

DOI: 10.3969/j.issn.1673-6141.2024.01.005

基于混核极限学习机的道路高排放源识别方法

段培杰^{1,2}, 李泽瑞^{2*}, 李鲲³, 许镇义², 吕钊⁴, 康宇^{2,3}

(1 安徽大学人工智能学院, 安徽 合肥 230601;
2 合肥综合性国家科学中心人工智能研究院, 安徽 合肥 230088;
3 中国科学技术大学先进技术研究院, 安徽 合肥 230088;
4 安徽大学计算机科学与技术学院, 安徽 合肥 230601)

摘要: 由于道路高排放源所产生的污染气体对环境危害巨大, 因此实现对高排放源的准确识别具有重要意义。而传统的基于限值划分的识别方法及新兴的人工智能识别方法在模型选择、评价指标、识别性能等方面都存在一定的改进空间, 因此针对以上问题, 提出一种基于混核极限学习机的道路高排放源识别方法。该方法使用道路遥感监测设备获取的移动源遥测数据, 在核极限学习机的基础上融合不同核函数, 可提升模型鲁棒性及道路高排放源识别性能。针对合肥市蜀山区真实道路遥测数据上的分析结果表明, 该方法相比于其他方法具有较高的F1分数以及较低的漏报率、虚警率, 证实了该方法在高排放源识别中的有效性。因此, 该方法有助于对交通路网中高排放车辆进行高效识别, 为进一步提升城市空气质量提供支撑。

关键词: 高排放识别; 混合核函数; 极限学习机; 道路遥感监测

中图分类号: X831

文献标识码: A

文章编号: 1673-6141(2024)01-062-011

On-road high-emitter identification method based on mixed kernel extreme learning machine

DUAN Peijie^{1,2}, LI Zeru^{2*}, LI Kun³, XU Zhenyi², LYU Zhao⁴, KANG Yu^{2,3}

(1 School of Artificial Intelligence, Anhui University, Hefei 230601, China;
2 Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China;
3 Institute of Advanced Technology, University of Science and Technology of China, Hefei 230088, China;
4 School of Computer Science and Technology, Anhui University, Hefei 230601, China)

Abstract: Since the pollution gases produced by on-road high-emitters are significantly harmful to the environment, it is of great significance to identify on-road high-emitters accurately. However, there is still a relatively large promotion space for identifying high-emitters in model selection, evaluation metrics, recognition performance, and other aspects, for both traditional cutpoint-based methods and emerging

基金项目: 国家自然科学基金 (62103125, 62033012), 安徽省博士后研究人员科研活动资助经费 (BSH202103)

作者简介: 段培杰 (2000-), 安徽阜阳人, 硕士研究生, 主要从事机器学习及其在大气环境监测等方面的应用研究。

E-mail: peijieduan@stu.ahu.edu.cn

导师简介: 李泽瑞 (1991-), 女, 河南焦作人, 博士, 副研究员, 硕士生导师, 主要从事机器学习及其在大气环境监测等方面的应用研究。

E-mail: lizr@iai.ustc.edu.cn

收稿日期: 2022-06-20; 修改日期: 2022-09-02

*通信作者。

artificial intelligence-based methods. Therefore, to address the above issues, a method for on-road high-emitters identification is proposed based on mixed kernel extreme learning machine. The method uses mobile source telemetry data obtained from on-road remote sensing detection equipment, and integrates different kernel functions on the basis of kernel extreme learning machine, which can improve the robustness of the model and the recognition performance of on-road high-emitters. The experimental results on remote sensing data collected from the traffic network in Shushan District of Hefei City, China, show that, compared with the other methods, the proposed method has a higher F1-Score, lower missing alarm rate and false alarm rate, which confirms the effectiveness of the method in high-emitter identification. It is indicated that the proposed method can help to identify high-emitters in the traffic road network efficiently and provide basic support for further improving urban air quality.

Key words: high-emitter identification; mixed kernel function; extreme learning machine; on-road remote sensing

0 引言

近年来,随着经济社会的快速发展,空气污染不断加剧,严重危害人类健康。当前我国移动源污染已成为空气污染的重要来源,因此实现对高排放移动源的有效识别具有重要意义^[1]。现有的识别方法是制定相关法规,规定车辆定期前往检测机构进行尾气分析,进而判断车辆是否符合排放标准。然而,这种检测方式费时费力且效率较低,更无法反映车辆在道路上实际运行时的情况。相比而言,道路遥感监测设备可以在不干扰正常驾驶的情况下,对车辆运行时产生的尾气成分进行检测,同时还能得到一些外部环境信息,如风速、温度等^[2]。

