

基于自步集成学习的非道路移动机械排气烟度预测模型

许小伟¹, 肖喆¹, 王明达², 祝能¹, 钱枫¹, 肖宁强³

(1. 武汉科技大学汽车与交通工程学院, 湖北 武汉 430065;

2. 中国环境科学研究院, 北京 100012; 3. 微宏动力系统(湖州)有限公司, 浙江 湖州 313000)

摘要:非道路移动机械逐步成为减排潜力挖掘的重点。本研究对山东省 16 个地级市共计 13 413 台非道路移动机械开展排气烟度实测工作, 通过分析机械排气烟度值和烟度检测合格率随机械类型、使用年限、功率的分布情况及所属经济圈的变化规律, 绘制山东省排气烟度污染空间分布图; 在分析单变量影响因素的基础上, 选取 6 类易获取的机械属性作为特征参数, 基于自步集成学习算法构建叉车烟度检测结果预判模型。结果表明: 非道路移动机械标准从国 I 提升至国 II、国 II 提升至国 III, 排气烟度分别减小了 33% 和 50%; 推土机、沥青摊铺机、工业钻探设备以及叉车的整体排放状况较差; 山东省三大经济圈的机械保有量从大到小依次为胶东 > 省会 > 鲁西南, 机械整体排放状况从好到差依次为省会 > 胶东 > 鲁西南; 自步集成学习算法在决策树和 K 近邻两个分类器下获得了所有模型评价指标的最高值, 其在不平衡数据上具有更好的分类性能。

关键词:非道路移动机械; 排气烟度; 不平衡学习; 自步集成学习; 预测模型

中图分类号: X511

文献标志码: A

Prediction model for exhaust smoke of non-road vehicles with self-paced ensemble learning

XU Xiaowei¹, XIAO Zhe¹, WANG Mingda², ZHU Neng¹, QIAN Feng¹, XIAO Ningqiang³

(1. School of Automobile and Traffic Engineering, Wuhan University of Science and Technology,

Wuhan 430065, China; 2. Chinese Research Academy of Environmental Sciences, Beijing 100012, China;

3. Microvast (Huzhou) Co. Ltd, Huzhou 313000, China)

Abstract: Non-road vehicles have gradually become the focus of emission reduction work. In this study, exhaust smoke was tested on 13 413 non-road vehicles in 16 cities in Shandong province. By analyzing the variation law of the distribution of exhaust smoke values and smoke test pass rates of different vehicle types, different service life periods, different powers, and different economic circles they belong to, the spatial distribution map of exhaust smoke pollution in Shandong province was drawn. Six types of easily obtainable machinery properties were selected as the feature parameters based on the analysis of the influence of single variables. Finally, the prediction model for exhaust smoke detection results of forklifts was constructed on the basis of the self-paced ensemble learning algorithm. The results show that with the increase of the non-road vehicles standards from National I to National II, and then from National II to National III, the exhaust smoke has been reduced by 33% and 50% respectively. Bulldozers, asphalt pavers, industrial drilling equipments and forklifts have poor overall emission status. The three major economic circles in Shandong province can be ordered as Jiaodong, provincial capital, and southwestern in terms of machinery holdings, whereas they can be ordered as provincial capital, Jiaodong and southwestern Shandong in

收稿日期: 2021-12-29

基金项目: 国家自然科学基金项目(51975426); 武汉市科技计划项目(2019010701011393)

作者简介: 许小伟(1983—), 男, 湖北武汉人, 副教授, 博士生导师, 主要从事汽车电子控制与故障诊断、汽车排放控制技术、汽车智能网联技术等方面的研究. E-mail: xuxiaowei@wust.edu.cn

王明达(1982—), 男, 北京人, 工程师, 硕士, 主要从事机动车污染物排放检测研究, 本文通信作者.

E-mail: wangmingda@vecc-mee.org.cn

terms of the overall emissions of non-road vehicles. With the highest evaluation indexes under the decision tree and K nearest neighbor classifiers, SPE algorithm has better classification performance on unbalanced data.

