

DOI:10.13382/j.jemi.B2508326

嵌入式人工智能甲醛传感器抗交叉干扰研究*

赵纯源

(中国科学院城市环境研究所区域与城市生态安全全国重点实验室 厦门 361012)

摘要:针对甲醛电化学传感器在室内环境监测过程中易受乙醇气体交叉干扰的问题,本研究以嵌入式人工智能芯片 RV1126 作为运算平台上,以电化学甲醛传感器和乙醇传感器作为感知器件获取环境要素输入研制甲醛检测仪器硬件系统,同时结合嵌入式人工智能技术,在设备端以算法补偿的方式来提高甲醛检测仪抗乙醇交叉干扰的能力。本研究以网格采样法获取数据集,并以此数据为基础,分别以线性方程回归法、常规机器学习回归法、神经网络回归法搭建回归模型。通过对各种模型预测的均方根误差进行对比实验,发现线性方程回归法搭建的模型误差最大约 $600 \mu\text{g}/\text{m}^3$,以常规机器学习回归法搭建的模型可以达到约 $100 \mu\text{g}/\text{m}^3$,而神经网络回归法搭建的模型预测精度最高,其预测误差均方差可以达到 $30 \mu\text{g}/\text{m}^3$ 。而根据世界卫生组织的《室内空气质量指南》室内空气中甲醛的长期平均浓度不应超过 $80 \mu\text{g}/\text{m}^3$,本研究研制的甲醛检测仪及以神经网络回归法搭建的模型能够在有一定浓度的乙醇气体交叉干扰的环境下实现对甲醛污染的检测要求。

关键词: 气体传感器;嵌入式人工智能;交叉干扰;甲醛传感器

中图分类号: TH89;TN98

文献标识码: A

国家标准学科分类代码: 520.2

Research on cross-interference resistance for embedded artificial intelligence formaldehyde sensors

Zhao Chunyuan

(The Institute of Urban Environment Chinese Academy of Sciences, National Key Laboratory of Regional and Urban Ecological Security, Xiamen 361012, China)

Abstract: To address the issue of cross-interference from ethanol gas in electrochemical formaldehyde sensors during indoor environmental monitoring, this study developed a formaldehyde detection system using the embedded AI chip RV1126 as the computing platform, with electrochemical formaldehyde and ethanol sensors as sensing components for environmental data acquisition. At the same time, combined with embedded artificial intelligence technology, algorithm compensation is used at the device end to improve the formaldehyde detection instrument's ability to resist ethanol cross-interference. This study obtained a dataset using grid sampling method, and based on this data, regression models were constructed using linear equation regression, conventional machine learning regression methods, and neural network regression methods. Comparative experiments on root-mean-square error (RMSE) revealed that the linear regression model exhibited the highest prediction error ($\approx 600 \mu\text{g}/\text{m}^3$), conventional machine learning models achieved $\approx 100 \mu\text{g}/\text{m}^3$ error, while the neural network regression model demonstrated superior accuracy with an MSE of $30 \mu\text{g}/\text{m}^3$. According to the World Health Organization's Indoor Air Quality Guidelines, which recommend a long-term average formaldehyde concentration threshold of $80 \mu\text{g}/\text{m}^3$, the proposed formaldehyde detector, combined with the neural network model, effectively fulfills detection requirements in environments with ethanol cross-interference, ensuring reliable formaldehyde pollution monitoring.

Keywords: gas sensor; embedded artificial intelligence; cross-interference; formaldehyde sensor

0 引言

随着社会的发展,各种人造建筑材料、二次加工的复合材料等用于室内建设,然而材料中所采用的粘接剂、添加剂等常释放出甲醛^[1]等有害物质。有研究表明,居住在甲醛超标的室内环境将提高患癌风险^[2]。为了治理室内甲醛污染首先需要对甲醛浓度进行监测,而目前便携式甲醛检测设备常采用电化学传感器作为其感知器件^[3-4]。电化学传感器虽具灵敏度高、体积小巧,但易受其他气体的交叉干扰^[5-6],如在室内环境中,食物发酵或酒精制品挥发等产生乙醇蒸汽常对甲醛检测产生干扰。

为了准确地对室内甲醛浓度监测,需要解决传感器交叉干扰的问题。目前传感器抗交叉干扰的方法常见有物理改进法、电路改进法、算法补偿法 3 种,物理改进法如增加前置过滤器滤出干扰气体以避免干扰气体对传感器的交叉干扰;电路改进法如增加补偿电极,以工作电极获取总信号,以补偿电极获取干扰信号,通过补偿电极信号对工作电极信号进行补偿,从而提高传感器的抗交叉干扰能力;算法补偿法,主要是同时采集多种要素传感器输出然后进行算法补偿^[7],实现检测仪器抗交叉干扰的能力。物理改进法和电路改进法方法需要从传感器结构上进行改进,适合传感器生产厂商进行抗交叉干扰能力的提高;算法改进法适用于监测仪器制造厂家。

在气体传感器抗交叉干扰算法方面,目前常有构建函数法^[8]、线性变换法^[7]、人工智能模型法^[9],构建函数法将真实环境值与各传感器等效某一函数关系,该方法对设备算力要求较小,适合于低成本传感器,但是需要人为设定函数关系,当监测多元环境变量时,将难以实现。线性变换法可以理解为构建函数法中的一种特例,将传感器设定在线性工作区间,以线性变换求解模型参数。该方法算法简单,易理解,但是要求各个传感器都工作在线性区间,导致传感器的量程较窄,且由于忽略了非线性变化,从而引入了拟合误差。

