



(12)发明专利申请

(10)申请公布号 CN 107016233 A

(43)申请公布日 2017.08.04

(21)申请号 201710148529.1

(22)申请日 2017.03.14

(71)申请人 中国科学院计算技术研究所

地址 100190 北京市海淀区中关村科学院
南路6号

(72)发明人 陈益强 胡春雨 忽丽莎 谢涛
王晋东

(74)专利代理机构 北京泛华伟业知识产权代理
有限公司 11280

代理人 王勇

(51)Int.Cl.

G06F 19/00(2011.01)

G06K 9/62(2006.01)

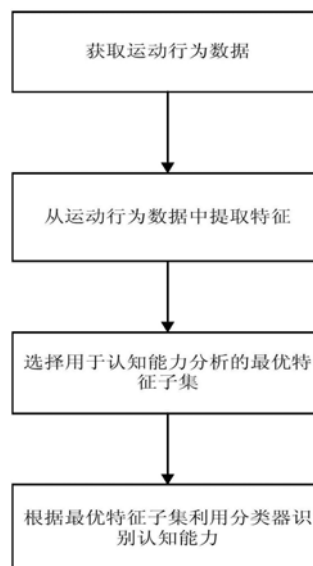
权利要求书2页 说明书8页 附图3页

(54)发明名称

运动行为和认知能力的关联分析方法及系统

(57)摘要

本发明提供一种运动行为和认知能力的关联分析方法。该方法包括：从运动行为数据中提取反映行为特点的运动特征集；采用迭代随机森林从所述运动特征集中选出用于检测认知能力的最优特征子集；利用所述最优特征子集采用分类器获得运动行为和认知能力的关联分析结果。根据本发明的方法，能够有效的分析运动行为和认知能力的关联，尤其对于小样本不均衡的数据集，能够提高认知障碍患者的识别率。



1. 一种运动行为和认知能力的关联分析方法, 包括:

步骤1: 从运动行为数据中提取反映行为特点的运动特征集;

步骤2: 采用迭代随机森林从所述运动特征集中选出用于分析认知能力的最优特征子集;

步骤3: 利用所述最优特征子集采用分类器获得运动行为和认知能力的关联分析结果。

2. 根据权利要求1所述的方法, 其中, 步骤2包括:

步骤21: 通过多次重复随机采样将所述运动特征集划分为多个训练集和多个测试集;

步骤22: 根据所述多个训练集构建多个随机森林;

步骤23: 在所述多个测试集上检验所构建的随机森林的性能, 根据每个特征出现在随机森林的个数和在随机森林中出现的频率选出一部分重要特征;

步骤24: 根据所述一部分重要特征重新构建随机森林, 重复步骤22和步骤23进行迭代, 以选出用于认知能力分析的最优特征子集。

3. 根据权利要求2所述的方法, 其中, 在步骤23中, 采用下式来选择所述一部分重要特征:

每个特征 f_j 出现的随机森林个数 N_j ; 和

每个特征 f_j 在随机森林中出现的频率 F_j ;

其中,

$$F_j = \sum_{s=1}^S \sum_{b=1}^B \frac{N_{bj}^s}{N_b^s}$$

N_{bj}^s 和 N_b^s 分别表示在S个随机森林的B棵树中分裂属性为 f_j 的内部节点数目以及所有内部节点的个数。

4. 根据权利要求2所述的方法, 其中, 步骤24包括:

将 $r = \left\lfloor n(t)/2 \right\rfloor$ 与预设阈值进行比较, 其中, $n(t)$ 为当前第t次迭代的特征数目, 如果r

大于所述阈值, 则继续进行迭代并将下一次迭代的特征数目 $n(t+1)$ 设置为r; 如果r小于所述阈值, 则终止迭代并选择当前的 $n(t)$ 个特征作为最优特征子集。

5. 根据权利要求1所述的方法, 其中, 步骤3包括以所述最优特征子集作为输入向量, 以是否患有认知能力疾病或认知能力等级作为目标向量, 利用加权极限学习机来获得运动行为和认知能力的关联分析结果。

6. 根据权利要求1所述的方法, 其中, 所述运动行为数据使用Kinect、智能手机、智能手表、肌电仪采集获得。

7. 根据权利要求1所述的方法, 其中, 所述运动特征集包括步态特征、灵活度特征和平衡性特征。

8. 根据权利要求1至7任一项所述的方法, 其中, 采用分类精度、灵敏度、特异性或几何均值来衡量所述迭代随机森林的性能。

9. 一种运动行为和认知能力的关联分析系统, 包括:

