

基于 DBSCAN 和 BP_Adaboost 的农机作业地块划分方法

李亚硕^{1,2} 赵 博^{1,2} 王长伟^{1,2} 徐名汉^{1,2} 伟利国^{1,2} 庞在溪^{1,2}

(1. 中国农业机械化科学研究院集团有限公司, 北京 100083;
2. 土壤植物机器系统技术国家重点实验室, 北京 100083)

摘要: 农业机械(农机)在多个地块作业,费用和效率有时需按地块统计,现有的农机监控系统仅能记录农机定位信息和作业状态信息,难以实现地块的自动精准划分。本文通过研究轨迹点属性特征,分析作业地块数量不确定性和轨迹点分布规律,采用基于密度聚类方法(Density-based spatial clustering of applications with noise, DBSCAN)和分类器集成算法(BP_Adaboost)结合的方法划分地块。根据 DBSCAN 算法对农机轨迹点多数有效、识别错误集中的特点,结合 BP_Adaboost 算法挖掘多维度信息关联、容错能力强、分类效果好等优势,先利用 DBSCAN 得到初步的轨迹点状态类别,再利用 BP_Adaboost 算法建立训练模型对农机轨迹点状态精准识别,根据时间序列和类别标记划分地块。本文方法既解决了只依靠阈值和经纬度信息聚类不准确的问题,也减少了大量样本标记工作。利用该方法轨迹点状态识别准确率达 96.75%,地块划分准确率为 97.74%。

关键词: 农业机械; 作业轨迹; 地块划分; 密度聚类算法; 分类器集成算法

中图分类号: TP399 文献标识码: A 文章编号: 1000-1298(2023)01-0037-08 OSID: 

Land Division Method for Agricultural Machinery Operation
Based on DBSCAN and BP_Adaboost

LI Yashuo^{1,2} ZHAO Bo^{1,2} WANG Changwei^{1,2} XU Minghan^{1,2} WEI Ligu^{1,2} PANG Zaixi^{1,2}

(1. Chinese Academy of Agricultural Mechanization Sciences Group Co., Ltd., Beijing 100083, China
2. State Key Laboratory of Soil-Plant-Machinery System Technology, Beijing 100083, China)

Abstract: Agricultural machinery operates in multiple plots, and the cost and efficiency sometimes need to be counted according to the plots. The existing agricultural machinery monitoring system can only record the positioning information and operation status information of agricultural machinery, which is difficult to realize the automatic and accurate division of plots. By studying the attribute characteristics of track points, the uncertainty of the number of work plots and the distribution law of track points were analyzed, and the combination of density clustering method (density-based spatial clustering of applications with noise, DBSCAN) and weak classifier integration algorithm (BP_Adaboost) were used to divide the plots. According to the characteristics that DBSCAN method is effective for most agricultural machinery trajectory points and the recognition error is concentrated, combined with BP_Adaboost method to mine multi-dimensional information association, strong fault tolerance, good classification effect and other advantages. Firstly, DBSCAN was used to obtain the preliminary track point state category, and then the method of BP_Adaboost was used to establish a training model to accurately identify the track point state of agricultural machinery, and divide the land mass according to time series and category markers. The method not only solved the problem of inaccurate clustering only relying on threshold and longitude and latitude information, but also reduced a lot of sample labeling work. Using this method, the accuracy of track point state recognition was 96.75%, and the accuracy of plot division was 97.74%.

Key words: agricultural machinery; operation track; parcel division; density clustering algorithm; classifier ensemble algorithm

0 引言

随着农业机械化作业的普遍推广、农业机械化水平的提升,物联网技术在农机生产管理中的作用日益突出。目前,通过机载物联网设备和无线通信技术,监管人员可以实时获取农机作业位置、作业状态和作业效率,并通过监管平台进行任务分配和补贴结算等,满足了农机监管部门和农机合作社管理人员对所属农机的实时监管和作业规划需求^[1-4]。农机管理人员按时间或工作量给机手发放工资,按地块面积向农户收取费用或向相关部门申请补贴,因此地块的精准划分是工资发放、费用收取、补贴申请、面积计算、效率评估等的基础。深松作业可以根据深松机入土深度判断是否在田间作业,但其它如收获、植保、打捆等作业难以根据机载传感器判断作业状态。无法强制农机驾驶员只在田间才打开机载设备,其他时间关闭设备。因此需要根据得到的轨迹点信息自动判断农机轨迹点状态和划分地块。

使用机载定位设备可以记录作业轨迹信息,实时发送后端处理,得到行驶轨迹和作业分布。根据上传的农机行驶轨迹,可在监管平台的地图中人工划分地块。但一台农机每天作业多个地块,每天多台农机同时作业,一个作业季的地块数量非常多,人工划分耗时耗力,并且由于地图精度问题,人工标记边界容易出现误差。