Key words: non-road vehicle; exhaust smoke; imbalance learning; self-paced ensemble learning; prediction model

大气污染是我国主要的环境问题之一。虽然近年来我国机动车大气污染物减排工作取得了重要进展,但非道路移动机械排放污染管控相对滞后^[1-4]。非道路移动源逐步成为“打赢蓝天保卫战”^[5]减排潜力挖掘的重点。目前关于非道路移动机械实测烟度的研究^[6-9]测试样本容量小(1 000 台左右),统计分析时重点关注各类机械排气烟度值的大小,没有研究合格率,忽略了检测标准对不同机械排气烟度限值的差异,并且选用的统计分析方法仍基于单变量条件,没有进一步探讨各类机械属性之间的相互作用对最终排气烟度检测结果的影响。

秦之涓等^[10]基于决策树算法建立了综合性机动车环检结果预测模型。在搭建模型过程中,发现数据集为不平衡数据集,这符合现实工业数据集的特点,即某一类样本数量显著多于另一类。传统机器学习模型为追求总体分类精度最大化,对多数类产生“偏好”,忽略了实际携带信息更多的少数类。针对该问题,有研究者提出采用 SMOTE^[11]、ADASYN^[12] 等重采样策略来重新平衡样本分布,但该方法只解决了类间数量不平衡问题,未注意到不平衡学习的难点在于类间重叠和内部噪声。为此,一些研究者将集成学习技术引入不平衡数据的分类问题中。自步集成学习(self-paced ensemble, SPE)算法^[13]作为一种欠采样加集成的学习策略,不仅能够不平衡学习中得到更好的分类性能,还能提高模型对噪声的鲁棒性。

非道路移动机械具有保有量大、流动性强、难以集中监管的特点。国内各地针对非道路移动机械的管控办法多为监督性尾气抽测或直接对尾气黑烟现象拍照执法^[14]。然而,机械出现排黑烟现象时,其排放已急剧恶化,而随机抽检也存在机械漏检的问题。山东省作为我国经济和社会建设较为活跃的省份之一,机动车和非道路移动机械保有量均居于国内前列。已有研究建立了山东省机动车排放清单^[15],但有关非道路移动机械的研究却少有报道。

本研究在 13 413 条大样本数据集上,进一步挖掘单因素条件下,不同使用年限、不同机械类型以及不同发动机功率的机械之间排气烟度值和检测合格率的差异;结合山东省经济圈的划分,在空间维度上分析不同经济圈机械类型的分布情况与烟度检测合格率的分布规律;引入自步集成学习算法构建叉车排气烟度检测结果预判模型,以期为非道路移动源需重点管控的区域和机械提供依据和参考。

1 数据与方法

1.1 数据来源

本研究使用的数据为山东省非道路移动机械监督性尾气抽检数据,在全省范围内共计抽检 13 413 台机械,检测时间为 2019—2020 年,所有机械均处于正常作业工况。我国对非道路移动机械执行《非道路移动柴油机械排气烟度限值及测量方法》(GB 36886—2018)^[16],不同排放标准以及不同功率的机械对应不同的排气烟度限值,机械的排气烟度若超过该限值,检测结果将被判定为不合格。本研究抽检的 13 413 台机械中,合格机械 10 834 台,不合格机械 2 579 台,合格机械数量远大于不合格机械,属于不平衡数据。被检测机械的排放等级分布如图 1(a)所示,机械类型分布如图 1(b)所示。其中,由于推土机、沥青摊铺机、工业钻探设备、发电机组、空气压缩机和机场地勤设备数量较少,统归为其他分类,总占比 1%。

1.2 模型构建方法与评价指标

1.2.1 自步集成学习原理

自步集成学习由 Liu 等^[13]提出,该算法旨在设计一种不依赖于距离度量的、基于模型的自步优化机制,从而达到提高模型效率、泛化能力和鲁棒性的目的。该算法创新性地引入了“分类硬度分布”的概念,样本 (x, y) 的分类硬度不仅与样本本身的特性相关,还受所选训练模型 F 的影响。 (x, y) 的分类硬度函数可表示为 $H = H(x, y, F)$,其中 F 表示当前模型。这样算得的分类硬度函数 H 既包含了数据噪声信息,还考虑了模型容量,在训练时能够更真实地反映模型在当前数据集上的拟合情况。模型训练过程中,结合分类硬度

