

文章编号: 2095-2163(2022)12-0100-04

中图分类号: TP391

文献标志码: A

基于优化随机森林的下肢外骨骼步态识别

陆康, 何志琴

(贵州大学 电气工程学院, 贵阳 550025)

摘要: 在下肢外骨骼的长期使用中, 足底压力传感器可能会因为频繁挤压和摩擦而导致灵敏度下降或失效。针对这一问题, 提出了仅使用大腿和小腿角度、角速度为输入来源的下肢外骨骼步态识别方法。利用惯性测量单元测量大腿和小腿角度、角速度, 提取时域特征训练随机森林分类模型进行步态识别, 并采用随机森林变量重要性结合皮尔逊相关系数进行输入特征筛选, 保证识别的准确率, 同时提高算法运行效率。通过实验验证, 本文方法的准确率达到93.2%, 运行耗时更少, 可以用于下肢外骨骼的步态识别。

关键词: 下肢外骨骼; 步态识别; 随机森林; 特征选择

Lower limb exoskeletons gait recognition based on optimized random forest

LU Kang, HE Zhiqin

(School of Electrical Engineering, Guizhou University, Guiyang 550025, China)

[Abstract] In the long-term use of the lower limb exoskeleton, the plantar pressure sensor may lose sensitivity or fail due to frequent squeezing and friction. To address this problem, a gait recognition method for the lower limb exoskeleton using only thigh and calf angles and angular velocities as input sources is proposed. Inertial measurement units are used to measure thigh and calf angles and angular velocities, extract time-domain features to train a random forest classification model for gait recognition, and use the importance of random forest variables combined with Pearson correlation coefficients for screening input features to ensure the accuracy of recognition and also improve the efficiency of the algorithm operation. Through experimental validation, the accuracy of this paper's method reaches 93.2%. The algorithm could run with less time consumption, which can be used for gait recognition of lower limb exoskeleton.

[Key words] lower limb exoskeletons; gait recognition; random forest; feature selection

0 引言

近年来, 由于下肢外骨骼在提高人体负重能力以及辅助下肢残障人士进行康复训练等方面的广阔应用前景, 对下肢外骨骼的研究逐渐成为新的热点。

下肢外骨骼在不同的步态阶段使用不同的控制模式, 因此, 步态识别是下肢外骨骼控制系统中的重要环节, 以往的步态识别主要以足底压力传感器为主要信息来源, 使用阈值法进行步态识别, 根据设置的阈值将脚跟和脚掌分为触地状态和非触地状态。Catalfamo 等人^[1]和 Lopez 等人^[2]使用每次实验中足底压力的最大值和最小值的计算阈值。Mariani 等人^[3]将体重的 5% 设置为阈值。阈值法在离线分析时效果较好, 但当在线使用时, 不同步速或不同体重对应的阈值不同, 阈值法难以适应这种变化。

随着机器学习的兴起, 研究者们提出了各种基于学习的步态识别方法。Park 等人^[4]根据 GRF 训

练了人工神经网络(ANN)来进行步态识别。Attal 等人^[5]使用无监督的多元回归隐马尔可夫模型(MRHMM)进行步态识别, 平均识别率为 83.21%。刘薛勤等人^[6]使用多个柔性薄膜压力传感器采集足底压力, 利用支持向量机(SVM)进行步态识别, 识别准确率达到 90%。然而, 由于足底压力传感器会经常受到挤压磨损, 导致其寿命有限, 并不适用于长期使用。

因此, 一些研究者提出了仅使用行走时的下肢角度的步态识别方法。Liu 等人^[7]建立了基于多层感知器的步态识别模型, 平均相位正确率(CRP)为 87.22%。Ma 等人^[8]采集了髌膝关节角度, 利用核递归最小二乘法(KRLS)实现了步态识别, 测试平均准确率为 86%, 但由于 KRLS 无法直接完成多分类, 需要训练多个 KRLS 模型来实现步态识别。

基于当前的研究现状, 为提高下肢外骨骼步态识别准确度以及识别效率, 同时减少传感器的需求,

基金项目: 黔科合支撑[2021]一般 442。

作者简介: 陆康(1998-), 男, 硕士研究生, 主要研究方向: 下肢外骨骼步态预测; 何志琴(1971-), 女, 教授, 主要研究方向: 机器人驱动控制技术。

收稿日期: 2022-06-21

本文仅使用 IMU 测量的大腿、小腿角度和角速度, 提出了基于随机森林多分类能力的步态识别方法, 避免了足底压力传感器的敏感性以及耐用性有限的问题, 并利用随机森林变量重要性以及皮尔逊相关系数进行输入特征筛选, 提高了步态识别准确度和算法的运行效率。

