

doi:10.16104/j.issn.1673-1891.2021.03.015

基于关联规则挖掘的赋权特征选择跌倒检测研究

谢 静, 魏 星, 蒋秀林, 陈春燕

(蚌埠医学院公共基础学院, 安徽 蚌埠 233030)

摘 要:针对穿戴式跌倒检测中特征属性过多,传统的降维和 PCA 赋权降维方法仅考虑了单个特征的重要度,未将特征组合的重要性考虑在内的问题,提出一种改进的基于关联规则挖掘的赋权特征选择方法。在特征降维过程中引入关联规则挖掘,根据频繁项集的支持度和置信度为特征赋权,实现特征选择。利用真实的数据集进行对比实验,结果表明:该方法分类的正确率、特异度和灵敏度均高于未经降维的原始数据集和经 PCA 赋权选择的数据集。该方法不仅能够实现对高维数据的降维,而且考虑了特征间的关联规则,能够得到更为优质的特征向量数据集。

关键词:关联规则;赋权;特征;跌倒检测

中图分类号:TP274;TP212.9 **文献标志码:**A **文章编号:**1673-1891(2021)03-0070-05

Study on Fall Detection by Weighting Feature Selection Based on Association Rules Mining

XIE Jing, WEI Xing, JIANG Xiulin, CHEN Chunyan

(School of Public Basic Education, Bengbu Medical College, Bengbu, Anhui 233030, China)

Abstract: With respect to the problem that too many feature attributes existed in wearable fall detectors and the existing dimensionality reduction and PCA methods did not take the importance of feature combination into account, an improved weighting feature selection method based on association rule mining is proposed. According to the support and confidence of frequent item sets, this method can select the feature attributes based on the association rules mining. The experimental results show that the accuracy, specificity and sensitivity of this method are higher than those of the original data set without dimensionality reduction and the data set selected by PCA. A better feature vector data set can be obtained, because the proposed method can not only reduce the dimensionality of high-dimensional data, but also consider the association rules among features.

Keywords: association rules; weighting; feature; fall detection

0 引言

伴随中国人口老龄化的加剧,老年人的健康问题已成为社会关注的焦点,老年人活动状态监测已成为多学科的研究热点^[1-3]。数据显示,超过 65 周岁的老年人有 40% 每年至少会跌倒一次^[4]。跌倒不仅会造成诸如肌肉损伤、骨骼断裂、关节扭伤等身体伤害,还会产生心理上的一系列严重后果^[5-6],这些伤害不仅给老年人及其家庭,同时也给社会带来了严重的经济负担^[7],因此,越来越多的学者投入到老年人的跌倒检测及预防研究中。

目前,使用较为广泛的是基于传感器的可穿戴设备跌倒检测^[8]。因传感器采集的信号无法直接用于行为分类训练,故需要在数据处理前提取合理有效的行为特征;又因跌倒检测对时效性的要求较高,故常采用时域分析法直接提取运动信号特征。常用的时域特征统计量有最大值、最小值、均值、方差、标准差、能耗、偏度、峰度等,如表 1 所示。

从表 1 可见,时域特征较多,而在人体行为采集时,最常用的是三轴加速度计和三轴陀螺仪,这 2 个传感器每个采样点可采集的向量 \mathbf{x} 含有 6 个变量,即 $\mathbf{x} = (a_x, a_y, a_z, w_x, w_y, w_z)$,其中 a_x, a_y 和 a_z 分别

收稿日期:2021-04-06

基金项目:国家重点研发计划(No.2018AAA0100500);安徽高校自然科学基金重点项目(KJ2019A0325);蚌埠医学院自然科学基金重点项目(BYKY1704ZD、BYKY2019022ZD)。

作者简介:谢静(1985—),女,安徽滁州人,讲师,硕士,研究方向:医疗物联网。

表示 X, Y 和 Z 轴上的加速度分量, w_x, w_y 和 w_z 分别表示三轴上的角速度分量。选择常见的 8 个特征统计量—— $max, min, avg, var, Sk, Ku, avgT, \Delta$ 计算可获得每个采样点含有特征 6×8 种。因采样向量中的 a_x, a_y, a_z 分别表示 X, Y 和 Z 轴上的加速度值, 无法直观地反映合成的运动结果, 为此, 引入合加速度 $a, a = \sqrt{a_x^2 + a_y^2 + a_z^2}$, 以 a 的峰度值为中心点, 记录该时刻点前后各 3 s 内的信号, 因采集频率为 25 Hz, 故每次记录由 151 个采样点组成, 采样矩阵为 151×7 , 而每个变量又有 8 个特征统计量, 将会产生 $151 \times 7 \times 8$ 的特征集。产生的这些特征中有的参数对判断跌倒和非跌倒的作用较弱, 且影响分类训练的算法复杂度和时间复杂度, 故需要在分类训练前剔除这些不相关和冗余特征, 即需要进行降维处理。

