女性,5名男性,年龄在10到22岁之间。每次癫痫的开始和结束都在一个单独的文件中进行了注释。Bonn数据集由5个健康人和5个癫痫病患者的脑电数据构成,共包含5个数据集(Z、O、N、F、S)。Z、O:五个健康志愿者睁眼和闭眼时的脑电信号;N、F:癫痫患者发作间歇期海马结构处N和病灶区F的脑电;S:癫痫患者发作期病灶处的脑电。

[0084] (2)预处理:使用MATLAB EEG LAB工具包将所有物理生物数据转换为MATLAB数据,再采用平稳小波变换(SWT)和独立分量分析(ICA)相结合的方法(SWT+ICA),利用EMD技术,从预处理的Boon脑电图信号中提取并检测人眼眨眼,使用带通滤波器(0.5-30Hz)对数据进行滤波。

[0085] 步骤2:将预处理后的数据进行特定通道选择,研究表明C3-C4通道在癫痫发作分类中提供了90.77%的高准确度,大脑的P3和P4位置更容易癫痫发作,并且它位于脑血管流域区域,这是脑损伤的高风险,因此选择(F3-C3), (C3-P3), (F4-C4), (C4-P4)这四个通道,对该4个通道进行平均,得到Avergared_{EEG}[n]的值。

[0086] (3)对Avergared_{EEG}[n]的值进行特征提取。具体来说,采用自回归建模(AR)技术,使用Burg AR估计器产生一个稳定的模型,计算反射系数。提取出PSD的峰值频率、PSD的峰值、一阶谱矩、二阶谱矩。计算基于连通度的特征提取和熵的特征,其中基于连通度的特征提取分别计算左右半脑的信息。最后通过计算得出从时域、频域、小波域提取的特征。

[0087] (4)把提取出来的多种特征进行模糊特征选择。首先提取的特征数据集输入 x_t 的每一列计算基于类的标准偏差(σ_1^1 & σ_1^2)和均值(μ_1^1 & μ_1^2)。接着把得到的值通过计算获取类标签保持清晰的模糊数据集。然后使用FDM算法生成模糊可辨识矩阵:

[0088] 算法伪代码如下:

Fuzzy Discernibility Matrix (FDM) Algorithm:

Input: A ← Conditional Attributes

D ← Decision Attributes

Initialize: DA ← {A, D}

Initialize: FDM ← { Φ } #empty fuzzy discernibility matrix

Initialize: classNum ← uniqueClass (D)

[0089]

FD ← fuzzification(DA) #fuzzification of feature

extracted dataset

[Row, Column] ← sizeof (DA)

for i ← 1 to Row do

for j ← 1 to Row do

if i > j

```

then tempset ← {} #computed only the lower half of
FDM

if DA(i, Column) ≠ DA(j, Column) then
    for z ← 1 to (Column - 1) do
        p ← classNum
        t ← DissM*(FD(i, P), FD(j, P))

#featurewise dissimilarity
[0090]         tempset ← tempset ∪ t

    end for

    FDM(i, j) ← tempset

    end if

    end if

    end for

    end for

return FDM

```

[0091] 最后进行特征子集的约简计算；

[0092] 用于特征子集的约简计算：模糊可辨矩阵的每个单元 $FDM(i, j)$ 是一个不同值的向量称为“可辨向量”，对所有的可分辨向量进行求和，并将得到的可分辨向量除以 $(m^2 - m) / 2$ ，按降序对可分辨向量进行排序，并使用用户输入选择前 K 个信息量最大的特性作为约简，以下为约简计算的伪代码：

[0093] Algorithm: Reduct

Inputs:A← Conditional Attributes

D← Descision Attributes

Inputs:FDM, n*n discernibility matrix

(where n is total no.of instances)

Inputs:s, vector of lengthof(A)

Initialize:s ← {0}

Initialize:Reduct ← {} #selected subset of features

[0094] $s \leftarrow \left(\sum_{\forall i, j \in (n^2 - n) / 2} FDM(i, j) / (n^2 - n) / 2 \right)$

where FDM(i, j) is a vector of dissimilar indices

with size equals to lengthof(A)