1 相关原理

1.1 随机森林

随机森林是一种经典的集成学习算法, 具有训练速度快、泛化性较好等优点。该算法建立在 Bagging 策略上, 以大量决策树作为基础学习器, 在完成分类任务时, 训练多个分类决策树, 通过投票方式确定最终的分类^[9]。在本文中, 随机森林将用来根据提取的时域特征进行步态识别。

1.2 特征变量降维

在本文中, 使用大腿、小腿 IMU 测量得到的角度值作为步态识别的输入来源, 同时为了提高分类精度, 使用滑动窗口的方式将一定时间长度的历史角度作为输入, 但此时存在输入变量维数较高、变量冗余的问题, 因此结合随机森林的变量重要性和皮尔逊相关系数进行输入特征降维^[10]。

随机森林具有评估变量重要性的能力, 在随机森林的每个决策树建立过程中, 有大约 1/3 的数据样本没有作为测试集, 称为袋外 (Out-of-bag, OOB) 数据, 这些数据可以用来评估模型, 本文中根据 OOB 数据采用平均准确度 (Mean - decrease - accuracy, MDA) 下降来评估输入变量的重要性^[11], 主要步骤如下:

(1) 使用 OOB 数据计算每个决策树的分类准确率 A_1 。

(2) 对 OOB 数据中某一输入特征加入随机扰动, 再计算分类准确率 A_2 。

(3) 计算 2 次分类准确率的差值, 即: $\Delta A = A_1 - A_2$ 。

(4) 求随机森林中所有决策树的差值的平均值, 即 MDA。

MDA 越大, 表明对应的输入变量改变后, 随机森林的准确率下降越大, 因此该输入变量对步态识别准确率的影响越大。

计算出输入变量的重要性后, 还需要剔除其中的相关性较大的冗余特征, Pearson 相关系数是分析变量相关性常用的方法。本文根据 Pearson 相关系数, 按照 2 个变量相关性高时、保留变量重要性大的

特征为原则, 进行特征变量的降维。

2 数据采集及步态划分

2.1 步态数据采集及处理

为了采集步态数据, 进行了步态数据采集实验。数据采集设备如图 1 所示。由图 1 可以看到, 一名实验者穿戴数据采集设备进行了数据采集实验, 用粘性绑带将 4 个维特智能公司生产的型号为 JY931 的九轴陀螺仪固定在实验对象的大腿和小腿处, 采集实验对象大腿、小腿的角度和角速度。实验对象脚底穿着足底压力采集鞋, 在每只鞋的前掌和足跟处分别安装有一个大洋微型压力传感器, 采集足底压力数据, 传感器的分布见图 1。数据的采样频率为 100 Hz, 使用截止频率为 20 Hz 的二阶巴特沃斯低通滤波器进行数据滤波, 消除采集数据中的噪声和干扰, 数据经过处理后被存放在 SD 卡中, 便于进行后续分析。

实验者在跑步机上进行了 5 次行走实验, 每次行走实验中按高速、中速、低速来随机切换速度, 不指定目标速度, 切换的速度以及切换时机出于实验者的自由选择。为保证采集数据的有效性, 每次实验前都要进行传感器位置及状态检查, 确保 IMU 牢固绑缚在同一位置, 避免位置不同可能带来的测量误差。以右腿的步态数据为研究对象, 剔除不完整的步态周期数据, 以及跑步机启动停止时的数据, 共获得 104 915 个样本点的数据。



图 1 数据采集设备

Fig. 1 Data collection equipment

2.2 步态划分

在步态识别的实验中, 需要进行步态划分来生成步态数据对应的标签。在传统的步态划分中, 一

个步态周期被定义为一次足跟触地到下一次足跟触地的时间,随后根据脚跟和脚掌的触地状态将步态划分为脚跟着地(Heel Strike, HS)、平地站立(Flat Stance, FS)、脚跟离地(Heel Off, HO)和摆动(Swing, SW),在本文中使用1、2、3、4作为标签来对应这4个状态,步态划分的规则见表1。

表1 步态划分规则

Tab. 1 Gait division rules

步态阶段	1(HS)	2(FS)	3(HO)	4(SW)
脚跟	触地	触地	离地	离地
脚掌	离地	触地	触地	离地

划分步态需要确定脚跟和脚掌的触地状态,本文中根据文献[2]提出的阈值方法来区分脚跟、脚掌的触地和离地状态。通过公式(1)计算阈值:

$$T = T_{\min} + \alpha(T_{\max} - T_{\min}) \quad (1)$$

其中, T_{\max} 和 T_{\min} 分别为一次实验中采集到的每个步态周期中足底压力的最大值、最小值的平均值; α 是一个比例因子,用于调整阈值,来补偿足底压力的差异性。