随着嵌入式技术的发展,X86 平台上训练的人工智能技术模型可以在嵌入式设备上运行,本研究采用人工智能算法,以嵌入式人工智能芯片为检测设备运算芯片,实现甲醛检测设备的研制以及抗交叉干扰算法研究。

1 监测仪器设计与制作

1.1 嵌入式人工智能

目前微型环境监测设备主要采用嵌入式芯片作为中央处理器。而嵌入式芯片上实现人工智能算法主要有 3 种^[10],一种将算法模型部署在服务器,嵌入式系统通过服务器提供的 API 接口进行数据交换^[11],从而实现人工

智能运算的功能,但是该方法要求嵌入式设备能够联网,且不适用于时效性要求较高的领域;第二种是直接嵌入在嵌入式设备上设计人工智能模型,如在嵌入式 linux 系统基础上结合 OpenCV 人工智能运算库^[12-13],实现简单的机器学习模型训练和推理,该方法直接采用 C 或 C++ 语言编写模型,其优点模型易于与硬件设备融合,但是难以调用最新的基于 python 代码编写的人工智能算法;第三种是在 X86 平台上训练人工智能模型,然后将模型通过芯片厂家提供的工具进行裁剪转换并下载到嵌入式设备上^[14],该方法的优点是可以借用目前先进如 tensorflow, pytorch 等的算法框架,和强大运算力的 GPU 实现人工智能模型训练,同时又不占用宝贵的嵌入式硬件资源。如目前瑞芯微厂家推出的 RV1126 芯片以及意法半导体推出 STM32F429 芯片采用第三种方案实现人工智能模型在嵌入式设备上的部署。

1.2 甲醛检测仪

本研究以瑞芯微公司的 RV1126 作为嵌入式人工智能芯片,该芯片采用 Cortex-A7 架构的 CPU 同时集成有 NPU 神经网络处理单元,通过嵌入式 Linux 系统实现各个外设驱动。在环境感知方面,本研究以温湿、甲醛浓度、乙醇浓度作为环境感知要素。其中甲醛浓度的监测采用电化学传感器 ZE08K-CH2O,如图 1(a)所示,该传感器采用串口输出浓度值,量程为 0~5 mg/m³,乙醇浓度监测采用电化学乙醇传感器 ZE30-C2H5OH,如图 1(b)所示,该传感器也是采用串口输出浓度值,量程也是 0~5 mg/m³,甲醛和乙醇传感器出厂前都分别做了标定。



(a) 乙醇传感器
(a) Ethanol sensor

(b) 甲醛传感器
(b) Formaldehyde sensor

图 1 气体传感器

Fig. 1 Gas sensor

各个传感器的信号流程图如图 2 所示,甲醛传感器、乙醇传感器、温湿度传感器分别采集环境要素值后,经过数据预处理进行去噪后作为输入变量传输给人工智能模型,经过模型进行矫正运算后输出甲醛真实浓度值。

甲醛检测仪的气路如图 3 所示,被测气体分别经过甲醛传感器和乙醇传感器进行数据采集后,汇集到温湿度传感器,最后由排气口排出到仪器外。

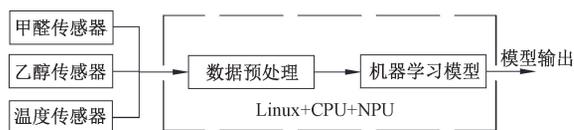


图 2 甲醛检测仪信号流程图

Fig. 2 Formaldehyde detector signal flow diagram

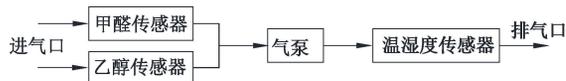


图 3 甲醛检测仪气路图

Fig. 3 Formaldehyde detector gas flow diagram

甲醛监测仪实物如图 4 所示。从内部电路板可以看出,本气体检测仪器采用国产 ARM 芯片瑞芯微 RV1126 作为人工智能运算平台,被测气体分别经过电化学甲醛传感器和电化学乙醇传感器后由气泵通过温湿度传感器排除。各个传感器采集的环境数据经过 RV1126 芯片运算后输出真实的环境数据。

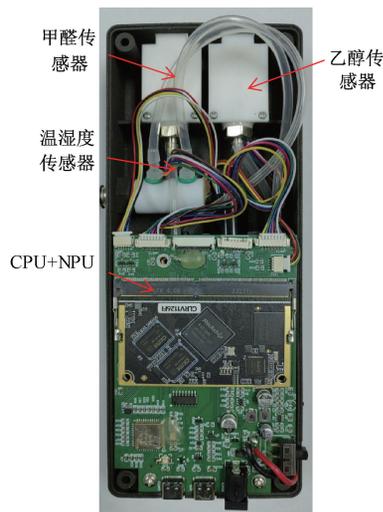


图 4 甲醛检测仪

Fig. 4 Formaldehyde detector

2 传感器数据采集

2.1 数据采集环境搭建

传感器数据采集系统结构如图 5 所示,通过气体调配器对甲醛标气、乙醇标气及高纯氮气进行浓度调配,输出流速恒定且气体浓度可调的混合气体,并通过气体混合预热器充分混合和温度控制后送入甲醛检测仪。环境检测仪和气体混合预热器共同置于温度控制舱,确保数据采集过程中的温度恒定。通过标定控制电脑现对气体调配器、温度控制等统一管理,形成各种真实环境,并同时读取甲醛检测仪的输出,实现原始数据的获取。