地块划分的前提是农机轨迹点状态的确定。目前,轨迹点数据处理多采用聚类算法,根据轨迹点聚簇判断状态和划分地块。田间作业轨迹点密集,道路行驶轨迹点稀疏,利用这一特点通过 DBSCAN 聚类的方法区分田间作业和道路行驶轨迹^[5-6]。但聚类方法的邻域半径和邻域点数不易确定,不同农机、不同地形作业轨迹相差很大。有研究人员利用时空信息对轨迹进行密度聚类^[7-9],由于轨迹点具有时序性,将时间作为聚类信息对准确率的提升作用不大。利用空间索引和网格密度聚类^[10-11],有效提升了运行速度,但网格尺寸和密度阈值参数敏感。尽管聚类方法对农机田间作业轨迹效果较好,但对道路转移行驶轨迹容易误识别,因此单纯依靠聚类方法,很难提升农机轨迹点状态识别率。

根据轨迹点单个特征设定阈值,容易遗漏深层的内部相互关系。依靠神经网络等深度学习方法进行训练,在多维特征组成的高维空间寻找有效的内部关联性,进而有效识别农机轨迹点状态,再根据时序关系划分地块。但训练样本的标记耗时太大,并且有新农机作业轨迹识别需要重新标记训练,推广性不强。

本文结合农机空间运行轨迹的特点,选择农机轨迹点经度、纬度、速度、近邻点相互关系等作为特征,先用 DBSCAN 算法进行聚类,得到初步的聚类结果和类别标记;利用 BP_Adaboost 算法对轨迹点进行训练和识别;将田间与道路交界处附近轨迹点标记为道路轨迹并添加到训练样本中重复训练,直至训练样本无新增。以期解决海量数据标记困难、不同地块轨迹差异大、田间与道路交界处轨迹点易误识别等问题。

1 农机作业地块划分

通过安装在农机上的北斗定位接收装置自动记录农机运行全过程轨迹,获取农机经纬度位置信息、速度信息等。农机在行进和停车时,机载定位设备都会发送位置信息,因此,农机行驶轨迹点分道路行驶轨迹点、田间作业轨迹点和停车点。农机从车库行驶至目的地块以及在两地块间转移时,一般沿道路单向行驶,速度较快,轨迹点稀疏;在田间作业时,在地块内往返行驶,速度偏慢,轨迹点密集;当农机因故停车时,轨迹点以农机为中心,在小范围内呈散点簇分布。划分地块的依据是区分田间作业轨迹点和道路行驶轨迹点,因此需先处理停车点和其他干扰轨迹点。再根据农机轨迹点的特性,对不同轨迹点信息进行分析试验,选取有用特征。

1.1 数据预处理

由于定位精度原因,农机在停车时,轨迹点不是集中在一个点,而是散落在农机附近小范围内,称为农机停车散点,速度为 0 或者很小值。当定位偏差过大或者设备故障导致多个轨迹点未发送时,连续接收到的两点间距过大,称为漂移点。停车散点和漂移点如图 1 所示。



图 1 停车散点和漂移点

Fig. 1 Parking scatter and drift point

有效农机行驶轨迹点是道路行驶轨迹点和田间作业轨迹点,农机停车散点和漂移点会影响轨迹状

态识别,在数据处理前应去除。为保持图像数据完整性,本文在数据处理时不计算停车散点和漂移点,在地图中,将这两类点标记为道路行驶点。为了防止将田间作业过程中临时停车轨迹点误删除,停车散点判断依据设置为:连续很多点的速度小于某一阈值。将漂移点的上一轨迹点作为前段轨迹终点,下一轨迹点作为下段轨迹起点。

由于采集器故障、信号差、断电等原因,会造成连续两轨迹点之间间隔很远。当某一轨迹点与前一轨迹点距离较大,而与下一轨迹点间隔正常时,认为两轨迹点之间存在轨迹点丢失。将上一轨迹点作为前段轨迹终点,下一轨迹点作为下段轨迹起点。

1.2 特征选取

农机行驶轨迹点包含时间、经纬度、速度等信息,农机田间作业轨迹和道路行驶轨迹交叉,依据轨迹点属性信息,可利用阈值设定、聚类、分类等方法区分轨迹点状态,找到有效的属性信息是取得理想效果的前提。图 2 为轨迹点近邻点数差异图,其中近邻区域半径为当日轨迹点平均距离的 3 倍。

