

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2021 年 2 月 25 日 (25.02.2021)



(10) 国际公布号
WO 2021/031817 A1

(51) 国际专利分类号:
G06K 9/62 (2006.01) **G06K 9/00** (2006.01)

(21) 国际申请号: PCT/CN2020/105630

(22) 国际申请日: 2020 年 7 月 29 日 (29.07.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910775783.3 2019年8月21日 (21.08.2019) CN

(71) 申请人: 深圳壹账通智能科技有限公司(ONE CONNECT SMART TECHNOLOGY CO., LTD. (SHENZHEN)) [CN/CN]; 中国广东省深圳市前海深港合作区前湾一路1号A栋201室, Guangdong 518000 (CN)。

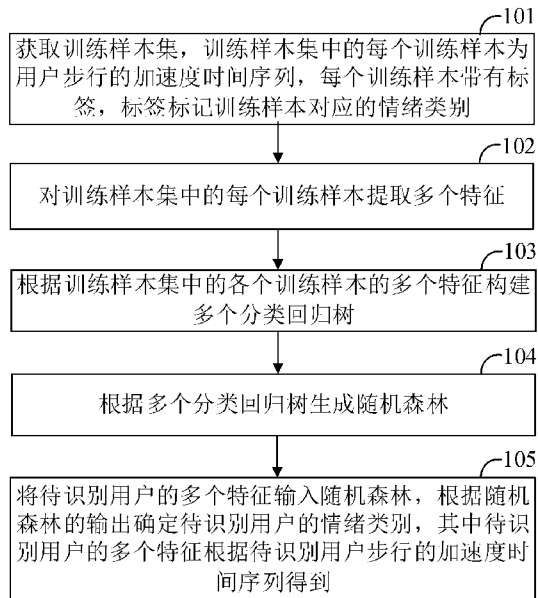
(72) 发明人: 刘利 (LIU, Li); 中国广东省深圳市前海深港合作区前湾一路1号A栋201室, Guangdong 518000 (CN)。

(74) 代理人: 深圳市赛恩倍吉知识产权代理有限公司(SHENZHEN SCIENBIZIP INTELLECTUAL PROPERTY AGENCY CO.,LTD.); 中国广东省深圳市龙华新区龙观东路83号荣群大厦9楼, Guangdong 518109 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,

(54) Title: EMOTION RECOGNITION METHOD AND DEVICE, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 情绪识别方法、装置、计算机装置及存储介质



101 Obtain a training sample set, each training sample in the training sample set being an acceleration time sequence of walking of a user, each training sample having a label, and the label labeling an emotion category corresponding to the training sample

102 Extract a plurality of features of each training sample in the training sample set

103 Construct a plurality of classification and regression trees according to the plurality of features of each training sample in the training sample set

104 Generate a random forest according to the plurality of classification and regression trees

105 Input a plurality of features of a user to be recognized to the random forest, and determine an emotion category of said user according to the output of the random forest, wherein the plurality of features of said user is obtained according to the acceleration time sequence of the walking of said user

图 1

(57) Abstract: The present application provides an emotion recognition method and device, a computer device, and a storage medium. The emotion recognition method comprises: obtaining a training sample set, each training sample in the training sample set being an acceleration time sequence of walking of a user, each training sample having a label, and the label labeling an emotion category corresponding to the training sample; extracting a plurality of features of each training sample in the training sample set; constructing a plurality of classification and regression trees according to the plurality of features of each training sample in the training sample set; generating a random forest according to the plurality of classification and regression trees; inputting a plurality of features of a user to be

MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

recognized to the random forest, and determining an emotion category of said user according to the output of the random forest, wherein the plurality of features of said user is obtained according to the acceleration time sequence of the walking of said user. The present application implements the recognition of an emotion of a user according to acceleration data during a walking process of the user.

(57) 摘要: 本申请提供了一种情绪识别方法、装置、计算机装置及存储介质。所述情绪识别方法包括: 获取训练样本集, 所述训练样本集中的每个训练样本为用户步行的加速度时间序列, 每个训练样本带有标签, 所述标签标记所述训练样本对应的情绪类别; 对所述训练样本集中的每个训练样本提取多个特征; 根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树; 根据所述多个分类回归树生成随机森林; 将待识别用户的多个特征输入所述随机森林, 根据所述随机森林的输出确定所述待识别用户的情绪类别, 其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。本申请实现了根据用户步行过程中的加速度数据识别用户情绪。

情绪识别方法、装置、计算机装置及存储介质

本申请要求于 2019 年 8 月 21 日提交中国专利局、申请号为 201910775783.3、发明名称为“情绪识别方法、装置、计算机装置及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及人工智能技术领域，具体涉及一种情绪识别方法、装置、计算机装置及存储介质。

背景技术

随着人工智能技术的进步，情绪识别已是目前人工智能领域中最活跃的研究课题之一。其目的是对人类的图像序列进行检测、跟踪和识别，更科学地解释人类行为。情绪识别可以应用于生活的各个方面：游戏厂商可以智能分析玩家的情绪，根据不同表情针对性地和玩家交互，提高游戏的体验；相机厂商可以利用该项技术捕捉人类表情，比如当需要一张微笑或者生气的照片时，可以捕获被拍人员的面部表情并快速完成拍照工作；政府或社会学家可以在公共场合安装摄像头，分析整个社会群体的表情和肢体动作以了解人们的生活工作压力；商厦可以根据顾客对商品的购物时的动作及表情视频，对产品做相关的市场调查。

在实际应用中，发明人意识到单纯基于人脸表情的情绪识别研究已遇到瓶颈，一方面，基于实验室视角的正面人脸表情识别已达到极高识别率，但相关算法在应用于自然态人脸表情识别时却识别率较低；另一方面，在运动中人的肢体动作和情绪也有着强相关关系，肢体动作同样是人们获取情绪的重要线索，在很多应用场合中，能够为情绪识别提供有效的帮助。因此，如果能从人的肢体动作识别出人的情绪变化，是对情绪识别技术的一个重大补充，并对今后人类情感智能识别相关应用的发展具有重要价值。

发明内容

鉴于以上内容，有必要提出一种情绪识别方法、装置、计算机装置及存储介质，其可以提高情绪识别的场景适应性。

本申请的第一方面提供一种情绪识别方法，其中，所述方法包括：

获取训练样本集，所述训练样本集中的每个训练样本为用户步行的加速度时间序列，每个训练样本带有标签，所述标签标记所述训练样本对应的情绪类别；

对所述训练样本集中的每个训练样本提取多个特征；

根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树；

根据所述多个分类回归树生成随机森林；

将待识别用户的多个特征输入所述随机森林，根据所述随机森林的输出确定所述待识别用户的情绪类别，其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

本申请的第二方面提供一种计算机装置，其中，所述计算机装置包括处理器，所述处理器用于执行存储器中存储的计算机可读指令以实现以下步骤：

获取训练样本集，所述训练样本集中的每个训练样本为用户步行的加速度时间序列，每个训练样本带有标签，所述标签标记所述训练样本对应的情绪类别；

对所述训练样本集中的每个训练样本提取多个特征；

根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树；

根据所述多个分类回归树生成随机森林；

将待识别用户的多个特征输入所述随机森林，根据所述随机森林的输出确定所述待识别用户的情绪类别，其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

本申请的第三方面提供一种存储介质，所述存储介质上存储有计算机可读指令，其中，所述计算机可读指令被处理器执行时实现以下步骤：

获取训练样本集，所述训练样本集中的每个训练样本为用户步行的加速度时间序列，每个训练样本带有标签，所述标签标记所述训练样本对应的情绪类别；

对所述训练样本集中的每个训练样本提取多个特征；

根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树；

根据所述多个分类回归树生成随机森林；

将待识别用户的多个特征输入所述随机森林，根据所述随机森林的输出确定所述待识别用户的情绪类别，其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

本申请的第四方面提供一种情绪识别装置，其中，所述装置包括：

获取模块，用于获取训练样本集，所述训练样本集中的每个训练样本为用户步行的加速度时间序列，每个训练样本带有标签，所述标签标记所述训练样本对应的情绪类别；

提取模块，用于对所述训练样本集中的每个训练样本提取多个特征；

构建模块，用于根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树；

生成模块，用于根据所述多个分类回归树生成随机森林；

识别模块，用于将待识别用户的多个特征输入所述随机森林，根据所述随机森林的输出确定所述待识别用户的情绪类别，其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

