

基于农机运行轨迹点多维特征的作业状态辨识方法与试验

李保忠^{1,2} 王伟鹏² 周国民^{3,4} 徐培² 申及²

(1. 中国农业科学院农业信息研究所, 北京 100081; 2. 洛阳智能农业装备研究院有限公司, 洛阳 471000;

3. 三亚中国农业科学院国家南繁研究院, 三亚 572024; 4. 农业农村部南京农业机械化研究所, 南京 210014)

摘要: 农业机械(农机)的作业状态是评估农业机械化效率和精准管理的关键指标,为实现数据管理平台对农机作业状态的精准监测,本文提出一种基于农机运行轨迹多维特征的作业状态辨识方法。首先,基于物联网现代信息技术支持的大数据管理平台,研究了农机作业空间轨迹点的潜在特征,分析了速度、平均加速度、转向率和轨迹点密度等特征的分布规律。其次,依据各特征的分布特点和作业状态辨识需求,采用多策略分箱处理方法对特征进行定量划分,并引入证据权重(Weight of evidence, WOE)和信息值(Information value, IV)方法量化不同特征对农机作业状态的影响权重,从而评估关键特征对作业状态的识别能力。最后,基于农机空间轨迹点的多维关键特征,结合BP神经网络与AdaBoost的融合算法对农机作业状态进行辨识。试验结果表明,所提出的算法模型在农机作业状态预测中的准确率高达97.3%,表明基于农机多维特征的辨识方法可准确辨识农机的作业状态。

关键词: 农业机械; 空间运行轨迹; 多策略分箱; 作业状态辨识; BP-AdaBoost

中图分类号: O657.3; S126

文献标识码: A

文章编号: 1000-1298(2025)12-0150-08

OSID:



Operation Status Recognition Method and Experiment Based on Multidimensional Features of Agricultural Machinery Spatial Track

LI Baozhong^{1,2} WANG Weipeng² ZHOU Guomin^{3,4} XU Pei² SHEN Ji²

(1. Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China

2. Luoyang Smart Agricultural Equipment Institute Co., Ltd., Luoyang 471000, China

3. National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya 572024, China

4. Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China)

Abstract: The operation state of agricultural machinery is a key indicator for assessing the efficiency of agricultural mechanization and precise management. To realize the precise monitoring of the operation state of farm machinery by the data management platform, an operation state identification method was proposed based on the multidimensional features of the spatial track of farm machinery. Firstly, based on the big data management platform supported by the modern information technology of Internet of Things, the potential characteristics of the trajectory points in the operating space of agricultural machinery were studied, and the distribution laws of the characteristics such as speed, acceleration, steering rate and distribution density of trajectory points were analyzed. Secondly, based on the distribution characteristics of each feature and the demand for operation state recognition, a multi-strategy split-box processing method was used to quantitatively divide the features, and weight of evidence (WOE) and information value (IV) methods were introduced to quantify the influence weight of different features on the operation state of the farm machinery, so as to assess the impact on operation state the key features of the recognition ability were evaluated. Finally, based on the multidimensional key features of the spatial trajectory points of the farm machinery, the fusion algorithm of BP neural network and AdaBoost was combined to recognize the operation state of the farm machinery. The experimental results showed that the accuracy of the proposed algorithm model in the prediction of the operation state of agricultural machinery was as high as 97.3%, indicating that the recognition method based on the multidimensional features of

收稿日期: 2025-04-11 修回日期: 2025-07-01

基金项目: 国家农机装备创新中心计划项目(2024A02)

作者简介: 李保忠(1990—),男,博士生,洛阳智能农业装备研究院有限公司工程师,主要从事农机装备智能化研究,E-mail: libaozhong@chiaic.com

通信作者: 周国民(1969—),男,研究员,博士生导师,主要从事农业科学数据治理与共享研究,E-mail: zhouguomin@caas.cn

agricultural machinery can accurately recognize the operation state of agricultural machinery.

Key words: agricultural machinery; spatial track; multi-strategy binning; operation status recognition; BP - AdaBoost

0 引言

随着农业现代化和机械化水平的不断提升,物联网技术在农业生产中的作用日益重要。现代信息技术在农机生产监测和管理过程中的应用也变得愈加广泛^[1-2]。目前,监管人员已经可以通过农机搭载的智能终端信息监测设备对农机进行定位^[3-4]、追踪^[5-6]和远程监测^[7-8],并开始对农机作业运行中产生的实时监测数据进行分析,研究农机作业状态,进而研究农机作业效率^[9-10]、农机作业区域识别^[11-12]、田间作业路径规划^[13-14]等。然而,现有的农机作业状态辨识方法通常仅依赖单一或少数几种特征进行分析^[15]。事实上,农机作业运行过程中产生了多种空间运动轨迹数据来反映农机的不同运行状态。通过结合多种特征参数进行分析,能够更精准地推算出农机作业状态,因此迫切需要对农机空间运行轨迹数据进行深度挖掘,探究多种特征与农机作业状态的潜在规律,构建农机作业状态辨识模型,为智能农业系统提供重要的信息支撑。

针对上述问题,本文设计一种基于农机多维信息的作业状态辨识方法。以空间数据挖掘为基础,结合农机空间运行轨迹的特点,选择农机速度、平均加速度、转向率、轨迹点密度等作为特征。首先,采用 WOE 算法对特征进行编码,将其转化为数值特征;然后,通过 IV 方法量化特征对农机作业状态辨识的影响权重,明确关键特征;最后,依据多维关键特征,基于 BP 神经网络与 AdaBoost 集成算法,构建多维特征融合的作业状态辨识模型。

