



(12)发明专利申请

(10)申请公布号 CN 108742517 A

(43)申请公布日 2018.11.06

(21)申请号 201810259825.3

(22)申请日 2018.03.27

(71)申请人 重庆邮电大学

地址 400065 重庆市南岸区黄桷垭崇文路2号

(72)发明人 王强强 赵德春 王怡 李舒粤

(74)专利代理机构 北京同恒源知识产权代理有限公司 11275

代理人 赵荣之

(51)Int.Cl.

A61B 5/00(2006.01)

A61B 5/04(2006.01)

A61B 5/0476(2006.01)

权利要求书2页 说明书8页 附图2页

(54)发明名称

一种基于Stacking单导联脑电睡眠自动分期方法

(57)摘要

本发明涉及一种基于Stacking单导联脑电睡眠自动分期方法,属于机器学习算法领域。该方法包括步骤:S1:睡眠脑电信号预处理;S2:睡眠脑电多特征提取和筛选;S3:机器学习分类;S4:睡眠自动分期。本发明能够获取自适应阈值的小波函数和IIR滤波函数结合的滤波方法对脑电信号进行降噪处理,有效地提高信号的信噪比;能够对特征算法进行优化筛选获取一种新的特征参数组,作为睡眠分期的特征;能够获取一种新的准确率高的多特征与集成学习算法组合作为睡眠分期方法。



1. 一种基于Stacking单导联脑电睡眠自动分期方法,其特征在于:该方法包括以下步骤:

- S1:睡眠脑电信号预处理;
- S2:睡眠脑电多特征提取和筛选;
- S3:机器学习分类;
- S4:睡眠自动分期。

2. 根据权利要求1所述的一种基于Stacking单导联脑电睡眠自动分期方法,其特征在于:所述步骤S1具体为:使用自适应阈值的小波函数和IIR滤波函数结合的滤波方法对脑电信号进行降噪处理,提高信号的信噪比,小波函数使用自适应阈值和软阈值函数,小波降噪的参数为小波基选择'db4',小波分解层数为5层;IIR滤波函数使用20阶巴特沃斯滤波函数,滤波频率为5-30HZ。

3. 根据权利要求1所述的一种基于Stacking单导联脑电睡眠自动分期方法,其特征在于:所述步骤S2具体为:采用特征算法,得出特征数据组;特征算法来自于时域、时频域和非线性域三个领域;时域采用的特征算法包括:基于Hjorth三个参数活动性,移动性,复杂性,峰度、偏度、基于 α 、 β 、 θ 、 δ 方差、基于 α 、 β 、 θ 、 δ 最大幅值;时频域采用的特征算法包括:基于 α 、 β 、 θ 、 δ 的小波能量,基于 δ/α 的小波能量比;非线性采用的特征算法包括:模糊熵、样本熵、分形维度,基于 α 、 β 、 θ 、 δ 的样本熵,复杂度、赫斯特指数、Tsallis熵,最大李雅科夫指数、排列熵;

特征中 α 、 β 、 θ 、 δ 节律波由小波分解进行五层分解得到;模糊熵和样本熵参数为嵌入维度 m 和相似容限 r ;利用相关系数方法找出模糊熵、样本熵和排列熵最优的参数;其中模糊熵的 m 和 r 分别为1和0.15,样本熵的 m 和 r 分别为2和0.15,排列熵参数中嵌入维数为2,延迟时间为10;利用C-C办法计算最大李雅科夫指数中时间延迟为17、平均周期256为和嵌入维数为1.5;

特征筛选算法使用Fisher Score特征筛选算法,利用Fisher Score选择特征的基本流程,即先求出每个特征的Fisher Score分数,然后设定一个阈值,若某个特征的Fisher Score分数大于阈值,则选择这个特征;否则,舍弃该特征,本算法阈值设定为0.3;

利用特征筛选算法从30个特征数据组进行筛选,选择模糊熵、Hjorth的复杂性(complexity)、基于 β 和 δ 的节律波样本熵、分形维度、排列熵、基于 β 、 θ 和 δ 节律波的小波能量、基于 β 节律波的方差。

4. 根据权利要求1所述的一种基于Stacking单导联脑电睡眠自动分期方法,其特征在于:所述步骤S3具体为:采用的分类算法是基于Stacking集成学习算法,通过元分类器或元回归聚合多个分类或回归模型;基础层次模型(level model)基于完整的训练集进行训练,然后元模型基于基础层次模型的输出进行训练;基础层次由不同的学习算法组成,stacking集成是异构的;

基础层采用的元学习算法包括K最近邻(K-NearestNeighbor,KNN)分类算法、随机森林(RandomForest,RF)、极端随机树(Extremely Randomized Trees,ERT)、梯度提升决策树(Gradient Boosting Decision Tree,GBDT)和多层感知机(Multi-layer Perceptron,MLP);