用于从运动行为数据中提取反映行为特点的运动特征集的装置;

用于采用迭代随机森林从所述运动特征集中选出用于检测认知能力的最优特征子集的装置；

用于利用所述最优特征子集采用分类器获得运动行为和认知能力的关联分析结果的装置。

10. 根据权利要求9所述的系统,还包括:

用于通过多次重复随机采样将所述运动特征集划分为多个训练集和多个测试集的装置;

用于根据所述多个训练集构建多个随机森林的装置;

用于在所述多个测试集上检验所构建的随机森林的性能,根据每个特征出现在随机森林的个数和在随机森林中出现的频率选出一部分重要特征的装置;

用于根据所述一部分重要特征重新构建随机森林进行迭代,以选出用于认知能力分析的最优特征子集的装置。

运动行为和认知能力的关联分析方法及系统

技术领域

[0001] 本发明涉及机器学习领域,尤其涉及进行运动行为与认知能力关联分析的方法及系统。

背景技术

[0002] 神经退行性疾病(Neurodegenerative disease)是威胁老年人生命健康的重大慢性疾病之一。传统的医学观点将神经退行性疾病分为两大类:一类是认知障碍疾病,这类疾病患者可能有记忆、语言或记忆力方面的障碍,而其运动功能是保留的;另一类是运动障碍疾病,这类疾病患者运动功能受损而认知功能却是完全保留的。近年来多个临床医学研究表明,认知障碍通常会伴随着运动功能混乱,而运动功能障碍也常常伴随有认知功能紊乱。

[0003] 在现有技术中,在认知能力评估中,很少存在有效进行运动行为评估的工具,例如,爱丁堡运动评估(Edinburgh Motor Assessment)主要应用于老年痴呆疾病的诊断,为认知障碍评估提供了简单的运动行为监测工具。又如,共济失调评定量表(SARA)以及统一帕金森评定量表(UPDRS)是目前认知障碍检测中被大家广泛接受的几种行为评估量表。然而,这些行为评测的结果通常是由医护人员进行记录和分析的,从而带来了一系列的问题,例如,测试记录工作十分繁琐,加重了医护人员的工作负担,以及医护人员的主观性通常影响行为评估测试的分析结果。

[0004] 此外,为了将运动行为与认知能力之间的关联分析自动化、客观化,计算机领域与医学领域的研究人员进行了很多共同的尝试。例如,将加速度、陀螺仪、磁力计等可穿戴传感器以及三维摄像机、智能手机等不同的传感器设备用来收集运动行为数据;将机器学习方法应用于运动行为与认知能力的关联分析等。

[0005] 然而,现有技术中的这些方法在通用性、针对性以及适用性方面尚存在一些不足。例如,由于在运动行为与认知能力的关联分析中,通常很难大规模采集数据,并且由于认知障碍患者数目通常远远小于非认知障碍患者,采集的数据还具有不均衡性的特点,传统的机器学习算法通常不适合具有不均衡特性的小样本数据集,因此,导致小类别(认知障碍患者)的识别精确度不高等问题。

发明内容

[0006] 因此,本发明的目的在于克服上述现有技术的缺陷,提供一种针对不均衡小样本数据集的运动行为与认知能力之间的关联分析方法,为挖掘运动行为与认知疾病之间的关联关系提供通用有效的手段。

[0007] 根据本发明的第一方面,提供了一种运动行为和认知能力的关联分析方法。该方法包括:

[0008] 步骤1:从运动行为数据中提取反映行为特点的运动特征集;

[0009] 步骤2:采用迭代随机森林从所述运动特征集中选出用于分析认知能力的最优特征子集;

[0010] 步骤3:利用所述最优特征子集采用分类器获得运动行为和认知能力的关联分析结果。

[0011] 在一个实施例中,步骤2包括:

[0012] 步骤21:通过多次重复随机采样将所述运动特征集划分为多个训练集和多个测试集;

[0013] 步骤22:根据所述多个训练集构建多个随机森林;

[0014] 步骤23:在所述多个测试集上检验所构建的随机森林的性能,根据每个特征出现在随机森林的个数和在随机森林中出现的频率选出一部分重要特征;

[0015] 步骤24:根据所述一部分重要特征重新构建随机森林,重复步骤22和步骤23进行迭代,以选出用于认知能力分析的最优特征子集。

[0016] 在一个实施例中,在步骤23中,采用下式来选择所述一部分重要特征:每个特征 f_j 出现的随机森林个数 N_j ;和每个特征 f_j 在随机森林中出现的频率 F_j ;