1 农机多维信息的作业状态辨识方法

1.1 农机作业多维信息的分布特点

农机作业班次中,不同的作业状态(如田间作业、空行转移、停歇等)会导致农机的多维信息表现出不同的分布特征。在进行正常田间作业时,农机速度通常保持在较低且稳定的水平,同时由于局部往复运动,轨迹点在空间上会呈现明显的聚集分布,轨迹点密度较大,且转向操作频繁;在地块间的空行转移阶段,农机的速度波动较大,转向次数较少,导致转向率显著下降,轨迹点一般呈现为线状分布;当农机处于停歇或维修检查状态时,远程监测到的数据特征趋于稳定,空间轨迹点的位置几乎保持不变,在小范围内呈散点簇分布。

基于上述特征分析,通过挖掘农机作业运行数据,可以提取农机速度、平均加速度、转向率和空间轨迹点密度分析农机作业状态,量化其多维特征与农机作业状态的重要程度,为作业状态辨识模型的构建提供理论依据。

1.2 农机作业状态关键特征分析方法

1.2.1 数据预处理

1.2.1.1 数据清洗

农机上安装农机的智能终端设备,能够自动记录农机运行全过程的特征数据,包括速度、经纬度、方位角等信息。由于定位精度问题,农机处于停歇状态时,轨迹点会呈现一定范围的散布,而非集中于单一位置,该现象会导致一定范围内轨迹点数量增大,影响单位面积轨迹点的密度监测。此外,定位偏差较大或设备故障时,可能导致轨迹点的间距异常增大,该异常点称为轨迹漂移点。农机轨迹漂移点示例如图 1 所示。



图 1 农机轨迹漂移点示例

Fig. 1 Example of drift point in trajectory of farm machine

为保证数据质量和可靠性,对无效轨迹点进行处处理,包括农机停歇散点与行驶漂移点。当农机处于停歇状态下,其轨迹点在小范围内密集,且运行速度基本为零,通过多次实验验证设置合适的速度阈值,并判断连续多个点的速度是否小于该阈值,从而对有效数据进行提取;当相邻轨迹点间距过大时,判断当前轨迹点与下一轨迹点间距是否超过前两个轨迹点间距,若间距数值表现为突变,则对该轨迹点数据剔除,避免对后续计算造成干扰。

1.2.1.2 下采样处理

农机运行轨迹点基本特征包含时间、速度、经纬度和方位角等。通过对轨迹点的基础特征进一步计算,可以得到第 i 个轨迹点的平均加速度、转向率和点密度等特征指标。其中,点密度是指每 10 m^2 内的轨迹点数量,而平均加速度和转向率的计算式为

$$a_i = \frac{\Delta v_i}{\Delta t_i} \quad (1)$$

$$\omega_i = \frac{\Delta \alpha_i}{\Delta t_i} \quad (2)$$

式中 a_i ——轨迹点的加速度

Δv_i ——轨迹点的速度变化量

Δt_i ——轨迹点的时间变化量

ω_i ——轨迹点的转向率

$\Delta \alpha_i$ ——轨迹点的方位角变化量

本文采用 WOE 和 IV 方法量化评估速度、平均加速度、转向率和轨迹点密度 4 种轨迹点特征对农机作业状态的重要程度。由于 WOE 值的计算依赖于农机作业状态的正负样本比例,当特征数据中样本不均衡可能导致 WOE 值失真。数据偏向多数类时,WOE 值可能过度反应多数类的特征,从而低估少数类的影响。对于 IV 值,不均衡数据会使稀有类别的特性不明显,导致特征的区分能力被低估,同时使 IV 值对主导类别的样本比例过度依赖,从而忽视少数类别对模型的潜在贡献,进而低估特征变量的真实预测能力。通过上述分析,本文通过合理分配正负样本比例,优化分组策略解决不均衡的问题,提升 WOE 和 IV 的可靠性及特征评估的准确性。

1.2.1.3 特征字段分箱

针对农机作业状态的多维特征,本文采用多种分箱方法,以准确反映各特征在不同作业状态下的分布规律。考虑到农机在作业过程中速度的波动性、不确定性及大数据量特征,本文引入决策树分箱方法进行数据预处理。根据速度特征与作业状态的分布情况,动态划分高、低速度段,能够自适应地捕捉作业状态与速度之间的非线性关系,从而实现对不同作业状态下速度特征的精确区分;平均加速度反映农机作业过程中的稳定性及动态变化,采用等宽分箱方法将平均加速度的取值范围划分为若干等间隔的区间,能够有效捕捉平均加速度的不同变化水平,处理平均加速度的数值连续性,并保证分箱后的数据均衡性;在农机局部往复作业过程中,转向频率较高,为精准捕捉频繁转向行为,采用等频分箱方法将数据划分为具有相同样本数量的多个区间,确保转向频率较高的区间得到充分表达;农机轨迹点的密度在作业状态和非作业状态下差异显著,为捕捉轨迹点密度的分布特征,同样采用等宽分箱方法对轨迹点密度进行分箱。