已有相关学者开展了使用道路遥感监测设备进行高排放移动源的识别研究。早期的研究是通过设定设备采集到的尾气遥测数据中CO、HC、NO等污染成分排放限值的方式来识别高排放源。Bishop等^[3]将CO排放限值设置为3%,并对高于限值的车辆进行维修,使CO的排放量减少了50%,证明了该方法的有效性。Pujadas等^[4]以每种污染物排放浓度的前5%作为分割点,在车辆比功率(VSP)小于0.02 kW/kg时,对应的CO、HC、NO的排放限值为:与CO₂的体积浓度比值分别为0.18、0.008和0.0088。在香港,相关部门综合考虑各种因素,将CO、HC及NO的排放限值保守设置为2%、0.05%、0.15%^[5]。近年来,随着人工智能技术的发展,不少学者开始使用表征能力更强的机器学习算法进行识别。Guo等^[6]使用遥测数据中车辆的速度、加速度及CO、HC、NO_x、CO₂的浓度、烟羽量共7维特征,建立了反向传播神经网络,达到81.63%的准确度。曾君等^[7]结合主成分分析法、K最近邻思想、遗传算法,并引入VPS特征,进一步提高了预测准确度。Li等^[8]考虑到遥测数据中高排放类别与正常排放类别的不平衡性,利用代价敏感思想,建立加权极限学习机模型,并设计了一种主动选择样本算法,用于模型更新。许镇义等^[9]提出一种基于深度特征聚类的高排放源识别模型,首先分析不同污染物排放的影响特征,对特征进行聚类分析并获取高排放类别标签,然后训练基于深度森林的移动污染源模型,自动识别高排放源。

上述基于限值划分的传统方式只考虑到了尾气中的几种污染气体成分。然而,在实际检测工作中,气体浓度还受周围许多环境因素的影响,如温度、湿度等^[10]。并且,限值的设定极大地依赖人工经验,因此该

方法具有一定的局限性和不稳定性。而基于人工智能技术的识别方法可以利用包括气体浓度及外界因素在内的多维数据识别高排放源。但现有的相关研究并不充分,在模型选择、评价指标、识别性能等方面有进一步的提升空间。

本文以合肥地区道路移动源为研究对象,设计了一种基于混核极限学习机的道路高排放源识别方法。极限学习机(ELM)是一种泛化能力强、识别精度高的网络模型,被广泛应用于各种场景^[11,12]。但是模型中的随机映射部分对其分类能力有一定的限制。本方法在ELM的基础上引入核函数,用核映射替代随机映射,达到将样本映射到较高维空间的目的,特别是使用了混合核函数,使得映射空间更加丰富,从而进一步提高分类能力。最后在合肥市蜀山区仰桥路收集的真实道路遥测数据上进行了实验,实验结果验证了该方法的有效性,并且与其他传统分类算法相比,该方法具有更高的F1分数。

1 道路遥感监测技术原理与设备

道路遥感监测技术是近些年来较为流行的车辆排放检测技术之一,与传统技术相比,具有检测速度快、检测成本低、效率高的优点^[13]。遥感监测的光学原理基于朗伯-比尔定律,当光线穿过尾气烟羽时,会被其中的气体分子选择性吸收而发生强度和结构的变化,将其与原光谱进行对比即可得到吸收光谱,进而分析出尾气烟羽中气体的种类与浓度。朗伯-比尔定律可表示为

$$\frac{I(\lambda)}{I_0(\lambda)} = \exp(-\delta cl), \quad (1)$$

式中 $I(\lambda)$ 表示透射光光强, $I_0(\lambda)$ 代表入射光光强, δ 为衰减系数, c 为某一气体的浓度, l 为吸收层厚度。基于这一原理,在道路遥感监测设备中,发射器发出红外线和紫外线激光,并穿过尾气烟羽到达接收器,根据激光强度的变化可计算出尾气中CO、HC和NO等污染物的浓度。

道路遥感监测设备(图1)通常由尾气检测器、速度/加速度检测器、车牌识别器、工控机、反射带、辅助设备等组成。其中,尾气检测器及反射带用来测量过往车辆产生的尾气烟羽中污染物种类及浓度,速度/加速度检测器用于获取车辆经过时的速度和加速度,车牌识别器可以拍摄并识别车牌号,辅助设备用于记录环境中的温度、湿度、风向、风速等信息,上述数据由工控机处理并存储在数据库中。

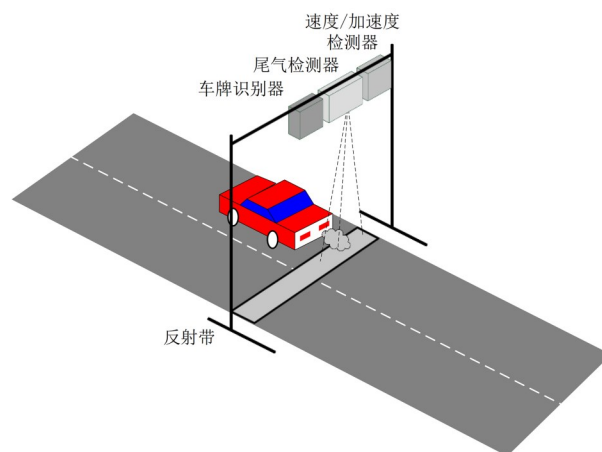


图1 道路遥感监测设备示意图

Fig. 1 Diagram of on-road remote sensing detection equipment

2 道路高排放源识别模型构建

2.1 极限学习机

ELM 是一种单隐层前馈神经网络, 网络结构由输入层、隐含层、输出层组成, 如图 2 所示。由于其输入层与隐含层之间的权重随机生成, 隐含层与输出层之间的权重通过计算 Moore-Penrose 广义逆得到, 因此该网络具有较高的学习效率及泛化能力。