3 实验与分析

3.1 实验参数设置

在 Matlab R2021a 中设计了本文的步态识别模型,按照 7:3 的比例将数据集随机划分为训练集和测试集。实验分析数据发现输入窗口越大,步态识别的准确率越高,考虑到算法的耗时问题以及参考文献[12]的经验,将输入窗口设置为 17,可以较好地保持算法的实时性,提取窗口内大腿、小腿角度和角速度的时域特征,包括:最大值、最小值、均值、方差和标准差,总共 20 个输入特征变量。随机森林的棵数和最小叶子节点大小经过网格搜索确定为 50 和 1。

3.2 评价指标

本文提出的步态识别方法使用准确度(Precision)和召回率(Recall)来进行评估^[5],推得的数学定义公式分别见如下:

$$Precision = \frac{T_p}{T_p + F_n} \quad (2)$$

$$Recall = \frac{T_p}{T_p + F_p} \quad (3)$$

其中, T_p 、 F_p 和 F_n 分别表示某一预测标签预测正确的数量、真实标签被误分为其他类的数量以及预测标签预测错误的数量。准确度和召回率越高,表明步态识别的效果越好。

3.3 特征变量筛选

输入变量的重要性如图 2 所示。图 2 中,输入量编号 1~5 表示大腿角度的最大值、最小值、均值、方差和标准差,6~10 表示大腿角速度的最大值、最小值、均值、方差和标准差,编号 11~20 表示小腿对应的输入量。从图 2 中可以看出,所有输入变量的重要性都大于 1,都具有一定的重要性,最高的 3 个变量编号为 6、8、9 分别对应大腿角速度的最大值、均值、方差。

对这 20 个输入变量计算 Pearson 相关系数,图 3 为输入变量的相关系数的部分结果。由图 3 可以看出,部分变量之间存在较大的相关性,根据 2 个变量相关性高时、保留变量重要性大的特征为原则,最终筛选出来 9 个输入变量,编号分别为 1、2、4、6、7、9、13、15、19,大大降低了输入变量的维数。

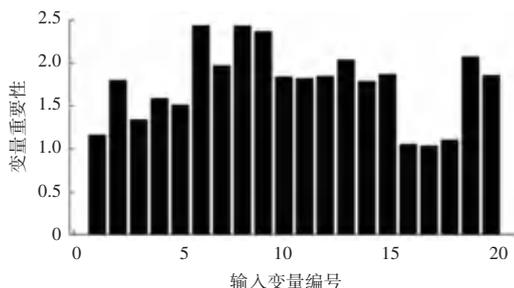


图2 输入变量的重要性

Fig. 2 Importance of input variables

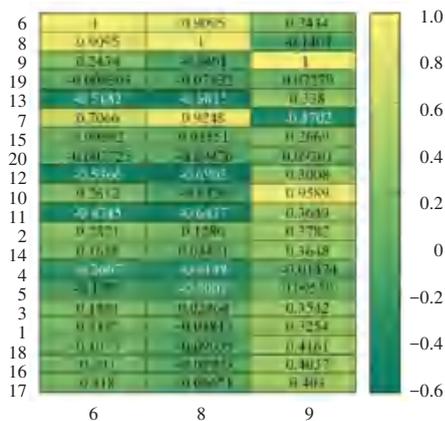


图3 部分输入变量的 Pearson 相关系数

Fig. 3 Pearson's correlation coefficient for a subset of input variables

3.4 仿真实验

为验证本文方法的性能,使用筛选前后的特征分别训练随机森林,并将文献[8]的一对多支持向量机(SVM)作为对比,在验证集上的步态识别的准确率及总预测时间见表 2。由表 2 分析可知,本文方法的步态识别准确率达到 93.2%,略低于使用全部特征的随机森林,与一对多 SVM 相比准确率提

高了 3.8%, 文献[4]以 90% 识别率作为评价步态识别方法能否用于实际步态识别的阈值, 因此本文方法可以用于步态识别。此外, 本文方法在测试集上的总预测耗时也是最低, 有助于提高步态识别的实时性。

表 2 步态识别方法对比

Tab. 2 Comparison of gait recognition methods

算法	总体准确率/%	预测耗时/s
SVM	89.4	15.00
RF	93.4	0.64
优化 RF	93.2	0.56

本文方法在测试集中取得的步态识别混淆矩阵如图 4 所示。由图 4 可以看到, 矩阵的最右一列为准确度, 最下一行为召回率。数据表明, 对标签 4、即摆动态的识别效果最好, 准确度达到了 97.9%, 对标签 2、即平地支撑的识别效果最差, 可能是由于支撑态时大腿、小腿角度变化较大, 造成识别困难, 并且支撑态的样本也较少, 不利于算法识别率的提高。图 5 则为 SVM 和本文方法对部分样本的步态识别结果。