基于上述分箱方法,能够更有效地处理特征数据的分布,确保每个分箱内的数据分布合理,提高 WOE 和 IV 计算结果的可靠性和有效性。

1.2.2 特征参数重要度分析方法

WOE 和 IV 方法能够揭示农机作业轨迹点属性与作业状态(1/0)的相关性^[16]。WOE 用于量化不同特征分箱在区分正负样本方面的贡献,其中正样本代表农机作业状态,负样本代表农机非作业状态。

其中 WOE 计算公式为

$$WOE_i = \ln \left(\frac{Good_i / Good_{total}}{Bad_i / Bad_{total}} \right) \quad (3)$$

式中 WOE_i ——第 i 个分箱的 WOE 值

$Good_i$ ——第 i 个分箱中属于正样本的数量

$Good_{total}$ ——所有样本中正样本的数量

Bad_i ——第 i 个分箱中属于负样本的数量

Bad_{total} ——所有样本中负样本的数量

WOE 值的计算提供了速度、平均加速度、转向率、轨迹点密度等单个特征不同分箱对农机作业状态预测的度量,但仅能反映特征分箱与作业状态之间的关系,无法量化整体特征对农机作业状态预测的重要性。为评估不同特征的整体预测能力,进一步计算 IV 值,即

$$IV = \sum_{i=1}^n ((GoodRatio_i - BadRatio_i) WOE_i) \quad (4)$$

式中 n ——分箱的数量

$GoodRatio_i$ ——分箱 i 中正样本所占比例

$BadRatio_i$ ——分箱 i 中负样本所占比例

当 $IV < 0.02$ 时,特征对作业状态预测的能力可以忽略不计;当 $0.02 \leq IV < 0.1$ 时,特征对作业状态的预测能力较弱;当 $0.1 \leq IV < 0.3$ 时,特征信息对作业状态预测的能力属于中等水平;而当 $IV \geq 0.3$ 时,特征对作业状态预测的能力最强。

1.3 基于 BP-AdaBoost 的作业状态智能辨识算法

BP-AdaBoost 是一种集成学习方法,其中弱分类器由 BP 神经网络构成^[17-18]。该方法通过迭代训练多个 BP 神经网络,并结合 AdaBoost 算法^[19-20]对分类器进行加权集成,以提高整体模型的性能。

BP-Adaboost 算法步骤如下:

(1) 初始化样本权重。从空间轨迹点样本中随机选取 m 组训练数据,初始化测试数据的分布权重值,公式为

$$D_i(i) = \frac{1}{m} \quad (5)$$

式中 $D_i(i)$ ——第 i 个样本分布权重

根据构建的神经网络结构,初始化 BP 神经网络的权值和阈值。

(2) 训练弱预测器并计算误差。对于每轮迭代 t ,使用当前样本分布权重 $D_t(i)$ 训练 BP 神经网络,得到弱预测器 $h_t(x_i)$ 。随后,使用该预测器计算每

个样本的预测误差绝对值 $e(i)$, 公式为

$$e(i) = |h_i(x_i) - y_i| \quad (i=1, 2, \dots, m) \quad (6)$$

式中 x_i ——输入变量

y_i ——样本的真实标签

计算误差和 e_i 为

$$e_i = \sum_{i=1}^m D_i(i) \quad (e(i) > \Phi) \quad (7)$$

式中 Φ ——预测误差的判断阈值

(3) 计算弱预测器的权重系数。根据式(5)中得出的 e_i , 计算弱预测器的权重 w_i , 其计算公式为

$$w_i = \frac{1}{2} \ln \left(\frac{1 - e_i}{e_i} \right) \quad (8)$$

(4) 调整样本权重。根据权重系数调整下一轮训练样本的权重, 调整公式为

$$D_{i+1}(i) = \frac{D_i(i)}{B_i} \exp(-w_i y_i h_i(x_i)) \quad (9)$$

式中 B_i ——归一化因子

保证权重比例不变, 使样本的分布权重和为 1。

(5) 强预测器构建。利用 BP 神经网络训练 T 轮后得到 T 组弱预测器 $f(h_i(x), w_i)$, 由其加权组合得到强状态预测器 $H(x)$, 计算公式为

$$H(x) = \sum_{i=1}^T w_i f(h_i(x), w_i) \quad (10)$$

在模型训练中, 初始学习率设定为 0.001, 输入层包含多个参数, 由本文分析的关键特征组成。由于各输入参数的单位不一致且数值范围相差较大, 提前对其进行归一化处理。

输出层由 2 个参数组成, 分别表示轨迹点的状态。输出层中某节点输出为 1 时, 代表农机处于作业状态, 输出为 0 时代表农机处于非作业状态。

隐含层作为 BP 神经网络重要组成部分^[21], 其节点数 l 计算式为

$$l = \sqrt{n + m_1} + a \quad (0 \leq a \leq 10) \quad (11)$$

式中 n ——输入层节点数

m_1 ——输出层节点数

由式(11)可确定本文模型隐含层节点数为 9。每一轮迭代中, AdaBoost 算法依据前一轮分类器的错误率调整样本权重, 从而使后续分类器聚焦于难以分类的样本。通过加权投票机制, 形成强分类器以进行决策, BP-AdaBoost 能够有效减少农机作业样本过拟合现象并增强模型的泛化能力。