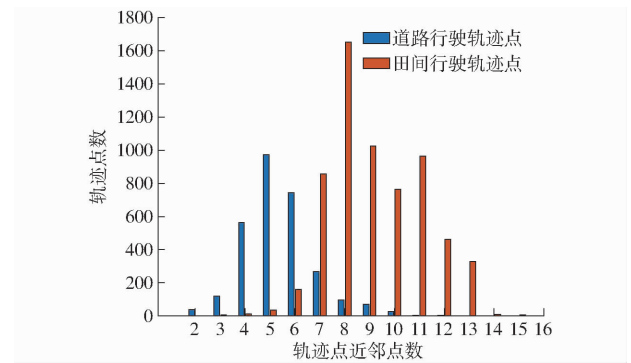


图 2 轨迹点近邻点数差异柱状图
Fig. 2 Histogram of difference in number of adjacent points of track points

以轨迹点为中心、以 r 为半径的圆范围内近邻轨迹点的个数,田间作业轨迹点多于道路行驶轨迹点。研究发现,当 r 为每天农机作业轨迹点平均距离 3 倍时,田间作业轨迹点和道路行驶轨迹点在 r 范围内的近邻点数量区分性最好。当 r 过小时,田

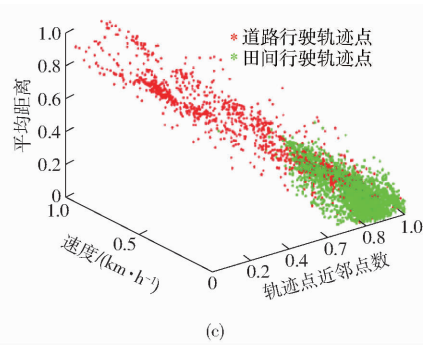
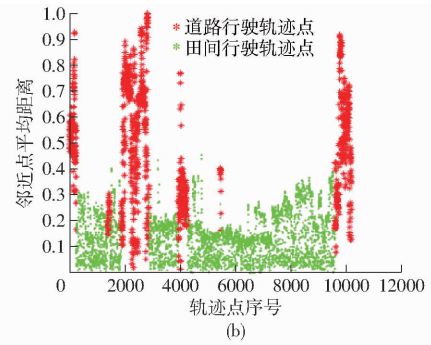
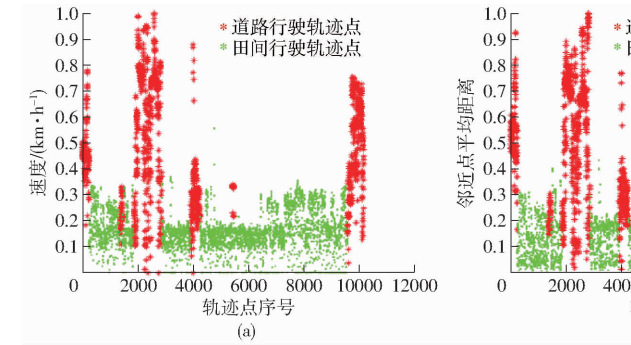


图 3 不同特征下轨迹点分布
Fig. 3 Distribution of trajectory points under different characteristics

间作业轨迹点近邻点数量急剧下降,当 r 过大时,道路行驶轨迹点近邻点数量增加,都会造成两类轨迹点近邻点数量重合部分增加。如图 2 所示,田间作业轨迹点周围轨迹点数量集中在 7~9 个之间,道路行驶轨迹点周围轨迹点数量集中在 4~6 个之间,有较好的区分性。因此该特征可以用于判断轨迹点状态。以速度和近邻点平均距离等为特征,不同状态轨迹点分布如图 3 所示。

由图 3 可以看出,田间作业轨迹点速度慢且与附近点距离小,虽然有一定的区分性,但是重叠交叉部分也很多,依靠某一、两个特征难以达到预期效果。在一定半径范围内,田间作业轨迹点附近轨迹点数要多于道路行驶轨迹点附近轨迹点数。当选用速度、近邻点数和近邻点平均距离作为特征时,轨迹点分布如图 3c 所示,虽然仍有部分重叠,但区分性较单个特征要好。在多特征组成的高维空间内,田间作业轨迹点和道路行驶轨迹点更具区分性。但单纯依靠阈值作为判断依据,普适性不强,且阈值不易确定。应在有效特征选取的基础上,自学习内部关联性,进而判断轨迹点状态。

本研究选取轨迹点经纬度、速度、近邻点平均距离、近邻点数、轨迹点与近邻点距离作为特征,利用 BP_Adaboost 训练模型识别轨迹点状态。样本标记工作量过大,先利用 DBSCAN 聚类方法得到初步类别标记,再进行训练。

1.3 DBSCAN 聚类

密度聚类算法^[12-14]假设聚类结构能通过样本分布的紧密程度确定,从样本密度的角度来考察样本之间的可连接性,并基于可连接样本不断扩展聚类簇以获得最终的聚类结果。农机田间作业轨迹点紧密,道路行驶轨迹点稀疏,利用密度聚类方法不仅可以区分田间作业轨迹和道路行驶轨迹,还可以将不同地块的轨迹点聚为不同类。DBSCAN 基于“邻域”参数(ϵ, N_{minPts})来判断轨迹点数据的紧密程度, ϵ 为以每个轨迹点为中心的半径范围, N_{minPts} 是轨迹点要成为一个簇中心,周围轨迹点距离小于 ϵ 的最