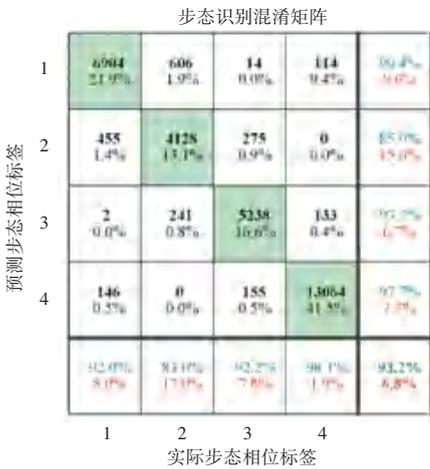


图 4 测试集步态识别的混淆矩阵

Fig. 4 Confusion matrix for test set gait recognition

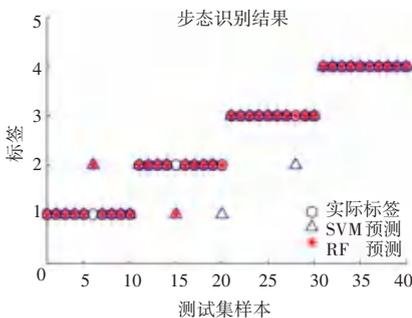


图 5 部分样本的分类结果图

Fig. 5 Classification results for parts of the samples

4 结束语

针对足底压力传感器在下肢外骨骼步态识别中的耐用性较差的问题, 设计了仅使用 IMU 信号的步态识别方法。使用随机森林设计了步态识别模型, 并且结合随机森林变量重要性和 Pearson 相关系数进行了输入变量筛选。

仿真结果表明, 本文方法的识别准确率高、耗时少, 为研究下肢外骨骼的步态识别提供了新的思路与参考。

参考文献

- [1] CATALFAMO P, MOSER D, GHOUSSAYNI S, et al. Detection of gait events using an F-Scan in-shoe pressure measurement system [J]. *Gait & Posture*, 2008, 28(3): 420-426.
- [2] LOPEZ-MEYER P, FULK G D, SAZONOV E S. Automatic detection of temporal gait parameters in poststroke individuals [J]. *IEEE Transactions on Information Technology in Biomedicine*, 2011, 15(4): 594-601.
- [3] MARIANI B, HOSSEIN R, XAVIER C, et al. Quantitative estimation of foot-flat and stance phase of gait using foot-worn inertial sensors [J]. *Gait & Posture*, 2013, 37(2): 229-234.
- [4] PARK J, KOO S, KIM C. Classification of standing and walking states using ground reaction forces [J]. *Sensors*, 2021, 21(6): 2145.
- [5] ATTAL F, YACINE A, ABDELGHANI C, et al. Automatic recognition of gait phases using a multiple regression hidden Markov model [J]. *IEEE/ASME Transactions on Mechatronics*, 2018, 23(4): 1597.
- [6] 刘薛勤, 刘宁, 苏中, 等. 基于足底压力感知的智能步态识别方法研究 [J]. *系统仿真学报*, 2021, 33(11): 2572-2578.
- [7] LIU Duxin, WU Xinyu, DU Wenbin, et al. Gait phase recognition for lower-limb exoskeleton with only joint angular sensors [J]. *Sensors*, 2016, 16(10): 1579.
- [8] MA Yue, WU Xinyu, WANG Can, et al. Gait phase classification and assist torque prediction for a lower limb exoskeleton system using kernel recursive least-squares method [J]. *Sensors*, 2019, 19(24): 5449.
- [9] 倪景峰, 李振, 乐晓瑞, 等. 基于随机森林的阻变型通风网络故障诊断方法 [J]. *中国安全生产科学技术*, 2022, 18(04): 34-39.
- [10] 刘翔, 谢涛, 王训洪, 等. 基于改进随机森林算法的岩石爆破块度预测 [J]. *矿业研究与开发*, 2022, 42(07): 25-29.
- [11] 吴立志, 陈振南, 张鹏. 基于随机森林算法的城市火灾风险评估研究 [J]. *灾害学*, 2021, 36(04): 54-60.
- [12] HUANG Liping, ZHENG Jianbin, HU Huacheng. Online gait phase detection in complex environment based on distance and multi-sensors information fusion using inertial measurement units [J]. *International Journal of Social Robotics*, 2022, 14(2): 413-428.