本申请以带有情绪类别标签的用户步行的加速度时间序列为训练样本，根据各个训练样本生成随机森林，利用所述随机森林对待识别用户的加速度时间序列进行识别。本申请实现了根据用户步行过程中的加速度数据识别用户的情绪。

附图说明

图 1 是本申请实施例提供的情绪识别方法的流程图。

图 2 是本申请实施例提供的情绪识别装置的结构图。

图 3 是本申请实施例提供的计算机装置的示意图。

具体实施方式

为了能够更清楚地理解本申请的上述目的、特征和优点，下面结合附图和具体实施例对本申请进行详细描述。需要说明的是，在不冲突的情况下，本申请的实施例及实施例中的特征可以相互组合。

在下面的描述中阐述了很多具体细节以便于充分理解本申请，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

除非另有定义，本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同。本文中在本申请的说明书中所使用的术语只是为了描述具体的实施例的目的，不是旨在于限制本申请。

优选地，本申请的情绪识别方法应用在一个或者多个计算机装置中。所述计算机装置是一种能够按照事先设定或存储的指令，自动进行数值计算和/或信息处理的设备，其硬件包括但不限于微处理器、专用集成电路(Application Specific Integrated Circuit, ASIC)、可编程门阵列(Field-Programmable Gate Array, FPGA)、数字处理器(Digital Signal Processor, DSP)、嵌入式设备等。

所述计算机装置可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。所述计算机装置可以与用户通过键盘、鼠标、遥控器、触摸板或声控设备等方式进行人机交互。

实施例一

图1是本申请实施例一提供的情绪识别方法的流程图。所述情绪识别方法应用于计算机装置。

本申请情绪识别方法涉及机器学习，用于根据用户步行过程中的加速度数据识别所述用户的情绪。

如图1所示，所述情绪识别方法包括：

101，获取训练样本集，所述训练样本集中的每个训练样本为用户步行的加速度时间序列，每个训练样本带有标签，所述标签标记所述训练样本对应的情绪类别。

可以在预设时间内，通过用户的手腕和/或脚踝上的加速度传感器采集用户步行过程中加速度数据，根据所述加速度数据得到所述加速度时间序列。每个加速度时间序列可以包括预设数量的加速度数据，例如100个加速度数据。或者，每个加速度时间序列可以包括预设时间（例如60秒）内的加速度数据。所述加速度数据可以是X轴、Y轴或Z轴方向的加速度数据，从而得到X轴、Y轴或Z轴方向的加速度时间序列。

例如，通过用户步行过程中用户手腕上的加速度传感器采集预设数量（例如100个）的X轴方向的加速度数据，将采集的预设数量的X轴方向的加速度数据组成一个加速度时间序列，得到一个训练样本。又如，通过用户步行过程中用户脚踝上的加速度传感器按照预设时间间隔采集预设时长（例如60秒）内X轴方向的加速度数据，将采集的预设时长内X轴方向的加速度数据组成一个加速度时间序列，得到一个训练样本。

每个训练样本对应一个标签，用于标识情绪类别。所述情绪类别可以包括正面情绪（激动、开心）、中性情绪（平静）或负面情绪（悲伤、哀愁）。所述标签可以是数字，例如1、2、3。例如，若用户的情绪为正面情绪，对应的标签为3；若用户的情绪为中性情绪，对应的标签为2；若用户的情绪为负面情绪，对应的标签为1。

用户处于不同情绪时，用户步行的加速度数据不同。可以在用户不同情绪时采集用户的加速度数据，得到带有不同标签的训练样本。

通过采集用户步行的加速度数据获得的多个训练样本组成所述训练样本集。

所述训练样本集可以包括多个用户的训练样本，即多个用户步行的加速度时间序列。或者，所述训练样本集可以包括一个用户的训练样本，即一个用户步行的加速度时间序列。

102，对所述训练样本集中的每个训练样本提取多个特征。

对所述训练样本集中的每个训练样本提取多个特征是对每个训练样本提取多个相同的特征。

所述多个特征可以包括加速度时间序列的标准差、平均值、峰值、偏态系数、FFT系数、功率谱密度平均值、功率谱密度标准偏差、坐标轴系数。

加速度时间序列的偏态系数是加速度时间序列分布不对称的度量值。若一个训练样本是对称的，则所述偏态系数等于0；若一个训练样本是左偏的，则所述偏态系数小于0；若一个训练样本是右偏的，则所述偏态系数大于0。

加速度时间序列的FFT系数是对所述加速度时间序列进行FFT（Fast Fourier

Transform, 快速傅里叶变换) 变换得到的系数, 可以取第 2 维到第 32 维的 FFT 系数。

若训练样本为 X 轴方向的加速度时间序列, 则对应的坐标轴系数为:

$$\rho_{YZ} = \frac{\text{cov}(Y, Z)}{\sqrt{D(Y)}\sqrt{D(Z)}}$$

其中, $\text{cov}(Y, Z)$ 为所述训练样本 Y 轴方向的加速度时间序列与所述训练样本 Z 轴方向的加速度时间序列的协方差, $D(Y)$ 为所述训练样本 Y 轴方向的加速度时间序列的方差, $D(Z)$ 为所述训练样本 Z 轴方向的加速度时间序列的方差;

若所述训练样本为 Y 轴方向的加速度时间序列, 则所述坐标轴系数为 ρ_{XZ} , ρ_{XZ} 的计算公式可以参照上述 ρ_{YZ} ;

若所述训练样本为 Z 轴方向的加速度时间序列, 则所述坐标轴系数为 ρ_{XY} , ρ_{XY} 的计算公式可以参照上述 ρ_{YZ} 。

在本实施例中, 可以对所述训练样本集中的每个训练样本的多个特征进行归一化处理, 得到每个训练样本归一化后的多个特征。

所述对所述训练样本集中的每个训练样本的多个特征进行归一化处理可以包括:

选择所述训练样本集中的第 i 个训练样本;

选择所述第 i 个训练样本的第 j 个特征;

计算所述训练样本集中所述第 j 个特征的均值 U_j 和方差 σ_j ;

对所述第 i 个训练样本的第 j 个特征进行归一化计算:

$$B_{ij} = \frac{b_{ij} - U_j}{\sigma_j}$$

其中, B_{ij} 为所述第 i 个训练样本的第 j 个特征归一化后的值, b_{ij} 为所述第 i 个训练样本的第 j 个特征归一化前的值。i=1,2,...,N, N 为所述训练样本集中训练样本的数量。j=1,2,...,M; M 为每个训练样本的特征的数量。

所述第 i 个训练样本的第 j 个特征是指所述第 i 个训练样本的多个特征中的第 j 个特征。

在对所述训练样本集中的每个训练样本提取多个特征之前, 所述方法还包括:

对所述训练样本集中的每个训练样本进行预处理。

所述对每个训练样本集中的每个训练样本进行预处理包括:

对所述训练样本进行降噪; 和/或

对所述训练样本中的缺失值进行填充; 和/或

对所述训练样本中的异常值进行修正。

具体地, 对所述训练样本进行降噪可以包括: 对所述训练样本进行移动平均降噪。

可以按照以下公式对训练样本进行移动平均降噪:

$$\text{output}[i] = \frac{1}{w} \sum_{j=0}^{w-1} \text{input}[i+j]$$

其中, $\text{output}[i]$ 是所述训练样本 (即加速度时间序列) 中第 i 个加速度数据对应的输出, w 为常量, 取值为 3 或 5, $\text{input}[i+j]$ 是所述训练样本中第 i+j 个加速度数据。

还可以采用其他方法对所述训练样本进行降噪。例如, 可以对所述训练样本进行小波降噪。

具体地, 所述对所述训练样本中的缺失值进行填充可以包括: 取所述训练样本中所述缺失值的前后若干个加速度数据 (例如缺失值的前 5 个和后 5 个加速度数据), 用所述缺失值的前后若干个加速度数据的平均值填充所述缺失值。或者, 可以采用 K-最近邻算法, 确定距离有缺失值的训练样本最近的 K 个训练样本 (例如根据欧式距离确定距离有缺失值的训练样本最近的 K 个训练样本), 将 K 个训练样本的数据加权平均来估计该训练样本的缺失值。或者, 可以采用其他方法填补所述缺失值。例如, 可以通过回归拟合