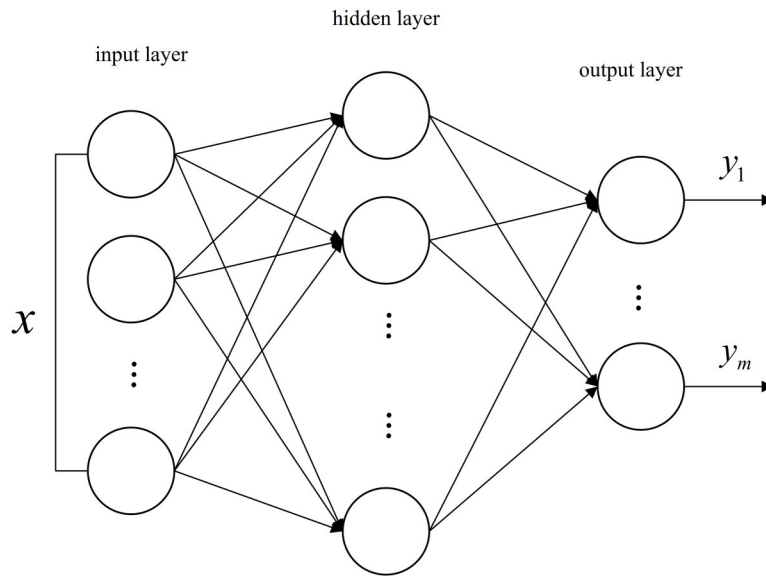


图 2 ELM 网络结构图
Fig. 2 Diagram of ELM network structure

对于任意的 N 个不同的样本 $\{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^N$, $\mathbf{x}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{id}]^T$, 代表 d 维输入数据, $\mathbf{t}_i = [\mathbf{t}_{i1}, \mathbf{t}_{i2}, \dots, \mathbf{t}_{im}]^T$ 代表输入数据对应的类别, m 表示类别数量, 具有 L 个隐含层节点的 ELM 模型预测结果可表示为

$$\mathbf{f}(\mathbf{x}_j) = \sum_{i=1}^L \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i), j = 1, \dots, N, \tag{2}$$

式中 β_i 为连接第 i 个隐含层节点和输出层各节点间的权重向量, $g(\cdot)$ 为激活函数, \mathbf{w}_i 为连接第 i 个隐含层节点和输入层各节点间的权重向量, b_i 为第 i 个隐含层节点的偏置。式 (2) 可以表示为

$$\mathbf{H}\beta = \mathbf{T}, \tag{3}$$

式中 $\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \dots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \dots & g(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix}_{N \times L}$, 为隐含层输出矩阵; $\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}$, 为隐含层输出权重矩阵,

$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m}$; \mathbf{T} 为目标矩阵。令 $\mathbf{h}(\mathbf{x})$ 为输入层到隐含层的映射函数, 对于数据 \mathbf{x}_i ,

$\mathbf{h}(\mathbf{x}_i) = [g(\mathbf{w}_1 \cdot \mathbf{x}_i + b_1) \quad \dots \quad g(\mathbf{w}_L \cdot \mathbf{x}_i + b_L)]$, \mathbf{H} 可表示为

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(x_1) \\ \vdots \\ \mathbf{h}(x_N) \end{bmatrix}. \quad (4)$$

为提高ELM的泛化能力与稳定性,引入正则化因子 C ,平衡模型的经验风险与结构化风险:

$$\min \|\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\|^2 + \frac{C}{2} \|\boldsymbol{\beta}\|^2. \quad (5)$$

输出权重 $\boldsymbol{\beta}$ 的解为

$$\boldsymbol{\beta} = \begin{cases} \mathbf{H}^T \left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{T}, & N \leq L \\ \left(\frac{\mathbf{I}}{C} + \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{T}, & N > L \end{cases}, \quad (6)$$

式中 \mathbf{I} 表示单位矩阵。因此,当 $L > N$ 时,ELM分类器的输出为

$$\mathbf{f}(\mathbf{x}) = \mathbf{h}(\mathbf{x})\boldsymbol{\beta} = \mathbf{h}(\mathbf{x})\mathbf{H}^T \left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{T}. \quad (7)$$

2.2 混核极限学习机

上述ELM模型中,样本被随机映射到 L 维空间。由于随机映射具有一定的不确定性并且映射空间维度有限,基于此,可以在ELM的基础上引入核函数,使用核映射代替随机映射,从而将数据映射到高维可分空间,称为核极限学习机(KELM)^[14]。将KELM的核矩阵 $\boldsymbol{\Omega}$ 定义为

$$\boldsymbol{\Omega} = \mathbf{H}\mathbf{H}^T = \begin{bmatrix} \mathbf{h}(x_1) \cdot \mathbf{h}(x_1) & \cdots & \mathbf{h}(x_1) \cdot \mathbf{h}(x_N) \\ \vdots & \ddots & \vdots \\ \mathbf{h}(x_N) \cdot \mathbf{h}(x_1) & \cdots & \mathbf{h}(x_N) \cdot \mathbf{h}(x_N) \end{bmatrix} = \begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_N) \\ \vdots & \ddots & \vdots \\ K(x_N, x_1) & \cdots & K(x_N, x_N) \end{bmatrix}. \quad (8)$$