第二层采用通用的线性逻辑回归分类方法,即特征信号先通过基础层五个异构的元学

习器,各自得出分类结果;在第二层中输入基于第一层产生的结果,通过线性分类算法得出最后的分类结果。

5.根据权利要求1所述的一种基于Stacking单导联脑电睡眠自动分期方法,其特征在于:所述步骤S4具体为:利用机器学习算法对睡眠状态进行分期后,对分期的结果进行评估,判定分期方法的优劣;评估参数有混淆矩阵、分类准确度和Kapper系数;

混淆矩阵基本形式是基于二分类,同时适用多分类情况;真阳性(True Positive,TP)是正类预测为正类数,假阴性(False Negative,FN)是正类预测为负类数,假阳性(False Positive,FP)是将负类预测为正类,真阴性(True Negative,TN)是将负类预测为负类;基于R&K睡眠分期标准,将睡眠分为六期:Awake、N-REM stage 1、N-REM stage 2、N-REM stage 3、N-REM stage 4和REM;

分类准确度(Accuracy,ACC)定义准确分类的样本数与样本总数的比值,从混淆矩阵得出:

$$ACC = \frac{TP + TN}{TP + FN + FP + TN}$$

TP为真阳性样本数,TN为真阴性样本数,FN为假阴性样本数,FP为假阳性样本数;

利用kappa系数用衡量专家分期和本方法分期结果的吻合程度,计算公式为:

$$\kappa = \frac{ACC - ACC_0}{1 - ACC_0}$$

ACC为分类准确度,机会水平 $ACC_0 = 1/N_Y$, N_Y 为分类数目;Kappa系数为0时,表示分类准确度为机会水平程度,Kappa系数为1表示最好的分类效果。

一种基于Stacking单导联脑电睡眠自动分期方法

技术领域

[0001] 本发明属于机器学习算法领域,涉及一种基于Stacking单导联脑电睡眠自动分期方法。

背景技术

[0002] 睡眠研究对于人的身心健康和日常工作具有重要意义。睡眠分期是了解睡眠状况和睡眠质量评价的关键,基于睡眠分期的睡眠评分同时也是精神病学和神经病学最重要的诊断方法。

[0003] Rechtschaffen和Kates根据人体睡眠过程中脑电图 (Electroencephalogram, EEG)、眼电图 (Electrooculogram, EOG)、肌电图 (Electromyography, EMG) 中的眼球运动情况和肌肉张大的变化等因素,提出R&K分期标准,把睡眠分为觉醒期、非快速眼动睡眠期 (NREM) 和快速眼动睡眠期 (REM),其中NREM又可分为睡眠I期、睡眠II期、睡眠III期和睡眠IV期。2007年AASM协会,推出了AASM睡眠标准,AASM标准将睡眠分为五期,分别是清醒期、非快速眼动期 (NREM,又分为1期,2期,3期)、快速眼动期 (REM)。

[0004] 目前对于睡眠自动分期研究主要在于脑电信号。脑电信号幅度非常微弱。频率范围0.5-50hz。脑电信号幅值一般在50 μ V左右,高于100 μ V一般可以看做是噪音。脑电信号是一种随机性很强的非平稳信号。影响脑电的因素很多,表现的很多规律和意义至今仍没有认识。不得不利用统计处理技术如方差、均值等来检测和估计它的特征。脑电信号同时也是非线性信号。传统意义上的信号处理方法都是建立在线性的理论上的,因此,基于传统信号处理,会不可避免的丢失很多原始信号所带的信息。所以,非线性动力学、信息论、混沌理论发展起来迅速应用到脑电的分析研究中。非线性的研究越来越在脑电信号中。所以基于脑电信号特征,研究者在对睡眠脑电信号进行特征提取中,一般从时域特征、频域特征、时频域特征和非线性特征四个方面来提取单个特征或者多个特征。

[0005] 睡眠脑电特征通过特征算法提取出来之后,需要利用机器学习算法或模型进行进一步的分析和处理。机器学习按照输入情况大致可以分为两类:监督式学习和非监督式学习。睡眠脑电分期大多采用监督式学习,即输入部分(训练集和测试集)包括特征数据列表和特征数据标签值列表。特征数据标签是睡眠分期状态,一般由人工专家提供。机器学习按照输出情况可分为两类:分类算法和回归算法。睡眠分期状态是离散的,属于分类问题。目前常用于睡眠分期的方法有支持向量机 (Support vector machine, SVM)、人工神经网络 (Artificial Neural Network, ANN)、决策树 (decision tree, DT)、随机森林 (Random forest, RF)。