的方法或者插值法对所述缺失值进行填补。

对所述训练样本中的异常值进行修正的方法可以与填补缺失值的方法相同。例如，可以取所述训练样本中所述异常值的前后若干个加速度数据（例如异常值的前 5 个和后 5 个加速度数据），用所述异常值的前后若干个加速度数据的平均值修正所述异常值。或者，可以采用 K-最近邻算法，确定距离有异常值的训练样本最近的 K 个训练样本（例如根据欧式距离确定距离有异常值的训练样本最近的 K 个训练样本），将 K 个训练样本的数据加权平均来估计该训练样本的异常值。或者，可以采用其他方法修正所述异常值。例如，可以通过回归拟合的方法或者插值法对所述异常值进行修正。

可以理解，修正异常值的方法可以不同于填补缺失值的方法。

103，根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树。

在一具体实施例中，所述根据所述训练样本集的各个训练样本的多个特征构建多个分类回归树可以包括：

- (1) 从所述训练样本集中随机选取 Q 个训练样本作为待分类样本；
- (2) 从所述待分类样本的多个特征中随机选取 K 个特征；
- (3) 确定所述待分类样本的所述 K 个特征中最优的切分特征和切分点，根据所述最优的切分特征和切分点将所述待分类样本划分为两个子集；
- (4) 计算划分的每个子集中的待分类样本的标签的均值；
- (5) 对于划分的每个子集，重复执行步骤 (2) - (4)，直至满足预设停止条件。

所述最优的切分特征和切分点可以根据如下目标函数确定：

$$\min_{j,s} \left[\sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right]$$

上式表示遍历待分类样本的 K 个特征（即切分特征 j）的所有特征值（即切分点 s），根据平方误差最小准则找到最优的切分特征和切分点。

其中， x_i 是所述待分类样本中的第 i 个训练样本， y_i 是 x_i 的标签。

R_1 、 R_2 是根据待分类样本的 K 个特征中的一个切分特征和一个切分点得到的两个子集， $R_1(j,s) = \{x|x^{(j)} \leq s\}$ ， $R_2(j,s) = \{x|x^{(j)} > s\}$ ， $x^{(j)}$ 是待分类样本的第 j 个特征的特征值。也就是说， $R_1(j,s)$ 是第 j 个特征的特征值小于或等于 s 的待分类样本的集合， $R_2(j,s) = \{x|x^{(j)} > s\}$ 是第 j 个特征的特征值大于 s 的待分类样本的集合。

$$c_1 = \frac{1}{N_1} \sum_{x_1 \in R_1(j,s)} y_1, \quad c_2 = \frac{1}{N_2} \sum_{x_2 \in R_2(j,s)} y_2,$$

其中， N_1 是子集 R_1 中待分类样本的个数， N_2 是子集 R_2 中待分类样本的个数。

所述满足预设停止条件可以包括：

若所述子集中待分类样本的数量小于或等于第一预设值，则满足预设停止条件；或者

若所述最优的切分特征和切分点对应的平方误差小于第二预设值，则满足预设停止条件，其中

$$r = \sum_{x_i \in R_1(j^*,s^*)} (y_i - c_1)^2 + \sum_{x_i \in R_2(j^*,s^*)} (y_i - c_2)^2; \text{ 或者}$$

若随机选取的所有 K 个特征中不同特征的数量大于或等于第三预设值，则满足预设停止条件。

执行一次上述步骤 (1) - (5)，可以得到一个分类回归树。也就是说，每次从所有训练样本中随机选取 Q 个训练样本作为待分类样本（即执行步骤 (1)），对所述待分类样本进行逐层划分（即执行步骤 (2) - (5)），根据各次划分的切分特征和切分点，得到一个分类回归树。分类回归树的根节点对应初始的待分类样本，分类回归树的每个叶子节点对应一个不再划分的子集。分类回归树的输出为叶子节点对应的输出，即划分到叶子节点的待分类样本的标签的均值。

重复执行上述步骤 (1) - (5)，即可得到多个分类回归树。

104, 根据所述多个分类回归树生成随机森林。

将多个分类回归树组成所述随机森林, 不同分类回归树之间相互独立。所述随机森林的输入就是所述随机森林中每个分类回归树的输入; 所述随机森林的输出为所述随机森林中所有分类回归树的输出的平均值。

在本实施例中, 所述根据所述多个分类回归树生成随机森林包括:

对所述多个分类回归树进行剪枝处理;

根据剪枝处理后的所述多个分类回归树生成所述随机森林。

对所述多个分类回归树进行剪枝处理包括:

(1) 从所述多个分类回归树选择一个分类回归树记为 T_0 ;

(2) 初始化参数: $k=0$, $T=T_0$, $\alpha = +\infty$;

(3) 在分类回归树 T 中自下而上地对各内部节点 (非叶子节点) t 计算 $C(T_t)$, $|T_t|$ 以及

$$g(t) = \frac{C(t) - C(T_t)}{|T_t| - 1}$$

$$\alpha = \min(\alpha, g(t))$$

其中, T_t 表示以 t 为根节点的子树, $C(t)$ 是根据划分到内部节点 t 的待分类样本得到的预测误差, $C(T_t)$ 是根据划分到子树 T_t 的待分类样本得到的预测误差, $C(t)$ 是根据划分到 t 节点的待分类样本得到的预测误差, $|T_t|$ 是所述子树 T_t 的叶子节点的数量;

(4) 自上而下地遍历内部节点 t , 若 $g(t)=\alpha$, 进行剪枝, 剪去所述节点 t 的子树, t 变为叶子节点, 剔除剪去的子树的 $g(t)$, $\alpha = \min(g(t))$, 并计算叶子节点 t 对应区域中的待分类样本的标签的均值, 得到分类回归树 T ;

(5) 对参数进行赋值: $k=k+1$, $T_k=T$;

(6) 若 T 不是由根节点单独构成的树, 则回到步骤 (4);

(7) 采用交叉验证法在子树序列 $T_0, T_1, \dots, T_n$ 中选择最优子树 T_α 。

105, 将待识别用户的多个特征输入所述随机森林, 根据所述随机森林的输出确定所述待识别用户的情绪类别, 其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

具体地, 所述随机森林中的每个分类回归树以待识别用户的多个特征作为输入, 根据所述待识别用户的多个特征对所述待识别用户进行分类, 得到所述分类回归树的输出; 计算所述随机森林中所有分类回归树的输出的平均值, 得到所述随机森林的输出; 根据所述随机森林的输出确定所述待识别用户的情绪类别。

可以选择与所述随机森林的输出差值最小的标签对应的情绪类别作为所述待识别用户的情绪类别。

待识别用户可以包含在训练样本对应的用户中。例如, 所述训练样本集包括用户 A 的训练样本, 所述待识别用户为用户 A。或者, 所述训练样本集包括用户 A、用户 B、用户 C、用户 D 的训练样本, 所述待识别用户为用户 A。

或者, 待识别用户可以不包含在训练样本对应的用户中。例如, 所述训练样本集包括用户 A、用户 B、用户 C、用户 D 的训练样本, 所述待识别用户为用户 E。

实施例一的情绪识别方法以带有情绪类别标签的用户步行的加速度时间序列为训练样本, 根据各个训练样本生成随机森林, 利用所述随机森林对待识别用户的加速度时间序列进行识别。实施例一实现了根据用户步行过程中的加速度数据识别所述用户的情绪。

实施例二

图 2 是本申请实施例二提供的情绪识别装置的结构图。所述情绪识别装置 20 应用于计算机装置。所述情绪识别装置 20 根据用户步行过程中的加速度数据识别所述用户的情绪。如图 2 所示, 所述情绪识别装置 20 可以包括获取模块 201、提取模块 202、构建

模块 203、生成模块 204、识别模块 205。

获取模块 201，用于获取训练样本集，所述训练样本集中的每个训练样本为用户步行的加速度时间序列，每个训练样本带有标签，所述标签标记所述训练样本对应的情绪类别。

可以在预设时间内，通过用户的手腕和/或脚踝上的加速度传感器采集用户步行过程中加速度数据，根据所述加速度数据得到所述加速度时间序列。每个加速度时间序列可以包括预设数量的加速度数据，例如 100 个加速度数据。或者，每个加速度时间序列可以包括预设时间（例如 60 秒）内的加速度数据。所述加速度数据可以是 X 轴、Y 轴或 Z 轴方向的加速度数据，从而得到 X 轴、Y 轴或 Z 轴方向的加速度时间序列。

