



(12)发明专利申请

(10)申请公布号 CN 110543832 A

(43)申请公布日 2019. 12. 06

(21)申请号 201910745397.X

(22)申请日 2019.08.13

(71)申请人 同济大学

地址 200092 上海市杨浦区四平路1239号

(72)发明人 何良华 任强

(74)专利代理机构 上海科盛知识产权代理有限公司

公司 31225

代理人 叶敏华

(51)Int.Cl.

G06K 9/00(2006.01)

G06K 9/62(2006.01)

G06N 3/04(2006.01)

G06N 3/08(2006.01)

A61B 5/00(2006.01)

A61B 5/0476(2006.01)

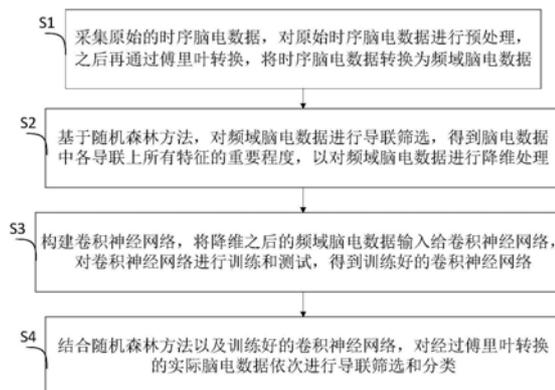
权利要求书2页 说明书5页 附图6页

(54)发明名称

基于随机森林和卷积神经网络的脑电数据分类方法

(57)摘要

本发明涉及一种基于随机森林和卷积神经网络的脑电数据分类方法,包括:S1、采集原始的时序脑电数据,并将时序脑电数据转换为频域脑电数据;S2、基于随机森林方法,对频域脑电数据进行导联筛选,以对频域脑电数据进行降维处理;S3、构建卷积神经网络,将降维之后的频域脑电数据输入给卷积神经网络,对卷积神经网络进行训练和测试,得到训练好的卷积神经网络;S4、结合随机森林方法以及训练好的卷积神经网络,对经过傅里叶转换的实际脑电数据依次进行导联筛选和分类。与现有技术相比,本发明利用随机森林筛选重要的特征,降低了脑电数据维度,减少了计算量;结合卷积神经网络对降维的脑电数据进行拟合,保证了脑电数据分类准确率。



1. 一种基于随机森林和卷积神经网络的脑电数据分类方法,其特征在于,包括以下步骤:

S1、采集原始的时序脑电数据,对原始时序脑电数据进行预处理,之后再通过傅里叶转换,将时序脑电数据转换为频域脑电数据;

S2、基于随机森林方法,对频域脑电数据进行导联筛选,得到脑电数据中各导联上所有特征的重要程度,以对频域脑电数据进行降维处理;

S3、构建卷积神经网络,将降维之后的频域脑电数据输入给卷积神经网络,对卷积神经网络进行训练和测试,得到训练好的卷积神经网络;

S4、结合随机森林方法以及训练好的卷积神经网络,对经过傅里叶转换的实际脑电数据依次进行导联筛选和分类。

2. 根据权利要求1所述的一种基于随机森林和卷积神经网络的脑电数据分类方法,其特征在于,所述步骤S1中对原始时序脑电数据进行预处理包括对脑电数据进行滤波和去除伪迹的操作。

3. 根据权利要求1所述的一种基于随机森林和卷积神经网络的脑电数据分类方法,其特征在于,所述步骤S2具体包括以下步骤:

S21、建立随机森林;

S22、获取频域脑电数据的灰度图,并将每个频率下的灰度图转换成对应的一维向量,以作为随机森林的输入;

S23、随机森林对所有的输入进行导联筛选,计算特征的重要程度;

S24、根据特征的重要程度,对频域脑电数据进行降维处理。

4. 根据权利要求3所述的一种基于随机森林和卷积神经网络的脑电数据分类方法,其特征在于,所述步骤S21具体包括以下步骤:

S211、定义样本数量为T,从所有样本中放回的随机抽样T个样本,将随机抽样的T个样本作为决策树根节点处的样本,以训练决策树;

S212、定义样本集中每个样本的特征数量为F,则每次决策树的节点需要进行分裂时,首先从这F个特征中选出f个特征,然后采用信息增益、信息增益比或基尼指数这三种策略,从f个特征中选出一个特征作为该节点的分裂属性,其中, $f \ll F$;

S213、重复步骤S212的过程,以进行决策树节点的分裂,直到无法再继续分裂为止,完成一颗决策树的训练;

S214、完成T棵决策树的训练后,由这T棵训练好的决策树构成随机森林。

5. 根据权利要求4所述的一种基于随机森林和卷积神经网络的脑电数据分类方法,其特征在于,所述步骤S22中随机森林的输入具体为W维的向量:

$$W=A \times B$$

其中,A表示导联的个数,B表示一个导联上所有频率的数量。

6. 根据权利要求5所述的一种基于随机森林和卷积神经网络的脑电数据分类方法,其特征在于,所述步骤S23和步骤S24中特征具体为导联上对应频率的数据。

7. 根据权利要求6所述的一种基于随机森林和卷积神经网络的脑电数据分类方法,其特征在于,所述步骤S23中计算特征的重要程度的具体过程为:

S231、使用袋外数据计算随机森林中每一棵决策树的误差err1;

S232、在所有袋外数据的相应特征位置加入噪声扰动,并计算此时决策树的误差err2;