[0006] Baha等利用C3-A2的单导联脑电信号,再利用小波系数、排列熵和赫斯特指数作为特征算法,使用随机森林分类算法对睡眠进行六期分类,算法的准确率为97.03%。Arthur等利用C3和C4的脑电数据,再利用复杂度和相关系数,使用隐马尔科夫 (HMM) 模型对睡眠进行三期分类,算法的准确率为80%。Luay等利用Pz-Oz脑电信号,再利用小波系数,使用决策树分类算法对睡眠进行六期分类,算法的准确率为75%。Salih等利用C4-A1脑电信号,再利

用韦尔奇分析作为特征算法,使用决策树分类算法对睡眠进行六期分类,算法的准确率82.15%。Thiago等利用Pz-0z脑电信号,再利用方差、偏度、峰度,使用随机森林分类算法对睡眠进行六期分类,算法的准确率为90.5%。Farideh等利用Pz-0z脑电信号,再利用小波包系数,使用人工神经网络分类算法对睡眠进行五期分类,算法的准确率为93%。Zhu等利用Pz-0z脑电信号,再利用不同可视图 (Difference Visibility Graph,DVG) 作为特征算法,使用支持向量机 (Support Vector Machine,SVM) 分类算法对睡眠进行六期分类,算法的准确率为87.5%。kaveh等利用Fpz-Cz脑电信号,再利用RDSTFT作为特征算法,使用随机森林进行4期睡眠分类,算法准确度为92.5%。刘志勇等利用Pz-0z脑电信号,再利用符号熵指数、去趋势波动指数、 δ 频率能量作为特征,使用SVM对睡眠进行五期分类,算法准确度为92.87%。林秀晶等利用Pz-0z脑电信号,再利用均值、方差、频率能量、频谱熵,使用SVM对睡眠进行五期分类,算法准确度为80.1%。

[0007] 现有技术的不足包括以下几点:

[0008] (1) 睡眠脑电信号的预处理

[0009] 脑电是一种随机性,非平稳性很高的生理信号,不同的节律、不同心理改变都会影响脑电波的变化。因此,EEG具有很高的时变敏感性,其信号极易被无关噪音污染。从而形成各种EEG噪音和伪迹。常见的伪迹有包括来及仪器的干扰和来自人体的如眼电伪迹 (ocular artifact)、肌电伪迹 (muscle artifact)。如何对常见的噪音进行有效的消除和尽可能避免在信号中产生眼电伪迹是预处理的关键。

[0010] (2) 睡眠分期算法准确度

[0011] 可用于睡眠脑电信号的特征很多,如时域上的均值、方差、时频域上的不同节律波的小波系数、样本熵、Reyi系数、排列熵等。基于时域和时频域上的特征参数很容易选择。但是基于非线性的算法的如样本熵、关联维数、最大Lyapunov熵等进行算法提取时如嵌入维数、延迟时间等特征参数的选择十分重要。目前没有统一的原则针对于信号是睡眠脑电信号时特征参数选择。特征选择之后,需要对特征进行优化和筛选。不同的选择原则有不同特征选择方案。如针对于分期算法高效性原则,特征算法中不应有关联维数、最大Lyapunov熵等运算周期长的特征算法。如针对于分期算法的准确度原则,需要对特征进行特征筛选工作。其目的在于找出和原信号相关系数高的特征算法。对特征算法进行如何的筛选方法决定了分期算法的准确度。

[0012] 可应用于睡眠分期的机器学习算法很多,如支持向量机、随机森林、神经网络等。如何选择合适的机器学习分类器来对脑电信号进行分类。且在每个机器学习算法内部参数很多,如何利用一定准则对参数进行调优,从而提高算法准确度,是目前研究的关键。

发明内容

[0013] 有鉴于此,本发明的目的在于提供一种基于Stacking单导联脑电睡眠自动分期方法,提高睡眠分期方法的准确率和分期效率,再利用睡眠分期结果设计睡眠质量评价模型,提供一种科学的参考依据,使个体和医师能够利用本发明得到睡眠分期的结果了解睡眠质量状况和某些睡眠疾病发病可能性。

[0014] 为达到上述目的,本发明提供如下技术方案:

[0015] 一种基于Stacking单导联脑电睡眠自动分期方法,包括以下步骤:

[0016] S1:睡眠脑电信号预处理;

[0017] S2:睡眠脑电多特征提取和筛选;

[0018] S3:机器学习分类;

[0019] S4:睡眠自动分期。

[0020] 进一步,所述步骤S1具体为:使用自适应阈值的小波函数和IIR滤波函数结合的滤波方法对脑电信号进行降噪处理,提高信号的信噪比,小波函数使用自适应阈值和软阈值函数,小波降噪的参数为小波基选择‘db4’,小波分解层数为5层;IIR滤波函数使用20阶巴特沃斯滤波函数,滤波频率为5-30HZ。

