

基于最优二叉决策树分类模型的奶牛运动行为识别

王 俊¹, 张海洋¹, 赵凯旋¹, 刘 刚²

(1.河南科技大学农业装备工程学院, 洛阳 471003;

2.中国农业大学现代精细农业系统集成研究教育部重点实验室, 北京 100083)

摘 要: 针对奶牛行为分类过程中决策树算法构建主观性强、阈值选取无确定规则, 易导致分类精度差的问题, 本文提出一种基于最优二叉决策树分类模型的奶牛运动行为识别方法, 首先选取描述奶牛腿部三轴加速度数值大小、对称性、陡峭程度、变异程度、不确定性及夹角的 24 个统计特征量, 其次通过构建 ROC (receiver operating characteristic, ROC) 曲线获得各统计特征量的最佳行为类别分组方式及最优阈值, 然后利用信息增益作为最优二叉决策树划分属性的选择标准, 最终构建最优二叉决策树分类模型对奶牛运动行为进行分类识别。试验结果表明, 该分类模型能够有效区分奶牛的站立、平躺、慢走、快走、站立动作、躺卧动作 6 种运动行为, 平均准确率、平均精度、平均 $F1$ 值分别为 76.47%、76.83%、76.57%, 相较传统的 ID3 (iterative dichotomiser 3, ID3) 决策树算法分别高 5.71、5.4 和 5.61 个百分点, 分别高于 K-means 聚类算法 7.51、8.02 和 7.77 个百分点, 优于支持向量机算法 6.77、6.72 和 6.57 个百分点。该方法可为提高奶牛行为分类精度提供有效的理论支撑。

关键词: 数据采集; 数据处理; 算法; 行为分类; 三轴加速度计; 无线腿部传感器; ROC 曲线; 二叉决策树

doi: 10.11975/j.issn.1002-6819.2018.18.025

中图分类号: S24; TP274.2

文献标志码: A

文章编号: 1002-6819(2018)-18-0202-09

王 俊, 张海洋, 赵凯旋, 刘 刚. 基于最优二叉决策树分类模型的奶牛运动行为识别[J]. 农业工程学报, 2018, 34(18): 202-210. doi: 10.11975/j.issn.1002-6819.2018.18.025 http://www.tcsae.org

Wang Jun, Zhang Haiyang, Zhao Kaixuan, Liu Gang. Cow movement behavior classification based on optimal binary decision-tree classification model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(18): 202-210. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2018.08.025 http://www.tcsae.org

0 引 言

奶牛行为是奶牛健康和福利水平的重要指标, 实时判别奶牛运动行为可及早发现疾病、提升福利水平^[1-3]。例如, 奶牛患蹄病时, 站立时间明显减少^[4-6]; 平躺时间过长, 奶牛患病可能性较大^[7-9]; 奶牛发情时, 慢走时间增长, 可达正常状态下的 6 倍, 快走行为出现频次增加, 影响产奶量^[10-13]; 站立动作、躺卧动作行为出现的频率, 与奶牛疾病、舒适度均有较强的关联性^[14-15]。实时判别上述 6 类奶牛行为强度 (数量、持续时间、频率), 有助于及时掌握奶牛的健康水平、发情状况, 为提升奶牛福利水平提供一种参考^[16-18]。

目前, 奶牛行为判别主要依靠人工观测, 劳动强度大、工作效率低、主观性强。近十年来, 相关研究人员在奶牛行为自动监测技术领域开展了大量研究^[19-21], 其中, 三轴加速度计以其低廉的价格、强大的抗干扰能力和准确的数据测量精度, 被广泛应用于奶牛运动行为识别^[22-24]。大量研究表明, 在奶牛站立、平躺、慢走等动作时, 加速度计坐标轴的指向不可避免会发生偏移, 但获取各类行为的数据特征仍具有一定的差异性, 可应用

机器学习算法判别奶牛行为类别^[25-27]。常用的奶牛行为识别的机器学习算法可分为无监督学习和监督学习, 无监督学习无需训练样本, 直接对数据进行建模分类。常用的无监督学习算法为 K-means 聚类算法。尹令等^[28]采用 K-means 聚类算法能够较好区分奶牛静止和动态行为特征, 但对于多种动态的行为分类效果不明显。无监督学习的优点是人为干涉小, 缺点是对奶牛行为异常数据敏感。监督学习是通过训练已知奶牛行为数据构建分类模型, 再将此模型用于新数据集的预测。相较于无监督学习, 监督学习具有抗噪声能力强和行为判别准确率高等优点。常用的监督学习算法为支持向量机 (support vector machine, SVM) 算法和决策树 (decision-tree) 算法。Martiskainen 等^[29]采用脖颈式固定方式, 应用支持向量机分类模型, 对奶牛的站立、平躺、反刍、进食、跛行 5 种运动行为进行识别, 平均精度为 78%, 但其算法复杂度较高, 实用性较差。与支持向量机算法相比, 决策树算法具有运算复杂度低、稳定性好、直观易懂等优点。Diosdado 等^[30]采用决策树分类模型, 判别奶牛进食、站立、平躺 3 种运动行为, 分类平均灵敏度与平均精度分别为 83.94%、78.56%, 然而难以判断奶牛的发情情况; Gupta 等^[31]采用决策树算法对奶牛跛足进行预测, 灵敏度为 79.67%, 仅能对单一行为进行判断; Abell 等^[32]运用决策树算法对奶牛平躺、站立、行走、爬跨 4 种行为进行识别, 平均精度为 81%, 但分类类别数量有限。已有决策树模型构建时, 由于分裂属性阈值选取无确定规则、全局优化效果差, 且随着分类类别的增加, 奶牛行为整

收稿日期: 2018-06-04 修订日期: 2018-08-10

基金项目: “十三五”国家重点研发计划课题 (2018YFD0500705); 国家自然科学基金 (61771184); 河南省科技攻关项目 (172102210040); 河南省创新型科技人才建设项目 (184200510017)

作者简介: 王 俊, 男, 山西晋中人, 系主任, 博士, 主要从事精细农业系统集成研究。Email: wj@haust.edu.cn

体识别率不可避免地会降低。

针对传统决策树算法无法合理选取最优阈值而影响奶牛运动行为分类准确率的问题, 提出一种最优二叉决策树的奶牛运动行为识别方法, 该方法首先选取奶牛腿部加速度的 24 个统计特征量, 其次通过 ROC (receiver operating characteristic, ROC) 曲线获得各统计特征量的最佳行为类别分组方式及最优阈值, 然后利用信息增益作为最优二叉决策树分类模型划分属性的选择标准, 最终构建最优二叉决策树分类模型, 进行奶牛运动行为的分类与决策, 有效提升了行为判别的准确率。

1 测量装置与仪器

1.1 腿部无线传感器

腿部无线传感器由微处理器、三轴加速度计、无线收发器、电源模块组成。通过腿部无线传感器可以获取奶牛运动三轴方向的加速度数据, 并通过无线收发器将数据发送给接收器, 然后经串口发送至上位机。