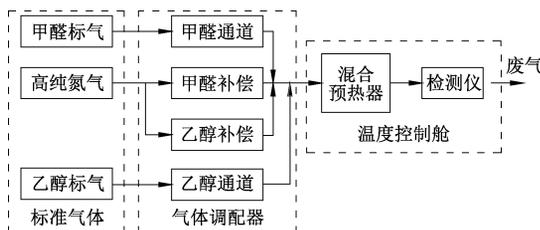


图 5 数据采集系统气路流程图

Fig. 5 Data acquisition system gas flow diagram

2.2 数据采集

为了便于研究过程的实现,本研究在标准温度 25 °C 环境,并在补偿气体的配合下,使混合气体以 400 mL/min 恒定流速输送到气体监测仪,具体控制方法如式(1)所示。

$$a_{\text{甲}} + a_{\text{甲补}} + a_{\text{乙}} + a_{\text{乙补}} = 400 \text{ mL/min} \quad (1)$$

其中, $a_{\text{甲}}$ 表示甲醛气路体积流速, $a_{\text{甲补}}$ 表示甲醛气路的氮气补偿体积流速, $a_{\text{乙}}$ 表示乙醇气路体积流速, $a_{\text{乙补}}$ 表示乙醇气路的氮气补偿体积流速。通过控制补偿通道的氮气流速,确保总输出流速的恒定。混合气体的浓度调制采用图 6 所示的网格化控制法进行调配,分别控制甲醛和乙醇的体积流量,使混合气体中乙醇和甲醛分别从 0 mg/m³ 以 1 mg/m³ 递增至 5 mg/m³,并在每个网格点分别采集甲醛传感器和乙醇传感器的输出。

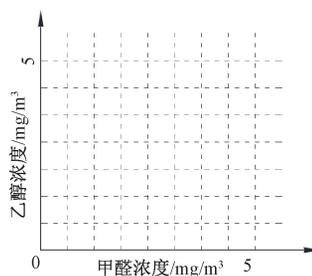


图 6 混合气体网格化调配

Fig. 6 Grid allocation of mixed gases

采样数据共输出数据 36 组如表 1 所示,从表可以看出,虽然传感器出厂前已做标定,但是在标气下,其输出值与标气真实浓度具有明显差距,如当只通入 5 000 μg/m³ 浓度的乙醇时,乙醇传感器输出浓度为 824 μg/m³,同时,甲醛传感器显示有 512 μg/m³ 浓度输出。两种传感器的输出与实际真实环境都有差距,因此,无论是从单一气体的准确性还是传感器的交叉干扰方面,都需要对传感器进行二次建模标定。

将表 1 中的采集数据绘制三维散点图如图 7(a) 所示,从散点图可以看出,传感器的数据输出并不是平面分布,而是平缓变化的曲面分布。由于实验条件限制,采集数据量较少,难以支撑较复杂的神经网络等模型训练。

为了便于模型比较,本研究分别在水平和垂直方向,采用逐次线性插值法将采样数据进行加密扩充^[15],插值后的

数据集如图 7(b) 所示,如图插值后数据集仍保持均匀平滑的曲面分布。

表 1 采样数据

Table 1 Sampled data

序号	甲醛 传感器输出	乙醇 传感器输出	甲醛 真实浓度	乙醇 真实浓度	序号	甲醛 传感器输出	乙醇 传感器输出	甲醛 真实浓度	乙醇 真实浓度
1	12	48	0	0	19	1 491	1 006	3 000	0
2	10	161	0	999	20	1 429	1 291	3 000	1 001
3	113	311	0	2 002	21	1 448	1 587	3 003	2 001
4	243	477	0	3 000	22	1 498	1 931	3 000	2 999
5	369	640	0	4 000	23	1 565	2 288	3 001	3 999
6	512	824	0	5 000	24	1 630	2 659	3 001	5 000
7	586	310	1 001	0	25	1 860	1 634	3 999	0
8	565	457	1 001	995	26	1 786	1 916	4 000	1 008
9	621	599	1 002	1 986	27	1 786	2 210	4 000	2 003
10	718	760	1 001	3 000	28	1 816	2 562	3 998	3 000
11	823	1 063	1 002	4 002	29	1 856	2 922	4 000	3 998
12	932	1 423	999	5 000	30	1 891	3 300	4 002	5 000
13	1 060	596	2 001	0	31	2 162	2 288	5 000	0
14	1 009	736	2 001	999	32	2 076	2 567	5 000	996
15	1 036	959	2 001	1 999	33	2 050	2 864	5 000	2 003
16	1 106	1 300	1 998	3 000	34	2 050	3 218	5 000	2 999
17	1 192	1 652	1 999	3 999	35	2 058	3 586	5 000	4 001
18	1 281	2 016	1 999	5 000	36	2 061	3 952	5 000	5 000