S233、计算特征的重要程度为:

$$w_f = \frac{1}{T} \cdot \sum_{i=1}^T (i_err2 - i_err1)$$

其中, w_f 表示特征f的重要程度,T为随机森林中决策树的总棵数,i表示随机森林中的第i棵决策树, i_err2 表示在袋外数据相应特征位置加入噪声扰动后计算的第i棵决策树的误差, i_err1 表示使用袋外数据计算的第i棵决策树的误差。

8.根据权利要求1所述的一种基于随机森林和卷积神经网络的脑电数据分类方法,其特征在于,所述步骤S3具体包括以下步骤:

S31、构建卷积神经网络,所述卷积神经网络包括依次连接的卷积层、最大池化层和全连接层;

S32、将降维之后的频域脑电数据分为训练集和测试集;

S33、将训练集输入卷积神经网络,以预设的训练轮数进行训练;

S34、每完成一次预设的训练轮数,则将测试集输入卷积神经网络,记录一次卷积神经网络输出分类结果的准确率;

S35、重复步骤S34,并判断卷积神经网络输出分类结果的准确率是否收敛,若收敛,则该卷积神经网络已经训练好,否则返回步骤S33。

9.根据权利要求8所述的一种基于随机森林和卷积神经网络的脑电数据分类方法,其特征在于,所述步骤S31中卷积层由三个子卷积层构成,三个子卷积层的卷积核大小与步长均一致,全连接层由第一全连接层和第二全连接层组成,所述第一全连接层的神经元个数与随机森林的输入的向量维数一致,所述第二全连接层的神经元个数为脑电数据分类结果的类别数。

基于随机森林和卷积神经网络的脑电数据分类方法

技术领域

[0001] 本发明涉及脑电信号的数据分析技术领域,尤其是涉及一种基于随机森林和卷积神经网络的脑电数据分类方法。

背景技术

[0002] 利用机器学习和深度学习方法对脑电信号进行分类是近年来脑电信号研究领域的一个热门方向,但由于脑电数据的维度高、计算量大,导致影响分类结果的数据特征过多,因此,采用机器学习方法对脑电数据进行分析时,通常无法找到一个合适的模型对脑电数据进行分类;采用深度学习的方法虽然能很好的拟合模型,可是随着数据规模的增大,必然会使计算量呈指数级增加。

[0003] 综上所述,在面对维度高、计算量大的脑电数据时,现有技术没有从脑电数据中提取出对分类结果产生重要作用的特征,机器学习和深度学习方法均不能在保证分类准确率的前提下,实现快速输出分类结果的目的。

发明内容

[0004] 本发明的目的就是为了解决上述现有技术存在的缺陷而提供一种基于随机森林和卷积神经网络的脑电数据分类方法。

[0005] 本发明的目的可以通过以下技术方案来实现:一种基于随机森林和卷积神经网络的脑电数据分类方法,包括以下步骤:

[0006] S1、采集原始的时序脑电数据,对原始时序脑电数据进行预处理,之后再通过傅里叶转换,将时序脑电数据转换为频域脑电数据;

[0007] S2、基于随机森林方法,对频域脑电数据进行导联筛选,得到脑电数据中各导联上所有特征的重要程度,以对频域脑电数据进行降维处理;

[0008] S3、构建卷积神经网络,将降维之后的频域脑电数据输入给卷积神经网络,对卷积神经网络进行训练和测试,得到训练好的卷积神经网络;

[0009] S4、结合随机森林方法以及训练好的卷积神经网络,对经过傅里叶转换的实际脑电数据依次进行导联筛选和分类。

[0010] 优选地,所述步骤S1中对原始时序脑电数据进行预处理包括对脑电数据进行滤波和去除伪迹的操作。

[0011] 优选地,所述步骤S2具体包括以下步骤:

[0012] S21、建立随机森林;

[0013] S22、获取频域脑电数据的灰度图,并将每个频率下的灰度图转换成对应的一维向量,以作为随机森林的输入;

[0014] S23、随机森林对所有的输入进行导联筛选,计算特征的重要程度;

[0015] S24、根据特征的重要程度,对频域脑电数据进行降维处理。

[0016] 优选地,所述步骤S21具体包括以下步骤:

[0017] S211、定义样本数量为T,从所有样本中放回的随机抽样T个样本,将随机抽样的T个样本作为决策树根节点处的样本,以训练决策树;

[0018] S212、定义样本集中每个样本的特征数量为F,则每次决策树的节点需要进行分裂时,首先从这F个特征中选出f个特征,然后采用信息增益、信息增益比或基尼指数这三种策略,从f个特征中选出一个特征作为该节点的分裂属性,其中, $f \ll F$;

[0019] S213、重复步骤S212的过程,以进行决策树节点的分裂,直到无法再继续分裂为止,完成一颗决策树的训练;

[0020] S214、完成T棵决策树的训练后,由这T棵训练好的决策树构成随机森林。

[0021] 优选地,所述步骤S22中随机森林的输入具体为W维的向量:

[0022] $W = A \times B$

[0023] 其中,A表示导联的个数,B表示一个导联上所有频率的数量。

[0024] 优选的,所述步骤S23和步骤S24中特征具体为导联上对应频率的数据。

[0025] 优选地,所述步骤S23中计算特征的重要程度的具体过程为:

[0026] S231、使用袋外数据计算随机森林中每一棵决策树的误差err1;

