



(12)发明专利申请

(10)申请公布号 CN 109512442 A

(43)申请公布日 2019. 03. 26

(21)申请号 201811571326.4

(22)申请日 2018.12.21

(71)申请人 杭州电子科技大学

地址 310018 浙江省杭州市下沙高教园区2
号大街

(72)发明人 曾虹 杨晨 孔万增 吴振华
张佳明

(74)专利代理机构 杭州君度专利代理事务所
(特殊普通合伙) 33240

代理人 朱月芬

(51)Int.Cl.

A61B 5/18(2006.01)

A61B 5/0476(2006.01)

A61B 5/00(2006.01)

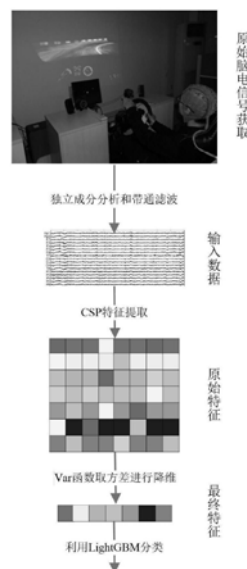
权利要求书2页 说明书5页 附图2页

(54)发明名称

一种基于LightGBM的EEG疲劳状态分类方法

(57)摘要

本发明公开了一种基于LightGBM的EEG疲劳状态分类方法。本发明以共空间模式作为特征提取方法,梯度boosting框架LightGBM作为分类器,通过对EEG信号的分析,对驾驶员的疲劳程度进行分类,实现疲劳、清醒、中性三种状态的区分。首先获取数据并预处理;其次通过CSP对脑电数据进行特征提取,后对脑电特征进行降维;然后划定实验的训练集和测试集,再构建分类模型分类。针对精神状态预测,本发明获得的结果更好,且从时间消耗角度,本发明拥有较快的运行速度,这对后期应用于实时数据分析提供了基础。总之,本发明在精神状态预测方面具有较好的性能,以期在实际的脑机交互中有着广泛的应用前景。



1.一种基于LightGBM的EEG疲劳状态分类方法,其特征在于以共空间模式作为特征提取方法,梯度boosting框架作为分类器,通过对EEG信号的分析,对驾驶员的疲劳程度进行分类,实现疲劳、清醒、中性三种状态的区分,具体实现包括如下步骤:

步骤1.获取数据:

搭建模拟驾驶平台:设计模拟驾驶实验,还原驾驶过程场景;同时选取多名被试驾驶员,并让他们分别进行模拟驾驶实验,采集多名被试驾驶员在模拟驾驶实验中的原始脑电信号I、眼电信号I和心电信号I;

根据现有研究对眼电信号I和心电信号I进行分析,得出被试驾驶员的心率和每分钟的眨眼次数;以此为依据,为不同时间段内的原始脑电信号I打上“疲劳”、“清醒”、“中性”这三种状态的标签;

步骤2.数据预处理:

对采集到的原始脑电信号I均进行数据预处理,获取处理后的脑电信号II,预处理包括独立成分分析和带通滤波;

步骤3.通过CSP对脑电数据进行特征提取

针对脑电信号II,每个样本表示为 $X \times S$ 的矩阵W;其中,X是通道数,S是每一个通道的采样点数;规则化空间协方差如式(1)所示:

$$C = \frac{WW^T}{\text{trace}(WW^T)} \quad (1)$$

其中,trace()表示矩阵的对角元素之和;为了将两类方差分开,通过对训练数据中的两类样本的协方差之和进行平均,分别得到各自的平均协方差 C_d 和 C_t ,进而得到混合的空间协方差: $C_c = C_d + C_t$,将混合的空间协方差 C_c 分解为 $C_c = E_c \lambda_c E_c^T$ 形式,其中 E_c 是协方差矩阵的特征向量, λ_c 是特征值构成的对角阵;将特征值进行降序排列,按式(2)进行白化变换后获得P:

$$P = \sqrt{\lambda_c^{-1}} E_c^T \quad (2)$$

根据 PC_cP^T 对应的特征值为1,对 C_d 和 C_t 进行如下变换: $S_d = PC_dP^T$, $S_t = PC_tP^T$,则 S_d 和 S_t 具有共同的特征向量,当 $S_d = B\lambda_d B^T$ 时,有 $S_t = B\lambda_t B^T$, $\lambda_d + \lambda_t = I$;其中I是单位向量矩阵;因为对应的两个特征值之和总是1,所以当特征向量B对于 S_d 有最大的特征值时,对于 S_t 有最小的特征值;由此,能够得到投影矩阵:

$$P_M = (B^T P)^T \quad (3);$$

将疲劳和中性状态分别和清醒状态做投影,求得投影矩阵 P_A 和 P_B ,最终的投影矩阵为: $P_N = P_A + P_B$ (4)

对所有的实验样本,按照式(4)获取投影矩阵,得到所需的脑电特征F;

$$F = P_N W \quad (5)$$

步骤4.对脑电特征进行降维

使用方差var函数,对于每个实验样本,计算各个通道中数据的方差,对脑电特征进行降维;

步骤5.划定实验的训练集和测试集

针对每个被试驾驶员,将其所有脑电特征随机打散,并从中抽取80%作为训练集,记作Train_i,剩下的20%作为测试集,记作为Test_i,其中i表示第i个被试;为了避免不同类别样本测试和训练数据比例不同所带来的误差,对于三种状态的样本,抽取的训练和测试数据严格遵循4:1的比例;

从每一个被试被试驾驶员的所有脑电特征中各随机抽取出80%的数据,并将这些数据组合成跨被试的疲劳状态分类训练集,剩下的20%脑电特征组合成跨被试的疲劳状态分类测试集;

步骤6. 分类

LightGBM中的主要参数包括num_leaves、num_trees和learning_rate;其中num_trees代表生成树的总数目,而num_leaves代表每棵生成树上叶子的数目;

构建一个基于LightGBM的分类模型,将num_leaves设置为63,learning_rate设置为0.01.num_trees则根据测试的结果进行动态调整;

使用训练集训练LightGBM模型,然后利用训练好的模型预测出未被训练的测试集所对应的类别;最后将预测分类结果与这些特征向量的实际类别进行比较,得到疲劳状态的分类准确率。

一种基于LightGBM的EEG疲劳状态分类方法

技术领域

[0001] 本发明属于生物特征识别领域中的脑电信号识别领域,具体涉及一种基于LightGBM的EEG疲劳状态分类方法。

背景技术

[0002] 驾驶疲劳是造成交通事故的重要原因,据美国国家公路交通安全委员会的数据显示:美国每年由于驾驶疲劳引发交通事故而造成的经济损失在125亿美元以上。疲劳并没有明显的症状,但通常表现为嗜睡、劳累或虚弱等。疲劳会导致驾驶员在驾驶过程中警觉性严重下降,使他们犯错从而导致交通事故。研究一种准确率高、时间空间复杂度低的精神状态分类方法,能够为减少驾驶疲劳所引发的交通事故奠定算法基础。

[0003] 目前,针对疲劳检测这一问题,学术界开展了大量的的研究工作,总结下来,主要有以下几个方面:1)基于响应时间和注意力的心理活动测试;2)眼部运动参数的检测,如眼球扫视运动、眨眼率等;3)通过问卷调查方式的主动检测;4)通过生物电信号,如脑电、眼电、肌电以及心电等,进行疲劳驾驶状态检测。除此之外,近年来,转向握柄压力、皮肤电导率、血容量脉冲等指标,也被用于疲劳驾驶检测研究。