2.3 数据插值误差分析

设 $x_0 < x < x_1 < x_2$, 则插值函数如式(2)所示, 依据 Aitken 逐次线性插值^[16]余项式, 如式(3)所示。

$$I(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0) \quad (2)$$

$$R(x) = \frac{I_{0,1}(x) - I_{0,2}(x)}{x_1 - x_2}(x - x_1) \quad (3)$$

其中, I 为 1 次拉个朗日插值公式, 如式(4)和式(5)所示。

$$I_{0,1}(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0) \quad (4)$$

$$I_{0,2}(x) = f(x_0) + \frac{f(x_2) - f(x_0)}{x_2 - x_0}(x - x_0) \quad (5)$$

而本采集样本为等距离网格化采样, 设距离为 h , 如式(6)所示。

$$h = x_{n+1} - x_n \quad (6)$$

则插值余项式可表示为式(7), 其中当 $x = \frac{1}{2}(x_0 + x_1)$, 式(7)取得极值, 如式(8)所示。

$$R(x) = \frac{-1}{4h}(2f(x_1) - f(x_0) - f(x_2))(x - x_0)(x - x_1) \quad (7)$$

$$R(x) < \left| \frac{1}{16}(2f(x_1) - f(x_0) - f(x_2)) \right| \quad (8)$$

基于式(8), 对整个采样网格插值误差取最大值为, 如式(9)所示。

$$R_{max} = \max(R_{ij}(x)) \quad (9)$$

经计算经计算甲醛最大误差为 $6.9 \mu\text{g}/\text{m}^3$, 乙醇最大插值误差为 $17.2 \mu\text{g}/\text{m}^3$, 各项误差都低于甲醛监测最低要求 $80 \mu\text{g}/\text{m}^3$, 所以, 本实验所采用的采集密度及插值方法满足模型训练的误差要求。

3 算法研究

3.1 实验方法和步骤

首先以量程点采样数据构建多元线性回归模型, 并以图 7(a) 中的原始采样数据对模型进行预测测试。

然后将图 7(b) 中插值后的采样数据以 4 : 1 进行随机抽取, 分别构建训练数据集和测试数据集。以训练集数据对常规机器学习回归、神经网络回归等人工智能模型进行训练。

最后以均方根误差 RMSE 作为模型误差评估指标对各个模型进行预测精度对比, RMSE 的算法如式(10)所示。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}^{(i)} - y^i)^2} \quad (10)$$

1) 线性方程回归建模实验

线性函数法和线性变化法都属于线性回归建模法。

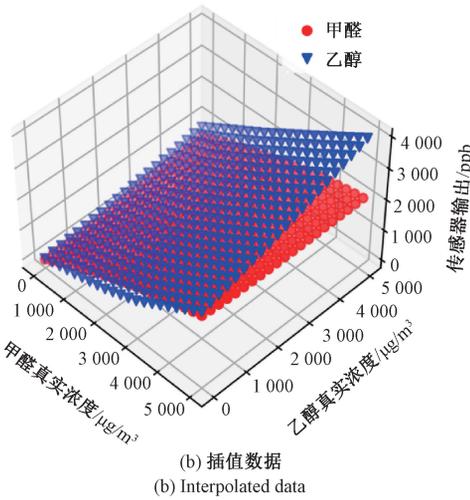
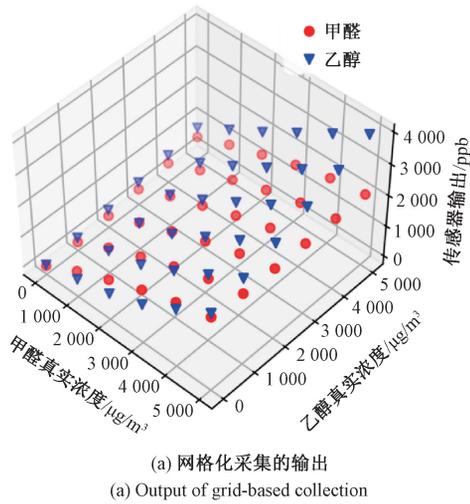


图 7 数据集构建

Fig. 7 Data set construction

本实验以市面上常见的线性传感器标定方法^[17],即选取传感器量程点所采集数据作为标定数据,并以此构建线性回归模型进行传感器标定。由于本实验在温度、流速等其他条件固定的前提下研究甲醛和乙醇两种气体污染,所以可以构建二元输入变量的线性平面空间,其函数模型可以设置为式(11)所示。

$$y_{\text{甲}} = ax_{\text{甲}} + bx_{\text{乙}} + c \quad (11)$$

其中, $y_{\text{甲}}$ 是甲醛体积浓度, $x_{\text{甲}}$ 、 $x_{\text{乙}}$ 分别是甲醛和乙醇传感器的输出,分别在量程点选取三组($y_{\text{甲}}$, $x_{\text{甲}}$, $x_{\text{乙}}$)并建立 3 个等式,即可求解 3 个参数 a 、 b 、 c ,如式(12)所示。

$$\begin{pmatrix} y_{\text{甲}1} \\ y_{\text{甲}2} \\ y_{\text{甲}3} \end{pmatrix} = \begin{pmatrix} x_{\text{甲}1} & x_{\text{乙}1} & 1 \\ x_{\text{甲}2} & x_{\text{乙}2} & 1 \\ x_{\text{甲}3} & x_{\text{乙}3} & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad (12)$$

本实验分别取 3 个量程点 (0, 12, 48), (0, 512, 824), (5 000, 2 061, 3 952), 构建 3 个等式如式(13)所示。

$$\begin{cases} 12a + 48b + c = 0 \\ 512a + 824b + c = 0 \\ 2\,061a + 3\,952b + c = 5\,000 \end{cases} \quad (13)$$