函数返回的信息,引入自步因子 α 来指导模型更多地关注对当前集成贡献较大的样本。

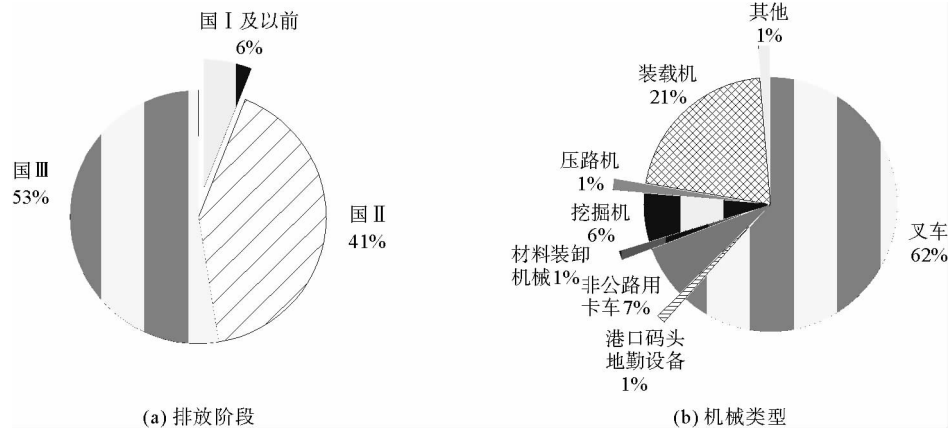


图 1 测试机械属性分布情况

Fig. 1 Distribution of test machinery properties

整个训练过程可以看作是欠采样加集成的串行训练模式。首先根据样本硬度值将多数类样本分为 k 个分桶,在第一次迭代过程中,假设每个分桶对分类硬度的贡献近似平衡,给予每个分桶相同的分类硬度值。与传统的集成算法(如 AdaBoost)相比,该方法能有效避免训练后期样本噪声的严重影响,同时保留了边界样本的有效信息。之后的迭代过程中,每个分桶的硬度值是动态更新的,第 l 个分桶的分类硬度 B_l 值可表示为:

$$B_l = \{(x, y) \mid \frac{l-1}{k} \leq H(x, y, F) \leq \frac{l}{k}\}, H \in [0, 1].$$

随着训练过程的推进,能被模型正确分类的“简单”样本数量快速增长,此时对每个分桶执行均衡化采样策略已不是最优选择。为提高训练后期基分类器的泛化能力并使模型关注更加难以识别的样本,引入自步因子 α 来降低包含较多“简单”样本分桶的采样权重,根据重新分配的采样权重对每个分桶的样本进行采样,用得到的新数据集训练下一个分类器。算法步骤可描述如下。

算法 1 自步集成学习

输入:训练集 Δ ,分类硬度函数 H ,基分类器 f ,基分类器数目 n ,分桶数目 k 。

初始化: Π 为训练集中的少数类, N 为训练集中的多数类。

对多数类进行随机欠采样得到 N'_0 ,用 N'_0 和 Π 训练基分类器 f_0 ,其中 $|N'_0| = |\Pi|$ 。

for $i = 1:n$

集成分类器: $F_i(x) = \frac{1}{i} \sum_{j=0}^{i-1} f_j(x)$ 。

将多数类样本的分类硬度分布 $H(x, y, F_i)$ 分为 k 个分桶: B_1, B_2, \dots, B_k 。

给予每个分桶相等的分类硬度值: $h_l = \sum_{s \in E} H(x_s, y_s, F_i) / |B_l|, \forall l = 1, 2, \dots, k$ 。

更新自步因子: $\alpha = \tan\left(\frac{i\pi}{2n}\right)$ 。

对每个分桶重新分配取权重: $p_l = \frac{1}{h_l + \alpha}, \forall l = 1, 2, \dots, k$ 。

对第 l 个分桶进行下采样: $\frac{p_l}{\sum_m p_m} \cdot |\Pi|$ 。

将下采样后得到的新数据集用来训练 f_i 。

end

输出: $F(X) = \frac{1}{n} \sum_{m=1}^n f_m(x)$ 。

1.2.2 模型性能评价指标

混淆矩阵(如表 1)将样本实际值和模型预测值关联起来,是衡量模型分类准确度最基本、最直观的方法。混淆矩阵的行代表真实类别,列代表预测值,主对角线表示预测正确的样本。

与标准机器学习不同,不平衡学习更注重对少数类的识别,这意味着模型可能会将一部分多类样本误划分为少类,牺牲了模型整体的准确率。通用的评价指标如准确率、召回率、ROC 曲线等都是基于模型整体性能的,会对多数类产生“偏好”。为避免此问题,研究引入 G-mean、F1-score、MCC 三类综合指标来评价不平衡学习的表现^[17],各指标含义以及公式如表 2 所示。

表 1 混淆矩阵示例

Table 1 Confusion matrix example

真实值	预测值	
	0	1
0	T_N	F_P
1	F_N	T_P

表 2 模型评价参数

Table 2 Parameters of model evaluation

评价指标	含义	公式
准确率(A)	所有分类正确的样本点与总样本点个数的比值	$A = \frac{T_P + T_N}{T_P + F_P + T_N + F_N}$
精确率(P)	被预测为正的样本中实际为正的比值	$P = \frac{T_P}{T_P + F_P}$
召回率(R)	正样本中被预测为正的比值,衡量的是模型的查全率	$R = \frac{T_P}{T_P + F_N}$
特异度(S)	实际为负的样本中被预测为负的比值,衡量的是模型对负样本的识别能力	$S = \frac{T_N}{T_N + F_P}$
G-mean(G)	正例准确率与负例准确率的综合指标	$G = \sqrt{R \times S}$
F1-score(F)	精确率和召回率的加权调和平均	$F = \frac{2PR}{P + R}$
MCC(M)	描述实际分类与预测分类之间的相关系数	$M = \frac{T_P \times T_N - F_P \times F_N}{\sqrt{(T_P + F_P)(T_P + F_N)(T_N + F_P)(T_N + F_N)}}$

