

基于扩展孤立森林的个性化跌倒检测研究

熊文滔, 郑建立

(上海理工大学健康科学与工程学院, 上海 200093)

✉xwt9888@foxmail.com; zhengjianli163@163.com



摘要: 随着中国人口老龄化趋势的加剧, 跌倒检测成为医疗健康领域的一个研究热点。针对传统基于监督学习的跌倒检测算法中难以获取真实的老年人跌倒数据作为跌倒检测方法的基础训练数据及未考虑个性化适配问题, 文章提出了一种基于扩展孤立森林的个性化跌倒检测模型, 将跌倒视为一个二分类异常检测问题, 通过可穿戴传感器采集大量老年人日常行为数据, 经预处理和滑动窗口特征提取后, 利用无监督的扩展孤立森林算法对每个老年人的日常行为数据进行单独建模, 当数据不符合正常行为模式时, 模型判定为跌倒。通过公开数据集SisFall验证模型效果, 实验结果表明基于扩展孤立森林的个性化跌倒检测模型的平均识别准确率可达96.76%, 平均敏感度和特异度分别为97.91%和94.72%, 具有良好的性能表现。

关键词: 跌倒检测; 扩展孤立森林; 个性化; 异常检测; 可穿戴设备

中图分类号: TP391 **文献标识码:** A

Research on Personalized Fall Detection based on Extended Isolation Forest

XIONG Wentao, ZHENG Jianli

(School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

✉xwt9888@foxmail.com; zhengjianli163@163.com

Abstract: With the increasing trend of China's aging population, fall detection has become a research focus in the field of public health. In traditional fall detection algorithm based on supervised learning, it is difficult to obtain the real fall data of the elderly as the basic training data of the fall detection method, and it fails to consider the personalized adaptation problem. In view of these problems, this paper proposes a personalized fall detection model based on the extended Isolated Forest, where fall is regarded as a binary classification anomaly detection problem, and a large number of daily behavior data of the elderly is collected through wearable sensors. After preprocessing and sliding window feature extraction, the daily behavior data of each elderly person is modeled separately using the unsupervised extended Isolated Forest algorithm. When the data does not conform to the normal behavior pattern, the model is determined as falling. The effect of the model is verified by the public dataset SisFall. The experimental results show that the average recognition accuracy of the personalized fall detection model based on the extended Isolated Forest is 96.76%, and the average sensitivity and specificity are 97.91% and 94.72%, respectively, with good performance.

Keywords: fall detection; extended isolation forest; personalization; anomaly detection; wearable device

1 引言(Introduction)

根据国家统计局2021年发布的第七次全国人口普查结果显示: 全国人口共14.11亿, 其中60岁以上的人口为2.64亿, 大约占总人口的18.7%, 与2010年相比上升了5.44%^[1], 中国正逐步迈入老龄化社会。在过去的几年里, 所有与老年人健

康相关的研究都得到了极大的关注, 跌倒检测就是其中的一个热门的研究领域。随着年龄的增长, 老年人的身体机能、认知能力和感官功能逐渐衰退, 因此极易发生跌倒事件。据统计, 约1/3的老年人每年会发生一次意外跌倒^[2], 跌倒也成为老年人因伤致死的重要原因^[3]。如果老年人在跌倒后能尽早

地被发现并治疗, 将对其后期的康复有很大的帮助^[4]。因此, 设计一个准确且高效的跌倒检测算法十分必要。

2 研究现状(Research status)

目前的跌倒检测算法主要分为三大类: 基于视频的跌倒检测算法、基于环境传感器的跌倒检测算法和基于可穿戴设备的跌倒检测算法。基于视频的跌倒检测算法主要依靠安装在屋内或者室外的摄像头对老年人的行为图像数据进行采集, 并通过机器视觉的方法判断老人出现的行为是否为跌倒行为^[5]。该算法的优点是准确率高, 但是运算开销大, 有应用盲区。基于环境传感器的跌倒检测主要是指在老年人的活动场所内安装红外线^[6]、超声波^[7]等环境传感器, 通过判断老年人周围环境状况的变化进而判断其是否跌倒。该算法对设备使用者的隐私侵犯风险较小, 但由于需要在环境中安装大量的传感器, 因此系统的便捷性不够, 使用场景受限。基于可穿戴设备的跌倒检测算法是指通过穿戴在身体某个部位的传感器收集原始数据后, 使用跌倒检测算法进行处理, 并判断设备使用者是否跌倒。与基于视频和环境传感器的跌倒检测算法相比, 可穿戴设备易安装、成本低, 并且受应用场景限制小。