少个数。DBSCAN 可以去除噪声点,并且不必提前规定聚类个数,聚类步骤如下:

- (1)从数据集中任意选取一个轨迹点 p 。
- (2)如果对于参数 ε 和 $N_{\min Pts}$,选一个轨迹点 p ,找到所有密度可达的核心对象,生成聚类簇。
- (3)如果选取的轨迹点 p 是边缘点,选取另一个轨迹点。
- (4)重复步骤(2)、(3),直到所有轨迹点被处理。

农机道路行驶轨迹一般为单向,田间作业为往返行驶,具有明显的区分性,因此 DBSCAN 方法对这种轨迹状态区分有不错的效果。但部分道路区段由于农机折返行驶、转向、减速等原因导致轨迹点密度变大,使该处轨迹被误识别为田间作业轨迹。不同农机行驶速度和轨迹不同,田间作业轨迹间隔也不同,DBSCAN 算法对于输入参数 ε 和 $N_{\min Pts}$ 是敏感的,使用固定的 ε 和 $N_{\min Pts}$ 不能很好解决所有农机轨迹状态识别问题。

1.4 类别不平衡处理

利用分类器训练模型,再对轨迹点进行识别,可以避免聚类算法中 ε 和 $N_{\min Pts}$ 不易确定的问题。但在农机行驶轨迹中,田间作业轨迹点数量要远多于道路行驶轨迹点数量,特别情况下比例可达 10:1。用同样数量田间作业轨迹点数据进行训练,样本数量不足;田间作业轨迹点选用过多,会造成类别不平衡,都影响识别准确率,因此需要增加道路行驶轨迹点数据量。SMOTE 算法^[15-17]是在少数类样本和其近邻的少数类样本之间进行随机线性插值生成新样本,来达到平衡数据集的目的。算法原理如下:

- (1)对于每一个少数类样本 $X_i(i=1,2,\cdots,n)$,根据欧氏距离计算出最近邻的 k 个少数类样本($Y_1, Y_2, \cdots, Y_k$)。
- (2)从 k 个最近邻样本中随机选择若干个样本,在每个选出的样本 Y_i 和原样本 X_i 之间进行随机线性插值,生成新样本 S_{new} 。插值方法计算式为
$$S_{\text{new}} = X_i + \text{rand}(0,1)(Y_i - X_i) \tag{1}$$
式中 $\text{rand}(0,1)$ ——(0,1)区间内的随机数
- (3)将新生成的样本加入原数据集中。SMOTE 算法是对随机过采样的一种改进方法,它简单有效,并且避免了过拟合的问题。根据选取的特征,对新的数据样本利用 BP_Adaboost 训练模型。

1.5 BP_Adaboost 算法

Adaboost 是一种迭代算法,其核心思想是针对同一个训练集训练不同的弱分类器,然后把这些弱分类器集合起来构成一个更强的分类器,而在 BP_

Adaboost 中弱分类器即 BP 神经网络^[18-19]。

BP 神经网络由输入层、隐含层和输出层构成^[20],是一种从输入到输出的映射,通过用已知的模式对网络进行训练,就会得到一个能反映输入到输出之间精确映射关系的网络,特点是信号前向传播,误差反向传播。

建模过程中,初始学习率设置为 0.000 1,既可以保证快速顺利找到损失函数的最小值,又不致于产生过大反复振荡。

其中,输入层为轨迹点特征,包括轨迹点经纬度、速度、半径 r 内近邻轨迹点个数、与近邻点平均距离和到近邻点的距离。由于各输入变量的单位不一致且数值范围相差较大,提前依据最大最小原则进行归一化。

输出层包含 2 个神经元,分别对应轨迹点状态。输出层中某神经元输出为 1 时,代表农机在田间作业,输出为 -1 时代表农机在道路行驶。

确定隐含层神经元个数^[21]计算式为

$$h = \sqrt{n_1 + n_2} + a \quad (0 \leq a \leq 10) \tag{2}$$

- 式中 h ——隐含层神经元个数
- n_1 ——输入层神经元个数
- n_2 ——输出层神经元个数
- a ——可变系数

由式(2)可确定隐含层神经元个数为 8。Adaboost 算法本身是通过改变数据分布来实现的,它根据每次训练集中每个样本的分类是否正确,以及上次的总体分类的准确率,来确定每个样本的权值。将修改过权值的新数据集送到下层分类器进行训练,最后将每次得到的分类器融合在一起,作为最后的决策分类器。