2 试验

2.1 大数据平台

本文采用的大数据平台由全国主要农区的管理单位统一协调, 农机作业信息采集系统由各试验测

试站具体实施。平台涵盖了试验田管理、样机作业监测以及作业过程中的多维数据采集, 实时记录农机的动态运行数据, 并具备高效的数据存储、处理与分析功能, 支持大数据量的实时处理及历史数据的深度挖掘与分析。各试验站负责数据的实时上传、处理与存储, 确保数据的准确性与时效性。

2.2 作业状态参数数据获取

农机配备的智能检测终端通过集成全球导航卫星系统获取农机的位置信息和作业速度, 并使用状态采集传感器实时监测农机作业过程中的运行数据, 包括时间、方位角等信息。最后, 采集的数据信息通过无线传输模块上传至数据服务器。

为确保数据的多样性与可靠性, 本文试验数据分别来源于河南省信阳市罗山县竹竿镇、漯河市临颍县巨陵镇、平顶山市叶县常村镇、南阳市唐河县古城乡和新野县施庵镇。数据采集时间为 2024 年 5 月 13—28 日, 采样间隔为 2 s, 最终共采集总数据量为 128 946 条。为确保数据精度和稳定性, 本文试验的采样数据总量为 128 782。

2.3 特征参数分箱计算结果

基于农机作业数据监测平台和农业大数据融合技术, 结合农机农艺要求, 针对部分实测数据进行筛选与分析。首先对采集的基础特征信息进行数据清洗与处理, 提取出平均加速度、转向率以及轨迹点密度 3 种特征信息, 并经数据预处理后, 构建了包含 128 782 条样本的完整数据集; 随后, 采用多策略分箱方法对数据区间进行划分, 划分结果如图 2 所示。

并进一步计算了相应分箱区间的 WOE 和 IV 值, 结果如表 1~4 所示。当分箱区间内正样本数为 0 时, 则认为该分箱基本不具备分析能力, 因此将其 WOE 和 IV 值分别设置为零。

对速度特征数据进行决策树分箱处理后, 结果如表 1 所示。从表 1 可知, 速度特征对作业状态的区分能力在中等速度区间 ($[3.5, 5.5)$ m/s) 表现最为显著。相比之下, 在极低或极高速度区间, 由于样本数量较少或分布差异不足, 速度特征的区分能力相对较弱。最终计算得出的 IV 值为 2.047, 表明速度特征对作业状态的预测能力较强。

对于平均加速度特征, 如表 2 所示, 虽然部分分箱的 WOE 值较高, 能够有效区分农机作业状态与非作业状态, 但 IV 值接近于零, 表明该特征的整体预测能力较弱。在预测能力计算过程中, 高 WOE 值并不一定伴随高 IV 值。在部分情况下, WOE 值较高的分箱由于样本量较少, 其对整体模型的贡献有限。因此, 平均加速度特征在作业状态辨识中的作业可以忽略。

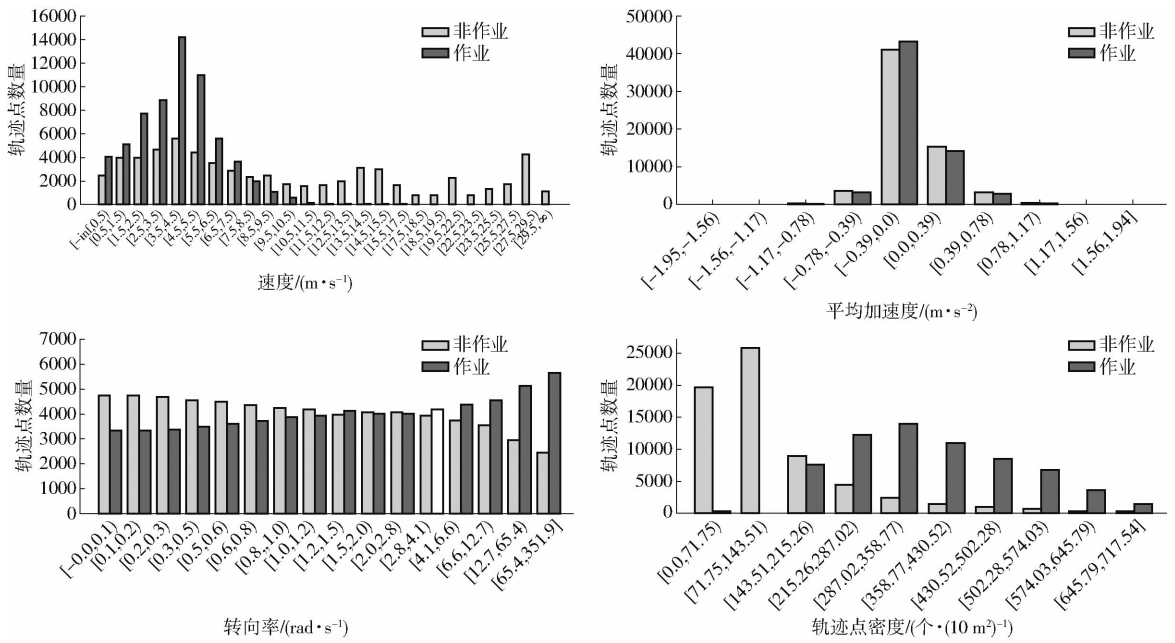


图2 特征参数分箱数据分布图

Fig.2 Feature bin distribution plots

表1 速度分箱对作业状态预测能力的评估结果

Tab.1 Results of assessment of ability of speed binning to predict operational states