[0027] S232、在所有袋外数据的相应特征位置加入噪声扰动,并计算此时决策树的误差err2;

[0028] S233、计算特征的重要程度为:

[0029]
$$w_f = \frac{1}{T} \cdot \sum_{i=1}^T (i_err2 - i_err1)$$

[0030] 其中, w_f 表示特征f的重要程度,T为随机森林中决策树的总棵数,i表示随机森林中的第i棵决策树, i_err2 表示在袋外数据相应特征位置加入噪声扰动后计算的第i棵决策树的误差, i_err1 表示使用袋外数据计算的第i棵决策树的误差。

[0031] 优选地,所述步骤S3具体包括以下步骤:

[0032] S31、构建卷积神经网络,所述卷积神经网络包括依次连接的卷积层、最大池化层和全连接层;

[0033] S32、将降维之后的频域脑电数据分为训练集和测试集;

[0034] S33、将训练集输入卷积神经网络,以预设的训练轮数进行训练;

[0035] S34、每完成一次预设的训练轮数,则将测试集输入卷积神经网络,记录一次卷积神经网络输出分类结果的准确率;

[0036] S35、重复步骤S34,并判断卷积神经网络输出分类结果的准确率是否收敛,若收敛,则该卷积神经网络已经训练好,否则返回步骤S33。

[0037] 优选地,所述步骤S31中卷积层由三个子卷积层构成,三个子卷积层的卷积核大小与步长均一致,全连接层由第一全连接层和第二全连接层组成,所述第一全连接层的神经元个数与随机森林的输入的向量维数一致,所述第二全连接层的神经元个数为脑电数据分类结果的类别数。

[0038] 与现有技术相比,本发明使用随机森林与卷积神经网络的组合模型对脑电数据进行分类,首先采用随机森林计算所有特征对于分类结果的重要程度,并根据计算的特征的重要程度进行导联筛选,筛选出重要的特征,降低了数据维度,大大减少了计算量;然后基于卷积神经网络优良的数据拟合性,利用反向传播算法去学习适合脑电数据的参数,以实

现脑电数据分类的目的,通过结合随机森林与卷积神经网络,能够在保证分类准确率的前提下,降低脑电数据的维度与计算量,从而提高脑电数据分类的速度。

附图说明

- [0039] 图1为本发明的方法流程图;
- [0040] 图2为随机森林的工作流程示意图;
- [0041] 图3为实施例中八个随机森林对应的分类准确率示意图;
- [0042] 图4为实施例中包含100棵决策树的随机森林特征重要程度的可视化示意图;
- [0043] 图5为实施例中八个随机森林的特征重要程度平均值的可视化示意图;
- [0044] 图6为实施例中求取特征重要程度平均值后的导联重要程度示意图;
- [0045] 图7为实施例中导联筛选之后八个随机森林对应的分类准确率示意图;
- [0046] 图8为实施例中卷积神经网络的结构图;
- [0047] 图9为实施例中导联筛选之后卷积神经网络的输出分类准确率示意图。

具体实施方式

[0048] 下面结合附图和具体实施例对本发明进行详细说明。

[0049] 如图1所示,一种基于随机森林和卷积神经网络的脑电数据分类方法,包括以下步骤:

[0050] S1、采集原始的时序脑电数据,对原始时序脑电数据进行预处理,之后再通过傅里叶转换,将时序脑电数据转换为频域脑电数据;

[0051] S2、基于随机森林方法,对频域脑电数据进行导联筛选,得到脑电数据中各导联上所有特征的重要程度,以对频域脑电数据进行降维处理;

[0052] S3、构建卷积神经网络,将降维之后的频域脑电数据输入给卷积神经网络,对卷积神经网络进行训练和测试,得到训练好的卷积神经网络;

[0053] S4、结合随机森林方法以及训练好的卷积神经网络,对经过傅里叶转换的实际脑电数据依次进行导联筛选和分类。

[0054] 本实施例中,采集的原始时序脑电数据包括56个正样本和54个负样本,在数据采集条件相同的条件下,均为闭眼时的静息态脑电数据,其中,正样本为受试者在正常情况下的脑电数据,负样本为受试者在有外界噪声干扰情况下的脑电数据。每个受试者均采集180秒的数据,采集完原始时序脑电数据之后,首先对数据进行滤波、去除伪迹等预处理操作,之后使用快速傅里叶变换提取脑电数据中的频率信息,针对频率提取之后的脑电数据,本实施例还进行了min-max归一化处理,即:

$$[0055] \quad x = \frac{x - \min}{\max - \min}$$

[0056] 其中,min表示脑电数据集中特征x的最小值,max表示脑电数据集中特征x的最大值,使用min-max的方法将数据进行归一化,使得预处理的数据被限定在一定的范围内(比如[0,1]或者[-1,1]),从而消除奇异样本数据导致的不良影响。

[0057] 步骤S2中随机森林的工作流程如图2所示,随机森林由很多棵决策树所构成,森林的建立方式完全随机,因此被称作随机森林。森林构建完成之后,每次有新样本输入时,森

林中的每棵决策树都会对这个样本进行分类判断,并以一种投票的方式来进行决策,将票数最多的类别作为最后的分类结果;

[0058] 随机森林的基分类器是决策树,决策树是一种树型结构,决策树的叶节点代表了相应的分类类别。使用决策树进行分类的过程就是从树的根节点出发,通过特征中的不同属性选出决策树的不同分支,最终到达叶节点,结束分类的过程。