微处理器采用美国 TI 公司的超低功耗芯片 MSP430F149IMP, 是一种 16 位具有精简指令集的混合型单片机, 内部含有 60 kB 的 FLASH 和 2 kB 的 RAM, QFP 封装; 三轴加速度计采用美国 ADI 公司的高分辨率芯片 ADXL345, 因奶牛运动幅度较小, 测量范围设置为 $\pm 2\text{ g}$; 无线收发器采用挪威 Chipcon 公司的 CC1101 芯片, 其最大传输数率达 500 kbps, 配置 CC1101 的工作频率为 433 MHz, 此频段的波长较长, 可以增加 CC1101 的通信距离; 电源模块采用深圳德力普公司生产的 18650 型锂电池, 容量为 2800 mA·h。

腿部无线传感器尺寸为 $75\text{ mm} \times 48\text{ mm} \times 30\text{ mm}$, 总质量为 205 g, 采样频率设为 4 Hz, 整个腿部无线传感器封装在一个密闭防水盒中, 通过尼龙绑带固定于奶牛右后腿关节以下处, 令三轴加速度计的 X 轴、Y 轴、Z 轴分别指向牛腿向下方向、牛腿前进方向、垂直牛腿向外方向, 奶牛腿部无线传感器及固定方式如图 1 所示。

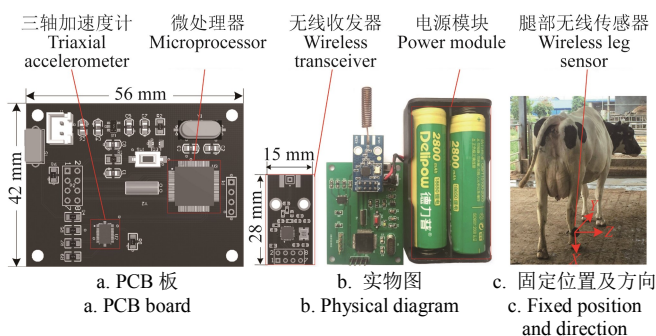


图 1 腿部无线传感器及固定方式

Fig.1 Wireless leg sensor and its fixed method

1.2 传输数据包格式及工作流程

腿部无线传感器的传输数据包共 21 字节, 包含 4 字节前导码、2 字节同步字、1 字节数据长度、1 字节设备地址、11 字节数据信息和 2 字节 CRC 校验码。前导码为一个交互的 1、0 序列, 用于数据的位同步; 同步字用于数据传输中对齐数据, 以及判断数据是否有效; 数据长

度设定为 11 字节数据; 数据信息的内容为三轴加速度计各方向加速度数据; CRC 是循环冗余校验码, 保证数据的正确传输。设置数据包传输速率为 38.4 kbps。

在采集试验数据过程中, 事先为每个腿部无线传感器分配唯一地址 (ID_1, ID_2, \dots, ID_n)。上电后, 接收器依次向各腿部无线传感器发送其地址, 腿部无线传感器接收地址后与自身地址相比较, 若地址不相同, 接收器访问下一个传感器地址; 若地址相同, 则腿部无线传感器将加速度数据包通过无线收发器传输到接收器。

1.3 腿部无线传感器性能测试

为了验证腿部无线传感器功耗性能, 于 2017 年 9 月 6 日 8:00—17:00 进行时长 9 h 的功耗测试。测试前, 在传感器供电端串联 100 Ω 电阻, 并通过示波器 (RIGOL DS1000Z) 记录工作电压和休眠电压波形。分别选取 0~3 h、4~6 h、7~9 h 共 3 个时间段中任意 2 s 内的波形, 结果如图 2 所示, 0~3 h 时间段的工作时峰值电压为 4.12 V, 4~6 h 时间段的工作时峰值电压为 4.00 V, 7~9 h 时间段的工作时峰值电压为 3.84 V, 工作时峰值电压仅下降 0.28 V; 休眠电压为 3.04 V, 9 h 期间工作电流约为 37.60 mA。通过进一步测试, 在 2 节 18650 锂电池电量充满状态下, 传感器处于稳定工作状态下采集加速度数据时长可达 60 h, 满足传感器试验期间的工作要求。

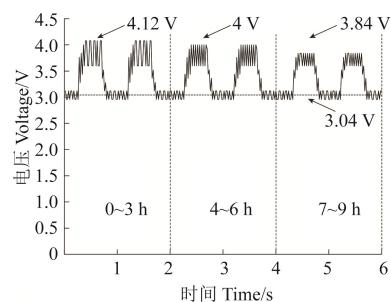


图 2 腿部无线传感器工作时电压变化情况

Fig.2 Voltage change when wireless leg sensor is working

为测试牛舍环境下腿部无线传感器与接收器间通信的丢包率水平, 通过改变腿部无线传感器距接收器的距离, 统计不同距离时接收器实际收取的数据包数量, 计算通信丢包率。腿部无线传感器与接收器距离分别设置为 1、15、30、45、60、75、90 m。丢包率测试结果如表 1 所示, 在 45 m 的传输距离以下时, 丢包率保持较低的水平, 在 2.5% 以内; 当距离大于 45 m 时, 丢包率迅速增加; 传输距离为 90 m 时, 丢包率仅为 6.33%, 整体丢包率较为理想。

2 最优二叉决策树算法设计

2.1 选取奶牛腿部加速度的统计特征量

为深入挖掘奶牛运动三轴加速度数据所蕴含的丰富信息, 应选取具有代表性的统计特征量, 全面反映其数据特征, 以提高奶牛行为识别率。选取的统计特征量具体包括加速度传感器各轴的均值、标准差、偏度、峰度、最大值、最小值, 以及总体标准差、能量、幅度 (signal magnitude area, SMA)、矢量幅度 (signal vector magnitude,

SVM)、熵、倾角,共 24 个统计特征量^[33],其中,平均值体现奶牛腿部各个方向运动变化的趋势;偏度反映奶牛腿部各个方向加速度数据总体分布的对称信息;峰度表征奶牛腿部各个方向加速度数据总体分布密度曲线在其峰值附近的陡峭程度;最大值、最小值表示剧烈运动的瞬间加速度变化量;标准差、总体标准差呈现奶牛腿部运动的瞬间最大变化趋势;能量、幅度、矢量幅度综合反映三轴运动变化的大小;熵综合反映三轴运动变化的不确定性;倾角体现奶牛腿部运动加速度与自然坐标系间夹角大小。

表 1 牛舍环境下无线腿部传感器丢包率测试
Table 1 Packet loss rate test of wireless leg sensor in cowshed environment

距离 Distance/m	发送数据包数量 Number of packets sent	接收数据包数量 Number of received packets	丢包率 Packet loss rate/%
1	300	296	1.33
15	300	295	1.67
30	300	293	2.33
45	300	293	2.33
60	300	286	4.67
75	300	285	5.00
90	300	281	6.33