2 单变量条件下排气烟度及检测合格率分布

影响工程机械排放状况的因素包括机械出厂年份、机械功率、油品差异及发动机构造等^[8-9]。本研究重点分析机械出厂年份、机械功率、机械类型等易获取的机械属性与机械排气烟度以及检测合格率的关系。按照国家标准 GB 36886—2018,各地市针对非道路移动机械的管控问题出台了不同的地方政策,导致山东省工程机械排放状况在空间分布上存在差异性。结合山东省三大经济圈的划分,对不同经济圈的机械类型分布以及机械整体排放状况进行分析。

2.1 不同使用年限机械排气烟度及检测合格率分布

机械使用年限常被当作衡量机械排放水平最直观的因素,分析不同使用年限机械的排气烟度特征及烟度检测合格率。如图 2 所示,随着机械使用年限的减小,平均烟度整体呈下降趋势,检测合格率呈上升趋势。其中,2009 年和 2016 年分别是国 II、国 III 标准正式实施年份。可知,非道路移动机械标准从国 I 提升至国 II,排气烟度整体减小 33%;从国 II 提升至国 III,排气烟度整体减小约 50%。由此可见,非道路移动机械排放标准收严,对降低排气烟度有直接作用。

2.2 不同机械类型排气烟度及检测合格率分布

不同机械类型排气烟度及检测合格率如图 3 所示。首先,挖掘机、港口码头地勤设备、机场地勤设备三类机械检测合格率较高,检测时发现这三类机械尾气后处理装置加装改造占比高。其次,占比最多的叉车排气烟度偏高,合格率偏低,误差棒较窄,表明叉车类机械整体排放较差。再次,推土机、沥青摊铺机、工业钻探设备三类机械排气烟度高,合格率最低,仅在 70%左右。经分析得知,该三类机械工作环境恶劣,且缺乏日

常的维护保养,导致机械排放状况差。

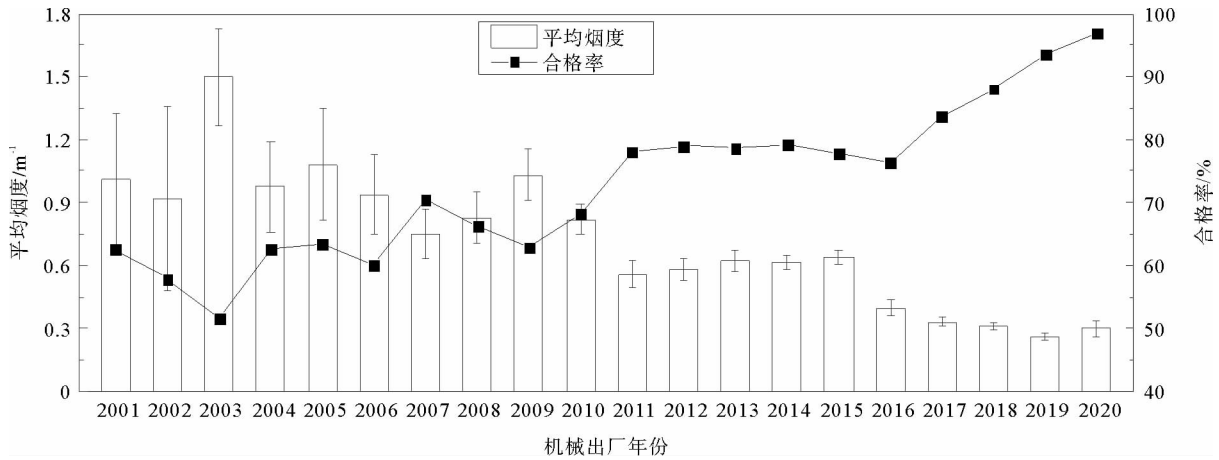


图2 不同出厂年份机械的排气烟度和合格率分布

Fig. 2 Distribution of exhaust smoke and qualified rate of non-road vehicles with different manufacturing years

2.3 不同功率机械排气烟度及检测合格率分布

根据国家标准 GB 36886—2018 对发动机功率的划分类型,将被检测机械发动机功率划分为 7 个功率区间,如表 3 所示,各功率区间机械的平均烟度和检测合格率如图 4(a)所示。