表 1 常用的特征统计量

特征值	符号	特征值	符号
最大值	max	能耗	E
最小值	min	偏度	Sk
均值	avg	峰度	Ku
方差	var	相关系数	r
标准差	Std	均值趋势	$avgT$
标准差趋势	SdT	最大插值	Δ

在特征降维时, 为获取合理的、更具代表性的特征, 将弱化甚至是抛弃一些不太重要的特征, 这个过程将会造成信息缺失。为尽可能避免信息的缺失, 一种类似于样本袋装法 (Bagging) 并结合特征选择与集成的算法——特征袋装法 (Attribute Bagging, AB)^[9] 应运而生。该方法与常见的样本袋装法的区别在于, 它先将全部属性分成若干子集, 然后利用这些子集对分类器进行训练, 最后将分类的结果集成。这种方法通过在特征选择部分产生多个不同的特征子集以保证数据信息的完整性。现常用的 AB 算法有随机特征选择袋装法 (Randomly Selected Attribute Bagging, RSAB) 和赋权特征选择袋装法 (Weight-selected Attribute Bagging, WSAB)^[10-11]。RSAB 即在原有特征集中随机选取得到每一个特征子集, 这种随机选取的方式极大地增加了重要特征遗漏的概率。针对此缺陷, WSAB 将特征的权重考虑在内, 先对特征的重要度进行评估, 如 Li 等^[12] 引入线性支持向量机 (Linear Support Vector Machine, LSVM) 和主成分分析法 (PCA) 来对特征的重要度进行评估并赋权, 孙勤等^[13] 提出结合过滤式和封装式的特征赋权降维方法。这些方法极大地降低了重要特征遗漏的概率, 为特征降维提供了新的思路。但是这些对特征重要度进行评估的方法仅关

注单一特征的重要度, 未将特征间的关联关系和特征组合的重要度对分类的影响考虑在内。

因此本文将关联规则挖掘 (Association Rules Mining) 算法 Apriori 引入到赋权特征选择袋装法 (简称 AWSAB) 中, 在评估特征的重要度时, 将特征间的关联关系考虑在内, 提出一种基于关联规则挖掘的赋权特征选择方法。本文首先对关联规则挖掘的相关知识进行介绍, 然后结合 Apriori 算法对随机特征袋装法进行改进, 结合特征的支持度和置信度对特征进行赋权, 最后将改进后的算法应用于跌倒检测, 结合几种对比实验结果对算法的性能进行分析。

1 基于关联规则挖掘的赋权特征选择方法

1.1 关联规则挖掘

关联规则挖掘^[14] 的目的是在事务数据集中找出所有的频繁模式、没有直接表示出来的各项之间的关联关系、偶然结构^[15] 及发现每一个项集的出现规律。通常, 关联规则挖掘分为 2 步: 第 1 步, 在全部项集中找出所有的频繁项集 (Frequent Itemsets); 第 2 步, 由这些高频项集产生规则 (Association Rules)。

第 1 步的目的是找出频繁项集。频繁即某一项集的出现频率相对于所有项集而言必须达到某一水平。这里的出现频率即为支持度 (Support)。假设 X, Y 是不同的项集, D 是事务集合, 则称 D 中包含 X 的数目与 D 中事务总数之比为 X 在 D 中的支持度^[16]。当项集的支持度高于事先设定的阈值 (即支持度阈值 Minimum Support, 表示最低重要性) 时, 认为这个项集是频繁项集。关联规则 $X \Rightarrow Y$ 的支持度记为 $\sup(X \cup Y)$ 。

第 2 步的目的是产生关联规则 (Association Rules)。关联规则的产生要依赖于上一步产生的频繁项集。在最小置信度 (Minimum Confidence, 表示关联规则的最低可靠性) 的条件门槛下, 若一规则所求得置信度 (Confidence) 大于最小置信度, 称此规则为关联规则。这里的置信度 $conf$ 表示 Y 在包含 X 的事务中出现的频率, 计算式如下:

$$conf(X \Rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)} \quad (1)$$

当 $conf(X \Rightarrow Y)$ 较高时, 表示当 X 出现时, Y 出现的概率较高。本文在进行特征选择时, 会结合已有特征和置信度, 调整待选特征的重要度。