将式(8)代入式(7),并结合式(4),KELM的输出可表示为

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} K(\mathbf{x}, x_1) \\ \vdots \\ K(\mathbf{x}, x_N) \end{bmatrix}^T \left(\frac{\mathbf{I}}{C} + \boldsymbol{\Omega} \right)^{-1} \mathbf{T}. \quad (9)$$

在KELM算法中,无需人为设置模型隐含层节点的数量 L 以及输入层到隐含层的特征映射函数 $\mathbf{h}(\mathbf{x})$,只需了解核函数 $K(\mathbf{x}, \mathbf{y})$ 的具体形式,就可以求出KELM输出函数的值。核函数的选择有多种,如高斯核函数 $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / 2\sigma^2)$,拉普拉斯核函数 $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\| / \sigma)$,多项式核函数 $K(\mathbf{x}, \mathbf{y}) = (\varphi \mathbf{x} \cdot \mathbf{y} + 1)^\gamma$,Sigmoid核函数 $K(\mathbf{x}, \mathbf{y}) = \tanh(\varphi \mathbf{x} \cdot \mathbf{y} + \gamma)$ 。

然而,单个核函数鲁棒性较差,分类精度有一定的局限性^[15],因此可以在KELM的基础上使用两个不同的核函数进行组合得到混合核函数,以此提高分类能力,称为混核极限学习机(MKELM)^[16]。假设有两个有效核函数 K_1, K_2 ,则二者的线性组合仍然是一个有效的核函数,可表示为

$$K_3 = \mu K_1 + (1 - \mu) K_2, \quad (10)$$

式中 μ 是权重系数,并且 $\mu \in [0, 1]$ 。

3 实验结果与分析

3.1 数据统计与可视化

使用的数据是由合肥市蜀山区仰桥路上架设的一套道路遥感监测设备采集到的真实数据。采集时间为 2017 年 5 月 1—31 日, 每当有车辆通过时, 设备中的发射器发出的红外线和紫外线激光会被车辆的尾气吸收一部分, 通过分析吸收光谱便可得到各污染物的浓度。数据样本共包括 11 维特征: CO、HC、NO、CO₂ 浓度及 VSP、速度、加速度、风速、风向、温度、车牌颜色, 全部作为模型的输入变量。实验共采集到 55211 个遥测数据样本, 与车辆年检站数据库匹配预处理后, 获取到 4092 个有效数据样本及对应的类别。其中高排放移动源数量为 218, 占总样本数的 5.33%, 其余为正常排放移动源, 数量为 3874, 占总样本数的 94.67%。为直观显示数据不同类别间样本数量的差异, 使用常用的降维方法 t-SNE^[17] 将 11 维实验数据降维至 2 维后进行可视化, 如图 3 所示, 其中绿色点代表正常排放源, 红色点代表高排放源。

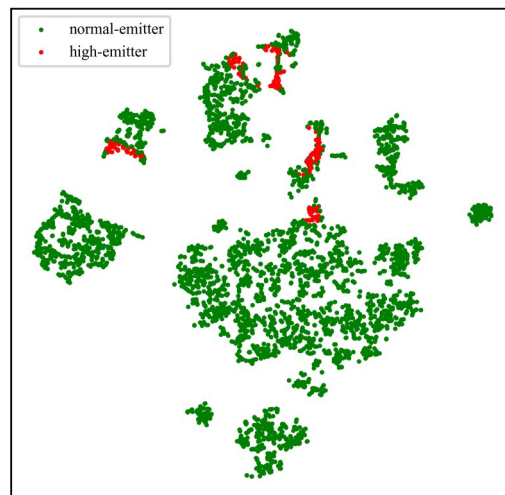


图 3 使用 t-SNE 降维后的数据分布情况可视化图

Fig. 3 Visualization diagram of data distribution after dimensionality reduction using t-SNE

3.2 实验评价指标

由图 3 可看出本实验数据集具有高度类别不平衡性, 由于常用的评价指标准确率无法真实表现出算法在不平衡数据集上的识别效果, 因此本研究采用的评价指标为漏报率 R_{MA} 、虚警率 R_{FA} 及 F1 分数 S_{F1} 。其中漏报率定义为在高排放样本中未被检测出的比例, 虚警率定义为模型预测为高排放的样本中正常排放被误判为高排放的比例, F1 分数对两者进行综合, 可以较为客观地衡量算法对高排放源的识别能力。

漏报率、虚警率、F1 分数的计算公式分别为

$$R_{MA} = \frac{N_{he} - T_{he}}{N_{he}}, \quad (11)$$

$$R_{FA} = \frac{F_{he}}{T_{he} + F_{he}}, \quad (12)$$

$$S_{F1} = 2 \frac{(1 - R_{MA})(1 - R_{FA})}{(1 - R_{MA}) + (1 - R_{FA})}, \quad (13)$$