对式(5)联立求解后得出结果如式(14)所示。

$$\begin{cases} a = -10.72 \\ b = 6.91 \\ c = -202.89 \end{cases} \quad (14)$$

得出回归模型如式(15)所示。

$$y_{\text{甲}} = -10.72x_{\text{甲}} + 6.91x_{\text{乙}} - 202.89 \quad (15)$$

2) 常规机器学习回归实验

机器学习技术中的回归方法较多,如 Logistic 回归、K 近邻回归、决策树回归、提升树回归等^[18]。本研究选取决策树^[19]、K 近邻^[20]、随机森林^[21]、支持向量^[22] 4 种常规回归进行预测分布和均方差 RMES 分析。各项机器学习的主要超参数如表 2 所示。

表 2 超参数

Table 2 Hyperparameters

序号	模型	超参数
1	SVR	核函数 kernel:liner; 阶数 degree=3, 核系数 gamma="scale", 独立项 coef0=0.0, 停止条件容差 tol=1e-3, 误差项惩罚参数 C=1.0, 惩罚范围 epsilon=0.1,
2	随机森林	树的数量 n_estimators=100, 评价标准 criterion="squared_error", 最大深度 max_depth=None, 最小样本拆分量 min_samples_split=2, 点最小样本数 min_samples_leaf=1, 点最小权重因子 in_weight_fraction_leaf=0.0, 最大特征数量 max_features=1.0, 最大叶节点数 max_leaf_nodes=None, 最小杂质减少量 min_impurity_decrease=0.0, 提升 bootstrap=True,
3	KNN	近邻数量:nNeighbors=5 权重函数 weights=uniform 算法类型:algorithm=auto 叶大小:leaf_size=30 距离度量:p=2,欧氏距离 度量标准 metric="minkowski", 度量参数 metric_params=None,
4	决策树	评价标准 criterion="squared_error", 分流器 splitter="best", 最大深度 max_depth=None, 拆分最小样本数 min_samples_split=2, 节点最小样本数 min_samples_leaf=1, 叶片最小重量分数 min_weight_fraction_leaf=0.0,

3) 基于神经网络回归实验

神经网络实验主要涉及到数据预处理、构建模型、模型嵌入式转化 3 个部分。模型下载到嵌入式 AI 芯片后,模型的网络结构将在芯片的 NPU 中运行,而数据预处理及模型预测后的逻辑推理等部分将在芯片的 CPU 中

进行。

数据预处理:本实验的数据预处理主要是数据归一化处理,由于甲醛和乙醇传感器的量程范围为 0~5 000 $\mu\text{g}/\text{m}^3$,因此,本实验采用将传感器数据除以 5 000,实现将模型输入数据压缩到 0~1 之间。

构建模型:本实验以 TensorFlow 为框架,综合考虑模型的深度及宽度,以及在训练过程中模型的收敛等因素,最终以 RELU 核函数搭建出神经网络,将模型在 X86 平台上训练并保存为 TFLITE 模型。采用 Netron 软件将 TFLITE 模型进行图形化输出,如图 8(a)所示,图中模型参数为浮点型,且模型的输入向量未初始化。

模型嵌入式转化:本实验在 Linux 系统下,采用瑞芯微公司提供的模型转化工具 RKNN-Toolkit,将 TFLITE 模型的网络结构转化为 RKNN 模型。通过 Netron 软件将 RKNN 模型进行图形化输出,如图 8(b)所示,图中模型参数变为整形,模型的输入向量已经初始化。在转化过程中主要涉及到模型量化,即将模型中的权重和激活值由浮点数转化为低比特的定点数据。本实验采用动态定点 16 位量化法来对模型参数进行量化,其量化函数如式(16)所示,其中 $quant$ 代表量化后的数, $float_num$ 代表浮点数, fl 是左移的位数。RKNN-Toolkit 以模型的输入样本作为量化数据集,结合量化函数,自动将模型的权重和激活值量化到 $[-32\ 767, 32\ 767]$ 之间。

$$quant = cast_to_i16(round(float_num \times 2^{fl})) \quad (16)$$

RKNN 模型下载到 RV1126 芯后将运行于 NPU 上,CPU 通过 API 接口函数对 RKNN 模型进行调用,实现模型的推理计算。

3.2 实验结果及数据分析

1) 预测结果

各个回归模型的预测分布如图 9 所示,图 9(a)~(f)分别为线性方程回归预测分布图、SVR 回归预测分布图、决策树回归预测分布图、KNN 回归预测分布图、随机森林回归预测分布图、神经网络回归预测分布图。

从图 9(a)可以看出线性方程回归预测点分布最离散,及预测误差最大,常规机器学习模型预测在整个量程区间可以保持相关性,但是相对于神经网络模型,其预测值分布较为离散。通过模型预测分布对比图可看出神经网络回归模型的预测一致性和准确性,相对于其他模型最高。

2) 预测误差分析

对线性方程回归、常规机器学习回归、神经网络模型预测均方误差 RMSE 进行对比如图 10 所示。

通过图 10 可以看出,线性方程回归法预测误差最高,达到 6 251,主要原因传感器的响应实际是非线性变化,而线性方程回归法直接取量程点构建线性模型,引入了拟合误差。常规机器学习法相对于线性方程回归法其