2.2 计算各统计特征量的最优阈值

1) 首先,对于任一统计特征量,将奶牛运动行为划分为 2 个行为组,一组为正类行为组,另一组为负类行为组。将该统计特征量的最小值与最大值作为阈值取值范围的下限和上限,并确立阈值个数为 n ,在阈值可取范围内,通过连续改变阈值将训练数据集中的奶牛运动行为划分为预测正类行为或预测负类行为,计算不同阈值下的真阳性率(true positive rate, TPR)和假阳性率(false positive rate, FPR)。

$$TPR = \frac{TP}{TP+FN} \quad (1)$$

$$FPR = \frac{FP}{FP+TN} \quad (2)$$

式中 TP (true positive) 表示实际为正类行为被预测成正类行为的样本数; FN (false negative) 表示实际为正类行为被预测成负类行为的样本数; FP (false positive) 表示实际为负类行为被预测成负类行为的样本数; TN (true negative) 表示实际为负类行为被预测成正类行为的样本数。

2) 根据步骤 1) 得到不同阈值下的 TPR 与 FPR,以 FPR 为横坐标,TPR 为纵坐标构建 ROC 曲线,并计算目标函数值 J 。

$$J = \max \left\{ \int_0^1 f(x) dx \right\} \\ = \max \left\{ \sum_{i=1}^n \frac{h}{2} (f((i-1)h) + f(ih)) \right\} \quad (3)$$

式中 J 表示 ROC 曲线下面积的最大值, n 为阈值的个数, h 为步长, $h=1/n$, 函数 $f(\cdot)$ 表示第 i 个阈值的真阳性率值构成的函数,即 ROC 曲线。目标函数值 J 对应的统计特征量值即为该行为类别分组方式下的最优阈值。

3) 最后,依次调整正类行为、负类行为 2 组中的行

为类别,改变分组情况,计算各分组情况下该统计特征量的目标函数值,通过比较得到最大的目标函数值,则此目标函数值对应的分组划分情况为该统计特征量的最佳行为类别分组方式,对应的阈值为最优阈值。

2.3 计算各统计特征量的信息增益

二叉决策树由一个根结点、若干个内部结点和叶结点构成。叶结点为分类结果,对应奶牛运动行为;根结点和内部结点为统计特征量,对应于奶牛腿部加速度的统计特征量^[34]。在各级结点上选择统计特征量时,选取信息增益作为分类准则,信息增益越大,即表示使用该统计特征量进行奶牛运动行为划分所获得的不确定性减小的程度愈小。

信息熵表示统计特征量的不确定度,其定义为

$$Entropy(D) = - \sum_{k=1}^K p_k \log_2 p_k \quad (4)$$

式中 D 表示训练数据集, k 表示行为类别 $\{k=1,2,\dots,K\}$, p_k 表示训练数据集 D 中第 k 类行为数据所占的比例。

统计特征量的信息增益为

$$Gain(D, a) = Entropy(D) - \sum_{v=1}^V \frac{|D^v|}{|D|} Entropy(D^v) \quad (5)$$

式中 a 表示奶牛腿部三轴加速度的统计特征量, V 表示统计特征量 a 的可能取值 $\{\alpha^1, \alpha^2, \dots, \alpha^V\}$, 统计特征量 a 中只有“是”与“否”2 种分支,则 $V=2$, D^v 表示训练数据集 D 中在统计特征量 a 上取值为 a^v 的样本。

2.4 构建最优二叉决策树分类模型

最优二叉决策树分类模型采用自顶向下的递归构造方法,其构造方法如下:

1) 针对 6 种奶牛运动行为,根据 ROC 曲线获得奶牛腿部三轴加速度数据 24 个统计特征量的最佳行为类别分组方式及最优阈值,并计算 24 个统计特征量的信息增益。

2) 选取具有最高信息增益的统计特征量作为最优二叉决策树模型根结点,并由该统计特征量的最优阈值建立 2 个分支,分别为具有最高信息增益的统计特征量的最佳行为类别分组方式中的正类行为组和负类行为组。

3) 分别将各分支内奶牛运动行为划分为正类行为组和负类行为组,递归调用步骤 1) 与 2) 的方法建立二叉决策树模型的分支,直到所有分组中仅包含 1 类行为后停止,最终构建泛化能力强的二叉决策树分类模型。

3 试验与结果分析

试验区位于河南省南阳市蒲山镇杨营新村三色鸽奶牛养殖场 (33°05'50.64" N、112°32'25.32" E), 该养殖场共有黑白花类荷斯坦奶牛 700 余头,分别饲养在 5 座双列半开放式标准化牛舍内,牛舍长 150 m,宽 31 m。选取 8 头健康成年荷斯坦奶牛 (胎次 = 3±0, 泌乳期 = 50.27±6.35 d, 日产奶量 = 31.05±2.46 kg, 体质量 = 678.46±34.18 kg) 作为试验对象,饲养区域为 25 m × 13 m 的独立划分区域,于 2017 年 9 月 11 日至 21 日共

10 d 进行数据采集, 每日每头牛采集 4 h 数据, 采集时间为每天 10:00—14:00。

3.1 奶牛运动行为数据获取

采用视频监控录像比对方式观察奶牛行为活动, 使用日本 Sony 公司生产的 ILCE-7RM2 摄像机(1920 × 1080 像素) 进行视频录制, 并同步采集无线腿部传感器行为信息。此外, 通过安装耳标方式对选取的奶牛进行标号, 标号与无线腿部传感器地址一一对应。数据采集过程中, 奶牛的行为改变为自然行为, 无外界干扰。

试验采集奶牛站立、平躺、慢走、快走、站立动作、躺卧动作 6 种运动行为。奶牛运动行为详细描述如表 2 所示。因试验过程中数据丢包及其他因素影响, 试验共采集 13 954 组原始数据, 其中, 运动行为持续时间超过 4 s 的有效数据有 13 057 组数据, 用于决策树算法模型构建及测试。躺卧动作和站立动作的行为转换时间通常超过 4 s, 采用持续时间超过 4 s 的数据可有效包含奶牛行为活动的整个过程, 保证行为数据的完整性和可辨别性。采集的 6 种运动行为数据详细构成如表 3 所示。

表 2 奶牛运动行为详细描述

Table 2 Descriptions of cows behavior in detail

行为类别	详细描述
Behavior category	Detailed description
站立 Standing	奶牛四条腿完全直立
平躺 Lying	奶牛躺在卧床隔间里
慢走	奶牛以四足交替方式平缓向前行进(行进速度小于或等于 1 步/s)
快走	奶牛以四足交替方式快速向前行进(行进速度大于 1 步/s)
站立动作	奶牛躯体向前冲, 前腿先弯曲跪地, 之后后腿站立, 后躯升起, 然后前躯升起成站立状态的过程
躺卧动作	奶牛弯曲前腿, 前躯先俯下, 然后后躯降落成平躺状态的过程
Lying down	