式中 N_{he} 表示高排放样本数量, T_{he} 表示正确识别的高排放源数量, F_{he} 表示被算法误判为高排放源的正常排放源数量。性能优秀的模型应具有较高的 F1 分数、较低的漏报率、虚警率。

3.3 实验结果分析

MKELM 中的核函数是由两个核函数组合而成, 由于不同的核函数对应不同的映射空间, 核函数种类的选择对实验结果有重要影响, 因此首先通过实验比较不同核函数性能的优劣。图 4 是使用了单个不同种类核函数的 KELM 的决策区域图, 图中绿色点与红色点分别代表正常排放源与高排放源, 浅绿色区域与浅红色区域分别代表该区域内的移动源会被模型预测为正常排放和高排放。从图 4 (a)、图 4 (b) 中可看出有大量绿色点落入浅红色区域, 代表大量正常排放源被误识别为高排放源。相比而言, 图 4 (c) 与图 4 (d) 中各颜色区域内大部分为相同颜色的点, 识别结果明显优于前者。

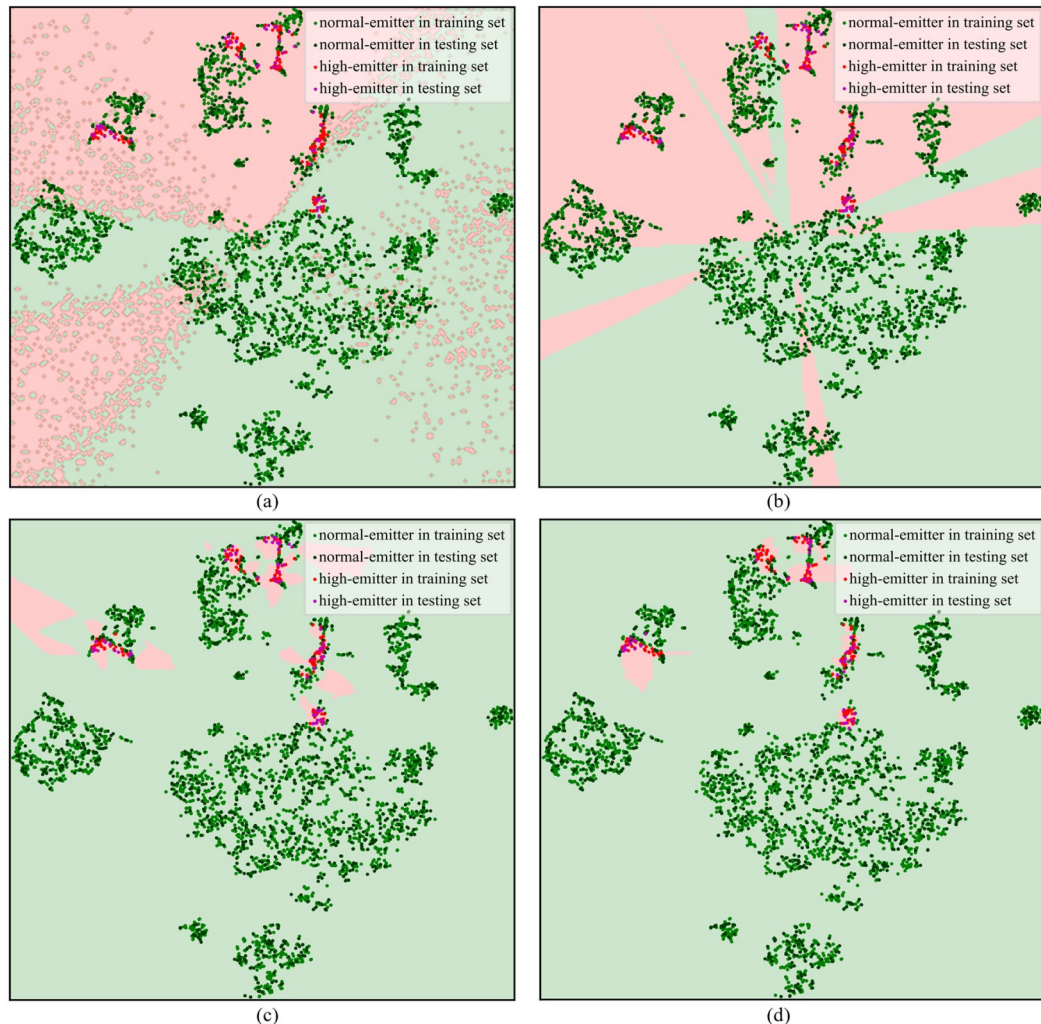


图 4 使用了四种不同核函数的 KELM 决策区域图。(a) 多项式核函数; (b) Sigmoid 核函数;
(c) 高斯核函数; (d) 拉普拉斯核函数

Fig. 4 Diagram of KELM decision region with four different kernel functions. (a) Polynomial kernel function; (b) Sigmoid kernel function; (c) Gaussian kernel function; (d) Laplacian kernel function