整体来看,1~5 功率区间(10~130 kW)内,烟度检测合格率随发动机功率增加呈下降趋势;6~7 功率区间机械排气烟度值最低,合格率最高。局部来看,1 功率区间排气烟度值大,合格率反而高,分析原因在于小功率机械排放限值高于大功率机械。在排放阶段相同的前提下,3~7 功率区间限值相同,其中国 II 及以前机械排气烟度限值为 1.61 m^{-1} ,国 III 机械限值为 0.8 m^{-1} 。3 功率区间排气烟度值高,同时合格率也高。进一步分析发现,3 功率区间国 II 及以前机械占比达 85%,较 4 功率区间(49%)和 5 功率区间(62%)高。因此,从平均值来看,3 功率区间限值高于 4、5 两功率区间,机械检测合格率高。

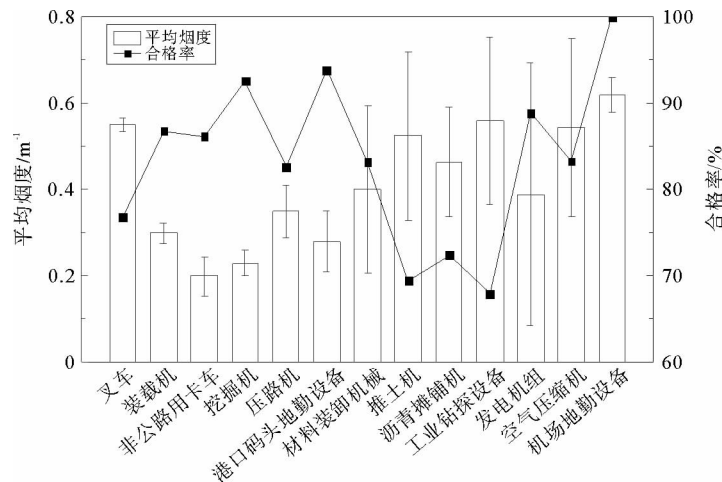


图3 不同类型机械的排气烟度和合格率分布

Fig. 3 Distribution of exhaust smoke and qualified rate of different types of non-road vehicles

表3 发动机功率分类

Table 3 Categories of engine power

序号	1	2	3	4	5	6	7
功率区间/kW	[10,19)	[19,37)	[37,56)	[56,75)	[75,130)	[130,560)	≥ 560

图 4(a)中 1 功率区间误差棒较大,将 1 功率区间样本排气烟度分布绘制图 4(b)。由图 4(b)可见,1 功率区间样本点分布较为集中,约 80% 机械的排气烟度在 2.5 m^{-1} 以内,60% 机械的排气烟度在 1.0 m^{-1} 以内。误差棒较大的原因是 7 台不合格机械中,仅有 1 台机械排气烟度为 3.61 m^{-1} ,其余 6 台机械排气烟度

均大于 7 m^{-1} , 2 台机械排气烟度达到烟度计的最大测量值 16 m^{-1} 。由此可见, 1 功率区间限值高, 机械排放一旦超过限值, 将成为高污染机械。

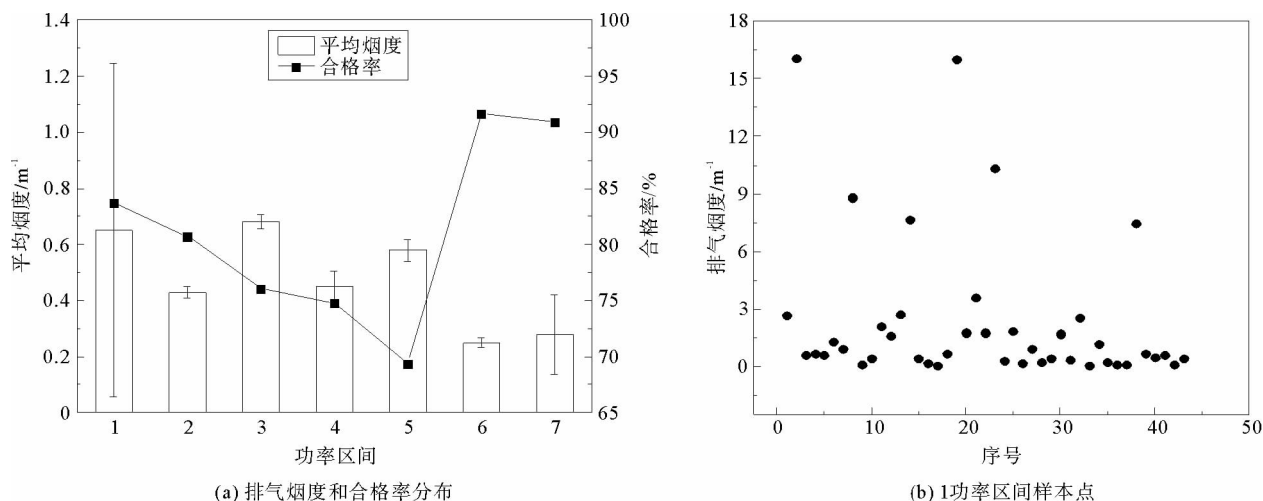


图 4 不同功率机械的排气烟度和合格率分布

Fig. 4 Distribution of exhaust smoke and qualified rate of non-road vehicles in different power ranges