K ← userinput(<<lengthof(A)) #k- best discriminating
features

index ← sort(s, descend) #sort discernibility vector in
descending

Reduct ← A(index(K))

return Reduct

[0095] (5) 采用贝叶斯优化算法对支持向量机的超参数进行优化, 利用SVM进行分类, 该分类器针对最佳结果进行了优化。

[0096] (6) 统计准确率, 得出最后的结果。

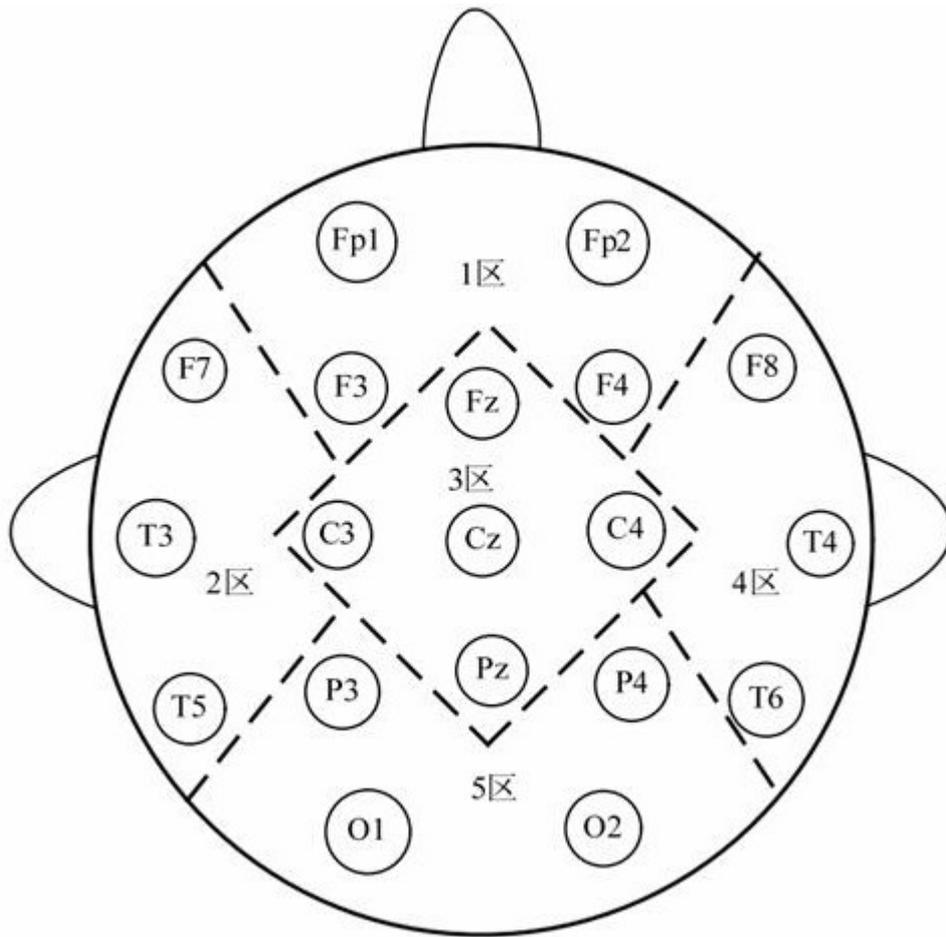


图1

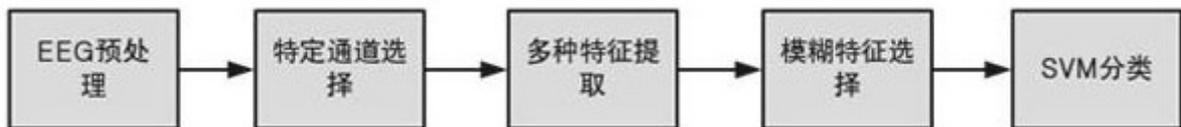


图2

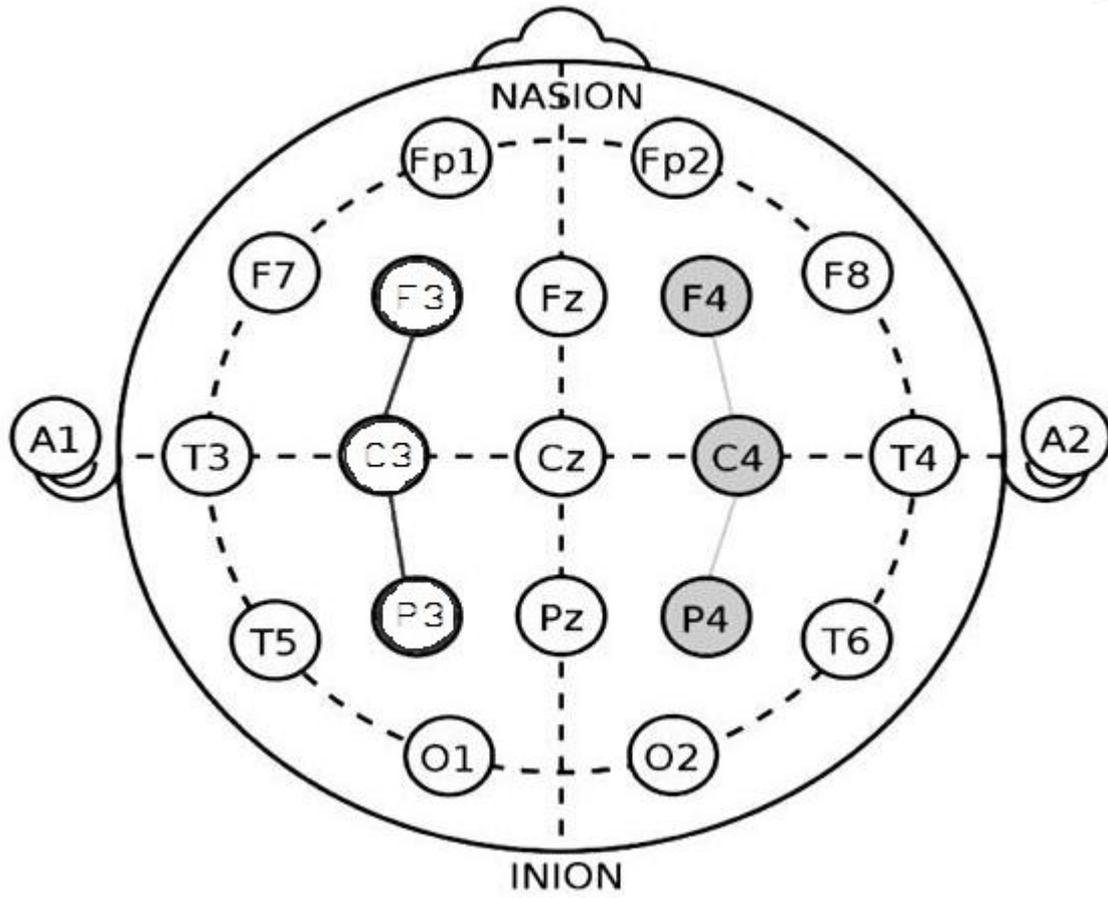


图3

专利名称(译)	基于机器学习模糊特征选择的癫痫发作脑电信号分类方法		
公开(公告)号	CN110558977A	公开(公告)日	2019-12-13
申请号	CN201910848404.9	申请日	2019-09-09
申请(专利权)人(译)	西北大学		