表 3 试验采集的奶牛腿部加速度数据

Table 3 Experimental data collected on cow leg acceleration

行为类别 Behavior category	采集数据时长 Data acquisition time/h	行为数据构成 Behavioral data structure							>4 s
		<4 s	4 s	5 s	6 s	7 s	> 7 s	合计 Total	
站立 Standing	12.49	49	72	553	627	826	1 710	3 837	3 716
平躺 Lying	15.79	51	108	507	516	724	1 340	3 246	3 087
慢走 Normal walking	5.51	93	207	393	509	435	759	2 396	2 096
快走 Active walking	4.16	71	55	527	487	503	481	2 124	1 998
站立动作 Standing up	2.08	7	94	403	379	192	124	1 199	1 098
躺卧动作 Lying down	1.92	16	74	384	351	235	92	1 152	1 062
合计 Total	41.95	287	610	2 767	2 869	2 915	4 506	13 954	13 057

3.2 构建二叉决策树分类模型

训练数据集所占比例对树模型算法的准确率有较大影响。文献[35]中分析训练数据集所占比例对决策树模型

算法准确率的影响, 得出训练数据集的比例选择为 70%~90%时, 决策树算法的准确率较高。故选取 13 057 条有效数据中 9 140 条数据(占总有效数据 70%比例) 作为训练数据集, 用于构建二叉决策树模型; 3 917 条数据(占总有效数据 30%比例) 用作测试数据集, 评估二叉决策树模型性能。

构建二叉决策树分类模型时, 首先, 针对 6 种奶牛运动行为, 计算任一统计特征量的不同行为类别分组方式的目标函数值, 则可得目标函数值个数 m 。

$$m = \sum_{i=1}^6 C_6^i = 64 \quad (6)$$

比较获得最大目标函数值, 可知此统计特征量的最佳行为类别分组方式及最优阈值。同理, 计算得到 24 个统计特征量最佳行为类别分组方式及最优阈值, 各统计特征量的最大目标函数值及最优阈值如表 4 所示, 各统计特征量最大目标函数值对应的 ROC 曲线如图 3 所示。

其次, 根据式(4)和式(5)计算各统计特征量的信息增益, 计算结果如表 5 所示, 其中, 总体标准差的信息增益最高为 0.52, 选取该统计特征量作为最优二叉决策树的根结点。总体标准差所对应的最佳行为类别分组方式中正类行为组包括快走、站立动作、躺卧动作行为, 负类行为组包括慢走、平躺、站立行为, 最优阈值为 0.09 g, 由该属性阈值建立 2 个分支, 分别包含所划分的正类行为组和负类行为组。

最后, 对 2 个分支内运动行为类别递归调用该方法逐步建立最优二叉决策树分类模型的分支, 直至所有分支仅包含 1 类运动行为为止, 最终构建最优阈值选取的二叉决策树分类模型。

构造的最优二叉决策树分类模型如图 4 所示, 对于快走、站立动作、躺卧动作 3 种运动行为特征, 具有最高信息增益的统计特征量为 Y 轴均值, 最优阈值为 0.17 g; 对于慢走、平躺、站立 3 种运动行为特征, 统计特征量中信息增益最高的是 Y 轴标准差, 最优阈值为 0.05 g; 对于站立动作、躺卧动作 2 种运动行为特征, 具有最高信息增益的统计特征量为 X 轴均值, 最优阈值为 -0.21 g; 对于平躺、站立 2 种运动行为特征, 统计特征量中信息增益最高的是幅度, 最优阈值为 0.13 g。

3.3 分类结果

采用最优二叉决策树模型对测试数据集进行分类, 分类结果如表 6 所示, 846 个站立行为、826 个平躺行为、476 个慢走行为、481 个快走行为、257 个站立动作行为、229 个躺卧行为被正确分类, 其中, 站立、平躺行为的分类正确性较高, 但也存在误判为其他行为的情况; 慢走行为主要被误判为站立、快走行为(占慢走行为观测总数的 18.79%); 快走行为容易被判断为站立、慢走行为(占快走行为观测总数的 18.43%); 站立动作与躺卧动作行为容易混淆(分别占站立动作行为观测总数的 15.53%、躺卧动作行为观测总数的 14.92%)。

表 4 6 种行为类别下各统计特征量的最大目标函数值及最优阈值

Table 4 Maximum objective function value and optimal threshold of each statistical feature corresponding to 6 behavior categories								
统计特征量 Statistical feature	最大目标函数值 Maximum objective value	最优阈值 Optimal threshold	统计特征量 Statistical feature	最大目标函数值 Maximum objective value	最优阈值 Optimal threshold	统计特征量 Statistical feature	最大目标函数值 Maximum objective value	最优阈值 Optimal threshold
X 轴均值 X-axis mean	0.89	-0.53 g	Z 轴峰度 Z-axis kurtosis	0.81	0.01	Y 轴最小值 Minimum value of Y-axis	0.82	0.28 g
Y 轴均值 Y-axis mean	0.91	0.42 g	X 轴标准差 Standard deviation of X-axis	0.82	0.06 g	Z 轴最小值 Minimum value of Z-axis	0.73	0.13 g
Z 轴均值 Z-axis mean	0.83	0.14 g	Y 轴标准差 Standard deviation of Y-axis	0.79	0.12 g	总体标准差 Overall standard deviation	0.91	0.09 g
X 轴偏度 X-axis skewness	0.90	0.06	Z 轴标准差 Standard deviation of Z-axis	0.75	0.02 g	能量 Energy	0.89	0.21
Y 轴偏度 Y-axis skewness	0.79	0.11	X 轴最大值 Maximum value of X-axis	0.82	-0.25 g	幅度 Signal magnitude area	0.88	0.14 g
Z 轴偏度 Z-axis skewness	0.74	0.02	Y 轴最大值 Maximum value of Y-axis	0.69	0.49 g	矢量幅度 Signal vector magnitude	0.89	0.03 g
X 轴峰度 X-axis kurtosis	0.72	0.04	Z 轴最大值 Maximum value of Z-axis	0.78	0.37 g	熵 Entropy	0.76	0.26
Y 轴峰度 Y-axis kurtosis	0.86	0.07	X 轴最小值 Minimum value of X-axis	0.86	-0.67 g	倾角 Inclination	0.87	15.73°