[0004] 在上述用于疲劳检测的各种指标中,EEG分析方法被认为是最便捷及有效的。作为一种有效的对神经活动进行间接测量的工具,EEG广泛应用于神经科学,认知科学,认知心理学,神经科学和心理生理学研究中。另一方面,驾驶行为涉及运动、推理、视听觉加工、决策、感知和识别等多种行为,还受到情绪、焦虑和其他许多心理因素的影响,这些与驾驶有关的身体和精神活动都反映在脑电信号中。因此,将脑电EEG信号作为疲劳检测的一个可靠指标是完全可行的。

[0005] 近年来,学术界已经提出了许多使用EEG进行疲劳检测的方法,比如,Correa等人采用时频及小波分析方法,基于EEG信号,针对不同嗜睡阶段,提出了一套自动检测方法检测清醒状态和嗜睡状态,分别获得了87.4%和83.6%的检测准确率。Khushaba等人提出了一种基于模糊互信息小波包变换的特征提取方法用于预测嗜睡状态。Mu等人采用四种不同的熵,包括频谱熵、近似熵、采样熵以及模糊熵,来提取EEG特征用于疲劳驾驶的检测。Wali等人融合了离散小波包变换和快速傅立叶变换对驾驶员在驾驶过程中注意力分散层级进行分类,分类准确率达到85%以上。Fu等人提出了一种基于隐马尔可夫的动态疲劳检测模型,用于估计驾驶员疲劳状态,得到了92.5%的准确率。

[0006] 尽管上述方法已经取得了优秀的表现,但是,针对驾驶员疲劳状态监测这一问题,如何利用合适的模型,通过对EEG信号的分析,获得鲁棒性强、准确率高的检测性能,是目前的一大重要挑战。原因如下:首先,从脑电信号本身考虑,EEG是一种非稳定、随机性强的信号,且随着时间变化,对同一被试或不同被试采集的EEG信号往往具有较大的差异性;其次,EEG信号的低信噪比特征往往会影响到检测的准确率。第三,随着脑电采集设备的不断进步,脑电信号逐渐向特征的多维度和复杂化发展,时间及空间消耗也是需要考虑的一大问题。

发明内容

[0007] 为了克服上述现有技术中存在的准确率低、鲁棒性差、时间空间消耗较大等问题，本发明提供了一种基于LightGBM的EEG疲劳状态分类方法。

[0008] 本发明采用的技术方案是：

[0009] 本发明以共空间模式作为特征提取方法，梯度boosting框架作为分类器，通过对EEG信号的分析，对驾驶员的疲劳程度进行分类，实现疲劳、清醒、中性三种状态的区分，具体实现包括如下步骤：

[0010] 步骤1. 获取数据：

[0011] 搭建模拟驾驶平台：设计模拟驾驶实验，还原驾驶过程场景；同时选取多名被试驾驶员，并让他们分别进行模拟驾驶实验，采集多名被试驾驶员在模拟驾驶实验中的原始脑电信号I、眼电信号I和心电信号I；

[0012] 根据现有研究对眼电信号I和心电信号I进行分析，得出被试驾驶员的心率和每分钟的眨眼次数；以此为依据，为不同时间段内的原始脑电信号I打上“疲劳”、“清醒”、“中性”这三种状态的标签；

[0013] 步骤2. 数据预处理：

[0014] 对采集到的原始脑电信号I均进行数据预处理，获取处理后的脑电信号II，预处理包括独立成分分析和带通滤波；

[0015] 步骤3. 通过CSP对脑电数据进行特征提取

[0016] 针对脑电信号II，每个样本表示为 $X \times S$ 的矩阵W；其中，X是通道数，S是每一个通道的采样点数；规则化空间协方差如式(1)所示：

$$[0017] \quad C = \frac{WW^T}{\text{trace}(WW^T)} \quad (1)$$

[0018] 其中， $\text{trace}()$ 表示矩阵的对角元素之和；为了将两类方差分开，通过对训练数据中的两类样本的协方差之和进行平均，分别得到各自的平均协方差 C_d 和 C_t ，进而得到混合的空间协方差： $C_c = C_d + C_t$ ，将混合的空间协方差 C_c 分解为 $C_c = E_c \lambda_c E_c^T$ 形式，其中 E_c 是协方差矩阵的特征向量， λ_c 是特征值构成的对角阵；将特征值进行降序排列，按式(2)进行白化变换后获得P：