基于可穿戴传感器的跌倒监测算法按照检测方法的不同可以分为两大类: 阈值法和机器学习法。阈值法是通过比较特征值和选定阈值的大小关系判断是否出现跌倒行为。这种方法的漏报误报率较高, 检测结果不准确。基于机器学习的算法将跌倒检测看作一个分类问题, 使用机器学习算法构建跌倒检测模型。季祥^[8]基于惯性传感器计算人体的俯仰角与横滚角, 结合足底压力, 利用支持向量机判断是否出现跌倒行为。段美玲等^[9]、赵举等^[10]利用加速度传感器结合双向长短期记忆网络算法进行特征提取与分类判别。YACCINREMA等^[11]使用加速度计和陀螺仪数据, 设计了一个基于决策树的跌倒检测模型。虽然目前研究人员已基于可穿戴传感器提出了多种跌倒检测算法, 但是这些算法仍然存在以下问题: 一是缺乏老年人的跌倒数据。目前, 被大量使用的监督学习算法主要针对跌倒行为进行建模和检测, 这对跌倒数据的质量要求较高。老年人跌倒是一种高危的行为, 因此在现实中很难获取老年人真实的跌倒数据, 大部分实验都是采用年轻志愿者参与跌倒实验获得的数据进行训练, 这必然导致实验结果与真实情况出现偏差。因此, 被大量使用的监督学习算法只是在理论上取得较好的结果, 但是在实际应用中的效果不佳。二是使用通用模型, 不考虑不同个体之间的差异。由于不同人的身高、体重及行为习惯不同, 即使在做出相同的跌倒动作时, 也会有很大的差别^[12]。图1展示了三个志愿者模拟“走路时不小心向后滑倒”这一动作时三轴合加速度的变化情况, 其中横轴表示时间, 纵轴表示合加速度数据, 虚线框内部为跌倒时采集的数据。从图1中可以清晰地看出, 不同的人在做出相同跌倒行为时, 加速度的极值、方差和跌倒持续时间具

有很大的差异, 如果忽略这些差异, 那么跌倒检测的准确率很难得到保障。

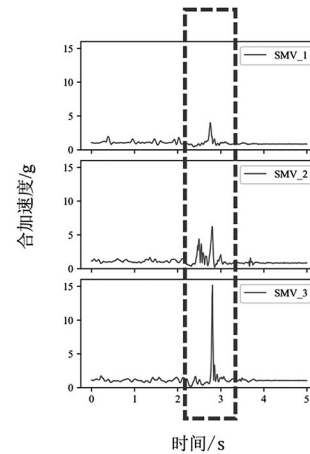


图1 不同的志愿者在做出相同跌倒动作时的合加速度

Fig.1 The sum vector magnitude of different volunteers when they perform the same falling action

为解决以上问题, 本文提出了一种基于扩展孤立森林的跌倒检测算法。针对老年人跌倒数据获取困难的问题, 研究人员使用数据挖掘中的异常检测技术, 将跌倒检测看作一个二分类异常检测问题, 使用扩展孤立森林算法对大量易于采集的正常行为数据进行建模, 当数据不符合正常行为数据特征的时候, 即可判定出现跌倒行为。针对个性化适配问题, 本文使用无监督的建模方式, 为每个人采集到的正常行为序列数据建立不同的跌倒检测模型, 解决了不同个体的个性化适配问题。本文在公开数据集SisFall上验证了算法的可行性和优越性。

3 相关方法(Correlation method)

3.1 扩展孤立森林算法

孤立森林(Isolation Forest, IForest)算法是由TING等^[13]提出的一种基于集成学习的无监督异常检测算法, 其主要依赖于特征选择和循环随机分割; 它并不是直接处理所有的数据点, 而是使用超平面对数据进行随机分割, 这就使得其具有线性的时间复杂度以及较低的内存需求。算法将异常数据标记为“少数且容易被孤立”的数据, 即分布稀疏且离群的点, 这与跌倒数据发生少且与正常行为数据差异较大的特征相符。

尽管标准的孤立森林算法的计算效率很高, 但由于其切割超平面是轴平行的, 这就导致在具有多个聚类中心的数据集中, 算法会因为切割条件的重叠而难以检测到局部异常点, 从而导致模型的检测精度下降^[14]。HARIRI等^[15]提出了一种扩展孤立森林(Extended Isolation Forest, EIF)算法, 其使用具有随机斜率的超平面替代轴平行的超平面, 克服了标准孤立森林的局限性。