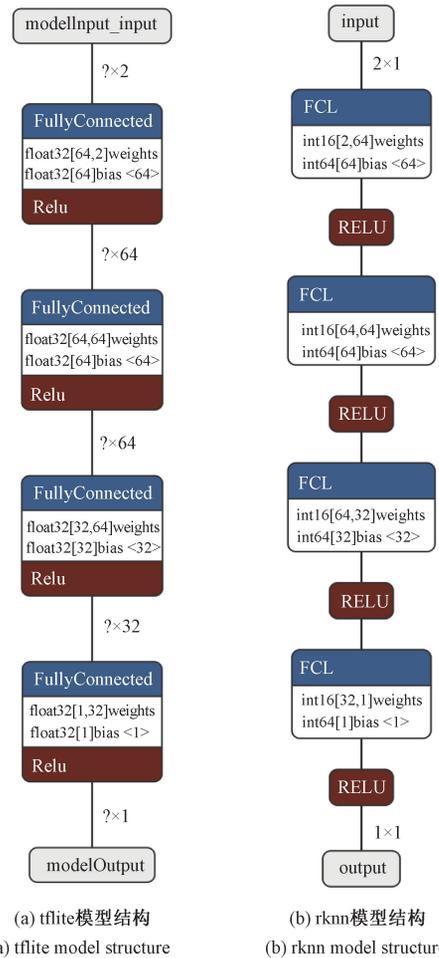


图 8 神经网络结构图(Netron)
Fig. 8 Neural network structure diagram (Netron)

误差大幅度地降低到 100~300 左右,其原因一方面模型的训练引入了更多的标定样本点,降低了样本采样的随机误差;另一方面是机器学习法可以处理非线性回归模型。神经网络法相对于机器学习法其误差更低,达到 36,其原因主要是神经网络法在处理多参数非线性模型上更加精细。

3) 回归模型嵌入式实现条件分析

各种回归模型嵌入式设备实现所需条件对比如表 3 所示。

表 3 模型嵌入式实现条件
Table 3 Conditions for embedded implementation of models

序号	模型名称	硬件	调用库
1	满量程线性回归	普通单片机	无
2	支持向量回归	ARM	openCv
3	决策树回归	ARM	openCv
4	随机深林回归	ARM	openCv
5	K 近邻回归	ARM	openCv
6	神经网络回归	ARM+NPU	rknn