例如，通过用户步行过程中用户手腕上的加速度传感器采集预设数量（例如 100 个）的 X 轴方向的加速度数据，将采集的预设数量的 X 轴方向的加速度数据组成一个加速度时间序列，得到一个训练样本。又如，通过用户步行过程中用户脚踝上的加速度传感器按照预设时间间隔采集预设时长（例如 60 秒）内 X 轴方向的加速度数据，将采集的预设时长内 X 轴方向的加速度数据组成一个加速度时间序列，得到一个训练样本。

每个训练样本对应一个标签，用于标识情绪类别。所述情绪类别可以包括正面情绪（激动、开心）、中性情绪（平静）或负面情绪（悲伤、哀愁）。所述标签可以是数字，例如 1、2、3。例如，若用户的情绪为正面情绪，对应的标签为 3；若用户的情绪为中性情绪，对应的标签为 2；若用户的情绪为负面情绪，对应的标签为 1。

用户处于不同情绪时，用户步行的加速度数据不同。可以在用户不同情绪时采集用户的加速度数据，得到带有不同标签的训练样本。

通过采集用户步行的加速度数据获得的多个训练样本组成所述训练样本集。

所述训练样本集可以包括多个用户的训练样本，即多个用户步行的加速度时间序列。或者，所述训练样本集可以包括一个用户的训练样本，即一个用户步行的加速度时间序列。

提取模块 202，用于对所述训练样本集中的每个训练样本提取多个特征。

对所述训练样本集中的每个训练样本提取多个特征是对每个训练样本提取多个相同的特征。

所述多个特征可以包括加速度时间序列的标准差、平均值、峰值、偏态系数、FFT 系数、功率谱密度平均值、功率谱密度标准偏差、坐标轴系数。

加速度时间序列的偏态系数是加速度时间序列分布不对称的度量值。若一个训练样本是对称的，则所述偏态系数等于 0；若一个训练样本是左偏的，则所述偏态系数小于 0；若一个训练样本是右偏的，则所述偏态系数大于 0。

加速度时间序列的 FFT 系数是对所述加速度时间序列进行 FFT（Fast Fourier Transform，快速傅里叶变换）变换得到的系数，可以取第 2 维到第 32 维的 FFT 系数。

若训练样本为 X 轴方向的加速度时间序列，则对应的坐标轴系数为：

$$\rho_{YZ} = \frac{\text{cov}(Y, Z)}{\sqrt{D(Y)}\sqrt{D(Z)}}$$

其中， $\text{cov}(Y, Z)$ 为所述训练样本 Y 轴方向的加速度时间序列与所述训练样本 Z 轴方向的加速度时间序列的协方差， $D(Y)$ 为所述训练样本 Y 轴方向的加速度时间序列的方差， $D(Z)$ 为所述训练样本 Z 轴方向的加速度时间序列的方差；

若所述训练样本为 Y 轴方向的加速度时间序列，则所述坐标轴系数为 ρ_{XZ} ， ρ_{XZ} 的计算公式可以参照上述 ρ_{YZ} ；

若所述训练样本为 Z 轴方向的加速度时间序列，则所述坐标轴系数为 ρ_{XY} ， ρ_{XY} 的计算公式可以参照上述 ρ_{YZ} 。

在本实施例中，可以对所述训练样本集中的每个训练样本的多个特征进行归一化处理，得到每个训练样本归一化后的多个特征。

所述对所述训练样本集中的每个训练样本的多个特征进行归一化处理可以包括：

选择所述训练样本集中的第 i 个训练样本；

选择所述第 i 个训练样本的第 j 个特征；

计算所述训练样本集中所述第 j 个特征的均值 U_j 和方差 σ_j ；

对所述第 i 个训练样本的第 j 个特征进行归一化计算：

$$B_{ij} = \frac{b_{ij} - U_j}{\sigma_j}$$

其中， B_{ij} 为所述第 i 个训练样本的第 j 个特征归一化后的值， b_{ij} 为所述第 i 个训练样本的第 j 个特征归一化前的值。 $i=1,2,\dots,N$ ， N 为所述训练样本集中训练样本的数量。 $j=1,2,\dots,M$ ； M 为每个训练样本的特征的数量。

所述第 i 个训练样本的第 j 个特征是指所述第 i 个训练样本的多个特征中的第 j 个特征。

在对所述训练样本集中的每个训练样本提取多个特征之前，所述方法还包括：

对所述训练样本集中的每个训练样本进行预处理。

所述对每个训练样本集中的每个训练样本进行预处理包括：

对所述训练样本进行降噪；和/或

对所述训练样本中的缺失值进行填充；和/或

对所述训练样本中的异常值进行修正。

具体地，对所述训练样本进行降噪可以包括：对所述训练样本进行移动平均降噪。

可以按照以下公式对训练样本进行移动平均降噪：

$$\text{output}[i] = \frac{1}{w} \sum_{j=0}^{w-1} \text{input}[i+j]$$

其中， $\text{output}[i]$ 是所述训练样本（即加速度时间序列）中第 i 个加速度数据对应的输出， w 为常量，取值为 3 或 5， $\text{input}[i+j]$ 是所述训练样本中第 $i+j$ 个加速度数据。

还可以采用其他方法对所述训练样本进行降噪。例如，可以对所述训练样本进行小波降噪。

具体地，所述对所述训练样本中的缺失值进行填充可以包括：取所述训练样本中所述缺失值的前后若干个加速度数据（例如缺失值的前 5 个和后 5 个加速度数据），用所述缺失值的前后若干个加速度数据的平均值填充所述缺失值。或者，可以采用 K-最近邻算法，确定距离有缺失值的训练样本最近的 K 个训练样本（例如根据欧式距离确定距离有缺失值的训练样本最近的 K 个训练样本），将 K 个训练样本的数据加权平均来估计该训练样本的缺失值。或者，可以采用其他方法填补所述缺失值。例如，可以通过回归拟合的方法或者插值法对所述缺失值进行填补。

对所述训练样本中的异常值进行修正的方法可以与填补缺失值的方法相同。例如，可以取所述训练样本中所述异常值的前后若干个加速度数据（例如异常值的前 5 个和后 5 个加速度数据），用所述异常值的前后若干个加速度数据的平均值修正所述异常值。或者，可以采用 K-最近邻算法，确定距离有异常值的训练样本最近的 K 个训练样本（例如根据欧式距离确定距离有异常值的训练样本最近的 K 个训练样本），将 K 个训练样本的数据加权平均来估计该训练样本的异常值。或者，可以采用其他方法修正所述异常值。例如，可以通过回归拟合的方法或者插值法对所述异常值进行修正。

可以理解，修正异常值的方法可以不同于填补缺失值的方法。

构建模块 203，用于根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树。

在一具体实施例中，所述根据所述训练样本集的各个训练样本的多个特征构建多个分类回归树可以包括：

- (1) 从所述训练样本集中随机选取 Q 个训练样本作为待分类样本；
 - (2) 从所述待分类样本的多个特征中随机选取 K 个特征；
 - (3) 确定所述待分类样本的所述 K 个特征中最优的切分特征和切分点，根据所述最优的切分特征和切分点将所述待分类样本划分为两个子集；
 - (4) 计算划分的每个子集中的待分类样本的标签的均值；
 - (5) 对于划分的每个子集，重复执行步骤 (2) - (4)，直至满足预设停止条件。
- 所述最优的切分特征和切分点可以根据如下目标函数确定：

$$\min_{j,s} \left[\sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right]$$

上式表示遍历待分类样本的 K 个特征 (即切分特征 j) 的所有特征值 (即切分点 s)，根据平方误差最小准则找到最优的切分特征和切分点。

其中， x_i 是所述待分类样本中的第 i 个训练样本， y_i 是 x_i 的标签。

R_1 、 R_2 是根据待分类样本的 K 个特征中的一个切分特征和一个切分点得到的两个子集， $R_1(j,s) = \{x | x^{(j)} \leq s\}$ ， $R_2(j,s) = \{x | x^{(j)} > s\}$ ， $x^{(j)}$ 是待分类样本的第 j 个特征的特征值。也就是说， $R_1(j,s)$ 是第 j 个特征的特征值小于或等于 s 的待分类样本的集合， $R_2(j,s) = \{x | x^{(j)} > s\}$ 是第 j 个特征的特征值大于 s 的待分类样本的集合。

$$c_1 = \frac{1}{N_1} \sum_{x_i \in R_1(j,s)} y_i, \quad c_2 = \frac{1}{N_2} \sum_{x_i \in R_2(j,s)} y_i,$$