[0021] 进一步,所述步骤S2具体为:采用特征算法,得出特征数据组;特征算法来自于时域、时频域和非线性域三个领域;时域采用的特征算法包括:基于Hjorth三个参数活动性,移动性,复杂性,峰度、偏度、基于 α 、 β 、 θ 、 δ 方差、基于 α 、 β 、 θ 、 δ 最大幅值;时频域采用的特征算法包括:基于 α 、 β 、 θ 、 δ 的小波能量,基于 δ/α 的小波能量比;非线性采用的特征算法包括:模糊熵、样本熵、分形维度,基于 α 、 β 、 θ 、 δ 的样本熵,复杂度、赫斯特指数、Tsallis熵,最大李雅科夫指数、排列熵;

[0022] 特征中 α 、 β 、 θ 、 δ 节律波由小波分解进行五层分解得到;模糊熵和样本熵参数为嵌入维度 m 和相似容限 r ;利用相关系数方法找出模糊熵、样本熵和排列熵最优的参数;其中模糊熵的 m 和 r 分别为1和0.15,样本熵的 m 和 r 分别为2和0.15,排列熵参数中嵌入维数为2,延迟时间为10;利用C-C办法计算最大李雅科夫指数中时间延迟为17、平均周期256为和嵌入维数为1.5;

[0023] 特征筛选算法使用Fisher Score特征筛选算法,利用Fisher Score选择特征的基本流程,即先求出每个特征的Fisher Score分数,然后设定一个阈值,若某个特征的Fisher Score分数大于阈值,则选择这个特征;否则,舍弃该特征,阈值设定为0.3;

[0024] 从特征数据,选择模糊熵、Hjorth的复杂性(complexity)、基于 β 和 δ 的节律波样本熵、分形维度、排列熵、基于 β 、 θ 和 δ 节律波的小波能量、基于 β 节律波的方差。

[0025] 进一步,所述步骤S3具体为:采用的分类算法是基于Stacking集成学习算法,通过元分类器或元回归聚合多个分类或回归模型;基础层次模型(level model)基于完整的训练集进行训练,然后元模型基于基础层次模型的输出进行训练;基础层次由不同的学习算法组成,stacking集成是异构的;

[0026] 基础层采用的元学习算法包括K最近邻(K-NearestNeighbor,KNN)分类算法、随机森林(RandomForest,RF)、极端随机树(Extremely Randomized Trees,ERT)、梯度提升决策树(Gradient Boosting Decision Tree,GBDT)和多层感知机(Multi-layer Perceptron,MLP);

[0027] 第二层采用通用的线性逻辑回归分类方法,即特征信号先通过基础层五个异构的元学习器,各自得出分类结果;在第二层中输入基于第一层产生的结果,通过线性分类算法得出最后的分类结果。

[0028] 进一步,所述步骤S4具体为:利用机器学习算法对睡眠状态进行分期后,对分期的结果进行评估,判定分期方法的优劣;评估参数有混淆矩阵、分类准确度和Kapper系数;

[0029] 混淆矩阵基本形式是基于二分类,同时适用多分类情况;真阳性(True Positive, TP)是正类预测为正类数,假阴性(False Negative, FN)是正类预测为负类数,假阳性

(False Positive,FP)是将负类预测为正类,真阴性(True Negative,TN)是将负类预测为负类;基于R&K睡眠分期标准,将睡眠分为六期:Awake、N-REM stage 1、N-REM stage 2、N-REM stage 3、N-REM stage 4和REM;

[0030] 分类准确度(Accuracy,ACC)定义准确分类的样本数与样本总数的比值,从混淆矩阵得出:

$$[0031] \quad ACC = \frac{TP + TN}{TP + FN + FP + TN}$$

[0032] TP为真阳性样本数,TN为真阴性样本数,FN为假阴性样本数,FP为假阳性样本数;

[0033] 利用kappa系数用衡量专家分期和本方法分期结果的吻合程度,计算公式为:

$$[0034] \quad \kappa = \frac{ACC - ACC_0}{1 - ACC_0}$$

[0035] ACC为分类准确度,机会水平 $ACC_0 = 1/N_Y$, N_Y 为分类数目;Kappa系数为0时,表示分类准确度为机会水平程度,Kappa系数为1表示最好的分类效果。

[0036] 本发明的有益效果在于:

[0037] (1) 获取自适应阈值的小波函数和IIR滤波函数结合的滤波方法对脑电信号进行降噪处理,有效地提高信号的信噪比;

[0038] (2) 对特征算法进行优化筛选获取一种新的特征参数组,作为睡眠分期的特征;

[0039] (3) 获取一种新的准确率高的多特征与集成学习算法组合作为睡眠分期算法。

附图说明

[0040] 为了使本发明的目的、技术方案和有益效果更加清楚,本发明提供如下附图进行说明:

[0041] 图1为本发明主要流程图;

[0042] 图2为原始信号与降噪后信号对比图;

[0043] 图3为本发明中采用的Stacking模型。

具体实施方式

[0044] 下面将结合附图,对本发明的优选实施例进行详细的描述。

[0045] 图1为本发明主要流程图,解决的技术问题:

[0046] (1) 脑电信号的预处理

[0047] 脑电信号的特点是高度随机性、高度的非平衡性、高度的非线性性,脑电信号的降噪处理是脑电信号处理关键。需要对传统的巴特沃斯滤波器和目前流行的小波降噪算法进行对比,同时通过比较不同的小波基函数、小波的阈值函数、小波层数、小波阈值大小。研究出新的适合睡眠脑电降噪算法。

[0048] 获取自适应阈值的小波函数和IIR滤波函数结合的滤波方法对脑电信号进行降噪处理,有效的提高信号的信噪比。

[0049] (2) 睡眠分期算法设计与评估

[0050] ①如何选择特征参数和如何对特征参数进行筛选

[0051] 对于得出的多特征值,需要对特征值进行筛选,理由如下:特征值并不是都能体现

出脑电信号的真实信息,需要找出主要的、关联性好的特征参数。特征算法提取的难度、时间效率有所区别,基于时域的特征算法提取速度很快,但是特征不能很好的表征睡眠信息,而关联维数、分型维数等可以很好的表征睡眠信息但是提取速度较慢。需要对其进行平衡处理。本发明使用Fisher score特征筛选算法对特征进行筛选,选择的标准在于是否提高了分期算法的准确度。

[0052] ②如何选取机器学习算法进行分期

[0053] 通过特征向量,利用机器学习算法来对其进行分类,目前的机器学习算法很多,如神经网络、支持向量机、决策树、随机森林、梯度提升决策树(Gradient Boosting Decision Tree,GBDT)等。本发明使用Stacking集成学习算法进行分类,利用机器学习常见的评估方法进行比较,得出最适合本方法的机器学习方法。常用的评估参数有混淆矩阵、分类准确度、和Kapper系数。

[0054] 本发明主要研究内容包括以下几个方面:

[0055] (1)睡眠脑电信号预处理

[0056] (2)睡眠脑电多特征提取和特征筛选

[0057] (3)机器学习算法的选择和性能评估

[0058] (4)睡眠自动分期

[0059] 1、数据选择

[0060] 本方法验证的数据来自于PhysioNet website公共数据库。数据名为Sleep-EDF。利用61个多导睡眠仪(polysomnograms,PSGs)在1987-1991和1994分别两次记录睡眠信号。第一次采集睡眠信号时,采集了20个无睡眠方面病史的正常成年人20个小时睡眠信号,其中年龄在23到25之间,男女各占一半。信号采样频率为100Hz。使用脑电导联Fpz-Cz和Pz-0z代替传统的C4-A1和C3-A2。数据中带有专家的人工分期标准,每30s判断一次睡眠分期状态。本研究采用Pz-0z单脑电导联,把睡眠按照R&K准则分为六期,分别为Awake、N1、N2、N3、N4、REM。从10个人提取到本次使用的原始数据如下表1所示:

[0061] 表1本研究采用的睡眠数据分布图

[0062]

睡眠分期	Awake	N1	N2	N3	N4	REM	Total
数据(30S)	3853	826	1185	991	831	537	8233

[0063] 2、信号预处理

[0064] 使用自适应阈值的小波函数和IIR滤波函数结合的滤波方法对脑电信号进行降噪处理,有效地提高信号的信噪比,小波降噪的主要参数:小波基选择'db4',小波分解层数为5层、使用自适应阈值和软阈值函数。IIR滤波函数使用20阶巴特沃斯滤波函数,滤波频率为5-30HZ。

[0065] 图2为原始信号与降噪后信号的对比图。

[0066] PER值为降噪信号的方差/原始信号的模.ERR为降噪信号和原始信号差的模。本方法中PER值为0.93.ERR值为142.7。从数值上看,降噪算法达到了很好的滤波效果。

[0067] 从图上可以看出原始信号上,局部上混杂较为严重,通常这是噪音的标志。在降噪

后的信号中,有效的降低了噪音。与原信号相比,细节更为具体。降噪函数同时有效的消除了原信号具有的较多的孤立尖峰点。由此可以看出,降噪函数可以明显的提高信号的纯净度。

[0068] 3、特征提取和特征筛选

[0069] 本研究使用15个特征算法,得出30个特征数据组。特征算法来自于时域、时频域和非线性域三个领域。表2为30特征算法的分布。本文在这里不详细对每个特征算法原理和提取细节进行阐述,只对重要的影响特征算法优劣的参数进行叙说。特征中 α 、 β 、 θ 、 δ 节律波由小波分解进行五层分解得到。模糊熵和样本熵参数为嵌入维度 m 和相似容限 r 。本研究利用相关系数方法找出模糊熵、样本熵和排列熵最优的参数。其中模糊熵的 m 和 r 分别为1和0.15,样本熵的 m 和 r 分别为2和0.15,排列熵参数中嵌入维数为2,延迟时间为10。利用C-C办法计算最大李雅科夫指数中时间延迟为17、平均周期256为和嵌入维数为1.5。