[0017] 其中,

$$[0018] \quad F_j = \sum_{s=1}^S \sum_{b=1}^B \frac{N_{bj}^s}{N_b^s}$$

[0019] N_{bj}^s 和 N_b^s 分别表示在S个随机森林的B棵树中分裂属性为 f_j 的内部节点数目以及所有内部节点的个数。

[0020] 在一个实施例中,步骤24包括:将 $r = \left\lfloor \frac{n(t)}{2} \right\rfloor$ 与预设阈值进行比较,其中, $n(t)$

为当前第t次迭代的特征数目,如果r大于所述阈值,则继续进行迭代并将下一次迭代的特征数目 $n(t+1)$ 设置为r;如果r小于所述阈值,则终止迭代并选择当前的 $n(t)$ 个特征作为最优特征子集。

[0021] 在一个实施例中,步骤3包括以所述最优特征子集作为输入向量,以是否患有认知能力疾病或认知能力等级作为目标向量,利用加权极限学习机来获得运动行为和认知能力的关联分析结果。

[0022] 在一个实施例中,所述运动行为数据使用Kinect、智能手机、智能手表、肌电仪采集获得。

[0023] 在一个实施例中,所述运动特征集包括步态特征、灵活度特征和平衡性特征。

[0024] 在一个实施例中,采用分类精度、灵敏度、特异性或几何均值来衡量所述迭代随机森林的性能。

[0025] 根据本发明的第二方面,提供了一种运动行为和认知能力的关联分析系统。该系统包括:用于从运动行为数据中提取反映行为特点的运动特征集的装置;用于采用迭代随机森林从所述运动特征集中选出用于检测认知能力的最优特征子集的装置;用于利用所述最优特征子集采用分类器获得运动行为和认知能力的关联分析结果的装置。

[0026] 在一个实施例中,本发明的系统还包括:用于通过多次重复随机采样将所述运动特征集划分为多个训练集和多个测试集的装置;用于根据所述多个训练集构建多个随机森林的装置;用于在所述多个测试集上检验所构建的随机森林的性能,根据每个特征出现在

随机森林的个数和在随机森林中出现的频率选出一部分重要特征的装置;用于根据所述一部分重要特征重新构建随机森林进行迭代,以选出用于认知能力分析的最优特征子集的装置。

[0027] 与现有技术相比,本发明的优点在于,能够有效地分析运动行为和认知能力之间的关联关系,尤其对于小样本不均衡的数据集,能够提高认知障碍患者的识别率。

附图说明

[0028] 以下附图仅对本发明作示意性的说明和解释,并不用于限定本发明的范围,其中:

[0029] 图1示出了根据本发明一个实施例的运动行为和认知能力关联分析的方法流程图。

[0030] 图2示出了根据本发明一个实施例的迭代随机森林方法的示意图。

[0031] 图3示出了具有L个隐层节点的单隐层前向神经网络的极限学习机的结构。

具体实施方式

[0032] 为了对本发明的技术特征、目的和效果有更加清楚的理解,现参照附图对本发明提出的运动行为和认知能力的关联分析方法作进一步详细说明。

[0033] 图1示出了根据本发明一个实施例的运动行为和认知能力关联分析的方法流程图。具体包括以下步骤:

[0034] 第一步、获取运动行为数据

[0035] 用于采集运动行为数据的设备有多种,例如,最常用的设备包括Kinect、智能手机、智能手表、肌电仪等。使用者可根据不同种类的认知疾病、采集目的等确定所要采集的运动行为种类,并选择适合的数据采集设备。例如,根据统一帕金森评定量表采集与帕金森疾病相关的运动行为;根据Tinetti运动测试(TMT, Tinetti Mobility Test)量表采集与脑部损伤(例如中风等)有关的运动行为

[0036] 采集到的运动行为数据可以呈现为不同的形式,例如,智能手表和智能手机以文本形式记录用户的加速度等运动数据;Kinect以图像的形式采集用户的骨骼数据,Kinect包含深度传感器和RGB摄像头,其采集到的数据主要分为两部分:深度传感器数据和RGB摄像头数据。其中,RGB摄像头数据可以用于动作的分割和标定,深度传感器数据则可以用于获取每种运动行为的骨骼关节坐标。

[0037] 第二步、从运动行为数据中提取运动特征

[0038] 运动行为特征提取的目的是,挖掘所获取的原始运动行为数据中的隐藏信息,从而挖掘运动行为与认知能力之间的潜在关联。

[0039] 通常,运动行为特征可分为两类:信号相关特征以及行为相关特征。信号相关特征主要从物理学角度分析数据的潜在特性,通常包括时域特征和频域特征,时域特征指所有与时间相关的特征,例如,均值、方差、过零率、最大值和最小值等。频域特征通常用于发现信号的周期特性,例如,直流分量、幅度以及功率谱密度等。而运动行为相关特征更侧重于挖掘动作的潜在属性。在本文中,主要以运动行为相关特征来阐述本发明的原理。