Lu 等^[8]对两类功率区间(大于 130 kW 的大功率起重机和小于 56 kW 的小功率叉车)的研究发现, 大功率机械排气烟度高于小功率机械。然而, 在本研究 13 413 台机械实测排气烟度的样本集上, 大于 130 kW 的大功率机械(6~7 功率区间)排气烟度最低, 合格率最高。进一步分析发现, 6~7 功率区间装载机占比较多, 合格率低, 的叉车占比较少 1~5 功率区间少, 导致该区间平均烟度低。

2.4 机械排气烟度空间分布特征

根据 2020 年 2 月山东省人民政府印发的《贯彻落实〈中共中央、国务院关于建立更加有效的区域协调发展新机制的意见〉的实施方案》中提出的“三圈”, 将 16 个地级市划分为省会、胶东和鲁南三大经济圈。其中: 省会经济圈包括济南市、淄博市、泰安市、德州市、滨州市、东营市、聊城市共 7 市, 胶东经济圈包括青岛市、潍坊市、日照市、烟台市、威海市共 5 市, 鲁南经济圈包括临沂市、枣庄市、济宁市、菏泽市共 4 市。为平衡各经济圈内的城市数, 本研究将省会经济圈西南部的聊城市归类到鲁南经济圈, 称作鲁西南经济圈。

三大经济圈抽检的机械数量分别为: 省会 4 372 台、胶东 5 028 台、鲁西南 4 013 台, 各经济圈抽检比均在 10% 左右。可以推测出: 胶东经济圈机械保有量最大, 其次是省会经济圈, 鲁西南经济圈机械保有量最少, 这与三大经济圈经济体量排序基本一致。分析各经济圈机械类型分布, 如图 5 所示, 三大经济圈企业内部机械占比最多, 均超过 80%, 其次是建筑工地机械占比稍高(6%~14%), 胶东经济圈由于地处山东半岛, 港口码头机械占比超 10%, 较其他两大经济圈突出。

分析三大经济圈被检测机械的排放等级分布以及不同排放等级机械烟度检

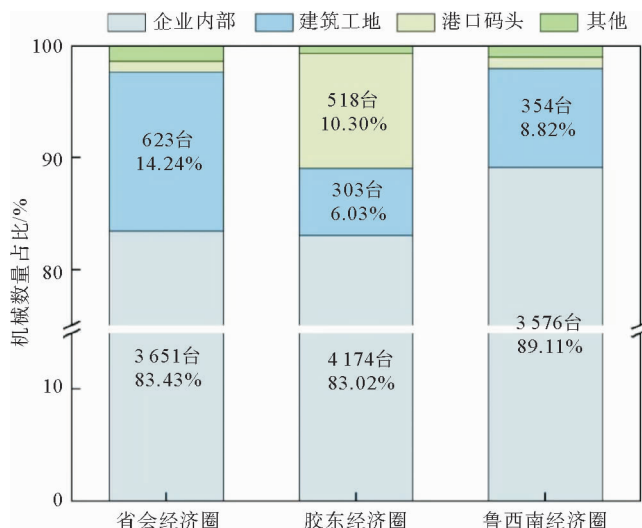


图 5 三大经济圈机械属性分布

Fig. 5 Distribution of vehicle types in the three economic circles

放阶段、不同功率的机械烟度限值不同,选取烟度限值为 1.61 m^{-1} ,即功率在 $[37, 560]$ kW 范围内的叉车数据,共计 2 492 台。结合本数据集上叉车的特征差异,最终筛选出传动形式、额定起重量、额定功率、使用场所、排放标准、地区等 6 类属性参与构建模型,各变量属性及赋值情况见表 5。

表 5 叉车常见机械属性

Table 5 Common mechanical properties of forklifts

机械属性	变量类型	赋值情况
传动形式	分类	1=动液传动,2=机械传动
额定起重量	连续	实际值
额定功率	连续	实际值
使用场所	分类	1=货运物流,2=其他,3=轻加工,4=机械制造,5=机械加工,6=冶金矿工,7=化工,8=建筑建材
排放标准	分类	1=国 I 及以上,2=国 II,3=国 III
地区	分类	1=省会,2=胶东,3=鲁西南