在一个二维空间中的EIF算法切割方式如图2所示。对于图2(a)中的异常点 x_0 , 只需要少数几次切割就能到达二叉树的叶子

节点，而对于图2(b)中的正常点 x_1 ，则需要更多的切割次数。

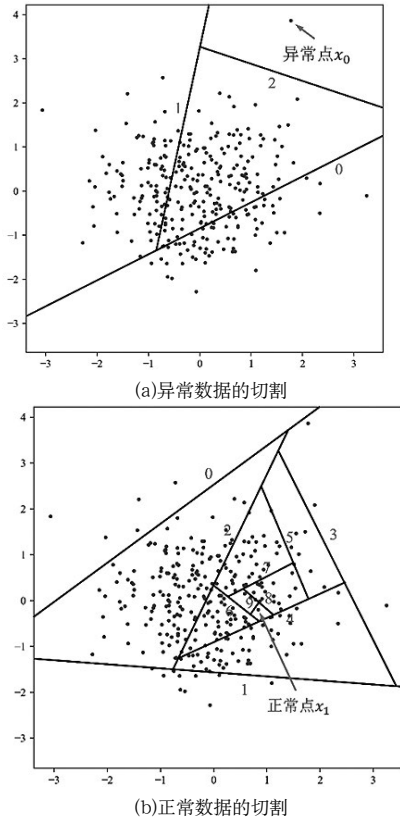


图2 二维空间中EIF算法区分异常和正常数据

Fig.2 Anomaly and normal data distinguished by EIF algorithm in 2D space

EIF由多棵孤立二叉树(Isolation Tree, iTree)组成，构建单棵iTree的具体实现方案如下。

步骤1：给定一个具有 Q 维特征的数据集 X ，随机选择 ψ 个样本点作为 X 的子样本集 X' ，放入孤立树的根节点。

步骤2：使用空间随机向量 $\vec{w} = (i_1, i_2, \dots, i_k)$ 表示一个空间的随机斜率，其中 $i \in [0, 1)$ ，使用 $\vec{p} = (n_1, n_2, \dots, n_k)$ 表示空间的随机截距。

步骤3：对于数据集 X' 中的点 $\vec{x}$ 的切割条件表示为 $(\vec{x} - \vec{p}) \cdot \vec{w} \leq 0$ ，以所选切割条件为基础构造一个超平面，将当前节点的数据空间划分成2个子空间，将满足切割条件的数据放在当前节点的左子树，将不满足切割条件的数据放在当前节点的右子树。

步骤4：在子节点中重复步骤2和步骤3，不断构造新的子节点，当数据不能被继续分割或者已经达到树的最大设定深度时，单棵iTree构建完毕。

由于单棵iTree的切割过程完全随机，为了使结果收敛，利用Ensemble的思想重复构建 n 棵iTree，形成EIF。

在检测阶段，EIF算法通过计算异常分数判断待检测样本点的异常程度。对于一个待检测的样本点 x_i ，其异常分数计算公式如下：

$$\text{score}(x_i) = 2 - \frac{E(h(x_i))}{c(n)} \quad (1)$$

其中， $h(x_i)$ 代表iTree的根节点到叶子节点的路径长度，

$E(h(x_i))$ 为节点 x_i 在多棵iTree上路径长度的期望值， $c(n)$ 为多棵iTree的平均查找深度，其计算公式如下：

$$c(n) = \begin{cases} 2H(n-1) - \frac{2(n-1)}{n}, & n > 2 \\ 1, & \text{otherwise} \end{cases} \quad (2)$$

其中， $H(n-1)$ 为调和函数，可被估计为 $\ln(n-1) + \xi$ ， ξ 为欧拉常数，其值为0.5772156649。

最后得到的异常分数将落在0—1的区间，如果异常分数越接近1，则表示该点异常的可能性越大，如果异常分数越接近0，则表示该点正常的可能性越大。

3.2 滑动窗口特征提取

跌倒是一个随机动态的过程，仅用单一时间点的数据很难完全反映跌倒的全过程。为了确保模型的训练效果以及计算的准确率，本文使用滑动窗口对采集到的数据进行特征提取。图3表明不同滑动窗口大小对模型最终跌倒检测准确率的影响，横轴表示选择的窗口大小，纵轴表示最终跌倒检测的准确率。