其中， N_1 是子集 R_1 中待分类样本的个数， N_2 是子集 R_2 中待分类样本的个数。

所述满足预设停止条件可以包括：

若所述子集中待分类样本的数量小于或等于第一预设值，则满足预设停止条件；或者

若所述最优的切分特征和切分点对应的平方误差小于第二预设值，则满足预设停止条件，其中

$$r = \sum_{x_i \in R_1(j^*,s^*)} (y_i - c_1)^2 + \sum_{x_i \in R_2(j^*,s^*)} (y_i - c_2)^2; \text{ 或者}$$

若随机选取的所有 K 个特征中不同特征的数量大于或等于第三预设值，则满足预设停止条件。

执行一次上述步骤 (1) - (5)，可以得到一个分类回归树。也就是说，每次从所有训练样本中随机选取 Q 个训练样本作为待分类样本 (即执行步骤 (1))，对所述待分类样本进行逐层划分 (即执行步骤 (2) - (5))，根据各次划分的切分特征和切分点，得到一个分类回归树。分类回归树的根节点对应初始的待分类样本，分类回归树的每个叶子节点对应一个不再划分的子集。分类回归树的输出为叶子节点对应的输出，即划分到叶子节点的待分类样本的标签的均值。

重复执行上述步骤 (1) - (5)，即可得到多个分类回归树。

生成模块 204，用于根据所述多个分类回归树生成随机森林。

将多个分类回归树组成所述随机森林，不同分类回归树之间相互独立。所述随机森林的输入就是所述随机森林中每个分类回归树的输入；所述随机森林的输出为所述随机森林中所有分类回归树的输出的平均值。

在本实施例中，所述根据所述多个分类回归树生成随机森林包括：

对所述多个分类回归树进行剪枝处理；

根据剪枝处理后的所述多个分类回归树生成所述随机森林。

对所述多个分类回归树进行剪枝处理包括：

(1) 从所述多个分类回归树选择一个分类回归树记为 T_0 ；

(2) 初始化参数： $k=0$ ， $T=T_0$ ， $\alpha = +\infty$ ；

(3) 在分类回归树 T 中自下而上地对各内部节点 (非叶子节点) t 计算 $C(T_t)$ ， $|T_t|$

以及

$$g(t) = \frac{C(t) - C(T_t)}{|T_t| - 1}$$

$$\alpha = \min(\alpha, g(t))$$

其中, T_t 表示以 t 为根节点的子树, $C(t)$ 是根据划分到内部节点 t 的待分类样本得到的预测误差, $C(T_t)$ 是根据划分到子树 T_t 的待分类样本得到的预测误差, $C(t)$ 是根据划分到 t 节点的待分类样本得到的预测误差, $|T_t|$ 是所述子树 T_t 的叶子节点的数量;

(4) 自上而下地遍历内部节点 t , 若 $g(t)=\alpha$, 进行剪枝, 剪去所述节点 t 的子树, t 变为叶子节点, 剔除剪去的子树的 $g(t)$, $\alpha = \min(g(t))$, 并计算叶子节点 t 对应区域中的待分类样本的标签的均值, 得到分类回归树 T ;

(5) 对参数进行赋值: $k=k+1$, $T_k=T$;

(6) 若 T 不是由根节点单独构成的树, 则回到步骤 (4);

(7) 采用交叉验证法在子树序列 $T_0, T_1, \dots, T_n$ 中选择最优子树 T_α 。

识别模块 205, 用于将待识别用户的多个特征输入所述随机森林, 根据所述随机森林的输出确定所述待识别用户的情绪类别, 其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

具体地, 所述随机森林中的每个分类回归树以待识别用户的多个特征作为输入, 根据所述待识别用户的多个特征对所述待识别用户进行分类, 得到所述分类回归树的输出; 计算所述随机森林中所有分类回归树的输出的平均值, 得到所述随机森林的输出; 根据所述随机森林的输出确定所述待识别用户的情绪类别。

可以选择与所述随机森林的输出差值最小的标签对应的情绪类别作为所述待识别用户的情绪类别。

待识别用户可以包含在训练样本对应的用户中。例如, 所述训练样本集包括用户 A 的训练样本, 所述待识别用户为用户 A。或者, 所述训练样本集包括用户 A、用户 B、用户 C、用户 D 的训练样本, 所述待识别用户为用户 A。

或者, 待识别用户可以不包含在训练样本对应的用户中。例如, 所述训练样本集包括用户 A、用户 B、用户 C、用户 D 的训练样本, 所述待识别用户为用户 E。

实施例二的情绪识别装置 20 以带有情绪类别标签的用户步行的加速度时间序列为训练样本, 根据各个训练样本生成随机森林, 利用所述随机森林对待识别用户的加速度时间序列进行识别。实施例二实现了根据用户步行过程中的加速度数据识别所述用户的情绪。

实施例三

本实施例提供一种存储介质, 该存储介质上存储有计算机可读指令, 该计算机可读指令被处理器执行时实现上述情绪识别方法实施例中的步骤, 例如图 1 所示的 101-105:

101, 获取训练样本集, 所述训练样本集中的每个训练样本为用户步行的加速度时间序列, 每个训练样本带有标签, 所述标签标记所述训练样本对应的情绪类别;

102, 对所述训练样本集中的每个训练样本提取多个特征;

103, 根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树;

104, 根据所述多个分类回归树生成随机森林;

105, 将待识别用户的多个特征输入所述随机森林, 根据所述随机森林的输出确定所述待识别用户的情绪类别, 其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

或者, 该计算机可读指令被处理器执行时实现上述装置实施例中各模块的功能, 例如图 2 中的模块 201-205:

获取模块 201, 用于获取训练样本集, 所述训练样本集中的每个训练样本为用户步行的加速度时间序列, 每个训练样本带有标签, 所述标签标记所述训练样本对应的情绪类别;

提取模块 202，用于对所述训练样本集中的每个训练样本提取多个特征；

构建模块 203，用于根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树；

生成模块 204，用于根据所述多个分类回归树生成随机森林；

识别模块 205，用于将待识别用户的多个特征输入所述随机森林，根据所述随机森林的输出确定所述待识别用户的情绪类别，其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

实施例四

图 3 为本申请实施例四提供的计算机装置的示意图。所述计算机装置 30 包括存储器 301、处理器 302 以及存储在所述存储器 301 中并可在所述处理器 302 上运行的计算机程序 303，例如情绪识别程序。所述处理器 302 执行所述计算机程序 303 时实现上述情绪识别方法实施例中的步骤，例如图 1 所示的 101-105：

101，获取训练样本集，所述训练样本集中的每个训练样本为用户步行的加速度时间序列，每个训练样本带有标签，所述标签标记所述训练样本对应的情绪类别；

102，对所述训练样本集中的每个训练样本提取多个特征；

103，根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树；

104，根据所述多个分类回归树生成随机森林；

105，将待识别用户的多个特征输入所述随机森林，根据所述随机森林的输出确定所述待识别用户的情绪类别，其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

或者，该计算机程序被处理器执行时实现上述装置实施例中各模块的功能，例如图 2 中的模块 201-205：

获取模块 201，用于获取训练样本集，所述训练样本集中的每个训练样本为用户步行的加速度时间序列，每个训练样本带有标签，所述标签标记所述训练样本对应的情绪类别；

提取模块 202，用于对所述训练样本集中的每个训练样本提取多个特征；

构建模块 203，用于根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树；

生成模块 204，用于根据所述多个分类回归树生成随机森林；

识别模块 205，用于将待识别用户的多个特征输入所述随机森林，根据所述随机森林的输出确定所述待识别用户的情绪类别，其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