分箱/($m \cdot s^{-1}$)	样本总数	正样本数	WOE 值	IV 值
$[-inf, 0.5)$	6 579	4 076	0.488	0.012
$[0.5, 1.5)$	9 123	5 126	0.249	0.004
$[1.5, 2.5)$	11 861	7 836	0.666	0.039
$[2.5, 3.5)$	13 534	8 855	0.638	0.041
$[3.5, 4.5)$	19 888	14 278	0.934	0.126
$[4.5, 5.5)$	15 382	10 954	0.906	0.092
$[5.5, 6.5)$	9 118	5 587	0.459	0.015
$[6.5, 7.5)$	6 473	3 629	0.244	0.003
$[7.5, 8.5)$	4 412	2 019	-0.170	0.001
$[8.5, 9.5)$	3 661	1 094	-0.853	0.020
$[9.5, 10.5)$	2 318	584	-1.088	0.019
$[10.5, 11.5)$	1 792	182	-2.180	0.048
$[11.5, 12.5)$	1 653	46	-3.553	0.086
$[12.5, 13.5)$	2 065	32	-4.152	0.129
$[13.5, 14.5)$	3 119	27	-4.741	0.226
$[14.5, 15.5)$	3 045	13	-5.452	0.256
$[15.5, 17.5)$	1 727	24	-4.262	0.111
$[17.5, 18.5)$	772	6	-4.849	0.057
$[18.5, 19.5)$	771	3	-5.545	0.066
$[19.5, 22.5)$	2 261	11	-5.321	0.185
$[22.5, 23.5)$	767	1	-6.641	0.079
$[23.5, 25.5)$	1 359	6	-5.418	0.113
$[25.5, 27.5)$	1 695	1	-7.435	0.195
$[27.5, 29.5)$	4 274	0	0	0
$[29.5, \infty)$	1 133	1	-7.032	0.124
总计				2.047

表2 平均加速度分箱对作业状态预测能力的评估结果

Tab.2 Results of assessment of ability of acceleration binning to predict operational states

分箱/($m \cdot s^{-2}$)	样本总数	正样本数	WOE 值	IV 值
$[-1.95, -1.56)$	4	1	-1.099	3×10^{-5}
$[-1.56, -1.17)$	88	15	-1.582	1.4×10^{-3}
$[-1.17, -0.78)$	717	303	-0.312	5×10^{-4}
$[-0.78, -0.39)$	6 876	3 240	-0.115	7×10^{-4}
$[-0.39, 0)$	84 186	43 090	0.047	1.5×10^{-3}
$[0, 0.39)$	29 752	14 408	-0.063	9×10^{-4}
$[0.39, 0.78)$	6 227	2 967	-0.094	4×10^{-4}
$[0.78, 1.17)$	859	349	-0.379	9×10^{-4}
$[1.17, 1.56)$	71	18	-1.080	6×10^{-4}
$[1.56, 1.94]$	2	0	0	0
总计				0.007

表3 转向率分箱对作业状态预测能力的评估结果

Tab.3 Results of assessment of ability of steering rate binning to predict operational states

分箱/($rad \cdot s^{-1}$)	样本总数	正样本数	WOE 值	IV 值
$[0, 0.1)$	8 049	3 326	-0.351	7.6×10^{-3}
$[0.1, 0.2)$	8 049	3 300	-0.364	8.2×10^{-3}
$[0.2, 0.3)$	8 048	3 356	-0.335	7×10^{-3}
$[0.3, 0.5)$	8 049	3 494	-0.265	4.4×10^{-3}
$[0.5, 0.6)$	8 049	3 590	-0.217	2.9×10^{-3}
$[0.6, 0.8)$	8 049	3 678	-0.173	1.9×10^{-3}
$[0.8, 1.0)$	8 049	3 852	-0.086	5×10^{-4}
$[1.0, 1.2)$	8 049	3 898	-0.063	2×10^{-4}
$[1.2, 1.5)$	8 049	4 086	0.031	6×10^{-5}
$[1.5, 2.0)$	8 049	4 011	-0.007	3×10^{-6}
$[2.0, 2.8)$	8 048	4 016	0.004	1×10^{-6}
$[2.8, 4.1)$	8 049	4 151	0.063	2.47×10^{-3}
$[4.1, 6.6)$	8 049	4 351	0.163	1.65×10^{-3}
$[6.6, 12.7)$	8 049	4 532	0.254	4×10^{-3}
$[12.7, 65.4)$	8 049	5 131	0.564	1.9×10^{-2}
$[65.4, 351.9]$	8 049	5 619	0.838	4.2×10^{-2}
总计				0.102