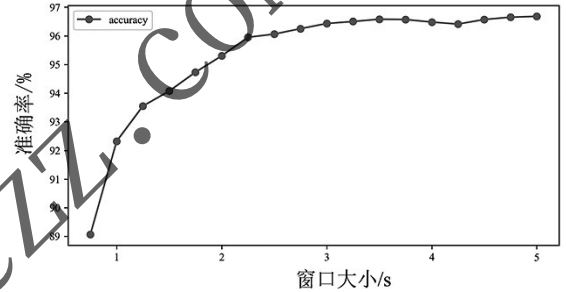


图3 不同滑动窗口大小对准确率的影响

Fig.3 Effect of different sliding window sizes on accuracy

从图3中可以看出，较小的滑动窗口无法包含完整的跌倒动作，因此可能存在漏报或者误报的情况。在滑动窗口选择大于3.5 s后，由于单个滑动窗口能够包含完整的跌倒动作，因此模型跌倒检测的准确率可以维持在一个较高的水平。但是，为了保证跌倒检测的及时性，滑动窗口也不宜选择过大，因此本文选择的滑动窗口大小为3.5 s。

本文使用佩戴在腰部的可穿戴三轴加速度传感器进行数据采集。为减少量纲差异，对滑动窗口内采集的数据进行归一化处理。归一化计算公式如下：

$$x_{\text{new}} = \frac{x - \mu}{\sigma} \quad (3)$$

其中， x 代表当前待归一化数据， μ 代表窗口内样本的均值， σ 代表窗口内样本的标准差。

在跌倒行为发生时，往往不是沿着单个加速度计轴方向跌倒的，使用原始三轴加速度直接进行跌倒检测判定的效果不够好，因此研究人员通过轴合成的方法引入以下三个新特征。

(1)加速度幅度矢量和定义如下：

$$SMV = \sqrt{a_x^2 + a_y^2 + a_z^2} \quad (4)$$

(2)信号 x 轴和 z 轴的幅度矢量和定义如下：

$$HOR = \sqrt{a_x^2 + a_z^2} \quad (5)$$

(3)信号 y 轴和 z 轴的幅度矢量和定义如下：

$$VER = \sqrt{a_y^2 + a_z^2} \quad (6)$$

其中, a_x 、 a_y 、 a_z 代表三轴加速度计在 x 轴、 y 轴和 z 轴方向的加速度数据。

从滑动窗口内提取模型所需特征: 选择窗口内加速度幅度矢量和的均值 $mean(SMV)$ 和标准差 $std(SMV)$ 衡量整体运动状况, 选择窗口内 x 轴和 z 轴幅度矢量和的均值 $mean(HOR)$ 和标准差 $std(HOR)$ 衡量水平方向的运动状况, 选择窗口内 y 轴和 z 轴幅度矢量和的均值 $mean(VER)$ 和标准差 $std(VER)$ 衡量竖直方向的运动状况。

3.3 跌倒检测算法框架

本文提出的基于EIF算法的跌倒检测算法的训练和检测流程如图4所示。

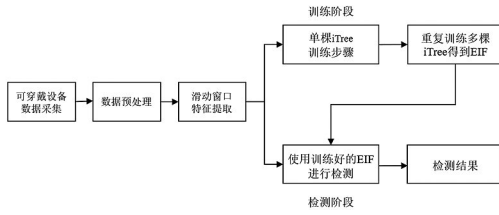


图4 跌倒检测模型结构

Fig.4 Fall detection model structure

从传感器直接采集到的数据中通常会有许多不需要的信号噪声, 为减小噪声对跌倒检测结果的影响, 同时考虑到计算的便捷性, 本文使用截止频率为5 Hz的四阶IIR Butterworth低通滤波器对原始数据进行滤波预处理。

在训练阶段, 首先通过佩戴在人体腰部的可穿戴设备进行数据采集, 然后进行数据预处理, 使用滑动窗口对预处理后的数据进行下采样和归一化并提取特征, 将得到的特征使用EIF算法训练得到可供使用的跌倒检测模型。在检测阶段, 同样进行数据采集预处理和特征提取, 然后通过训练好的EIF算法模型计算异常分数, 并判断是否出现跌倒行为。

4 实验验证(Experimental verification)

4.1 实验环境

采用python3.8编写实验代码, 使用EIF模块^[15]实现扩展孤立森林算法, 操作系统为Ubuntu 20.04.4 LTS。训练时设置的参数如表1所示。