1.2 基于关联规则挖掘的赋权特征选择算法

该算法是通过引入 Apriori 算法以实现随机

特征袋装法的改进。首先对特征进行关联规则挖掘,找出频繁项集,然后再结合特征的支持度和置信度对特征进行赋权。算法的具体流程如图 1 所示。

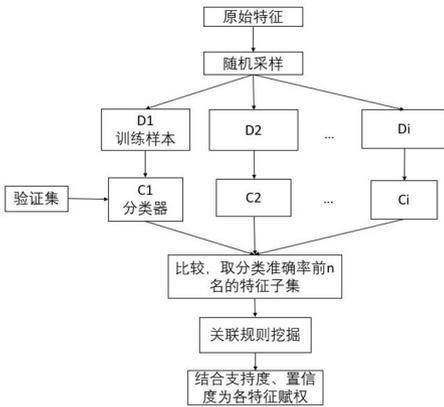


图 1 算法流程

基于关联规则挖掘的特征选择算法的详细解释如下:

输入:原始特征向量集 E 。

输出:赋权特征向量集 E' 。

1)对采集的原始数据进行运算,生成时域下的特征,对生成的特征进行随机采样,获得 i 个特征训练样本,并将这些训练样本输入分类器进行训练。

2)对验证集的分类结果进行比较,取分类结果正确率排名前 n 名的训练样本特征子集。

3)采用关联规则挖掘算法 Apriori 对获得的特征子集进行关联规则挖掘,得符合支持度阈值的频繁项集的支持度和置信度。

4)取频繁项集中支持度最高的特征作为第 1 个特征,然后结合其他特征的支持度和置信度选取第 2 特征,第 3 特征等。即第 2 特征的选取要结合该特征的支持度及第 2 特征基于第 1 特征的置信度,依此类推,获得按重要性从高到低排列的特征子集,然后再为各特征赋相应的权值。

2 实验结果及分析

为验证基于关联规则挖掘的赋权特征选择算法的有效性和正确性,本文以基于径向基核函数的支持向量机(Support Vector Machines, SVM)作为分类器,以原始特征向量集、主成分分析法赋权降维后的特征向量集以及采用本文方法进行降维后的特征向量集在分类器上的分类性能作为评价指标。具体比较方案设计如下:

方案 1:初始特征向量集+SVM。

方案 2:初始特征向量集+PCA 赋权特征选择

+ SVM。

方案 3:初始特征向量集+基于关联规则挖掘的赋权特征选择(AWSAB)+SVM。

2.1 实验数据采集

2019 年 3—12 月,5 位(3 男 2 女)身体健康的在校本科生参与实验数据采集,年龄 22~25 岁,体重 51.2~63.1 kg,身高 163~182 cm。测试者在铺有 30 cm 厚的软垫上对跌倒动作进行模拟,每位测试者将手环佩戴在左手手腕部位。模拟的跌倒动作包括前向跌倒、后向跌倒及侧向跌倒。此外,测试者也对日常行为动作进行了采集。日常行为包括上下楼梯、坐在椅子上、蹲下、躺到床上和正常行走等,表 2 具体描述了测试者模拟的动作。

表 2 动作行为描述

分类	动作	描述
	步行	测试者以正常速度的行走
	上楼梯	测试者以正常速度上楼梯
	下楼梯	测试者以正常速度下楼梯
日常行为	蹲下	测试者快速蹲下
	坐下	测试者快速坐到 50 cm 高的椅子上
	躺下	测试者躺倒一个 50 cm 高的硬板床上
跌倒行为	向后跌倒	测试者从步行到向后跌倒在一个 30 cm 的垫子上
	向前跌倒	测试者从步行到向前跌倒在一个 30 cm 的垫子上
	侧向跌倒	测试者从步行到侧向跌倒在一个 30 cm 的垫子上

测试者采集的动作不包括跑步,因为老人很少有跑步行为,故未将其考虑在内。实验要求测试者模拟各动作 20 次,排除明显异常数据 46 组,得到跌倒动作 271 组,日常行为数据 583 组。