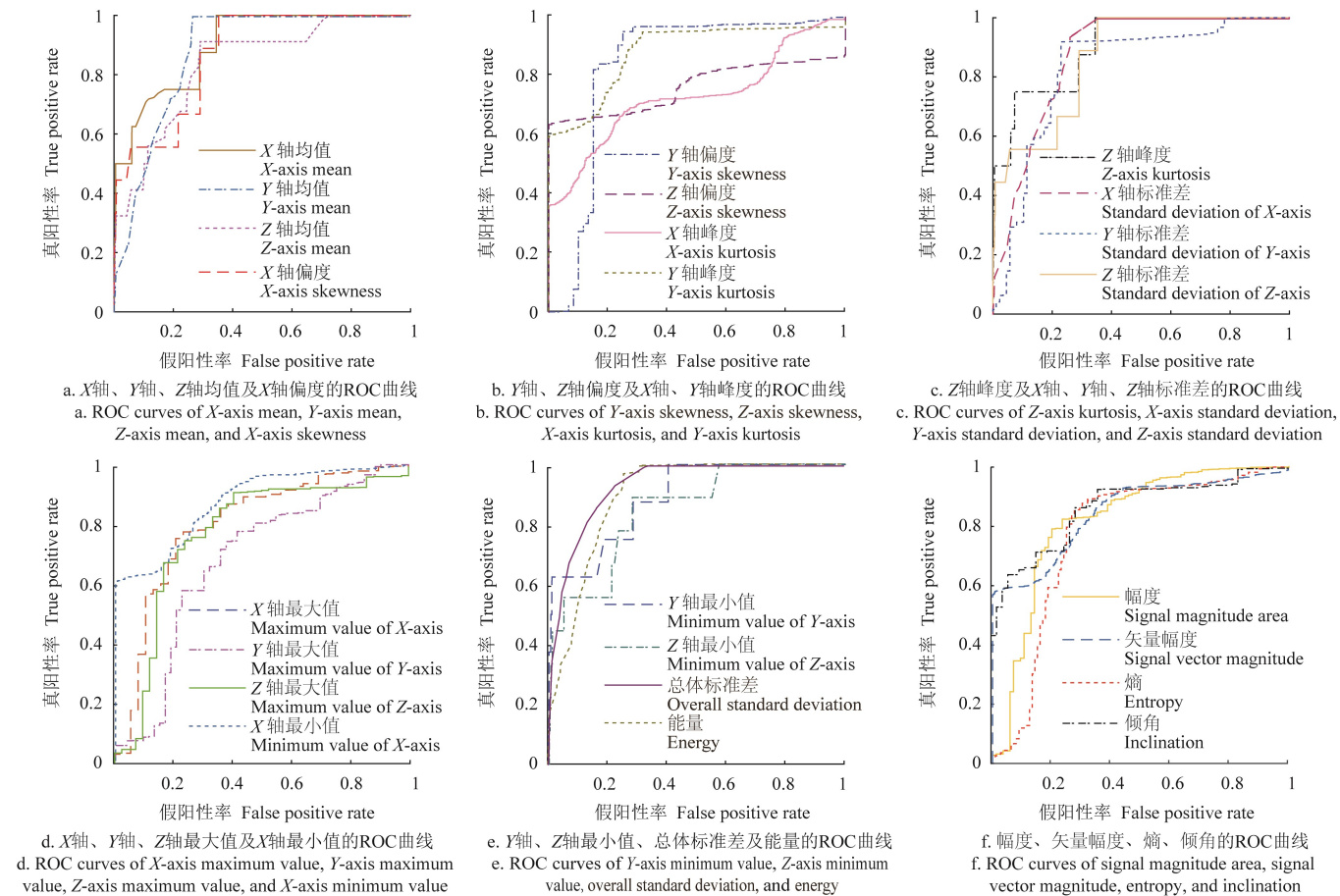


图 3 各统计特征量的最大目标函数值对应的 ROC 曲线
Fig.3 ROC curve corresponding to maximum objective function value of each statistical feature

混淆矩阵分类结果运用灵敏度、精度、F1 值 3 种性能指标来度量，灵敏度、精度、F1 值的定义如下：

灵敏度= $\frac{TP}{TP+FN} \times 100\%$ (7)

精度= $\frac{TP}{TP+FP} \times 100\%$ (8)

F1= $\frac{2 \times TP}{2 \times TP+FP+FN} \times 100\%$ (9)

灵敏度表示通过分类模型各个行为被正确识别为该行为的比例。站立、平躺行为的灵敏度较高，均超过 85%。

精度反映各行被正确预测的比例，站立、平躺、慢走行为的精度均超过 80%；站立动作、躺卧动作行为的精度偏低。6 种行为的平均精度为 76.83%，且均超过 65%。

F1 值是灵敏度和精度的调和平均值，相当于灵敏度和精度的综合评价指标，平躺行为的 F1 值最高；站立动作、躺卧动作行为的 F1 值偏低，分别低于平躺行为 18.28 个百分点、20.01 个百分点。

试验表明，最优二叉决策树模型对行为的识别率与

奶牛的运动状态有关，奶牛处于静止状态（站立、平躺行为）时，分类的灵敏度、精度、F1 值较高；奶牛处于运动状态（慢走、快走、站立动作、躺卧动作行为）时，分类的灵敏度、精度、F1 值偏低。其次，最优二叉决策树模型的灵敏度、精度、F1 值和数据量有密切联系，站立、平躺、慢走、快走行为的数据训练量较高，分类的灵敏度、精度、F1 值偏高；站立动作、躺卧动作行为的训练数据样本较少，分类的灵敏度、精度、F1 值较低，因此需要获取更多的站立动作与躺卧动作行为的数据样本用于算法学习。

表 5 6 种行为类别下各统计特征量的信息增益
Table 5 Information gain for each statistical feature under six behavior categories

统计特征量 Statistical feature	信息增益 Information gain	统计特征量 Statistical feature	信息增益 Information gain	统计特征量 Statistical feature	信息增益 Information gain
X 轴均值 X-axis mean	0.43	Z 轴峰度 Z-axis kurtosis	0.33	Y 轴最小值 Minimum value of Y-axis	0.43
Y 轴均值 Y-axis mean	0.32	X 轴标准差 Standard deviation of X-axis	0.19	Z 轴最小值 Minimum value of Z-axis	0.31
Z 轴均值 Z-axis mean	0.37	Y 轴标准差 Standard deviation of Y-axis	0.28	总体标准差 Overall standard deviation	0.52
X 轴偏度 X-axis skewness	0.46	Z 轴标准差 Standard deviation of Z-axis	0.31	能量 Energy	0.47
Y 轴偏度 Y-axis skewness	0.27	X 轴最大值 Maximum value of X-axis	0.42	幅度 Signal magnitude area	0.36
Z 轴偏度 Z-axis skewness	0.41	Y 轴最大值 Maximum value of Y-axis	0.38	矢量幅度 Signal vector magnitude	0.32
X 轴峰度 X-axis kurtosis	0.32	Z 轴最大值 Maximum value of Z-axis	0.27	熵 Entropy	0.46
Y 轴峰度 Y-axis kurtosis	0.29	X 轴最小值 Minimum value of X-axis	0.43	倾角 Inclination	0.36