示例性的，所述计算机程序 303 可以被分割成一个或多个模块，所述一个或者多个模块被存储在所述存储器 301 中，并由所述处理器 302 执行，以完成本方法。所述一个或多个模块可以是能够完成特定功能的一系列计算机可读指令段，该指令段用于描述所述计算机程序 303 在所述计算机装置 30 中的执行过程。例如，所述计算机程序 303 可以被分割成图 2 中的获取模块 201、提取模块 202、构建模块 203、生成模块 204、识别模块 205，各模块具体功能参见实施例二。

所述计算机装置 30 可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。本领域技术人员可以理解，所述示意图 3 仅仅是计算机装置 30 的示例，并不构成对计算机装置 30 的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件，例如所述计算机装置 30 还可以包括输入输出设备、网络接入设备、总线等。

所称处理器 302 可以是中央处理单元(Central Processing Unit, CPU)，还可以是其他通用处理器、数字信号处理器(Digital Signal Processor, DSP)、专用集成电路(Application Specific Integrated Circuit, ASIC)、现成可编程门阵列(Field-Programmable Gate Array,

FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器 302 也可以是任何常规的处理器等, 所述处理器 302 是所述计算机装置 30 的控制中心, 利用各种接口和线路连接整个计算机装置 30 的各个部分。

所述存储器 301 可用于存储所述计算机程序 303, 所述处理器 302 通过运行或执行存储在所述存储器 301 内的计算机程序或模块, 以及调用存储在存储器 301 内的数据, 实现所述计算机装置 30 的各种功能。所述存储器 301 可主要包括存储程序区和存储数据区, 其中, 存储程序区可存储操作系统、至少一个功能所需的应用程序(比如声音播放功能、图像播放功能等)等; 存储数据区可存储根据计算机装置 30 的使用所创建的数据(比如音频数据)等。此外, 存储器 301 可以包括易失性和非易失性存储器, 例如硬盘、内存、插接式硬盘, 智能存储卡(Smart Media Card, SMC), 安全数字(Secure Digital, SD)卡, 闪存卡(Flash Card)、至少一个磁盘存储器件、闪存器件或其他存储器件。

所述计算机装置 30 集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用, 可以存储在一个计算机可读存储介质中。基于这样的理解, 本申请实现上述实施例方法中的全部或部分流程, 也可以通过计算机程序来指令相关的硬件来完成, 所述的计算机程序可存储于一存储介质中, 该计算机程序在被处理器执行时, 可实现上述各个方法实施例的步骤。其中, 所述计算机程序包括计算机可读指令代码, 所述计算机可读指令代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读存储介质可以包括: 能够携带所述计算机可读指令代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、存储器、只读存储器(ROM, Read-Only Memory)、随机存储器(RAM, Random Access Memory)等。所述计算机可读存储介质可以是非易失性, 也可以是易失性的。

在本申请所提供的几个实施例中, 应该理解到, 所揭露的系统, 装置和方法, 可以通过其它的方式实现。例如, 以上所描述的装置实施例仅仅是示意性的, 例如, 所述模块的划分, 仅仅为一种逻辑功能划分, 实际实现时可以有另外的划分方式。

所述作为分离部件说明的模块可以是或者也可以不是物理上分开的, 作为模块显示的部件可以是或者也可以不是物理模块, 即可以位于一个地方, 或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

另外, 在本申请各个实施例中的各功能模块可以集成在一个处理模块中, 也可以是各个模块单独物理存在, 也可以两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现, 也可以采用硬件加软件功能模块的形式实现。

上述以软件功能模块的形式实现的集成的模块, 可以存储在一个计算机可读存储介质中。上述软件功能模块存储在一个存储介质中, 包括若干指令用以使得一台计算机设备(可以是个人计算机, 服务器, 或者网络设备等)或处理器(processor)执行本申请各个实施例所述方法的部分步骤。

对于本领域技术人员而言, 显然本申请不限于上述示范性实施例的细节, 而且在不背离本申请的精神或基本特征的情况下, 能够以其他的具体形式实现本申请。因此, 无论从哪一点来看, 均应将实施例看作是示范性的, 而且是非限制性的, 本申请的范围由所附权利要求而不是上述说明限定, 因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本申请内。不应将权利要求中的任何附关联图标记视为限制所涉及的权利要求。此外, 显然“包括”一词不排除其他模块或步骤, 单数不排除复数。系统权利要求中陈述的多个模块或装置也可以由一个模块或装置通过软件或者硬件来实现。第一, 第二等词语用来表示名称, 而并不表示任何特定的顺序。

最后应说明的是, 以上实施例仅用以说明本申请的技术方案而非限制, 尽管参照较佳实施例对本申请进行了详细说明, 本领域的普通技术人员应当理解, 可以对本申请的

技术方案进行修改或等同替换，而不脱离本申请技术方案的精神和范围。

权 利 要 求 书

1. 一种情绪识别方法, 其中, 所述方法包括:

获取训练样本集, 所述训练样本集中的每个训练样本为用户步行的加速度时间序列, 每个训练样本带有标签, 所述标签标记所述训练样本对应的情绪类别;

对所述训练样本集中的每个训练样本提取多个特征;

根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树;

根据所述多个分类回归树生成随机森林;

将待识别用户的多个特征输入所述随机森林, 根据所述随机森林的输出确定所述待识别用户的情绪类别, 其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

2. 如权利要求 1 所述的情绪识别方法, 其中, 所述多个特征包括如下各项的任意组合:

加速度时间序列的标准差、平均值、峰值、偏态系数、FFT 系数、功率谱密度平均值、功率谱密度标准偏差、坐标轴系数。

3. 如权利要求 1 所述的情绪识别方法, 其中, 所述方法还包括:

对所述训练样本集中的每个训练样本的多个特征进行归一化处理, 得到每个训练样本归一化后的多个特征;

所述根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树包括:

根据每个训练样本归一化后的多个特征构建所述多个分类回归树。

4. 如权利要求 1 所述的情绪识别方法, 其中, 所述对所述训练样本集中的每个训练样本提取多个特征之前, 所述方法还包括:

对所述训练样本进行降噪; 和/或

对所述训练样本中的缺失值进行填充; 和/或

对所述训练样本中的异常值进行修正。

5. 如权利要求 4 所述的情绪识别方法, 其中, 所述对所述训练样本进行降噪包括:

按照以下公式对所述训练样本进行移动平均降噪:

$$\text{output}[i] = \frac{1}{w} \sum_{j=0}^{w-1} \text{input}[i+j]$$

其中, $\text{output}[i]$ 是所述训练样本中第 i 个加速度数据对应的输出, w 为常量, 取值为 3 或 5, $\text{input}[i+j]$ 是所述训练样本中第 $i+j$ 个加速度数据。

6. 如权利要求 1 所述的情绪识别方法, 其中, 所述根据所述训练样本集的各个训练样本的多个特征构建多个分类回归树包括:

从所述训练样本集中随机选取 Q 个训练样本作为待分类样本;

从所述待分类样本的多个特征中随机选取 K 个特征;

确定所述待分类样本的所述 K 个特征中最优的切分特征和切分点, 根据所述最优的切分特征和切分点将所述待分类样本划分为两个子集;

计算划分的每个子集中的待分类样本的标签的均值;

对于划分的每个子集, 重复执行所述从所述待分类样本的多个特征中随机选取 K 个特征至所述计算划分的每个子集中的待分类样本的标签的均值, 直至满足预设停止条件。

7. 如权利要求 1 所述的情绪识别方法, 其中, 所述根据所述多个分类回归树生成随机森林包括:

对所述多个分类回归树进行剪枝处理;

根据剪枝处理后的所述多个分类回归树生成所述随机森林。

8. 一种计算机装置, 其中, 所述计算机装置包括处理器, 所述处理器用于执行存储器中存储的计算机可读指令以实现以下步骤:

获取训练样本集，所述训练样本集中的每个训练样本为用户步行的加速度时间序列，每个训练样本带有标签，所述标签标记所述训练样本对应的情绪类别；

对所述训练样本集中的每个训练样本提取多个特征；

根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树；

根据所述多个分类回归树生成随机森林；

将待识别用户的多个特征输入所述随机森林，根据所述随机森林的输出确定所述待识别用户的情绪类别，其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