[0040] 在特征提取中,使用者可根据自身的目的和采集的运动行为数据,提取与其研究相关的特征。例如,根据采集的动作种类,可将运动行为相关特征划分为:步态特征,如步

长、步宽、步速等；身体平衡性特征，如起立速度、起立过程中躯干长度变化等；以及身体灵活性特征，如运动轨迹的平滑性、运动速度的变化等。

[0041] 在此步骤中，所选出的运动特征集能够更好的反映行为特点，从而构成了用于认知能力分析的原始运动特征集。

[0042] 第三步、选择用于认知能力分析的最优特征子集。

[0043] 此步骤的目的是选择出能够用于分析特定疾病的最优特征子集。

[0044] 为了适合具有小样本特性的认知障碍疾病的检测，在本发明中，采用迭代随机森林算法来衡量所提取原始运动特征集的重要性并从中选择出最有效的特征。图2示出了根据本发明一个实施例的迭代森林的算法的示意图。

[0045] 随机森林是由许多决策树集成在一起的分器，迭代随机森林是一个迭代的过程。概括而言，图2的过程包括：利用提取的原始特征集构建多个随机森林；在每次迭代过程中，根据特征在随机森林中出现的频率和次数对每个特征进行排序，以选择出一部分重要特征构建新的随机森林；最后，在分类性能不再提升时，选择出用于认知能力分析的最优特征子集。

[0046] 具体而言，迭代随机森林的算法包括以下步骤：

[0047] 步骤S210：初始化。

[0048] 在此步骤中，完成初始化训练集、输入向量以及衡量随机森林的分类性能的指标。

[0049] 用 $T = (x_i, t_i) \in R^{n(t)} \times R^m, i = 1, 2, \dots, N$ 表示初始训练集，其中 $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T$ 是具有 n 个特征的输入向量， $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T$ 是 m 个对应的目标向量。 $n(t)$ 是第 t 次迭代时的特征数目，在初始化阶段，输入向量的特征维数是 n ，即 $n(0) = n$ 。

[0050] 输入向量是上述根据运动行为数据提取的特征数据，以脑小血管疾病检测为例（脑小血管疾病是指病理改变主要累及颅内小血管的一组疾病），输入向量可包括步速、右步速均值、右步速方差、左步速均值、左步速方差、右步长均值、右步长方差、起立速度均值、起立速度方差、坐下速度均值、坐下速度方差等在内的特征。

[0051] 目标向量可设定为是否患脑小血管疾病，例如，用数值0表示未患病，用数值1表示患病。在另一个实施例中，针对不同的疾病检测问题，目标向量还可以设定为表示病情的严重程度，用病情等级来作为目标向量。根据实际需要设定不同的目标向量，可以实现认知能力全面有效的检测。

[0052] 简言之，初始化过程包括设置迭代索引 $t=0$ ，初始训练集 $T(0) = T$ ，初始化平均分类精度 $\overline{acc} = 0$ 等。

[0053] 应理解的是，在此实施例中，采用了平均分类精度来校验随机森林模型，即分类正确的样本占有所有样本的比例。而在另外的实施例中，还可以采用灵敏度、特异性或几何均值的准则来检验随机森林的性能。灵敏度是指患病样本被正确识别为患病样本占有所有被识别为患病样本的比例；特异性是指健康人被正确识别为健康人占有所有被识别为健康人的样本的比例；几何均值是灵敏度与特异性的乘积的平方根。根据认知能力疾病的检测特征来设置不同的检验准则，能够使迭代随机森林模型更有针对性。

[0054] 步骤220：随机采样以构建训练集和测试集。

[0055] 从训练集 $T(t)$ 中随机采样 S 次，每次随机采样 N_r 个样本，得到 S 个训练集 $T^s(t) = \{x_i, t_i\} \in R^{n(t)} \times R^m, i = 1, 2, \dots, N_r$ 以及 S 个测试集 $V^s(t) = \{x_i, t_i\} \in R^{n(t)} \times R^m, i = N_r + 1, N_r +$

$2, \dots, N, s=1, 2, \dots, S$ 。获得的训练集用来构建随机森林模型, 测试集用来评估随机森林模型的精确度。

[0056] 步骤S230: 利用训练集构建随机森林。

[0057] 根据步骤S220得到的 S 个训练集 $T^s(t)$ 构建 S 个随机森林, 将每个随机森林中的决策树数目设定为 B , 决策树中的每个分裂节点有 $\lfloor \sqrt{n(t)} \rfloor$ 个候选特征。