通过等频分箱方法对转向率进行处理并分析其特征重要性,结果如表3所示,表明大多数区间对作

业状态的区分能力较弱,仅在较高区间($[12.7, 65.4)$ rad/s 和 $[65.4, 351.9)$ rad/s)表现出一定程度的信息增益。

由表 4 可知,农机轨迹点密度对作业状态的决策能力较强。高 WOE 值和 IV 值的区间($[287.02, 358.77)$ (个 $\cdot(10\text{ m}^2)^{-1}$))表现出显著的区分能力,能够有效区分作业状态,表明其对作业状态的预测能力较强。相对而言,低 IV 值的区间($[143.51, 215.26)$ (个 $\cdot(10\text{ m}^2)^{-1}$))缺乏明显的区分作用,由于作业与非作业状态的比例差异较小,难以提供有效信息。负 WOE 值和较高 IV 值的区间($[0, 71.75)$ (个 $\cdot(10\text{ m}^2)^{-1}$))尽管非作业状态样本占比较多,但仍能有效区分农机作业状态,表明其在特征关键性分析中具有关键作用。高 WOE 值但低 IV 值的区间($[645.79, 717.54)$ (个 $\cdot(10\text{ m}^2)^{-1}$))尽管分布差异明显,但由于样本量较少,对整体模型的影响有限。

表 4 轨迹点密度分箱对作业状态预测能力的评估结果

Tab.4 Results of assessment of ability of distribution density of trajectory points binning to predict operational states

分箱/(个 $\cdot(10\text{ m}^2)^{-1}$)	样本总数	正样本数	WOE 值	IV 值
$[0, 71.75)$	25 798	2	-9.465	3.791
$[71.75, 143.51)$	19 800	195	-4.611	1.389
$[143.51, 215.26)$	16 351	7 507	-0.164	0.003
$[215.26, 287.02)$	16 640	12 221	1.017	0.123
$[287.02, 358.77)$	16 091	13 774	1.783	0.317
$[358.77, 430.52)$	12 265	10 811	2.006	0.292
$[430.52, 502.28)$	9 176	8 295	2.242	0.258
$[502.28, 574.03)$	7 356	6 745	2.401	0.229
$[574.03, 645.79)$	3 710	3 498	2.803	0.143
$[645.79, 717.54]$	1 595	1 343	1.673	0.028
总计				6.573

2.4 特征重要性分析结果

对本次试验农机的多维特征进行特征权重排序,农机轨迹点密度是区分作业状态能力最强的特征,其 IV 值为 6.573,显著高于其他特征。轨迹点密度在农机作业样本与非作业样本中的分布差异尤为突出,特别是在区间 $[0, 71.75)$ (个 $\cdot(10\text{ m}^2)^{-1}$) 和 $[71.75, 143.51)$ (个 $\cdot(10\text{ m}^2)^{-1}$) 内,IV 值较高,即预测能力强能够区分作业和非作业样本。上述区间数据表明,轨迹点密度变化能够直观反映农机在作业过程中区域覆盖的强弱程度。因此,该特征可以作为区分作业与非作业状态的最关键特征。

速度特征参数的 IV 值为 2.047,仅次于轨迹点密度,表明其具有较强的作业状态区分能力。速度变化与作业状态密切相关,在中高速范围内($[3.5, 4.5)$ m/s 和 $[4.5, 5.5)$ m/s)的 WOE 值较高,表明

速度在中高速区间内对作业状态的区分作用较为显著。该特征表现与农机在作业时保持稳定的中等速度,而非作业状态下速度波动较大或静止的特点相符。

转向率特征整体区分能力较弱,但在特定高值区间($[65.4, 351.9]$ rad/s)内 WOE 值达到 0.838,表明转向率在特定场景下能够提供额外的信息。在高转向率的往复式作业过程中,转向率特征能够成为有效的关键特征。尽管其单独信息增益较小,但与其他特征联合使用可能提高作业状态的预测精度。相比之下,平均加速度的 IV 值极低,仅为 0.007,表明其对作业状态的区分能力几乎可以忽略不计。

综上所述,轨迹点密度和速度是区分作业状态的关键特征,其中轨迹点密度表现出显著的区分能力,而速度则在特定区间内发挥重要作用。转向率在特定场景下提供了辅助信息,但其单独的区分能力较弱。平均加速度的贡献几乎可以忽略不计,对作业状态的区分作用有限。因此,选择轨迹点密度、速度和转向率 3 个特征作为作业状态辨识模型的输入。

2.5 作业状态辨识结果

经特征参数重要性分析后,选取农机空间轨迹点密度、速度和转向率作为特征参数,并对作业轨迹点进行标记。将 80% 的数据作为训练样本,利用 BP-AdaBoost 算法训练模型,最终对所有轨迹点进行识别。图 3 为农机工作 16 h 的轨迹点作业状态辨识效果图,红色轨迹点为作业状态,白色点为非作业状态,黄色圈出部分为辨识错误的轨迹点。

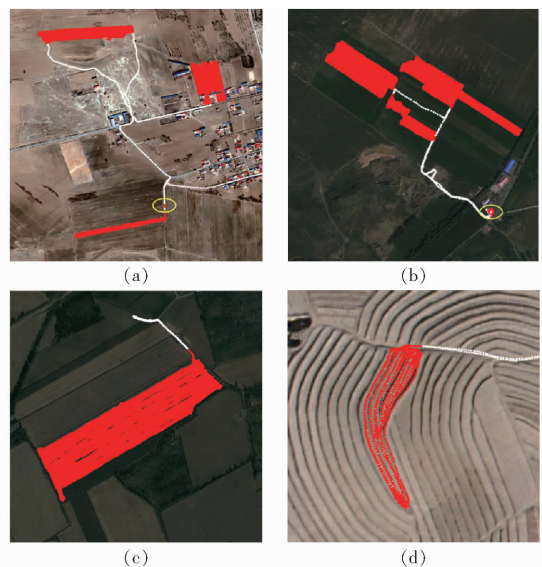


图 3 农机作业状态辨识效果图

Fig.3 Agricultural machinery operation state identification effect diagrams

由图3可得,本文提出的农机轨迹点作业状态辨识方法能够准确识别绝大多数轨迹点的作业状态,仅有极少轨迹点出现误判。在图3a中,误辨识的位置主要出现在农机即将拐入作业区域的阶段。此时,农机速度相较于道路行驶时明显降低,该变化干扰了算法对轨迹点作业状态的判断。在图3b右下角黄色标记区域,由于农机在该区域多次往返行驶并频繁转向,使得该区域的轨迹点密度和转向率特征与作业状态下的特征相似,导致轨迹点被误判为作业状态。而在图3c、3d中,均未出现错误识别轨迹点。