[0059] 本实施例使用随机森林方法进行导联筛选的具体过程为:

[0060] 一、随机森林输入的每个样本都是一个由多个特征组成的向量,因此需要将傅里叶变换后的灰度图拉成一维向量,作为随机森林的输入。直观上来看,随机森林的输入为1024维的向量(32导联x 32个频率值),每个特征代表了每个导联上对应频率的数据;

[0061] 本实施例中,建立了八个不同决策树棵数的随机森林,森林中决策树的棵树从40棵开始,每次增加20棵,一直增加到180棵,图3中所示八个随机森林对应的分类准确率,横坐标为随机森林中决策树的棵树,纵坐标为分类的准确率,图3中这八个随机森林模型分类准确率均在98%以上,当设置决策树棵数为80时,其对应的分类准确率最高,达到99%,这是由于随机森林参数少,从而能在小型数据集上产生优良的表现;

[0062] 使用随机森林对所有数据进行分类之后,可以得到训练好的模型对每个特征的敏感程度,将上述八个随机森林对每个特征计算的重要程度进行可视化,获取到每个特征的重要程度之后,由于输入数据自身的特点,每个特征都有其特定的空间和频域上的意义,将所有特征重要程度还原成 32×32 的图片之后可将其进行可视化,图4为100棵决策树的随机森林特征重要程度的可视化示意图,由图4可知,第2、5、7、20、22这5个导联在特征的分类中起到了主导作用,图中颜色较暗的位置对应的值较小,说明输入的样本中对应位置的特征对分类结果的贡献不大,即特征重要程度不高,而颜色较亮的位置是随机森林中对分类结果影响较大的特征;

[0063] 二、随机森林由于其初始化构建的过程中存在随机性,这可能导致多次重复运行得到的特征重要程度不同,因此较好的方式是使用不同配置的随机森林求取特征重要程度的平均值,因此在得到不同特征对分类结果的重要程度之后,可以利用此结果对输入的数据进行降维,将降维之后的数据再次使用随机森林进行分类,通过将降维前后分类结果进行对比就可以评估出选取的特征对分类结果的重要程度,即由于随机森林中决策树自身存在的随机性,每个随机森林模型给出的特征重要程度可能不同,为了降低随机误差对特征重要程度计算带来的影响,本实施例将上述八个随机森林给出的特征重要程度进行了平均,并从中找出了对最终分类结果影响较大的5个导联,图5所示为八个随机森林平均特征重要程度的可视化示意图,横坐标表示不同的频率,纵坐标表示不同的导联,同样沿着第一个维度求和,算出每个导联上的所有特征对分类结果的重要程度,图6所示为求取平均值之后各个导联上的特征对分类结果的贡献,横坐标表示不同的导联,纵坐标表示导联对分类结果的重要程度,从图6可以明显看出,第2、5、7、20、22这5个导联对分类结果的影响较大,这其中第20个导联最为重要,从图5中也能看到,第20个导联的亮度最大;另一方面,观察图5中这5行相应位置的亮度,我们可以看到这5行中频率数值小的像素点亮度更大,这说明导联上的低频成份对分类的影响更大;

[0064] 三、选出这5个导联之后,对脑电数据进行降维处理,从1024维的特征中筛选出了160维(5个导联x 32维特征),再次使用随机森林对筛选后的数据进行分类,图7所示为导联

筛选之后八个随机森林对应的分类准确率,图7中横坐标为随机森林中决策树的棵树,纵坐标为分类的准确率,从图7中可以看出,八个随机森林的分类准确率均在98%以上,对比图3未经过导联筛选的八个随机森林的分类准确率,仅降低了0.4%左右,这说明使用随机森林进行特征筛选方法是合理的,通过这种方式,特征数量由1024维降低至160维,同时保证了分类准确率。

[0065] 步骤S3中构建的卷积神经网络结构如图8所示,网络中包含三个卷积层和两个全连接层,relu被用作网络中的激活函数,每个卷积层之前都使用batchnorm(批标准化层)进行标准化,网络中所有卷积层中卷积核大小都为3,步长设置为1,降采样操作使用最大池化来完成,池化操作的大小为 2×2 ,第一个全连接层中神经元的个数为1024,第二个全连接层中神经元的个数为2(即设置的分类结果的类别数)。

[0066] 本实施例在网络的末端并没有使用全局平均池化,而是在最后一个卷积层之后连接了两个全连接层,主要是出于两方面考虑:首先,采集的原始脑电数据包含了54个正样本和56个负样本,因此虽然全连接层会增加一些参数量,但是训练数据的大量增加使得这些参数并不会影响模型训练;此外,由于分类结果的类别数被设置为2,如果使用全局平均池化的方式进行二分类,那么最后一个卷积层中卷积核的数量就要设置为2,通过图8可以看到,经过S6之后输出的特征图大小为 $4 \times 4 \times 128$,若直接使用卷积层将特征图减少到两张,信息量损失较大,因此在最后的特征图之后使用了两个全连接层。