2.2 评价指标

文章从识别准确度(Accuracy),灵敏度(Sensitivity)、特异度(Specificity)对跌倒检测的性能进行评估。准确度 Ac 表示所有行为被正确检出的比例,灵敏度 Se 表示所有跌倒动作的检出率,特异度 Sp 表示所有非跌倒行为的检出率。 Ac, Se, Sp 的定义如式 2~4 所示,其中 TP, FP, FN, TN 的含义如表 3 所示。

$$Ac = \frac{TP+TN}{TP+FP+TN+FN} \tag{2}$$

$$Se = \frac{TP}{TP+FN} \tag{3}$$

$$Sp = \frac{TN}{TN+FP} \quad (4)$$

表 3 各指标及其含义

指标	含义
TP	真阳性,阳性样本被判断为阳性
FP	假阳性,阴性样本被判断为阳性
TN	真阴性,阴性样本被判断为阴性
FN	假阴性,阳性样本被判断为阴性

2.3 实验结果比较及分析

实验中将跌倒行为标签设置为 1,非跌倒行为标签设置为 0,不同实验方案的实验结果有较大差异。方案 1 采用 SVM 分类器直接对原始特征集进行分类,得到的分类性能在 3 种方案中是最低的;方案 2 先对原始特征集用 PCA 方法进行特征赋权,再依据权重选择特征,再利用 SVM 对降维后的特征集进行分类,分类性能优于第 1 种方案;方案 3 对原始特征集先用本文的方法进行赋权特征选择,再对选择后的特征进行 SVM 分类,其分类性能明显优于前 2 种方案。3 种方法的 Ac, Se, Sp 的具体数据如表 4 所示。3 种方案的分类混淆矩阵如图 2 所示。

表 4 实验结果对比

方案	Ac	Se	Sp
原始特征集+SVM 分类器	89.7	88.3	90.4
原始特征集+PCA 赋权特征选择+SVM 分类器	93.4	93.0	93.6
原始特征集 + AWSAB + SVM 分类器	98.5	98.4	98.6

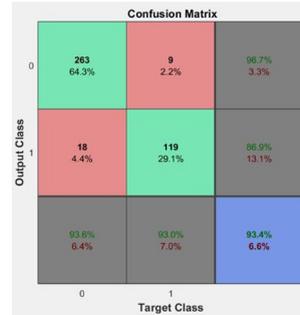
3 结论

本文针对行为检测中特征数据的高维特性,将



a. 本文方法(方案 3)

b. 方案 1



c. 方案 2

图 2 实验结果混淆矩阵

关联规则挖掘算法引入到特征降维过程中,形成了基于关联规则挖掘的赋权特征选择方法,并将此方法应用到跌倒检测实践中。该方法通过关联规则挖掘,不仅能够对单特征的重要度进行排序,而且能够识别出特征组合的重要度,再结合支持度和关联规则对特征及特征组合进行赋权。本文以原始特征集经 SVM 分类、原始特征集经 PCA 赋权特征选择再经 SVM 分类,这 2 种方法的实验结果和本文方法作为对比。对比结果表明,在原始数据和分类器相同的情况下,本文方法识别的准确度、灵敏度和特异度均高于其他 2 种方法。

参考文献:

- [1] 朱图陵.辅助技术在老年人康复中的应用[J].中国康复理论与实践,2017,23(8):971-975.
- [2] KHAN S S, HOEY J. Review of fall detection techniques: a data availability perspective [J]. Medical Engineering Physics, 2017, 39: 12-22.
- [3] RODRIGUEZ-MARTIN D, CARLOS P, MONSONIS A S, et al. A wearable inertial measurement unit for long-term monitoring in the dependency care area [J]. Sensors, 2013, 13(10): 14079-14104.
- [4] NOURY N, FLEURY A, RUMEAU P, et al. Fall detection—principles and methods [C]// Donna Hudson. 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Lyon, France: IEEE, 2007.
- [5] TINETTI M E, WILLIAMS C. Falls, injuries due to falls, and the risk of admission to a nursing home [J]. New England Journal of Medicine, 1997, 337(18): 1279-1284.
- [6] ZHAO Guoru, MEI Zhanyong, LIANG Ding, et al. Exploration and implementation of a pre-impact fall recognition method based on an inertial body sensor network [J]. Sensors, 2012, 12(11): 15338-15355.
- [7] LU Xianling, WANG Hongbin, WANG Yingying, et al. Human fall detection method based on acceleration sensor [J]. Journal of Computer Applications, 2013, 30(4): 1109-1111.
- [8] CHEN Long, MA Ting, LIU Songsong, et al. Posture estimation by Bayesian Network with Belief Propagation [C]// IEEE Re-

gion 10. IEEE International Conference of IEEE Region 10. Xi' an, China;IEEE,2013.