[0070] 表2本研究采用的特征算法

[0071]

特征领域

	基于 Hjorth 三个参数活动性、移动性、复杂性、峰度、偏度
时域	基于 α 、 β 、 θ 、 δ 方差
	基于 α 、 β 、 θ 、 δ 最大幅值
时频域	基于 α 、 β 、 θ 、 δ 的小波能量
	基于 δ/α 的小波能量比
	模糊熵、样本熵、分形维度
	基于 α 、 β 、 θ 、 δ 的样本熵
非线性	复杂度、赫斯特指数、Tsallis 熵
	最大李雅科夫指数、排列熵

[0072] 特征筛选算法使用常见的Fisher Score特征筛选算法,利用Fisher Score选择特征的基本流程是,先求出每个特征的Fisher Score分数,然后设定一个阈值,若某个特征的Fisher Score分数大于阈值,则选择这个特征;否则,舍弃该特征。Fisher Score特征筛选计算方法简单,准确性高、可操作性强、节省运算时间。故本研究采用该特征筛选办法,阈值设定为0.3。

[0073] 从30个特征数据,选择模糊熵、Hjorth的复杂性(complexity)、基于 β 和 δ 的节律波样本熵、分形维度、排列熵、基于 β 、 θ 和 δ 节律波的小波能量、基于 β 节律波的方差。

[0074] 4、机器学习分类

[0075] 本研究采用的分类算法是基于Stacking集成学习算法。Stacking是一种集成学习技术,通过元分类器或元回归聚合多个分类或回归模型。基础层次模型(level model)基于完整的训练集进行训练,然后元模型基于基础层次模型的输出进行训练。基础层次通常由不同的学习算法组成,因此stacking集成通常是异构的。本研究采用的Stacking模型如图3所示。

[0076] 基础层采用五个异构的元学习算法包括K最近邻(K-NearestNeighbor,KNN)分类算法、随机森林(RandomForest,RF)、极端随机树(Extremely Randomized Trees,ERT)、梯度提升决策树(Gradient Boosting Decision Tree,GBDT)和多层感知机(Multi-layer Perceptron,MLP)。

[0077] 第二层采用通用的线性逻辑回归分类方法。特征信号先通过基础层五个异构的元学习器,各自得出分类结果。在第二层中输入基于第一层产生的结果,通过线性分类算法得出最后的分类结果。

[0078] 5、分类评估

[0079] 利用机器学习算法对睡眠状态进行分期后,需要对分期的结果进行评估,来判定分期算法的优劣。常用的评估参数有混淆矩阵、分类准确度和Kapper系数。

[0080] 混淆矩阵基本形式是基于二分类,同时适用多分类情况。真阳性 (True Positive, TP) 是正类预测为正类数 (准确判断为正类)。假阴性 (False Negative, FN) 是正类预测为负类数。假阳性 (False Positive, FP) 是将负类预测为正类。真阴性 (True Negative, TN) 是将负类预测为负类 (准确判断为负类)。基于R&K睡眠分期标准,将睡眠分为六期:Awake, N-REM 1, N-REM 2, N-REM 3, N-REM 4, REM。建立基于睡眠六期的混淆矩阵如表3。

[0081] 表3睡眠六期的混淆矩阵

[0082]

		本算法分组					
		Awake	N1	N2	N3	N4	REM
专家 分组	Awake	1135	14	4	0	0	2
	N1	0	236	6	0	0	6
	N2	0	8	342	0	0	5
	N3	0	0	2	284	11	0
	N4	0	0	2	6	242	0
	REM	0	3	10	3	0	145
	ACC	98.27%	95.16%	96.33%	95.62%	96.8%	91.93%
Total ACC							96.67%

[0083] 从对睡眠的混淆矩阵中可以看出,算法对于各期的睡眠准确率都大于90%,对于清醒期和睡眠期主要的分期,算法达到了98.27%。可以认为算法进行高效准确的分期。同时,在比较难分期的N-REM stage 2和REM期,算法在这两期主要的错分类数据也集中在于此。但是算法在这个期间的分期准确度是可观的,超过了90%。大于一般的睡眠分期算法。

[0084] 分类准确度 (Accuracy, ACC) 定义了准确分类的样本数与样本总数的比值,可以从混淆矩阵得出

$$[0085] \quad ACC = \frac{TP + TN}{TP + FN + FP + TN}$$

[0086] 本分类算法分期准确度为96.67%,高于最近的研究现状中睡眠六期的准确度。本文利用kappa系数用来衡量专家分期和本研究算法分期结果的吻合程度。计算公式为:

$$[0087] \quad \kappa = \frac{ACC - ACC_0}{1 - ACC_0}$$

[0088] ACC为分类准确度,机会水平 $ACC_0 = 1/N_Y$, N_Y 为分类数目。Kappa系数为0时,表示分类准确度为机会水平程度,Kappa系数为1意味着最好的分类效果。本方法的kappa系数为0.96。说明本算法进行的睡眠分期的效果具有很高的一致性,睡眠分期准确度具有可靠的