当前申请(专利权)人(译)	西北大学		
[标]发明人	王海 张侃 高岭 郭红波 郑勇 赵悦蓉 郑杰 杨旭东		
发明人	王海 张侃 高岭 郭红波 郑勇 赵悦蓉 郑杰 杨旭东		
IPC分类号	A61B5/0476 A61B5/00		
CPC分类号	A61B5/0476 A61B5/4094 A61B5/7203 A61B5/725 A61B5/7253 A61B5/7267		
代理人(译)	刘华		
外部链接	Espacenet SIPO		

摘要(译)

一种基于机器学习模糊特征选择的癫痫发作脑电信号分类方法，步骤一、对原始的EEG数据进行预处理，先进行SWT+ICA再进行带通滤波器处理；步骤二、对预处理后的数据依照已有研究支撑进行特定通道选择；步骤三、对所选出的通道上的数据进行多种特征提取；步骤四、计算每个实验对于不同类的隶属度，依次通过FDM算法和特征子集的约简算法进行模糊特征选择；步骤五、采用优化的SVM进行癫痫信号分类。通过本发明，用户可以获得较高的癫痫分类准确率，另一方面，本发明提出的模糊特征选择思想也可以应用到脑电信号分类的其他领域，比如想象运动脑信号分类，为基于EEG信号的分类和识别提供了借鉴。