开始时,每个样本对应的权重是相同的,即其中 n 为样本个数,在此样本下训练出第 1 个弱分类器,对于分类错误的样本,加大其对应的权值,而对于分类正确的样本,降低其权重。从而得到一个新的样本分布,再次对样本进行训练,得到第 2 个弱分类器,以此类推, T 次循环得到 T 个弱分类器。最后将弱分类器联合起来,使用加权的投票机制代替平均投票机制。让分类效果好的弱分类器具有较大的权重,而分类效果差的分类器具有较小的权重。

算法步骤如下:

- (1)从样本空间中随机选择 m 组训练数据,初始化测试数据的分布权值

$$D(i) = \frac{1}{m} \tag{3}$$

根据确定好的神经网络结构,初始化 BP 神经网络权值和阈值。

(2) 用训练数据训练 BP 神经网络并且预测训练数据输出, 得到预测序列 $g(t)$ 的预测误差和 e_i, e_i 的计算公式为

$$e_i = \sum_{i=1}^m D_i(i) \quad (g(t) \neq y) \quad (4)$$

式中 $g(t)$ ——预测分类结果

y ——期望分类结果

(3) 根据步骤(2)中得到的 e_i , 计算序列的权重 α_i , 权重计算公式为

$$\alpha_i = \frac{1}{2} \ln \left(\frac{1 - e_i}{e_i} \right) \quad (5)$$

(4) 调整下一轮训练样本的权重, 调整公式为

$$D_{i+1}(i) = \frac{D_i(i)}{\beta_i} \exp(-\alpha_i y_i g_i(x_i)) \quad (i = 1, 2, \dots, m) \quad (6)$$

式中 β_i ——归一化因子

保证在权重比例不变的情况下使分布权值和为 1。

(5) 用 BP 神经网络训练 T 轮后得到 T 组弱分类函数 $f(g_i, \alpha_i)$, 由 T 组弱分类函数 $f(g_i, \alpha_i)$ 组合得到了强分类函数 $h(x)$ 。

选取农机运行轨迹点经纬度、速度、近邻点平均距离等作为特征, 对作业轨迹点进行标记, 选取 80% 作为训练样本, 利用 BP_Adaboost 算法得到训练模型, 对所有轨迹点进行识别。BP_Adaboost 算法对轨迹点识别效果要好于只依靠经纬度信息聚类的 DBSCAN 算法, 但农机多, 工作时间长, 产生的轨迹点多, 对轨迹点标记耗时耗力, 容易出错。不同农机不同作业地块行驶轨迹不同, 对新的农机轨迹识别需要重新标记, 推广适用性不强。

1.6 DBSCAN + BP_Adaboost 算法

研究大量农机作业轨迹数据发现, 由于农机在道路行驶过程中折返行驶、转向、减速等原因导致轨迹点密度变大, 使该处轨迹被误识别为田间作业轨迹, 另外部分田间轨迹与道路轨迹交界处也容易将道路行驶轨迹点识别为田间作业轨迹点。总体而言, 聚类算法的误识别集中在将道路轨迹误识别为田间轨迹。DBSCAN 方法对大多数轨迹点分类准确, 利用聚类得到的类别信息, 通过 BP_Adaboost 集成弱分类器训练, 并将类别变换处附近轨迹点权重设置一个小的置信值, 得到分类器模型后对轨迹点状态进行识别。再对得到的轨迹点进行处理后重新训练, 直至识别结果不发生变化。

1.7 地块划分

将识别后带类别标记的轨迹点按时序遍历, 当轨迹点类别标记为田间作业轨迹时, 放入地块序列, 查找后续轨迹点, 若类别标记不发生变化, 放入同一

个序列; 若类别标记发生变化, 查找到下一个地块标记轨迹点, 放入新的地块序列。当下一地块轨迹点距离超过 5 倍平均距离时, 说明遇到轨迹点丢失, 将下一轨迹点放入新地块序列。

遍历完所有轨迹点后, 得到多个地块序列, 每个序列内的轨迹点属于同一个地块。为防止同一地块因出现断点或轨迹点状态识别错误导致被误分成多个地块, 计算各地块轨迹点与其他地块轨迹点距离最小值, 若小于 3 倍平均距离, 将 2 地块合并为同一地块。直至地块数量无变化。

具体步骤如下:

(1) 按时序遍历所有带状态标记的轨迹点。

(2) 找到第 1 个地块标记轨迹点, 放入地块序列中。

(3) 将其后所有地块标记轨迹点放入同一序列, 直至查找到道路标记轨迹点。

(4) 继续遍历, 查找到下一个地块标记轨迹点, 计算该点到上一地块序列轨迹点最小距离是否小于 3 倍平均距离, 若小于, 则将该点放入上一地块序列; 否则, 放入新地块序列。