[0058] 由步骤S220和步骤S230可知, 通过对整个训练集重复采样, 可以获得多个训练集和多个测试集, 对每个训练集建立一个随机森林, 能够避免单次样本集划分而导致的随机性, 这种重复构建随机森林的方式, 更适合于小样本数据集。

[0059] 步骤S240: 利用测试集检验随机森林的精度。

[0060] 利用测试集 $V^s(t)$ 对构建的随机森林进行校验, 以得到平均分类精度 $\overline{acc}(t)$ 。如果 $\overline{acc}(t)$ 比 \overline{acc} 高, 将 \overline{acc} 设置为 $\overline{acc}(t)$ 并跳至步骤S250, 否则终止迭代过程, 并选择当前的 $\lfloor \sqrt{n(t)} \rfloor$ 个特征作为最重要的特征。

[0061] 步骤S250: 根据特征的重要性进行排序。

[0062] 此步骤的目的是, 根据对于认知能力评估的重要性将当前迭代的特征进行排序。

[0063] 在一个实施例中, 为了度量每个特征 $f_j, j=1, 2, \dots, n(t)$ 的重要性, 引入两个变量: 1) f_j 出现的随机森林个数 N_j ; 以及2) f_j 在随机森林中出现的频率 F_j , 其中 N_j 可以直接数出来, 而 F_j 可以通过公式(1)计算得到:

$$[0064] \quad F_j = \sum_{s=1}^S \sum_{b=1}^B \frac{N_{bj}^s}{N_b^s} \quad (1)$$

[0065] 其中, N_{bj}^s 和 N_b^s 分别表示在 S 个随机森林的 B 棵树中分裂属性为 f_j 的内部节点数目以及所有内部节点的个数。

[0066] 根据 N_j 和/或 F_j 对 $n(t)$ 个特征进行降序排列, 即可获得 $n(t)$ 个特征的重要程度, 这是由于一个比较重要的特征会在用不同的训练集构建不同的随机森林时被更多的森林更多的树选中, 因此, 具有较大的 N_j 和较大的 F_j 的特征更加重要。

[0067] 步骤260: 选择最优的特征子集。

[0068] 此步骤的目的在于, 根据上述排序结果进一步筛选出用于评估认知能力的最优特征子集, 以排除一部分非关键性的特征。

[0069] 例如, 选择排在前 r 位的特征作为最有效的特征, 其中, 较大的 r 值会导致冗余的循环, 而较小的 r 值可能会导致丢掉重要的特征。在一个实施例中, 将 r 值设定为当前特征数量的一半, 即 $r = \lfloor n(t)/2 \rfloor$, 如果 r 的值大于预设的特征的最小数目 n_{\min} , 将 $n(t+1)$ 设置为 r , 重新构建随机森林进行迭代, 否则, 终止迭代过程并选择当前 $n(t)$ 个特征作为最优特征子集。

[0070] 在实际的应用中, 可根据待分析的认知障碍的种类或样本的数目来设置合适的特征的最小数目。

[0071] 最后在判断为需要继续迭代的情况下, 设置 $t=t+1$, 并转至步骤S220, 执行循环迭代, 直到最终筛选出最优特征子集, 即随机森林的分类性能不再提高。

[0072] 通过迭代随机森林算法可以有效的筛选出用于评估认知障碍疾病的最优特征。下

表1示出了采用迭代随机森林方法选出的对脑小血管疾病检测最有效的5维特征。

[0073] 表1:采用迭代随机森林方法选出的特征

[0074]	特征类型	运动行为	特征描述
	灵活度特征	指鼻	只用右手食指指鼻的运动速度方差
	灵活度特征	下肢灵活度	完全抬起右腿的抬腿速度均值
	灵活度特征	快速轮替	右手手心向下抬起速度均值
	步态特征	三米步行转身 180°	右步结束帧左髋 z 角度均值
	步态特征	三米步行转身 180°	右步结束帧左髋 x 角度方差

[0075] 第四步、根据最优特征子集利用分类器来识别认知能力。

[0076] 为了根据选出的最优特征子集来构建认知能力识别的模型,可采用极限学习机、加权支持向量机、支持向量机等分类器。

[0077] 在一个实施例中,为了解决认知疾病数据集存在的不均衡问题,采用了加权极限学习机来进行认知能力分类。加权极限学习机是对极限学习机的一种改进。图3示意了极限学习机的结构,其是一种单隐层前向神经网络模型。