$$[0019] \quad P = \sqrt{\lambda_c^{-1}} E_c^T \quad (2)$$

[0020] 根据 $PC_c P^T$ 对应的特征值为1，对 C_d 和 C_t 进行如下变换： $S_d = PC_d P^T$ ， $S_t = PC_t P^T$ ，则 S_d 和 S_t 具有共同的特征向量，当 $S_d = B \lambda_d B^T$ 时，有 $S_t = B \lambda_t B^T$ ， $\lambda_d + \lambda_t = I$ ；其中I是单位向量矩阵；因为对应的两个特征值之和总是1，所以当特征向量B对于 S_d 有最大的特征值时，对于 S_t 有最小的特征值；由此，能够得到投影矩阵：

$$[0021] \quad P_M = (B^T P)^T \quad (3) ;$$

[0022] 将疲劳和中性状态分别和清醒状态做投影，求得投影矩阵 P_A 和 P_B ，最终的投影矩阵为： $P_N = P_A + P_B$ (4)

[0023] 对所有的实验样本，按照式(4)获取投影矩阵，得到所需的脑电特征F；

$$[0024] \quad F = P_N W \quad (5)$$

[0025] 步骤4. 对脑电特征进行降维

[0026] 使用方差var函数,对于每个实验样本,计算各个通道中数据的方差,对脑电特征进行降维;

[0027] 步骤5.划定实验的训练集和测试集

[0028] 针对每个被试驾驶员,将其所有脑电特征随机打散,并从中抽取80%作为训练集,记作Train_i,剩下的20%作为测试集,记作为Test_i,其中i表示第i个被试;为了避免不同类别样本测试和训练数据比例不同所带来的误差,对于三种状态的样本,抽取的训练和测试数据严格遵循4:1的比例;

[0029] 从每一个被试被试驾驶员的所有脑电特征中各随机抽取出80%的数据,并将这些数据组合成跨被试的疲劳状态分类训练集,剩下的20%脑电特征组合成跨被试的疲劳状态分类测试集;

[0030] 步骤6.分类

[0031] LightGBM中的主要参数包括num_leaves、num_trees和learning_rate;其中num_trees代表生成树的总数目,而num_leaves代表每棵生成树上叶子的数目;

[0032] 构建一个基于LightGBM的分类模型,将num_leaves设置为63,learning_rate设置为0.01.num_trees则根据测试的结果进行动态调整;

[0033] 使用训练集训练LightGBM模型,然后利用训练好的模型预测出未被训练的测试集所对应的类别;最后将预测分类结果与这些特征向量的实际类别进行比较,得到疲劳状态的分类准确率。

[0034] 与现有技术相比,本发明的有益效果是:

[0035] 1) 针对精神状态预测,本发明可以得到比其他分类方法更好的结果;

[0036] 2) 高维度特征的会影响LightGBM的运行速度和内存消耗,但最终分类准确率并不依赖于高维度特征;

[0037] 3) 从时间消耗角度,本发明拥有较快的运行速度,这对后期应用于实时数据分析提供了基础。

[0038] 总之,本发明在精神状态预测方面具有较好的性能,以期在实际的脑机交互中有着广泛的应用前景。

附图说明

[0039] 图1为本发明流程图;

[0040] 图2为脑电通道图。

具体实施方式

[0041] 下面结合附图对本发明的较佳实施例进行详细阐述,以使本发明的优点和特征能更易于被本领域技术人员理解,从而对本发明的保护范围做出更为清楚明确的界定。

[0042] 请参阅图1和图2,本发明实施方式包括如下步骤:

[0043] 步骤1.获取数据:

[0044] 搭建模拟驾驶平台:设计合理的模拟驾驶实验,尽可能地还原驾驶时的场景。同时选取通过多名被试驾驶员,并让他们分别进行模拟驾驶实验,采集多名被试驾驶员在模拟驾驶实验中的原始脑电信号I。同时,被试驾驶员的眼电信号和心电信号也被一并采集。

[0045] 据研究,随着人进入疲劳状态,眼动的次数不断减少,而眨眼的频率却不断增加;另外,疲劳状态下,心率降低。通过对眼电信号和心电信号的分析,可以得出被试驾驶员的心率和每分钟的眨眼次数。以此为依据,为不同时间段内的脑电信号打上“疲劳”、“清醒”、“中性”这三种状态的标签。

[0046] 步骤2.数据预处理:

[0047] 对采集到的各类别的原始脑电信号均进行数据预处理,获取处理后的脑电信号 Π ,预处理包括独立成分分析和带通滤波,目的是减少伪迹的干扰,提高信噪比,从而提高特征提取的准确性。独立成分分析和带通滤波属于本领域技术人员所熟知的常规技术,故不详解。

[0048] 步骤3.通过CSP对脑电数据进行特征提取

[0049] 共空间算法能够找到最优空间投影使两类信号的功率最大,因此它能估计出两个空间滤波器来提取任务相关信号成分,并且同时去除任务不相关成分和噪声。共空间模式使用的方法是基于两个协方差矩阵的同时对角线化。

[0050] 针对脑电信号 Π ,每个样本(trail)可表示为 $X \times S$ 的矩阵 W ,其中, X 是通道数, S 是每一个通道的采样点数。规则化空间协方差如式(1)所示:

$$[0051] \quad C = \frac{WW^T}{\text{trace}(WW^T)} \quad (1)$$

[0052] 其中, $\text{trace}()$ 表示矩阵的对角元素之和。为了将两类方差分开,通过对训练数据中的两类样本的协方差之和进行平均,分别得到各自的平均协方差 C_d 和 C_t ,进而得到混合的空间协方差: $C_c = C_d + C_t$,将空间协方差 C_c 分解为 $C_c = E_c \lambda_c E_c^T$ 形式,其中 E_c 是协方差矩阵的特征向量, λ_c 是特征值构成的对角阵。将特征值进行降序排列,按式(2)进行白化变换后获得 P :

$$[0053] \quad P = \sqrt{\lambda_c^{-1}} E_c^T \quad (2)$$

[0054] 根据 $PC_c P^T$ 对应的特征值为1,对 C_d 和 C_t 进行如下变换: $S_d = PC_d P^T$, $S_t = PC_t P^T$,则 S_d 和 S_t 具有共同的特征向量,当 $S_d = B \lambda_d B^T$ 时,有 $S_t = B \lambda_t B^T$, $\lambda_d + \lambda_t = I$ 。其中 I 是单位向量矩阵。因为对应的两个特征值之和总是1,所以当特征向量 B 对于 S_d 有最大的特征值时,对于 S_t 有最小的特征值。由此,能够得到投影矩阵:

$$[0055] \quad P_M = (B^T P)^T \quad (3)$$

[0056] 由于实验中用到了三种状态,因此通过设计一种针对三分类的CSP特征提取方法来对状态进行特征提取。

[0057] 将疲劳和中性状态分别和清醒状态做投影,求得投影矩阵 P_A 和 P_B ,最终的投影矩阵为: $P_N = P_A + P_B$ (4)

对所有的实验样本(包括训练和测试),按照式(4)获取投影矩阵,得到所需的脑电特征 F 。

$$[0058] \quad F = P_N W \quad (5)$$

[0059] 步骤4.对脑电特征进行降维

[0060] 脑电特征的高维度特性会增大LightGBM训练计算过程中的时间和空间损耗,而且通过我们的实验测试,我们得出一个结论:LightGBM并不像深度学习那样依赖高维度的数据特征,对特征降维后,加快了训练速度,降低了内存消耗,且最终准确率不会发生太大变