即使在一些非本领域的分类任务中, 使用多项式核函数效果较好^[18, 19], 但上述实验结果表明, 在道路高排放源识别任务上, 高斯核函数及拉普拉斯核函数更有优势。因此在接下来的实验中将对比高斯核函数及拉普拉斯核函数进行组合, 主要考虑以下三个新生成的核函数:

$$K_{G+G} = \mu K_{\text{Gaussian}} + (1 - \mu) K_{\text{Gaussian}}, \quad (14)$$

$$K_{G+L} = \mu K_{\text{Gaussian}} + (1 - \mu) K_{\text{Laplacian}}, \quad (15)$$

$$K_{L+L} = \mu K_{\text{Laplacian}} + (1 - \mu) K_{\text{Laplacian}}, \quad (16)$$

在式 (14) 及式 (16) 中均使用了种类相同但宽度 σ 不同的核函数进行组合。

接下来进行相关对比实验, 首先使用上述三种混合核函数进行实验, 判断核函数的最优种类组合方式。其次, 为证明模型中所使用混合核函数的有效性, 与未使用核函数的 ELM 及使用了单个核函数的 KELM 进行对比。最后本实验对比了主流机器学习方法支持向量机 (SVM)、随机森林 (RF), 来证明本研究所使用的 MKELM 方法的优越性。为使实验结果更具说服力, 对每种方法进行 5 折交叉验证, 即首先将数据集分为 5 份, 随后轮流挑选其中 1 份作为测试集, 剩余 4 份作为训练集。记录下这 5 次不同划分情况下的实验结果, 统计平均值与标准差。实验结果如表 1 所示, 表中单元格格式为: 均值 \pm 标准差。对比表中八种方法识别结果可以看出, 使用了高斯与拉普拉斯混合核函数的 MKELM (G+L) 具有最高的 F1 分数、最低的虚警率与较低的漏报率, 分别为 81.40%、15.50% 和 21.10%。该方法相比于仅使用单个高斯核函数或拉普拉斯核函数的 KELM (G)、KELM (L), F1 分数分别提高了 1.54%、1.66%, 比未使用核函数的 ELM 提升了 22.90%, 证明了利用核函数将数据映射到高维空间后进行分类的有效性。同时, 使用了混合核函数的 MKELM (G+G)、MKELM (G+L) 及 MKELM (L+L) 的 F1 分数均高于使用单个核函数的 KELM (G) 和 KELM (L), 主要是由于混合核函数相比于单个核函数的映射空间更为丰富, 提高了对数据的分类能力。此外, MKELM (G+L) 的 F1 分数分别高于 SVM 及 RF 2.00% 和 2.92%, 说明了该方法相比于主流机器学习模型识别效果也更好。

表 1 八种方法识别结果对比

Table 1 Comparison of identification results of eight methods

Method	$S_{F1}/\%$	$R_{FA}/\%$	$R_{MA}/\%$
SVM	79.40 \pm 2.24	17.87 \pm 4.76	22.44 \pm 7.98
RF	78.48 \pm 3.83	15.88 \pm 7.59	25.94 \pm 5.45
ELM	58.50 \pm 4.82	40.00 \pm 11.07	41.44 \pm 4.58
KELM (G)	79.86 \pm 4.73	20.00 \pm 8.38	19.53 \pm 7.15
KELM (L)	79.74 \pm 4.53	19.78 \pm 8.33	20.01 \pm 6.81
MKELM (G+G)	81.10 \pm 1.76	16.48 \pm 4.92	20.78 \pm 4.85
MKELM (L+L)	80.59 \pm 3.35	15.86 \pm 6.77	22.20 \pm 5.46
MKELM (G+L)	81.40 \pm 2.56	15.50 \pm 5.22	21.10 \pm 5.14

图 5 为使用不同混合核函数的 MKELM (G+G)、MKELM (L+L) 及 MKELM (G+L) 的决策区域图。由图中可看出, 这三种方法识别效果都比较好, 各颜色区域内大部分为相同颜色的点, 三者之间预测结果较为接近。但相比而言, MKELM (G+L) 的决策区域图 5 (c) 相比于图 5 (a) 和图 5 (b), 预测范围更加紧凑, 降低了虚警的风险。同时 MKELM (G+L) 实验结果的 F1 分数分别高于 MKELM (G+G)、MKELM (L+L) 约

0.30%、0.81%，主要是由于不用种类核函数进行组合时综合了各个核函数的优点。以上表明，使用了高斯与拉普拉斯混合核函数的MKELM (G+L) 预测精度高、误差小，能够广泛应用于实际道路高排放源识别任务中。