表1 算法参数设置

Tab.1 Algorithm parameter settings

参数名称	参数值
iTree个数	100
单棵iTree最大深度	10
采样数	256

4.2 数据集介绍

本文使用公开数据集SisFall^[16]作为跌倒检测的研究数据集, 数据集中的所有记录均由放置在志愿者腰部的传感单元进行数据采集, 该单元由两个加速度计和一个陀螺仪组成,

设置采样频率为200 Hz, 本文使用其中一个加速度计数据进行实验。数据集包含38名志愿者的1,798例跌倒和2,740例日常行为活动的传感器数据。志愿者分布情况见表2, 行为数据分布情况见表3。38名志愿者中有23位青年人(年龄小于30岁)和15位老年人(年龄大于60岁)。23位年龄小于30岁的青年人和一位60岁的柔道运动员执行15种跌倒活动和19种日常行为活动并采集数据, 其余14位老年人只采集19种日常行为活动数据。

表2 志愿者分布情况

Tab.2 Distribution of volunteers

类别	人数/人	身高/m	体重/kg
老年人	15	1.50—1.71	50.0—102.0
青年人	23	1.49—1.83	41.5—80.5

表3 行为数据分布情况

Tab.3 Distribution of human activity data

动作	总数/人	类别数量/种	重复次数/次	采样时间/s
跌倒	1,798	15	5	15
非跌倒	2,740	19	1—5	12—100

4.3 实验及结果分析

由于本实验使用EIF算法对每一个人的日常行为数据进行无监督的个性化建模, 因此在个体层面划分数据集。将每一个人经预处理后的日常行为活动数据种的60%作为训练集、40%作为测试集, 同时将动作慢走、快走、慢跑、快跑这些只执行了一次的动作放入训练集, 将所有的跌倒动作作为测试集。经数据集划分后共有1,719组训练数据及2,819组测试数据。所有的实验都将重复进行五次, 确保实验结果的可信度。

训练完成后在测试集上对结果进行检测。跌倒检测为一个二分类问题, 模型只需区分跌倒和非跌倒动作两种类别, 判断结果可能存在真阳性(True Positive, TP)、真阴性(True Negative, TN)、假阳性(False Positive, FP)、假阴性(False Negative, FN)四种情况。在此基础上计算以下模型评价指标: 准确率(Accuracy, Acc)反映了模型的总体精确度, 敏感性(Sensitivity, Sen)反映了模型对跌倒动作的识别能力, 特异性(Specificity, Spe)反映了模型对日常行为活动(非跌倒)的识别能力。各个指标的计算公式如下:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (7)$$

$$Sen = \frac{TP}{TP + FN} \times 100\% \quad (8)$$

$$Spe = \frac{TN}{TN + FP} \times 100\% \quad (9)$$

图5展示了非个性化模型与个性化模型对于分类精度的影响。非个性化模型指的是使用非本人的日常行为数据训练而成的模型, 个性化模型是指使用大量用户个人的日常行为数据训练出的模型。

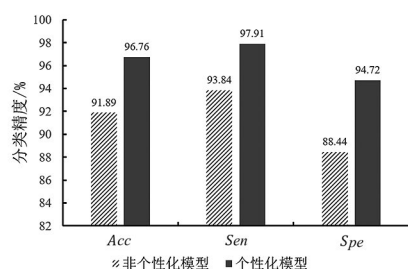


图5 个性化与非个性化模型结果对比

Fig.5 Personalized and non-personalized model results comparison

从图5中可以看出,由于个体存在差异性(身高、体重、行为习惯等),所以使用个性化模型分类结果要好于使用非个性化的模型。具体来说,平均准确率从91.89%提升到了96.76%,平均敏感度从93.84%提升到了97.91%,平均特异度从88.44%提升到了94.72%。

表4展示了其他常见的无监督异常检测算法与本文使用的EIF算法在跌倒检测模型中的性能对比。从表4可以看出,在跌倒检测任务中,EIF算法的准确率、敏感度和特异度均是最高。

表4 EIF算法与其他异常检测算法性能比较

Tab.4 Performance comparison between EIF algorithm and other anomaly detection algorithms

模型	Acc/%	Sen/%	Spe/%
IForest	84.03	88.20	76.68
ECOD ^[17]	86.02	89.54	79.82
COPOD ^[18]	86.80	89.87	81.39
LOF ^[19]	95.14	95.82	93.92
EIF(本文)	96.76	97.91	94.72