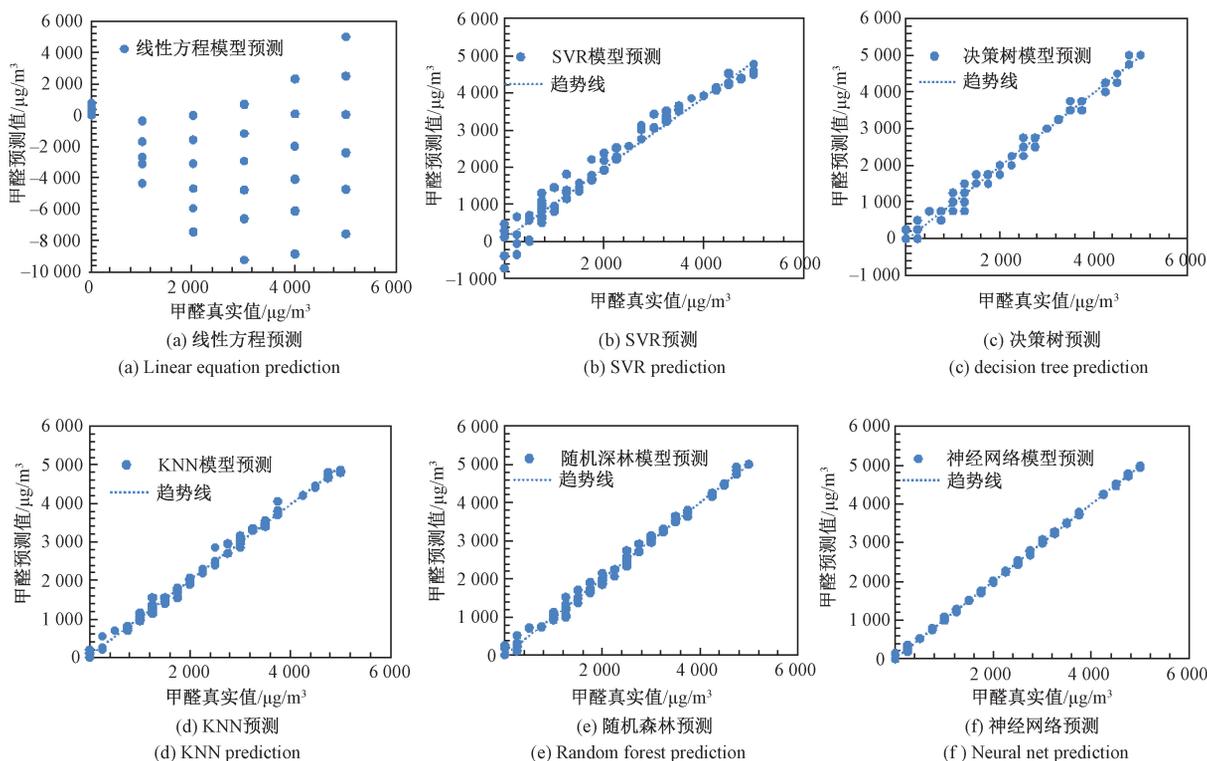


图 9 常规机器学习预测分布图

Fig. 9 The distribution of conventional machine learning prediction

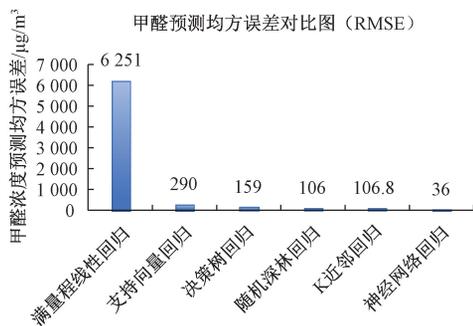


图 10 模型预测均方根误差对比图

Fig. 10 Comparison of mean root error of model prediction

从表 3 可以看出,线性方程回归对硬件要求最低,在低成本的普通单片机上即可实现回归模型运算;常规机器学习算法的嵌入式实现,主要以嵌入式 linux 系统环境采用 C/C++ 语言调用 OpenCV 的 ML 库来实现,因此需要采用 ARM 芯片;神经网络回归模型的嵌入式实现,需要 X86 平台上进行模型训练,以及模型转化工具包实现在嵌入式设备上的部署,因此需要专用的人工智能芯片实现模型的落地。

4 结论

针对便携式甲醛检测仪器的抗交叉干扰研究,物理

过滤法需要对干扰气体进行过滤,导致结构复杂,难以实现设备便携化改进,电路补偿法需要对传感器本身进行改动,而仪器开发者在于集成传感器,难以改变传感器本身。本研究以便携式气体检测仪常用的电化学传感器为基础,探索合适的算法以提高传感器的抗交叉干扰能力。

本研究在温度、流速其他环境要素恒定的前提下,网格化地采集甲醛乙醇二要素传感器输出,并构建数据集,并在此数据集上对线性方程回归、支持向量回归、决策树回归、随机森林回归、K 近邻回归、神经网络回归进行模型精度对比实验及嵌入式实现难度分析。综合模型误差实验及模型嵌入式平台部署的实现方式,本研究得出如下结论:线性回归模型实现算法简单,对硬件要求较低,可以在普通单片机上实现模型的运算,但是该模型忽略了传感器的非线性变化,模型误差较大,适合低成本低精度监测场合;支持向量回归、决策树回归等常规机器学习回归算法相对线性回归建模在解决非线性能力上有所提高,模型预测精度也大幅提高,在模型嵌入式上需要借助 openCV 等人工智能库,因此需要 ARM 芯片作为硬件平台,相对线性回归模型,常规机器学习模型部署成本将明显提高;神经网络回归在预测精度上相对常规机器学习,其解决多要素、非线性能力更强大、预测精度也更高,但是在嵌入式设备实现方面,需要专用的人工智能芯片,模型部署成本将进一步提高;三类模型模型各有优

缺点,分别适合不同的应用场景,若追求低成本实现方式可以选择线性方程法,如果追求高精度监测,可以选择神经网络法,如果选择折中方案,可以选择支持向量机回归、决策树回归等常规机器学习法。

本研究针对甲醛传感器的交叉干扰因素,构建多要素便携式甲醛检测仪。结合各传感器响应特性,通过网格化采样、逐次线性插值等方法,生成足够多的样本数据,并以此数据训出神经网络模型,最后通过对模型裁剪、量化等转化,使模型能够运行于嵌入式设备上,提高了便携式甲醛检测仪抗交叉干扰能力。本研究为小样本学习、神经网络模型嵌入式化、以及气体传感器抗交叉干扰算法提供了一种解决方案,可供其他研究者借鉴。在未来的研究中将探索更适合的插值方法、神经网络算法,以更少的标定样本训练出更加精确的算法模型,简化传感器的标定过程并提高传感器的检测精度。