化。因此,我们使用方差var函数,对于每个样本,计算各个通道中数据的方差,对脑电特征进行降维。

[0061] 步骤5.划定实验的训练集和测试集

[0062] 针对每个被试驾驶员,将其所有脑电特征随机打散,并从中抽取80%作为训练集,记作Train_i,剩下的20%作为测试集,记作为Test_i,其中 $i=1;2;\cdots;10$ 。i表示第i个被试。为了避免不同类别样本测试和训练数据比例不同所带来的误差,对于这三种状态的样本,我们所抽取的训练和测试数据严格遵循4:1的比例。

[0063] 不同个体间的脑电信号差异很大,而这些差异会影响最终的分类结果。为了进一步验证模型性能,需要用到跨被试的数据集。我们从每一个被试的所有脑电特征中各随机抽取出80%的数据,并将这些数据组合成跨被试的疲劳状态分类训练集,剩下的20%脑电特征组合成跨被试的疲劳状态分类测试集。

[0064] 步骤6.分类

[0065] LightGBM中的主要参数有num_leaves、num_trees和learning_rate等,其中num_trees代表生成树的总数目,而num_leaves代表每棵生成树上叶子的数目。较小的learning_rate和较大的num_trees可以在一定程度上提高最终的准确率,但会增大时间和空间上的开销。

[0066] 构建一个基于LightGBM的分类模型,将num_leaves设置为63,learning_rate设置为0.01.num_trees则根据测试的结果进行动态调整。

[0067] 使用训练集训练LightGBM模型,然后利用训练好的模型预测出未被训练的测试集所对应的类别。最后将预测分类结果与这些特征向量的实际类别进行比较,得到疲劳状态的分类准确率。单个被试的平均分类准确率为95.31%,跨被试的分类准确率为91.67%。