(5) 重复步骤(3)、(4), 直至遍历完所有轨迹点。

(6) 计算某一地块序列内轨迹点与其他地块序列轨迹点距离最小值是否小于 3 倍平均距离, 若小于, 按时序合并 2 个地块序列。

(7) 重复步骤(6), 直至地块序列数量不发生变化。

2 实验

2.1 数据获取

实验采用某县 50 台农机在 2020 年 10 月 15—26 日期间所有作业数据作为数据集, 作业地块数量为 5 304。在农机上安装采集设备, 每 5 s 采集一次作业点数据, 总轨迹点数为 5 196 782, 通过无线通信发送至后端服务器。每个轨迹点包含农机作业点经纬度、速度等信息。实验采用经纬度、速度、3 倍平均距离内近邻点数量、轨迹点周围 20 个点距离作为特征, 利用 DBSCAN 对轨迹点聚类, 将得到的特征及类标记用 BP_Adaboost 算法训练模型, 识别农机运行状态。

2.2 轨迹状态识别及地块划分

将获取到的轨迹点数据进行如下步骤处理:

(1) 将采集的轨迹点做预处理, 并提取特征。

(2) 使用 DBSCAN 算法对轨迹点聚类, 得到道路轨迹和田间轨迹。

(3) 将前后标记发生变化的轨迹点及前后 10

个轨迹点权重减半。

(4)使用 smote 方法增加道路行驶轨迹点样本数量。

(5)选取 80% 轨迹点作为训练样本,全部轨迹点为测试样本。

(6)选取轨迹点经纬度、速度、3 倍平均距离内近邻点数量、轨迹点周围 n 个点距离作为特征,利用 BP_Adaboost 训练数据集,并对测试样本进行轨迹状态识别。

(7)将识别的每天轨迹点中前后标记发生变化的轨迹点及前后 10 个轨迹点标记为道路轨迹点,并放入训练样本中。

(8)重复步骤(6)、(7),直至训练样本数量不发生变化。

(9)根据时序关系和轨迹点类别划分地块。

对数据采用本文方法和 DBSCAN 方法分别做实验,将轨迹点加载到地图中,根据地图中轨迹点位置和行驶轨迹,人工对轨迹点状态进行标记。总轨迹点数为 5 196 782,其中田间作业轨迹点数量为 3 162 468,道路行驶轨迹点数量为 2 034 314。轨迹点状态识别结果如表 1 所示,其中田间作业轨迹点记为 T,道路行驶轨迹点记为 D。

表 1 轨迹点状态各方法识别结果

Tab.1 Recognition results of each method of track point state

方法	轨迹点状态	识别为 T 个数	识别为 D 个数
DBSCAN 聚类算法	T	2 866 948	295 520
	D	50 066	1 984 248
DBSCAN + BP_Adaboost 算法	T	3 026 529	135 939
	D	32 957	2 001 357

轨迹点识别正确的数量为田间轨迹点识别为田间轨迹点个数与道路轨迹点识别为道路轨迹点个数之和。轨迹点状态识别准确率 P 为

$$P = \frac{T_{11} + T_{22}}{T} \times 100\%$$

(7)

式中 T_{11} ——田间作业轨迹点识别正确数量
 T_{22} ——道路行驶轨迹点识别正确数量
 T ——轨迹点总数量

用 DBSCAN 算法对轨迹点状态识别准确率为 93.35%,用 DBSCAN + BP_Adaboost 算法对轨迹点状态识别准确率为 96.75%。因此本文算法得到的轨迹点状态识别准确率高于 DBSCAN 算法。

当 DBSCAN 算法选取不同参数时,轨迹点识别准确率如图 4 所示。

由图 4 可以看出,本文方法依托于 DBSCAN 算

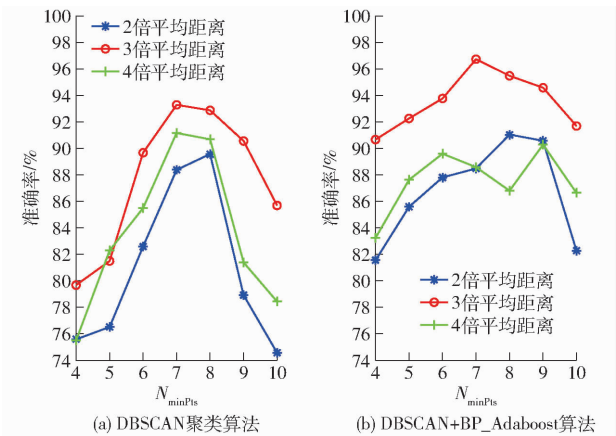


图 4 不同参数下两算法轨迹点识别准确率

Fig.4 Accuracy of two methods for track point recognition under different parameters

法,走势和 DBSCAN 算法有相关性,整体优于 DBSCAN 算法,并且准确率变化受参数影响小于 DBSCAN 算法。

使用密度聚类方法识别轨迹点状态的同时,根据点簇聚集关系划分不同地块。使用本研究方法得到轨迹点状态后,需再按照地块划分步骤进行划分,得到地块划分效果如图 5 所示。