参考价值。

[0089] 本方法提出的特征筛选方法同样适用于其他医学生理信号,有助于发现高效的特征算法。本研究算法可以为其准备一个可视化界面应用,增加适用性。本方法的成功表明了特征参数的多样性会有效地提高脑电信号及其他生理信号的识别精度,具有广泛的应用前景。

[0090] 最后说明的是,以上优选实施例仅用以说明本发明的技术方案而非限制,尽管通过上述优选实施例已经对本发明进行了详细的描述,但本领域技术人员应当理解,可以在形式上和细节上对其作出各种各样的改变,而不偏离本发明权利要求书所限定的范围。

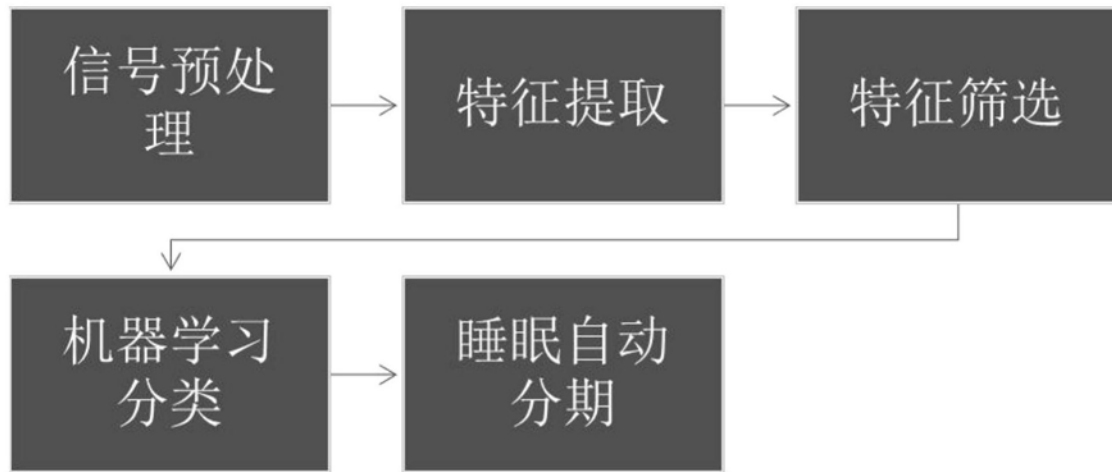


图1

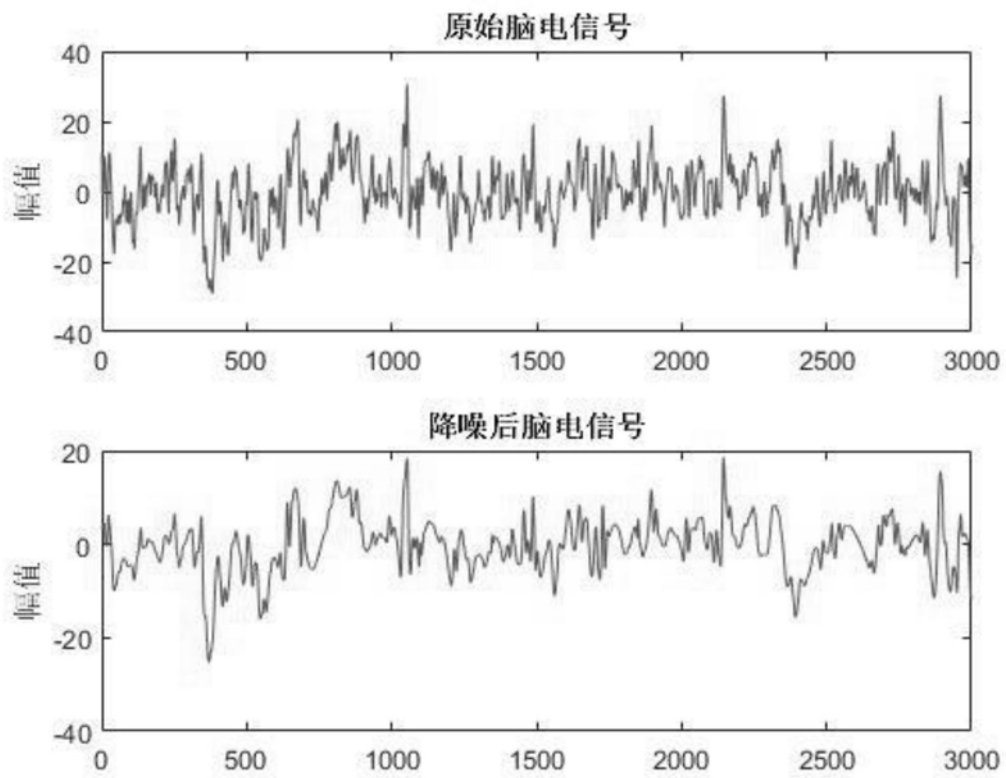


图2

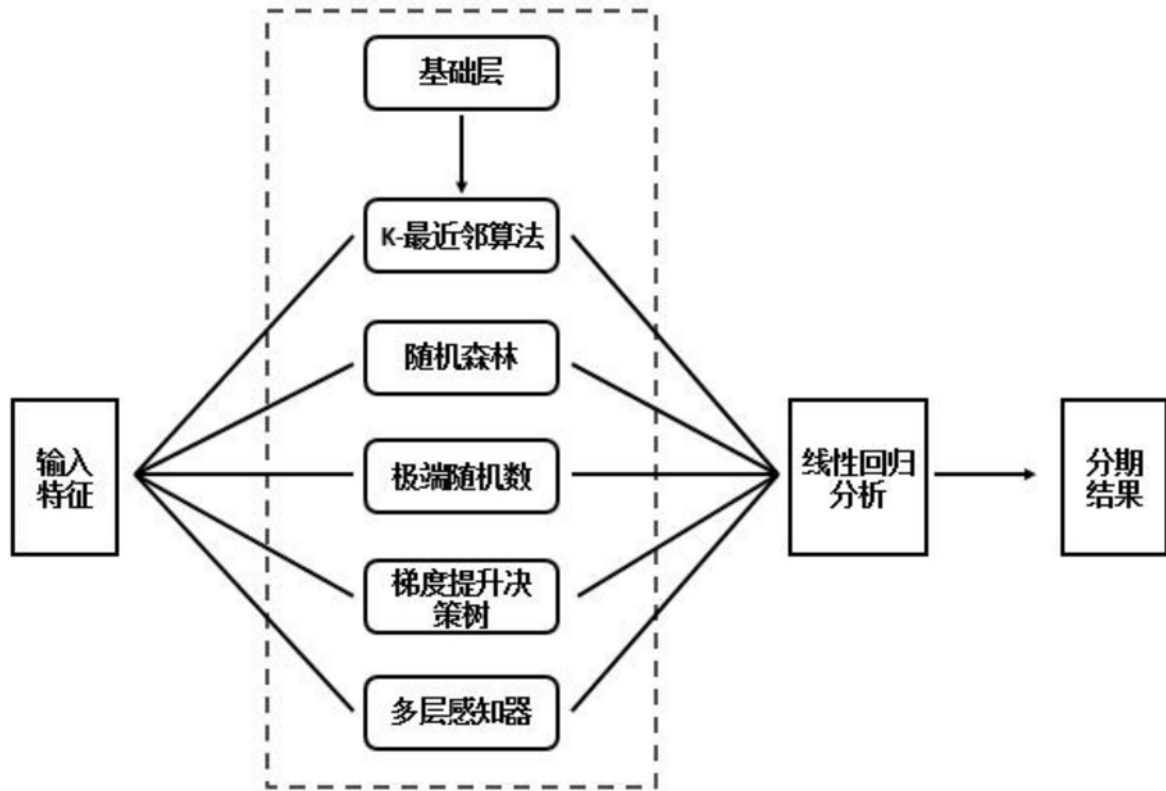


图3

专利名称(译)	一种基于Stacking单导联脑电睡眠自动分期方法		
公开(公告)号	CN108742517A	公开(公告)日	2018-11-06
申请号	CN201810259825.3	申请日	2018-03-27
[标]申请(专利权)人(译)	重庆邮电大学		
申请(专利权)人(译)	重庆邮电大学		
当前申请(专利权)人(译)	重庆邮电大学		
[标]发明人	王强强 赵德春 王怡 李舒粤		
发明人	王强强 赵德春 王怡 李舒粤		
IPC分类号	A61B5/00 A61B5/04 A61B5/0476		
CPC分类号	A61B5/4812 A61B5/04012 A61B5/0476		
外部链接	Espacenet SIPO		

摘要(译)

本发明涉及一种基于Stacking单导联脑电睡眠自动分期方法，属于机器学习算法领域。该方法包括步骤：S1：睡眠脑电信号预处理；S2：睡眠脑电多特征提取和筛选；S3：机器学习分类；S4：睡眠自动分期。本发明能够获取自适应阈值的小波函数和IIR滤波函数结合的滤波方法对脑电信号进行降噪处理，有效地提高信号的信噪比；能够对特征算法进行优化筛选获取一种新的特征参数组，作为睡眠分期的特征；能够获取一种新的准确率高的多特征与集成学习算法组合作为睡眠分期方法。