[0078] 一般性的,极限学习机网络输出函数为 $f(x) = h(x)\beta$,其中 $h(x)$ 是特征映射函数, β 是隐层节点输出函数,给定任意 N 个样本 $(x_i, t_i) \in R^n \times R^m, i = 1, \dots, N, x_i = (x_{i1}, \dots, x_{im})^T$ 是一个 $n \times 1$ 维的输入向量, $t_i = (t_{i1}, \dots, t_{im})^T$ 是一个 $m \times 1$ 维的目标向量,极限学习机的优化问题可以用下式表示:

$$\begin{aligned}
 & \min_{\beta, \xi} \frac{1}{2} \|\beta\|^2 + \frac{C}{2} \sum_{i=1}^N \|\xi_{:,i}\|^2 \\
 & \text{s.t. } \beta^T \cdot h(x_i) = t_i - \xi_{:,i}, i = 1, \dots, N
 \end{aligned} \quad (2)$$

[0079]

[0080] 其中, ξ 是一个 $m \times N$ 维的松弛变量, $\xi_{:,i}$ 表示 ξ 的第 i 列, C 表示惩罚参数,用于平衡最大泛化能力(即公式2的第一项)和最小训练误差(即公式2的第二项)。

[0081] 加权极限学习机在极限学习机的基础上引入了权重矩阵 W ,其优化函数用下式表述:

$$\begin{aligned}
 & \min_{\beta, \xi} L_{PWELM} = \frac{1}{2} \|\beta\|^2 + \frac{CW}{2} \sum_{i=1}^N \|\xi_{:,i}\|^2 \\
 & \text{s.t. } \beta^T \cdot h(x_i) = t_i - \xi_{:,i}, i = 1, \dots, N
 \end{aligned} \quad (3)$$

[0082]

[0083] 其中, W 是一个 $N \times N$ 维的对角权重矩阵, W_{ii} 表示样本 x_i 的权重。通常,当样本 x_i 属于一个小类别时, W_{ii} 相对较大;相反,当样本 x_i 属于一个大类别时, W_{ii} 的值则相对较小。此时的目标优化函数可以表述为最大化边缘距离以及最小化加权累计误差。

[0084] 例如,采用上述的加权极限学习机对脑小血管疾病进行检测时,其输入向量即是选出的有效特征子集,其输出向量可以设置为是否患脑小血管疾病或患病等级。其具体过程属于现有技术,在此不再赘述。

[0085] 为了解决样本不均衡带来的小类别分类效果差的问题,可通过权重来调整样本对训练误差的贡献度。在一个实施例,将权重设定为相应类别样本个数的倒数,例如,病人数目为20,健康人数目为80,则将病人的权重设定为 $1/20$,将健康人的权重设定为 $1/80$ 。

[0086] 此外,在加权极限学习机中,还可设置隐层节点个数 L 和惩罚参数 C ,对于这两个参数,通常可以采用网格搜索方法进行设定,其中, C 的取值范围为 $\{2^{-24}, 2^{-23}, \dots, 2^{-24}, 2^{-25}\}$, L 的取值范围为 $\{2^0, 2^1, \dots, 2^{12}\}$,选取分类性能最好的实验结果对应的一组 (C, L) 的取值作为实验参数。

[0087] 综上所述,本发明采用的迭代随机森林方法,可以从原始特征集中选择出用于评估认知能力的最优特征子集,在该最优特征子集上所构建的分类或回归模型可以达到与特征选择前近似甚至更好的预测精度,其更适合于数据集规模很小的认知障碍疾病检测;进一步地,通过加权极限学习机进行认知能力分类,可以解决样本不均衡带来的小类别分类效果差的问题,从而提高了疾病检测的效率和准确性。

[0088] 为了验证根据本发明实施例的运动行为与认知关联分析方法的有效性,以运动行为与脑小血管疾病关联分析为例进行了实验。

[0089] 实验内容包括:使用的数据集分两个场景采集,在医疗环境中采集了20个人的数据,包括8个脑小血管疾病患者,12个健康人,在正常生活环境中采集了31个健康人的数据;从采集的数据集中提取了包括步态特征、平衡特征和灵活性特征在内的739维特征;针对医疗环境下采集的20个人的数据,根据迭代随机森林算法选出了最有效的5维特征;为了验证这5维特征对脑小血管疾病检测的有效性,采用极限学习机和支持向量机两种算法作为分类器对医疗环境下和正常环境中采集的共51人进行分类实验,与WEKA(怀卡托智能分析环境)中内嵌的六种特征选择方法所选出的五维特征进行对比,实验结果如表2所示。从表2可以看出,无论是采用极限学习机还是支持向量机作为分类器,迭代随机森林算法选出的5维特征比其他特征选择方法选出的5维特征更具代表性,对脑小血管疾病的检测更加有效。