[0067] 步骤S3中对卷积神经网络进行训练和测试的具体过程为:将降维之后的正样本和负样本数据分别以3:1的比例划分到训练集和测试集中,卷积神经网络使用Adam优化方法进行参数更新,学习率为0.05, L_2 权重衰减系数设置为0.5,同时使用指数衰减对学习率进行衰减,每200次迭代对学习率进行一次衰减,衰减系数为0.95,每训练3个epoch后,在测试集上进行一次分类准确率的测试,模型训练过程中测试结果如图9所示,具体为为卷积神经网络迭代次数的分类准确率,图9的横坐标代表分类准确率记录的次数,每三轮记录一次结果,纵坐标代表分类准确率。从图9可以看到,训练过程中分类准确率在 $95\% \pm 0.4\%$ 上下波动,经过5次记录之后,此时卷积神经网络收敛,在测试集上的准确率可以达到97%。

[0068] 为了进一步验证本发明方法的有效性,实施例还使用k折交叉验证的方法划分数据集,并统计最终对应的分类准确率:即将数据集等分成k份,将其中的k-1份用来训练,剩余的1份用来测试,然后取k次实验结果的均值作为最终的分类结果。本实施例中k的取值范围为2到7,表1所示为经过k折交叉验证的实施例的结果数据:

[0069] 表1

[0070]

k折	2	3	4	5	6	7
分类准确率	94.2%	94.7%	94.6%	94.7%	94.6%	94.5%

[0071] 从表1可以看出,交叉验证的结果较为稳定,准确率均在94%以上,从实施例的结果表明,相对于其他机器学习和深度学习的脑电分类算法,本发明提出的组合模型,首先利用随机森林筛选出重要的特征,降低了数据维度,减少了计算量,并且准确度只是些许下降;又利用了神经网络自身的拟合性,利用反向传播算法去学习适合脑电数据的参数,克服了根据数据选择模型对分类结果影响的未知性。