3.2 预测模型对比

为验证自步集成学习算法在本研究问题上的有效性,选用常见的 4 类分类器:决策树、K 近邻、支持向量机和 AdaBoost 进行对比。对于传统的不平衡学习方法,运用 SMOTE 或 ADASYN 对训练集过采样后,分别输入 4 类模型进行训练,再将模型在测试集上的误差作为最终模型应对现实场景中的泛化误差。值得注意的是,测试集是未经过采样技术处理的原始不平衡集合,对于自步集成学习则不需要对数据集做处理,直接将训练集输入模型,再将训练得到的模型用于测试集。自步集成学习可以理解作为一种集成学习方法,为此引入一种集成学习算法 OverBoost 作为对照。实验结果如表 6 所示。

结果表明:①在 4 种典型的

分类器下,自步集成学习算法有效提升了对不平衡数据的判别能力,与过采样技术相比,G-mean 最高提高 12.9%,准确率最高提高 12.8%,且该算法对原始数据集未做任何更改,保留了数据集真实的信息;②传统的过采样方法只在特定的分类器下(支持向量机)才能获得理想的结果,在其他分类器上的性能表现不如 OverBoost、SPE 两类集成学习算法;③对于非结构化的、包含较多离散型变量的工业数据集,决策树算法的整体性能优于其他 3 类典型分类器。

选取自步集成学习算法构建最终的预判断模型,以决策树作为基分类器,基分类器数目 $n=23$ 、分桶数目 $k=5$,模型在测试集上的其他评价指标如表 7 所示。模型的准确率为 70.0%,实际合格与不合格两类的召回率分别为 70.1%和 69.2%,即在保证模型整体性能的前提下,模型对 2 类机械的识别能力均较高,且较

表 6 不同算法在不平衡数据集上的表现

Table 6 Performance of different algorithms on unbalanced dataset

分类器	方式	G-mean	F1-score	MCC	准确率/%
决策树	SMOTE	0.524	0.506	0.190	57.2
	ADASYN	0.471	0.471	0.199	63.9
	OverBoost	0.580	0.574	0.338	68.6
	SPE	0.600	0.595	0.371	70.0
支持向量机	SMOTE	0.553	0.704	0.253	61.1
	ADASYN	0.547	0.537	0.266	63.6
	OverBoost	0.544	0.513	0.184	53.5
	SPE	0.553	0.544	0.272	63.2
AdaBoost	SMOTE	0.551	0.540	0.266	63.1
	ADASYN	0.543	0.532	0.251	62.3
	OverBoost	0.557	0.551	0.309	67.8
	SPE	0.569	0.564	0.318	67.1
K 近邻	SMOTE	0.547	0.545	0.309	68.6
	ADASYN	0.509	0.509	0.262	67.0
	OverBoost	—	—	—	—
	SPE	0.550	0.549	0.327	70.2

注:加粗值为该项指标的最优结果

为均衡,不会对多数类产生过分“偏好”。若监管部门仅对模型预测为不合格的机械进行实际排气烟度检测,则需检测车辆数为 $152+166=318$ (台),较检测整个样本(748台)减少了 57.5%,而此时漏检的不合格机械为 74 台,仅占总样本的 10%。

表 7 预判断模型评价

Table 7 Evaluation of prediction model results

评价指标	混淆矩阵		精确率/%	召回率/%	F1-score	类别数量	准确率/%
	预测-合格	预测-不合格					
实际-合格	356	152	82.8	70.1	0.759	508	70.0
实际-不合格	74	166	52.2	69.2	0.595	240	

4 结论

1) 不同使用年限机械排放状况差异明显,整体排序为:国 III > 国 II > 国 I 及以前,非道路移动机械检测标准收严能够有力减小机械排气烟度。挖掘机、港口码头地勤设备、机场地勤设备由于加装柴油颗粒过滤器装置,排放状况良好;推土机、沥青摊铺机、工业钻探设备以及叉车排放状况偏差。因此,无论是在企业使用还是执法部门监管时,都需要有针对性地重点关注上述类型的工程机械烟度排放情况。对于高排放机械,可采取加装柴油颗粒过滤器等尾气后处理装置来降低排放。

2) 在 13 413 台机械的大样本数据集上发现:机械功率与烟度检测合格率呈现出一定的规律性,即[10, 130)kW 功率范围内,机械烟度检测合格率随功率增大而减小,超过 130 kW 的大功率机械中叉车占比小,装载机占比大,排气烟度最低,合格率最高。实际测量得到的各类型机械平均额定功率普遍高于《非道路移动源排放清单编制技术指南(试行)》(公告 2014 年第 92 号)给出的推荐值。

3) 山东省三大经济圈非道路移动机械保有量胶东最多,鲁西南最少;鲁西南经济圈机械烟度检测合格率较其他两大经济圈低 15%,排气烟度污染偏高。

4) 在 4 种典型的分类器下,自步集成学习算法的整体表现优于传统的不平衡学习算法,有效地提升了分类算法对不平衡数据的判别能力。本研究提出的预判断模型,准确率为 70%,召回率为 69.2%,对不合格机械识别能力较高。

参考文献:

[1] 中华人民共和国生态环境部.非道路移动机械用柴油机排气污染物排放限值及测量方法(中国第三、四阶段):GB 20891—2014[S].北京:中国环境科学出版社,2014.