图 5 不同算法划分地块效果图

Fig.5 Effect drawing of dividing plots by two methods

图 5 中灰色点表示道路行驶轨迹点,彩色点表示田间作业轨迹点,不同颜色表示不同地块。由图 5a 可以看到,部分道路行驶轨迹点被识别为田间作业轨迹点,下方绿色区域多个不同地块被聚类成一个地块。由图 5b 可以看到,使用本文方法划分地块,准确率高于只用 DBSCAN 算法。个别地块紧邻,由于阈值设置原因,DBSCAN 算法无法将两地块区分。用本方法能识别出两地块间转移行驶轨迹,

进而可以很好地精细划分地块。

地块划分错误分为 4 种类型：非地块误识别为地块、地块未被识别出、多个地块被识别成一个地块和一个地块被识别成多个地块。

其中，多个地块被识别成一个地块可以归为地块未被识别出；一个地块被识别成多个地块，可以归为非地块误识别成地块。

因此地块识别准确率 R 为

$$R = \frac{T_0}{T_f} \times 100\% \tag{8}$$

式中 T_0 ——识别正确的地块数量

T_f ——总地块数量

错误识别量为

$$E_n = T_1 + T_2 \tag{9}$$

式中 T_1 ——非地块误识别成地块和一个地块被识别成多个地块的数量

T_2 ——地块未被识别出和多个地块被识别成一个地块的数量

另外

$$T_f \leq T_0 + T_1 + T_2 \tag{10}$$

由于错误识别的数量和作业地块的数量没有关联性，因此地块识别的准确程度由地块识别准确率和错误识别个数组成。两算法划分地块结果如表 2 所示。

表 2 不同算法划分地块结果

Tab. 2 Results of dividing plots by two methods

方法	地块类型	识别为地	识别为非
		块数	地块数
DBSCAN 聚类算法	地块	5 072	214
	非地块	378	
DBSCAN + BP_Adaboost 算法	地块	5 184	155
	非地块	184	

由表 2 可知，由于非地块区域为道路行驶轨迹

点，将道路行驶轨迹点误识别为田间作业轨迹点会错分成地块，而道路行驶轨迹点识别正确仍旧是轨迹点而不是地块，因此不存在非地块识别为非地块的数量。通过 DBSCAN 算法得到地块识别准确数为 5 072，准确率为 95.63%，误识别个数为 592；通过 DBSCAN + BP_Adaboost 算法得到地块识别准确数为 5 184，准确率为 97.74%，误识别个数为 339。因此在识别准确率和误识别个数方面，DBSCAN + BP_Adaboost 算法都优于 DBSCAN 聚类算法。

本文提出的先通过 DBSCAN 算法得到轨迹点类别，再对轨迹点通过 BP_Adaboost 算法训练，轨迹点状态识别准确率高于只使用 DBSCAN 算法，且省去了大量人工标记的时间。后期有新农机作业轨迹识别，可以直接使用本方法，适用性较强。地块的划分准确率高于基于密度聚类方法，且错分的地块数量也少于基于密度聚类的方法。

3 结论

(1) 分析了农机作业轨迹特征及轨迹点有用信息，选取农机轨迹点经纬度、速度、近邻点关系等作为特征，实验证明特征有效。

(2) 利用 DBSCAN 算法对农机轨迹点聚类，得到带类别标记的田间作业轨迹点和道路行驶轨迹点；再用 BP_Adaboost 算法建立训练模型对农机轨迹点进行识别；将容易误识别的道路和田间交界处附近轨迹点标记为道路行驶轨迹点，并添加到训练样本中重复训练，直至训练样本不再增加。轨迹点状态识别准确率达到 96.75%，较 DBSCAN 算法有很大提升。

(3) 本文算法划分地块较 DBSCAN 算法聚类后得到的地块更准确，错分的地块数量更少。无需对样本标记，节省大量人力，并且对新农机作业轨迹有适用性。

参 考 文 献

[1] 张丰兆, 刘瑞华, 倪育德, 等. 北斗卫星导航系统动态定位精度测试与分析[J]. 全球定位系统, 2018, 43(1): 43–48. ZHANG Fengzhao, LIU Ruihua, NI Yude, et al. Dynamic positioning accuracy test and analysis of BeiDou satellite navigation system[J]. Global Positioning System, 2018, 43(1): 43–48. (in Chinese)

[2] 郑一力, 赵燕东, 刘卫平, 等. 基于北斗卫星通信的林区小气候监测系统研究[J]. 农业机械学报, 2018, 49(2): 217–224. ZHENG Yili, ZHAO Yandong, LIU Weiping, et al. Forest microclimate monitoring system based on Beidou satellite[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(2): 217–224. (in Chinese)

[3] 李道亮. 物联网与智慧农业[J]. 农业工程, 2012, 2(1): 1–7. LI Daoliang. Internet of things and intelligent agriculture[J]. Agricultural Engineering, 2012, 2(1): 1–7. (in Chinese)

[4] CHEN Y F. Discussion on the relationship between intelligent control and agricultural internet of things in agricultural complex large system[J]. Agricultural Network Information, 2012, 2: 8–12.