[9] WASKE B, Van Der LINDEN S, BENEDIKTSSON J A, et al. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2010, 48(7): 2880-2889.

[10] CHIH-FONG T. Feature selection in bankruptcy prediction[J]. Knowledge-Based Systems, 2009, 22(2):120-127.

[11] PATHASARATHY R, VADLAMANI R. Financial distress prediction in banks using group method of data handling neural network, counter propagation neural network and fuzzy ARTMAP[J]. Knowledge-Based Systems, 2010. 23(8): 823-831.

[12] LI Jianwu, WEI Haizhou, HAO Wangli. Weight-selected attribute bagging for credit scoring[J]. Mathematical Problems in Engineering, 2013:1-13.

[13] 孙勤,蒋艳凤,胡维,等.多准则赋权排序与 CSVM 相结合的特征选择算法[J].计算机工程与应用,2018,54(3):125-130.

[14] AGRAWAL R, IMIELINSKI T, SWAMI A. Mining association rules between set of items in large databases[C]//? Proc. of the 1993 ACM SIGMOD Conference, Washington DC, USA:ACM, 1993.

[15] 王晓龙.基于关联规则属性约简的树增广朴素贝叶斯分类器及应用[D].长春:吉林大学,2014.

[16] 刘亚波.关联规则挖掘方法的研究及应用[D].长春:吉林大学,2005.

(上接第 50 页)

从图 5、6 得出,在机械臂运动轨迹发生抖动的情况下,模糊自适应 PID 机械臂控制系统对比传统 PID 机械臂控系统可以更快地进行调整达到稳态,对预设的运动轨迹有着更好的重合度,机械臂波动较小。2 种算法花费时间如表 1 所示。

表 1 2 种算法花费时间的对比

所用算法	PID 机械臂控制系统	模糊自适应 PID 机械臂控制系统
花费时间/s	7.716 1	7.583 1

4 结语

运用模糊数学原理,基于传统机械臂的 PID 控

制系统,提出了一种适用于机械臂控制的模糊自适应控制系统,该控制系统将机械臂运行中的误差变化率和误差值输入模糊控制系统,依靠模糊控制系统的输出参数的变化来实时对 PID 控制参数进行调整优化。在 PC 机上用 MATLAB2014a 进行了的仿真对比实验,模糊自适应 PID 机械臂控制系统在缩短过渡时间和减小调量方面相较于传统 PID 机械臂控制系统具有较高的优势,对在 STM32F103C8T6 控制芯片控制下的舵机驱动小型六轴械臂具有一定的应用价值。

参考文献:

[1] 毛龙锐,薛涛,曹秀芳,等.基于 MATLAB 的 7 自由度液压重载机械臂的工作空间分析[J].天津理工大学学报,2020,160(3):43-47.

[2] 刘士荣.机械臂轨迹跟踪控制研究进展[J].控制工程,2011(1):120-126+136.

[3] 姚月琴,王影星,张磊.刚性机械臂系统的抗饱和和神经网络动态面控制[J].机床与液压,2019,47(3):27-31.

[4] 田苗,史玉琴.基于 PID 技术的工业自动控制系统的优化改进研究[J].现代电子技术,2015(13):132-133.

[5] 陈家俊,贺云波.伺服系统中 PID 控制器参数整定的研究[J].机床与液压,2021,49(1):13-16.

[6] 舒蕾.基于光学偏振成像技术的机械臂运动结构稳定性检测[J].自动化与仪器仪表,2020(8):172-176.

[7] 李素若.机械臂动态响应传递函数矩阵模型[J].辽宁工程技术大学学报,2014(5):670-674.

[8] 王娜.基于 Simulink 的机械臂 PID 算法优化设计与实现[J].电子设计工程,2020,28(4):158-162.

[9] 邬合林.PID 在机械臂位置控制中的应用[J].冶金管理,2020(19):46-47.

[10] 宁煜.基于改进 PID 控制的三关节机械臂运动轨迹跟踪误差研究[J].机械设计与制造工程,2018(8):107-110.