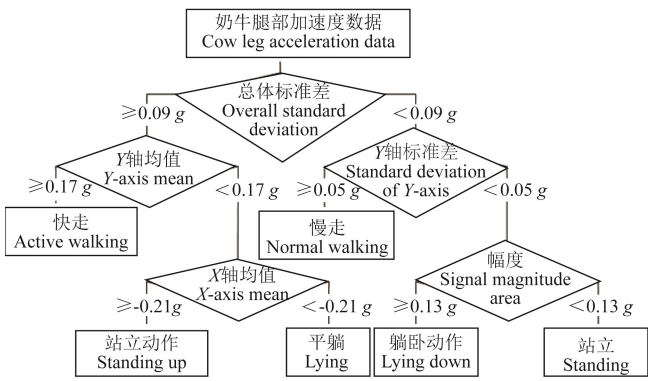


图 4 最优二叉决策树分类模型
Fig.4 Optimal binary decision-tree classification model

表 6 二叉决策树分类模型分类结果

观察行为 Observed behaviors	预测行为 Predicted behaviors						行为总数 Total number	性能指标 Performance index/%	
	站立 Standing	平躺 Lying	慢走 Normal walking	快走 Active walking	站立动作 Standing up	躺卧动作 Lying down	of observations	灵敏度 Sensitivity	精度 Precision
站立 Standing	846	50	13	9	26	11	955	88.59	83.35
平躺 Lying	43	826	15	21	32	24	961	85.95	86.04
慢走 Normal walking	47	25	476	74	5	17	644	73.91	84.25
快走 Active walking	68	17	49	481	8	12	635	75.75	74.46
站立动作 Standing up	7	18	3	36	257	59	380	67.63	67.81
躺卧动作 Lying down	4	24	9	25	51	229	342	66.96	65.06
合计 Total	1 015	960	565	646	379	352	3 917	76.47	76.83

3.4 算法性能比较

在相同的计算机软硬件环境下进行算法比较，其中操作系统为 Window 10，CPU 采用 Intel i5-8400 处理器，主频为 3.4 GHz，内存容量为 8 GB，算法仿真平台软件为 Matlab 2014 b，采用传统的 ID3 决策树算法^[36]、K-means 聚类算法^[37]、支持向量机算法^[29]与本文提出的最优二叉决策树算法进行对比，训练集和测试集和本文算法一致。ID3 决策树算法选取 X 轴均值、X 轴方差、总体方差、Z 轴均值、Y 轴均值 5 个统计特征量值作为分裂属性，属性阈值分别设置为-0.7 g、-0.01 g、0.1 g²、0.02 g、0.25 g，选取信息增益作为属性选取标准进行构建决策树模型；K-means 聚类算法从全部有效数据中任选 6 个数据点作为初始簇心，选取欧式距离作为相近度的评价指标进行模型构建；支持向量机算法采用一类对多类方法，首先构建 6 个二分分类器，每个分类器构造一个行为与其他 5 类行为的超平面，设置正则化常数 C=50，间隔 γ=1，算法分类结果如图 5 所示。

ID3 决策树算法分类的平均灵敏度为 70.76%，平均精度为 71.43%，平均 F1 值为 70.96%，分别低于最优二叉决策树算法 5.71、5.4 和 5.61 个百分点。其中，慢走、站立动作行为的灵敏度较低，相较于最优二叉决策树算法，分别低 10.56 和 10.52 个百分点。

K-means 聚类算法判别的平均灵敏度为 68.96%，与最优二叉决策树算法相比低 7.51 个百分点；平均精度为 68.81%，低于最优二叉决策树算法 8.02 个百分点，其中，站立动作、躺卧动作的精度偏低，分别低于最优二叉决策树算法 12.85 和 11.28 个百分点；平均 F1 值为 68.8%，低于最优二叉决策树算法 7.77 个百分点。

支持向量机分类的平均灵敏度为 69.7%，平均精度为 70.11%，平均 $F1$ 值为 69.82%，分别低于最优二叉决策树算法 6.77、6.72 和 6.57 个百分点。其中，站立、平躺行为的分类效果较好，相较最优二叉决策树算法，分别提升了 1.57、0.53 个百分点。慢走、快走、躺卧动作行为识别率偏低，慢走、躺卧动作行为的精度分别低于最优二叉决策树算法 18.4 和 12.21 个百分点；与最优二叉决策树相比，快走、躺卧动作的灵敏度分别低 14.49 和 12.21 个百分点。

通过对比可知，ID3 决策树算法在慢走、站立动作行为上判别效果不佳；K-means 聚类算法在站立动作、躺卧动作行为分类效果不理想；支持向量机算法尽管在站立、平躺行为的分类效果较好，但是在慢走、快走、躺卧动作行为的识别效果偏差。试验结果表明，与其他 3 种分类算法相比，本文提出的最优二叉决策树分类模型具有更高的行为识别率，可实现奶牛运动行为的准确识别。

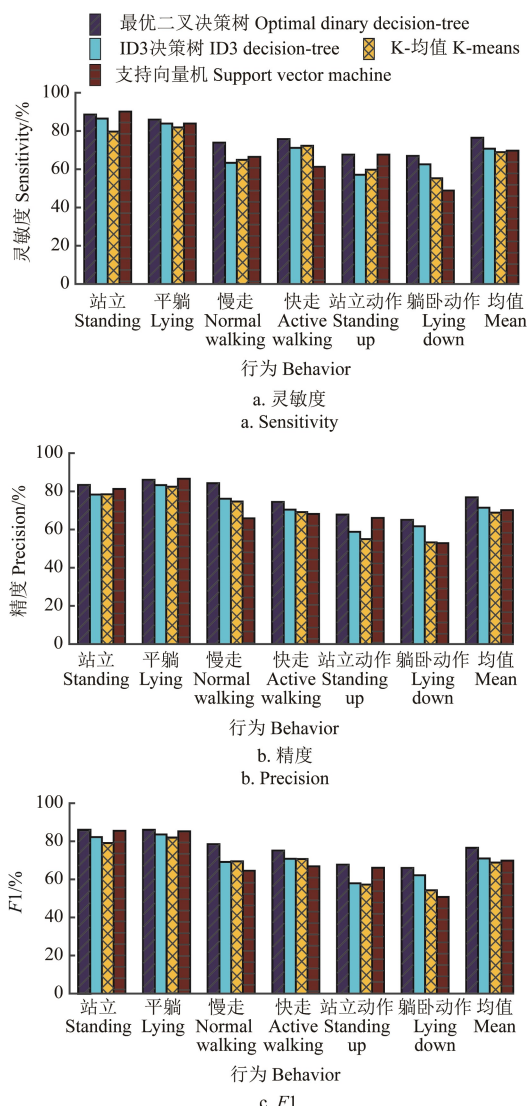


图5 算法分类性能指标比较

Fig.5 Comparison of performance indexes of algorithm classification

4 结论

本文开发一种无线腿部传感器设备，该传感器集成

微处理器 MSP430F149IMP、三轴加速度计 ADXL345、无线收发器 CC1101，具有功耗低、运行稳定、传输可靠的优点，满足采集奶牛腿部加速度数据的工作要求。

1) 提出了一种最优二叉决策树属性阈值选择方法，该方法基于 ROC (receiver operating characteristic, ROC) 曲线原理获取各统计特征量的最大目标函数值，得到各统计特征量的最佳行为类别分组方式及最优阈值，实现奶牛运动行为的高效分类。

2) 该算法能够有效区分奶牛站立、平躺、慢走、快走、站立动作、躺卧动作 6 类运动行为，分类的平均准确率、平均精度、平均 $F1$ 值分别为 76.47%、76.83%、76.57%，相较传统的 ID3 决策树算法分别高 5.71、5.4 和 5.61 个百分点。该算法具有分类规则简单、分类准确率高优点，可为提升奶牛行为分类水平提供参考。