[0090] 表2:各种特征选择方法的对比

[0091]

特征选择方法	支持向量机	极限学习机
方法1	90.67%	90.66%
方法2	87.79%	88.60%
方法3	86.74%	85.56%
方法4	87.36%	88.65%
方法5	89.01%	88.33%
方法6	86.74%	85.49%
迭代随机森林	90.79%	91.43%

[0092] 注:

[0093] 方法1:根据单个属性和类别的相关性进行选择。

[0094] 方法2:根据与分类有关的每一个属性的增益比进行评估。

[0095] 方法3:根据与分类有关的每一个属性的信息增益进行评估。

[0096] 方法4:根据OneR分类器评估属性。

[0097] 方法5:通过反复测试一个实例和其同类或不同类中最近的实例上的属性值进行评估。

[0098] 方法6:根据与分类有关的每一个属性的对称不稳定性进行评估。

[0099] 进一步地,为了说明加权极限学习机的有效性,采用极限学习机、支持向量机和加权支持向量机作为对比算法,以分类精度、灵敏度、特异性、几何均值作为性能测试标准,测试结果如表3所示。由表3可以看出,采用迭代随机森林算法选出的特征进行脑小血管疾病检测精度可以达到90%以上,此外,采用加权极限学习机进行分类在测试精度、测试敏感度、测试几何均值方面比其他三种方法表现要更好,更适用于不均衡数据集。

[0100] 表3:脑小血管疾病检测精度对比结果

[0101]

分类模型	分类精度	敏感度	特异性	几何均值
加权极限学习机	95.24%	94.73%	95.35%	94.69%
极限学习机	94.02%	78.93%	97.50%	86.18%
加权支持向量机	91.90%	87.27%	92.97%	89.63%
支持向量机	94.57%	81.77%	97.52%	88.55%

[0102] 上述实验结果表明,本文所提出的运动行为与认知能力关联分析方法能够有效挖掘出运动行为与认知能力之间的关联关系,为认知疾病的诊断工作提供有效的手段。

[0103] 以上已经描述了本发明的各实施例,上述说明是示例性的,并非穷尽性的,并且也不限于所披露的各实施例。在不偏离所说明的各实施例的范围和精神的情况下,对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择,旨在最好地解释各实施例的原理、实际应用或对市场中的技术改进,或者使本技术领域的其它普通技术人员能理解本文披露的各实施例。