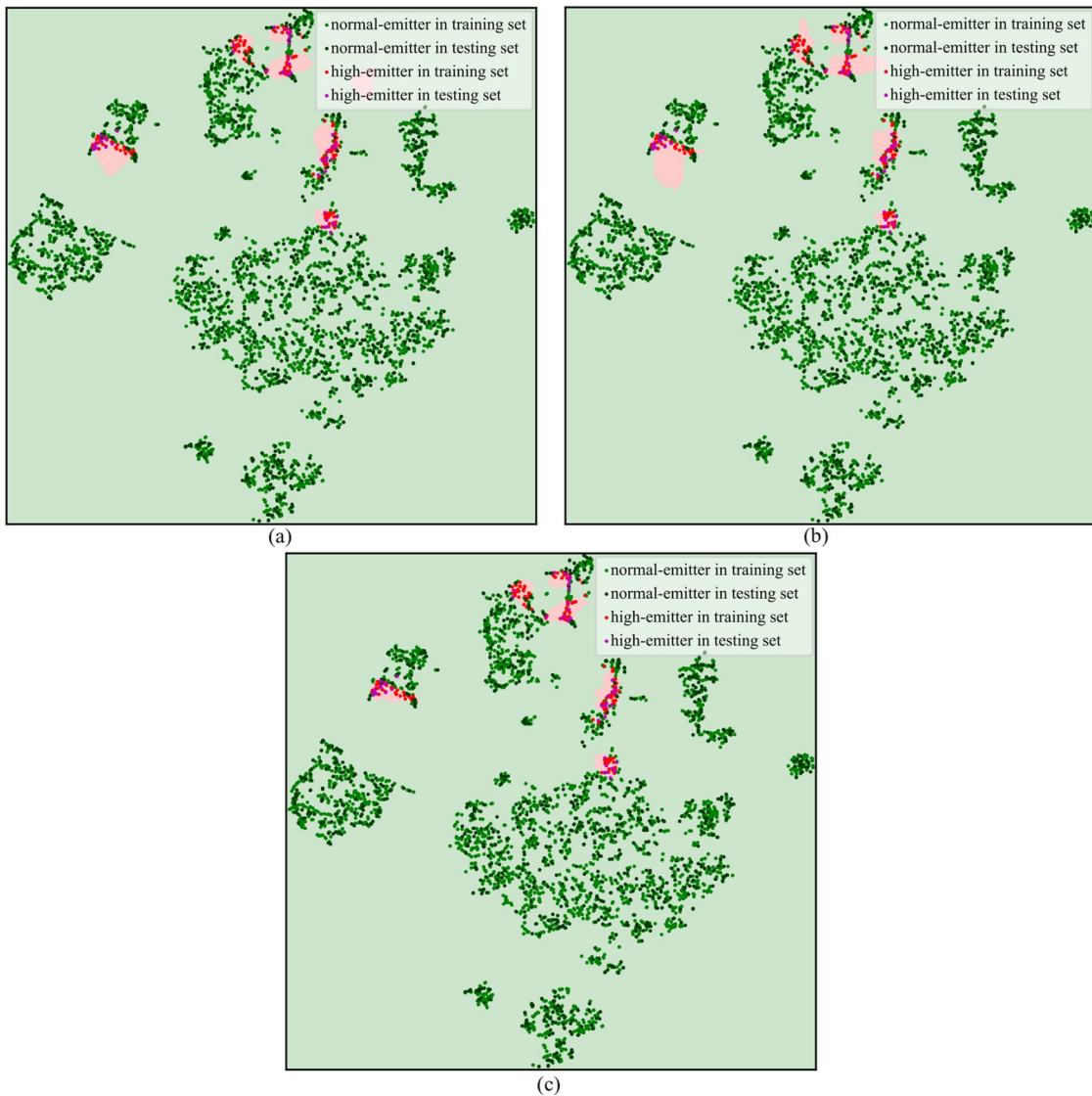


图5 使用了三种不同混合核函数的MKELM决策区域图。(a) 高斯与高斯混合核函数; (b) 拉普拉斯与拉普拉斯混合核函数; (c) 高斯与拉普拉斯混合核函数

Fig. 5 Diagram of MKELM decision region with four different mixed kernel functions. (a) Gaussian-Gaussian mixed kernel function; (b) Laplacian-Laplacian mixed kernel function; (c) Gaussian-Laplacian mixed kernel function

4 结 论

以合肥市蜀山区仰桥路道路遥感监测设备2017年5月采集的遥测数据为例，利用数据中CO、NO、CO₂、HC浓度及VSP、速度、加速度、风速、风向、温度、车牌颜色共11维特征构建了MKELM模型，对道路高排放源进行识别。实验结果表明，使用高斯与拉普拉斯混合核函数的MKELM(G+L)的F1分数达到81.40%，具有最优的识别能力。在实验中通过将MKELM (G+L)与ELM、KELM进行对比，证明了使用高斯核函数或拉

普拉斯核函数的有效性, 特别是使用了二者的混合核函数, 识别性能进一步提升。并且与 ELM、SVM、RF 的对比实验结果可以看出, MKELM (G+L) 优于这三种主流机器学习模型, 证明了所提出方法在高排放识别任务上的优越性。因此, 本研究有助于实现对高排放移动源的有效识别, 提升城市空气质量, 可为人工智能在大气环境监测领域上的应用提供一定的参考。

参考文献:

- [1] Fan W B, Chen J H, Qian J, *et al.* Effects of vehicle emissions on human's health [J]. *Chinese Journal of Environmental Management*, 2016, 8(1): 110-113.
范武波, 陈军辉, 钱 骏, 等. 机动车尾气对人体健康的危害 [J]. 中国环境管理, 2016, 8(1): 110-113.
- [2] Li Z R. *Study of Key Problems for Capturing High-Emitters in Road Network Remote Sensing System* [D]. Hefei: University of Science and Technology of China, 2020.
李泽瑞. 面向高排放源捕获的城市路网遥感监测系统关键问题研究 [D]. 合肥: 中国科学技术大学, 2020.
- [3] Bishop G A, Stedman D H, Peterson J E, *et al.* A cost-effectiveness study of carbon monoxide emissions reduction utilizing remote sensing [J]. *Air & Waste*, 1993, 43(7): 978-988.
- [4] Pujadas M, Domínguez-SáEZ A, De la Fuente J. Real-driving emissions of circulating Spanish car fleet in 2015 using RSD Technology [J]. *Science of the Total Environment*, 2017, 576: 193-209.
- [5] Huang Y H, Lee C K C, Yam Y S, *et al.* Rapid detection of high-emitting vehicles by on-road remote sensing technology improves urban air quality [J]. *Science Advances*, 2022, 8(5): eabl7575.
- [6] Guo H F, Zeng J, Hu Y M. Neural network modeling of vehicle gross emitter prediction based on remote sensing data [C]. *2006 IEEE International Conference on Networking, Sensing and Control*. April 23-25, 2006, Ft. Lauderdale, FL, USA. IEEE, 2006: 943-946.
- [7] Zeng J, Guo H F, Hu Y M. A PKGV-ANN model for vehicle high emitters identification based on remote sensing data [C]. *Proceedings of the 27th Chinese Control Conference*, Kunming, 2008: 171-175.
曾 君, 郭华芳, 胡跃明. 基于遥感监测的 PKGV-ANN 高排污车辆鉴别模型研究 [C]. 第二十七届中国控制会议论文集, 昆明, 2008: 171-175.
- [8] Li Z R, Kang Y, Lv W J, *et al.* High-emitter identification model establishment using weighted extreme learning machine and active sampling [J]. *Neurocomputing*, 2021, 441: 79-91.
- [9] Xu Z Y, Wang R J, Zhang C, *et al.* Automatic identification of high-emitting vehicle based on deep feature clustering [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2021, 21(6): 298-309.
许镇义, 王仁军, 张 聪, 等. 基于深度特征聚类的高排放移动污染源自动识别 [J]. 交通运输系统工程与信息, 2021, 21(6): 298-309.
- [10] Xu Z Y, Wang R B, Kang Y, *et al.* Analysis and prediction of main influencing factors in mobile source remote sensing [J]. *Journal of Atmospheric and Environmental Optics*, 2022, 17(2): 220-229.
许镇义, 王瑞宾, 康 宇, 等. 移动源排放遥测主要影响因素分析及预测 [J]. 大气与环境光学学报, 2022, 17(2): 220-229.
- [11] Huang G, Zhu Q, Siew C. Extreme learning machine: Theory and applications [J]. *Neurocomputing*, 2006, 70(1-3): 489-501.
- [12] Huang H, Lan H Y, Huang Y B. A detection method of SO₂ concentration based on DBN and ELM [J]. *Journal of Atmospheric and Environmental Optics*, 2020, 15(3): 207-216.
黄 鸿, 兰洪勇, 黄云彪. 基于深度信念网络和极限学习机的 SO₂ 浓度检测 [J]. 大气与环境光学学报, 2020, 15(3): 207-216.

- [13] Wang T D, Liu W Q, Zhang Y J, *et al.* Research of on-road vehicle emissions remote sensing system [J]. *Journal of Atmospheric and Environmental Optics*, 2007, 2(3): 198-202.
王铁栋, 刘文清, 张玉钧, 等. 机动车尾气遥测技术和应用研究 [J]. 大气与环境光学学报, 2007, 2(3): 198-202.
- [14] Huang G B, Zhou H M, Ding X J, *et al.* Extreme learning machine for regression and multiclass classification [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2012, 42(2): 513-529.
- [15] Zhang W J, Han D Z, Li K C, *et al.* Wireless sensor network intrusion detection system based on MK-ELM [J]. *Soft Computing*, 2020, 24(16): 12361-12374.
- [16] Wu D H, Wang Z L, Chen Y, *et al.* Mixed-kernel based weighted extreme learning machine for inertial sensor based human activity recognition with imbalanced dataset [J]. *Neurocomputing*, 2016, 190: 35-49.
- [17] Van der Maaten L, Hinton G. Visualizing data using t-SNE [J]. *Journal of Machine Learning Research*, 2008, 9(86): 2579-2605.
- [18] Xu X B, Zhao Y Q, Lu L B, *et al.* Research on vegetable and fruit image classification method based on LBP particle swarm optimization mixed kernel function ELM [J]. *Machine Design & Research*, 2021, 37(4): 15-20.
许学斌, 赵雨晴, 路龙宾, 等. 基于LBP的粒子群优化混合核函数ELM的蔬菜水果图像分类方法 [J]. 机械设计与研究, 2021, 37(4): 15-20.
- [19] He N H, Ding P, Ma F Y, *et al.* Research on transformer fault diagnosis method based on optimized extreme learning machine [J]. *Chinese Journal of Electron Devices*, 2021, 44(4): 897-902.
何宁辉, 丁培, 马飞越, 等. 基于优化极限学习机的变压器故障诊断方法研究 [J]. 电子器件, 2021, 44(4): 897-902.