9. 如权利要求 8 所述的计算机装置，其中，所述多个特征包括如下各项的任意组合：

加速度时间序列的标准差、平均值、峰值、偏态系数、FFT 系数、功率谱密度平均值、功率谱密度标准偏差、坐标轴系数。

10. 如权利要求 8 所述的计算机装置，其中，所述处理器执行所述计算机可读指令还用以实现以下步骤：

对所述训练样本集中的每个训练样本的多个特征进行归一化处理，得到每个训练样本归一化后的多个特征；

所述处理器执行所述计算机可读指令以实现所述根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树时，具体包括：

根据每个训练样本归一化后的多个特征构建所述多个分类回归树。

11. 如权利要求 8 所述的计算机装置，其中，所述处理器执行所述计算机可读指令以实现所述对所述训练样本集中的每个训练样本提取多个特征之前，还用以实现以下步骤：

对所述训练样本进行降噪；和/或

对所述训练样本中的缺失值进行填充；和/或

对所述训练样本中的异常值进行修正。

12. 如权利要求 11 所述的计算机装置，其中，所述处理器执行所述计算机可读指令以实现所述对所述训练样本进行降噪时，具体包括：

按照以下公式对所述训练样本进行移动平均降噪：

$$\text{output}[i] = \frac{1}{w} \sum_{j=0}^{w-1} \text{input}[i+j]$$

其中， $\text{output}[i]$ 是所述训练样本中第 i 个加速度数据对应的输出， w 为常量，取值为 3 或 5， $\text{input}[i+j]$ 是所述训练样本中第 $i+j$ 个加速度数据。

13. 如权利要求 8 所述的计算机装置，其中，所述处理器执行所述计算机可读指令以实现所述根据所述训练样本集的各个训练样本的多个特征构建多个分类回归树时，具体包括：

从所述训练样本集中随机选取 Q 个训练样本作为待分类样本；

从所述待分类样本的多个特征中随机选取 K 个特征；

确定所述待分类样本的所述 K 个特征中最优的切分特征和切分点，根据所述最优的切分特征和切分点将所述待分类样本划分为两个子集；

计算划分的每个子集中的待分类样本的标签的均值；

对于划分的每个子集，重复执行所述从所述待分类样本的多个特征中随机选取 K 个特征至所述计算划分的每个子集中的待分类样本的标签的均值，直至满足预设停止条件。

14. 如权利要求 8 所述的计算机装置，其中，所述处理器执行所述计算机可读指令以实现所述根据所述多个分类回归树生成随机森林时，具体包括：

对所述多个分类回归树进行剪枝处理；

根据剪枝处理后的所述多个分类回归树生成所述随机森林。

15. 一种存储介质，所述存储介质上存储有计算机可读指令，其中，所述计算机可读指令被处理器执行时实现以下步骤：

获取训练样本集，所述训练样本集中的每个训练样本为用户步行的加速度时间序列，每个训练样本带有标签，所述标签标记所述训练样本对应的情绪类别；

对所述训练样本集中的每个训练样本提取多个特征；

根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树；

根据所述多个分类回归树生成随机森林；

将待识别用户的多个特征输入所述随机森林，根据所述随机森林的输出确定所述待识别用户的情绪类别，其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

16. 如权利要求 15 所述的存储介质，其中，所述计算机可读指令被所述处理器执行还实现以下步骤：

对所述训练样本集中的每个训练样本的多个特征进行归一化处理，得到每个训练样本归一化后的多个特征；

所述计算机可读指令被所述处理器执行以实现所述根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树时，具体包括：

根据每个训练样本归一化后的多个特征构建所述多个分类回归树。

17. 如权利要求 15 所述的存储介质，其中，所述计算机可读指令被所述处理器执行以实现所述对所述训练样本集中的每个训练样本提取多个特征之前，还实现以下步骤：

对所述训练样本进行降噪；和/或

对所述训练样本中的缺失值进行填充；和/或

对所述训练样本中的异常值进行修正。

18. 如权利要求 17 所述的存储介质，其中，所述计算机可读指令被所述处理器执行以实现所述对所述训练样本进行降噪时，具体包括：

按照以下公式对所述训练样本进行移动平均降噪：

$$\text{output}[i] = \frac{1}{w} \sum_{j=0}^{w-1} \text{input}[i+j]$$

其中， $\text{output}[i]$ 是所述训练样本中第 i 个加速度数据对应的输出， w 为常量，取值为 3 或 5， $\text{input}[i+j]$ 是所述训练样本中第 $i+j$ 个加速度数据。

19. 如权利要求 15 所述的存储介质，其中，所述计算机可读指令被所述处理器执行以实现所述根据所述训练样本集的各个训练样本的多个特征构建多个分类回归树时，具体包括：

从所述训练样本集中随机选取 Q 个训练样本作为待分类样本；

从所述待分类样本的多个特征中随机选取 K 个特征；

确定所述待分类样本的所述 K 个特征中最优的切分特征和切分点，根据所述最优的切分特征和切分点将所述待分类样本划分为两个子集；

计算划分的每个子集中的待分类样本的标签的均值；

对于划分的每个子集，重复执行所述从所述待分类样本的多个特征中随机选取 K 个特征至所述计算划分的每个子集中的待分类样本的标签的均值，直至满足预设停止条件。

20. 一种情绪识别装置，其中，所述装置包括：

获取模块，用于获取训练样本集，所述训练样本集中的每个训练样本为用户步行的加速度时间序列，每个训练样本带有标签，所述标签标记所述训练样本对应的情绪类别；

提取模块，用于对所述训练样本集中的每个训练样本提取多个特征；

构建模块，用于根据所述训练样本集中的各个训练样本的多个特征构建多个分类回归树；

生成模块，用于根据所述多个分类回归树生成随机森林；

识别模块，用于将待识别用户的多个特征输入所述随机森林，根据所述随机森林的输出确定所述待识别用户的情绪类别，其中所述待识别用户的多个特征根据所述待识别用户步行的加速度时间序列得到。