5 结论(Conclusion)

本文提出了一种基于扩展孤立森林算法的跌倒检测模型,利用佩戴在人体腰部的三轴加速度传感器采集人的日常活动数据,使用滤波算法对原始数据进行降噪处理,通过滑动窗口进行特征提取,对每一个人提取好的特征数据使用无监督的EIF算法进行个性化建模。解决了传统监督学习算法获取老年人跌倒数据困难以及个性化适配的问题。该模型在公开数据集SisFall上进行实验,验证了算法的有效性,同时与其他算法相比具有更高的识别精度。

参考文献(References)

- [1] 国家统计局.第七次全国人口普查主要数据情况[EB/OL]. (2021-05-11)[2022-07-02].http://www.stats.gov.cn/tjsj/zxfb/202105/t20210510_1817176.html.
- [2] 吕艳,张萌,姜昊昊,等.采用卷积神经网络的老年人跌倒检测系统设计[J].浙江大学学报(工学版),2019,53(6):1130-1138.
- [3] REN L, PENG Y. Research of fall detection and fall prevention technologies: a systematic review[J]. IEEE Access, 2019, 7:77702-77722.
- [4] YANG Y, YANG H, LIU Z, et al. Fall detection system based on infrared array sensor and multi-dimensional feature fusion[J]. Measurement, 2022, 192:110870.

- [5] VISHNU C, DATLA R, ROY D, et al. Human fall detection in surveillance videos using fall motion vector modeling[J]. IEEE Sensors Journal, 2021, 21(15):17162-17170.
- [6] LIU Z, YANG M, YUAN Y, et al. Fall detection and personnel tracking system using infrared array sensors[J]. IEEE Sensors Journal, 2020, 20(16):9558-9566.
- [7] 张大钧,兰恒荣,吴幼龙.基于超声多普勒效应的浴室跌倒检测[J].上海师范大学学报(自然科学版),2018,47(2):225-229.
- [8] 季祥.基于惯性传感器的坐姿和摔倒检测系统设计[D].杭州:浙江工业大学,2020.
- [9] 段美玲,潘巨龙.基于双向LSTM神经网络可穿戴跌倒检测研究[J].广西师范大学学报(自然科学版),2022,40(3):141-150.
- [10] 赵举,郑建立.基于多传感器和Bi-LSTM的个性化跌倒检测研究[J].智能计算机与应用,2022,12(4):146-150,158.
- [11] YACCHIREMA D, DE PUGA J S, PALAU C, et al. Fall detection system for elderly people using IoT and ensemble machine learning algorithm[J]. Personal and Ubiquitous Computing, 2019, 23(5-6):801-817.
- [12] SANTOYO-RAMÓN J A, CASILARI-PÉREZ E, CANO-GARCÍA J M. A study on the impact of the users' characteristics on the performance of wearable fall detection systems[J]. Scientific Reports, 2021, 11(1):1-14.
- [13] TING K, LIU F, ZHOU Z. Isolation Forest[C]// ICDM 2008. Eighth IEEE International Conference on Data Mining. Los Alamitos, CA, USA: IEEE Computer Society, 2008:413-422.
- [14] 谢雨,蒋瑜,龙超奇.基于随机子空间的扩展隔离林算法[J].计算机应用,2021,41(6):1679-1685.
- [15] HARIRI S, KIND M C, BRUNNER R J. Extended isolation forest[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(4):1479-1489.
- [16] SUCERQUIA A, LÓPEZ J, VARGAS-BONILLA J. SisFall: a fall and movement dataset[J]. Sensors, 2017, 17(12):198.
- [17] LI Z, ZHAO Y, HU X, et al. ECOD: unsupervised outlier detection using empirical cumulative distribution functions[J]. IEEE Transactions on Knowledge and Data Engineering, 2022:1.
- [18] LI Z, ZHAO Y, BOTTA N, et al. COPOD: copula-based outlier detection[C]// 2020 IEEE International Conference on Data Mining (ICDM). IEEE, 2020:1118-1123.
- [19] BREUNIG M M, KRIEGEL H P, NG R T, et al. LOF: identifying density-based local outliers[C]// Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. New York, NY, USA: Association for Computing Machinery, 2000:93-104.

作者简介:

熊文滔(1998-),男,硕士生.研究领域:数据挖掘.

郑建立(1965-),男,博士,副教授.研究领域:医学信息系统与集成技术.