根据轨迹点的地理位置和行驶轨迹,对轨迹点状态进行标记。为了评估预测的准确程度,采用预测数据与真实数据的一致性作为评估标准。测试集总轨迹点数 D 为25 756,其中农机处于作业状态轨迹点数量和非作业状态轨迹点数量均为12 878。农机作业状态识别结果如表5所示,其中农机处于作业状态轨迹点记为P,非作业状态记为N。

表5 作业状态识别结果

Tab.5 Operational state identification results

作业状态情况	正确识别个数	错误识别个数
作业 P	12 630	248
非作业 N	12 431	447

轨迹点状态识别正确的数量为作业状态样本识别为作业状态样本个数与非作业状态样本识别为非作业状态个数之和。由表5可知,采用农机空间轨迹点密度、速度和转向率作为模型输入参数,结合神

经网络与AdaBoost算法的融合方法,正确辨识的数量为25 061,准确率达到97.3%。由此可见,基于BP-AdaBoost神经网络模型结果与实际作业状态基本一致,错误率仅为2.7%,可准确有效辨识农机的作业状态。

3 结论

(1)为实现数据管理平台对农机作业状态的精准监测,基于已获取的农机运行轨迹点特征数据,提出了一种融合多维特征的作业状态辨识方法。深入分析了农机作业轨迹的特征及轨迹点的有效信息,选取轨迹点速度、平均加速度、转向率及轨迹点密度作为典型特征,并研究了其分布规律。

(2)为验证这4种典型特征对农机作业状态的预测能力,基于WOE和IV方法,结合农机作业数据监测平台采集的数据,开展了特征相关性分析试验。结果表明,4种特征的IV值分别为2.047、0.007、0.102和6.573,其中轨迹点密度、速度和转向率对农机作业状态具有较强的辨识能力,而平均加速度的辨识能力较弱,可忽略不计。

(3)在此基础上,研究了BP-AdaBoost神经网络算法,并以轨迹点密度、速度和转向率作为输入变量,构建了农机作业状态辨识模型。通过农机作业数据平台对作业状态进行测试和验证,结果显示,该模型的识别准确率高达97.3%,表明基于轨迹点密度、速度和转向率3种关键特征,采用神经网络模型进行农机作业状态辨识具有较高的可行性和有效性。

参 考 文 献

- [1] 陈学庚,温浩军,张伟荣,等. 农业机械与信息技术融合发展现状与方向[J]. 智慧农业(中英文),2020,2(4):1-16. CHEN Xuegeng, WEN Haojun, ZHANG Weirong, et al. Advances and progress of agricultural machinery and sensing technology fusion[J]. Smart Agriculture, 2020, 2(4): 1-16. (in Chinese)
- [2] SHARMA K, SHIVANDU S K. Integrating artificial intelligence and Internet of Things (IoT) for enhanced crop monitoring and management in precision agriculture[J]. Sensors International, 2024,5: 100292.
- [3] 张漫,李宇寒,李世超,等. 农业机械导航技术研究进展[J]. 农业机械学报,2020,51(4):1-18. ZHANG Man, LI Yuhuan, LI Shichao, et al. Research progress of agricultural machinery navigation technology[J]. Transactions of the Chinese Society for Agricultural Machinery,2020,51(4):1-18. (in Chinese)
- [4] 解开婷,张兆国,王法安,等. 基于全质心WLS-HDS-TWR算法的无人驾驶农业机械协同定位算法研究[J]. 农业机械学报,2024,55(7):27-36,110. XIE Kaiting, ZHANG Zhaoguo, WANG Fa'an, et al. Cooperative localization algorithm for full center mass of WLS-HDS-TWR driverless agricultural machines[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(7):27-36,110. (in Chinese)
- [5] 段贤强,陶建峰,覃程锦,等. 变速条件下农业机械路径跟踪稳定控制方法[J]. 农业机械学报,2019,50(9):18-24,32. DUAN Xianqiang, TAO Jianfeng, QIN Chengjin, et al. Agricultural machinery path tracking stability control method under variable speed conditions[J]. Transactions of the Chinese Society for Agricultural Machinery,2019,50(9):18-24,32. (in Chinese)
- [6] 马莉,刘国威,丁世宏,等. 考虑输入饱和的农用拖拉机路径跟踪有限时间控制[J]. 控制与决策,2024,39(6):1943-1950. MA Li, LIU Guowei, DING Shihong, et al. Finite-time path tracking control of agricultural tractors subject to input saturation