[5] LIU Hui, QIAO Yujie, ZHAO Guofa, et al. Agricultural machinery abnormal trajectory recognition[J]. International Journal of Machine Learning and Computing, 2021, 11(4): 291–297.

[6] LIU L X, SONG J T, GUAN B, et al. Tra-DBScan: a algorithm of clustering trajectories[J]. Applied Mechanics &

- Materials, 2012, 121 – 126:4875 – 4879.
- [7] WU Di, DU Yunyan, YI Jiawei, et al. Density-based spatiotemporal clustering analysis of trajectories[J]. Journal of Geo-Information Science, 2015, 17(10): 1162 – 1171.
- [8] 刘大有,陈慧灵,齐红,等. 时空数据挖掘研究进展[J]. 计算机研究与发展,2013,50(2): 225 – 239.
LIU Dayou, CHEN Huiling, QI Hong, et al. Advances in spatiotemporal data mining[J]. Journal of Computer Research and Development, 2013, 50(2): 225 – 239. (in Chinese)
- [9] SHAW S, YU Hongbo, LEONARD S B. A space-time GIS approach to exploring large individual-based spatiotemporal datasets[J]. Transactions in GIS,2008,12(4):425 – 441.
- [10] 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)
- [11] JOKINEN J, RATY T, LINTONEN T. Clustering structure analysis in time-series data with density-based clusterability measure[J]. Automatica Sinica, IEEE/CAA Journal of Automatica Sinica, 2019, 6(6): 1332 – 1343.
- [12] 吐尔逊·买买提,谢建华. 基于 DBSCAN 的农机作业轨迹聚类研究[J]. 农机化研究,2017,39(4): 7 – 11.
TURSUN Mamat, XIE Jianhua. Research on clustering of agricultural machinery operation trajectory based on DBSCAN algorithm[J]. Journal of Agricultural Mechanization Research, 2017,39(4):7 – 11. (in Chinese)
- [13] SCITO V K R, MAJSTOROVI S, SABO K. A combination of RANSAC and DBSCAN methods for solving the multiple geometrical object detection problem[J]. Journal of Global Optimization, 2021, 79(3): 669 – 686.
- [14] XUE X, HUANG S, XIE J, et al. Resolvable cluster target tracking based on the DBSCAN clustering algorithm and labeled RFS[J]. IEEE Access, 2021, 9: 43364 – 43377.
- [15] 张天翼, 丁立新. 一种基于 SMOTE 的不平衡数据集重采样方法[J]. 计算机应用与软件, 2021, 38(9):273 – 279.
ZHANG Tianyi, DING Lixin. A new resampling method based on SMOTE for imbalanced data set[J]. Computer Applications and Software, 2021, 38(9):273 – 279. (in Chinese)
- [16] LIU C L,HSIEH P Y. Model-based synthetic sampling for imbalanced data[J]. IEEE Transactions on Knowledge and Data Engineering,2020,32(8): 1543 – 1556.
- [17] DOUZAS G,BACAO F, LAST F. Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE[J]. Information Sciences,2018,465:1 – 20
- [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] 李蓓, 张兴敢, 方晖. 一种改进的 BP – Adaboost 算法及在雷达多目标分类上的应用[J]. 南京大学学报(自然科学版), 2017, 53(5): 984 – 989.
LI Bei, ZHANG Xinggan, FANG Hui. An improved algorithm of BP – Adaboost and application of radar multi-target classification[J]. Journal of Nanjing University(Natural Science), 2017, 53(5): 984 – 989. (in Chinese)
- [20] 石惠娴,孟祥真,游煜成,等. 植物工厂地源热泵系统热负荷 BP 神经网络预测及验证[J]. 农业工程学报,2019,35(2): 196 – 202.
SHI Huixian, MENG Xiangzhen, YOU Yucheng, et al. Prediction and verification on heating load of source heat pump heating system based on BP neural network for plant factory[J]. Transactions of the CSAE, 2019, 35(2): 196 – 202. (in Chinese)
- [21] 包涵春. 基于 BP 神经网络能量预测的地面式超级电容储能系统能量管理策略[D]. 北京:北京交通大学,2020.
BAO Hanchun. Energy management strategy of wayside supercapacitor energy storage system based on BP neural network energy prediction[D]. Beijing: Beijing Jiaotong University, 2020. (in Chinese)