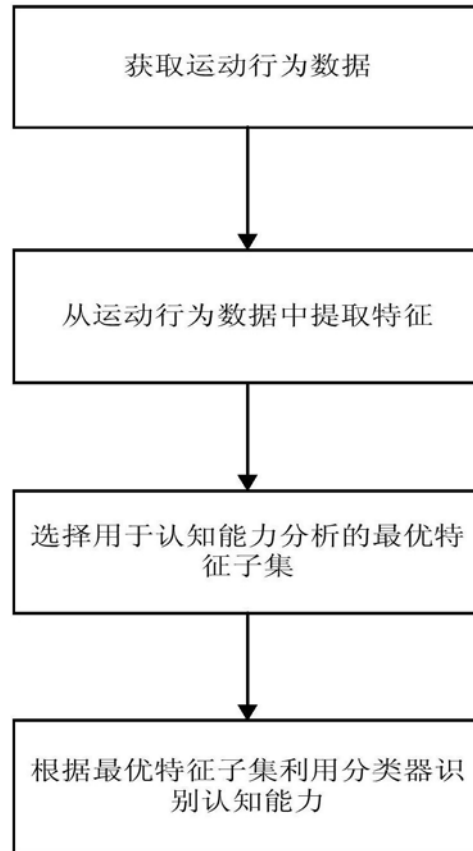


图1

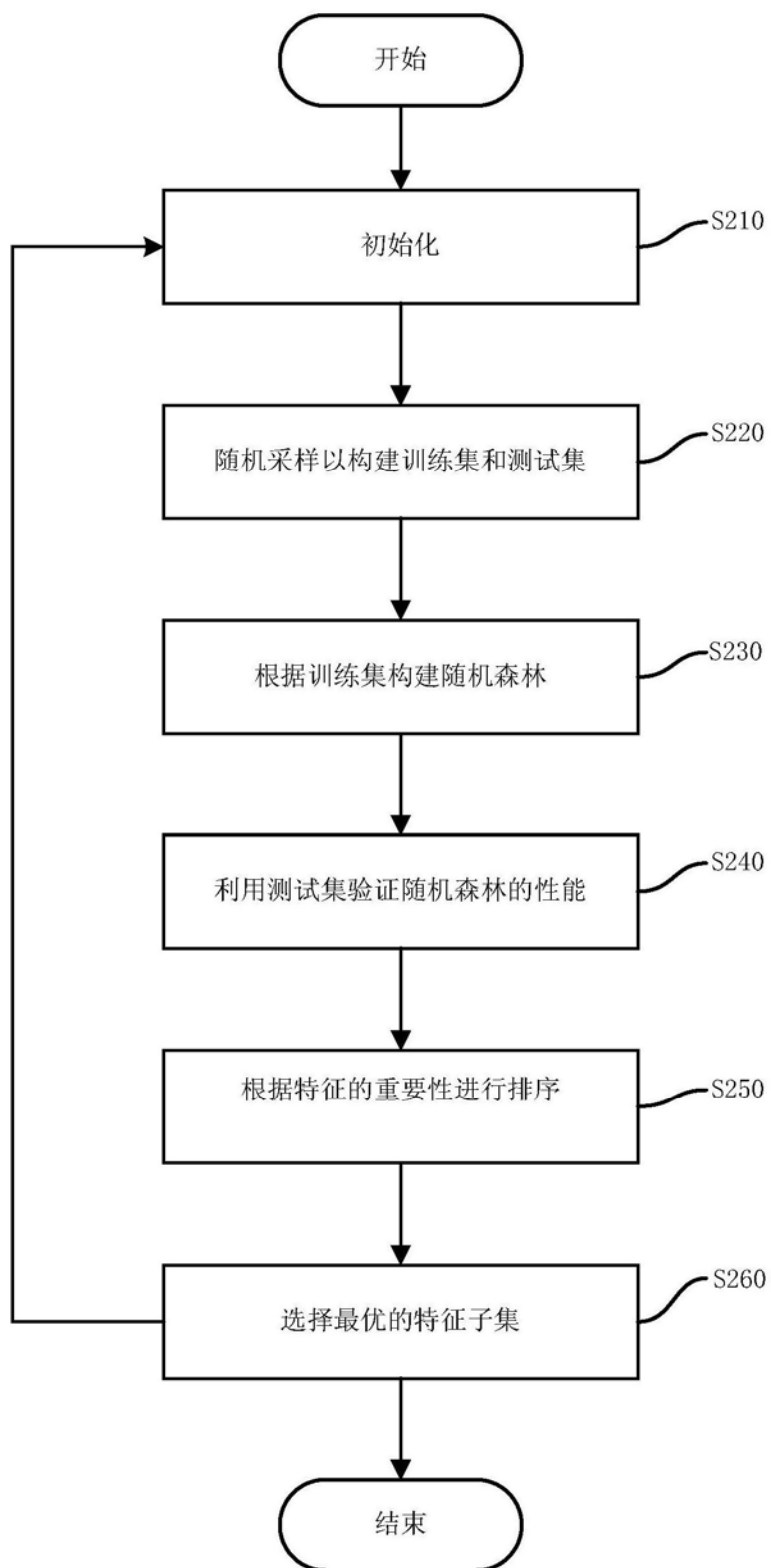


图2

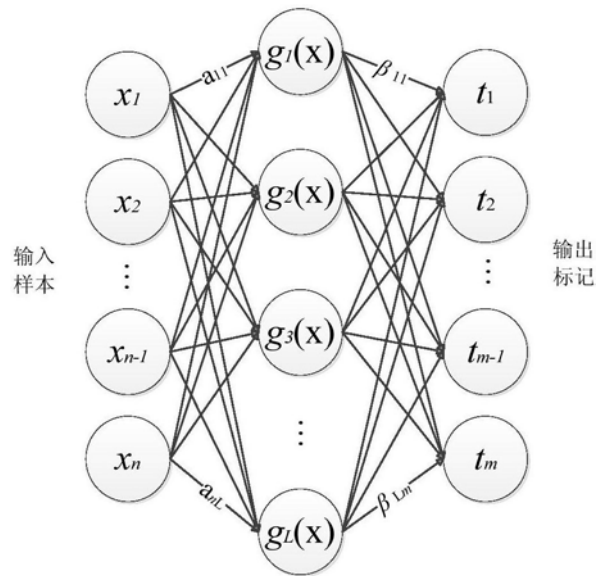


图3