- [J]. *Control and Decision*, 2024, 39(6): 1943 – 1950. (in Chinese)
- [7] 刘阳春, 苑严伟, 张俊宁, 等. 深松作业远程管理系统设计与试验[J]. *农业机械学报*, 2016, 47(增刊 1): 43 – 48.
LIU Yangchun, YUAN Yanwei, ZHANG Junning, et al. Design and experiment of remote management system for subsoiler[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2016, 47(Supp. 1): 43 – 48. (in Chinese)
- [8] 孟志军, 尹彦鑫, 罗长海, 等. 农机深松作业远程监测系统设计与实现[J]. *农业工程技术*, 2018, 38(18): 34 – 37.
MENG Zhijun, YIN Yanxin, LUO Changhai, et al. Design and realization of remote monitoring system for deep pine operation of agricultural machinery[J]. *Agricultural Engineering Technology*, 2018, 38(18): 34 – 37. (in Chinese)
- [9] 王培, 孟志军, 安晓飞, 等. 拖拉机功率与深松作业效率关系研究[J]. *农业机械学报*, 2019, 50(增刊 1): 87 – 90.
WANG Pei, MENG Zhijun, AN Xiaofei, et al. Relationship between agricultural machinery power and agricultural machinery subsoiling operation[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(Supp. 1): 87 – 90. (in Chinese)
- [10] DANDA R R. Innovations in agricultural machinery: assessing the impact of advanced technologies on farm efficiency[J]. *Journal of Artificial Intelligence and Big Data*, 2022, 2(1): 64 – 83.
- [11] 李翔, 潘瑜春, 赵春江, 等. 基于多年产量数据的精准农业管理分区提取与尺度效应评价[J]. *中国农业科学*, 2005, 38(9): 1825 – 1833.
LI Xiang, PAN Yuchun, ZHAO Chunjiang, et al. Delineation and scale effect of precision agriculture management zones using yield monitor data over four years[J]. *Scientia Agricultura Sinica*, 2005, 38(9): 1825 – 1833. (in Chinese)
- [12] ZHANG S, LIU H, CAO X, et al. Agricultural machinery movement trajectory recognition method based on two-stage joint clustering[J]. *Agriculture*, 2024, 14(12): 2294.
- [13] 陈凯, 解印山, 李彦明, 等. 多约束情形下的农机全覆盖路径规划方法[J]. *农业机械学报*, 2022, 53(5): 17 – 26, 43.
CHEN Kai, XIE Yinshan, LI Yanming, et al. Full coverage path planning method of agricultural machinery under multiple constraints[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(5): 17 – 26, 43. (in Chinese)
- [14] 桑浩博. 基于大数据分析的农业机械智能化路径规划研究[J]. *南方农机*, 2024, 55(增刊 1): 50 – 53, 72.
SANG Haobo. Research on intelligent path planning of agricultural machinery based on big data analysis[J]. *South Agricultural Machinery*, 2024, 55(Supp. 1): 50 – 53, 72. (in Chinese)
- [15] 王培, 孟志军, 尹彦鑫, 等. 基于农机空间运行轨迹的作业状态自动识别试验[J]. *农业工程学报*, 2015, 31(3): 56 – 61.
WANG Pei, MENG Zhijun, YIN Yanxin, et al. Automatic recognition algorithm of field operation status based on spatial track of agricultural machinery and corresponding experiment[J]. *Transactions of the CSAE*, 2015, 31(3): 56 – 61. (in Chinese)
- [16] 庄晓明. 大数据环境下提升风控效率的风险标签筛选方案优化设计[J]. *电子元器件与信息技术*, 2024, 8(5): 101 – 103.
ZHANG Xiaoming. Optimized design of risk label screening scheme to enhance risk control efficiency in big data environment[J]. *Electronic Component and Information Technology*, 2024, 8(5): 101 – 103. (in Chinese)
- [17] LI J, CHENG J, SHI J, et al. Brief introduction of back propagation (BP) neural network algorithm and its improvement[C] // *Advances in Computer Science and Information Engineering*, 2012: 553 – 558.
- [18] LI N, CHENG X, ZHANG S, et al. Recognizing human actions by BP AdaBoost algorithm under a hierarchical recognition framework[C] // *2013 IEEE International Conference on Acoustics Speech and Signal Processing*, 2013: 3407 – 3411.
- [19] 李亚硕, 赵博, 王长伟, 等. 基于 DBSCAN 和 BP_Adaboost 的农机作业地块划分方法[J]. *农业机械学报*, 2023, 54(1): 37 – 44.
LI Yashuo, ZHAO Bo, WANG Changwei, et al. Land division method for agricultural machinery operation based on DBSCAN and BP_Adaboost[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(1): 37 – 44. (in Chinese)
- [20] CHOUDHURY A, MONDAL A, SARKAR S. Searches for the BSM scenarios at the LHC using decision tree-based machine learning algorithms: a comparative study and review of random forest, AdaBoost, XGBoost and LightGBM frameworks[J]. *The European Physical Journal Special Topics*, 2024, 233: 2425 – 2463.
- [21] 温昌凯, 谢斌, 李若晨, 等. 拖拉机作业载荷数据平台设计与旋耕作业质量预测[J]. *农业机械学报*, 2020, 51(8): 372 – 381.
WEN Changkai, XIE Bin, LI Ruochen, et al. Construction of tractor working load data platform and prediction of rotary tillage quality[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(8): 372 – 381. (in Chinese)