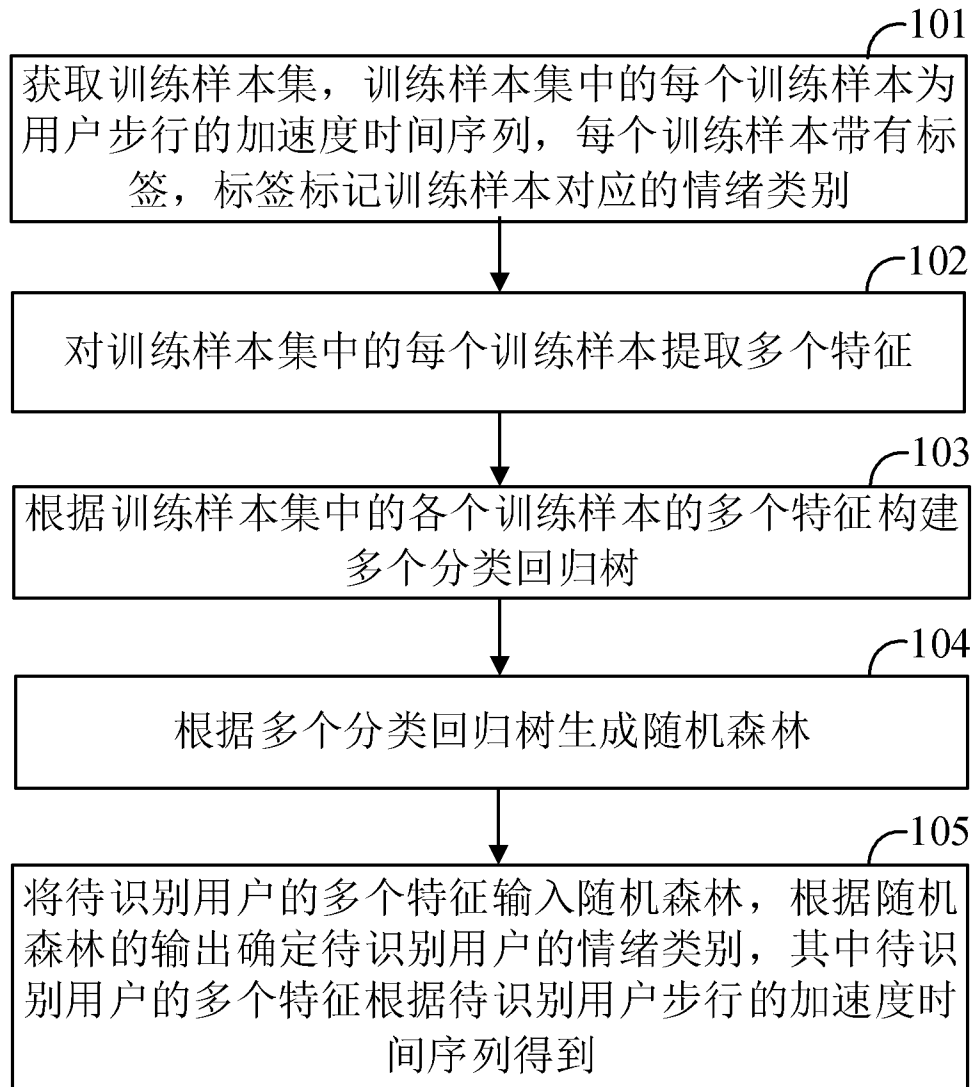


图 1

2/3

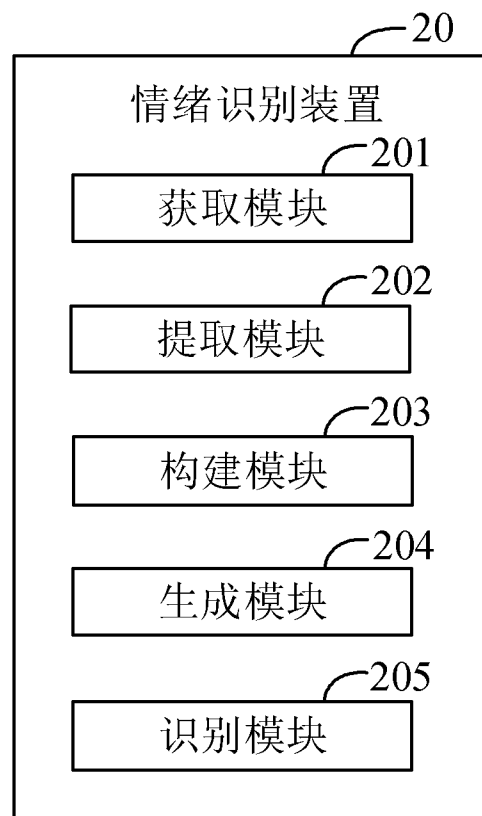


图 2

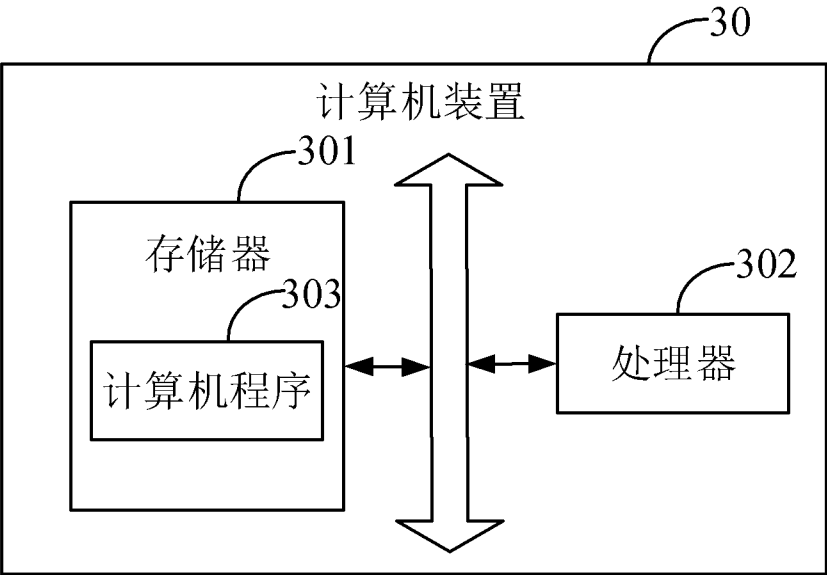


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/105630

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(2006.01)i; G06K 9/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K 9/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNTXT, CNKI, DWPI, SIPOABS: 情绪, 识别, 训练样本, 用户, 加速度, 时间序列, 标签, 特征, 分类回归树, 随机森林, emotion, recognition, training sample, user, acceleration, time sequence, label, feature, classification regression tree, random forest

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 109447324 A (SHENZHEN PERSONAL MOTION & PHYSICAL DATA CO., LTD.) 08 March 2019 (2019-03-08) claims 1-14, description paragraphs [0002]-[0177]	1-20
Y	CN 109492682 A (GUILIN UNIVERSITY OF ELECTRONIC TECHNOLOGY) 19 March 2019 (2019-03-19) description, paragraphs [0002]-[0110]	1-20
A	US 2012308971 A1 (SHIN HYUN SOON et al.) 06 December 2012 (2012-12-06) entire document	1-20
PX	CN 110705584 A (SHENZHEN ONECONNECT TECHNOLOGY CO., LTD.) 17 January 2020 (2020-01-17) description, paragraphs [0002]-[0257]	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

29 October 2020

Date of mailing of the international search report

04 November 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/105630

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109447324	A	08 March 2019	None			
CN	109492682	A	19 March 2019	None			
US	2012308971	A1	06 December 2012	KR	20120133979	A	11 December 2012
				KR	101840644	B1	22 March 2018
CN	110705584	A	17 January 2020	None			

国际检索报告

国际申请号

PCT/CN2020/105630

A. 主题的分类 G06K 9/62(2006.01)i; G06K 9/00(2006.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类																	
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) G06K 9/- 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNABS, CNTXT, CNKI, DWPI, SIPOABS: 情绪, 识别, 训练样本, 用户, 加速度, 时间序列, 标签, 特征, 分类回归树, 随机森林, emotion, recognition, training sample, user, acceleration, time sequence, label, feature, classification regression tree, random forest																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>Y</td> <td>CN 109447324 A (深圳个人数据管理服务有限公司) 2019年 3月 8日 (2019 - 03 - 08) 权利要求1-14, 说明书第[0002]-[0177]段</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>CN 109492682 A (桂林电子科技大学) 2019年 3月 19日 (2019 - 03 - 19) 说明书第[0002]-[0110]段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2012308971 A1 (SHIN HYUN SOON等) 2012年 12月 6日 (2012 - 12 - 06) 全文</td> <td>1-20</td> </tr> <tr> <td>PX</td> <td>CN 110705584 A (深圳壹账通智能科技有限公司) 2020年 1月 17日 (2020 - 01 - 17) 说明书第[0002]-[0257]段</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	Y	CN 109447324 A (深圳个人数据管理服务有限公司) 2019年 3月 8日 (2019 - 03 - 08) 权利要求1-14, 说明书第[0002]-[0177]段	1-20	Y	CN 109492682 A (桂林电子科技大学) 2019年 3月 19日 (2019 - 03 - 19) 说明书第[0002]-[0110]段	1-20	A	US 2012308971 A1 (SHIN HYUN SOON等) 2012年 12月 6日 (2012 - 12 - 06) 全文	1-20	PX	CN 110705584 A (深圳壹账通智能科技有限公司) 2020年 1月 17日 (2020 - 01 - 17) 说明书第[0002]-[0257]段	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
Y	CN 109447324 A (深圳个人数据管理服务有限公司) 2019年 3月 8日 (2019 - 03 - 08) 权利要求1-14, 说明书第[0002]-[0177]段	1-20															
Y	CN 109492682 A (桂林电子科技大学) 2019年 3月 19日 (2019 - 03 - 19) 说明书第[0002]-[0110]段	1-20															
A	US 2012308971 A1 (SHIN HYUN SOON等) 2012年 12月 6日 (2012 - 12 - 06) 全文	1-20															
PX	CN 110705584 A (深圳壹账通智能科技有限公司) 2020年 1月 17日 (2020 - 01 - 17) 说明书第[0002]-[0257]段	1-20															
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																	
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件													
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																
国际检索实际完成的日期 2020年 10月 29日		国际检索报告邮寄日期 2020年 11月 4日															
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 吴敏 电话号码 86-(010)-62411697															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/105630

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	109447324	A	2019年 3月 8日	无			
CN	109492682	A	2019年 3月 19日	无			
US	2012308971	A1	2012年 12月 6日	KR	20120133979	A	2012年 12月 11日
				KR	101840644	B1	2018年 3月 22日
CN	110705584	A	2020年 1月 17日	无			