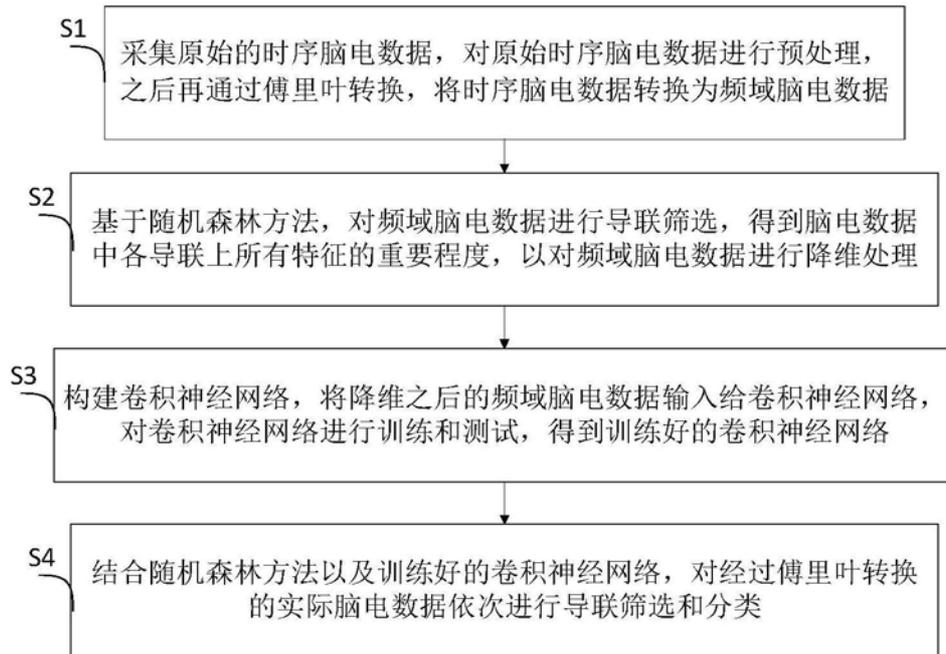


图1

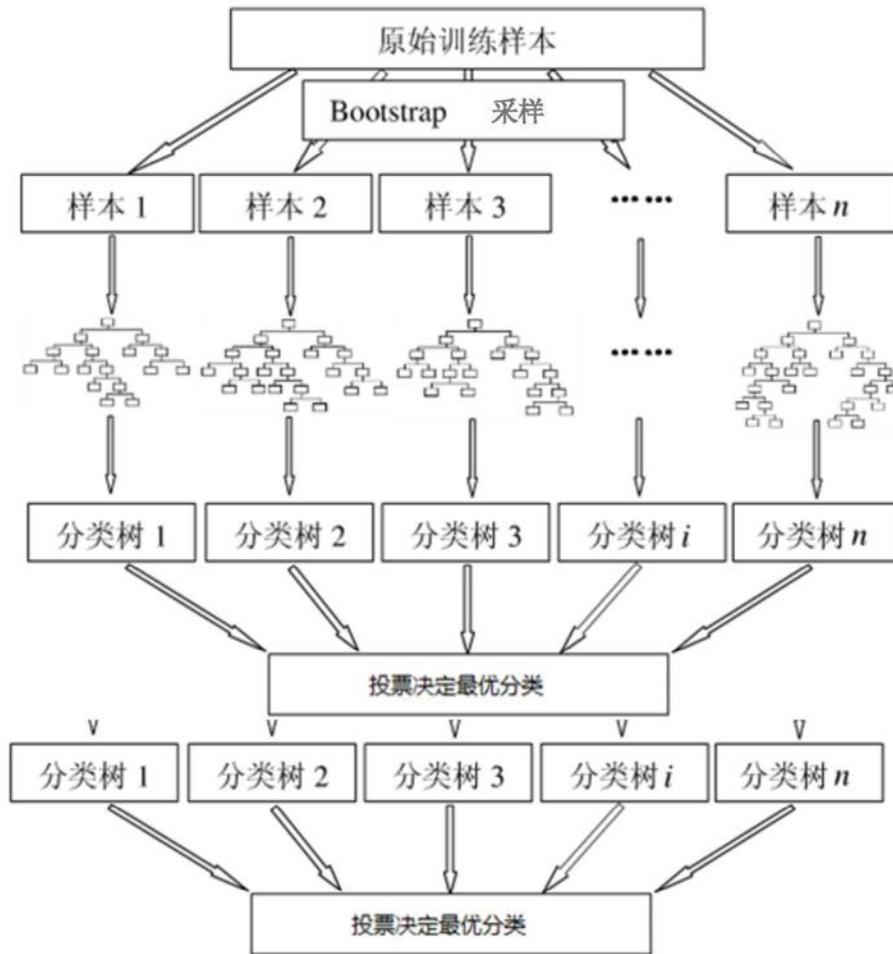


图2

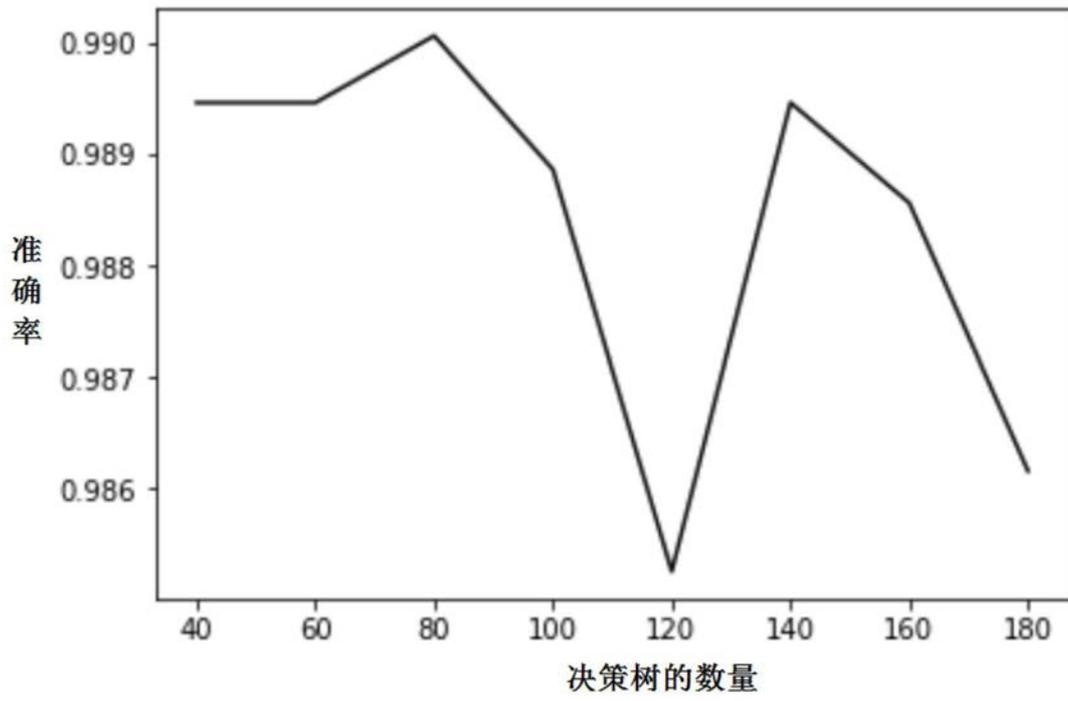


图3

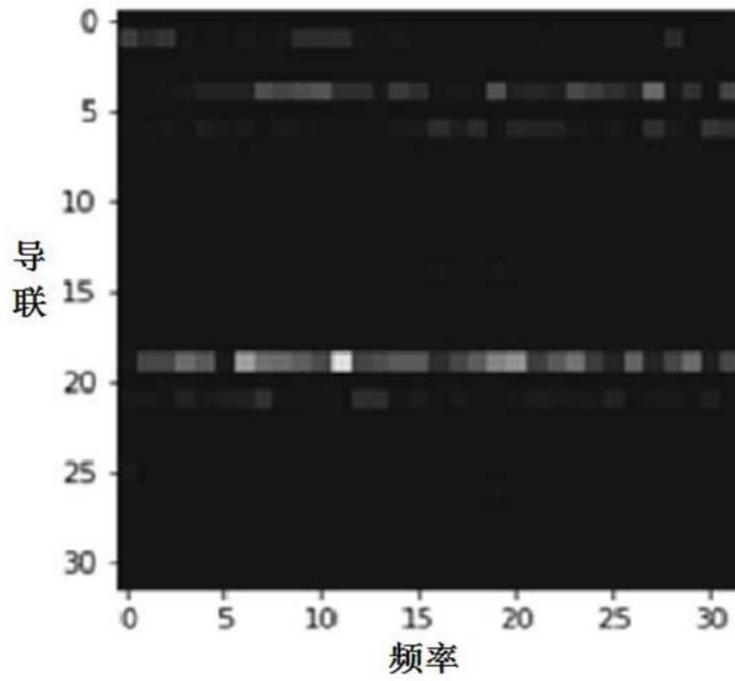


图4

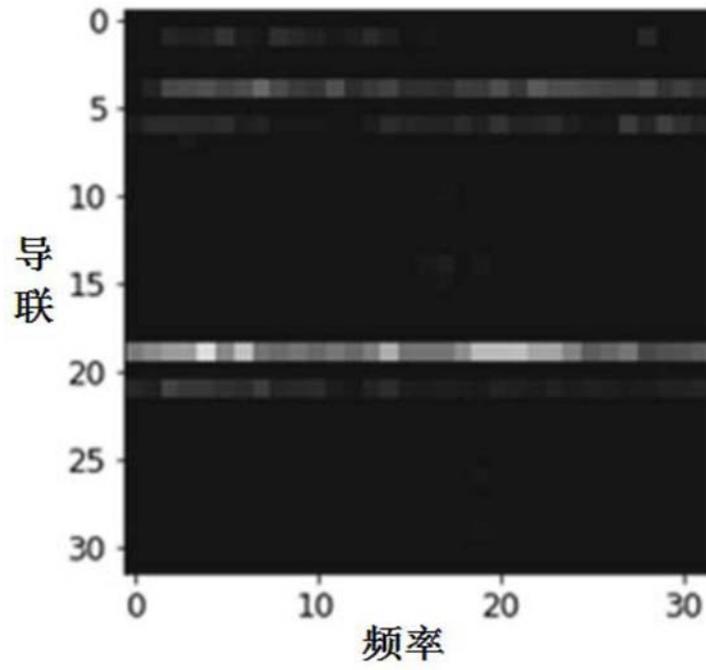


图5

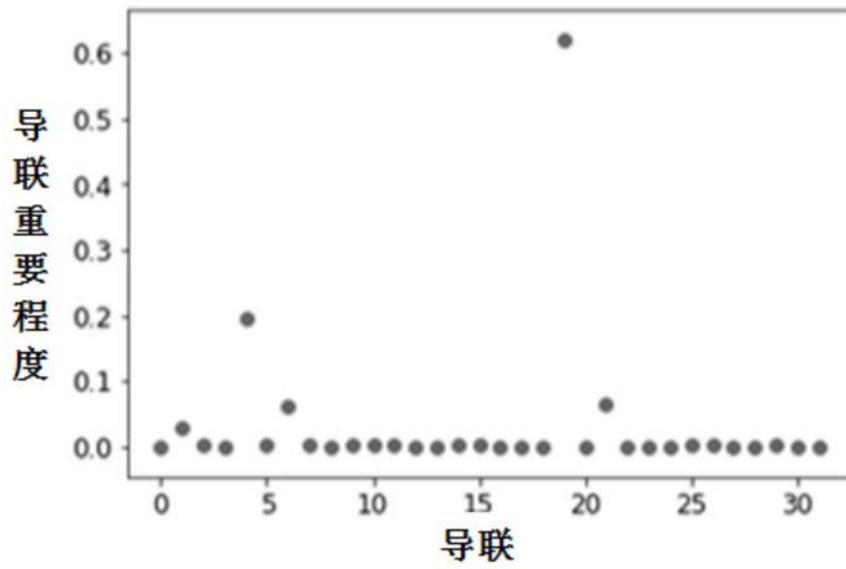


图6

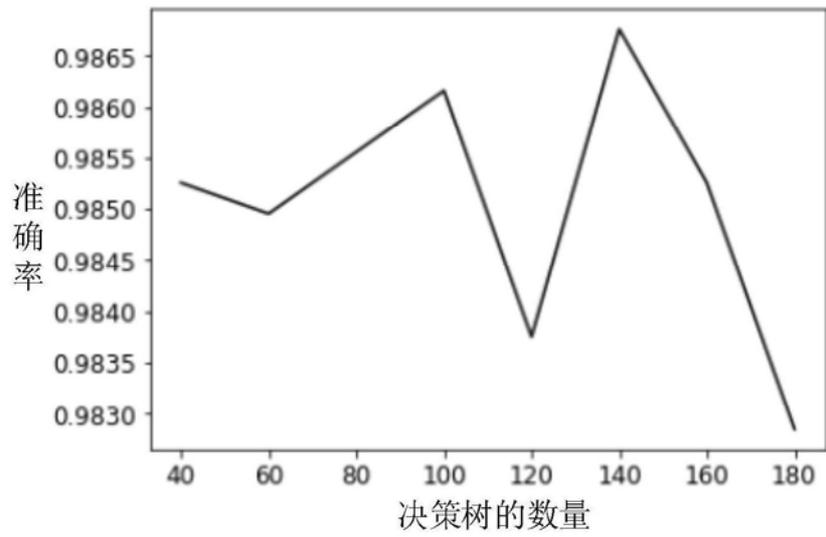


图7

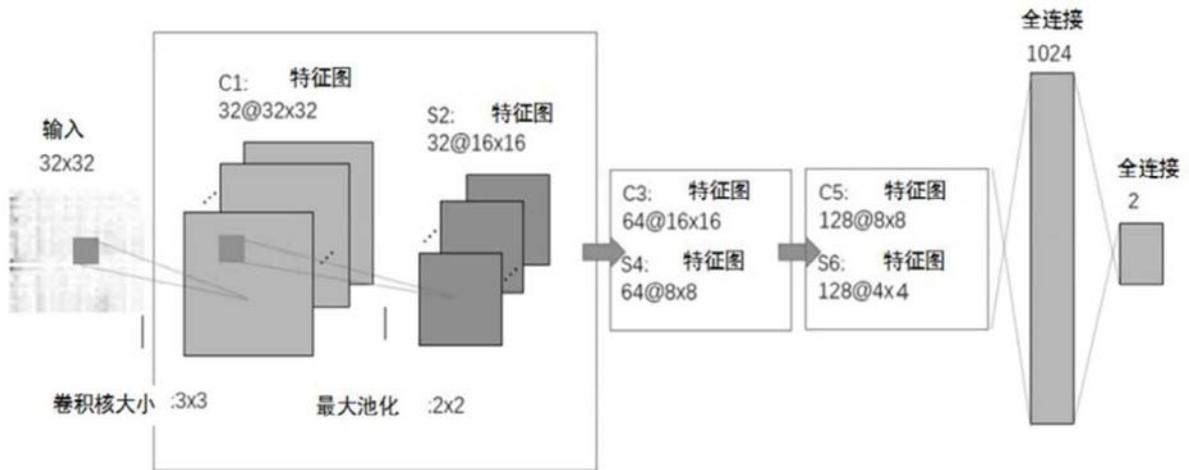


图8