参考文献

- [1] 洪淑娇. 民用建筑室内空气污染物来源及控制技术研究 [J]. 居舍, 2025(8):153-156.
HONG SH J. Study on the source and control technology of indoor air pollutants in civil buildings [J]. Dwellinghouse, 2025(8):153-156.
- [2] 潘扬松, 杞昊, 杜焯江, 等. 住宅室内甲醛污染健康危害及通风控制 [J]. 工业安全与环保, 2024, 50(11):78-82.
PAN Y S, QI H, DU ZH J, et al. Health hazards of indoor formaldehyde pollution and ventilation control in residential buildings [J]. Industrial Safety and Environmental Protection, 2024, 50(11):78-82.
- [3] CHATTOPADHYAY A, HUERTAS A, REBEIRO-HARGRAVE A, et al. Low-cost formaldehyde sensor evaluation and calibration in a controlled environment[J]. IEEE Sensors Journal, 2022, 22(12):11791-11802.
- [4] 葛亮, 徐一鹏, 肖小汀, 等. 燃气管网四氢噻吩浓度电化学在线检测仪设计[J]. 电子测量与仪器学报, 2025,39(7):247-258.
GE L, XU Y P, XIAO X T, et al. Design of gas pipeline tetrahydrothiophene concentration electrochemical online detector [J]. Journal of Electronic Measurement and Instrumentation, 2025,39(7):247-258.
- [5] 陈永冉. 煤矿气体检测设备交叉干扰及评判方法研究[J]. 工矿自动化, 2023,49(2):63-69,93.
CHEN Y R. Study on cross-interference and evaluation method of gas detection equipment in coal mine [J]. Journal of Mine Automation, 2023, 49(2):63-69,93.
- [6] 刘培源, 李剑, 夏春, 等. 多种气体对烟气分析仪中电化学传感器的交叉干扰研究 [J]. 计量科学与技术, 2023,67(7):62-67,52.
LIU P Y, LI J, XIA CH, et al. Study on cross interference of various gases to electrochemical sensors in flue gas analyzer[J]. Metrology Science and Technology, 2023, 67(7):62-67,52.
- [7] 曹利峰, 贾博文, 蔡宏忱, 等. 一种电化学传感器交叉干扰的补偿方法与实践 [J]. 中国环保产业, 2021(7):64-72.
CAO L F, JIA B W, CAI H CH, et al. A method and practice of compensation for cross interference of electrochemical sensors [J]. China Environmental Protection Industry, 2021(7):64-72.
- [8] 张国城, 赵红达, 赵俊秀, 等. 两种微型空气质量监测站气体交差干扰的数学模型 [J]. 计量技术, 2020(1):46-49.
ZHANG G CH, ZHAO H D, ZHAO J X, et al. Mathematical model of gas cross interference between two types of micro air quality monitoring stations [J]. Metrology Science and Technology, 2020(1):46-49.
- [9] ZHOU S, LIU N, SHEN C, et al. An adaptive Kalman filtering algorithm based on back-propagation (BP) neural network applied for simultaneously detection of exhaled CO and N₂O [J]. Spectrochim Acta Part A Molecular and Biomolecular Spectroscopy, 2019, 223:117332.
- [10] 陈华, 曲益明, 吴文豪, 等. 低功耗人工智能计算系统研究进展综述[J]. 功能材料与器件学报, 2024, 30(6):300-309.
CHEN H, QU Y M, WU W H, et al. Research progress review of low power artificial intelligence computing system[J]. Journal of Functional Materials and Devices, 2024, 30(6):300-309.
- [11] CALO S B, TOUNA M, VERMA D C, et al. Edge computing architecture for applying AI to IoT[C]. 2017 IEEE International Conference on Big Data (Big Data). IEEE, 2017:3012-3016.
- [12] WANG J B, LI Y, XIONG Y H, et al. Dress identification for camp security [C]. International Conference on Security with Intelligent Computing and Big-data Services. Cham: Springer International Publishing, 2018:661-672.
- [13] VIWATPINYO S, PHATCHUAY S. The automatic car to implementation of lane detective using raspberry Pi 3 model B on OpenCV[C]. 2022 International Conference on Cybernetics and Innovations (ICCI). IEEE, 2022: 1-5.
- [14] ZHU J, ZHOU Y, JIANG D, et al. Infrared image target detection system based on RV1126 [C]. Seventh Asia Pacific Conference on Optics Manufacture and 2021

- International Forum of Young Scientists on Advanced Optical Manufacturing (APCOM and YSAOM 2021). SPIE, 2022, 12166:1257-1261.
- [15] FLOATER M S, PATRIZI F. Transfinite mean value interpolation over polygons [J]. Numerical Algorithms, 2020, 85(3) :995-1003.
- [16] 许川佩, 江林, 黄喜军. 基于埃特金逐步插值滤波器的 TIADC 时间误差校准方法 [J]. 仪器仪表学报, 2023, 44(2) :93-100.
- XU CH P, JANG L, HUANG X J. A time error calibration method for TIADC based on Aitken step interpolation filter [J]. Chinese Journal of Scientific Instrument, 2023, 44(2) :93-100.
- [17] 冯大光, 冯思哲, 李兆星, 等. 基于多元线性回归模型的日光温室温度预测方法对比 [J]. 沈阳师范大学学报(自然科学版), 2025:43(1) ,75-81.
- FENG D G, FENG S ZH, LI ZH X, et al. Comparison of solar greenhouse temperature prediction methods based on multiple linear regression model [J]. Journal of Shenyang Normal University (Natural Science Edition), 2025:43(1) ,75-81.
- [18] HUANG J C, KO K M, SHU M H, et al. Application and comparison of several machine learning algorithms and their integration models in regression problems [J]. Neural Computing & Applications, 2020, 32 (10): 5461-5469.
- [19] CHEN J R, LIN Y H, LEU Y G. Predictive model based on decision tree combined multiple regressions [C]. 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). IEEE, 2017:1855-1858.
- [20] REN Z L, TANG Y Q, ZHANG W S. Quality-related fault diagnosis based on k-nearest neighbor rule for non-linear industrial processes [J]. International Journal of Distributed Sensor Networks, 2021, 17 (11): 12899-12912.
- [21] AO Y L, LI H Q, ZHU L P, et al. The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling [J]. Journal of Petroleum Science and Engineering, 2019, 174: 776-789.
- [22] KADYROVA N O, PAVLOVA L V. Comparative efficiency of algorithms based on support vector machines for regression [J]. Biofizika, 2015, 60(6) :1085-1098.

作者简介



赵纯源 (通信作者), 2006 年于四川理工学院获得学士学位, 2023 年于厦门理工学院获得硕士学位, 现为中国科学院城市环境研究所助理研究员, 主要研究方向为环境物联网与人工智能。

E-mail: zychina@126.com

Zhao Chunyuan (Corresponding author) received his B. Sc. degree in 2006 from Sichuan university of science & engineering, received his M. Sc. degree in 2023 from Xiamen university of science & engineering. Now he is an assistant researcher in the Institute of Urban Environment, Chinese Academy of Sciences. His main research interests include environmental internet of things and artificial intelligence.