[参考文献]

- [1] Zehner N, Umstätter C, Niederhauser J J, et al. System specification and validation of a noseband pressure sensor for measurement of ruminating and eating behavior in stable-fed cows[J]. Computers and Electronics in Agriculture, 2017, 136: 31–41.
- [2] Gonzalez L A, Bishop-Hurley G J, Handcock R N, et al. Behavioral classification of data from collars containing motion sensors in grazing cattle[J]. Computers and Electronics in Agriculture, 2015, 110: 91–102.
- [3] 田富洋, 王冉冉, 刘莫尘, 等. 基于神经网络的奶牛发情行为辨识与预测研究[J]. 农业机械学报, 2013, 44(增刊 1): 277–281.
Tian Fuyang, Wang Ranran, Liu Mochen, et al. Oestrus detection and prediction in dairy cows based on neural networks[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(Supp.1): 277–281. (in Chinese with English abstract)
- [4] Mattachini G, Riva E, Bisaglia C, et al. Methodology for quantifying the behavioral activity of dairy cows in freestall barns[J]. Journal of Animal Science, 2013, 91(10): 4899–4907.
- [5] Khanh P C P, Chinh N D, Cham T T, et al. Classification of cow behavior using 3-DOF accelerometer and decision tree algorithm[C]// International Conference on Biomedical Engineering, 2016: 45–50.
- [6] Arcidiacono C, Porto S M C, Mancino M, et al. Development of a threshold-based classifier for real-time recognition of cow feeding and standing behavioural activities from accelerometer data[J]. Computers and Electronics in Agriculture, 2017, 134: 124–134.
- [7] Thorup V M, Munksgaard L, Robert P E, et al. Lameness detection via leg-mounted accelerometers on dairy cows on four commercial farms[J]. Animal An International Journal of Animal Bioscience, 2015, 9(10): 1704–1712.
- [8] 沈维政, 郑爽, 初永良, 等. 基于 ADXL345 的奶牛活动量采集系统设计[J]. 东北农业大学学报, 2014, 45(10): 80–85.
Shen Weizheng, Zheng Shuang, Chu Yongliang, et al. Design of cow activity data acquisition system based on ADXL345[J]. Journal of Northeast Agricultural University, 2014, 45(10): 80–85. (in Chinese with English abstract)
- [9] 温长吉, 王生生, 赵昕, 等. 基于视觉词典法的母牛产前行为识别[J]. 农业机械学报, 2014, 45(1): 266–274.
Wen Changji, Wang Shengsheng, Zhao Xin, et al. Visual

- dictionary for cows sow behavior recognition[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(1): 266—274. (in Chinese with English abstract)
- [10] Smith D, Dutta R, Hellicar A, et al. Bag of class posteriors: A new multivariate time series classifier applied to animal behaviour identification[J]. Expert Systems with Applications, 2015, 42(7): 3774—3784.
- [11] Dutta R, Smith D, Rawnsley R, et al. Cattle behaviour classification using 3-axis collar sensor and multi-classifier pattern recognition[C]//2014 IEEE Sensors, 2014: 1272—1275.
- [12] 尹令, 洪添胜, 刘汉兴, 等. 结构相似子序列快速聚类算法及其在奶牛发情检测中的应用[J]. 农业工程学报, 2012, 28(15): 107—112.
Yin Ling, Hong Tiansheng, Liu Hanxing, et al. Subsequence clustering algorithm based on structural similarity and its application in cow estrus detection[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(15): 107—112. (in Chinese with English abstract)
- [13] Nielsen L R, Pedersen A R, Herskin M S, et al. Quantifying walking and standing behaviour of dairy cows using a moving average based on output from an accelerometer[J]. Applied Animal Behaviour Science, 2010, 127(1/2): 12—19.
- [14] Nadimi E S, Jørgensen R N, Blanes-Vidal V, et al. Monitoring and classifying animal behavior using zigbee-based mobile ad hoc wireless sensor networks and artificial neural networks[J]. Computers and Electronics in Agriculture, 2012, 82: 44—54.
- [15] Leroy T, Guarino M, Vranken E, et al. Automatic real-time monitoring of locomotion and posture behaviour of pregnant cows prior to calving using online image analysis[J]. Computers and Electronics in Agriculture, 2008, 64(1): 53—60.
- [16] Zhao K, Bewley J M, He D, et al. Automatic lameness detection in dairy cattle based on leg swing analysis with an image processing technique[J]. Computers and Electronics in Agriculture, 2018, 148: 226—236.
- [17] Smith D, Rahman A, Bishop-Hurley G J, et al. Behavior classification of cows fitted with motion collars: Decomposing multi-class classification into a set of binary problems[J]. Computers and Electronics in Agriculture, 2016, 131: 40—50.
- [18] 何东健, 刘冬, 赵凯旋. 精准畜牧业中动物信息智能感知与行为检测研究进展[J]. 农业机械学报, 2016, 47(5): 231—244.
He Dongjian, Liu Dong, Zhao Kaixuan. Review of perceiving animal information and behavior in precision livestock farming[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(5): 231—244. (in Chinese with English abstract)
- [19] 屈东东, 刘素梅, 吴金杰, 等. 群养奶牛体温实时监测系统设计与实现[J]. 农业机械学报, 2016, 47(增刊 1): 408—412.
Qu Dongdong, Liu Sumei, Wu Jinjie, et al. Design and implementation of monitoring system for multiple cows body temperature[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(Supp.1): 408—412. (in Chinese with English abstract)
- [20] Thompson A J, Weary D M, von Keyserlingk M A G. Technical note: Mining data from on-farm electronic equipment to identify the time dairy cows spend away from the pen[J]. Journal of Dairy Science, 2017, 100(5): 3975—3982.
- [21] Nielsen P P, Fontana I, Sloth K H, et al. Technical note: Validation and comparison of 2 commercially available activity loggers[J]. Journal of Dairy Science, 2018, 101(6): 5449—5453.
- [22] Watanabe N, Sakanoue S, Kawamura K, et al. Development of an automatic classification system for eating, ruminating and resting behavior of cattle using an accelerometer[J]. Grassland Science, 2008, 54(4): 231—237.
- [23] Resheff Y S, Rotics S, Harel R, et al. Accelerator: A web application for supervised learning of behavioral modes from acceleration measurements[J]. Movement Ecology, 2014, 2(1): 27—33.
- [24] Brown D D, Kays R, Wikelski M, et al. Observing the unwatchable through acceleration logging of animal behavior[J]. Animal Biotelemetry, 2013, 1(1): 20—36.
- [25] Shahriar M S, Smith D, Rahman A, et al. Detecting heat events in dairy cows using accelerometers and unsupervised learning[J]. Computers and Electronics in Agriculture, 2016, 128: 20—26.
- [26] Langrock R, King R, Matthiopoulos J, et al. Flexible and practical modeling of animal telemetry data: Hidden markov models and extensions[J]. Ecology, 2012, 93(11): 2336—2342.
- [27] 何东健, 孟凡昌, 赵凯旋, 等. 基于视频分析的犊牛基本行为识别[J]. 农业机械学报, 2016, 47(9): 294—300.
He Dongjian, Meng Fanchang, Zhao Kaixuan, et al. Recognition of calf basic behaviors based on video analysis[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(9): 294—300. (in Chinese with English abstract)
- [28] 尹令, 刘财兴, 洪添胜, 等. 基于无线传感器网络的奶牛行为特征监测系统[J]. 农业工程学报, 2010, 26(3): 203—208.
Yin Ling, Liu Caixing, Hong Tiansheng, et al. Design of system for monitoring dairy cattle's behavioral features based on wireless sensor networks[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2010, 26(3): 203—208. (in Chinese with English abstract)
- [29] Martiskainen P, Järvinen M, Skön J-P, et al. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines[J]. Applied Animal Behaviour Science, 2009, 119(2): 32—38.
- [30] Diosdado J A V, Barker Z E, Hodges H R, et al. Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system[J]. Animal Biotelemetry, 2015, 3(1): 15—28.
- [31] Gupta R K, Lathwal S S, Ruhil A P, et al. Lameness prediction in Karan Fries cross-bred cows using decision tree models[C]// International Conference on Computing for Sustainable Global Development. IEEE, 2015: 1166—1171.
- [32] Abell K M, Theurer M E, Larson R L, et al. Predicting bull behavior events in a multiple-sire pasture with video analysis, accelerometers, and classification algorithms[J]. Computers and Electronics in Agriculture, 2017, 136: 221—227.
- [33] Alvarenga F A P, Borges I, Palkovic L, et al. Using a three-axis accelerometer to identify and classify sheep behaviour at pasture[J]. Applied Animal Behaviour Science, 2016, 181: 91—99.
- [34] 于文婧, 刘晓娜, 孙丹峰, 等. 基于 HJ-CCD 数据和决策树法的干旱半干旱灌区土地利用分类[J]. 农业工程学报, 2016, 32(2): 212—219.
Yu Wenjing, Liu Xiaona, Sun Danfeng, et al. Land use classification in arid and semi-arid irrigated area based on