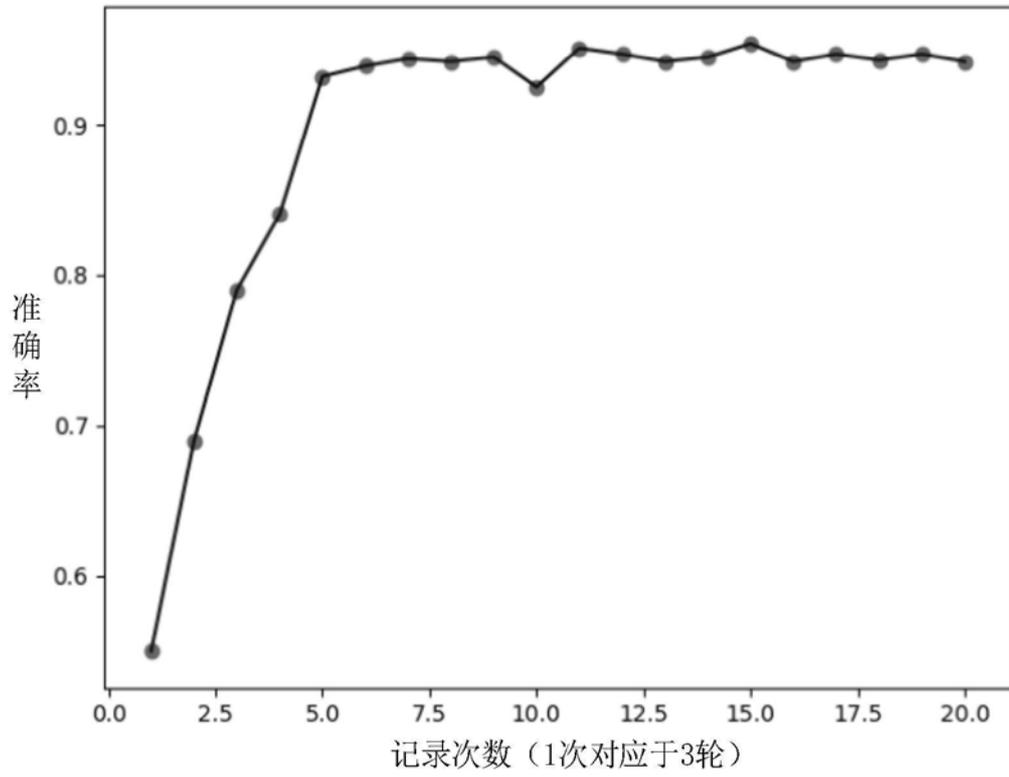


图9

专利名称(译)	基于随机森林和卷积神经网络的脑电数据分类方法		
公开(公告)号	CN110543832A	公开(公告)日	2019-12-06
申请号	CN201910745397.X	申请日	2019-08-13
[标]申请(专利权)人(译)	同济大学		
申请(专利权)人(译)	同济大学		
当前申请(专利权)人(译)	同济大学		
[标]发明人	何良华 任强		
发明人	何良华 任强		
IPC分类号	G06K9/00 G06K9/62 G06N3/04 G06N3/08 A61B5/00 A61B5/0476		
CPC分类号	A61B5/0476 A61B5/7264 A61B5/7282 G06K9/00503 G06K9/00523 G06K9/00536 G06K9/6232 G06K9/6282 G06N3/0454 G06N3/084		
代理人(译)	叶敏华		
外部链接	Espacenet SIPO		

摘要(译)

本发明涉及一种基于随机森林和卷积神经网络的脑电数据分类方法，包括：S1、采集原始的时序脑电数据，并将时序脑电数据转换为频域脑电数据；S2、基于随机森林方法，对频域脑电数据进行导联筛选，以对频域脑电数据进行降维处理；S3、构建卷积神经网络，将降维之后的频域脑电数据输入给卷积神经网络，对卷积神经网络进行训练和测试，得到训练好的卷积神经网络；S4、结合随机森林方法以及训练好的卷积神经网络，对经过傅里叶转换的实际脑电数据依次进行导联筛选和分类。与现有技术相比，本发明利用随机森林筛选重要的特征，降低了脑电数据维度，减少了计算量；结合卷积神经网络对降维的脑电数据进行拟合，保证了脑电数据分类准确率。