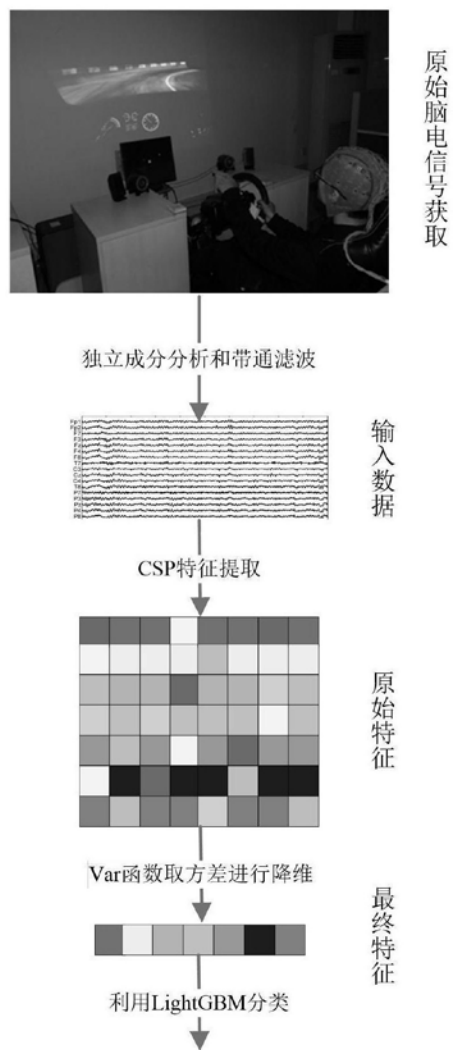


图1

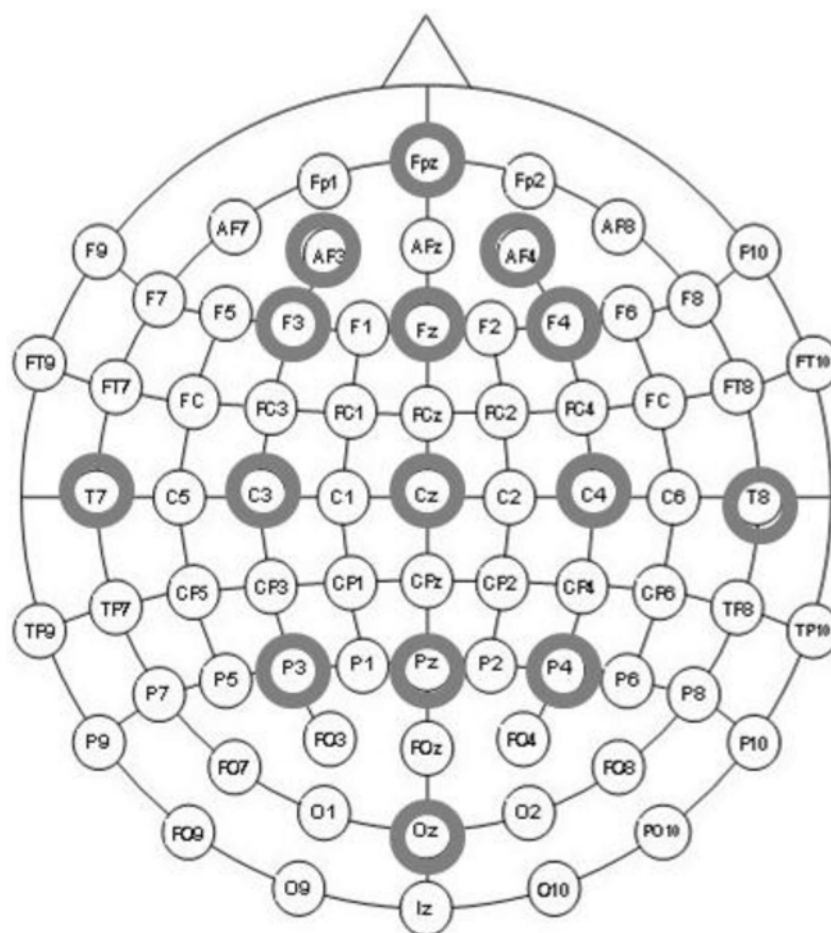


图2

专利名称(译)	一种基于LightGBM的EEG疲劳状态分类方法		
公开(公告)号	CN109512442A	公开(公告)日	2019-03-26
申请号	CN201811571326.4	申请日	2018-12-21
[标]申请(专利权)人(译)	杭州电子科技大学		
申请(专利权)人(译)	杭州电子科技大学		
当前申请(专利权)人(译)	杭州电子科技大学		
[标]发明人	曾虹 杨晨 孔万增 吴振华 张佳明		
发明人	曾虹 杨晨 孔万增 吴振华 张佳明		
IPC分类号	A61B5/18 A61B5/0476 A61B5/00		
CPC分类号	A61B5/18 A61B5/04012 A61B5/0476 A61B5/168 A61B5/7203 A61B5/7235 A61B5/7253 A61B5/7267 A61B5/7275 A61B2503/22		
代理人(译)	朱月芬		
外部链接	Espacenet SIPO		

摘要(译)

本发明公开了一种基于LightGBM的EEG疲劳状态分类方法。本发明以共空间模式作为特征提取方法，梯度boosting框架LightGBM作为分类器，通过对EEG信号的分析，对驾驶员的疲劳程度进行分类，实现疲劳、清醒、中性三种状态的区分。首先获取数据并预处理；其次通过CSP对脑电数据进行特征提取，后对脑电特征进行降维；然后划定实验的训练集和测试集，再构建分类模型分类。针对精神状态预测，本发明获得的结果更好，且从时间消耗角度，本发明拥有较快的运行速度，这对后期应用于实时数据分析提供了基础。总之，本发明在精神状态预测方面具有较好的性能，以期在实际的脑机交互中有着广泛的应用前景。