- HJ-CCD data and decision tree method[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2016, 32(2): 212–219. (in Chinese with English abstract)
- [35] 程华, 李艳梅, 罗谦, 等. 基于 C4.5 决策树方法的到港航班延误预测问题研究[J]. 系统工程理论与实践, 2014, 34(增刊 1): 239–247.
- Cheng Hua, Li Yanmei, Luo Qian, et al. Study on flight delay with C4.5 decision tree based prediction method[J]. Systems Engineering-Theory & Practice, 2014, 34(Supp.1): 239–247. (in Chinese with English abstract)
- [36] Steensels M, Antler A, Bahr C, et al. A decision-tree model to detect post-calving diseases based on rumination, activity, milk yield, bw and voluntary visits to the milking robot[J]. Animal, 2016, 10(9): 1493–1500.
- [37] Schwager M, Anderson D M, Butler Z, et al. Robust classification of animal tracking data[J]. Computers and Electronics in Agriculture, 2007, 56(1): 46–59.

Cow movement behavior classification based on optimal binary decision-tree classification model

Wang Jun¹, Zhang Haiyang¹, Zhao Kaixuan¹, Liu Gang²

(1. College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471003, China;

2. Key Laboratory for Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing 100083, China)

Abstract: Changes in behavioral activity are increasingly recognized as a useful indicator of dairy cows' health and welfare. The classifying of changes in behavioral activity can be useful in early detection and prevention of diseases, and monitoring dairy cows' behavioral activity helps farmers to take a comprehensive view of the dairy cows' estrus time. The aim of this study is to automatically measure and distinguish several behavior activities of dairy cows from accelerometer data. The study consists of 2 parts, namely, wireless leg sensor and binary decision-tree algorithm. The wireless leg sensor was designed to collect test data, which integrates microcontroller MSP430F149IMP, tri-axial accelerometer ADXL345, and radio frequency module CC1101 to meet the requirements of accurately collecting data of the acceleration of dairy cows, and long-term reliable transmission of data. The binary decision-tree algorithm was designed to classify the behavior of dairy cows. Firstly, 24 statistical features describing the magnitude, symmetry, steepness, variability, uncertainty and angle of the three-axis acceleration of cow legs were selected. Secondly, the best classification behavior category and optimal threshold of each statistical feature were obtained by constructing ROC(receiver operating characteristic) curve. Then the information gain is used as the selection criterion for the split attribute of the binary decision-tree model. Finally, a optimal binary decision tree classification model is constructed to classify and recognize the dairy cow motion behavior. Compared with the traditional binary decision-tree algorithm, the innovation of the algorithm is as follows: Firstly, the ROC curve principle is used to ensure the classification and threshold of each statistical feature to select the local optimal. Then the information gain is used as the split attribute selection standard, and the binary decision-tree classification model is constructed to realize the overall optimal classification of the behavior characteristics of the dairy cows. The results illustrate that the optimal binary decision-tree algorithm can accurately classify 6 types of biologically relevant behavior: standing (88.59% sensitivity, 83.35% precision, and 85.89% *F1* score), lying (85.59% sensitivity, 86.04% precision, and 86% *F1* score), normal walking (73.91% sensitivity, 84.25% precision, and 78.74% *F1* score), active walking (75.75% sensitivity, 74.46% precision, and 75.1% *F1* score), standing up (67.63% sensitivity, 67.81% precision, and 67.72% *F1* score), and lying down (66.96% sensitivity, 65.06% precision, and 65.99% *F1* score). The highest sensitivity was 88.59% for standing and the sensitivity was good for all classes of behavior except standing up and lying down. The best precision was achieved for standing, lying, and normal walking. The precision for active walking classification was slightly lower but substantially better than those for standing up and lying down. Standing and lying behavior were classified correctly to a high degree, but were also misclassified as other behavior. Normal walking was mainly misclassified as either standing or active walking (18.79% of the cases). Active walking was misclassified most often as standing or normal walking (18.43% of the cases). Standing up and lying down were mostly confused with each other (15.53% and 14.92% of the cases, respectively). The average sensitivity, the average precision and the average *F1* score of the classification are 76.47%, 76.83%, and 76.57% respectively. Compared with the traditional ID3 (iterative dichotomiser 3) decision-tree algorithm, they are increased by 5.71 percentage points, 5.4 percentage points and 5.61 percentage points respectively; they are increased by 7.51 percentage points, 8.02 percentage points and 7.77 percentage points respectively compared with the K-means clustering algorithm, and 6.77 percentage points, 6.72 percentage points and 6.57 percentage points respectively compared with the support vector machine algorithm. The experimental results show that the optimal binary decision-tree algorithm has the characteristics of simple classification rules and high classification accuracy. This research of the method can provide an effective theoretical support for improving the classification accuracy of dairy cow behavior.

Keywords: data acquisition; data processing; algorithms; behavior classification; tri-axial accelerometer; wireless leg sensor; receiver operating characteristic curve; binary Decision-tree