[2] 中华人民共和国生态环境部.非道路移动机械用柴油机排气污染物排放限值及测量方法(中国 I、II 阶段):GB 20891—2007[S].北京:中国环境科学出版社,2007.

[3] EUROPEAN PARLIAMENT.Regulation (EU) 2016/1628 of the European Parliament and of the Council of 14 September 2016 on requirements relating to gaseous and particulate pollutant emission limits and type-approval for internal combustion engines for non-road mobile machinery[EB/OL].(2016-09-16).<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R1628&qid=1646036204758>.

[4] WANG F,LI Z,ZHANG K,et al.An overview of non-road equipment emissions in China[J].Atmospheric Environment, 2016,132(5):283-289.

[5] 国务院.打赢蓝天保卫战三年行动计划[R].北京:国务院,2018.

[6] 左跃,穆德松,吕石成,等.黄骅港非道路移动机械排气烟度实测分析[J].皮革制作与环保科技,2021,2(20):120-123. ZUO Yue,MU Desong,LÜ Shicheng,et al.Actual measurement and analysis of exhaust smoke from non-road mobile machinery in Huanghua Port[J].Leather Manufacture and Environmental Technology,2021,2(20):120-123.

[7] 杨保华,崔焕星,刘树成,等.济南市在用典型工程机械分布及排气烟度特性[J].工程机械,2021,52(10):108-112. YANG Baohua,CUI Huanxing,LIU Shucheng,et al.Distribution and exhaust smoke characteristics of typical construction

- machinery in service in Jinan[J]. *Construction Machinery and Equipment*, 2021, 52(10):108-112.
- [8] LU J, HU Q Y, YANG Q, et al. Exhaust smoke of non-road vehicles in typical cities of eastern China[J]. *Environmental Science*, 2020, 41(1):133-138.
- [9] 陈秋宇, 周子航, 邓也, 等. 成都市典型非道路工程机械排气烟度分析[J]. *环境科学学报*, 2019, 39(11):3709-3715.
CHEN Qiuyu, ZHOU Zihang, DENG Ye, et al. Exhaust smoke analysis of typical non-road engineering machinery in Chengdu[J]. *Acta Scientiae Circumstantiae*, 2019, 39(11):3709-3715.
- [10] 秦之涓, 熊阳欣, 费怡, 等. 基于决策树的城市在用车环检首检结果预测模型研究[J]. *环境科学学报*, 2021, 41(4):1574-1583.
QIN Zhimei, XIONG Yangxin, FEI Yi, et al. Modelling the initial emissions inspection of urban in-use vehicles with decision trees[J]. *Acta Scientiae Circumstantiae*, 2021, 41(4):1574-1583.
- [11] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: Synthetic minority over-sampling technique[J]. *Journal of Artificial Intelligence Research*, 2002, 16:321-357.
- [12] HE H, BAI Y, GARCIA E A, et al. ADASYN: Adaptive synthetic sampling approach for imbalanced learning[C]// 2008 IEEE International Joint Conference on Neural Networks (IEEE world Congress on Computational Intelligence). New York: IEEE, 2008:1322-1328.
- [13] LIU Z N, CAO W, GAO Z, et al. Self-paced ensemble for highly imbalanced massive data classification[C]// 36th International Conference on Data Engineering (ICDE). Dallas: IEEE, 2020:841-852.
- [14] 山东省生态环境厅. 山东省非道路移动机械排气污染防治规定[EB/OL]. (2019-12-17). http://zfc.sdein.gov.cn/hjbhdfzfgz/202005/t20200516_3106328.html.
- [15] 孙世达. 山东省机动车污染物排放演变研究[D]. 济南: 济南大学, 2017.
SUN Shida. Study on evolution of vehicle emissions in Shandong Province[D]. Jinan: University of Jinan, 2017.
- [16] 中华人民共和国生态环境部. 非道路移动柴油机械排气烟度限值及测量方法: GB 36886—2018[S]. 北京: 中国环境科学出版社, 2018.
- [17] LIN H, LU Y, HAN X, et al. Adaptive scaling for sparse detection in information extraction[C]// 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne: Association for Computational Linguistics, 2018:1033-1043.

(责任编辑: 傅 游)